Changepoint detection in a nonparametric time series regression model

Dissertation

zur Erlangung des Doktorgrades
an der Fakultät für Mathematik, Informatik und Naturwissenschaften
Fachbereich Mathematik
der Universität Hamburg

vorgelegt von

Maria Mohr
Hamburg, 2018

Folgende Gutachter empfehlen die Annahme der Dissertation:
Prof. Dr. Natalie Neumeyer
Prof. Dr. Claudia Kirch

Tag der Disputation: 18. Oktober 2018

Contents

Notations and conventions

- $\mathbb{N}$ denotes the set of all natural numbers and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$
- $\mathbb{Z}$ denotes the set of all integers and $\mathbb{R}$ denotes the set of all real numbers
- $\boldsymbol{x} \in \mathbb{R}^d$ for some $d \in \mathbb{N}$ has the representation $\boldsymbol{x} = (x_1, \ldots, x_d)^T$
- $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ for some $n \in \mathbb{N}$ denotes a d-variate random sample with $\boldsymbol{X}_i = (X_{i1}, \ldots, X_{id})^T$ for $i \in \{1, \ldots, n\}$
- $\boldsymbol{x} \leq \boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is short for $x_j \leq y_j$ for all $j = 1, \ldots, d$
- $\boldsymbol{x} \wedge \boldsymbol{y}$ for $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ is short for $(x_1 \wedge y_1, \ldots, x_d \wedge y_d)$, where $x \wedge y := \min(x, y)$ for $x, y \in \mathbb{R}$
- $\lfloor x \rfloor := \max\{j \in \mathbb{Z} : j \leq x\}$ for $x \in \mathbb{R}$
- for $\boldsymbol{k}, \boldsymbol{i} \in \mathbb{N}_0^d$ and $\boldsymbol{x} \in \mathbb{R}^d$ let
  - $|\boldsymbol{k}| := \sum_{j=1}^{d} k_j$ and $\boldsymbol{k}! := k_1 \cdots k_d$
  - $\boldsymbol{x}^{\boldsymbol{k}} := x_1^{k_1} \cdots x_d^{k_d}$
  - $\binom{\boldsymbol{k}}{\boldsymbol{i}} := \frac{\boldsymbol{k}!}{(\boldsymbol{k}-\boldsymbol{i})!\boldsymbol{i}!}$ for $\boldsymbol{i} \leq \boldsymbol{k}$
- for $g : \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{0} := (0, \ldots, 0) \in \mathbb{N}_0^d$, $\boldsymbol{k} \in \mathbb{N}_0^d \backslash \{\boldsymbol{0}\}$ and $\boldsymbol{x} \in \mathbb{R}^d$ let
  - $D^{\boldsymbol{0}} g(\boldsymbol{x}) := g(\boldsymbol{x})$
  - $D^{\boldsymbol{k}} g(\boldsymbol{x}) := \frac{\partial^{|\boldsymbol{k}|} g}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}(\boldsymbol{x})$
  - $\int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} g(\boldsymbol{u}) d\boldsymbol{u} := \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} g(u_1, \ldots, u_d) du_1 \ldots du_d$
  - $\int g(\boldsymbol{u}) d\boldsymbol{u} := \int_{\mathbb{R}^d} g(\boldsymbol{u}) d\boldsymbol{u}$
- for $g : \mathbb{R} \to \mathbb{R}$ let $\int g(u) du := \int_{-\infty}^{\infty} g(u) du$
- i.i.d. is short for independent and identically distributed
- a.s. is short for almost surely
- $\sigma(Z)$ denotes the σ-algebra generated by the random variable Z
- $\mathcal{N}(\mu, \sigma^2)$ is a normal distribution with mean $\mu \in \mathbb{R}$ and variance $\sigma^2 \in \mathbb{R}$ with $\sigma^2 > 0$
- $\mathcal{N}_K(\boldsymbol{\mu}, \Sigma)$ is a K-dimensional normal distribution with mean vector $\boldsymbol{\mu} \in \mathbb{R}^K$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{K \times K}$
- $\stackrel{\mathcal{D}}{=}$ denotes equality in distribution
- $\stackrel{\mathcal{D}}{\to}$ denotes convergence in distribution
- $\stackrel{P}{\to}$ denotes convergence in probability
- $\rightsquigarrow$ denotes weak convergence

Introduction

When looking at time ordered statistical data - arising in such diverse fields as epidemiology, geology or econometrics - changepoint analysis can help to detect if and when changes have occurred. In the context of regression analysis, detecting and estimating changes of the relationship between regressor X_t and response variable Y_t for a given data set $\{(Y_t, X_t) : t = 1, \ldots, n\}$ of size $n \in \mathbb{N}$ may be a matter of particular interest. If a change exists, a single regression model will fit the data rather poorly. Figure 1 shows the scatter plot of simulated data points that follow a linear regression with different coefficients for the first and second half.[1] Clearly, fitting a linear regression model to the full data set will lead to misspecification.

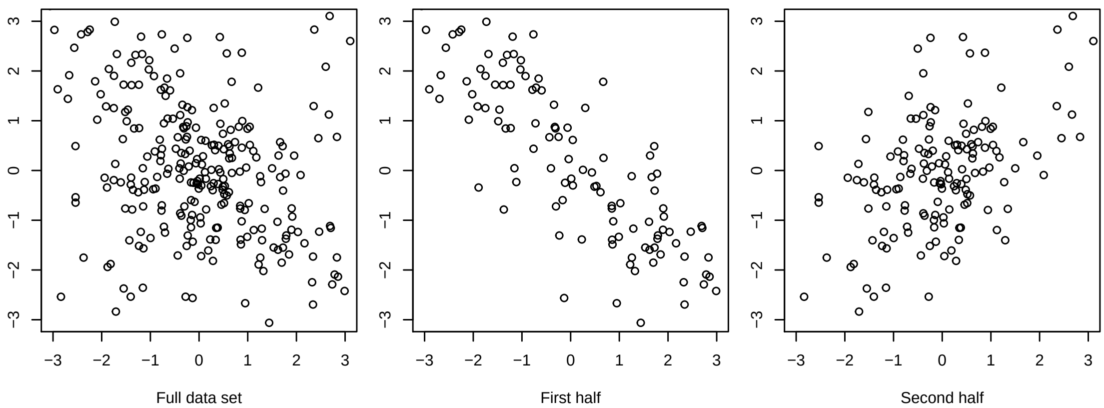

Figure 1: Scatter plots of simulated data with changepoint

These problems first arose in the 1950s in the context of quality control where changepoint analysis was used to analyze a machine's output in production for structural instability over time. Classical changepoint analysis is further used in the fields of epidemiology, biology and medicine to just mention a few (see [12] for more precise examples). Despite the fact that changepoint analysis has received much attention for decades, it is still the subject of current research. Classical methods and models have been extended in several directions. It was only a matter of time until the problem naturally moved in the time series context, such as autoregressive models that cover various more application fields. For instance economic and financial time series data is frequently affected by political and social events and therefore liable

[1] A sample of 300 data points was simulated according to the autoregression model from Subsection 6.1.3 on page 88 with break size $\Delta_0 = 1.4$.

to structural breaks. Furthermore, in the field of climate control, climate data is monitored and studied with regard to structural instability over time.

Figure 2 shows two data sets that have been extensively investigated in the changepoint framework. The left one shows the weekly differences of log-returns of the Dow Jones Industrial Average (DJIA) index between July 1st 1971 and August 2nd 1974. The right plot shows the annually flow of the river Nile in the city of Aswan, recorded between 1871 and 1970. Regarding the DJIA time series, there are several results that indicate a change in variance in the beginning of 1973. One possible explanation is the discovery of the Watergate scandal that took place at that time. Concerning the Nile data set results indicate that there is a structural change in the mean in 1898. In fact, in this year the first damn was built in Aswan.

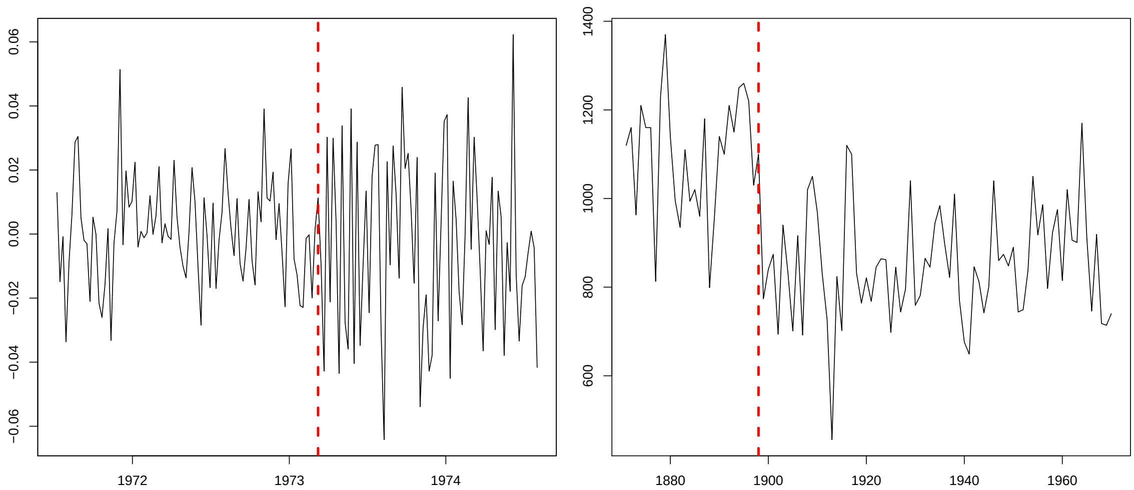

Figure 2: DJIA (left) and Nile (right) data with potential changepoints

This thesis makes a contribution to the field of changepoint analysis in nonparametric time series regression models that allow for heteroscedasticity. It focuses on the detection of possible changes rather than on their estimation. It is structured as follows. Chapter 1 gives a brief overview on relevant literature concerning changepoint analysis in regression models. The model under consideration is introduced and some basic definitions are given. In Chapter 2, the kernel estimators are presented and uniform rates of convergence are proven. Chapter 3 is the main part of this thesis. It contains the construction of a test to detect changes in the conditional mean function, which is based on the sequential marked empirical process of residuals. The limiting distribution under the null hypothesis of no changepoint and a stationarity assumption is proven, as well as a consistency result under a simple fixed alternative is given. Finally, to emphasize the advantage of the proposed procedure, related literature is discussed. The assumption of stationarity under the null plays an important role in the proof of the limiting distribution. However, a testing procedure that allows for occurring breaks in the variance may be of interest as well. In Chapter 4, a bootstrap version of the test is proposed as valid procedure in the case of non-stationary variances. Additionally, by extending already acquired methods, a test for change in the conditional variance function is obtained in Chapter 5. It contains a heuristic discussion of the limiting distribution under the null and consistency properties against changepoint alternatives. Finally, in Chapter 6, the finite sample performance of the tests, presented in this thesis, is investigated. Both level and power simulations in various time series models are conducted. Additionally, the tests are applied to the Nile and DJIA data sets that are given in Figure 2. Technical and auxiliary lemmata for all proofs can be found in Appendix A. A weak convergence result for sequential empirical processes with weakly dependent data is given in Appendix B. It is needed for the main result of this thesis, but may also be of interest on its own.

1 Fundamentals

In this introductory chapter, a literature review on changepoint detection in regression analysis will be given. Furthermore, the statistical model and the null hypothesis under consideration will be presented. Finally, some basic definitions concerning mixing conditions and weak convergence of empirical processes will be made, as well as the stochastic o and O notations will be introduced.

1.1 Literature review

Since the pioneering work of Page [58, 59] in the field of quality control, changepoint analysis has received much attention in the literature. Especially in the parametric framework, a lot of research has been done. In the most simple situation a possible mean change in otherwise independent and identically distributed (i.i.d.) random variables is considered. Later on this was extended to stability tests for the parameter in linear regression models. A popular procedure for changepoint detection in this context is the so called CUSUM[1] test, first proposed by Brown, Durbin and Evans [5], which is based on the fluctuation in the partial sum of residuals. It has been used quite extensively in the case of independent observations, see for example Csörgő and Horváth [12] and references mentioned therein. The problem naturally moved from i.i.d. into the time series context. Krämer, Ploberger and Alt [38] considered linear regression models with lagged dependent variables, Horváth [32] investigated linear autoregressive models and Bai [3] more generally studied ARMA[2] models, to name just a few. In contrast, relatively little work has been devoted to the nonparametric case. Most of the existing literature deals with the construction of estimates for both size and location of a possible changepoint (cf. Delgado and Hidalgo [16] for an overview) rather than testing for its existence. Kirch and Kamgaing [35] extented a CUSUM type test to non-linear autoregressive models, using neural networks. Both Hidalgo [29] and Honda [31] used nonparametric methods to test for changepoints in the regression function in nonparametric time series regression models with strictly stationary and absolutely regular data. Su and Xiao [71] extended these tests to not necessarily stationary and strongly mixing processes, allowing for heteroscedasticity and changes in the conditional variance function. However, their procedure does only seem to work for fixed changes in the variance. Su and White [70] proposed changepoint tests in partially linear time series models and Vogt [76] constructed a kernel-based L_2-test for structural change in the regression function in time-varying nonparametric regression models with locally stationary regressors. Hidalgo and

[1]CUSUM: cumulative sum

[2]ARMA: autoregressive moving-average

Dalla [30] proposed a test for smoothness in a nonparametric regression model with dependent data, based on the supremum of the difference between one-sided kernel regression estimates.

A related strand of the literature deals with changepoint detection in the error distribution of a regression model. In the parametric framework Koul [37] considered non-linear regression models and Ling [48] non-stationary AR models, to just mention a few. Neumeyer and Van Keilegom [54] constructed a test for change in the error distribution in nonparametric regression models with independent observations, while Selk and Neumeyer [66] obtained a procedure that allows for heteroscedastic autoregression models.

Instead of considering the whole error distribution, more specifically tests for changepoints in the unconditional error variance can be of interest as well. Lee, Na and Na [45] considered parametric autoregression models, as well as fixed design nonparametric regression models with strongly mixing errors using a CUSUM testing procedure. Chen and Tian [10] constructed a ratio test for changepoint detection in the variance in random design nonparametric regression models. Though, it does not allow for autoregressive effects, as a compact support of regressors is assumed.

Another interesting issue is the investigation of structural stability of the conditional variance function in heteroscedastic models. While again a lot of research has been devoted to the parametric case, notably for ARCH and GARCH[3] models (cf. Chen, Choi and Zhou [8] for an overview), just a few authors have considered nonparametric models. Chen, Choi and Zhou [8] for instance studied a nonparametric heteroscedastic time series model with a scale change in volatility. However, they assume a compact support of regressors, which is problematic when considering autoregression models.

1.2 The model

The aim of this thesis is to draw conclusions from an observed data set $\{(Y_t, \boldsymbol{X}_t) : 1 \leq t \leq n\}$ of size $n \in \mathbb{N}$ about the structural behavior of the whole process. The model under consideration is a regression model, that also allows for autoregressive effects and heteroscedasticity.

To this end let $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be a weakly dependent stochastic process, following the regression model

$$Y_t = m_t(\boldsymbol{X}_t) + U_t, \ t \in \mathbb{Z}, \tag{1.1}$$

where $\boldsymbol{X}_t$ is a d-dimensional covariate, Y_t is a one-dimensional response variable[4] and with unobservable innovations $(U_t)_{t\in\mathbb{Z}}$. Let $E[U_t|\mathcal{F}^t] = 0$ almost surely (a.s.), where $\mathcal{F}^t = \sigma(U_{j-1}, \boldsymbol{X}_j : j \leq t)$ and $\sigma(Z)$ denotes the σ-algebra generated by the random variable Z. For the unknown regression function $m_t : \mathbb{R}^d \to \mathbb{R}$, it therefore holds that

$$E[Y_t|\boldsymbol{X}_t = \boldsymbol{x}] = m_t(\boldsymbol{x}), \ t \in \mathbb{Z},$$

for all $\boldsymbol{x} \in \mathbb{R}^d$. Thus, m_t is the *conditional mean function* of Y_t conditioned on $\boldsymbol{X}_t$.

[3] (G)ARCH: (generalized) autoregressive conditional heteroscedasticity

[4] Sometimes the notation $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t \in \mathbb{Z}\}$ will be used as well.

Assuming $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ have been observed, the aim is to test the following null hypothesis

$$H_0 : m_t(\cdot) = m(\cdot),\ t = 1, \ldots, n, \tag{1.2}$$

for some function $m : \mathbb{R}^d \to \mathbb{R}$ not depending on the time of observation t against the alternative hypothesis

$$H_1 : \exists s_0 \in (0,1) : m_t(\cdot) = \begin{cases} m_{(1)}(\cdot), & t = 1, \ldots, \lfloor ns_0 \rfloor \\ m_{(2)}(\cdot), & t = \lfloor ns_0 \rfloor + 1, \ldots, n \end{cases}, \tag{1.3}$$

for some functions $m_{(1)}, m_{(2)} : \mathbb{R}^d \to \mathbb{R}$ with $m_{(1)} \not\equiv m_{(2)}$.

Note that $\boldsymbol{X}_t$ may include finitely many lagged values of Y_t, for instance $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-d})^T$. Hence, the model allows for autoregressive effects. By allowing the second moments of the innovations U_t conditioned on $\boldsymbol{X}_t$ to depend on $\boldsymbol{X}_t$, heteroscedastic models will also be covered.

1.3 Mixing conditions

Regarding the dependence of the process, some kind of weak dependence structure is needed. In what follows, the notion of strongly mixing triangular arrays and sequences will be introduced. In the mixing framework a so called *mixing coefficient* is defined, that in some sense measures the dependency of two segments of the sequence that are apart from each other in time t. The sequence is called *mixing* if this coefficient tends to zero, as t tends to infinity, i.e. heuristically speaking if the two segments behave asymptotically independent.

More precisely, let the following definition be introduced which can be found in [69].

Definition 1.1 (Strongly mixing triangular array)**.** For some triangular array of random variables $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ define

$$\alpha_n(t) := \begin{cases} \sup\limits_{1 \leq k \leq n-t} \ \sup\limits_{\substack{A \in \sigma(X_{n,j} : 1 \leq j \leq k) \\ B \in \sigma(X_{n,j} : k+t \leq j \leq n)}} |P(A \cap B) - P(A)P(B)|, & t \leq n-1 \\ 0, & t \geq n \end{cases} \tag{1.4}$$

and

$$\alpha(t) := \begin{cases} \sup\limits_{n \in \mathbb{N}} \alpha_n(t), & t \in \mathbb{N} \\ 1, & t = 0 \end{cases}.$$

Then $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ is called *strongly mixing* or *α-mixing*, if $\alpha(t) \to 0$ as $t \to \infty$. Furthermore, $\alpha(\cdot)$ is referred to as the *mixing coefficient* of $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$.

Remark. Note that within some proofs in this thesis, where results only for a fixed sample size $n \in \mathbb{N}$ are needed, properties of $\alpha_n(\cdot)$ as defined in (1.4), rather than those of $\alpha(\cdot)$, are used. In these cases $\alpha_n(\cdot)$ will also be referred to as the mixing coefficient of $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$.

For a sequence of random variables the strongly mixing notion simplifies in the following matter. This version can for example be found in [4].

Definition 1.2 (Strongly mixing sequence)**.** A sequence of random variables $\{X_t : t \in \mathbb{Z}\}$ is called *strongly mixing* or *α-mixing*, if

$$\alpha(t) := \sup_{k\in\mathbb{Z}} \sup_{\substack{A\in\sigma(X_j : j\leq k)\\ B\in\sigma(X_j : k+t\leq j)}} |P(A\cap B) - P(A)P(B)| \to 0,\ t\to\infty.$$

If the sequence is strictly stationary[5], then the mixing coefficient simplifies to

$$\alpha(t) = \sup_{\substack{A\in\sigma(X_j : j\leq 0)\\ B\in\sigma(X_j : t\leq j)}} |P(A\cap B) - P(A)P(B)|.$$

In Chapter 3, a test statistic will be constructed to test H_0 of no change in the regression function against changepoint alternatives as H_1. In Section 3.2, which concerns the asymptotic behavior of the test statistic under the null, the process $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R}\times\mathbb{R}^d : t \in \mathbb{Z}\}$ will be assumed to be strictly stationary. The strongly mixing assumption used therein is therefore meant in the sense of Definition 1.2. Section 3.3 deals with the behavior under some fixed alternatives. As the possible changepoint is assumed to depend on the sample size $n \in \mathbb{N}$, the process will then be viewed as a triangular array $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) \in \mathbb{R}\times\mathbb{R}^d : 1 \leq t \leq n, n \in \mathbb{N}\}$. The strongly mixing assumption is then meant in the sense of Definition 1.1.

Remark. Two different kinds of conditions on the mixing coefficient will be used in this thesis, namely

- polynomial mixing rates, i.e. $\alpha(t) \leq Bt^{-\beta}$ for all $t > 0$ and for some $\beta > 0$, $0 < B < \infty$ and
- geometric (exponential) mixing rates, i.e. $\alpha(t) \leq Aa^{-t}$ for all $t > 0$ and for some $a \in (1,\infty)$, $0 < A < \infty$.

In the following, examples will be given, that are strongly mixing with exponential mixing rates.

Example. (i) Let $(Y_t)_{t\in\mathbb{Z}}$ be strictly stationary following the AR(1) model

$$Y_t = aY_{t-1} + \varepsilon_t,\ t\in\mathbb{Z}$$

with $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and $|a| < 1$. Then $(Y_t)_{t\in\mathbb{Z}}$ is a linear process of the form $Y_t = \sum_{j=0}^{\infty} a^j \varepsilon_{t-j}$ (see for instance example 2.7 in [41]). As the coefficients in above series converge exponentially fast, $(Y_t)_{t\in\mathbb{Z}}$ is strongly mixing and possesses exponential mixing rates (see for instance [21] Subsection 2.6.1 (iii), p. 69). Note that this does not hold if for instance the innovations $(\varepsilon_t)_{t\in\mathbb{Z}}$ follow a binomial distribution as stated in [21] as well.

[5] A sequence $\{X_t : t \in \mathbb{Z}\}$ is called *strictly stationary* if $(X_{t_1},\dots,X_{t_k}) \overset{\mathcal{D}}{=} (X_{t_1+h},\dots,X_{t_k+h})$ for all $k \in \mathbb{N}$ and $t_1,\dots,t_k,h \in \mathbb{Z}$, where $\overset{\mathcal{D}}{=}$ denotes equality in distribution.

(ii) Causal and stationary ARMA processes have an MA(∞) representation with coefficients that decay exponentially fast (see for instance Remark 7.8 in [41]). Hence, they are also strongly mixing with exponential mixing rates, provided that the innovations are absolutely continuous.

(iii) Doukhan [18] considers the non-linear AR-ARCH process $(Y_t)_{t\in\mathbb{Z}}$ of the form

$$Y_t = m(Y_{t-1}) + \sigma(Y_{t-1})\varepsilon_t, \ t \in \mathbb{Z}$$

for an i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ with mean zero and variance one. Proposition 6 in Subsection 2.4.2.3 in [18] gives conditions on m, σ and the innovations that imply geometric ergodicity of $(Y_t)_{t\in\mathbb{Z}}$ (see [18], p. 89 for the definition). This implies the strong mixing property with exponential mixing rates (see [21] Subsection 2.6.1 (vi), p. 70).

(iv) Both Lu [50] and Liebscher [47] consider $(Y_t)_{t\in\mathbb{Z}}$ following the more general non-linear AR-ARCH equation

$$Y_t = m(Y_{t-1}, \dots, Y_{t-d}) + \sigma(Y_{t-1}, \dots, Y_{t-d})\varepsilon_t, \ t \in \mathbb{Z}$$

for an i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$ with mean zero and variance one. They both give sufficient conditions on m, σ and the innovations under which $(Y_t)_{t\in\mathbb{Z}}$ is geometric ergodic (see Theorem 1 in [50] and Theorem 4 in [47]). In the linear model

$$Y_t = a_1 Y_{t-1} + \dots + a_d Y_{t-d} + \sqrt{b_0 + b_1 Y_{t-1}^2 + \dots + b_d Y_{t-d}^2}\varepsilon_t, \ t \in \mathbb{Z},$$

where $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ε_t is independent of Y_j for all $j \le t-1$, the condition in [50] simplifies to

$$\left(\sum_{i=1}^d |a_i|\right)^2 + \sum_{i=1}^d b_i < 1.$$

Note that this however is not the weakest possible condition as can for instance easily be seen in the homoscedastic case, i.e. for $b_1 = \dots = b_d = 0$.

1.4 Weak convergence and empirical processes

The concept of weak convergence of a sequence of stochastic processes is a generalization of convergence in distribution of a sequence of random variables. Instead of random variables with values in Euclidean spaces, it concerns random elements that take values in more abstract metric spaces. The space that will be of interest in this thesis is the function space of all uniformly bounded real-valued functions. More precisely let T be an arbitrary set and let

$$l^\infty(T) := \left\{ z : T \to \mathbb{R} : \|z\|_\infty := \sup_{t\in T} |z(t)| < \infty \right\}.$$

This is a metric space with respect to the sup norm $\|\cdot\|_\infty$. Following the modern empirical process theory, well summarized in [75], this space will be equipped with the Borel σ-algebra, the smallest σ-algebra that contains all open sets.

A stochastic process $Z = \{Z(t) : t \in T\}$, defined on some underlying probability space, can be viewed as a random element in $l^\infty(T)$ if all sample paths are bounded. Let in the following all random objects (measurable or not measurable) be defined on the same underlying probability space $(\Omega, \mathcal{A}, P)$. The following definition can for example be found as Definition 1.3.3 in [75] in a more abstract version for random elements in general metric spaces.

Definition 1.3 (Weak convergence)**.** Let Z be measurable with values in $l^\infty(T)$. A sequence Z_n with values in $l^\infty(T)$ is said to *converge weakly* to Z if

$$E^*[H(Z_n)] \underset{n\to\infty}{\to} E[H(Z)]$$

for all bounded and continuous functions $H : l^\infty(T) \to \mathbb{R}$, where $E^*[X]$ denotes the outer expectation of a possibly non-measurable real valued mapping X. It will be denoted by $Z_n \underset{n\to\infty}{\rightsquigarrow} Z$.

Remark. As it can be seen for example by applying Theorem 1.5.7 and Theorem 1.5.4 in [75], it holds that Z_n converges weakly to Z in $l^\infty(T)$ if and only if the following two conditions hold

- *fidi convergence:* for all $K \in \mathbb{N}$ and all $t_1, \ldots, t_K \in T$

$$(Z_n(t_k))_{k=1,\ldots,K} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} (Z(t_k))_{k=1,\ldots,K},$$

  where $\overset{\mathcal{D}}{\to}$ denotes convergence in distribution.

- there exists a semi metric d on T, such that (T, d) is totally bounded and Z_n is *asymptotic equicontinuous*, i.e.

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} P^* \left(\sup_{\{t_1, t_2 \in T : d(t_1,t_2)<\delta\}} |Z_n(t_1) - Z_n(t_2)| > \epsilon \right) = 0$$

  for all $\epsilon > 0$, where $P^*(A)$ denotes the outer probability of a possibly non-measurable set A.

These two conditions are in many situations easier to verify.

Definition 1.4 (Empirical process)**.** Let $\{X_{n,t} : 1 \le t \le n, n \in \mathbb{N}\}$ be a triangular array of random variables with values in some measure space $\mathcal{X}$. For some measurable function $\varphi : \mathcal{X} \to \mathbb{R}$ and some $s \in [0,1]$ let

$$G_n(s,\varphi) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\varphi(X_{n,i}) - E[\varphi(X_{n,i})]\right), n \in \mathbb{N}.$$

For some function class $\mathcal{F}$ of measurable functions $\mathcal{X} \to \mathbb{R}$ the *(non-sequential) empirical process* indexed by $\mathcal{F}$ is defined as $\{G_n(1,\varphi) : \varphi \in \mathcal{F}\}, n \in \mathbb{N}$, and can be viewed as a sequence of random elements in $l^\infty(\mathcal{F})$. The *sequential empirical process* indexed by $[0,1] \times \mathcal{F}$ is defined as $\{G_n(s,\varphi) : s \in [0,1], \varphi \in \mathcal{F}\}, n \in \mathbb{N}$, and can be viewed as a sequence of random elements in $l^\infty([0,1] \times \mathcal{F})$.

For the empirical processes to converge, assumptions concerning the underlying process and the function class are needed. Assumptions on the underlying process concern dependency and distribution. Assumptions on the function class are often given in terms of the so called covering and bracketing numbers, which in some sense measure the size of the function class. The following definition introduces the bracketing notion and can be found for example as Definition 2.1.6 in [75].

Definition 1.5 (Bracketing number - first version)**.** Let $\mathcal{X}$ be a measure space, $\mathcal{F}$ some class of functions $\mathcal{X} \to \mathbb{R}$ and ρ some semi norm on $\mathcal{F}$. Given two functions $l, u : \mathcal{X} \to \mathbb{R}$ with $\rho(l) < \infty$ and $\rho(u) < \infty$, the set of all functions φ such that $l \leq \varphi \leq u$ is called *bracket* and denoted by $[l, u]$. For $\varepsilon > 0$ the ε-*bracket* is a bracket $[l, u]$ with $\rho(u - l) < \varepsilon$. The smallest number of ε-brackets needed to cover $\mathcal{F}$ is called the *bracketing number* and denoted by $N_{[\,]}(\varepsilon, \mathcal{F}, \rho)$.

In Appendix B, a different definition for bracketing number will be used to show a weak convergence result for empirical processes with weakly dependent random variables. It is based on Definition 2.1 in [2], but uses a slightly different notation.

Definition 1.6 (Bracketing number - second version)**.** 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 an $a^* \in \mathcal{A}$ and a $b^* \in \mathcal{B}$ such that

$$|\varphi - a^*| \leq b^*.$$

Then $N(\varepsilon)$ is called the *bracketing number* and denoted by $\tilde{N}_{[\,]}(\varepsilon, \mathcal{F}, \rho)$. The function a^* is referred to as the *(to φ) corresponding approximating function* and the function b^* as the *(to φ) corresponding bounding function.*

Remark. Note that $N_{[\,]}(2\varepsilon, \mathcal{F}, \rho) \leq \tilde{N}_{[\,]}(\varepsilon, \mathcal{F}, \rho)$ for all $\varepsilon > 0$. To see this, for $\varepsilon > 0$ and $\varphi \in \mathcal{F}$ let a^* and b^* as in Definition 1.6 be considered. Then $[l, u]$ with $l := a^* - b^*$ and $u := a^* + b^*$ is a 2ε-bracket containing φ as $l \leq \varphi \leq u$ and

$$\rho(u - l) = \rho(2b^*) = 2\rho(b^*) < 2\varepsilon.$$

On the other hand for an ε-bracket $[l, u]$ containing $\varphi \in \mathcal{F}$, by the choice of $a^* := \frac{u+l}{2}$ and $b^* := \frac{u-l}{2}$, it can be obtained that $|\varphi - a^*| \leq b^*$ and

$$\rho(b^*) = \rho\left(\frac{u - l}{2}\right) = \frac{1}{2}\rho(b^*) < \frac{\varepsilon}{2}.$$

However, as a^* is not necessarily element in $\mathcal{F}$, this does not necessarily lead to a valid approximating function as in Definition 1.6. Hence, the other inequality does not hold.

1.5 Stochastic o and O symbols

In analysis the o and O symbols are used to describe the limiting behavior of a function. It particularly simplifies notations when expressing rates of convergence. When talking about its stochastic versions, rates of convergence in probability are of interest.

For real valued sequences $(x_n)_{n\in\mathbb{N}}$ and $(y_n)_{n\in\mathbb{N}}$ let the following notations be introduced

$$x_n = o(y_n) :\Leftrightarrow \frac{x_n}{y_n} \underset{n\to\infty}{\to} 0$$

and

$$x_n = O(y_n) :\Leftrightarrow \exists C < \infty, N < \infty : \left|\frac{x_n}{y_n}\right| < C,\ \forall n \geq N.$$

For sequences of real valued random variables $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$ defined on some common probability space $(\Omega, \mathcal{A}, P)$ let the following notations be introduced

$$X_n = o_P(Y_n) :\Leftrightarrow \frac{X_n}{Y_n} \underset{n\to\infty}{\overset{P}{\to}} 0,$$

where $\overset{P}{\to}$ denotes convergence in probability. Furthermore, let

$$X_n = O_P(Y_n) :\Leftrightarrow \forall t > 0\ \exists C < \infty, N < \infty : P\left(\left|\frac{X_n}{Y_n}\right| > C\right) < t,\ \forall n \geq N.$$

For further information and some useful properties see Section 2.2 in [74].

2 Kernel estimation

To avoid additional model misspecification, nonparametric methods are used for the estimation of unknown functions. In contrast to parametric procedures, no specific assumptions on the form of the unknown functions are presumed. The only kind of conditions concern smoothness and uniform bounds on expanding compact sets. One of the most popular approaches in nonparametric statistics is the kernel estimation. Kernel density and regression estimators, that are used in this thesis, will be defined and uniform rates of convergence for strongly mixing and strictly stationary processes will be proven under regularity assumptions. The results will be compared with existing literature. Additionally, a uniform consistency result will be given for strongly mixing triangular array processes.

2.1 Definition

The nonparametric estimators, that will be used, are the *kernel density estimator*, its introduction can be traced back to Rosenblatt [65] in 1956 and Parzen [60] in 1962, and the *Nadaraya-Watson estimator*, independently proposed by Nadaraya [53] and Watson [79] in 1964.

Definition 2.1. Let $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t = 1, \dots, n\}$ be a sample of size $n \in \mathbb{N}$, $K : \mathbb{R}^d \to \mathbb{R}$ a function with $\int_{\mathbb{R}^d} K(\boldsymbol{x}) d\boldsymbol{x} = 1$, called a *kernel*, and h a positive real valued number, called a *bandwidth*. Let $K_h(\cdot) := \frac{1}{h^d} K\left(\frac{\cdot}{h}\right)$. The *kernel density estimator* is defined by

$$\hat{f}_n(\boldsymbol{x}) := \frac{1}{n} \sum_{i=1}^{n} K_h(\boldsymbol{X}_i - \boldsymbol{x}), \text{ for } \boldsymbol{x} \in \mathbb{R}^d. \tag{2.1}$$

The *Nadaraya-Watson estimator* is defined by

$$\hat{m}_n(\boldsymbol{x}) := \begin{cases} \dfrac{\frac{1}{n} \sum\limits_{i=1}^{n} K_h(\boldsymbol{X}_i - \boldsymbol{x}) Y_i}{\hat{f}_n(\boldsymbol{x})}, & \text{if } \hat{f}_n(\boldsymbol{x}) \neq 0\,, \\ 0, & \text{otherwise} \end{cases} \tag{2.2}$$

for $\boldsymbol{x} \in \mathbb{R}^d$, following standard literature (see for instance [73], p. 32). Furthermore, for $\boldsymbol{x} \in \mathbb{R}^d$ define

$$\hat{\sigma}_n^2(\boldsymbol{x}) := \begin{cases} \dfrac{\frac{1}{n} \sum\limits_{i=1}^{n} K_h(\boldsymbol{X}_i - \boldsymbol{x})(Y_i - \hat{m}_n(\boldsymbol{x}))^2}{\hat{f}_n(\boldsymbol{x})}, & \text{if } \hat{f}_n(\boldsymbol{x}) \neq 0\,. \\ 0, & \text{otherwise} \end{cases} \tag{2.3}$$

Remark.
- Rosenblatt [65] first introduced the kernel density estimator for $d = 1$ and $K(x) := \frac{1}{2}I(|x| \leq 1)$, the so called *rectangular kernel*, motivated by a discrete derivative version of the empirical distribution function.
- For i.i.d. or strictly stationary data, the density function f of $\boldsymbol{X}_i$ does not depend on i and $\hat{f}_n$ is a nonparametric estimator for f. Furthermore, $m(\boldsymbol{x}) := E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}]$ does not depend on i and $\hat{m}_n$ is a nonparametric estimator for m. Additionally, $\sigma^2(\boldsymbol{x}) := Var(Y_i|\boldsymbol{X}_i = \boldsymbol{x})$ then does not depend on i either and $\hat{\sigma}_n^2$ is an estimator for σ^2.
- Note that, if K is non-negative and therefore a probability density itself, then for fixed $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$, the function $\boldsymbol{x} \mapsto \hat{f}_n(\boldsymbol{x})$ is a probability density. In this case the function $\boldsymbol{x} \mapsto \hat{\sigma}_n^2(\boldsymbol{x})$ is also non-negative and therefore

$$\hat{\sigma}_n(\boldsymbol{x}) := \sqrt{\hat{\sigma}_n^2(\boldsymbol{x})} \text{ for } \boldsymbol{x} \in \mathbb{R}^d$$

  is well defined. Nevertheless, for technical reasons it is useful to allow the kernel to take negative values, as will be seen later on. Thus, both $\hat{f}_n$ and $\hat{\sigma}_n^2$ can take negative values for a finite sample size.
- The kernel density estimator $\hat{f}_n$ can be interpreted as a smooth version of the histogram of a sample. Additionally, $\hat{m}_n(\boldsymbol{x})$ is an average of such Y_i that the corresponding $\boldsymbol{X}_i$ lies in a neighborhood of $\boldsymbol{x}$ and $\hat{\sigma}_n^2(\boldsymbol{x})$ is an average of such $(Y_i - \hat{m}_n(\boldsymbol{x}))^2$ that the corresponding $\boldsymbol{X}_i$ lies in a neighborhood of $\boldsymbol{x}$. The bandwidth h determines the size of the neighborhood and should be chosen dependent on $n \in \mathbb{N}$. For a larger sample size, the bandwidth should be chosen smaller. Thus, in an asymptotic framework the bandwidth will always be a sequence of positive real valued numbers $(h_n)_{n\in\mathbb{N}}$ with $\lim_{n\to\infty} h_n = 0$.

2.2 Uniform rates of convergence

In this section, the performance of the estimators will be studied given a strictly stationary and α-mixing sequence $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t \in \mathbb{Z}\}$. In particular, it will be shown that the difference between the estimators and the unknown functions converges in probability to zero as $n \to \infty$ uniformly over some compact subset of $\mathbb{R}^d$ and with certain rates. A direct consequence will be the property of consistency for the estimators. Furthermore, uniform rates of convergence will be shown for the partial derivatives of the difference. Therefore, a multi-index notation for higher order partial derivatives is needed. For an index $\boldsymbol{k} = (k_1, \ldots, k_d) \in \mathbb{N}_0^d$ and $\boldsymbol{x} = (x_1, \ldots, x_d) \in \mathbb{R}^d$ let

$$\begin{aligned}
|\boldsymbol{k}| &:= k_1 + \cdots + k_d, \\
\boldsymbol{k}! &:= k_1! \cdots k_d!, \\
\boldsymbol{x}^{\boldsymbol{k}} &:= x_1^{k_1} \cdots x_d^{k_d}.
\end{aligned}$$

For a function $h : \mathbb{R}^d \to \mathbb{R}$, $\boldsymbol{x} \in \mathbb{R}^d$ and $\boldsymbol{k} \in \mathbb{N}_0^d \backslash \{\boldsymbol{0}\}$, where $\boldsymbol{0} := (0, \ldots, 0) \in \mathbb{N}_0^d$, let

$$D^{\boldsymbol{k}} h(\boldsymbol{x}) := \frac{\partial^{|\boldsymbol{k}|} h}{\partial x_1^{k_1} \ldots \partial x_d^{k_d}}(\boldsymbol{x}),$$

if all the $|\boldsymbol{k}|$-th partial derivatives of h exist in $\boldsymbol{x}$ and $D^{\boldsymbol{0}}h(\boldsymbol{x}) := h(\boldsymbol{x})$. A useful tool is Taylor's expansion. It will be used with Lagrange form of the remainder, which can be found for example in [36], p. 65. Let $r \in \mathbb{N}$. For an r times continuously differentiable function $h : \mathbb{R}^d \to \mathbb{R}$ and some $\boldsymbol{x}, \boldsymbol{a}$ in some open subset $\boldsymbol{U} \subset \mathbb{R}^d$ for which the line segment lies in $\boldsymbol{U}$ as well, there exists a $\boldsymbol{\xi}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{a}$, such that

$$h(\boldsymbol{x}) = \sum_{\substack{\boldsymbol{i}\in\mathbb{N}_0^d \\ 0\leq|\boldsymbol{i}|\leq r-1}} \frac{D^{\boldsymbol{i}}h(\boldsymbol{a})(\boldsymbol{x}-\boldsymbol{a})^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{\substack{\boldsymbol{i}\in\mathbb{N}_0^d \\ |\boldsymbol{i}|=r}} \frac{D^{\boldsymbol{i}}h(\boldsymbol{\xi})(\boldsymbol{x}-\boldsymbol{a})^{\boldsymbol{i}}}{\boldsymbol{i}!}. \tag{2.4}$$

For brevity reasons, the condition $\boldsymbol{i} \in \mathbb{N}_0^d$ in the bound of summation will be omitted most of the times. The first sum of (2.4) is the Taylor polynomial of order $r-1$ of h in $\boldsymbol{a}$. The second sum is the remainder term in Lagrange form. Another result, that will be used, is the so called Leibniz's formula for higher partial derivatives of the product of two functions. It can be found in [20], p. 13. For functions $u, v : \mathbb{R}^d \to \mathbb{R}$ and $\boldsymbol{k} \in \mathbb{N}_0^d$, it states that

$$D^{\boldsymbol{k}}(uv)(\boldsymbol{x}) = \sum_{\boldsymbol{i}\leq\boldsymbol{k}} \binom{\boldsymbol{k}}{\boldsymbol{i}} D^{\boldsymbol{i}}u(\boldsymbol{x})D^{\boldsymbol{k}-\boldsymbol{i}}v(\boldsymbol{x}), \tag{2.5}$$

where $\binom{\boldsymbol{k}}{\boldsymbol{i}} := \frac{\boldsymbol{k}!}{(\boldsymbol{k}-\boldsymbol{i})!\boldsymbol{i}!}$ for all $\boldsymbol{i} \leq \boldsymbol{k}$ which in turn is short for $i_j \leq k_j$ for all $j = 1, \ldots, d$.

As already mentioned, rates of convergence will be shown uniformly in $\boldsymbol{x}$. As the kernel estimators only perform well in regions where there are many observations and rather poorly on the edges and outside of the sample space, nice asymptotic properties can not be expected on the whole domain $\mathbb{R}^d$. This issue can be solved by a truncation of the domain to the following compact set $\boldsymbol{J}_n \subset \mathbb{R}^d$

$$\boldsymbol{J}_n := [-c_n, c_n]^d := \bigtimes_{i=1}^{d} [-c_n, c_n],$$

where $(c_n)_{n\in\mathbb{N}}$ is a positive sequence of real valued numbers converging to infinity as n tends to infinity. Dependent on the rate of convergence of c_n, uniform rates of convergence of the estimators on the set $\boldsymbol{J}_n$ can be established.

In what follows, regularity assumptions will be presented under which the uniform rates of convergence stated in Lemma 2.2 hold.

(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

$$\alpha(t) = O\left(t^{-\beta}\right),\ (t \to \infty) \text{ and } \beta > \frac{1+(b-1)(1+d)}{b-2}. \tag{2.6}$$

(M) For b from assumption **(P)** let

- $E[|Y_1|^b] < \infty$,
- $\boldsymbol{X}_1$ be absolutely continuous with density function $f : \mathbb{R}^d \to \mathbb{R}$ that satisfies $\sup_{\boldsymbol{x}\in\mathbb{R}^d} E[|Y_1|^b|\boldsymbol{X}_0 = \boldsymbol{x}]f(\boldsymbol{x}) < \infty$ and $\sup_{\boldsymbol{x}\in\mathbb{R}^d} f(\boldsymbol{x}) < \infty$, and

- $\exists j^* < \infty$ such that $\forall\, j \geq j^*$: $\sup_{\boldsymbol{x}_1,\boldsymbol{x}_j} E[|Y_1 Y_j||\boldsymbol{X}_1 = \boldsymbol{x}_1, \boldsymbol{X}_j = \boldsymbol{x}_j] f_{1j}(\boldsymbol{x}_1, \boldsymbol{x}_j) < \infty$, where f_{1j} is the density function of $(\boldsymbol{X}_1, \boldsymbol{X}_j)$.

Additionally, let $m : \mathbb{R}^d \to \mathbb{R}$ be the conditional mean function defined by

$$m(\boldsymbol{x}) := E[Y_1|\boldsymbol{X}_1 = \boldsymbol{x}] \text{ for all } \boldsymbol{x} \in \mathbb{R}^d.$$

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

(F1) For some $C < \infty$ and c_n from assumption **(J)** let $\boldsymbol{I}_n := [-c_n - Ch_n, c_n + Ch_n]^d$ and for some $r, l \in \mathbb{N}$ let

- f and $m : \mathbb{R}^d \to \mathbb{R}$ be $l + 1 + r$ times continuously differentiable,
- $\delta_n^{-1} := \inf_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x}) > 0$ for all $n \in \mathbb{N}$,
- $p_n := \max_{1\leq|\boldsymbol{k}|\leq l+1+r} \sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{k}} f(\boldsymbol{x})| < \infty$ for all $n \in \mathbb{N}$ and
- $q_n := \max_{0\leq|\boldsymbol{k}|\leq l+1+r} \sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{k}} m(\boldsymbol{x})| < \infty$ and $q_n > 0$ for all $n \in \mathbb{N}$.

(K) Let $K : \mathbb{R}^d \to \mathbb{R}$ be symmetric in each component with $\int_{\mathbb{R}^d} K(\boldsymbol{z})d\boldsymbol{z} = 1$. Additionally, for $r, l \in \mathbb{N}$ and C from assumption **(F1)** 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 $1 \leq |\boldsymbol{k}| \leq r - 1$,
- K have compact support $[-C, C]^d$,
- K be $l + 1$ times differentiable.

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 $\boldsymbol{u} \in \mathbb{R}^d$,
- $|L(\boldsymbol{u}) - L(\boldsymbol{u'})| \leq \Lambda\|\boldsymbol{u} - \boldsymbol{u'}\|$ for some $\Lambda < \infty$ and for all $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^d$.

(B1) With b and β from assumption **(P)** let

$$\frac{\log(n)}{n^\theta h_n^d} = o(1) \text{ for } \theta = \frac{\beta - 1 - d - \frac{1+\beta}{b-1}}{\beta + 3 - d - \frac{1+\beta}{b-1}}. \tag{2.7}$$

(B2) For δ_n, p_n, q_n and r, l from assumption **(F1)** let

$$\left(\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+1} \delta_n^{l+2} = O(1), \tag{2.8}$$

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

$$\left(\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+\eta} q_n \delta_n^{l+1+\eta} = o(1). \tag{2.9}$$

Remark.
- Assumptions **(P)**, **(M)**, parts of **(K)** and **(B1)** are reproduced from Hansen [26]. In particular, Theorem 2 in [26], which is a result on uniform rates of convergence for general kernel estimators, will be applied in the proof several times.

- The second bullet point in **(M)** controls the tail behavior of the conditional expectation $E[|Y_1|^b|\boldsymbol{X}_0 = \boldsymbol{x}]$ which can only increase to infinity at a slower rate than $f(\boldsymbol{x})^{-1}$. The last bullet point in **(M)** is a similar assumption for the joint density and conditional expectation.

- Kernel functions that satisfy the first bullet point in **(K)** are often referred to as *kernels of order* r. Together with the existence of r-th partial derivatives of f and m in **(F1)** it causes the bias to be of order $O(h_n^r)$. These are typical assumptions in the nonparametric framework.

- The condition $q_n > 0$ for all $n \in \mathbb{N}$ excludes the case of $m \equiv 0$. However, in this particular case uniform rates of convergence can be obtained as well, which will be pointed out within the proof.

- Assumption **(P)** specifies the dependence structure of the process $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ expressed in terms of the mixing notion. In particular, mixing coefficients with polynomial rates of convergence are allowed for.

Lemma 2.2. *Under the assumptions **(P)**, **(M)**, **(J)**, **(F1)**, **(K)**, **(B1)** and **(B2)** the following rates of convergence can be obtained.*

(i) *For* $\hat{f}_n(\boldsymbol{x})$ *from* (2.1) *in Definition 2.1, it holds that*

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

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

(ii) *For* $\hat{g}_n(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(\boldsymbol{X}_i - \boldsymbol{x})Y_i$ *as an estimator for* $g(\boldsymbol{x}) := m(\boldsymbol{x})f(\boldsymbol{x})$, *it holds that*

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

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

(iii) *For* $\hat{m}_n(\boldsymbol{x})$ *from* (2.2) *in Definition 2.1, it holds that*

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

Remark. Note that this result implies that $\hat{f}_n$ and $\hat{m}_n$ are consistent estimators for f and m respectively. For the Nadaraya-Watson estimator it particularly implies that

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

$$\sup_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\ 1\leq|\boldsymbol{k}|\leq l}} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)\right| = o_P(1)$$

and

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

as the assumptions of Lemma 2.2 concerning the bandwidth imply

$$\left(\sqrt{\frac{\log(n)}{nh_n^{d+2j}}} + h_n^r p_n\right) p_n^j q_n \delta_n^{j+1} = o(1),\ \forall\, j \in \{0,\dots,l\}.$$

Thus, the difference $\hat{m}_n - m$ can be embedded in a function class, containing smooth, uniformly bounded functions, that posses uniformly bounded partial derivatives up to order l with highest partial derivatives being Lipschitz of order η. This technique will be used in the proof of Lemma A.1 in Appendix A.

Proof. The key tool in proving Lemma 2.2 is an application of Hansen's Theorem 2 in [26]. With

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x})\right| \leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{f}_n(\boldsymbol{x}) - E\left[\hat{f}_n(\boldsymbol{x})\right]\right| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|E\left[\hat{f}_n(\boldsymbol{x})\right] - f(\boldsymbol{x})\right|$$

the proof of (i) (a) splits into two parts, which will be treated separately, beginning with the left term. A direct application of Theorem 2 in [26] results in

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{f}_n(\boldsymbol{x}) - E\left[\hat{f}_n(\boldsymbol{x})\right]\right| = O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}\right).$$

Concerning the right term, inserting the definition of $\hat{f}_n$, using integration by substitution and $\int_{\mathbb{R}^d} K(\boldsymbol{u})d\boldsymbol{u} = 1$ yields

$$\begin{aligned}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|E\left[\hat{f}_n(\boldsymbol{x})\right] - f(\boldsymbol{x})\right| &= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|E\left[\frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h_n}\right)\right] - f(\boldsymbol{x})\right| \\ &= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\int_{\mathbb{R}^d}\frac{1}{h_n^d}K\left(\frac{\boldsymbol{u}-\boldsymbol{x}}{h_n}\right)f(\boldsymbol{u})d\boldsymbol{u} - f(\boldsymbol{x})\right|\end{aligned}$$

$$= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) f(\boldsymbol{z}h_n + \boldsymbol{x}) d\boldsymbol{z} - f(\boldsymbol{x}) \right|$$
$$= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(f(\boldsymbol{z}h_n + \boldsymbol{x}) - f(\boldsymbol{x}) \right) d\boldsymbol{z} \right|.$$

Taylor's expansion of f in $\boldsymbol{x}$ up to order $r-1$ with Lagrange remainder term results in

$$f(\boldsymbol{z}h_n + \boldsymbol{x}) - f(\boldsymbol{x}) = \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} f(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}$$

for some $\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{z}h_n$. This and the assumptions **(K)** and **(F1)** furthermore lead to

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| E\left[\hat{f}_n(\boldsymbol{x})\right] - f(\boldsymbol{x}) \right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(f(\boldsymbol{z}h_n + \boldsymbol{x}) - f(\boldsymbol{x}) \right) d\boldsymbol{z} \right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(\sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} f(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right) d\boldsymbol{z} \right| \\
&\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} f(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z} \right| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z} \right| \\
&\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} f(\boldsymbol{x}) h_n^{|\boldsymbol{i}|}}{\boldsymbol{i}!} \underbrace{\int_{\mathbb{R}^d} \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) d\boldsymbol{z}}_{\substack{=0 \\ \forall\ 1\leq|\boldsymbol{i}|\leq r-1}} \right| + h_n^r \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \sum_{|\boldsymbol{i}|=r} \left| \frac{D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})}{\boldsymbol{i}!} \right| \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{\boldsymbol{i}!} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \left| D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) \right| \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{\boldsymbol{i}!} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \underbrace{\left| D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) \right| I\{\boldsymbol{z} \in [-C,C]^d\}}_{\leq \sup_{\boldsymbol{y}\in[-C,C]^d} \left| D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}}) \right|} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{\boldsymbol{i}!} \sup_{\substack{\boldsymbol{x}\in\boldsymbol{J}_n \\ \boldsymbol{y}\in[-C,C]^d}} \left| D^{\boldsymbol{i}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}}) \right| \int_{\mathbb{R}^d} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{\boldsymbol{i}!} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{I}_n} \left| D^{\boldsymbol{i}} f(\boldsymbol{x}) \right|}_{\substack{=O(p_n) \\ \forall|\boldsymbol{i}|=r}} \underbrace{\int_{\mathbb{R}^d} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z}}_{\substack{<\infty \\ \forall|\boldsymbol{i}|=r}} \\
&= O(h_n^r p_n),
\end{aligned}$$

which concludes the proof of the assertion in (i) (a).

For (i) (b) let $\boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq l+1$. Notice that by chain rule

$$D^{\boldsymbol{k}} \hat{f}_n(\boldsymbol{x}) = D^{\boldsymbol{k}} \left(\frac{1}{nh_n^d} \sum_{i=1}^n K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right) \right) = \frac{(-1)^{|\boldsymbol{k}|}}{nh_n^{d+|\boldsymbol{k}|}} \sum_{i=1}^n D^{\boldsymbol{k}} K\left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right). \quad (2.10)$$

With

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})-D^{\boldsymbol{k}}f(\boldsymbol{x})\right|\\
&\le \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})-E\left[D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})\right]\right|+\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|E\left[D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})\right]-D^{\boldsymbol{k}}f(\boldsymbol{x})\right|
\end{aligned}$$

the proof of (i) (b) splits into two parts again, which will be treated separately beginning with the left term. Denoting $\hat{\Psi}(\boldsymbol{x}) := \frac{1}{nh_n^d}\sum_{i=1}^n D^{\boldsymbol{k}}K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h_n}\right)$, it follows that

$$\begin{aligned}
\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})-E\left[D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})\right]\right| &= \frac{1}{h_n^{|\boldsymbol{k}|}}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{\Psi}(\boldsymbol{x})-E\left[\hat{\Psi}(\boldsymbol{x})\right]\right|\\
&= O_P\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{k}|}}}\right),
\end{aligned}$$

where the last step is again an application of Theorem 2 in [26]. Note that the assumptions for the used Theorem are satisfied by the assumptions of the Lemma. In particular, regularity assumptions on the partial derivatives $D^{\boldsymbol{k}}K$ are needed here. Concerning the right term, inserting the representation of $D^{\boldsymbol{k}}\hat{f}_n$ of (2.10) and using integration by substitution yields

$$\begin{aligned}
E\left[D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})\right] &= E\left[\frac{(-1)^{|\boldsymbol{k}|}}{nh_n^{d+|\boldsymbol{k}|}}\sum_{i=1}^n D^{\boldsymbol{k}}K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h_n}\right)\right]\\
&= (-1)^{|\boldsymbol{k}|}\int_{\mathbb{R}^d}\frac{1}{h_n^{d+|\boldsymbol{k}|}}D^{\boldsymbol{k}}K\left(\frac{\boldsymbol{u}-\boldsymbol{x}}{h_n}\right)f(\boldsymbol{u})d\boldsymbol{u}\\
&= (-1)^{|\boldsymbol{k}|}\int_{\mathbb{R}^d}\frac{1}{h_n^{|\boldsymbol{k}|}}D^{\boldsymbol{k}}K(\boldsymbol{z})f(\boldsymbol{z}h_n+\boldsymbol{x})d\boldsymbol{z}\\
&= (-1)^{2|\boldsymbol{k}|}\int_{\mathbb{R}^d}K(\boldsymbol{z})D^{\boldsymbol{k}}f(\boldsymbol{z}h_n+\boldsymbol{x})d\boldsymbol{z}.
\end{aligned}$$

For the last equality integration by parts was applied $|\boldsymbol{k}|$-times, as well as the assumptions on the kernel and its derivatives in **(K)** were used. More precisely, it is needed that the kernel and its derivatives vanish on the edge of their support. Note that this follows as K is $l+1$ times differentiable and has compact support. Taylor's expansion of $D^{\boldsymbol{k}}f$ in $\boldsymbol{x}$ up to order $r-1$ with Lagrange form of the remainder implies

$$\begin{aligned}
D^{\boldsymbol{k}}f(\boldsymbol{z}h_n+\boldsymbol{x})-D^{\boldsymbol{k}}f(\boldsymbol{x}) &= \sum_{|\boldsymbol{i}|=1}^{r-1}\frac{D^{\boldsymbol{i}}D^{\boldsymbol{k}}f(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}+\sum_{|\boldsymbol{i}|=r}\frac{D^{\boldsymbol{i}}D^{\boldsymbol{k}}f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}\\
&= \sum_{|\boldsymbol{i}|=1}^{r-1}\frac{D^{\boldsymbol{i}+\boldsymbol{k}}f(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}+\sum_{|\boldsymbol{i}|=r}\frac{D^{\boldsymbol{i}+\boldsymbol{k}}f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}
\end{aligned}$$

for some $\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{z}h_n$. Assumptions **(K)** and **(F1)** then imply

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|E\left[D^{\boldsymbol{k}}\hat{f}_n(\boldsymbol{x})\right]-D^{\boldsymbol{k}}f(\boldsymbol{x})\right|$$

$$
\begin{aligned}
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) D^{\boldsymbol{k}} f(\boldsymbol{z}h_n + \boldsymbol{x}) d\boldsymbol{z} - D^{\boldsymbol{k}} f(\boldsymbol{x}) \right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(D^{\boldsymbol{k}} f(\boldsymbol{z}h_n + \boldsymbol{x}) - D^{\boldsymbol{k}} f(\boldsymbol{x}) \right) d\boldsymbol{z} \right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(\sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{x}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right) d\boldsymbol{z} \right| \\
&\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{x}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z} \right| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z} \right| \\
&\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{x}) h_n^{|\boldsymbol{i}|}}{\boldsymbol{i}!} \underbrace{\int_{\mathbb{R}^d} \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) d\boldsymbol{z}}_{\substack{=0 \\ \forall 1 \leq |\boldsymbol{i}| \leq r-1}} \right| + h_n^r \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \sum_{|\boldsymbol{i}|=r} \left| \frac{D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})}{\boldsymbol{i}!} \right| \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \left| D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) \right| \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \int_{\mathbb{R}^d} \underbrace{\left| D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) \right| I\{\boldsymbol{z} \in [-C,C]^d\}}_{\leq \sup_{\boldsymbol{y}\in[-C,C]^d} |D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}})|} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\substack{\boldsymbol{x}\in\boldsymbol{J}_n \\ \boldsymbol{y}\in[-C,C]^d}} \left| D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}}) \right| \int_{\mathbb{R}^d} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{I}_n} \left| D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{x}) \right|}_{\substack{=O(p_n) \\ \forall |\boldsymbol{i}|=r, 1 \leq |\boldsymbol{k}| \leq l+1}} \underbrace{\int_{\mathbb{R}^d} \left| \boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z}) \right| d\boldsymbol{z}}_{<\infty} \\
&= O(h_n^r p_n),
\end{aligned}
$$

which completes the proof of (i) (b).

The outline of the proof of (ii) (a) is similar to the one before. Using

$$
\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})| \leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{g}_n(\boldsymbol{x}) - E[\hat{g}_n(\boldsymbol{x})]| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |E[\hat{g}_n(\boldsymbol{x})] - g(\boldsymbol{x})|,
$$

it again splits into two parts. Concerning the left term, an application of Theorem 2 in [26] yields

$$
\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{g}_n(\boldsymbol{x}) - E[\hat{g}_n(\boldsymbol{x})]| = O_P\left(\sqrt{\frac{\log(n)}{n h_n^d}} \right).
$$

For the right term, the law of total expectation and integration by substitution is used. Moreover, Taylor's expansion of g in $\boldsymbol{x}$ up to order $r-1$ results in

$$
g(\boldsymbol{z}h_n + \boldsymbol{x}) - g(\boldsymbol{x}) = \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} g(\boldsymbol{x}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) (\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}
$$

for some $\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{z}h_n$. Applying the assumptions in **(K)** and **(F1)**, it follows that

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |E[\hat{g}_n(\boldsymbol{x})] - g(\boldsymbol{x})| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| E\left[\frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h_n}\right) Y_i\right] - g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| E\left[\frac{1}{nh_n^d}\sum_{i=1}^n K\left(\frac{\boldsymbol{X}_i-\boldsymbol{x}}{h_n}\right) \underbrace{E[Y_i|\boldsymbol{X}_i]}_{=m(\boldsymbol{X}_i)\text{ a.s.}}\right] - g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} \frac{1}{h_n^d} K\left(\frac{\boldsymbol{u}-\boldsymbol{x}}{h_n}\right) \underbrace{m(\boldsymbol{u})f(\boldsymbol{u})}_{=g(\boldsymbol{u})} d\boldsymbol{u} - g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z})g(\boldsymbol{z}h_n+\boldsymbol{x})d\boldsymbol{z} - g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z})\left(g(\boldsymbol{z}h_n+\boldsymbol{x}) - g(\boldsymbol{x})\right)d\boldsymbol{z}\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z})\left(\sum_{|\boldsymbol{i}|=1}^{r-1}\frac{D^{\boldsymbol{i}}g(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r}\frac{D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}\right)d\boldsymbol{z}\right| \\
&\le \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z})\sum_{|\boldsymbol{i}|=1}^{r-1}\frac{D^{\boldsymbol{i}}g(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}d\boldsymbol{z}\right| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z})\sum_{|\boldsymbol{i}|=r}\frac{D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}d\boldsymbol{z}\right| \\
&\le \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \sum_{|\boldsymbol{i}|=1}^{r-1}\frac{D^{\boldsymbol{i}}g(\boldsymbol{x})h_n^{|\boldsymbol{i}|}}{\boldsymbol{i}!}\underbrace{\int_{\mathbb{R}^d}\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})d\boldsymbol{z}}_{\substack{=0\\ \forall\ 1\le|\boldsymbol{i}|\le r-1}}\right| + h_n^r\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\int_{\mathbb{R}^d}\sum_{|\boldsymbol{i}|=r}\left|\frac{D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})}{\boldsymbol{i}!}\right|\left|\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})\right|d\boldsymbol{z} \\
&= h_n^r\sum_{|\boldsymbol{i}|=r}\frac{1}{\boldsymbol{i}!}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\int_{\mathbb{R}^d}\left|D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})\right|\left|\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})\right|d\boldsymbol{z} \\
&= h_n^r\sum_{|\boldsymbol{i}|=r}\frac{1}{\boldsymbol{i}!}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\int_{\mathbb{R}^d}\underbrace{\left|D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})\right|I\{\boldsymbol{z}\in[-C,C]^d\}}_{\le\sup_{\boldsymbol{y}\in[-C,C]^d}\left|D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}})\right|}\left|\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})\right|d\boldsymbol{z} \\
&\le h_n^r\sum_{|\boldsymbol{i}|=r}\frac{1}{\boldsymbol{i}!}\sup_{\substack{\boldsymbol{x}\in\boldsymbol{J}_n\\ \boldsymbol{y}\in[-C,C]^d}}\left|D^{\boldsymbol{i}}g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}})\right|\int_{\mathbb{R}^d}\left|\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})\right|d\boldsymbol{z} \\
&\le h_n^r\sum_{|\boldsymbol{i}|=r}\frac{1}{\boldsymbol{i}!}\underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{I}_n}\left|D^{\boldsymbol{i}}g(\boldsymbol{x})\right|}_{\substack{\overset{(*)}{=}O(p_nq_n)\\ \forall|\boldsymbol{i}|=r}}\underbrace{\int_{\mathbb{R}^d}\left|\boldsymbol{z}^{\boldsymbol{i}}K(\boldsymbol{z})\right|d\boldsymbol{z}}_{\substack{<\infty\\ \forall|\boldsymbol{i}|=r}} \\
&= O(h_n^r p_n q_n),
\end{aligned}
$$

where $\sup_{\boldsymbol{x}\in\boldsymbol{I}_n}\left|D^{\boldsymbol{j}}m(\boldsymbol{x})\right| = O(q_n)$ and $\sup_{\boldsymbol{x}\in\boldsymbol{I}_n}\left|D^{\boldsymbol{j}}f(\boldsymbol{x})\right| = O(p_n)$ $\forall\boldsymbol{j}\le\boldsymbol{i}$ together imply $(*)$ for all $|\boldsymbol{i}|=r$. Hence, the assertion in (ii) (a) is shown.

For (ii) (b) let $\boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq l+1$. Again, by chain rule it holds that

$$D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) = \frac{(-1)^{|\boldsymbol{k}|}}{n h_n^{d+|\boldsymbol{k}|}} \sum_{i=1}^{n} D^{\boldsymbol{k}} K \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right) Y_i. \tag{2.11}$$

With

$$\begin{aligned}
&\sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) - D^{\boldsymbol{k}} g(\boldsymbol{x}) \right| \\
&\leq \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) - E\left[D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) \right] \right| + \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| E\left[D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) \right] - D^{\boldsymbol{k}} g(\boldsymbol{x}) \right|
\end{aligned}$$

the proof of (ii) (b) splits into two parts again, which will be treated separately, beginning with the left term. Denoting $\hat{\Psi}(\boldsymbol{x}) := \frac{1}{n h_n^d} \sum_{i=1}^n D^{\boldsymbol{k}} K \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right) Y_i$, it follows that

$$\begin{aligned}
\sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) - E\left[D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) \right] \right| &= \frac{1}{h_n^{|\boldsymbol{k}|}} \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| \hat{\Psi}(\boldsymbol{x}) - E\left[\hat{\Psi}(\boldsymbol{x}) \right] \right| \\
&= O_P\left(\sqrt{\frac{\log(n)}{n h_n^{d+2|\boldsymbol{k}|}}} \right),
\end{aligned}$$

where the last step is again an application of Theorem 2 in [26]. Note that the assumptions for the used Theorem are satisfied by the assumptions of the Lemma. Concerning the right term, inserting the presentation of $D^{\boldsymbol{k}} \hat{g}_n$ of (2.11) and using the law of total expectation and integration by substitution, it can be obtained that

$$\begin{aligned}
E\left[D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x}) \right] &= E\left[\frac{(-1)^{|\boldsymbol{k}|}}{n h_n^{d+|\boldsymbol{k}|}} \sum_{i=1}^{n} D^{\boldsymbol{k}} K \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right) Y_i \right] \\
&= E\left[\frac{(-1)^{|\boldsymbol{k}|}}{n h_n^{d+|\boldsymbol{k}|}} \sum_{i=1}^{n} D^{\boldsymbol{k}} K \left(\frac{\boldsymbol{X}_i - \boldsymbol{x}}{h_n} \right) \underbrace{E[Y_i | \boldsymbol{X}_i]}_{= m(\boldsymbol{X}_i) \text{ a.s.}} \right] \\
&= (-1)^{|\boldsymbol{k}|} \int_{\mathbb{R}^d} \frac{1}{h_n^{d+|\boldsymbol{k}|}} D^{\boldsymbol{k}} K \left(\frac{\boldsymbol{u} - \boldsymbol{x}}{h_n} \right) \underbrace{m(\boldsymbol{u}) f(\boldsymbol{u})}_{= g(\boldsymbol{u})} d\boldsymbol{u} \\
&= (-1)^{|\boldsymbol{k}|} \int_{\mathbb{R}^d} \frac{1}{h_n^{|\boldsymbol{k}|}} D^{\boldsymbol{k}} K(\boldsymbol{z}) g(\boldsymbol{x} + \boldsymbol{z} h_n) d\boldsymbol{z} \\
&= (-1)^{2|\boldsymbol{k}|} \int_{\mathbb{R}^d} K(\boldsymbol{z}) D^{\boldsymbol{k}} g(\boldsymbol{z} h_n + \boldsymbol{x}) d\boldsymbol{z}.
\end{aligned}$$

The last equality follows again by integration by parts $|\boldsymbol{k}|$-times and the assumptions on the kernel and its derivatives in **(K)**. Taylor's expansion of $D^{\boldsymbol{k}} g$ in $\boldsymbol{x}$ up to order $r-1$ with Lagrange form of the remainder yields

$$D^{\boldsymbol{k}} g(\boldsymbol{z} h_n + \boldsymbol{x}) - D^{\boldsymbol{k}} g(\boldsymbol{x}) = \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}} D^{\boldsymbol{k}} g(\boldsymbol{x}) (\boldsymbol{z} h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}} D^{\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}) (\boldsymbol{z} h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}$$

$$= \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}$$

for some $\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}}$ on the line segment between $\boldsymbol{x}$ and $\boldsymbol{x}+\boldsymbol{z}h_n$. Using additionally the assumptions **(K)** and **(F1)**, it follows that

$$\begin{aligned}
&\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| E\left[D^{\boldsymbol{k}} \hat{g}_n(\boldsymbol{x})\right] - D^{\boldsymbol{k}} g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) D^{\boldsymbol{k}} g(\boldsymbol{z}h_n + \boldsymbol{x}) d\boldsymbol{z} - D^{\boldsymbol{k}} g(\boldsymbol{x})\right| \\
&= \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(D^{\boldsymbol{k}} g(\boldsymbol{z}h_n + \boldsymbol{x}) - D^{\boldsymbol{k}} g(\boldsymbol{x})\right) d\boldsymbol{z}\right| \\
&= \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \left(\sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}\right) d\boldsymbol{z}\right| \\
&\leq \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{x})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z}\right| + \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \int_{\mathbb{R}^d} K(\boldsymbol{z}) \sum_{|\boldsymbol{i}|=r} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})(\boldsymbol{z}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} d\boldsymbol{z}\right| \\
&\leq \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \sum_{|\boldsymbol{i}|=1}^{r-1} \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{x}) h_n^{|\boldsymbol{i}|}}{\boldsymbol{i}!} \underbrace{\int_{\mathbb{R}^d} z^{\boldsymbol{i}} K(\boldsymbol{z}) d\boldsymbol{z}}_{\substack{=0 \\ \forall 1\leq |\boldsymbol{i}|\leq r-1}} \right| + h_n^r \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \int_{\mathbb{R}^d} \sum_{|\boldsymbol{i}|=r} \left| \frac{D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})}{\boldsymbol{i}!}\right| \left|\boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z})\right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \int_{\mathbb{R}^d} \left| D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})\right| \left|\boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z})\right| d\boldsymbol{z} \\
&= h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \int_{\mathbb{R}^d} \underbrace{\left| D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{z}})\right| I\{\boldsymbol{z}\in[-C,C]^d\}}_{\leq \sup_{\boldsymbol{y}\in[-C,C]^d} |D^{\boldsymbol{i}+\boldsymbol{k}} f(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}})|} \left|\boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z})\right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \sup_{\substack{\boldsymbol{x}\in \boldsymbol{J}_n \\ \boldsymbol{y}\in[-C,C]^d}} \left| D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{y}})\right| \int_{\mathbb{R}^d} \left|\boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z})\right| d\boldsymbol{z} \\
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \frac{1}{|\boldsymbol{i}|} \underbrace{\sup_{\boldsymbol{x}\in \boldsymbol{I}_n} \left| D^{\boldsymbol{i}+\boldsymbol{k}} g(\boldsymbol{x})\right|}_{\substack{\overset{(*)}{=} O(p_n q_n) \\ \forall |\boldsymbol{i}|=r, 1\leq |\boldsymbol{k}|\leq l+1}} \underbrace{\int_{\mathbb{R}^d} \left|\boldsymbol{z}^{\boldsymbol{i}} K(\boldsymbol{z})\right| d\boldsymbol{z}}_{<\infty} \\
&= O(h_n^r p_n q_n),
\end{aligned}$$

where $\sup_{\boldsymbol{x}\in \boldsymbol{I}_n} \left|D^{\boldsymbol{j}+\boldsymbol{k}} m(\boldsymbol{x})\right| = O(q_n)$ and $\sup_{\boldsymbol{x}\in \boldsymbol{I}_n} \left|D^{\boldsymbol{j}+\boldsymbol{k}} f(\boldsymbol{x})\right| = O(p_n)$ for all $\boldsymbol{j} \leq \boldsymbol{i}, 1 \leq |\boldsymbol{k}| \leq l+1$ together imply $(*)$ for all $|\boldsymbol{i}| = r$ and $1 \leq |\boldsymbol{k}| \leq l+1$. Hence, the assertion in (ii) (b) is shown.

For the proof of (iii), the results from (i) and (ii) will be used. Concerning (iii) (a), it can be obtained that

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})| = \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \frac{\hat{g}_n(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} - m(\boldsymbol{x})\right| = \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| \frac{\frac{\hat{g}_n(\boldsymbol{x})}{f(\boldsymbol{x})} - \frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} m(\boldsymbol{x})}{\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}}\right|$$

$$= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})m(\boldsymbol{x})\right)}{\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}} \right|$$
$$\leq \frac{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})m(\boldsymbol{x})\right) \right|}{\inf_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right|}.$$

The numerator and denominator will be considered separately, beginning with the numerator. By adding and subtracting $\frac{g(\boldsymbol{x})}{f(\boldsymbol{x})}$ and using $g(\boldsymbol{x}) = m(\boldsymbol{x})f(\boldsymbol{x})$, it can be obtained that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})m(\boldsymbol{x})\right) \right|$$
$$\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) \right| + \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right) \right|$$
$$\leq \underbrace{\frac{1}{\inf_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x})}}_{=O(\delta_n)} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right|}_{\overset{(ii)(a)}{=} O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n q_n\right)} + \underbrace{\frac{1}{\inf_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x})}}_{=O(\delta_n)} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |m(\boldsymbol{x})|}_{=O(q_n)} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left|f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right|}_{\overset{(i)(a)}{=} O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right)}$$
$$= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right) q_n \delta_n\right).$$

Note that the last equality does not hold if $q_n = 0$ for all $n \in \mathbb{N}$. Instead, in this case it can be obtained that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})m(\boldsymbol{x})\right) \right| = O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}} \delta_n\right).$$

It is left to show that the denominator is bounded away from zero which can be seen in the following way,

$$\inf_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right| = \inf_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| 1 - \frac{f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right|$$
$$\geq \inf_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| 1 - \left| \frac{f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right| \right|$$
$$\geq \inf_{\boldsymbol{x}\in\boldsymbol{J}_n} \left(1 - \left| \frac{f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right| \right)$$
$$= 1 - \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| \frac{f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})} \right|$$
$$\geq 1 - \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} \left| f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x}) \right|}_{\overset{(i)(a)}{=} O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right)} \underbrace{\frac{1}{\inf_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x})}}_{=O(\delta_n)}$$

$$= 1 + O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right)\delta_n\right)$$
$$= 1 + o_P(1),$$

where the last equality is implied by condition (2.9) in **(B2)**. Putting these results together,

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})| = \frac{O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right) q_n\delta_n\right)}{1 + o_P(1)}$$
$$= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}} + h_n^r p_n\right) q_n\delta_n\right)$$

completes the proof of (iii) (a). Note that for $q_n = 0$ for all $n \in \mathbb{N}$ (i.e. for $m \equiv 0$), it can be obtained that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})| = O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}\delta_n\right). \tag{2.12}$$

For (iii) (b) let $\boldsymbol{k} \in \mathbb{N}_0^d$ with $1 \leq |\boldsymbol{k}| \leq l + 1$. As seen in the first part, it holds that

$$\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x}) = \left(\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) + \frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right)\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-1}.$$

Applying Leibniz's formula (2.5), it can be obtained that

$$D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)$$
$$= \sum_{\boldsymbol{i}\leq\boldsymbol{k}} D^{\boldsymbol{i}}\left(\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) + \frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right) D^{\boldsymbol{k}-\boldsymbol{i}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-1}. \tag{2.13}$$

The two factors in the summands of (2.13) will be treated separately, beginning with the left one. A repeated application of Leibniz's formula yields

$$D^{\boldsymbol{i}}\left(\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) + \frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right)$$
$$= D^{\boldsymbol{i}}\left(\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right)\right) + D^{\boldsymbol{i}}\left(\frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right)$$
$$= \sum_{\boldsymbol{j}\leq\boldsymbol{i}} D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})}\right) D^{\boldsymbol{i}-\boldsymbol{j}}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) + \sum_{\boldsymbol{j}\leq\boldsymbol{i}} D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\right) D^{\boldsymbol{i}-\boldsymbol{j}}\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right).$$

Additionally, for all $\boldsymbol{j} \leq \boldsymbol{i} \leq \boldsymbol{k}$ it holds that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})}\right)\right| = O(p_n^{|\boldsymbol{j}|}\delta_n^{|\boldsymbol{j}|+1}). \tag{2.14}$$

To see this, using both product and chain rule, it can be obtained that $D^{\boldsymbol{j}}\left(f(\boldsymbol{x})\right)^{-1}$ is a sum of products, each containing one factor of the form $(f(\boldsymbol{x}))^{-s}$ for some $s \in \mathbb{N}$ with $2 \le s \le |\boldsymbol{j}|+1$ and at most $|\boldsymbol{j}|$ factors being of the form $D^{\boldsymbol{s}} f(\boldsymbol{x})$ for some $\boldsymbol{s} \in \mathbb{N}_0^d$ with $\boldsymbol{s} \le \boldsymbol{j}$. Using

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \frac{1}{f(\boldsymbol{x})^s} = \frac{1}{\inf\limits_{\boldsymbol{x}\in \boldsymbol{J}_n} f(\boldsymbol{x})^s} = O(\delta_n^s),\ \forall\ s \in \mathbb{N},\ 2 \le s \le |\boldsymbol{j}|+1$$

and

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} |D^{\boldsymbol{s}} f(\boldsymbol{x})| \le \sup_{\boldsymbol{x}\in \boldsymbol{I}_n} |D^{\boldsymbol{s}} f(\boldsymbol{x})| = O(p_n),\ \forall\ \boldsymbol{s} \in \mathbb{N}_0^d, \boldsymbol{s} \le \boldsymbol{j}$$

from assumption **(F1)** therefore results in (2.14). Furthermore,

$$\begin{aligned}\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\right)\right| &\le \sum_{\boldsymbol{s}\le \boldsymbol{j}} \underbrace{\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{s}}\left(\frac{1}{f(\boldsymbol{x})}\right)\right|}_{\overset{(2.14)}{=} O(p_n^{|\boldsymbol{s}|}\delta_n^{|\boldsymbol{s}|+1})} \underbrace{\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{j}-\boldsymbol{s}} m(\boldsymbol{x})\right|}_{\overset{\textbf{(F1)}}{=} O(q_n)} \\ &= O(p_n^{|\boldsymbol{j}|} q_n \delta_n^{|\boldsymbol{j}|+1})\end{aligned}$$

holds. Together with (i) (b) and (ii) (b), it can be concluded that for all $\boldsymbol{i} \le \boldsymbol{k}$

$$\begin{aligned}&\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{i}}\left(\frac{1}{f(\boldsymbol{x})}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right) + \frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right)\right| \\ &\le \sum_{\boldsymbol{j}\le \boldsymbol{i}} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})}\right)\right| \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{i}-\boldsymbol{j}}\left(\hat{g}_n(\boldsymbol{x}) - g(\boldsymbol{x})\right)\right| \\ &+ \sum_{\boldsymbol{j}\le \boldsymbol{i}} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{j}}\left(\frac{1}{f(\boldsymbol{x})} m(\boldsymbol{x})\right)\right| \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{i}-\boldsymbol{j}}\left(f(\boldsymbol{x}) - \hat{f}_n(\boldsymbol{x})\right)\right| \\ &= \sum_{\boldsymbol{j}\le \boldsymbol{i}} O(p_n^{|\boldsymbol{j}|}\delta_n^{|\boldsymbol{j}|+1}) O_P\left(\sqrt{\frac{\log(n)}{n h_n^{d+2|\boldsymbol{i}-\boldsymbol{j}|}}} + h_n^r p_n q_n\right) \\ &+ \sum_{\boldsymbol{j}\le \boldsymbol{i}} O(p_n^{|\boldsymbol{j}|} q_n \delta_n^{|\boldsymbol{j}|+1}) O_P\left(\sqrt{\frac{\log(n)}{n h_n^{d+2|\boldsymbol{i}-\boldsymbol{j}|}}} + h_n^r p_n\right) \\ &= O_P\left(\left(\sqrt{\frac{\log(n)}{n h_n^{d+2|\boldsymbol{i}|}}} + h_n^r p_n\right) p_n^{|\boldsymbol{i}|} q_n \delta_n^{|\boldsymbol{i}|+1}\right).\end{aligned}$$

Concerning the second factor in the summands of (2.13), it is left to show for all $\boldsymbol{i} \le \boldsymbol{k}$

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left| D^{\boldsymbol{i}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-1}\right| = O_P(1). \tag{2.15}$$

Again, by product and chain rule it can be seen that $D^{\boldsymbol{i}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-1}$ is a sum of products, each containing one factor of the form $\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-j}$ for some $j \in \mathbb{N}$ with

$2 \leq j \leq |\boldsymbol{i}|+1$ and at most $|\boldsymbol{i}|$ factors being of the form $D^{\boldsymbol{j}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)$ for some $\boldsymbol{j} \in \mathbb{N}_0^d$ with $\boldsymbol{j} \leq \boldsymbol{i}$. The assertion of (2.15) is therefore an immediate consequence of

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-j}\right| = O_P(1),\ \forall\, j \in \mathbb{N}, 2 \leq j \leq |\boldsymbol{i}|+1 \tag{2.16}$$

and

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)\right| = O_P(1),\ \forall\, \boldsymbol{j} \in \mathbb{N}_0^d, \boldsymbol{j} \leq \boldsymbol{i}. \tag{2.17}$$

Concerning (2.16), let $j \in \mathbb{N}$ with $2 \leq j \leq |\boldsymbol{i}|+1$. Because of

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-j}\right| = \frac{1}{\inf\limits_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{j}\right|},$$

it is to show that the denominator of the term on the right hand side is bounded away from zero. Applying the binomial theorem, it can be seen that

$$\begin{aligned}
\inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{j}\right| &= \inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(1+\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{j}\right| \\
&= \inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\sum_{s=0}^{j}\binom{j}{s}1^{j-s}\left(\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{s}\right| \\
&= \inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|1+\sum_{s=1}^{j}\binom{j}{s}\left(\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{s}\right| \\
&\geq \inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|1-\sum_{s=1}^{j}\binom{j}{s}\left|\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right|^{s}\right| \\
&\geq \inf_{\boldsymbol{x}\in\boldsymbol{J}_n}\left(1-\sum_{s=1}^{j}\binom{j}{s}\left|\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right|^{s}\right) \\
&= 1-\sum_{s=1}^{j}\binom{j}{s}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right|^{s} \\
&\geq 1-\sum_{s=1}^{j}\binom{j}{s}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})\right|^{s}\frac{1}{\inf\limits_{\boldsymbol{x}\in\boldsymbol{J}_n} f(\boldsymbol{x})^{s}} \\
&= 1+\sum_{s=1}^{j}\binom{j}{s}O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right)^{s} O(\delta_n^s) \\
&= 1+O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right) O(\delta_n^j) \\
&= 1+o_P(1),
\end{aligned}$$

where the last equality is implied by condition (2.9) in **(B2)**. Therefore,

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)^{-j}\right| = \frac{1}{1+o_P(1)} = O_P(1)$$

holds which is the assertion in (2.16). Finally, the statement in (2.17) will be proven. Let $\boldsymbol{j}\in\mathbb{N}_0^d$ with $\boldsymbol{j}\leq\boldsymbol{i}$. Then

$$\begin{aligned}
\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}}\left(\frac{\hat{f}_n(\boldsymbol{x})}{f(\boldsymbol{x})}\right)\right| &= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}}\left(1+\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right)\right| \\
&= \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}}\left(\frac{\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})}{f(\boldsymbol{x})}\right)\right| \\
&= \sum_{\boldsymbol{s}\leq\boldsymbol{j}}\binom{\boldsymbol{j}}{\boldsymbol{s}}\underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{s}}\left(\hat{f}_n(\boldsymbol{x})-f(\boldsymbol{x})\right)\right|}_{\overset{\text{(i),(b)}}{=}O_P\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{s}|}}}+h_n^r p_n\right)}\underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{j}-\boldsymbol{s}}\left(\frac{1}{f(\boldsymbol{x})}\right)\right|}_{\overset{(2.14)}{=}O(p_n^{|\boldsymbol{j}-\boldsymbol{s}|}\delta_n^{|\boldsymbol{j}-\boldsymbol{s}|+1})} \\
&= \sum_{\boldsymbol{s}\leq\boldsymbol{j}}\binom{\boldsymbol{j}}{\boldsymbol{s}}O_P\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{s}|}}}+h_n^r p_n\right)O(p_n^{|\boldsymbol{j}-\boldsymbol{s}|}\delta_n^{|\boldsymbol{j}-\boldsymbol{s}|+1}) \\
&= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{j}|}}}+h_n^r p_n\right)p_n^{|\boldsymbol{j}|}\delta_n^{|\boldsymbol{j}|+1}\right) \\
&= O_P(1),
\end{aligned}$$

where the last equality holds for all $0\leq|\boldsymbol{j}|\leq l+1$ due to condition (2.8) in **(B2)**. Coming back to (2.13), it has been proven that

$$\begin{aligned}
\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x})-m(\boldsymbol{x})\right)\right| &= \sum_{\boldsymbol{i}\leq\boldsymbol{k}}O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{i}|}}}+h_n^r p_n\right)p_n^{|\boldsymbol{i}|}q_n\delta_n^{|\boldsymbol{i}|+1}\right)O_P(1) \\
&= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{k}|}}}+h_n^r p_n\right)p_n^{|\boldsymbol{k}|}q_n\delta_n^{|\boldsymbol{k}|+1}\right),
\end{aligned}$$

for all $1\leq|\boldsymbol{k}|\leq l+1$ which is the statement of (iii) (b). Note that in the case of $q_n=0$ for all $n\in\mathbb{N}$, it can be obtained that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x})-m(\boldsymbol{x})\right)\right| = O_P\left(\sqrt{\frac{\log(n)}{nh_n^{d+2|\boldsymbol{k}|}}}p_n^{|\boldsymbol{k}|}\delta_n^{|\boldsymbol{k}|+1}\right),\ \forall\ 1\leq|\boldsymbol{k}|\leq l+1.$$

For (iii) (c) let $\boldsymbol{k}\in\mathbb{N}_0^d$ with $|\boldsymbol{k}|=l$. Then

$$\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in\boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}}\frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x})-m(\boldsymbol{x})\right)-D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y})-m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}}$$

$$= \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| > (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}} \tag{2.18}$$

$$+ \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| \leq (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}}. \tag{2.19}$$

Both (2.18) and (2.19) will be investigated separately. Starting with (2.18), it can be obtained that

$$\begin{aligned}
&\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| > (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}} \\
&\leq (p_n\delta_n)^{\eta} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| > (p_n\delta_n)^{-1}}} \left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right| \\
&\leq 2(p_n\delta_n)^{\eta} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)\right| \\
&\overset{(iii)(b)}{=} 2(p_n\delta_n)^{\eta} O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2l}}} + h_n^r p_n\right) p_n^l q_n \delta_n^{l+1}\right) \\
&= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2l}}} + h_n^r p_n\right) p_n^{l+\eta} q_n \delta_n^{l+1+\eta}\right) \\
&= o_P(1),
\end{aligned}$$

where the last equality is implied by (2.9) in **(B2)**. Concerning (2.19), the mean value theorem (see for instance [36], p. 56) will be used to obtain

$$\begin{aligned}
&\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| \leq (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}} \\
&= \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| \leq (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta-1}\|\boldsymbol{x}-\boldsymbol{y}\|} \\
&\leq (p_n\delta_n)^{\eta-1} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n, \boldsymbol{x}\neq \boldsymbol{y} \\ \|\boldsymbol{x}-\boldsymbol{y}\| \leq (p_n\delta_n)^{-1}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|} \\
&\leq (p_n\delta_n)^{\eta-1} \max_{|\boldsymbol{i}|=l+1} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left|D^{\boldsymbol{i}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right)\right| \\
&\overset{(iii)(b)}{=} (p_n\delta_n)^{\eta-1} O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+1} q_n \delta_n^{l+2}\right) \\
&= O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}} + h_n^r p_n\right) p_n^{l+\eta} q_n \delta_n^{l+1+\eta}\right) \\
&= o_P(1),
\end{aligned}$$

where the last equality holds due to condition (2.9) in **(B2)**. Note that in the particular case of $q_n = 0$ for all $n \in \mathbb{N}$ it can be obtained that

$$\sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{x}) - m(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{m}_n(\boldsymbol{y}) - m(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}} = O_P\left(\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}}p_n^{l+\eta}\delta_n^{l+1+\eta}\right)$$
$$= o_P(1),$$

for all $|\boldsymbol{k}| = l$.

□

Remark. Results for $\hat{\sigma}_n^2$ can be obtained in a similar matter. To see this, let the following notations be introduced. Let

$$\tilde{g}(\boldsymbol{x}) := E[Y_i^2|\boldsymbol{X}_i = \boldsymbol{x}]f(\boldsymbol{x})$$

and

$$\hat{\tilde{g}}_n(\boldsymbol{x}) := \frac{1}{n}\sum_{i=1}^{n} K_{h_n}(\boldsymbol{X}_i - \boldsymbol{x})Y_i^2.$$

Then it holds that

$$\sigma^2(\boldsymbol{x}) = E[Y_i^2|\boldsymbol{X}_i = \boldsymbol{x}] - E[Y_i|\boldsymbol{X}_i = \boldsymbol{x}]^2 = \frac{\tilde{g}(\boldsymbol{x})}{f(\boldsymbol{x})} - m^2(\boldsymbol{x})$$

and

$$\hat{\sigma}_n^2(\boldsymbol{x}) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} K_{h_n}(\boldsymbol{X}_i - \boldsymbol{x})Y_i^2}{\hat{f}_n(\boldsymbol{x})} - \hat{m}_n^2(\boldsymbol{x}) = \frac{\hat{\tilde{g}}_n(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} - \hat{m}_n^2(\boldsymbol{x}),$$

for $\hat{f}_n(\boldsymbol{x}) \neq 0$. Thus, it holds that

$$\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x}) = \left(\frac{\hat{\tilde{g}}_n(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} - \frac{\tilde{g}(\boldsymbol{x})}{f(\boldsymbol{x})}\right) + \left(m^2(\boldsymbol{x}) - \hat{m}_n^2(\boldsymbol{x})\right).$$

To obtain uniform rates of convergence for the first summand, additional assumptions, that imply uniform convergence rates for the difference $\hat{\tilde{g}}_n - \tilde{g}$ and its partial derivatives, need to be made. In particular, the moment assumptions in **(M)** need to be extended to Y_1^2 and the smoothness assumptions in **(F1)** need to be extended to σ^2. Under suitable conditions, it then can be shown that

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left|\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})\right| = o_P(1),$$

$$\sup_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\ 1\leq|\boldsymbol{k}|\leq l}} \sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left|D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})\right)\right| = o_P(1)$$

and

$$\max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\ |\boldsymbol{k}|=l}} \sup_{\substack{\boldsymbol{x},\boldsymbol{y}\in \boldsymbol{J}_n\\ \boldsymbol{x}\neq\boldsymbol{y}}} \frac{\left|D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{x})\right) - D^{\boldsymbol{k}}\left(\hat{\sigma}_n^2(\boldsymbol{y}) - \sigma^2(\boldsymbol{y})\right)\right|}{\|\boldsymbol{x}-\boldsymbol{y}\|^{\eta}} = o_P(1).$$

These properties are needed to construct a test for change in the conditional variance function that is motivated in Chapter 5. The main part of this thesis, however, is the construction of a changepoint test in the conditional mean function in Chapter 3. Rates of convergence for $\hat{\sigma}_n^2$ are not required for this part and a more detailed discussion is therefore omitted.

2.3 Related results

Related results to Lemma 2.2 can be found in several papers that use uniform rates of convergence for kernel regression estimators and possibly its derivatives. To start with the simplest case, the paper of Akritas and Van Keilegom [1] is to mention. For $d = 1$ and a sequence of i.i.d. random variables, they showed uniform convergence rates for the Nadaraya-Watson estimator and its first derivative, as well as a Lipschitz condition for the first derivative comparable to assertion (iii) (c) in Lemma 2.2 (see Proposition 3,4 and 5 in [1]). For $d \geq 1$ Neumeyer and Van Keilegom [55] proved similar results for the local polynomial kernel regression estimator and its partial derivatives up to order d in an i.i.d. model (see Lemma A.1 in [55]). Concerning the dependence structure of the underlying process, Dette, Pardo-Fernández and Van Keilegom [17] extended existing results for $d = 1$ to β-mixing strictly stationary sequences. They showed uniform rates of convergence for the Nadaraya-Watson estimator and its first derivative, as well as the Lipschitz condition for the first derivative (see proof of Lemma 1 in [17]). Selk and Neumeyer [66] showed similar results for $d = 1$ and an α-mixing, not necessarily stationary underlying process (see Lemma 1 in [66]). Using local polynomial regression estimation, Neumeyer, Omelka and Hudecova were able to show results for $d \geq 1$ and β-mixing strictly stationary data (see Lemma 1 in [56]).

The paper most related to Lemma 2.2 is possibly Hansen's [26] and therefore the results will be compared in more detail. Theorem 2 in [26] gives uniform rates of convergence for general kernel estimators under a possibly multidimensional and α-mixing underlying process. As an application, Hansen proves uniform rates of convergence for the kernel density estimator and its partial derivatives comparable to the results in Lemma 2.2 (i) (a) and (b) (see Theorem 6 in [26]). In comparison to Hansen's proof, the result in Lemma 2.2 holds without the assumption of uniform bounded partial derivatives of $f(\cdot)$. Instead, the less restrictive condition on p_n in **(F1)** is imposed. Further, the proof is more clear in notations regarding the partial derivatives using multi-index notations. Uniform rates for the Nadaraya-Watson estimator as assertion (iii) (a) in Lemma 2.2 were also given in Theorem 8 in [26]. The new result is an improvement insofar as a faster rate can be obtained by imposing the existence of more partial derivatives and by choosing a kernel of higher order. Hansen assumed the existence of second order partial derivatives and a kernel of order two and therefore obtains a bias of order $O(h_n^2)$. Furthermore, by imposing assumptions on p_n and q_n in **(F1)** the assumption on uniformly bounded partial derivatives of $f(\cdot)$ and $f(\cdot)m(\cdot)$ made in his Theorem 8 are relaxed. The uniform rates of convergence for the partial derivatives of $\hat{m}_n(\cdot)$ obtained in (iii) (b) and (c) are new to the best of our knowledge.

2.4 Non-stationary observations

A main assumption in Lemma 2.2 is strict stationarity of the underlying stochastic process. However, kernel estimators can be defined and asymptotic properties can be obtained in more general situations. Let again K be some kernel function and

$(h_n)_{n\in\mathbb{N}}$ be a sequence of bandwidths. For a sample $(Y_{n,1}, \boldsymbol{X}_{n,1}), \ldots, (Y_{n,n}, \boldsymbol{X}_{n,n})$ of size $n \in \mathbb{N}$ observed from a triangular array process $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : t = 1, \ldots, n, n \in \mathbb{N}\}$, let then

$$\hat{f}_n(\boldsymbol{x}) := \frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{X}_{n,i} - \boldsymbol{x}}{h_n}\right)$$

and

$$\hat{m}_n(\boldsymbol{x}) := \frac{\frac{1}{nh_n^d} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{X}_{n,i} - \boldsymbol{x}}{h_n}\right) Y_{n,i}}{\hat{f}_n(\boldsymbol{x})}, \text{ if } \hat{f}_n(\boldsymbol{x}) \neq 0,$$

and $\hat{m}_n(\boldsymbol{x}) := 0$, otherwise. It can be shown that under suitable conditions, $\hat{m}_n(\boldsymbol{x})$ still consistently estimates some non-stochastic object that will be denoted by $\bar{m}_n(\boldsymbol{x})$. Moreover, this convergence in fact holds uniformly over some expanding compact set $\boldsymbol{J}_n$. This result is stated in Lemma 2.3 and holds under the following assumptions.

(P)' Let $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : t = 1, \ldots, n, n \in \mathbb{N}\}$ be a strongly mixing triangular array with mixing coefficient $\alpha(\cdot)$. For some $b > 2$ let

$$\alpha(t) = O(t^{-\beta}),\ (t \to \infty) \text{ and } \beta > \frac{1 + (b-1)(1+d)}{b-2}.$$

(M)' For b from assumption **(P)'** let

- $\sup_{n\in\mathbb{N}} \sup_{1\le i\le n} E[|Y_{n,i}|^b] < \infty$,
- for all $1 \le i \le n$ and $n \in \mathbb{N}$, $\boldsymbol{X}_{n,i}$ be absolutely continuous with density function $f_{n,i} : \mathbb{R}^d \to \mathbb{R}$ that satisfies

  $$\sup_{n\in\mathbb{N}} \sup_{1\le i\le n} \sup_{\boldsymbol{x}\in\mathbb{R}^d} E[|Y_{n,i}|^b | \boldsymbol{X}_{n,i} = \boldsymbol{x}] f_{n,i}(\boldsymbol{x}) < \infty$$

  and

  $$\sup_{n\in\mathbb{N}} \sup_{1\le i\le n} \sup_{\boldsymbol{x}\in\mathbb{R}^d} f_{n,i}(\boldsymbol{x}) < \infty,$$

- there exist an $N \ge 0$ such that

  $$\sup_{n\in\mathbb{N}} \sup_{|i-j|\ge N} \sup_{\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^d} E[|Y_{n,i}Y_{n,j}| | \boldsymbol{X}_{n,i} = \boldsymbol{x}, \boldsymbol{X}_{n,j} = \boldsymbol{y}] f_{n,ij}(\boldsymbol{x}, \boldsymbol{y}) < \infty,$$

  where $f_{n,ij}$ is the density function of $(\boldsymbol{X}_{n,i}, \boldsymbol{X}_{n,j})$.

Additionally, for all $1 \le i \le n$ and $n \in \mathbb{N}$ let $m_{n,i} : \mathbb{R}^d \to \mathbb{R}$ be the conditional mean function defined by

$$m_{n,i}(\boldsymbol{x}) := E[Y_{n,i} | \boldsymbol{X}_{n,i} = \boldsymbol{x}] \text{ for all } \boldsymbol{x} \in \mathbb{R}^d.$$

(J)' Let $(c_n)_{n\in\mathbb{N}}$ be a positive sequence of real valued numbers satisfying $c_n = O\left(\log(n)^{\frac{1}{d}}\right)$ and $\boldsymbol{J}_n := [-c_n, c_n]^d$.

(F1)' For some $C < \infty$ and c_n from assumption **(J)'** let $\boldsymbol{I}_n := [-c_n - Ch_n, c_n + Ch_n]^d$ and let for $1 \leq i \leq n$ and $n \in \mathbb{N}$,

- $f_{n,i}$ and $m_{n,i} : \mathbb{R}^d \to \mathbb{R}$ be continuously differentiable,
- $\delta_n^{-1} := \inf_{\boldsymbol{x} \in \boldsymbol{J}_n} \inf_{1 \leq i \leq n} f_{n,i}(\boldsymbol{x}) > 0$ for all $n \in \mathbb{N}$,
- $p_n := \max_{|\boldsymbol{k}|=1} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} \sup_{1 \leq i \leq n} |D^{\boldsymbol{k}} f_{n,i}(\boldsymbol{x})| < \infty$ for all $n \in \mathbb{N}$ and
- $q_n := \max_{0 \leq |\boldsymbol{k}| \leq 1} \sup_{\boldsymbol{x} \in \boldsymbol{I}_n} \sup_{1 \leq i \leq n} |D^{\boldsymbol{k}} m_{n,i}(\boldsymbol{x})| < \infty$ for all $n \in \mathbb{N}$.

(K)' Let $K : \mathbb{R}^d \to \mathbb{R}$ be symmetric in each component with $\int_{\mathbb{R}^d} K(\boldsymbol{z}) d\boldsymbol{z} = 1$ and for C from assumption **(F)'** let

- K have compact support $[-C, C]^d$,
- $|K(\boldsymbol{u})| < \infty$ for all $\boldsymbol{u} \in \mathbb{R}^d$,
- $|K(\boldsymbol{u}) - K(\boldsymbol{u'})| \leq \Lambda \|\boldsymbol{u} - \boldsymbol{u'}\|$ for some $\Lambda < \infty$ and for all $\boldsymbol{u}, \boldsymbol{u'} \in \mathbb{R}^d$.

(B)' With b and β from assumption **(P)'** let

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

For δ_n, p_n, q_n from assumption **(F)'** let

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

Lemma 2.3. *Let the assumptions **(P)'**, **(M)'**, **(J)'**, **(F1)'**, **(K)'** and **(B)'** hold and let for all $n \in \mathbb{N}$, $\bar{m}_n : \mathbb{R}^d \to \mathbb{R}$ be defined by*

$$\bar{m}_n(\boldsymbol{x}) := \frac{\frac{1}{n} \sum_{i=1}^n f_{n,i}(\boldsymbol{x}) m_{n,i}(\boldsymbol{x})}{\frac{1}{n} \sum_{i=1}^n f_{n,i}(\boldsymbol{x})}. \tag{2.20}$$

Then it holds that

$$\sup_{\boldsymbol{x} \in \boldsymbol{J}_n} |\hat{m}_n(\boldsymbol{x}) - \bar{m}_n(\boldsymbol{x})| = o_P(1). \tag{2.21}$$

The proof is similar to the proof of Lemma 2.2. As the underlying process is not strictly stationary, Hansen's uniform convergence rates in [26] can not be used. Instead, Kristensen's Theorem 1 in [42] can be applied. It is a generalization of the aforementioned uniform convergence rates to possibly heterogeneous triangular arrays. Concerning the bias term, the proof only requires minor modifications. The details are omitted.

Remark. A similar result for

$$\hat{\sigma}_n^2(\boldsymbol{x}) := \frac{\frac{1}{n}\sum\limits_{i=1}^{n} K_h(\boldsymbol{X}_{n,i} - \boldsymbol{x})(Y_{n,i} - \hat{m}_n(\boldsymbol{x}))^2}{\hat{f}_n(\boldsymbol{x})}, \text{ if } \hat{f}_n(\boldsymbol{x}) \neq 0,$$

and $\hat{\sigma}_n^2(\boldsymbol{x}) := 0$ otherwise, can be obtained. Under suitable conditions it can be shown that

$$\sup_{\boldsymbol{x}\in \boldsymbol{J}_n} \left|\hat{\sigma}_n^2(\boldsymbol{x}) - \bar{\sigma}_n^2(\boldsymbol{x})\right| = o_P(1),$$

where

$$\bar{\sigma}_n^2(\boldsymbol{x}) := \frac{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})\left(\sigma_{n,i}^2(\boldsymbol{x}) + m_{n,i}^2(\boldsymbol{x})\right)}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})} - \left(\frac{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x}) m_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})}\right)^2$$

and $\sigma_{n,i}^2(\boldsymbol{x}) := \mathrm{Var}(Y_{n,i}|\boldsymbol{X}_{n,i} = \boldsymbol{x})$. Note that for $m_{n,i} \equiv m$, for all $1 \leq i \leq n$, $n \in \mathbb{N}$ and some $m : \mathbb{R}^d \to \mathbb{R}$ not depending on i and n, as it will be assumed in Chapter 5, this simplifies to

$$\bar{\sigma}_n^2(\boldsymbol{x}) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})\sigma_{n,i}^2(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})}.$$

3 Changepoint test in the conditional mean function

In this chapter, a test for a change in the conditional mean function will be constructed. To do that, a suitable test statistic will be considered. Under the null hypothesis of no change and some regularity assumptions, the limiting distribution of the test statistic will lead to the necessary critical value to construct a test of some asymptotic level. The consistency of the test against changepoint alternatives will be studied. Furthermore, the case of one-dimensional covariates will be discussed in more detail, as it will lead to a distribution free limiting distribution. Finally, some remarks on related literature will be made.

3.1 Definition of the test statistic

Consider model (1.1) on page 6. To test the null hypothesis (1.2), the following cumulative sum of residuals is used

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

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)** on page 16 and $\lfloor x \rfloor := \max\{k \in \mathbb{Z} : k \leq x\}$ for all $x \in \mathbb{R}$. The process

$$\hat{T}_n := \left\{ \hat{T}_n(s, \boldsymbol{z}) : s \in [0, 1], \boldsymbol{z} \in \mathbb{R}^d \right\}$$

can be viewed as a random element in $l^\infty([0, 1] \times \mathbb{R}^d)$ and is referred to as the *sequential marked empirical process of residuals*. Under H_0 and regularity assumptions, it will be shown that it converges weakly to a centered Gaussian process

$$\{G_0(s, \boldsymbol{z}) : s \in [0, 1], \boldsymbol{z} \in \mathbb{R}^d\}.$$

Using the continuous mapping theorem (see for instance Theorem 1.3.6 in [75]) it then can be concluded that

$$T_{n1} := \sup_{\boldsymbol{z} \in \mathbb{R}^d} \sup_{s \in [0,1]} \left| \hat{T}_n(s, \boldsymbol{z}) \right| \xrightarrow[n \to \infty]{\mathcal{D}} \sup_{\boldsymbol{z} \in \mathbb{R}^d} \sup_{s \in [0,1]} \left| G_0(s, \boldsymbol{z}) \right|.$$

The test statistic T_{n1} is called *Kolmogorov-Smirnov* test statistic. A test of asymptotic level $\alpha \in (0, 1)$, based on T_{n1} can be constructed by rejecting the null if T_{n1} exceeds the $(1 - \alpha)$-quantile of the limiting distribution. Thus, the asymptotic behavior of the test statistic under the null is of great importance and will be studied in the next section. Note that different test statistics, that are also based on $\hat{T}_n$, will be constructed in Section 3.4. They will be denoted by T_{n2}, T_{n3} and T_{n4}.

3.2 Asymptotic behavior under the null

This section contains the proof of the limiting distribution of the process $\hat{T}_n$ under H_0 and regularity assumptions. It splits into two main parts. First, it will be shown that $\hat{T}_n$ decomposes into a dominating term and a remainder term that is negligible. Secondly, the weak convergence of the dominating term will be shown. Both proofs require some sophisticated methods that need empirical processes and some parts are rather technical. For reasons of clarity and comprehensibility, most of the technical results and proofs are outsourced to Appendix A. A new weak convergence result for sequential empirical processes with dependent data will also be applied. It can be found in Appendix B.

In what follows, the regularity assumptions, under which the limiting distribution can be obtained, are displayed.

(G) Let $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ be a strictly stationary, strongly mixing process with mixing coefficient $\alpha(\cdot)$ such that

$$\alpha(t) = O\left(a^{-t}\right),\ (t \to \infty) \tag{3.1}$$

for some $a \in (1, \infty)$.

Remark. Note that this assumption means that the mixing coefficient decays at a geometric rate. This is strictly stronger than the polynomial rate of decay assumed in **(P)** in Chapter 2. More precisely, under condition (3.1) condition (2.6) in **(P)** on page 15 holds for arbitrary large β. As Hansen pointed out in [26], then condition (2.7) on the bandwidth in **(B1)** on page 16 simplifies to the less restrictive condition

$$\frac{\log(n)}{nh_n^d} = o(1). \tag{3.2}$$

(U) For some $\gamma > 0$ and some even $Q > (d+1)(2+\gamma)$, and $\mathcal{F}^t := \sigma(U_{j-1}, \boldsymbol{X}_j : j \leq t)$, let for $(U_t)_{t\in\mathbb{Z}}$ the following hold

- $E[U_t|\mathcal{F}^t] = 0$ a.s. for all $t \in \mathbb{Z}$,
- $E[U_t^2|\boldsymbol{X}_t] = \sigma^2(\boldsymbol{X}_t)$ a.s. for all $t \in \mathbb{Z}$ and
- $E\left[|U_t|^{Q\frac{2+\gamma}{2}}|\boldsymbol{X}_t\right] \leq c(\boldsymbol{X}_t)^Q$ a.s. for all $t \in \mathbb{Z}$, for some functions $c, \sigma^2 : \mathbb{R}^d \to \mathbb{R}$ with

$$\int \bar{c}(\boldsymbol{u})dF(\boldsymbol{u}) \leq M, \bar{c}(\boldsymbol{u}) := \max\left\{\sigma^2(\boldsymbol{u}), c(\boldsymbol{u})^2, \ldots, c(\boldsymbol{u})^Q\right\},$$

for some $M < \infty$.

(F2) For m from assumption **(M)**, q_n from assumption **(F1)**, c_n from assumption **(J)** and C from assumption **(K)**, let for all $\boldsymbol{k} \in \mathbb{N}_0^d$ with $|\boldsymbol{k}| = 2$

$$\sup_{\boldsymbol{x}\in[-c_n-2h_nC, c_n+2h_nC]^d} \left|D^{\boldsymbol{k}}m(\boldsymbol{x})\right| = O(q_n).$$

Remark. Note that from assumption **(F1)** on page 16, it already holds for all $|\boldsymbol{k}| = 2$ that

$$\sup_{\boldsymbol{x}\in[-c_n-h_nC,c_n+h_nC]^d} \left|D^{\boldsymbol{k}}m(\boldsymbol{x})\right| = O(q_n).$$

(B3) For l, p_n, q_n, δ_n from assumption **(F1)** and η from assumption **(B2)**, let h_n satisfy the following conditions

$$\frac{\log(n)^{3+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}} q_n^2\delta_n^2 = o(1), \tag{3.3}$$

$$\frac{\log(h_n)}{\sqrt{nh_n^d}} = o(1), \tag{3.4}$$

$$\sqrt{n}h_n^r p_n q_n = o(1), \tag{3.5}$$

$$\log(n)^3 h_n q_n^2 = o(1). \tag{3.6}$$

Remark.

- Note that condition (3.3) implies condition (3.2) on page 38.
- In order to satisfy (3.3), 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.
- If q_n and δ_n only have a $\log(n)$ rate, namely if there exist $r_1, r_2 \geq 0$, such that

$$q_n = O\left(\log(n)^{r_1}\right) \text{ and } \delta_n = O\left(\log(n)^{r_2}\right),$$

then condition (3.3) simplifies to

$$\frac{\log(n)^{3+\frac{d}{l+\eta}+2r_1+2r_2}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}} = o(1).$$

For faster rates of q_n and δ_n, namely if there exist $r_1, r_2 \geq 0$ and $s_1, s_2 \geq 0$ such that

$$q_n = O\left(\log(n)^{r_1}n^{s_1}\right) \text{ and } \delta_n = O\left(\log(n)^{r_2}n^{s_2}\right),$$

then condition (3.3) is

$$\frac{\log(n)^{3+\frac{d}{l+\eta}+2r_1+2r_2}}{\sqrt{n^{1-\frac{d}{l+\eta}-4(s_1+s_2)}h_n^d}} = o(1).$$

A necessary condition then is $1 > \frac{d}{l+\eta} + 4(s_1 + s_2)$, which only allows for small s_1, s_2 and large l, depending on the dimension d.

- In order to satisfy both (3.3) and (3.5) at the same time, the order of the kernel needs to be large depending on both the dimension and the smoothness assumptions at the same time. In particular, $r > \frac{d}{2}\frac{l+\eta}{l+\eta-d}$ is a necessary condition.
- Condition (3.4) is implied by (3.3), 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.

Theorem 3.1 (Decomposition)**.** *Suppose that **(G)**, **(U)**, **(M)**, **(J)**, **(F1)**, **(F2)**, **(K)**, **(B1)**, **(B2)** and **(B3)** 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* $\boldsymbol{z} \in \mathbb{R}^d$, *where* $T_n(s,\boldsymbol{z}) := \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i I\{\boldsymbol{X}_i \le \boldsymbol{z}\}$.

Proof. Inserting the definition of $\hat{T}_n$ and $Y_i = m(\boldsymbol{X}_i) + U_i$ for all $i = 1,\dots,n$ under the null, it can be obtained that

$$\begin{aligned}
\hat{T}_n(s,\boldsymbol{z}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(Y_i - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\\
&= T_n(s,\boldsymbol{z}) + \frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\} + o_P(1),
\end{aligned}$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where the last equality is an application of Lemma A.4 on page 142. Applying Lemma A.1 on page 113, inserting the definition of $\hat{m}_n$ and using $Y_i = m(\boldsymbol{X}_i) + U_i$ under the null, yields

$$\begin{aligned}
&\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i)\right)\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\}\\
&\overset{L.A.1}{=} s\sqrt{n}\int_{\mathbb{R}^d}\left(m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x})\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x} + o_P(1)\\
&= s\sqrt{n}\int_{\mathbb{R}^d}\left(m(\boldsymbol{x}) - \frac{1}{n}\sum_{i=1}^{n}K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)Y_i\frac{1}{\hat{f}_n(\boldsymbol{x})}\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x} + o_P(1)\\
&= s\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\int_{\mathbb{R}^d}\left(m(\boldsymbol{x}) - m(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} \qquad (3.7)\\
&- s\frac{1}{\sqrt{n}}\sum_{i=1}^{n}U_i\int_{\mathbb{R}^d}K_{h_n}(\boldsymbol{x}-\boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} + o_P(1), \qquad (3.8)
\end{aligned}$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Concerning (3.7), Lemma A.2 on page 124 states that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \int_{\mathbb{R}^d} (m(\boldsymbol{x}) - m(\boldsymbol{X}_i)) K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i) \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}\} \frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x} = o_P(1),$$

uniformly in $\boldsymbol{z} \in \mathbb{R}^d$. Concerning (3.8), applying Lemma A.3 on page 131 and Lemma A.4 on page 142, it can be obtained that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \int_{\mathbb{R}^d} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i) \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}\} \frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x} \\
&\overset{L.A.3}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} + o_P(1) \\
&\overset{L.A.4}{=} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} + o_P(1),
\end{aligned}$$

uniformly in $\boldsymbol{z} \in \mathbb{R}^d$. Putting the results together, the assertion of Theorem 3.1 is obtained.

□

Before stating the weak convergence result for $\{T_n(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$, let the following notations be introduced. For all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^d$ let

$$\boldsymbol{x} \wedge \boldsymbol{y} := (x_1 \wedge y_1, \ldots, x_d \wedge y_d),$$

where $x \wedge y := \min\{x, y\}$ for all $x, y \in \mathbb{R}$. Let furthermore for $g : \mathbb{R}^d \to \mathbb{R}$

$$\int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} g(\boldsymbol{u}) d\boldsymbol{u} := \int_{-\infty}^{x_d} \cdots \int_{-\infty}^{x_1} g(u_1, \ldots, u_d) du_1 \ldots du_d.$$

Theorem 3.2 (Weak convergence of T_n)**.** *Suppose that the assumptions* ***(G)*** *and* ***(U)*** *are satisfied. Then under H_0 it holds that*

$$T_n := \left\{T_n(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\right\} \underset{n\to\infty}{\rightsquigarrow} G := \left\{G(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\right\}$$

in $l^\infty([0,1] \times \mathbb{R}^d)$, where G is a centered Gaussian process 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 $\Sigma : \mathbb{R}^d \to \mathbb{R}, \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} \sigma^2(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u}$.

Remark. The proof of Theorem 3.2 is essentially an application of Corollary B.3. Note that concerning the dependence structure of the underlying process, assumption **(A1)** from Theorem B.1 needs to be verified. It is however less restrictive than **(G)**. In particular Theorem 3.2 can also be proven under $\big\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t \in \mathbb{Z}\big\}$ being strictly stationary and strongly mixing with

$$\sum_{t=1}^{\infty} t^{Q-2} \alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty,$$

for some $\gamma > 0$ and some even $Q > d(2+\gamma)$ satisfying assumption **(U)** (cf. assumption **(A1)** on page 160).

Proof. First notice that due to assumption **(G)** and under the null restriction $(U_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ 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 $(U_1, \boldsymbol{X}_1)$. By defining

$$\mathcal{F} := \{(u, \boldsymbol{x}) \mapsto uI\{\boldsymbol{x} \leq \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\},$$

the process T_n in $l^\infty([0,1] \times \mathbb{R}^d)$ can thus be identified with the process

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

in $l^\infty([0,1] \times \mathcal{F})$. Notice that for all $\varphi \in \mathcal{F}$, it holds that $\int \varphi dP = 0$ as $E[U_t|\mathcal{F}^t] = 0$ for all $t \in \mathbb{Z}$. It is therefore sufficient to prove the weak convergence of G_n. This will be shown by an application of Corollary B.3 on page 161 in Appendix B. Hence, assumptions **(A1)**, **(A2)** of Theorem B.1 and assumption **(A3)** of Corollary B.3 will be verified for the process $(U_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ and the function class $\mathcal{F}$, as well as the convergence of all finite dimensional distributions of G_n will be shown.

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 restriction as measurable functions maintain mixing properties (see [21] Subsection 2.6.1 (ii), p. 69).

To show condition **(A2)** on the function class $\mathcal{F}$, the choice of approximating functions and bounding functions, as in Definition 1.6 on page 11, will be discussed in more detail. Note that the semi norm ρ simplifies to the $L_2(P)$ norm and the semi metric d simplifies to the $L_{Q\frac{2+\gamma}{2}}(P)$ metric as the underlying process is not a triangular array but a strictly stationary sequence. Denote with $\bar{c}$ from assumption **(U)**

$$h : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \bar{c}(\boldsymbol{x})f(\boldsymbol{x}),$$

and for all $i = 1, \ldots, d$

$$\begin{aligned} h_i &: \mathbb{R} \to \mathbb{R} \\ &x \mapsto \int \cdots \int h(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_d. \end{aligned}$$

Let furthermore

$$H : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty}, \boldsymbol{x}]} h(\boldsymbol{t}) d\boldsymbol{t},$$

and for all $i = 1, \ldots, d$

$$H_i : \mathbb{R} \to \mathbb{R}, x \mapsto \int_{-\infty}^{x} h_i(t) dt.$$

Let $\varepsilon > 0$ and choose for all $i = 1, \ldots, d$ some $N_i = N_i(\varepsilon) \in \mathbb{N}$ and

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

namely a partition of $\mathbb{R}$, such that

$$H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right) \leq \frac{\varepsilon^2}{d},\ \forall\, j_i = 1,\ldots,N_i,\ i = 1,\ldots,d. \tag{3.9}$$

Since H_i is continuous and $H_i(-\infty) = H(-\boldsymbol{\infty}) = 0$ and $H_i(\infty) = H(\boldsymbol{\infty}) \leq M$ 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

$$\boldsymbol{z_j} := \left(z_{j_1,1},\ldots,z_{j_d,d}\right),$$

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

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

and bounding functions

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

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

$$|\varphi - a_{\boldsymbol{j}}| \leq b_{\boldsymbol{j}}.$$

Furthermore, it holds that

$$\|b_{\boldsymbol{j}}\|_{L_2(P)} \leq \varepsilon \text{ and } \max_{2\leq i\leq Q}\left(\int |b_{\boldsymbol{j}}|^{i\frac{2+\gamma}{2}}dP\right)^{\frac{1}{2}} \leq \varepsilon,\ \forall\, \boldsymbol{j} \in \mathop{\times}_{i=1}^{d}\{1,\ldots,N_i\}. \tag{3.10}$$

To see this let $\boldsymbol{j} \in \times_{i=1}^d\{1,\ldots,N_i\}$ and consider

$$\begin{aligned}
\|b_{\boldsymbol{j}}\|^2_{L_2(P)} &= E\left[|U_t|^2\left(I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\right\}\right)\right] \\
&= E\Bigg[\underbrace{E[|U_t|^2|\boldsymbol{X}_t]}_{=\sigma^2(\boldsymbol{X}_t)\text{ a.s.}}\left(I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\right\}\right)\Bigg] \\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]\setminus(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \sigma^2(\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u} \\
&\leq \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]\setminus(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \bar{c}(\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u} \\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \bar{c}(\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u} - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \bar{c}(\boldsymbol{u})f(\boldsymbol{u})d\boldsymbol{u} \\
&= H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}),
\end{aligned}$$

and for all $i = 2,\ldots,Q$ by Jensen's inequality and **(U)**, it holds that

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

hence

$$\begin{aligned}
\int |b_{\boldsymbol{j}}|^{i\frac{2+\gamma}{2}} dP &= E\left[|U_t|^{i\frac{2+\gamma}{2}}\left(I\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\} - I\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\}\right)\right] \\
&= E\Bigg[\underbrace{E\left[|U_t|^{i\frac{2+\gamma}{2}}|\boldsymbol{X}_t\right]}_{\leq c(\boldsymbol{X}_t)^i \text{ a.s.}}\left(I\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\} - I\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\}\right)\Bigg] \\
&\leq \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]\setminus(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} c(\boldsymbol{u})^i f(\boldsymbol{u}) d\boldsymbol{u} \\
&\leq \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]\setminus(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \bar{c}(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u} \\
&= H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}).
\end{aligned}$$

An investigation of the (same) upper bound in both cases leads to

$$\begin{aligned}
H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}) &= H\left(z_{j_1,1}, \ldots, z_{j_d,d}\right) - H\left(z_{j_1-1,1}, \ldots, z_{j_d-1,d}\right) \\
&= H\left(z_{j_1,1}, \ldots, z_{j_d,d}\right) - H\left(z_{j_1-1,1}, z_{j_2,2}, \ldots, z_{j_d,d}\right) \\
&+ H\left(z_{j_1-1,1}, z_{j_2,2}, \ldots, z_{j_d,d}\right) - \ldots \\
&+ \\
&\vdots \\
&+ \cdots - H\left(z_{j_1-1,1}, \ldots, z_{j_{d-1}-1,d-1}, z_{j_d,d}\right) \\
&+ H\left(z_{j_1-1,1}, \ldots, z_{j_{d-1}-1,d-1}, z_{j_d,d}\right) - H\left(z_{j_1-1,1}, \ldots, z_{j_d-1,d}\right) \\
&\overset{(*)}{\leq} \sum_{i=1}^{d}\left(H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right)\right) \\
&\overset{(3.9)}{\leq} d\frac{\varepsilon^2}{d} = \varepsilon^2.
\end{aligned}$$

To see the validity of $(*)$, only the first summand will be considered as the other ones work completely similar. It holds that

$$\begin{aligned}
&H\left(z_{j_1,1}, \ldots, z_{j_d,d}\right) - H\left(z_{j_1-1,1}, z_{j_2,1}, \ldots, z_{j_d,d}\right) \\
&= \int_{-\infty}^{z_{j_d,d}} \cdots \int_{-\infty}^{z_{j_2,2}} \int_{-\infty}^{z_{j_1,1}} h(u_1, \ldots, u_d) du_1 \ldots du_d - \int_{-\infty}^{z_{j_d,d}} \cdots \int_{-\infty}^{z_{j_2,2}} \int_{-\infty}^{z_{j_1-1,1}} h(u_1, \ldots, u_d) du_1 \ldots du_d \\
&\leq \int_{z_{j_1-1,1}}^{z_{j_1,1}} \underbrace{\left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} h(u_1, \ldots, u_d) du_2 \ldots du_d\right)}_{=h_1(u_1)} du_1 \\
&= \int_{z_{j_1-1,1}}^{z_{j_1,1}} h_1(u_1) du_1 \\
&= H_1\left(z_{j_1,1}\right) - H_1\left(z_{j_1-1,1}\right),
\end{aligned}$$

which proves (∗). Therefore, the inequalities in (3.10) have been shown for all $\boldsymbol{j} \in \times_{i=1}^{d}\{1,\dots,N_i\}$. Since $N_i = O(\varepsilon^{-2})$ for all $i = 1,\dots,d$, it holds that

$$\tilde{N}_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}\right) \leq \left| \mathop{\times}_{i=1}^{d} \{1,\dots,N_i\} \right| = \prod_{i=1}^{d} N_i = O\left(\varepsilon^{-2d}\right).$$

As $Q > d(2+\gamma)$ holds, assumption **(A2)** 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 $\mathcal{F}$ that fulfills

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

and additionally, it holds that

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

What is left to show, is the convergence of all finite dimensional distributions. It is to show that

$$\begin{pmatrix} T_n(s_1, \boldsymbol{z}_1) \\ \vdots \\ T_n(s_K, \boldsymbol{z}_K) \end{pmatrix} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}_K\left(\boldsymbol{0}, W\right) \tag{3.11}$$

for all $K \in \mathbb{N}$ and all collections $s_1,\dots,s_K \in [0,1]$ and $\boldsymbol{z}_1,\dots,\boldsymbol{z}_K \in \mathbb{R}^d$, where $W = (W_{ij})_{i,j=1,\dots,K}$ is the covariance matrix with entries

$$W_{ij} := (s_i \wedge s_j)\Sigma(\boldsymbol{z}_i \wedge \boldsymbol{z}_j),\ \forall\ i,j = 1,\dots,K.$$

Using the Cramér-Wold device, (3.11) is equivalent to

$$\sum_{j=1}^{K} \lambda_j T_n(s_j, \boldsymbol{z}_j) \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}(0, w^2),\ \forall\ (\lambda_1,\dots,\lambda_K) \in \mathbb{R}^K, \tag{3.12}$$

with variance $w^2 := \sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\lambda_{j_1}\lambda_{j_2}(s_{j_1}\wedge s_{j_2})\Sigma(\boldsymbol{z}_{j_1}\wedge\boldsymbol{z}_{j_2})$. The random variable of interest in (3.12) can be written as

$$\sum_{j=1}^{K}\lambda_j T_n(s_j,\boldsymbol{z}_j) = \sum_{j=1}^{K}\lambda_j\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_j\rfloor}U_i I\{\boldsymbol{X}_i\le\boldsymbol{z}_j\} = \sum_{i=1}^{n}\xi_{n,i},$$

where

$$\xi_{n,i} := \frac{1}{\sqrt{n}}U_i\sum_{j=1}^{K}\lambda_j I\{\boldsymbol{X}_i\le\boldsymbol{z}_j\}I\left\{\tfrac{i}{n}\le s_j\right\}.$$

To show (3.12), Rio's Corollary 1 in [64] will be used, which is a central limit theorem for strongly mixing triangular arrays. Following the notations in [64] 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[1] 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 coefficients of $\{\xi_{n,i} : 1\le i\le n, n\in\mathbb{N}\}$ defined as in (1.4) in Definition 1.1 on page 7. For $t\in(0,\infty)$ define $\tilde{\alpha}_n(t) := \tilde{\alpha}_n(\lfloor t\rfloor)$. Let its càdlàg inverse function be defined by

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

As for all $l=1,\dots,n$

$$V_{n,l} = \sum_{j_1=1}^{K}\sum_{j_2=1}^{K}\lambda_{j_1}\lambda_{j_2}\frac{\lfloor ls_{j_1}\rfloor\wedge\lfloor ls_{j_2}\rfloor}{n}\Sigma(\boldsymbol{z}_{j_1}\wedge\boldsymbol{z}_{j_2}),$$

holds, it can be obtained that

$$\limsup_{n\to\infty}\max_{1\le l\le n}\frac{V_{n,l}}{V_{n,n}} = \limsup_{n\to\infty}\max_{1\le l\le n}\frac{\sum\limits_{j_1=1}^{K}\sum\limits_{j_2=1}^{K}\lambda_{j_1}\lambda_{j_2}\lfloor ls_{j_1}\rfloor\wedge\lfloor ls_{j_2}\rfloor\Sigma(\boldsymbol{z}_{j_1}\wedge\boldsymbol{z}_{j_2})}{\sum\limits_{j_1=1}^{K}\sum\limits_{j_2=1}^{K}\lambda_{j_1}\lambda_{j_2}\lfloor ns_{j_1}\rfloor\wedge\lfloor ns_{j_2}\rfloor\Sigma(\boldsymbol{z}_{j_1}\wedge\boldsymbol{z}_{j_2})} < \infty,$$

which is condition (a) in Corollary 1 in [64]. Concerning condition (b) in aforementioned corollary, it needs to be shown that

$$V_{n,n}^{-\frac{3}{2}}\sum_{i=1}^{n}\int_0^1\tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right)Q_{n,i}^2(x)\inf\left\{\tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right)Q_{n,i}(x),\sqrt{V_{n,n}}\right\}dx\to 0. \tag{3.13}$$

By Markov's inequality, it holds that for all $t>0$ and with $q := Q\frac{2+\gamma}{2}$

$$P(|\xi_{n,i}|>t) = P\left(\left|\frac{1}{\sqrt{n}}U_i\sum_{j=1}^{K}\lambda_j I\{\boldsymbol{X}_i\le\boldsymbol{z}_j\}I\left\{\tfrac{i}{n}\le s_j\right\}\right|>t\right)$$

[1] càdlàg: right continuous with existing left limits (French: *continué à droite, limité à gauche*)

$$
\begin{aligned}
&\leq t^{-q} E\left[\left|\frac{1}{\sqrt{n}} U_i \sum_{j=1}^{K} \lambda_j I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\} I\left\{\tfrac{i}{n} \leq s_j\right\}\right|^q\right] \\
&= t^{-q} n^{-\frac{q}{2}} E\left[\underbrace{E[|U_i|^q|\boldsymbol{X}_i]}_{\leq c^Q(\boldsymbol{X}_i) \text{ a.s.}} \left|\sum_{j=1}^{K} \lambda_j I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\} I\left\{\tfrac{i}{n} \leq s_j\right\}\right|^q\right] \\
&\leq t^{-q} n^{-\frac{q}{2}} \left(\sum_{j=1}^{K} |\lambda_j|\right)^q \underbrace{\int c^Q(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u}}_{\leq M} \\
&\leq t^{-q} n^{-\frac{q}{2}} \tilde{M},
\end{aligned}
$$

for $M < \infty$ from assumption **(U)** and $\tilde{M} := (\sum_{j=1}^{K} |\lambda_j|)^q M$. Hence, for $u > 0$ fixed and for all $t > 0$ with $P(|\xi_{n,i}| > t) > u$, it holds that $u \leq t^{-q} n^{-\frac{q}{2}} \tilde{M}$. Solving the inequality for t, results in

$$
t \leq u^{-\frac{1}{q}} n^{-\frac{1}{2}} \tilde{M}^{\frac{1}{q}}, \ \forall\, t > 0 \text{ with } P(|\xi_{n,i}| > t) > u.
$$

It therefore also holds for $\sup\{t > 0 : P(|\xi_{n,i}| > t) > u\}$, i.e.

$$
Q_{n,i}(u) \leq u^{-\frac{1}{q}} n^{-\frac{1}{2}} \tilde{M}^{\frac{1}{q}}, \ \forall\, u > 0.
$$

For fixed $n \in \mathbb{N}$, it holds that $\xi_{n,i} = g_i(U_i, \boldsymbol{X}_i)$ with

$$
g_i(u, \boldsymbol{x}) := \frac{1}{\sqrt{n}} u \sum_{j=1}^{K} \lambda_j I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} I\left\{\tfrac{i}{n} \leq s_j\right\},
$$

where g_i is measurable for all $i \in \mathbb{N}$. As Bradley [4] states, $\tilde{\alpha}_n(\cdot)$ can therefore be bounded by the mixing coefficient of $\{(U_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$ which has the same properties as the mixing coefficient $\alpha(\cdot)$ of $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$. From assumption **(G)** it holds for all $t > 0$ that $\alpha(t) \leq Aa^{-t}$ for some $0 < A < \infty$ and some $a \in (1, \infty)$. Analogous to before, it can be shown that

$$
\tilde{\alpha}_n^{-1}(u) \leq \tilde{A} - \log_a(u), \ \forall\, u > 0,
$$

where $\tilde{A} := \log_a(A)$. Furthermore,

$$
V_{n,n} = \sum_{j_1=1}^{K} \sum_{j_2=1}^{K} \lambda_{j_1} \lambda_{j_2} \frac{\lfloor n s_{j_1} \rfloor \wedge \lfloor n s_{j_2} \rfloor}{n} \Sigma(\boldsymbol{z}_{j_1} \wedge \boldsymbol{z}_{j_2}) \underset{n \to \infty}{\to} w^2 < \infty \tag{3.14}
$$

holds. Putting the results together, it can be obtained that

$$
\begin{aligned}
&V_{n,n}^{-\frac{3}{2}} \sum_{i=1}^{n} \int_0^1 \tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right) Q_{i,n}^2(x) \inf\left\{\tilde{\alpha}_n^{-1}\left(\tfrac{x}{2}\right) Q_{i,n}(x), \sqrt{V_{n,n}}\right\} dx \\
&\leq \tilde{M}^{\frac{2}{q}} V_{n,n}^{-\frac{3}{2}} \frac{1}{n} \sum_{i=1}^{n} \int_0^1 \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} \inf\left\{\left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} n^{-\frac{1}{2}} \tilde{M}^{\frac{1}{q}}, \sqrt{V_{n,n}}\right\} dx
\end{aligned}
$$

$$= \frac{1}{\sqrt{n}} \tilde{M}^{\frac{2}{q}} V_{n,n}^{-\frac{3}{2}} \int\limits_0^1 \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} \inf\left\{\left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} \tilde{M}^{\frac{1}{q}}, \sqrt{n}\sqrt{V_{n,n}}\right\} dx.$$

Now let the sequence $(b_n)_{n\in\mathbb{N}}$ be defined by

$$b_n := \sup\left\{k \in (0,1) : \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} \tilde{M}^{\frac{1}{q}} > \sqrt{n}\sqrt{V_{n,n}},\ \forall\, x \le k\right\}.$$

Note that $b_n \to 0$ as $n \to \infty$ by construction. Then

$$\begin{aligned} &\frac{1}{\sqrt{n}} \tilde{M}^{\frac{2}{q}} V_{n,n}^{-\frac{3}{2}} \int\limits_{b_n}^1 \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} \inf\left\{\left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} \tilde{M}^{\frac{1}{q}}, \sqrt{n}\sqrt{V_{n,n}}\right\} dx \\ &= \frac{1}{\sqrt{n}} \tilde{M}^{\frac{3}{q}} V_{n,n}^{-\frac{3}{2}} \int\limits_{b_n}^1 \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right)^2 x^{-\frac{3}{q}} dx \underset{n\to\infty}{\to} 0, \end{aligned}$$

where the integral exists as $q = Q\frac{2+\gamma}{\gamma} > 3$. Furthermore,

$$\begin{aligned} &\frac{1}{\sqrt{n}} \tilde{M}^{\frac{2}{q}} V_{n,n}^{-\frac{3}{2}} \int\limits_0^{b_n} \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} \inf\left\{\left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{1}{q}} \tilde{M}^{\frac{1}{q}}, \sqrt{n}\sqrt{V_{n,n}}\right\} dx \\ &= \tilde{M}^{\frac{2}{q}} V_{n,n}^{-1} \int\limits_0^{b_n} \left(\tilde{A} - \log_a\left(\tfrac{x}{2}\right)\right) x^{-\frac{2}{q}} dx \to 0, \end{aligned}$$

as $(b_n)_{n\in\mathbb{N}}$ is a null sequence, which finally proves the validity of (3.13). Applying Corollary 1 in [64], it holds that $\frac{1}{\sqrt{V_{n,n}}} \sum_{i=1}^n \xi_{n,i} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \mathcal{N}(0,1)$. With (3.14) the assertion in (3.12) follows.

□

Corollary 3.3 (Weak convergence of $\hat{T}_n$)**.** *Suppose that the assumptions of Theorem 3.1 are satisfied. Then under H_0 it holds that*

$$\hat{T}_n(s, \boldsymbol{z}) \underset{n\to\infty}{\rightsquigarrow} G_0 := \left\{G_0(s, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\right\}$$

in $l^\infty([0,1] \times \mathbb{R}^d)$, where G_0 is a centered Gaussian process with

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

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

Proof. With Theorem 3.2 and an application of the continuous mapping theorem, it can be obtained that

$$\left\{T_n(s, \boldsymbol{z}) - sT_n(1, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\right\} \underset{n\to\infty}{\rightsquigarrow} \left\{G_0(s, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\right\},$$

where $G_0(s, \boldsymbol{z}) := G(s, \boldsymbol{z}) - sG(1, \boldsymbol{z})$ for all $s \in [0, 1]$, $\boldsymbol{z} \in \mathbb{R}^d$ and G is a centered Gaussian process resulting from Theorem 3.2 possessing the following covariance function

$$\operatorname{Cov}\big(G(s_1, \boldsymbol{z}_1), G(s_2, \boldsymbol{z}_2)\big) = (s_1 \wedge s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2), \tag{3.15}$$

for all $s_1, s_2 \in [0, 1]$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$. Hence, G_0 is again a centered Gaussian process and possesses the following covariance function

$$\begin{aligned}\operatorname{Cov}\left(G_0(s_1, \boldsymbol{z}_1), G_0(s_2, \boldsymbol{z}_2)\right) &= \operatorname{Cov}(G(s_1, \boldsymbol{z}_1) - s_1 G(1, \boldsymbol{z}_1), G(s_2, \boldsymbol{z}_2) - s_2 G(1, \boldsymbol{z}_2)) \\ &\overset{(3.15)}{=} (s_1 \wedge s_2 - s_1 s_2)\Sigma(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2),\end{aligned}$$

for all $s_1, s_2 \in [0, 1]$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$. Applying Theorem 3.1 and Slutsky's theorem (see Example 1.4.7 in [75]) yields the assertion of Corollary 3.3.

□

It will be seen in Section 3.4 that in the case of $d = 1$, the Kolmogorov-Smirnov test statistic T_{n1} can be standardized such that the resulting limiting distribution is free of nuisance parameters.

For $d > 1$ this is not the case. As the covariance function contains the unknown Σ, critical values for the changepoint test using the limiting distribution of T_{n1} can in general only be estimated. One possible choice to do that, is to use an analogue to the empirical distribution function, namely

$$\hat{\Sigma}_n(\boldsymbol{x}) := \frac{1}{n}\sum_{j=1}^{n} \hat{\sigma}_n^2(\boldsymbol{X}_j) I\{\boldsymbol{X}_j \leq \boldsymbol{x}\},$$

where $\hat{\sigma}_n^2$ is the Nadaraya-Watson estimator for σ^2 defined in (2.3) in Definition 2.1 on page 13.

A different option to construct a test without knowing the limiting distribution of the test statistic, is to use a bootstrap version of it to estimate quantiles of the test statistic itself under the null. A valid bootstrap procedure in the considered setting will be motivated in Chapter 4. It is a motivation for the case of non-stationary variances, i.e. where the assumption of strict stationarity under the null is not satisfied, such that Corollary 3.3 is not applicable. This bootstrap procedure can also be used in the case of $d > 1$, when the limiting distribution is not known, due to the fact that Σ is unknown.

3.3 Consistency analysis

In this section, the behavior of T_{n1} will be investigated under alternatives. In particular, it will be shown that the Kolmogorov-Smirnov type test is consistent against a simple fixed alternative as in (1.3) on page 7, in the sense that the rejection probability converges to one as the sample size converges to infinity, given that the alternative holds.

As the changepoint is supposed to depend on the sample size, the process should be viewed as a triangular array $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : t = 1, \ldots, n, n \in \mathbb{N}\}$. Rewriting model (1.1) on page 6, the following is considered

$$Y_{n,t} = m_{n,t}(\boldsymbol{X}_{n,t}) + U_{n,t},\ t = 1, \ldots, n,\ n \in \mathbb{N} \tag{3.16}$$

with

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

for some functions $m_{(1)}, m_{(2)} : \mathbb{R}^d \to \mathbb{R}$ with $m_{(1)} \not\equiv m_{(2)}$.

Theorem 3.4. *Let the assumptions of Lemma 2.3 on page 34 in Chapter 2 hold. Additionally,*

(U)' *for some functions $\sigma^2_{n,t} : \mathbb{R}^d \to \mathbb{R}$ and $\mathcal{F}_n^t := \sigma(U_{n,j-1}, \boldsymbol{X}_{n,j} : j \le t)$ let*

$$E[U_{n,t}|\mathcal{F}_n^t] = 0 \text{ and } E[U^2_{n,t}|\boldsymbol{X}_{n,t}] = \sigma^2_{n,t}(\boldsymbol{X}_{n,t}) \text{ a.s.}$$

for all $1 \le t \le n$ and $n \in \mathbb{N}$,

(F2)' *let $\inf\limits_{n\in\mathbb{N}} \inf\limits_{1\le i\le n} f_{n,i}(\boldsymbol{u}) > 0$ for all $\boldsymbol{u} \in \mathbb{R}^d$,*

(D)' *for all $s \in (0,1]$ let there exist a function $\bar{f}^{(s)} : \mathbb{R}^d \to \mathbb{R}$ such that*

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

(I)' *for $\bar{\sigma}^2 := \sup\limits_{n\in\mathbb{N}} \sup\limits_{1\le i\le n} \sigma^2_{n,i}$ and $\bar{f} := \sup\limits_{n\in\mathbb{N}} \sup\limits_{1\le i\le n} f_{n,i}$ and b from assumption* ***(P)'*** *on page 33 let*

$$\int \bar{\sigma}^2(\boldsymbol{x})\bar{f}(\boldsymbol{x})d\boldsymbol{x} < \infty, \tag{3.17}$$

$$\int \left|m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})\right|^b \bar{f}(\boldsymbol{x})d\boldsymbol{x} < \infty. \tag{3.18}$$

Then the test based on T_{n1} is consistent against H_1, meaning that

$$P(T_{n1} > c_0|H_1) \underset{n\to\infty}{\to} 1,$$

where c_0 is the critical value of the limiting distribution of T_{n1} under the null.

Remark. The estimated average $\bar{m}_n$ defined in Lemma 2.3 simplifies under H_1 to

$$\bar{m}_n(\boldsymbol{x}) = \frac{\frac{1}{n}\sum\limits_{i=1}^n f_{n,i}(\boldsymbol{x})m_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^n f_{n,i}(\boldsymbol{x})}$$

$$= \frac{m_{(1)}(\boldsymbol{x})\frac{1}{n}\sum\limits_{i=1}^{\lfloor ns_0\rfloor} f_{n,i}(\boldsymbol{x}) + m_{(2)}(\boldsymbol{x})\frac{1}{n}\sum\limits_{i=\lfloor ns_0\rfloor+1}^{n} f_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})}$$

$$= \left(m_{(1)}(\boldsymbol{x}) - m_{(2)}(\boldsymbol{x})\right) \frac{\frac{1}{n}\sum\limits_{i=1}^{\lfloor ns_0\rfloor} f_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})} + m_{(2)}(\boldsymbol{x})$$

and under assumption **(D)'** therefore converges for all fixed $\boldsymbol{x} \in \mathbb{R}^d$ to

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

Proof. First, it will be shown that for fixed $\boldsymbol{z} \in \mathbb{R}^d$ and $s \in (0,1)$ with $s \leq s_0$, it holds that

$$\hat{T}_n(s,\boldsymbol{z}) = \sqrt{n}\Delta(s,\boldsymbol{z}) + o_P(\sqrt{n}), \tag{3.19}$$

where

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

Let therefore $\boldsymbol{z} \in \mathbb{R}^d$ and $s \in (0,1)$ with $s \leq s_0$ be fixed. It holds that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\hat{T}_n(s,\boldsymbol{z}) &= \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(Y_{n,i} - \hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(m_{(1)}(\boldsymbol{X}_{n,i}) + U_{n,i} - \hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(m_{(1)}(\boldsymbol{X}_{n,i}) - \bar{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} && (3.20) \\
&+ \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(\bar{m}_n(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} && (3.21) \\
&+ \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}U_{n,i}\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}. && (3.22)
\end{aligned}$$

First, (3.22) will be considered. Using assumption **(U)'** and (3.17) in **(I)'**, it can be obtained that

$$E\left[\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}U_{n,i}\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}\right] = 0,\ \forall\ n \in \mathbb{N},$$

and

$$
\begin{aligned}
Var\left(\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor} U_{n,i}\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\leq \boldsymbol{z}\}\right) &= \frac{1}{n^2}\sum_{i=1}^{\lfloor ns\rfloor} E\left[U_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\leq \boldsymbol{z}\}\right] \\
&= \frac{1}{n}\int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}]} \omega_n(\boldsymbol{u})\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}\sigma_{n,i}^2(\boldsymbol{u})f_{n,i}(\boldsymbol{u})d\boldsymbol{u} \\
&\leq \frac{1}{n}s\int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}]} \bar{\sigma}^2(\boldsymbol{u})\bar{f}(\boldsymbol{u})d\boldsymbol{u},\ (\forall\ n\in\mathbb{N}) \\
&\underset{n\to\infty}{\to} 0.
\end{aligned}
$$

Considering (3.21), by an application of Lemma 2.3, it can be obtained that

$$
\begin{aligned}
&\left|\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(\bar{m}_n(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\leq \boldsymbol{z}\}\right| \\
&\leq \frac{1}{n}\sum_{i=1}^{n}\left|\bar{m}_n(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i})\right|\omega_n(\boldsymbol{X}_{n,i}) \\
&\leq \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\bar{m}_n(\boldsymbol{x})-\hat{m}_n(\boldsymbol{x})\right| \\
&= o_P(1).
\end{aligned}
$$

Hence, the dominating term is (3.20). Inserting $\bar{m}_n$, it holds that

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(m_{(1)}(\boldsymbol{X}_{n,i})-\bar{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\leq \boldsymbol{z}\} \\
&= \frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}(m_{(1)}(\boldsymbol{X}_{n,i})-m_{(2)}(\boldsymbol{X}_{n,i}))\left(1-\frac{\frac{1}{n}\sum\limits_{j=1}^{\lfloor ns_0\rfloor}f_{n,j}(\boldsymbol{X}_{n,i})}{\frac{1}{n}\sum\limits_{j=1}^{n}f_{n,j}(\boldsymbol{X}_{n,i})}\right)\omega_n(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\leq \boldsymbol{z}\} \\
&=: \hat{\Delta}_n(s,\boldsymbol{z}),
\end{aligned}
$$

and the proof of (3.19) is therefore concluded if

$$
\hat{\Delta}_n(s,\boldsymbol{z}) \underset{n\to\infty}{\overset{P}{\to}} \Delta(s,\boldsymbol{z})
$$

holds. As

$$
E\left[\hat{\Delta}_n(s,\boldsymbol{z})\right] = \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}]}(m_{(1)}(\boldsymbol{u})-m_{(2)}(\boldsymbol{u}))\left(1-\frac{\frac{1}{n}\sum\limits_{j=1}^{\lfloor ns_0\rfloor}f_{n,j}(\boldsymbol{u})}{\frac{1}{n}\sum\limits_{j=1}^{n}f_{n,j}(\boldsymbol{u})}\right)\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}f_{n,i}(\boldsymbol{u})\omega_n(\boldsymbol{u})d\boldsymbol{u}
$$

and for all $n \in \mathbb{N}$

$$\begin{aligned}
&\left| (m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})) \left(1 - \frac{\frac{1}{n}\sum\limits_{j=1}^{\lfloor ns_0 \rfloor} f_{n,j}(\boldsymbol{u})}{\frac{1}{n}\sum\limits_{j=1}^{n} f_{n,j}(\boldsymbol{u})} \right) \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}(\boldsymbol{u}) \omega_n(\boldsymbol{u}) I\{\boldsymbol{u} \leq \boldsymbol{z}\} \right| \\
&= \left| m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u}) \right| \frac{\sum\limits_{j=\lfloor ns_0 \rfloor+1}^{n} f_{n,j}(\boldsymbol{u})}{\sum\limits_{j=1}^{n} f_{n,j}(\boldsymbol{u})} \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}(\boldsymbol{u}) \omega_n(\boldsymbol{u}) I\{\boldsymbol{u} \leq \boldsymbol{z}\} \\
&\leq \left| m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u}) \right| \frac{\lfloor ns \rfloor}{n} \bar{f}(\boldsymbol{u}) \\
&\leq s \left| m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u}) \right| \bar{f}(\boldsymbol{u}),
\end{aligned}$$

holds, using the dominated convergence theorem and assumption (3.18), it can be obtained that

$$E\left[\hat{\Delta}_n(s, \boldsymbol{z}) \right] \underset{n \to \infty}{\to} \Delta(s, \boldsymbol{z}).$$

Additionally, it will be shown that $Var(\hat{\Delta}_n(s, \boldsymbol{z})) \to 0$ as $n \to \infty$. For simplicity reasons let the following notation be introduced

$$g_n(\boldsymbol{X}_{n,i}) := \left(m_{(1)}(\boldsymbol{X}_{n,i}) - m_{(2)}(\boldsymbol{X}_{n,i}) \right) \left(1 - \frac{\frac{1}{n}\sum\limits_{j=1}^{\lfloor ns_0 \rfloor} f_{n,j}(\boldsymbol{X}_{n,i})}{\frac{1}{n}\sum\limits_{j=1}^{n} f_{n,j}(\boldsymbol{X}_{n,i})} \right) \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\},$$

where the dependency on s and $\boldsymbol{z}$ is not reflected in the notation of g_n. Then it holds that

$$\begin{aligned}
Var\left(\hat{\Delta}_n(s, \boldsymbol{z}) \right) &= Var\left(\frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} g_n(\boldsymbol{X}_{n,i}) \right) \\
&= \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} Var\left(g_n(\boldsymbol{X}_{n,i}) \right) \qquad (3.23) \\
&+ \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \sum_{\substack{j=1 \\ j \neq i}}^{\lfloor ns \rfloor} \left(E[g_n(\boldsymbol{X}_{n,i}) g_n(\boldsymbol{X}_{n,j})] - E[g_n(\boldsymbol{X}_{n,i})] E[g_n(\boldsymbol{X}_{n,j})] \right). \qquad (3.24)
\end{aligned}$$

Considering (3.23), condition (3.18) from **(I)'** is used to obtain

$$\begin{aligned}
&\frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} Var(g_n(\boldsymbol{X}_{n,i})) \\
&\leq \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} E[g_n(\boldsymbol{X}_{n,i})^2]
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}]} \left(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\right)^2 \left(\frac{\sum\limits_{j=\lfloor ns_0 \rfloor+1}^{n} f_{n,j}(\boldsymbol{u})}{\sum\limits_{j=1}^{n} f_{n,j}(\boldsymbol{u})} \right)^2 f_{n,i}(\boldsymbol{u}) \omega_n(\boldsymbol{u}) d\boldsymbol{u} \\
&\leq s \frac{1}{n} \underbrace{\int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}]} \left(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\right)^2 \bar{f}(\boldsymbol{u}) d\boldsymbol{u}}_{\overset{(3.18)}{<} \infty}, \ (\forall\ n \in \mathbb{N}) \\
&\underset{n\to\infty}{\longrightarrow} 0.
\end{aligned}
$$

Hence, the term in (3.23) is negligible. For (3.24) the covariance inequality for strongly mixing triangular arrays, stated in Lemma B.5 on page 163, will be applied. For $b > 2$ from assumption **(I)'** and for all $1 \leq i \leq n$ and $n \in \mathbb{N}$, it holds that

$$
E\left[|g_n(\boldsymbol{X}_{n,i})|^b\right] \leq \int \left|m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\right|^b \bar{f}(\boldsymbol{u}) d\boldsymbol{u} \overset{(3.18)}{<} \infty,
$$

It therefore holds for $i \neq j$ that

$$
E\left[|g_n(\boldsymbol{X}_{n,i}) g_n(\boldsymbol{X}_{n,j})|^{\frac{b}{2}}\right] \leq \left(E\left[|g_n(\boldsymbol{X}_{n,i})|^b\right] E\left[|g_n(\boldsymbol{X}_{n,j})|^b\right]\right)^{\frac{1}{2}} = O(1)
$$

and similarly

$$
E\left[|g_n(\boldsymbol{X}_{n,i}) g_n(\tilde{\boldsymbol{X}}_{n,j})|^{\frac{b}{2}}\right] = O(1)
$$

for an independent copy $\{\tilde{\boldsymbol{X}}_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ of $\{\boldsymbol{X}_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$. Hence, Lemma B.5 can be applied with $\delta := \frac{b-2}{2} > 0$ and $M_n = O(1)$ using the notations from Lemma B.5. Let $i > j$ (the case $j > i$ works analogously). Then the term in (3.24) can be bounded by

$$
\begin{aligned}
&\frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \sum_{\substack{j=1 \\ j<i}}^{\lfloor ns \rfloor} \left| E[g_n(\boldsymbol{X}_{n,i}) g_n(\boldsymbol{X}_{n,j})] - E[g_n(\boldsymbol{X}_{n,i})] E[g_n(\boldsymbol{X}_{n,j})] \right| \\
&\leq \frac{1}{n^2} \sum_{i=1}^{\lfloor ns \rfloor} \sum_{\substack{j=1 \\ j<i}}^{\lfloor ns \rfloor} \alpha(i-j)^{\frac{b-2}{b}} \\
&\leq \frac{1}{n} \underbrace{\sum_{t=1}^{\infty} \alpha(t)^{\frac{b-2}{b}}}_{\overset{(*)}{<} \infty} \\
&= o(1).
\end{aligned}
$$

The assertion in $(*)$ holds as for β from assumption **(P)'**, it holds that

$$
\sum_{t=1}^{\infty} \alpha(t)^{\frac{b-2}{b}} = \sum_{t=1}^{\infty} O\left(t^{-\beta \frac{b-2}{b}}\right) = O(1),
$$

where the last equality is implied by

$$\beta > \frac{1+(b-1)(1+d)}{b-2} \overset{\forall d \geq 1}{\geq} \frac{2b-1}{b-2} \geq \frac{b}{b-2}.$$

This finally completes the proof of (3.19). Note that analogously it can be shown that

$$\hat{T}_n(s,\boldsymbol{z}) = \sqrt{n} \int\limits_{(-\infty,\boldsymbol{z}]} (m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})) \left(1 - \frac{\bar{f}^{(s)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) \bar{f}^{(s_0)}(\boldsymbol{u}) d\boldsymbol{u} + o_P(\sqrt{n}).$$

holds for all $\boldsymbol{z} \in \mathbb{R}^d$ and $s \in (0,1)$ with $s > s_0$. To show the consistency of the test, choose $\boldsymbol{z}_0 \in \mathbb{R}^d$ such that

$$\Delta(s_0, \boldsymbol{z}_0) \neq 0.$$

The existence of such a $\boldsymbol{z}_0 \in \mathbb{R}^d$ can be argued in the following way. It holds that $\Delta(s_0, \boldsymbol{z})$ vanishes for all $\boldsymbol{z} \in \mathbb{R}^d$ if and only if

$$(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) \bar{f}^{(s_0)}(\boldsymbol{u}) = 0, \; \forall \boldsymbol{u} \in \mathbb{R}^d. \tag{3.25}$$

By assumption **(F2)'** it holds that $\bar{f}^{(s_0)}(\boldsymbol{u}) > 0$ and $\bar{f}^{(s_0)}(\boldsymbol{u}) < \bar{f}^{(1)}(\boldsymbol{u})$ for all $\boldsymbol{u} \in \mathbb{R}^d$. Hence, (3.25) holds if and only if

$$m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u}) = 0, \forall \; \boldsymbol{u} \in \mathbb{R}^d.$$

However, as $m_{(1)} \not\equiv m_{(2)}$ by assumption, $m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u}) \neq 0$ holds for some $\boldsymbol{u} \in \mathbb{R}^d$. Applying (3.19), it can finally be concluded that

$$T_{n,1} = \sup_{s \in [0,1]} \sup_{\boldsymbol{z} \in \mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right| \geq \left|\hat{T}_n(s_0,\boldsymbol{z}_0)\right| = \underbrace{\sqrt{n}|\Delta(s_0,\boldsymbol{z}_0)|}_{\underset{n\to\infty}{\to}\infty} + o_P(\sqrt{n}).$$

The test based on $T_{n,1}$ is therefore consistent against H_1.

□

Remark (special cases). Sometimes strict stationarity is assumed even under the alternative. Note that this particularly excludes the case of autoregressive models. However this assumption implies that for all $1 \leq t \leq n$ and $n \in \mathbb{N}$

$$f_{n,t} \equiv f,$$

for some $f : \mathbb{R}^d \to \mathbb{R}$ not depending on t and n. Under the assumptions of Theorem 3.4 and H_1, it then holds that

$$\hat{T}_n(s,\boldsymbol{z}) = \begin{cases} \sqrt{n} s(1-s_0) \displaystyle\int\limits_{(-\infty,\boldsymbol{z}]} \big(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\big) f(\boldsymbol{u}) d\boldsymbol{u} + o_P(\sqrt{n}), & s \leq s_0 \\ \sqrt{n} s_0(1-s) \displaystyle\int\limits_{(-\infty,\boldsymbol{z}]} \big(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\big) f(\boldsymbol{u}) d\boldsymbol{u} + o_P(\sqrt{n}), & s > s_0 \end{cases}.$$

Here it can be seen that using the marked empirical process of residuals is essential for the consistency of the test. To see this let $m_{(1)} - m_{(2)}$ be an odd function and let f be a symmetric density. Examples are considered in the simulations, see for instance Subsection 6.1.1 for an i.i.d. model or Subsection 6.1.2 for a time series (but not autoregressive) model. Now it holds that

$$\int \left(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\right) f(\boldsymbol{u}) d\boldsymbol{u} = 0,$$

although $m_{(1)} \not\equiv m_{(2)}$. The test based on $\hat{T}_n(s, \infty)$ is therefore not consistent against H_1. The test based on $\hat{T}_n(s, \boldsymbol{z})$ however, is consistent as proven in Theorem 3.4.

A second simplification is the assumption of strict stationarity before and right after the break, meaning that for all $1 \leq t \leq n$ and $n \in \mathbb{N}$

$$f_{n,t} \equiv f_{(1)} I\left\{\frac{t}{n} \leq s_0\right\} + f_{(2)} I\left\{\frac{t}{n} > s_0\right\},$$

for some functions $f_{(1)}, f_{(2)} : \mathbb{R}^d \to \mathbb{R}$ with $f_{(1)} \not\equiv f_{(2)}$. Then for all $\boldsymbol{z} \in \mathbb{R}^d$ and $s \leq s_0$, it can be shown that

$$\hat{T}_n(s, \boldsymbol{z}) = \sqrt{n} s(1 - s_0) \int\limits_{(-\infty, \boldsymbol{z}]} \left(m_{(1)}(\boldsymbol{u}) - m_{(2)}(\boldsymbol{u})\right) \frac{f_{(1)}(\boldsymbol{u}) f_{(2)}(\boldsymbol{u})}{s_0 f_{(1)}(\boldsymbol{u}) + (1 - s_0) f_{(2)}(\boldsymbol{u})} d\boldsymbol{u}$$
$$+ o_P(\sqrt{n}).$$

Remark (local alternatives). Often it is of interest to also consider local alternatives such as

$$H_{1l} : \exists s_0 \in (0,1) : m_{n,t}(\cdot) = \begin{cases} m(\cdot), & t = 1, \ldots, \lfloor n s_0 \rfloor \\ m(\cdot) + \frac{1}{\sqrt{n}} \tilde{m}(\cdot), & t = \lfloor n s_0 \rfloor + 1, \ldots, n \end{cases},$$

for some $m, \tilde{m} : \mathbb{R}^d \to \mathbb{R}$ not depending on t and n with $\tilde{m} \not\equiv 0$. For an example see Subsection 6.1.4. Then $\boldsymbol{z}_0 \in \mathbb{R}^d$ can be chosen such that

$$\Delta(s_0, \boldsymbol{z}_0) := \int\limits_{(-\infty, \boldsymbol{z}_0]} \tilde{m}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) \bar{f}^{(s_0)}(\boldsymbol{u}) d\boldsymbol{u} \neq 0.$$

It can furthermore be shown that under H_{1l}

$$\hat{T}_n(s_0, \boldsymbol{z}_0) = \Gamma_n(s_0, 1, \boldsymbol{z}_0) - \Gamma_n(1, s_0, \boldsymbol{z}_0) + \Delta(s_0, \boldsymbol{z}_0) + o_P(1)$$

holds for all $n \in \mathbb{N}$, where

$$\Gamma_n(s, t, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} \frac{\bar{f}^{(t)}(\boldsymbol{X}_{n,i})}{\bar{f}^{(1)}(\boldsymbol{X}_{n,i})} I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\},$$

for $s, t \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. The random variable $\Gamma_n(s_0, 1, \boldsymbol{z}_0) - \Gamma_n(1, s_0, \boldsymbol{z}_0)$ converges in distribution to a centered normal random variable. This suggests that the test can also detect local alternatives such as H_{1l}. Note that the process $\{\Gamma_n(s, t, \boldsymbol{z}) : s, t \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$ is also of interest in Section 4.2, see (4.4) on page 65.

3.4 Other tests and the one-dimensional case

In this section, other possible test statistics, that also use the process $\hat{T}_n$, will be presented. Secondly, the special case of one-dimensional covariates will be investigated. It will result in a distribution free limiting distribution.

The test for change in the regression function using T_{n1} is based on the supremum over all $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$ of $\hat{T}_n(s,\boldsymbol{z})$. In principle, every continuous functional of the process $\hat{T}_n$ can be considered. The most common ones additional to the Kolmogorov-Smirnov type are the so called *Cramér-von Mises* type functionals, like the following ones

$$T_{n2} := \sup_{\boldsymbol{z}\in\mathbb{R}^d} \int_0^1 \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 ds, \tag{3.26}$$

$$T_{n3} := \sup_{s\in[0,1]} \int_{\mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 w(\boldsymbol{z})d\boldsymbol{z}, \tag{3.27}$$

$$T_{n4} := \int_0^1 \int_{\mathbb{R}^d} \left|\hat{T}_n(s,\boldsymbol{z})\right|^2 w(\boldsymbol{z})d\boldsymbol{z}ds, \tag{3.28}$$

where $w : \mathbb{R}^d \to \mathbb{R}$ is some weighting function such that the integrals in (3.27) and (3.28) exist. Using Corollary 3.3 and the continuous mapping theorem, it follows that

$$T_{n2} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{\boldsymbol{z}\in\mathbb{R}^d} \int_0^1 |G_0(s,\boldsymbol{z})|^2 ds,$$

$$T_{n3} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]} \int_{\mathbb{R}^d} |G_0(s,\boldsymbol{z})|^2 w(\boldsymbol{z})d\boldsymbol{z},$$

$$T_{n4} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \int_{\mathbb{R}^d} |G_0(s,\boldsymbol{z})|^2 w(\boldsymbol{z})d\boldsymbol{z}ds.$$

If $d = 1$ holds, a nice limiting distribution for the test statistics T_{n1} to T_{n4} can be obtained. To see this, let

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

Furthermore, let

$$c := \Sigma(\infty) = \int \sigma^2(u)f(u)du < \infty.$$

The first fundamental theorem of calculus states that if $\sigma^2 f$ is continuous, then Σ is differentiable and for all $x \in \mathbb{R}$ it holds that

$$\frac{d\Sigma(x)}{dx} = \sigma^2(x)f(x).$$

Let next $\{K_0(s,t) : s \in [0,1], t \in [0,c]\}$ be a centered Gaussian process with covariance function

$$\mathrm{Cov}(K_0(s_1,t_1), K_0(s_2,t_2)) = (s_1 \wedge s_2 - s_1 s_2)(t_1 \wedge t_2).$$

Such a process is a two parameter Gaussian process that is a Brownian bridge in the first parameter and a Brownian motion in the second one. It is often referred to as a *Kiefer-Müller* process (see for instance Example A.2.13 in [75]). Then under continuity of σ^2, it holds that

$$\begin{aligned}
T_{n1} &\underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]}\sup_{z\in\mathbb{R}} |G_0(s,z)| \\
&\overset{\mathcal{D}}{=} \sup_{s\in[0,1]}\sup_{z\in\mathbb{R}} |K_0(s,\Sigma(z))| \\
&= \sup_{s\in[0,1]}\sup_{t\in[0,c]} |K_0(s,t)| \\
&\overset{\mathcal{D}}{=} \sup_{s\in[0,1]}\sup_{t\in[0,c]} \left|\sqrt{c}K_0\left(s,\tfrac{t}{c}\right)\right| \\
&= \sqrt{c}\sup_{s\in[0,1]}\sup_{t\in[0,1]} |K_0(s,t)|,
\end{aligned}$$

where continuity of Σ and the scaling property of the Brownian motion were used. Note that the first equality in distribution does not hold if $d>1$. Let furthermore

$$\hat{c}_n := \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{m}_n(X_i))^2\omega_n(X_i). \tag{3.29}$$

Then it holds that $\hat{c}_n \underset{n\to\infty}{\overset{P}{\to}} c$ and therefore by Slutsky's theorem

$$\frac{1}{\sqrt{\hat{c}_n}}T_{n1} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]}\sup_{t\in[0,1]} |K_0(s,t)|.$$

Note that any consistent estimator for the constant c can be used here. Similarly it can be seen that

$$\begin{aligned}
T_{n2} &\underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{z\in\mathbb{R}}\int_0^1 |G_0(s,z)|^2 ds \\
&\overset{\mathcal{D}}{=} \sup_{z\in\mathbb{R}}\int_0^1 |K_0(s,\Sigma(z))|^2 ds \\
&= \sup_{t\in[0,c]}\int_0^1 |K_0(s,t)|^2 ds \\
&\overset{\mathcal{D}}{=} \sup_{t\in[0,c]}\int_0^1 \left|\sqrt{c}K_0\left(s,\tfrac{t}{c}\right)\right|^2 ds \\
&= c\sup_{t\in[0,1]}\int_0^1 |K_0(s,t)|^2 ds
\end{aligned}$$

and therefore

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

For T_{n3} and T_{n4} by choosing the weighting function $w \equiv \sigma^2 f$, which however is unknown, and additionally using integration by substitution twice, it can be obtained that

$$T_{n3} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]}\int_{\mathbb{R}} |G_0(s,z)|^2\,\sigma^2(z)f(z)dz$$

$$\begin{aligned}
&\overset{\mathcal{D}}{=} \sup_{s\in[0,1]} \int_{\mathbb{R}} |K_0(s,\Sigma(z))|^2 \sigma^2(z) f(z) dz \\
&= \sup_{s\in[0,1]} \int_0^c |K_0(s,t)|^2 \, dt \\
&\overset{\mathcal{D}}{=} \sup_{s\in[0,1]} \int_0^c \left|\sqrt{c} K_0\left(s, \tfrac{t}{c}\right)\right|^2 dt \\
&\overset{\mathcal{D}}{=} c^2 \sup_{s\in[0,1]} \int_0^1 |K_0(s,t)|^2 \, dt.
\end{aligned}$$

Consequently, it holds that

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

Similarly, it can be shown that

$$T_{n4} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \int_{\mathbb{R}} |G_0(s,z)|^2 dz ds \overset{\mathcal{D}}{=} c^2 \int_0^1 \int_0^1 |K_0(s,t)|^2 dt ds,$$

and therefore

$$\frac{1}{\hat{c}_n^2} T_{n4} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \int_0^1 |K_0(s,t)|^2 dt ds.$$

Remark. With the above choice of weighting function, the test statistics T_{n3} and T_{n4} both contain the unknown quantity $\sigma^2 f$. Thus, to apply the tests they have to be modified. It holds that

$$\int_{\mathbb{R}} \left|\hat{T}_n(s,z)\right|^2 \sigma^2(z) f(z) dz = \int_{\mathbb{R}} \left|\hat{T}_n(s,z)\right|^2 \sigma^2(z) dF(z),$$

where F is the distribution function of X_1. Hence, the integral can be estimated by the sample mean $\frac{1}{n}\sum_{k=1}^n |\hat{T}_n(s,X_k)|^2 \sigma^2(X_k)$. Replacing $\sigma^2(X_k)$ by the Nadaraya-Watson estimator $\hat{\sigma}_n^2$ evaluated at X_k, an estimator for the above integral can be obtained by

$$\frac{1}{n}\sum_{k=1}^n \left|\hat{T}_n(s,X_k)\right|^2 \hat{\sigma}_n^2(X_k).$$

Hence, suitable modifications for T_{n3} and T_{n4} are

$$\sup_{s\in[0,1]} \frac{1}{n}\sum_{k=1}^n \left|\hat{T}_n(s,X_k)\right|^2 \hat{\sigma}_n^2(X_k) \tag{3.30}$$

and

$$\int_0^1 \frac{1}{n}\sum_{k=1}^n \left|\hat{T}_n(s,X_k)\right|^2 \hat{\sigma}_n^2(X_k) ds \tag{3.31}$$

respectively.

3.5 Related literature

The proposed test is a modification of Su and Xiao's CUSUM test in [71]. They consider a nonparametric time series regression model as in (1.1) and construct a test for change in the regression function. The modification is an improvement in two directions. First, Su and Xiao allow for heteroscedasticity and α-mixing data, and claim that their procedure also works for possibly non-stationary sequences even under the null of no change in the regression function. However, a result applicable to triangular arrays is necessary in this setting as possible changes in both the conditional mean and the variance function are assumed to depend on the sample size. It is not clear that the limiting distribution given in Theorem 3.1 in [71] is still valid in this case. For the proof of the limiting distribution of the new test statistic T_{n1} given in Corollary 3.3, strict stationarity under the null is needed. However, the case of non-stationary variances will be investigated and a bootstrap procedure will be proposed in Chapter 4. Secondly, as Su and Xiao's test is based on the cumulative sum of residuals, it is not consistent against simple alternatives in some cases as has been motivated in Section 3.3. To overcome this problem they use the following weighted version

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

where $w:\mathbb{R}^d\to\mathbb{R}$ is a weighting function. 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 consistency of their test as they point out themselves (see remarks to Theorem 3.2 in [71]). Depending on the alternative, it needs to be chosen appropriately in order to construct a consistent testing procedure. By using the marked empirical process of residuals, consistency can be obtained in a more generic way as it has been proven in Theorem 3.4. Possibly Stute [68] was one of the first researches to introduce the marked empirical process of residuals. The idea of using it in this setting however, stemmed from Burke and Bewa [6]. They consider an i.i.d. setting under the model (1.1) and also allow for heteroscedasticity. They construct a test for change in the regression function based on the following test statistic

$$\sqrt{n}\left(\frac{1}{\lfloor ns\rfloor}\sum_{i=1}^{\lfloor ns\rfloor}Y_iI\{\boldsymbol{X}_i\le\boldsymbol{z}\}-\frac{1}{n-\lfloor ns\rfloor}\sum_{i=\lfloor ns\rfloor+1}^{n}Y_iI\{\boldsymbol{X}_i\le\boldsymbol{z}\}\right).$$

It is the difference of two sample means marked with the indicator function. Their limiting distribution looks more complicated than the resulting one in Corollary 3.3 as it particularly contains the unknown regression function m. They propose a bootstrap version of the test as an approximation. Furthermore, the test does not allow for time series data. Well established results on empirical processes with i.i.d. random variables are used to proof the limiting distribution in their setting. In the dependent setup more sophisticated methods needed to be established to show the limiting distribution of $\hat{T}_n$.

4 Non-stationary variances

A key assumption in proving the weak convergence of $\hat{T}_n$ in Chapter 3 is the strict stationarity of the underlying process under the null. This particularly requires the conditional variance function $\sigma^2(\cdot)$, that also appears in the limiting distribution, to be stable in time. It would be of interest to construct a test that detects changes in the conditional mean function, even when the conditional variance function is also not stable. First, a brief literature review concerning this kind of settings is presented. A corresponding version of model (1.1), that allows for instabilities in the conditional variance function, will be introduced in this chapter. It will also be referred to as the model with *non-stationary variances*. Furthermore, the asymptotic behavior of the process $\hat{T}_n$ under the null in this generalized model will be discussed. Since the theory on weak convergence of sequential empirical processes with dependent and possibly non-stationary data is not available, this section rather serves as a motivation. However, similar techniques as used in the stationary setup will be suggested and references to the corresponding lemmata in Appendix A will be made. In contrast to the results under stationarity in Chapter 3, it turns out that the expected limiting distribution is rather complicated. Finally, a bootstrap version, that is a presumably valid testing procedure even under non-stationary variances, will be presented.

4.1 Literature review

Most literature assumes stationary variances of the error terms (unconditional or conditional) when testing for changes in regression, such as parameter instabilities in parametric models or changes in the nonparametric regression function in nonparametric models as considered in this thesis. However, as Wu [81] 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.

Pitarakis [62] considers estimation of and testing for a changepoint in the regression coefficient in linear regression models when one single changepoint in the error variance is present. The procedure does not allow for heteroscedasticity but for autoregressive effects. Perron and Zhou [61] consider a partial structural change linear regression model, meaning that some parameters of covariates are stable and some are unstable in time. It is a test for change in the regression coefficients allowing for a given number of changes in the error variance. They allow for heteroscedasticity but also assume stationary mixing regressors under the null, which rules out autoregression models. Kristensen's [43] semi-nonparametric approach to test parameter stability in possibly (autoregressive) time series data allows for potentially time-varying volatility but not for correlated errors and heteroscedasticity. Cai [7] considers a linear regression model with weakly dependent and strictly stationary exogenous variables which again excludes autoregressive effects but covers heteroscedasticity and instabilities in the conditional variance function. A different approach to model time series with structural instabilities can be obtained by using locally stationary processes as Vogt does in [76]. He allows for nonparametric hetereroscedastic models where both conditional mean and variance function vary over time. However, a main assumption is that they change smoothly over time satisfying some Lipschitz condition.

A few more recent results will be discussed a little more detailed. Xu [83] considers a linear regression model of the form $Y_t = \boldsymbol{X}_t^T \beta_t + v_t s_t$ with α-mixing, stationary and possibly conditional heteroscedastic error terms v_t, and with a mixing and stationary vector of regressors $\boldsymbol{X}_t$. The model thus is not applicable in the case of autoregression. The deterministic sequences β_t and s_t model instabilities in regression coefficient and error variance respectively. The proposed procedure uses the cumulative sum of OLS[1]-residuals. The limiting distribution includes the Brownian motion and some integral over the unknown function s_t. Critical values need to be estimated via Monte-Carlo simulations. Wu [81] considers a similar setting as in [83] but allows $\boldsymbol{X}_t$ to include lagged dependent variables in order to include autoregressive models. Instead of a CUSUM test, he uses an U-statistic for moment condition stability first proposed by [34]. Similar to the result of Xu, heteroscedasticity is allowed for, but the conditional variance function is assumed to be stable in time.

All these settings do not fit in the considered 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. As discussed in Section 3.5, Su and Xiao's [71] test allows for breaks in the conditional variance function. But their procedure does only seem to work for fixed breaks that do not depend on the sample size.

4.2 The model with non-stationary variances

In this section, the model in (1.1) will be modified in order to allow for non-stationary variances. As mentioned earlier, possible instabilities of the process should be modeled depending on the sample size n. Therefore, again the underlying process should be viewed as a triangular array process. Thus let $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : 1 \leq t \leq n, n \in \mathbb{N}\}$ such that

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

with

$$E[U_{n,t}|\mathcal{F}_n^t] = 0 \text{ and } E[U_{n,t}^2|\boldsymbol{X}_{n,t}] = \sigma_{n,t}^2(\boldsymbol{X}_{n,t}) \text{ a.s.},$$

for some functions $\sigma_{n,t}^2 : \mathbb{R}^d \to \mathbb{R}$ and $\mathcal{F}_n^t := \sigma(U_{n,j-1}, \boldsymbol{X}_{n,j} : j \leq t)$. This model therefore not only allows for heteroscedasticity but also for possible changes in $\sigma_{n,t}^2$.

[1]OLS: ordinary least squares

Let $\boldsymbol{X}_{n,t}$ be absolutely continuous with density function $f_{n,t}$. Given observations $(Y_{n,1}, \boldsymbol{X}_{n,1}), \ldots, (Y_{n,n}, \boldsymbol{X}_{n,n})$ consider the null hypothesis of no change in the conditional mean function

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

for some $m : \mathbb{R}^d \to \mathbb{R}$ not depending on t and n against the changepoint alternative

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

for functions $m_{(1)}, m_{(2)} : \mathbb{R}^d \to \mathbb{R}$ with $m_{(1)} \not\equiv m_{(2)}$. In this context the *stationary case* is always referring to the case where $f_{n,t}(\cdot) = f(\cdot)$ and $\sigma^2_{n,t}(\cdot) = \sigma^2(\cdot)$ for all $t = 1, \ldots, n$ and for some $f, \sigma^2 : \mathbb{R}^d \to \mathbb{R}$ not depending on t and n. The corresponding test statistic, using the sequential marked empirical process of residuals, then is

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

with residuals $\hat{U}_{n,t} := Y_{n,t} - \hat{m}_n(\boldsymbol{X}_{n,t})$. The Kolmogorov-Smirnov type test rejects the null, if

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

is large. Under H_0 it holds that

$$\begin{aligned} \hat{T}_n(s, \boldsymbol{z}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_{n,i} - \hat{m}_n(\boldsymbol{X}_{n,i})) \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} \\ &+ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (m(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i})) \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}. \end{aligned}$$

It is reasonable to proceed on the assumption that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} + o_P(1),$$

holds uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$ under suitable conditions. Note however, that the corresponding result for the stationary case is stated in Lemma A.4 on page 142 and its proof particularly needs the strict stationarity assumption. It uses a weak convergence result for sequential empirical processes, which is not available for non-stationary data to the best of our knowledge.

Furthermore, it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (m(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i})) \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}$$

$$\overset{(*)}{=} \sqrt{n} \int_{(-\infty,\boldsymbol{z}]} (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x}) \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}(\boldsymbol{x}) d\boldsymbol{x} + o_P(1)$$
$$= \sqrt{n} \int_{(-\infty,\boldsymbol{z}]} (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x}) \bar{f}^{(s)}(\boldsymbol{x}) d\boldsymbol{x} + o_P(1),$$

holds uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$ under suitable conditions and where $\bar{f}^{(s)}(\cdot) := \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}(\cdot) < \infty$ for all fixed $s \in (0,1]$. The assertion in $(*)$ is similar to Lemma A.1 on page 113. For the proof, $m - \hat{m}_n$ needs to be embedded in a smooth function class by using uniform rates of convergence on $\boldsymbol{J}_n$ for $m - \hat{m}_n$ and its partial derivatives. Note that the results from Lemma 2.2 can not be applied as they only hold for strictly stationary data. However, as suggested in Section 2.4 rates of convergence can be shown for non-stationary data under suitable conditions as well. Under the null the kernel density estimator $\hat{f}_n$ consistently estimates

$$\bar{f}_n(\cdot) := \frac{1}{n} \sum_{i=1}^{n} f_{n,i}(\cdot)$$

and $\hat{g}_n(\cdot) := \hat{m}_n(\cdot)\hat{f}_n(\cdot)$ consistently estimates $m(\cdot)\bar{f}_n(\cdot)$. Hence, $\hat{m}_n(\cdot)$ consistently estimates m and similar results can be obtained for the partial derivatives. The assertion in $(*)$ then follows by

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\hat{h}_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} - E\left[\hat{h}_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} \right] \right) = o_P(1)$$

and

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} E\left[\hat{h}_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} \right] = \sqrt{n} \int_{(-\infty,\boldsymbol{z}]} \hat{h}_n(\boldsymbol{x}) \frac{1}{n} \sum_{i=1}^{\lfloor ns \rfloor} f_{n,i}(\boldsymbol{x}) d\boldsymbol{x},$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, where $\hat{h}_n(\cdot) := (m(\cdot) - \hat{m}_n(\cdot))\omega_n(\cdot)$. Note that the corresponding assertion in the stationary case is Lemma A.1 on page 113. The key tool in proving it, is an exponential inequality for strongly mixing data (Theorem 2.1 in [46]), which does not need the strict stationarity assumption.

Proceeding by inserting the definition of $\hat{m}_n(\cdot)$ yields

$$\sqrt{n} \int_{(-\infty,\boldsymbol{z}]} \left(m(\boldsymbol{x}) - \frac{1}{n} \sum_{j=1}^{n} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j}) Y_{n,j} \frac{1}{\hat{f}_n(\boldsymbol{x})} \right) \omega_n(\boldsymbol{x}) \bar{f}^{(s)}(\boldsymbol{x}) d\boldsymbol{x}$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_{(-\infty,\boldsymbol{z}]} \left(m(\boldsymbol{x}) - m(\boldsymbol{X}_{n,j})\right) K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j}) \omega_n(\boldsymbol{x}) \frac{\bar{f}^{(s)}(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x}$$
$$- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_{n,j} \int_{(-\infty,\boldsymbol{z}]} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j}) \omega_n(\boldsymbol{x}) \frac{\bar{f}^{(s)}(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x}$$
$$= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int_{(-\infty,\boldsymbol{z}]} \left(m(\boldsymbol{x}) - m(\boldsymbol{X}_{n,j})\right) K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j}) \omega_n(\boldsymbol{x}) \frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})} d\boldsymbol{x} \qquad (4.1)$$

$$- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_{n,j} \int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j}) \omega_n(\boldsymbol{x}) \frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})} d\boldsymbol{x} + o_P(1), \tag{4.2}$$

where in the last step $\hat{f}_n$ was replaced by $\bar{f}^{(1)}$. Now it can be shown that the term in (4.1) is negligible uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Note that the corresponding result in the stationary case is Lemma A.2 on page 124. The main step of its proof is again an application of the exponential inequality in [46].

Concerning the term in (4.2) under suitable conditions it can be shown that

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

The corresponding assertion in the stationary case is stated in Lemma A.3 on page 131. The proof uses again the exponential inequality in [46]. Finally it is reasonable to assume that

$$\begin{aligned} \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_{n,j} \omega_n(\boldsymbol{X}_{n,j}) \frac{\bar{f}^{(s)}(\boldsymbol{X}_{n,j})}{\bar{f}^{(1)}(\boldsymbol{X}_{n,j})} I\{\boldsymbol{X}_{n,j} \le \boldsymbol{z}\} = \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U_{n,j} \frac{\bar{f}^{(s)}(\boldsymbol{X}_{n,j})}{\bar{f}^{(1)}(\boldsymbol{X}_{n,j})} I\{\boldsymbol{X}_{n,j} \le \boldsymbol{z}\} \\ + o_P(1) \end{aligned}$$

holds uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. However, the tools to proof this kind of conjecture are not available, as it requires a weak convergence result for sequential empirical processes with non-stationary data.

Putting the results together, this would lead to

$$\hat{T}_n(s, \boldsymbol{z}) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U_{n,i} \bar{g}^{(s)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} + o_P(1), \tag{4.3}$$

where $\bar{g}^{(s)}(\cdot) := \frac{\bar{f}^{(s)}(\cdot)}{\bar{f}^{(1)}(\cdot)}$. Even if both summands in (4.3) converge weakly, this certainly does not imply the weak convergence of $\hat{T}_n := \{\hat{T}_n(s, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$. The idea is to show weak convergence of the richer process

$$\left\{ \Gamma_n(s, t, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i} \bar{g}^{(t)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} : s, t \in [0,1], \boldsymbol{z} \in \mathbb{R}^d \right\} \tag{4.4}$$

to a centered Gaussian process $\{\Gamma(s, t, \boldsymbol{z}) : s, t \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$. This result again requires a weak convergence result for sequential empirical processes indexed in general function classes and with a weakly dependent and possibly non-stationary underlying triangular array process. Such a result is not available so far to the best of our knowledge and could not be obtained within this work. Note that the corresponding result in the stationary setup is Theorem 3.2 on page 41.

However, continuing under the conjecture that the process in (4.4) does converge weakly in $l^\infty([0,1]\times[0,1]\times\mathbb{R}^d)$, it follows from (4.3) that

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

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$, as $\bar{g}^{(1)}(\cdot) = 1$. By continuous mapping theorem the weak convergence of $\hat{T}_n$ to the centered Gaussian process $\{G_0(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$ follows. The covariance function of $\{\Gamma(s,t,\boldsymbol{z}) : s,t \in [0,1], z \in \mathbb{R}^d\}$ results in

$$\begin{aligned}
&\operatorname{Cov}(\Gamma_n(s_1,t_1,\boldsymbol{z}_1),\Gamma_n(s_2,t_2,\boldsymbol{z}_2))\\
&= E\left[\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor} U_{n,i}^2 \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\}\right]\\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z}_1\wedge\boldsymbol{z}_2]} \frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor} \sigma_{n,i}^2(\boldsymbol{u}) f_{n,i}(\boldsymbol{u})\bar{g}^{(t_1)}(\boldsymbol{u})\bar{g}^{(t_2)}(\boldsymbol{u})d\boldsymbol{u}\\
&\underset{n\to\infty}{\to} \int_{(-\boldsymbol{\infty},\boldsymbol{z}_1\wedge\boldsymbol{z}_2]} \bar{h}^{(s_1\wedge s_2)}(\boldsymbol{u})\bar{g}^{(t_1)}(\boldsymbol{u})\bar{g}^{(t_2)}(\boldsymbol{u})d\boldsymbol{u},
\end{aligned}$$

under suitable conditions and with $\bar{h}^{(s)}(\cdot) := \lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{\lfloor ns\rfloor}\sigma_{n,i}^2(\cdot)f_{n,i}(\cdot) < \infty$ for $s \in (0,1]$. The covariance function of $\{G_0(s,\boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d\}$ thus results in

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

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 3.3 on page 48 can be obtained. The convergence of the Kolmogorov-Smirnov test statistic follows by the continuous mapping theorem. Hence, under H_0 it can be shown that

$$\sup_{s\in[0,1]}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\hat{T}_n(s,\boldsymbol{z})\right| \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]}\sup_{\boldsymbol{z}\in\mathbb{R}^d}|G_0(s,\boldsymbol{z})|.$$

Note that under H_1 the model applies to the setting of Section 3.3. Thus under the assumptions of Theorem 3.4 on page 50, it holds that the test is consistent against H_1, i.e. under H_1 it holds that

$$\sup_{s\in[0,1]}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\hat{T}_n(s,\boldsymbol{z})\right| \underset{n\to\infty}{\overset{P}{\to}} \infty.$$

4.3 Bootstrap test

As the potential limiting distribution of $\hat{T}_n$ and therefore of T_{n1} is rather complicated and does not decompose in such a nice way as it does in the stationary case, the critical value using the asymptotic distribution can not be computed. An alternative procedure is to estimate the critical value using bootstrap. The idea is to construct a bootstrap version T_{n1}^* based on the sample $\mathcal{Y}_n := \{(Y_{n,1}, \boldsymbol{X}_{n,1}), \ldots, (Y_{n,n}, \boldsymbol{X}_{n,n})\}$ that somehow mimics the distribution of the original test statistic T_{n1} under the null without knowing its limiting distribution.

Given $\mathcal{Y}_n$ in each bootstrap replication $b \in \{1, \ldots, B\}$ a bootstrap version $T_{n1,b}^*$ of the test statistic T_{n1} is constructed as follows. First, define the bootstrap innovations by

$$U_{n,t}^* := \hat{U}_{n,t}\eta_t,$$

where $\{\eta_t\}$ are i.i.d. random variables, independent of $\mathcal{Y}_n$ with $E[\eta_0] = 0$, $E[\eta_0^2] = 1$ and $E[\eta_0^4] < \infty$. This so called *Wild Bootstrap* procedure was first introduced by Wu [80] and Liu [49] for linear regression with heteroscedasticity. It was used in time series context by Kreiß [39] and Hafner and Herwarzt [23] among others.[2] Then the bootstrap data is generated by

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

Note that if the original data follows an autoregression model, say $d = 1$ and $X_t = Y_{t-1}$, by the above choice the resulting bootstrap data does not follow the same structure. As was pointed out by Kreiß and Lahiri [40] this is still a reasonable choice in particular if the dependence structure of the underlying process does not show up in the asymptotic distribution. The bootstrap residuals are then defined by

$$\hat{U}_{n,t}^* := Y_{n,t}^* - \hat{m}_n^*(\boldsymbol{X}_{n,t}),$$

where $\hat{m}_n^*(\boldsymbol{x}) := \dfrac{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{n,j}}{h_n}\right) Y_{n,j}^*}{\sum_{j=1}^n K\left(\frac{\boldsymbol{x}-\boldsymbol{X}_{n,j}}{h_n}\right)}$. Then the bootstrap test statistic is defined by

$$T_{n1,b}^* := \sup_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} \left|\hat{T}_n^*(s, \boldsymbol{z})\right|,$$

where $\hat{T}_n^*(s, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \hat{U}_{n,i}^* \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\}$. For a given level $\alpha \in (0, 1)$ the bootstrap test then rejects the null, if

$$\frac{1}{B} \sum_{b=1}^{B} I\left\{T_{n1} \leq T_{n1,b}^*\right\} < \alpha.$$

Note that given $\mathcal{Y}_n$ the bootstrap variables $(\boldsymbol{X}_{n,1}, Y_{n,1}^*), \ldots, (\boldsymbol{X}_{n,n}, Y_{n,n}^*)$ fulfill the null restriction even if the original ones do not. To motivate that this method leads to a valid testing procedure, the asymptotic behavior of $\hat{T}_n^* := \{\hat{T}_n^*(s, \boldsymbol{z}) : s \in [0, 1], \boldsymbol{z} \in \mathbb{R}^d\}$ will be investigated in more detail. It will particularly be suggested that under H_0 the process $\hat{T}_n^*$ converges weakly in probability (see [22] for the definition) to G_0, which is the expected limiting distribution of $\hat{T}_n$ from Section 4.2. This ensures

[2] As an alternative to this, a triangular array $\{\eta_{n,t} : 1 \leq t \leq n\}$ (rather than an i.i.d. sequence) could be used to mimic the dependence structure of the innovations. This so called *Dependent Wild Bootstrap* was introduced by Shao [67].

that the rejection probability of the bootstrap test does not exceed the given level α under the null. Secondly, it will be conjectured that under H_1 the process $\hat{T}_n^*$ converges weakly in probability to some centered Gaussian process G_0^*. Note that this process particularly differs from G_0. Nevertheless, under H_1 it then holds that

$$\sup_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{T}_n(s,\boldsymbol{z})| \underset{n\to\infty}{\overset{P}{\to}} \infty \tag{4.5}$$

and

$$\sup_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{T}_n^*(s,\boldsymbol{z})| \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |G_0^*(s,\boldsymbol{z})| = O_P(1),$$

where here $\overset{\mathcal{D}}{\to}$ denotes convergence in distribution in probability. Thus, under H_1 the rejection probability of the bootstrap test converges to one as n tends to infinity. The bootstrap test therefore provides a consistent testing procedure for changes in the regression function even under non-stationary variances.

To motivate above conjectures, the limiting behavior of $\hat{T}_n^*$ under both H_0 and H_1 will be investigated. Let therefore P^* denote the probability conditioned on $\mathcal{Y}_n$[3] and let for any sequence of real valued random variables $(Z_n)_{n\in\mathbb{N}}$, $Z_n = o_{P^*}(1)$ be short for $P^*(|Z_n| > \epsilon) = o_P(1)$ for all $\epsilon > 0$. First, it can be obtained that

$$\begin{aligned}
\hat{T}_n^*(s,\boldsymbol{z}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_{n,i}^* - \hat{m}_n^*(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_{n,i}^* \omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\} \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (\hat{m}_n(\boldsymbol{X}_{n,i}) - \hat{m}_n^*(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \le \boldsymbol{z}\}.
\end{aligned}$$

Furthermore, it can be shown that

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

under suitable conditions. Inserting the definition of $\hat{m}_n^*$ yields

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

[3]Note that the notation P^* is also used for the different concept of outer probability, first noted in Chapter 1. It should be clear within the context which definition is intended.

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int\limits_{(-\infty,\boldsymbol{z}]} \left(\hat{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{X}_{n,j})\right) K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x} \\
&- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U^*_{n,j} \int\limits_{(-\infty,\boldsymbol{z}]} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})} d\boldsymbol{x} \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} \int\limits_{(-\infty,\boldsymbol{z}]} \left(\hat{m}_n(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{X}_{n,j})\right) K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})} d\boldsymbol{x} \qquad (4.6) \\
&- \frac{1}{\sqrt{n}} \sum_{j=1}^{n} U^*_{n,j} \int\limits_{(-\infty,\boldsymbol{z}]} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_{n,j})\omega_n(\boldsymbol{x})\frac{\bar{f}^{(s)}(\boldsymbol{x})}{\bar{f}^{(1)}(\boldsymbol{x})} d\boldsymbol{x} + o_{P^*}(1), \qquad (4.7)
\end{aligned}
$$

where in the last step $\hat{f}_n$ was replaced by $\bar{f}^{(1)}$. Concerning the term in (4.6), it can be shown that under suitable conditions

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

Concerning the term in (4.7), it can further be obtained that under suitable conditions

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

Thus a similar expansion as for $\hat{T}_n(s, \boldsymbol{z})$ in (4.3) on page 65 can be obtained in the bootstrap world as

$$
\begin{aligned}
\hat{T}^*_n(s, \boldsymbol{z}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U^*_{n,i}\omega_n(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} \\
&- \frac{1}{\sqrt{n}} \sum_{i=1}^{n} U^*_{n,i}\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(s)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}\} + o_{P^*}(1) \\
&= \Gamma^*_n(s, 1, \boldsymbol{z}) - \Gamma^*_n(1, s, \boldsymbol{z}) + o_{P^*}(1),
\end{aligned}
$$

where

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

Finally, the expected covariance functions of Γ_n and Γ^*_n under both H_0 and H_1 need to be compared. First, note that

$$
E[U^*_{n,i}U^*_{n,j}|\mathcal{Y}_n] = \begin{cases} 0 & i \neq j \\ \hat{U}^2_{n,i} & i = j, \end{cases}
$$

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 thus

$$
\begin{aligned}
&E\left[\Gamma_n^*(s_1,t_1,\boldsymbol{z}_1)\cdot\Gamma_n^*(s_2,t_2,\boldsymbol{z}_2)|\mathcal{Y}_n\right]\\
&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}\hat{U}_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\}\\
&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \qquad (4.8)\\
&+\frac{2}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}(m(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \qquad (4.9)\\
&+\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}(m(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i}))^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\}, \qquad (4.10)
\end{aligned}
$$

a.s. for fixed $s_1, s_2, t_1, t_2 \in [0,1]$ and $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathbb{R}^d$. Using the uniform rates of convergence of $\hat{m}_n$, both (4.9) and (4.10) converge to zero in probability. Concerning (4.8), it can be shown that

$$
\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \underset{n\to\infty}{\overset{P}{\to}} E\left[\Gamma(s_1,t_1,\boldsymbol{z}_1)\cdot\Gamma(s_2,t_2,\boldsymbol{z}_2)\right],
$$

where Γ is the expected limiting distribution of Γ_n in (4.4) on page 65. Thus $\hat{T}_n^*$ indeed converges weakly in probability to G_0 under H_0. If the null does not hold, then with

$$\hat{U}_{n,i} = m_{n,i}(\boldsymbol{X}_{n,i}) - \hat{m}_n(\boldsymbol{X}_{n,i}) + U_{n,i}$$

it holds that

$$
\begin{aligned}
&E\left[\Gamma_n^*(s_1,t_1,\boldsymbol{z}_1)\cdot\Gamma_n^*(s_2,t_2,\boldsymbol{z}_2)|\mathcal{Y}_n\right]\\
&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}\hat{U}_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\}\\
&=\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \qquad (4.11)\\
&+\frac{2}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}U_{n,i}(m_{n,i}(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i}))\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \qquad (4.12)\\
&+\frac{1}{n}\sum_{i=1}^{\lfloor ns_1\rfloor\wedge\lfloor ns_2\rfloor}(m_{n,i}(\boldsymbol{X}_{n,i})-\hat{m}_n(\boldsymbol{X}_{n,i}))^2\omega_n(\boldsymbol{X}_{n,i})\bar{g}^{(t_1)}(\boldsymbol{X}_{n,i})\bar{g}^{(t_2)}(\boldsymbol{X}_{n,i})I\{\boldsymbol{X}_{n,i}\le\boldsymbol{z}_1\wedge\boldsymbol{z}_2\} \qquad (4.13)
\end{aligned}
$$

a.s. 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 term in (4.11) again converges in probability to $E\left[\Gamma(s_1, t_1, \boldsymbol{z}_1) \cdot \Gamma(s_2, t_2, \boldsymbol{z}_2)\right]$. It can be shown that the term in (4.12) is negligible in probability. Finally consider the term in (4.13). Under suitable conditions it can be shown that

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

applying Lemma 2.3 on page 34 and with

$$\bar{m}_n(\cdot) = \frac{\frac{1}{n} \sum_{i=1}^{n} f_{n,i}(\boldsymbol{x}) m_{n,i}(\cdot)}{\frac{1}{n} \sum_{i=1}^{n} f_{n,i}(\cdot)}$$

from (2.20) on page 34. Under H_1 it holds that

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

for all $\boldsymbol{x} \in \mathbb{R}^d$, which was also discussed in the remark to Theorem 3.4 on page 50. Thus, it can be shown that

$$\begin{aligned}
&\frac{1}{n} \sum_{i=1}^{\lfloor ns_1 \rfloor \wedge \lfloor ns_2 \rfloor} (m_{n,i}(\boldsymbol{X}_{n,i}) - \bar{m}_n(\boldsymbol{X}_{n,i}))^2 \omega_n(\boldsymbol{X}_{n,i}) \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i}) \bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&= \frac{1}{n} \sum_{i=1}^{\lfloor ns_1 \rfloor \wedge \lfloor ns_2 \rfloor \wedge \lfloor ns_0 \rfloor} \left(m_{(1)}(\boldsymbol{X}_{n,i}) - m_{(2)}(\boldsymbol{X}_{n,i})\right)^2 \omega_n(\boldsymbol{X}_{n,i}) \left(1 - \bar{g}^{(s_0)}(\boldsymbol{X}_{n,i})\right)^2 \\
&\qquad \cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i}) \bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&+ \frac{1}{n} \sum_{i=\lfloor ns_1 \rfloor \wedge \lfloor ns_2 \rfloor \wedge \lfloor ns_0 \rfloor + 1}^{\lfloor ns_1 \rfloor \wedge \lfloor ns_2 \rfloor} \left(m_{(1)}(\boldsymbol{X}_{n,i}) - m_{(2)}(\boldsymbol{X}_{n,i})\right)^2 \omega_n(\boldsymbol{X}_{n,i}) \bar{g}^{(s_0)}(\boldsymbol{X}_{n,i})^2 \\
&\qquad \cdot \bar{g}^{(t_1)}(\boldsymbol{X}_{n,i}) \bar{g}^{(t_2)}(\boldsymbol{X}_{n,i}) I\{\boldsymbol{X}_{n,i} \leq \boldsymbol{z}_1 \wedge \boldsymbol{z}_2\} \\
&+ o_P(1).
\end{aligned}$$

It can be seen that these terms do not vanish but converge to some limit in probability. Thus the limiting distribution G_0^* then is not equal to G_0 and in particular depends on the changepoint s_0. Nevertheless the bootstrap test will still be a consistent procedure as under H_1 the original test statistic converges in probability to infinity while the bootstrap test statistic converges in distribution in probability to some non-degenerated limit.

5 Changepoint test in the conditional variance function

In this chapter, a test for change in the conditional variance function is suggested. The construction uses the idea that it can be seen as a conditional mean function in a new regression model. The test is therefore again based on the marked empirical process of new residuals. The asymptotic behavior under the null of no change in the conditional variance function will be established heuristically. In particular, methods of proof obtained earlier can be used. The ideas of a possible proof are based on the proofs of the lemmata in Appendix A. It can therefore be helpful to read the proofs first. Additionally, the behavior of the Kolmogorov-Smirnov type test statistic will be studied under a simple alternative. Finally, other tests based on Cramér-von Mises type test statistics will be presented and the special case of one-dimensional covariates will be investigated in more detail, as under this assumption a distribution free limiting distribution can be obtained.

5.1 Model and test statistic

Suppose that a changepoint test in the conditional mean function gave reason to assume stability in this function in model (1.1). Hence, let

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

where $E[U_t|\mathcal{F}^t] = 0$ a.s. and $m : \mathbb{R}^d \to \mathbb{R}$ is not depending on t. The verification of this assumption can be delicate. The test constructed in Chapter 3 only works under stationary variances. However, as suggested in Chapter 4 the bootstrap version of the changepoint test is a valid procedure to detect changes in the regression function even in the presence of changes in the variance function.

Furthermore, let the following representation for the innovations U_t hold

$$U_t = \sigma_t(\boldsymbol{X}_t)\varepsilon_t,\ t \in \mathbb{Z},$$

for some functions $\sigma_t : \mathbb{R}^d \to \mathbb{R}$ and an i.i.d. sequence $(\varepsilon_t)_{t\in\mathbb{Z}}$, such that ε_t is independent of $\boldsymbol{X}_j$ for all $j \leq t$ and fulfills $E[\varepsilon_1] = 0$, $E[\varepsilon_1^2] = 1$ and $E[\varepsilon_1^4] < \infty$. With these restrictions, σ_t^2 is the conditional variance function of Y_t conditioned on $\boldsymbol{X}_t$ as

$$\mathrm{Var}(Y_t|\boldsymbol{X}_t) = E[U_t^2|\boldsymbol{X}_t] = \sigma_t^2(\boldsymbol{X}_t)$$

holds almost surely. This also shows that σ_t^2 can be viewed as the regression function or conditional mean function in the new regression model

$$U_t^2 = \sigma_t^2(\boldsymbol{X}_t) + \xi_t,\ t \in \mathbb{Z},$$

with covariate $\boldsymbol{X}_t$, response variable U_t^2 and innovations

$$\xi_t := U_t^2 - \sigma_t^2(\boldsymbol{X}_t),\ t \in \mathbb{Z},$$

that satisfy

$$E[\xi_t|\boldsymbol{X}_t] = 0 \text{ and } E[\xi_t^2|\boldsymbol{X}_t] = \sigma_t^4(\boldsymbol{X}_t)E[(\varepsilon_t^2 - 1)^2]$$

almost surely. It will be investigated, whether the function $\sigma_t^2(\cdot)$ is stable in time t. Given observations $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ the null hypotheses

$$\tilde{H}_0 : \sigma_t^2(\cdot) = \sigma^2(\cdot),\ t = 1, \ldots, n,$$

for some $\sigma^2 : \mathbb{R}^d \to \mathbb{R}$ not depending on time t will therefore be considered.

The idea is again to base the test on the cumulative sum of residuals $U_t^2 - \hat{\sigma}_n^2(\boldsymbol{X}_t)$, where $\hat{\sigma}_n^2$ is the kernel estimator for the unknown σ^2 defined in (2.3) on page 13. The problem, that arises, is that U_t^2 is not observable either. It will therefore be replaced by the estimator $(Y_t - \hat{m}_n(\boldsymbol{X}_t))^2$ where $\hat{m}_n$ is the Nadaraya-Watson estimator for the unknown m defined in (2.2) on page 13. The CUSUM type test statistic again uses the marked empirical process of residuals, thus

$$\hat{\tilde{T}}_n := \left\{ \hat{\tilde{T}}_n(s, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d \right\},$$

where

$$\hat{\tilde{T}}_n(s, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left((Y_i - \hat{m}_n(\boldsymbol{X}_i))^2 - \hat{\sigma}_n^2(\boldsymbol{X}_i) \right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}$$

and the Kolmogorov-Smirnov type test statistic then is defined as

$$\tilde{T}_{n1} := \sup_{\boldsymbol{z} \in \mathbb{R}^d} \sup_{s \in [0,1]} \left| \hat{\tilde{T}}_n(s, \boldsymbol{z}) \right|. \tag{5.1}$$

Note that by using other functionals of $\hat{\tilde{T}}_n$, other tests can be obtained simultaneously. In Section 5.4, Cramér-von Mises type test statistics will be presented, which will be denoted by $\tilde{T}_{n2}$, $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$.

5.2 Asymptotic behavior under the null

To construct the test using $\tilde{T}_{n1}$, the limiting distribution of $\hat{\tilde{T}}_n$ under the null is of interest. This section will give a heuristic motivation under what kind of assumptions and with what kind of methods the convergence of the test statistic can be obtained.

Under $\tilde{H}_0$, strict stationary and α-mixing $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$ and under suitable conditions it can be shown that

$$\hat{\tilde{T}}_n(s, \boldsymbol{z}) = \tilde{T}_n(s, \boldsymbol{z}) - s\tilde{T}_n(1, \boldsymbol{z}) + o_P(1), \tag{5.2}$$

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

$$\tilde{T}_n(s, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}$$

and that

$$\tilde{T}_n := \left\{ \tilde{T}_n(s, \boldsymbol{z}) : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d \right\}$$

converges weakly in $l^\infty([0,1] \times \mathbb{R}^d)$ to a centered Gaussian process $\tilde{G}$ with

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

where $\tilde{\Sigma}(\boldsymbol{z}) := E[(\varepsilon_1^2 - 1)^2] \int_{(-\infty, \boldsymbol{z}]} \sigma^4(\boldsymbol{x}) f(\boldsymbol{x}) d\boldsymbol{x}$. Applying continuous mapping and Slutzky's theorem, it therefore follows that $\hat{\tilde{T}}_n$ converges weakly in $l^\infty([0,1] \times \mathbb{R}^d)$ to a centered Gaussian process $\tilde{G}_0$ with

$$\operatorname{Cov}\left(\tilde{G}_0(s_1, \boldsymbol{z}_1), \tilde{G}_0(s_2, \boldsymbol{z}_2)\right) = (s_1 \wedge s_2 - s_1 s_2)\tilde{\Sigma}(\boldsymbol{z}_1 \wedge \boldsymbol{z}_2).$$

The continuous mapping theorem then implies that

$$\tilde{T}_{n1} \underset{n \to \infty}{\overset{\mathcal{D}}{\to}} \sup_{\boldsymbol{z} \in \mathbb{R}^d} \sup_{s \in [0,1]} \left| \tilde{G}_0(s, \boldsymbol{z}) \right|.$$

The weak convergence of $\tilde{T}_n$ can be shown by using Corollary B.3. In what follows, it will be sketched how the proof of the decomposition, namely assertion (5.2), could work mentioning the kind of assumptions need to be made. Consider

$$\begin{aligned}
\hat{\tilde{T}}_n(s, \boldsymbol{z}) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left((Y_i - \hat{m}_n(\boldsymbol{X}_i))^2 - \hat{\sigma}_n^2(\boldsymbol{X}_i) \right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left((m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i) + U_i)^2 - \hat{\sigma}_n^2(\boldsymbol{X}_i) \right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i))^2 \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} && (5.3) \\
&+ \frac{2}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} U_i (m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{X}_i)) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} && (5.4) \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(U_i^2 - \hat{\sigma}_n^2(\boldsymbol{X}_i) \right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}. && (5.5)
\end{aligned}$$

Using the uniform rates of convergence of $m - \hat{m}_n$ obtained in Lemma 2.2 on page 17, the term in (5.3) is negligible uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. It can be shown that the term in (5.4) is as well negligible uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. To see this, define the function classes

$$\mathcal{F} := \{(u, \boldsymbol{x}) \mapsto uI\{\boldsymbol{x} \leq \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\}$$

and

$$\mathcal{H} := \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n),$$

from the proof of Lemma A.1 on page 113. There it is shown that for $\hat{h}_n(\boldsymbol{x}) := (m(\boldsymbol{x}) - \hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x})$ it holds that

$$P\left(\hat{h}_n \in \mathcal{H}\right) \underset{n\to\infty}{\to} 1.$$

Then the assertion is implied by

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

where $E[h(\boldsymbol{X}_i)\varphi(U_i, \boldsymbol{X}_i)] = 0$ holds for all $i = 1, \ldots, n$, $h \in \mathcal{H}$ and $\varphi \in \mathcal{F}$. Similarly to the proof of Lemma A.1, the interval $[0,1]$ can be covered with finitely many intervals and the function classes $\mathcal{F}$ and $\mathcal{H}$ can be covered with finitely many brackets. The suprema over $[0,1], \mathcal{F}$ and $\mathcal{H}$ can then be replaced by the maxima over finitely many objects, namely centers of intervals and lower and upper bounds of the brackets respectively. The maxima can then be bounded in probability using Liebscher's exponential inequality for α-mixing sequences stated in Theorem 2.1 in [46]. The resulting upper bound converges to zero as n tends to infinity. Up to this step the assumptions already made in Chapter 3 are sufficient.

Finally, concerning (5.5) using $\xi_t = U_t^2 - \sigma^2(\boldsymbol{X}_t)$ under $\tilde{H}_0$, it holds that

$$\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(U_i^2 - \hat{\sigma}_n^2(\boldsymbol{X}_i)\right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} + \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\sigma^2(\boldsymbol{X}_i) - \hat{\sigma}_n^2(\boldsymbol{X}_i)\right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}.
\end{aligned}$$

Similarly to the proof of A.4 on page 142, under suitable conditions, it can be shown that uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \xi_i \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} = \tilde{T}_n(s, \boldsymbol{z}) + o_P(1).$$

In particular, strict stationarity of $\{(\xi_t, \boldsymbol{X}_t) : t \in \mathbb{Z}\}$ and moment constraints on ξ_1 are needed here. Hence, moment constraints on both U_1^2 and $\sigma^2(\boldsymbol{X}_1)$ are necessary. Furthermore, similar to the proof of A.1 on page 113 and under suitable conditions it can be shown that uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\sigma^2(\boldsymbol{X}_i) - \hat{\sigma}_n^2(\boldsymbol{X}_i)\right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\}$$

$$= s\sqrt{n}\int \left(\sigma^2(\boldsymbol{x}) - \hat{\sigma}_n^2(\boldsymbol{x})\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x} + o_P(1).$$

For this step, uniform rates of convergence for $\sigma^2 - \hat{\sigma}_n^2$ and its partial derivatives need to be shown on $\boldsymbol{J}_n$ to embed this difference in some function class $\tilde{\mathcal{H}}$ of smooth functions. In Chapter 2, the remark on page 31 already suggested such properties for $\hat{\sigma}_n^2$. For the proof, analogues assumptions to the ones in Lemma 2.2 on page 17 need to be made. In particular, analogues conditions to **(F1)** on σ^2 are needed. Furthermore, moment constraints as in **(M)** are necessary for Y_1^2.

Continuing by inserting the definition of $\hat{\sigma}_n^2$, using $Y_i = m(\boldsymbol{X}_i) + U_i$ and finally $\xi_i = U_i^2 - \sigma^2(\boldsymbol{X}_i)$ under $\tilde{H}_0$, it holds that

$$\begin{aligned}
&\sqrt{n}\int \left(\sigma^2(\boldsymbol{x}) - \hat{\sigma}_n^2(\boldsymbol{x})\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x} \\
&= \sqrt{n}\int\left(\sigma^2(\boldsymbol{x}) - \frac{1}{n}\sum_{i=1}^n K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)(Y_i - \hat{m}_n(\boldsymbol{x}))^2\frac{1}{\hat{f}_n(\boldsymbol{x})}\right)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x} \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^n\int \left(\sigma^2(\boldsymbol{x}) - (Y_i - \hat{m}_n(\boldsymbol{x}))^2\right)K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} \\
&= -\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\int K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &(5.6)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\int \left(\sigma^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &(5.7)\\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n\int (m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{x}))^2K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &(5.8)\\
&+ \frac{2}{\sqrt{n}}\sum_{i=1}^n U_i\int (m(\boldsymbol{X}_i) - \hat{m}_n(\boldsymbol{x}))K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})I\{\boldsymbol{x} \le \boldsymbol{z}\}\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x}. &(5.9)
\end{aligned}$$

Concerning (5.6), similarly to Lemma A.3 on page 131 under suitable conditions, it can be shown that

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i \int\limits_{(-\infty,\boldsymbol{z}]} K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \xi_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\} + o_P(1), \\
&= \tilde{T}_n(1, \boldsymbol{z}) + o_P(1),
\end{aligned}$$

uniformly in $\boldsymbol{z} \in \mathbb{R}^d$ where the last equality can be shown by similar arguments as in Lemma A.4 on page 142. Additional assumptions, that need to be made for this step, need to ensure the existence of moments of ξ_1. Moment constraints on both U_1^2 and $\sigma^2(\boldsymbol{X}_1)$ are therefore necessary here. Concerning (5.7), similarly to the proof of Lemma A.2 on page 124 under certain conditions it can be shown that

$$\frac{1}{\sqrt{n}}\sum_{i=1}^n \int\limits_{(-\infty,\boldsymbol{z}]} \left(\sigma^2(\boldsymbol{x}) - \sigma^2(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{x} - \boldsymbol{X}_i)\omega_n(\boldsymbol{x})\frac{f(\boldsymbol{x})}{\hat{f}_n(\boldsymbol{x})}d\boldsymbol{x} = o_P(1),$$

uniformly in $\boldsymbol{z} \in \mathbb{R}^d$. An analogues assumption to **(F2)** has to be made for σ^2 in this step. Using the uniform convergence rates of $m - \hat{m}_n$, which also hold on the slightly larger set $\boldsymbol{I}_n = [-c_n - Ch_n, c_n + Ch_n]^d$, it can be shown that the term in (5.8) is negligible uniformly in $\boldsymbol{z} \in \mathbb{R}^d$. Finally, using similar methods as for the assertion in (5.4), it can be shown that the term in (5.9) is as well negligible uniformly in $\boldsymbol{z} \in \mathbb{R}^d$ under suitable conditions.

Putting the results together under suitable conditions, it can be shown that

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\sigma^2(\boldsymbol{X}_i) - \hat{\sigma}_n^2(\boldsymbol{X}_i) \right) \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} = -s\tilde{T}_n(1, \boldsymbol{z}) + o_P(1),$$

uniformly in $s \in [0,1]$ and $\boldsymbol{z} \in \mathbb{R}^d$. Hence, the assertion in (5.2) holds.

5.3 Consistency analysis

In this section, the asymptotic behavior of $\tilde{T}_{n1}$ defined in (5.1) under a simple fixed alternative will be investigated. It will be suggested that a similar behavior to the test in the conditional mean function can be obtained, which leads to a consistency result. Finally, it is indicated that the corresponding CUSUM tests are not consistent under some simple alternatives.

Let $\{(Y_{n,t}, \boldsymbol{X}_{n,t}) : t = 1, \ldots, n, n \in \mathbb{N}\}$ be a triangular array process satifying

$$Y_{n,t} = m(\boldsymbol{X}_{n,t}) + U_{n,t}, \; t = 1, \ldots, n, n \in \mathbb{N},$$

with regression function $m : \mathbb{R}^d \to \mathbb{R}$ and innovations such that $E[U_{n,t}|\mathcal{F}_n^t] = 0$ and $E[U_{n,t}^2|\boldsymbol{X}_{n,t}] = \sigma_{n,t}^2(\boldsymbol{X}_{n,t})$ almost surely. Given observation $(Y_{n,1}\boldsymbol{X}_{n,1}), \ldots (Y_{n,n}, \boldsymbol{X}_{n,n})$ consider the following alternative hypothesis

$$\tilde{H}_1 : \exists s_0 \in (0,1) : \sigma_{n,t}^2(\cdot) = \begin{cases} \sigma_{(1)}^2(\cdot), & t = 1, \ldots, \lfloor ns_0 \rfloor \\ \sigma_{(2)}^2(\cdot), & t = \lfloor ns_0 \rfloor + 1, \ldots, n \end{cases}, \; \forall \, n \in \mathbb{N},$$

for some functions $\sigma_{(1)}^2, \sigma_{(2)}^2 : \mathbb{R}^d \to \mathbb{R}$ with $\sigma_{(1)}^2 \not\equiv \sigma_{(2)}^2$. Let furthermore $\boldsymbol{X}_{n,t}$ be absolutely continuous with density function $f_{n,t}$. Note that again to verify that m does not depend on the time of observation t, the bootstrap test as valid testing procedure for change in the conditional mean function under non-stationary variances can be used.

As mentioned in the remark on page 35, it can be shown that $\hat{\sigma}_n^2$ is under $\tilde{H}_1$ and uniformly on $\boldsymbol{J}_n$ a consistent estimator for

$$\bar{\sigma}_n^2(\boldsymbol{x}) = \frac{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x}) \sigma_{n,i}^2(\boldsymbol{x})}{\frac{1}{n}\sum\limits_{i=1}^{n} f_{n,i}(\boldsymbol{x})}$$

$$= \left(\sigma^2_{(1)}(\boldsymbol{x}) - \sigma^2_{(2)}(\boldsymbol{x})\right) \frac{\frac{1}{n}\sum_{i=1}^{\lfloor ns_0 \rfloor} f_{n,i}(\boldsymbol{x})}{\frac{1}{n}\sum_{i=1}^{n} f_{n,i}(\boldsymbol{x})} + \sigma^2_{(2)}(\boldsymbol{x}).$$

Then under similar assumptions as in Theorem 3.4 and using the notations, used therein, it can be shown that for all $\boldsymbol{z} \in \mathbb{R}^d$ and $s \leq s_0$

$$\hat{\tilde{T}}_n(s, \boldsymbol{z}) = \sqrt{n} \int\limits_{(-\infty, \boldsymbol{z}]} \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) \bar{f}^{(s)}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) d\boldsymbol{u} + o_P(\sqrt{n})$$

and for all $\boldsymbol{z} \in \mathbb{R}^d$ and $s > s_0$

$$\hat{\tilde{T}}_n(s, \boldsymbol{z}) = \sqrt{n} \int\limits_{(-\infty, \boldsymbol{z}]} \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) \bar{f}^{(s_0)}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) d\boldsymbol{u} + o_P(\sqrt{n}).$$

By a similar argumentation as in the proof of Theorem 3.4 it holds for some $\boldsymbol{z}_0 \in \mathbb{R}^d$ that

$$\int\limits_{(-\infty, \boldsymbol{z}_0]} \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) \bar{f}^{(s_0)}(\boldsymbol{u}) \left(1 - \frac{\bar{f}^{(s_0)}(\boldsymbol{u})}{\bar{f}^{(1)}(\boldsymbol{u})}\right) d\boldsymbol{u} \neq 0$$

and therefore

$$\tilde{T}_{n1} \geq \left| \hat{\tilde{T}}_n(s_0, \boldsymbol{z}_0) \right| \underset{n \to \infty}{\overset{P}{\to}} \infty.$$

Note that using the marked empirical process of residuals is again essential for the consistency of the test. To see this, consider $\hat{\tilde{T}}_n(s, \infty)$ under $\tilde{H}_1$ and under the assumption that $f_{n,t} \equiv f$ for all $1 \leq t \leq n$, $n \in \mathbb{N}$ for some symmetric density function f and let $\sigma^2_{(1)} - \sigma^2_{(2)}$ be an odd function. For an example see Subsection 6.2.1 in the simulations. Then it can be shown that

$$\hat{\tilde{T}}(s, \infty) = \begin{cases} \sqrt{n} s(1 - s_0) \int \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) f(\boldsymbol{u}) d\boldsymbol{u} + o_P(\sqrt{n}), & s \leq s_0 \\ \sqrt{n} s_0(1 - s) \int \left(\sigma^2_{(1)}(\boldsymbol{u}) - \sigma^2_{(2)}(\boldsymbol{u})\right) f(\boldsymbol{u}) d\boldsymbol{u} + o_P(\sqrt{n}), & s > s_0 \end{cases}.$$

Now the integral vanishes although $\sigma^2_{(1)} \not\equiv \sigma^2_{(2)}$. Consequently, the test based on $\hat{\tilde{T}}_n(s, \infty)$ will not be consistent against $\tilde{H}_1$.

5.4 Other tests and the one-dimensional case

By using other continuous functionals of $\hat{\tilde{T}}_n$, different tests can be constructed instantly. Again Cramér-von Mises type tests are considered. Let therefore

$$\tilde{T}_{n2} := \sup_{\boldsymbol{z} \in \mathbb{R}^d} \int_0^1 \left| \hat{\tilde{T}}_n(s, \boldsymbol{z}) \right|^2 ds,$$

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

for some weighting function $w : \mathbb{R}^d \to \mathbb{R}$ such that the integrals used in $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ exist. Applying the continuous mapping theorem, it holds that

$$\tilde{T}_{n2} \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{\boldsymbol{z}\in\mathbb{R}^d} \int_0^1 \left|\tilde{G}_0(s,\boldsymbol{z})\right|^2 ds,$$
$$\tilde{T}_{n3} \xrightarrow[n\to\infty]{\mathcal{D}} \sup_{s\in[0,1]} \int_{\mathbb{R}^d} \left|\tilde{G}_0(s,\boldsymbol{z})\right|^2 w(\boldsymbol{z})d\boldsymbol{z},$$
$$\tilde{T}_{n4} \xrightarrow[n\to\infty]{\mathcal{D}} \int_0^1 \int_{\mathbb{R}^d} \left|\tilde{G}_0(s,\boldsymbol{z})\right|^2 w(\boldsymbol{z})d\boldsymbol{z}ds.$$

Similar to before in the case of $d = 1$ a nice limiting distribution can be obtained. To see this let $\tau^2(x) := E[(\varepsilon_1^2 - 1)^2]\sigma^4(x)$ and

$$\tilde{\Sigma} : \mathbb{R} \to \mathbb{R},\ x \mapsto \int_{(-\infty,x]} \tau^2(u)f(u)du.$$

Furthermore, let

$$\tilde{c} := \tilde{\Sigma}(\infty) = \int \tau^2(u)f(u)du < \infty.$$

Let τ^2 be continuous. Then $\tilde{\Sigma}$ is differentiable and for all $x \in \mathbb{R}$ it holds that

$$\frac{d\tilde{\Sigma}(x)}{dx} = \tau^2(x)f(x).$$

Similar to before using continuity of $\tilde{\Sigma}$ and the scaling property of the Brownian motion, it holds that

$$\tilde{T}_{n1} \xrightarrow[n\to\infty]{\mathcal{D}} \sqrt{\tilde{c}} \sup_{s\in[0,1]} \sup_{t\in[0,1]} |K_0(s,t)|,$$

where K_0 is again a Kiefer-Müller process. The constant $\tilde{c}$ can be written as

$$\begin{aligned}
\tilde{c} &= E[\tau^2(X_1)] \\
&= E[(\varepsilon_1^2 - 1)^2\sigma^4(X_1)] \\
&= E[(\varepsilon_1^2\sigma^2(X_1) - \sigma^2(X_1))^2] \\
&= E[((Y_1 - m(X_1))^2 - \sigma^2(X_1))^2].
\end{aligned}$$

Choosing

$$\hat{\tilde{c}}_n := \frac{1}{n}\sum_{i=1}^n \left(Y_i - \hat{m}_n(X_i))^2 - \hat{\sigma}_n^2(X_i)\right)^2 \omega_n(X_i),$$

where $\hat{m}_n$ and $\hat{\sigma}_n^2$ are the kernel estimators for m and σ^2 respectively, it can be shown that $\hat{\tilde{c}}_n \underset{n\to\infty}{\overset{P}{\to}} \tilde{c}$ holds and thus

$$\frac{1}{\sqrt{\hat{\tilde{c}}_n}}\tilde{T}_{n1} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \sup_{s\in[0,1]} \sup_{t\in[0,1]} |K_0(s,t)|.$$

Note that in principle any consistent estimator for $\tilde{c}$ can be used here. Similarly, it can be seen that

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

For $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ by choosing the weighting function $w \equiv \tau^2 f$, which however is unknown, and additionally using integration by substitution twice, it can be obtained that

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

and

$$\frac{1}{\hat{\tilde{c}}_n^2}\tilde{T}_{n4} \underset{n\to\infty}{\overset{\mathcal{D}}{\to}} \int_0^1 \int_0^1 |K_0(s,t)|^2 dt ds.$$

Remark. As the test statistics $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ both contain the unknown quantity $\tau^2 f$, to apply the tests they have to be modified. It holds that

$$\int_{\mathbb{R}} \left|\hat{\tilde{T}}_n(s,z)\right|^2 \tau^2(z) f(z) dz = \int_{\mathbb{R}} \left|\hat{\tilde{T}}_n(s,z)\right|^2 \tau^2(z) dF(z),$$

where F is the distribution function of X_t for all t. Hence, the integral can be estimated by the sample mean

$$\frac{1}{n}\sum_{k=1}^n |\hat{\tilde{T}}_n(s,X_k)|^2 \tau^2(X_k).$$

Using

$$\begin{aligned}
\tau^2(x) &= E[(\varepsilon_1^2-1)^2]\sigma^4(x) \\
&= E[(\varepsilon_1^2-1)^2\sigma^4(X_1)|X_1 = x] \\
&= E[((Y_1 - m(X_1))^2 - \sigma^2(X_1))^2|X_1 = x],
\end{aligned}$$

the quantity $\tau^2(x)$ can be estimated by

$$\hat{\tau}_n^2(x) := \frac{\frac{1}{n}\sum_{j=1}^n K_{h_n}(X_j - x)((Y_j - \hat{m}_n(x))^2 - \hat{\sigma}_n^2(x))^2}{\hat{f}_n(x)},$$

An estimator for the above integral can then be obtained by

$$\frac{1}{n}\sum_{k=1}^n \left|\hat{\tilde{T}}_n(s,X_k)\right|^2 \hat{\tau}_n^2(X_k).$$

Hence, suitable modifications for T_{n3} and T_{n4} are

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

and

$$\int_0^1 \frac{1}{n}\sum_{k=1}^{n} \left|\hat{\tilde{T}}_n(s, X_k)\right|^2 \hat{\tau}_n^2(X_k)ds$$

respectively.

6 Simulation study and application

To investigate the performance of the obtained tests, a simulation study is conducted, as well as applications to two real data sets are given in this chapter.

In the Monte-Carlo simulations both level and power simulations will be performed. Models under consideration include i.i.d. regression models, homoscedastic and heteroscedastic autoregression models, as well as other time series regression models. The performance of both tests in the conditional mean and variance function are investigated. For models, that satisfy the stationarity assumption under the null and have one-dimensional covariates, the critical values of the limiting distribution will be used. In the case of non-stationary variances or multidimensional covariates, the bootstrap test is applied with $B = 200$ bootstrap replications. All simulations are carried out for sample sizes $n \in \{100, 200, 300, 500\}$, a level of 5% and a break ratio of $s_0 := 0.5$ resulting in a changepoint of $k_0 := \frac{n}{2}$. Often s_0 is referred to as the changepoint as well. All tables show the rejection frequencies using 500 replications.

The real data sets under consideration are the annual flow of the river Nile in Aswan between 1871 and 1970 and the weekly closing values of the Dow Jones Industrial Average index between July 1st 1971 and August 2nd 1974. Both data sets have been extensively investigated in the changepoint framework and results can therefore be compared quite well with existing literature.

6.1 Tests in the conditional mean function

Starting with a simple i.i.d. setting, the focus will then be set on time series models, such as regression models with autoregressive exogenous variables, as well as homoscedastic and heteroscedastic autoregression models, namely models following AR(1), AR(1)-ARCH(1), AR(2), AR(2)-ARCH(1) and AR(2)-ARCH(2) processes. All testing procedures, where $d = 1$ and where the variance is stationary under the null, use the critical values of the limiting distributions of T_{n1}, T_{n2}, T_{n3} and T_{n4}, which is known, as obtained in Section 3.4. They will be compared with the tests based on $\hat{T}_n(s, \infty)$, namely with

$$\sup_{s\in[0,1]} |\hat{T}_n(s,\infty)| \text{ and } \int_0^1 |\hat{T}_n(s,\infty)|^2 ds,$$

which will be referred to as the KS and CM test, or tests based on KS and CM, respectively. For the AR(1)-ARCH(1) model with non-stationary variances, as well as the time series models with $d = 2$, the bootstrap version of the tests, introduced in Section 4.3 will be used. They will be referred to as T_{n1}^*, T_{n2}^*, T_{n3}^* and T_{n4}^*. They will also be compared with the bootstrap versions of the KS and CM tests.

6.1.1 A simple regression model

First, a simple i.i.d. regression model will be investigated. Let therefore

$$Y_t = m_t(X_t) + \varepsilon_t,\ t = 1, \dots, n$$

with covariates $(X_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ that are mutually independent. Consider two different choices for the regression function, namely

$$m_t(x) = x^2,\ t = 1, \dots, n \tag{6.1}$$

and

$$m_t(x) = \begin{cases} x^2, & t = 1, \dots, \lfloor ns_0 \rfloor \\ x^2 + x, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases}. \tag{6.2}$$

Note that model (6.1) satisfies the null hypothesis of no change in the regression function, while model (6.2) satisfies the alternative hypothesis of one change. Also note that in this case it holds for all fixed $s \in [0,1]$ and $z \in \mathbb{R}$ that

$$\hat{T}_n(s,z) = \begin{cases} -\sqrt{n}s(1-s_0) \int\limits_{-\infty}^{z} u\varphi(u)du + o_P(\sqrt{n}), & s \le s_0 \\ -\sqrt{n}s_0(1-s) \int\limits_{-\infty}^{z} u\varphi(u)du + o_P(\sqrt{n}), & s > s_0 \end{cases},$$

where φ is the standard normal density function. As

$$\int u\varphi(u)du = 0,$$

the tests based on $\hat{T}_n(s,\infty)$, namely the KS and CM tests are not consistent against H_1 in theory. Table 6.1 and 6.2 show the rejection frequencies of all tests and both models under H_0 and under H_1 respectively. Figure 6.1 is a visualization of the performance of the tests based on T_{n1} and T_{n2}, as well as the KS and CM tests under the alternative. It can be seen that under the null the rejection frequencies for all tests are near the level of 5%, however, for $n = 500$ the level is overestimated a little. Under H_1 even for the small sample size of $n = 100$ the tests based on T_{n1} to T_{n4} reject far more often than in 5% of all cases. Furthermore, it can be noted that for growing sample sizes the power of these tests increases rapidly, while the rejection frequency of the KS and CM tests remain near the level, which is consistent with the theoretical findings. To compare amongst the new tests, it is to say that T_{n2} and T_{n4} perform for smaller sample sizes a little better than T_{n1} and T_{n3}. For $n = 500$ they all reject the null in approximately 99% of the cases. In summary the tests obtained on this thesis work very well in this model and particularly outperform the CUSUM tests that are not consistent against H_1.

	$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0.052	0.066	0.042	0.062	0.058	0.056
200	0.046	0.064	0.052	0.054	0.042	0.032
300	0.066	0.062	0.052	0.052	0.068	0.040
500	0.090	0.080	0.070	0.070	0.074	0.050

Table 6.1: i.i.d. model under H_0

	$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0.180	0.264	0.232	0.254	0.070	0.064
200	0.580	0.692	0.616	0.686	0.086	0.066
300	0.786	0.860	0.808	0.860	0.086	0.066
500	0.990	0.990	0.990	0.988	0.056	0.048

Table 6.2: i.i.d. model under H_1

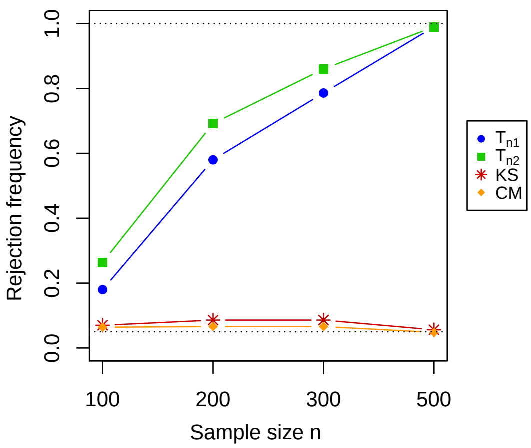

Figure 6.1: i.i.d. model under H_1

6.1.2 Heteroscedastic regression models with autoregressive exogenous variables

Consider the following heteroscedastic regression model

$$Y_t = m_t(X_t) + \sigma(X_t)\varepsilon_t,\ \ t = 1, \dots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and covariates $(X_t)_{t\in\mathbb{Z}}$ following the autoregression model

$$X_t = 0.4X_{t-1} + \xi_t,\ \ t \in \mathbb{Z}$$

with $(\xi_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ξ_t is independent of X_j for all $j \le t-1$. Let furthermore $(X_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ be mutually independent. The time series $(X_t)_{t\in\mathbb{Z}}$ is then a linear process of the form

$$X_t = \sum_{j=0}^{\infty} 0.4^j \xi_{t-j} \sim \mathcal{N}\left(0, \frac{1}{1-0.4^2}\right), \ \forall t \in \mathbb{Z}$$

and therefore in particular strictly stationary and strongly mixing with exponential mixing rates, see Example (i) on page 8. Let

$$\sigma^2(x) = 1 + 0.5x^2$$

be the conditional variance function and consider the conditional mean function

$$m_t(x) = \begin{cases} 0.5x, & t = 1, \dots, \lfloor ns_0 \rfloor \\ \left(0.5 + \Delta_0 e^{-0.8x^2}\right) x, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases},$$

with different break sizes $\Delta_0 \in \{0, 0.5, 1, 1.5, 1.5, 2, 2.5, 3, 3.5, 4\}$. For $\Delta_0 = 0$ the model satisfies the null hypothesis of no change in the regression function. The stationarity and mixing properties are inherited to $(Y_t)_{t\in\mathbb{Z}}$. For $\Delta_0 \neq 0$ the model satisfies the alternative hypothesis and the change occurs in k_0. Figure 6.2 shows the regression function under the null, as well as after the break for $\Delta_0 \in \{1, 2\}$.

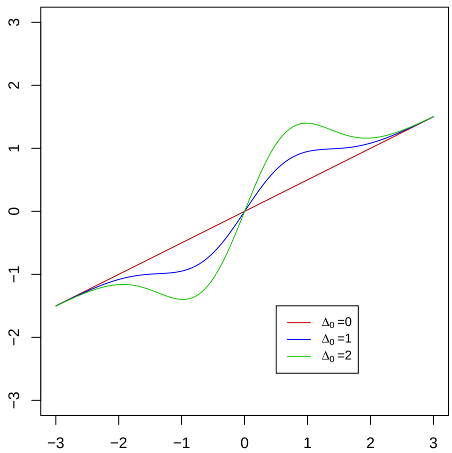

Figure 6.2: Regression function

Note that for all fixed $s \in [0,1]$ and $z \in \mathbb{R}$, it holds that

$$\hat{T}_n(s,z) = \begin{cases} -\sqrt{n} s(1-s_0)\Delta_0 \int\limits_{-\infty}^{z} e^{-0.8u^2} u\tilde{\varphi}(u)du + o_P(\sqrt{n}), & s \le s_0 \\ -\sqrt{n} s_0(1-s)\Delta_0 \int\limits_{-\infty}^{z} e^{-0.8u^2} u\tilde{\varphi}(u)du + o_P(\sqrt{n}), & s > s_0 \end{cases},$$

where $\tilde{\varphi}$ is the density function of X_t for all t, which is a centered normal random variable. Hence, the integral over the whole real line vanishes as $\tilde{\varphi}$ is an even and

$$x \mapsto e^{-0.8x^2}x$$

is an odd function. Consequently, the KS and CM tests are not consistent in theory. Table 6.3 shows the rejection frequencies in the models with break ratios $\Delta_0 \in \{0, 1, 2, 3\}$ for the tests based on T_{n1} to T_{n4}, as well as the KS and CM tests statistics. Additionally, Figure 6.3 is a visualization of the performance of T_{n1} and T_{n2}, as well as KS and CM for $n = 100$ and $n = 300$. Under the null ($\Delta_0 = 0$) the rejection frequencies for all tests are near the given level. Furthermore, it can be seen that with growing sample sizes and growing break ratios, the number of rejection generally increases. However, to just refer to Figure 6.3, it is clear that T_{n1} and T_{n2} perform much better than KS and CM. Even for the larger sample size of $n = 300$ and the large break ratio of $\Delta_0 = 4$, the rejection frequencies of KS and CM stay under 10%, while the tests based on T_{n1} and T_{n2} reject the null in 80% of all cases. The general behavior for $n = 100$ is comparable, though rejection frequencies grow less rapidly. Note that T_{n3} and T_{n4} also perform reasonably well. In summary, as the visualizations show most convincingly, the new tests again outperform the CUSUM tests by far under the alternative and also perform fairly well under the null hypothesis.

		$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	Δ_0	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0	0.038	0.086	0.068	0.074	0.050	0.064
	1	0.048	0.098	0.066	0.088	0.054	0.058
	2	0.096	0.158	0.128	0.146	0.080	0.066
	3	0.182	0.296	0.186	0.248	0.074	0.068
200	0	0.084	0.074	0.078	0.080	0.092	0.064
	1	0.086	0.116	0.106	0.104	0.064	0.050
	2	0.206	0.296	0.186	0.204	0.088	0.062
	3	0.500	0.628	0.362	0.448	0.058	0.042
300	0	0.054	0.072	0.076	0.074	0.054	0.038
	1	0.092	0.140	0.106	0.114	0.064	0.050
	2	0.334	0.446	0.258	0.296	0.086	0.060
	3	0.798	0.830	0.558	0.662	0.090	0.062
500	0	0.050	0.062	0.056	0.052	0.076	0.052
	1	0.182	0.230	0.170	0.158	0.088	0.068
	2	0.694	0.752	0.490	0.524	0.092	0.066
	3	0.978	0.990	0.850	0.932	0.104	0.086

Table 6.3: Regression with AR exogenous variables

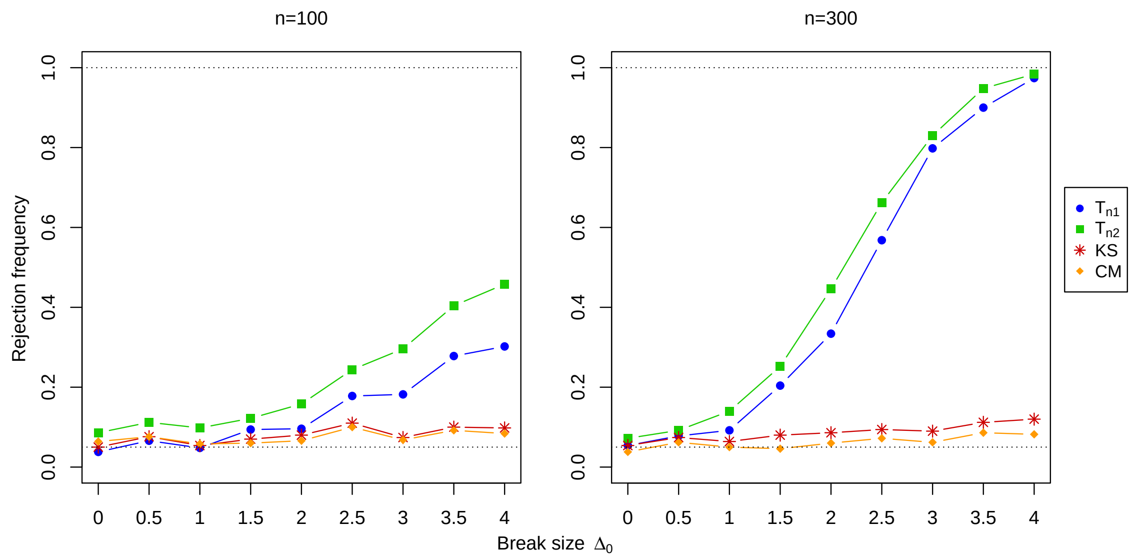

Figure 6.3: Regression with AR exogenous variables

6.1.3 Homoscedastic autogression models

Consider the following AR(1) model

$$Y_t = m_t(Y_{t-1}) + \varepsilon_t,\ t = 1, \ldots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and conditional mean function

$$m_t(x) = \begin{cases} -0.9x, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ (-0.9 + \Delta_0)\, x, & t = \lfloor ns_0 \rfloor + 1, \ldots, n \end{cases},$$

with different break sizes $\Delta_0 \in \{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8\}$. For $\Delta_0 = 0$ the model satisfies the null hypothesis of no change in the regression function. The time series $(Y_t)_{t\in\mathbb{Z}}$ is then a linear process of the form

$$Y_t = \sum_{j=0}^{\infty} (-0.9)^j \varepsilon_{t-j},\ \forall t \in \mathbb{Z}$$

and is therefore strictly stationary and strongly mixing as seen in Example (i) on page 8. For $\Delta_0 \neq 0$ the model fulfills the alternative hypothesis of one changepoint. Table 6.4 shows the rejection frequencies of the tests based on T_{n1}, T_{n2}, T_{n3} and T_{n4}, as well as the KS and the CM test statistics for $\Delta_0 \in \{0, 0.6, 1, 1.6\}$. Figure 6.4 visualizes the performance of T_{n1} and T_{n2}, as well as KS and CM for the sample sizes $n = 100$ and $n = 300$. It can be seen that the rejection frequencies for all tests are near the level of 5% under the null, i.e. for $\Delta_0 = 0$. To just refer to the visualization for $n = 100$, for $\Delta_0 \leq 0.8$ the rejection frequency is still relatively low for all the tests, though the tests based on T_{n1} and T_{n2} perform a little better. For $\Delta_0 > 0.8$ it increases with a steeper gradient for T_{n1} and T_{n2}. Note that in these cases the AR-coefficients before and after the break have a different sign. This explains the relatively large jump at this point. For $n = 300$ already for small breaks, namely for $\Delta_0 > 0.2$ the rejection frequency of T_{n1} and T_{n2} increases much faster then for KS and CM. Note that T_{n3} and T_{n4} also behave fairly well. In conclusion, the tests based on the marked empirical process perform as well as the CUSUM tests under the null and significantly better under the alternative.

		$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	Δ_0	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0	0.032	0.054	0.040	0.050	0.034	0.028
	0.6	0.116	0.148	0.146	0.166	0.078	0.068
	1	0.282	0.374	0.330	0.366	0.202	0.146
	1.6	0.776	0.812	0.728	0.810	0.504	0.414
200	0	0.046	0.066	0.044	0.060	0.050	0.050
	0.6	0.240	0.312	0.304	0.332	0.090	0.066
	1	0.594	0.668	0.606	0.656	0.180	0.102
	1.6	0.978	0.988	0.822	0.984	0.592	0.466
300	0	0.066	0.080	0.064	0.058	0.082	0.066
	0.6	0.368	0.460	0.434	0.478	0.090	0.060
	1	0.858	0.908	0.822	0.886	0.162	0.110
	1.6	0.996	0.998	0.996	0.994	0.714	0.538
500	0	0.060	0.080	0.060	0.058	0.062	0.056
	0.6	0.742	0.812	0.754	0.804	0.102	0.058
	1	0.988	0.990	0.988	0.994	0.232	0.122
	1.6	1	1	0.988	0.998	0.718	0.512

Table 6.4: AR(1) model

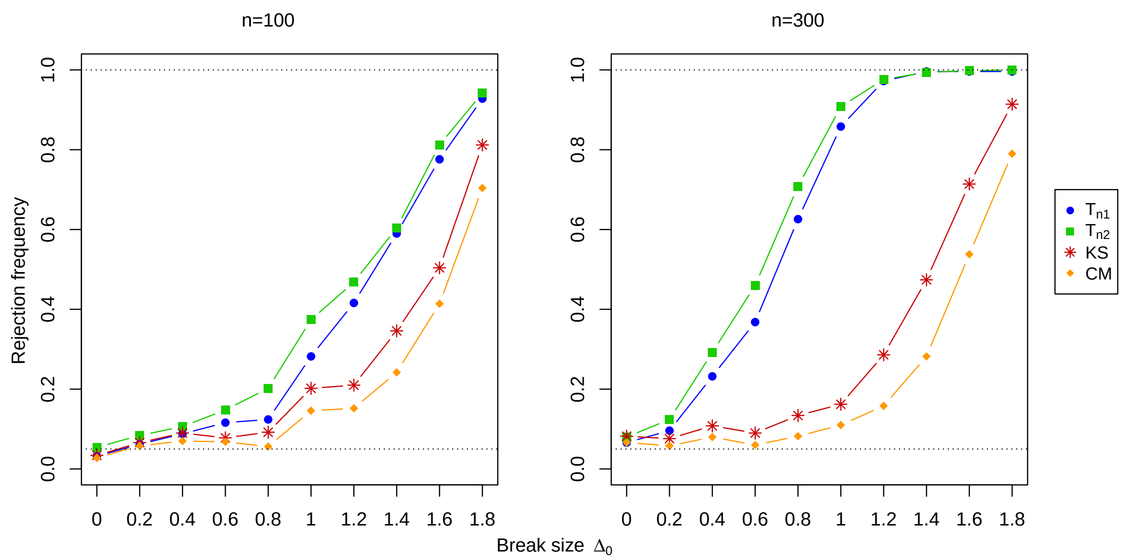

Figure 6.4: AR(1) model

6.1.4 A local alternative model

In Section 3.3, it has been motivated that the test using T_{n1} can still be a valid procedure in detecting local alternatives. To investigate this conjecture consider

$$Y_t = m_t(Y_{t-1}) + \varepsilon_t, \; t = 1, \dots, n,$$

where

$$m_t(x) = \begin{cases} -0.9x, & t = 1, \dots, \lfloor ns_0 \rfloor \\ \left(-0.9 + 5n^{-\frac{1}{2}}\right) x, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases}.$$

Note that as n tends to infinity, the conditional mean function converges to a function stable in time t. The model thus satisfies the local alternative H_{1l} as in the remark on page 56. Table 6.5 shows the rejection frequencies of T_{n1} to T_{n4}, as well as KS and CM under H_{1l}. Figure 6.5 is a visualization of the performance of the tests based on T_{n1}, T_{n2}, KS and CM. For small sample sizes the rejection frequencies are relatively low for all tests. They increase with increasing n for T_{n1} to T_{n4}. For $n = 500$ these tests reject in $20-30\%$ of all cases, where $T_{n2} - T_{n4}$ are more powerful than T_{n1}. The tests based on KS and CM do not recognize the alternative well, rejecting in less than 10% of all cases for all sample sizes under consideration.

	$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0.074	0.104	0.094	0.122	0.064	0.064
200	0.106	0.154	0.120	0.166	0.056	0.040
300	0.168	0.212	0.206	0.230	0.092	0.066
500	0.198	0.236	0.266	0.272	0.078	0.048

Table 6.5: AR(1) model under H_{1l}

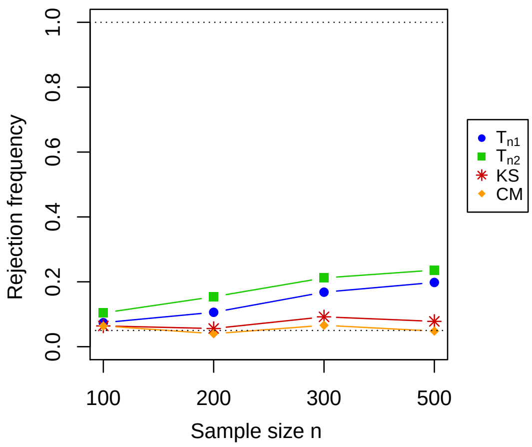

Figure 6.5: AR(1) model under H_{1l}

6.1.5 Heteroscedastic autoregression models

Consider the following AR(1)-ARCH(1) model

$$Y_t = m_t(Y_{t-1}) + \sigma(Y_{t-1})\varepsilon_t, \ t = 1, \dots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ε_t is independent of Y_j for all $j \leq t-1$. Consider the conditional mean function

$$m_t(x) = \begin{cases} -0.9x, & t = 1, \dots, \lfloor ns_0 \rfloor \\ (-0.9 + \Delta_0)x, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases},$$

with different break sizes $\Delta_0 \in \{0, 0.2, 0.4, 0.6, 0.8, 1, 1.2, 1.4, 1.6, 1.8\}$ and two different conditional variance functions, namely

$$\sigma^2(x) = 1 + 0.1x^2, \tag{6.3}$$

for a small influence of the variance and

$$\sigma^2(x) = 1 + 0.7x^2 \tag{6.4}$$

for a large influence of the variance. For $\Delta_0 = 0$ the model satisfies the null hypothesis of no changepoint. Note that for model (6.3) $(Y_t)_{t\in\mathbb{Z}}$ is strongly mixing with exponential mixing rates as $(-0.9)^2 + 0.1 < 1$ (see Example (iv) on page 9). However, the corresponding condition is not satisfied for model (6.4) as $(-0.9)^2 + 0.7 > 1$, and it is not clear if the process then possesses the required mixing properties. For $\Delta_0 \neq 0$ the model satisfies the alternative hypothesis of one changepoint in k_0. Table 6.6 and 6.7 show the rejection frequencies of T_{n1}, T_{n2}, T_{n3} and T_{n4}, as well as KS and the CM for $\Delta_0 \in \{0, 0.6, 1, 1.6\}$ in both models (6.3) and (6.4) respectively. Additionally, Figure 6.6 and 6.7 are visualizations of the performance of T_{n1} and T_{n2}, and compare them with the CUSUM tests based on the KS and CM test statistics for the sample sizes $n = 100$ and $n = 300$. A similar behavior as in the homoscedastic model can be observed. All tests approximately hold the level of 5% under the null hypothesis, i.e. for $\Delta_0 = 0$. In model (6.3) and for $n = 100$, the rejection frequency shows a steep increase for $\Delta_0 > 1$. Note that for these break ratios the sign of autoregression coefficient changes after the break. For $n = 300$ a steep increase can already be observed for $\Delta_0 > 0.2$. Note that all four tests, based on the modification, perform reasonably well, while T_{n2} is the most powerful. Comparing with the CUSUM tests, it is clear that both KS and CM tests succumb their contender, by showing a slower increase of rejections. In model (6.4) the general behavior is similar, although the rejection frequencies for all tests increase slower than in the first model. This is due to the fact that the variance is much larger causing more likely outliers. In particular, for the small sample size of $n = 100$, the rejection frequency does not exceed 50%. For larger sample sizes the rejection frequency increases, undermining the consistency of the tests even in this heteroscedastic model with a large influence of the variance. Here, the KS and CM tests also perform not as good as the tests based on the marked empirical process. In conclusion, the tests obtained in this thesis work fairly well in these heteroscedastic autoregression models under both small and large influence of the variance and furthermore again outperform the CUSUM tests, most evident from the visualizations.

		$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	Δ_0	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0	0.028	0.062	0.030	0.040	0.036	0.036
	0.6	0.088	0.154	0.146	0.168	0.076	0.064
	1	0.196	0.286	0.242	0.276	0.100	0.084
	1.6	0.732	0.782	0.666	0.736	0.486	0.382
200	0	0.048	0.052	0.058	0.062	0.070	0.044
	0.6	0.174	0.210	0.198	0.206	0.080	0.060
	1	0.458	0.566	0.414	0.476	0.132	0.088
	1.6	0.958	0.978	0.868	0.934	0.614	0.488
300	0	0.056	0.056	0.064	0.060	0.062	0.044
	0.6	0.352	0.420	0.350	0.384	0.090	0.070
	1	0.762	0.804	0.646	0.718	0.158	0.088
	1.6	0.994	0.988	0.980	0.990	0.648	0.412
500	0	0.054	0.052	0.048	0.042	0.060	0.042
	0.6	0.622	0.702	0.576	0.626	0.078	0.050
	1	0.980	0.976	0.922	0.964	0.192	0.102
	1.6	0.998	1	0.992	0.988	0.698	0.518

Table 6.6: AR(1)-ARCH(1) with $\sigma^2(x) = 1 + 0.1x^2$

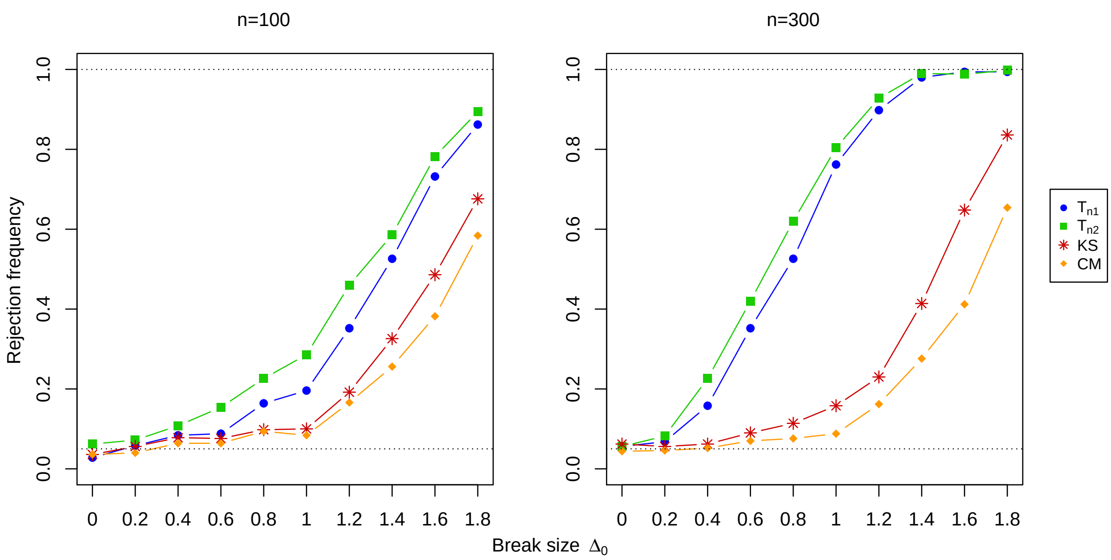

Figure 6.6: AR(1)-ARCH(1) with $\sigma^2(x) = 1 + 0.1x^2$

		$\hat{T}_n(s,z)$				$\hat{T}_n(s,\infty)$	
n	Δ_0	T_{n1}	T_{n2}	T_{n3}	T_{n4}	KS	CM
100	0	0.030	0.054	0.046	0.088	0.048	0.044
	0.6	0.058	0.104	0.086	0.108	0.060	0.066
	1	0.134	0.198	0.174	0.200	0.090	0.090
	1.6	0.380	0.462	0.320	0.398	0.210	0.182
200	0	0.052	0.076	0.066	0.080	0.054	0.050
	0.6	0.102	0.138	0.108	0.106	0.064	0.052
	1	0.222	0.280	0.194	0.224	0.088	0.092
	1.6	0.624	0.678	0.438	0.490	0.270	0.190
300	0	0.048	0.056	0.064	0.068	0.050	0.044
	0.6	0.106	0.178	0.100	0.112	0.042	0.038
	1	0.328	0.398	0.182	0.252	0.082	0.074
	1.6	0.780	0.800	0.530	0.571	0.308	0.202
500	0	0.028	0.050	0.056	0.070	0.036	0.028
	0.6	0.134	0.194	0.090	0.110	0.060	0.034
	1	0.430	0.494	0.196	0.236	0.096	0.064
	1.6	0.902	0.910	0.664	0.746	0.352	0.254

Table 6.7: AR(1)-ARCH(1) with $\sigma^2(x) = 1 + 0.7x^2$

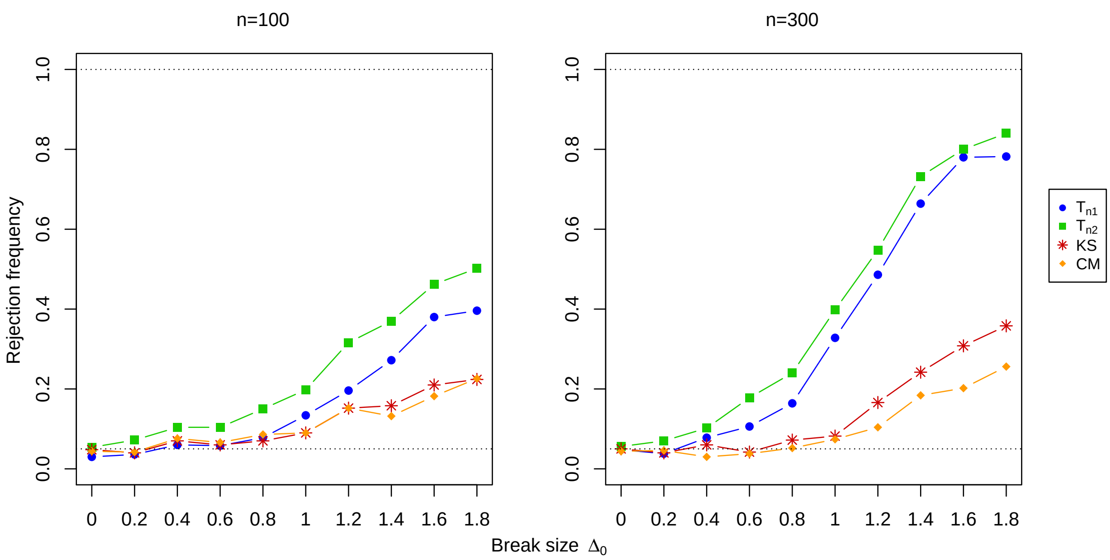

Figure 6.7: AR(1)-ARCH(1) with $\sigma^2(x) = 1 + 0.7x^2$

6.1.6 Heteroscedastic autoregression models with non-stationary variances

Consider the following AR(1)-ARCH(1) model

$$Y_t = m_t(Y_{t-1}) + \sigma_t(Y_{t-1})\varepsilon_t, \ t = 1, \dots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ε_t is independent of Y_j for all $j \leq t-1$. Consider the conditional variance function

$$\sigma_t^2(x) = \begin{cases} 1 + 0.1x^2, & t = 1, \dots, \lfloor nt_0 \rfloor \\ 1 + 0.8x^2, & t = \lfloor nt_0 \rfloor + 1, \dots, n \end{cases}$$

for some $t_0 \in (0,1)$. Consider two different choices for the conditional mean function, namely

$$m_t(x) = 0.9x, \ t = 1, \dots, n \tag{6.5}$$

and

$$m_t(x) = \begin{cases} 0.9x, & t = 1, \dots, \lfloor ns_0 \rfloor \\ -0.2x, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases}. \tag{6.6}$$

Note that model (6.5) satisfies the null hypothesis of no change in the regression function. However, even in this case the time series $(Y_t)_{t\in\mathbb{Z}}$ is not stationary, as a changepoint in the conditional variance function occurs in $\lfloor nt_0 \rfloor$. Model (6.6) satisfies the alternative hypothesis as an additional changepoint in the conditional mean function occurs in k_0. As suggested in Chapter 4, in these kind of models the bootstrap method from Section 4.3 can still be a valid testing procedure to test for changes in the regression function. Table 6.8 and 6.9 show the rejection frequencies of the bootstrap procedure using T_{n1}^*, T_{n2}^*, T_{n3}^* and T_{n4}^*, as well as the bootstrap version of the CUSUM tests KS and CM for $t_0 \in \{0.25, 0.5, 0.75\}$ under both the null and the alternative hypothesis respectively. Figure 6.8 shows a visualization of the performance of T_{n1}^*, T_{n2}^*, KS and CM under H_1. The level simulations show that all tests perform reasonably well under H_0, approximately holding the level of 5%. Furthermore, it can be seen that for all models and all tests the rejection frequency under H_1 exceeds the level, indicating that the changepoint is detected. With increasing sample size, the number of rejections increases rapidly for T_{n1}^* to T_{n4}^*, while it stays approximately constant for the bootstrap versions of KS and CM. This is most likely again 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 Section 4.3 on page 68, this is essential for the consistency property for the bootstrap tests as well. Comparing the three different models, it is to say that for $t_0 \in \{0.5, 0.75\}$ all tests perform better than for $t_0 = 0.25$. A possible explanation is that in t_0 the influence of the variance jumps from a small influence to a large influence, resulting in the fact that the tests can not detect the alternative as good for $s_0 > t_0$. As $s_0 = 0.5$ this is only the case for $t_0 = 0.25$. This is consistent with the results obtained for the AR(1)-ARCH(1) model in Subsection 6.1.5, where the tests detected occurring changepoints in the regression function less often, when the influence of the variance was larger. In conclusion, the bootstrap tests perform fairly well in these heteroscedastic autoregressive models with occurring change in the variance, as both level and power simulations indicate. Furthermore, they again outperform the bootstrap version of the CUSUM tests.

		$\hat{T}_n^*(s,z)$				$\hat{T}_n^*(s,\infty)$	
t_0	n	T_{n1}^*	T_{n2}^*	T_{n3}^*	T_{n4}^*	KS	CM
0.25	100	0.030	0.046	0.032	0.038	0.030	0.054
	200	0.044	0.044	0.054	0.044	0.056	0.040
	300	0.068	0.064	0.066	0.066	0.080	0.052
	500	0.060	0.052	0.064	0.050	0.058	0.046
0.50	100	0.068	0.048	0.074	0.060	0.068	0.056
	200	0.048	0.048	0.052	0.052	0.040	0.044
	300	0.066	0.050	0.056	0.056	0.056	0.046
	500	0.046	0.040	0.050	0.058	0.058	0.040
0.75	100	0.060	0.056	0.066	0.048	0.072	0.070
	200	0.052	0.050	0.054	0.048	0.048	0.054
	300	0.048	0.048	0.056	0.056	0.050	0.056
	500	0.034	0.040	0.060	0.048	0.046	0.056

Table 6.8: AR(1)-ARCH(1) with non-stationary variances under H_0

		$\hat{T}_n^*(s,z)$				$\hat{T}_n^*(s,\infty)$	
t_0	n	T_{n1}^*	T_{n2}^*	T_{n3}^*	T_{n4}^*	KS	CM
0.25	100	0.286	0.270	0.262	0.224	0.192	0.168
	200	0.472	0.492	0.390	0.406	0.210	0.146
	300	0.652	0.644	0.450	0.484	0.248	0.172
	500	0.878	0.868	0.590	0.694	0.264	0.194
0.50	100	0.420	0.438	0.410	0.402	0.316	0.256
	200	0.688	0.722	0.580	0.662	0.386	0.308
	300	0.868	0.894	0.734	0.838	0.378	0.292
	500	0.994	0.996	0.934	0.972	0.434	0.324
0.75	100	0.404	0.388	0.404	0.382	0.332	0.266
	200	0.638	0.636	0.596	0.626	0.290	0.212
	300	0.830	0.848	0.712	0.800	0.382	0.250
	500	0.986	0.988	0.926	0.972	0.350	0.202

Table 6.9: AR(1)-ARCH(1) with non-stationary variances under H_1

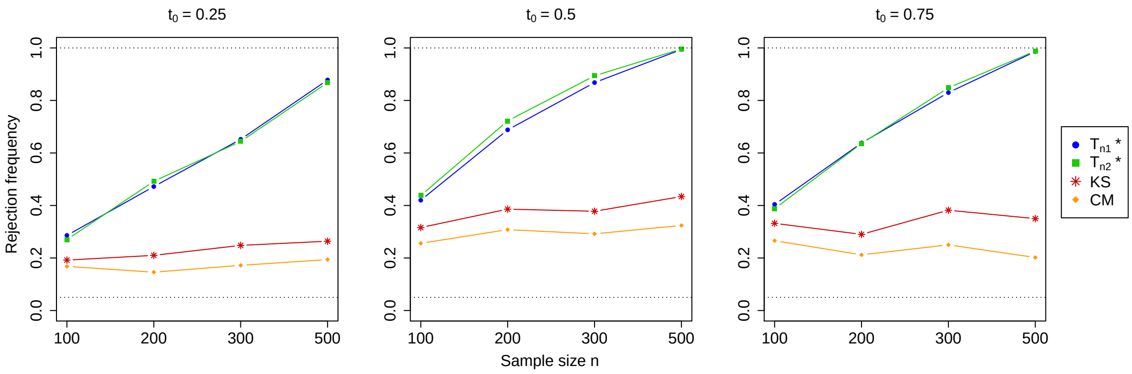

Figure 6.8: AR(1)-ARCH(1) with non-stationary variances under H_1

6.1.7 Heteroscedastic autoregression models of higher order

Let

$$Y_t = m_t(Y_{t-1}, Y_{t-2}) + \sigma(Y_{t-1}, Y_{t-2})\varepsilon_t,\ t = 1, \dots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1)$, such that ε_t is independent of Y_j for all $j \le t-1$. Consider two different choices for the conditional mean function, namely

$$m_t(x_1, x_2) = 0.9x_1 - 0.4x_2,\ t = 1, \dots, n \tag{6.7}$$

and

$$m_t(x_1, x_2) = \begin{cases} 0.9x_1 - 0.4x_2, & t = 1, \dots, \lfloor ns_0 \rfloor \\ -0.2x_1 - 0.4x_2, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases}. \tag{6.8}$$

Note that model (6.7) satisfies the null hypothesis, while model (6.8) satisfies the alternative hypothesis. Consider three different choices for the conditional variance function, namely

$$\sigma^2(x_1, x_2) = 1, \tag{6.9}$$

$$\sigma^2(x_1, x_2) = 1 + 0.4x_1^2 \tag{6.10}$$

and

$$\sigma^2(x_1, x_2) = 1 + 0.2x_1^2 + 0.2x_2^2. \tag{6.11}$$

For (6.9), an AR(2) model can be obtained. Under the null, the stationary solution is a causal linear process and hence strongly mixing with exponential mixing rates, see Example (ii) on page 9. Both (6.10) and (6.11) result in a heteroscedastic autoregression model, namely an AR(2)-ARCH(1) and AR(2)-ARCH(2) model respectively. Note that under the null, it is not clear if the required mixing properties are satisfied, as the result in [50] can not be applied (see Example (iv) on page 9). Furthermore, as mentioned in Section 3.4, the limiting distribution in these models is not known, as $d = 2$. Instead, the bootstrap procedure from Section 4.3 will be used. Table 6.10 and 6.11 show the rejection frequencies for all three models, when using the tests based on T^*_{n1} to T^*_{n4}, as well as the bootstrap versions of KS and CM under both H_0 and H_1 respectively. Figure 6.9 is a visualization of the performance of T^*_{n1}, T^*_{n2}, KS and CM under H_1. It can be seen that under H_0 the tests reject a little more often than in all other models so far, overestimating the level of 5% sometimes. Under the alternative the number of rejections increases rapidly for T^*_{n1} to T^*_{n4} with increasing n, while it stays relatively low for KS and CM. Also note that in all models the tests based T^*_{n3} and T^*_{n4} have most power. In summary, the bootstrap tests perform reasonably well and are therefore an acceptable alternative to the tests using critical values of the limiting distribution. Furthermore, in these specific models they outperform the bootstrap versions of the CUSUM tests.

		$\hat{T}^*_n(s,\boldsymbol{z})$				$\hat{T}^*_n(s,\infty)$	
Model	n	T^*_{n1}	T^*_{n2}	T^*_{n3}	T^*_{n4}	KS	CM
AR(2)	100	0.082	0.068	0.052	0.046	0.082	0.074
	200	0.074	0.072	0.046	0.036	0.064	0.072
	300	0.064	0.070	0.044	0.040	0.054	0.048
	500	0.076	0.058	0.064	0.056	0.068	0.060
AR(2)-ARCH(1)	100	0.076	0.060	0.070	0.062	0.094	0.068
	200	0.068	0.066	0.062	0.050	0.064	0.056
	300	0.084	0.098	0.064	0.060	0.086	0.096
	500	0.098	0.078	0.084	0.080	0.080	0.074
AR(2)-ARCH(2)	100	0.076	0.064	0.058	0.048	0.064	0.044
	200	0.086	0.078	0.056	0.064	0.074	0.074
	300	0.100	0.082	0.054	0.058	0.092	0.076
	500	0.082	0.068	0.056	0.064	0.076	0.056

Table 6.10: Heteroscedastic AR models of higher order under H_0

		$\hat{T}^*_n(s,\boldsymbol{z})$				$\hat{T}^*_n(s,\infty)$	
Model	n	T^*_{n1}	T^*_{n2}	T^*_{n3}	T^*_{n4}	KS	CM
AR(2)	100	0.124	0.110	0.134	0.110	0.080	0.070
	200	0.174	0.164	0.262	0.268	0.056	0.052
	300	0.284	0.308	0.440	0.454	0.096	0.070
	500	0.480	0.532	0.602	0.592	0.098	0.070
AR(2)-ARCH(1)	100	0.098	0.106	0.128	0.134	0.070	0.058
	200	0.184	0.194	0.220	0.220	0.106	0.072
	300	0.252	0.282	0.312	0.308	0.088	0.074
	500	0.476	0.484	0.504	0.520	0.120	0.078
AR(2)-ARCH(2)	100	0.096	0.104	0.102	0.102	0.072	0.050
	200	0.156	0.178	0.194	0.190	0.076	0.068
	300	0.226	0.236	0.302	0.298	0.108	0.074
	500	0.392	0.420	0.492	0.504	0.094	0.068

Table 6.11: Heteroscedastic AR models of higher order under H_1

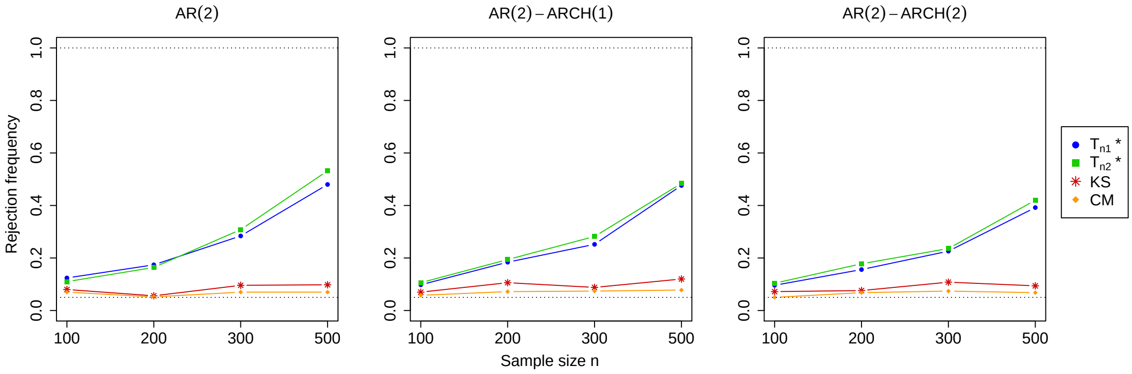

Figure 6.9: Heteroscedastic AR models of higher order under H_1

6.2 Tests in the conditional variance function

In this section, heteroscedastic models with and without change in the conditional variance function will be considered. Here, only models with stable conditional mean function are considered. To assure this assumption, the bootstrap test could be applied first to these kind of data sets. Tests in the conditional variance function use $\tilde{T}_{n1}$, $\tilde{T}_{n2}$, $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$. They will be compared with the corresponding KS and CM tests as well, meaning the Kolmogorov-Smirnov and Cramér-von Mises tests based on $\hat{\tilde{T}}(s,\infty)$ respectively.

6.2.1 Heteroscedastic regression models with autoregressive exogenous variables

Consider the following model

$$Y_t = m(X_t) + \sigma_t(X_t)\varepsilon_t,\ t = 1,\dots,n \tag{6.12}$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ and regressors $(X_t)_{t\in\mathbb{Z}}$ following the autoregression model

$$X_t = 0.4X_{t-1} + \xi_t,\ t = 1,\dots,n,$$

with $(\xi_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ξ_t is independent of X_j for all $j \le t-1$. Let furthermore $(X_t)_{t\in\mathbb{Z}}$ and $(\varepsilon_t)_{t\in\mathbb{Z}}$ be mutually independent. Note that $(X_t)_{t\in\mathbb{Z}}$ is a linear process of the form

$$X_t = \sum_{j=0}^{\infty} 0.4^j \xi_{t-j} \sim \mathcal{N}\left(0, \frac{1}{1-0.4^2}\right),\ \forall t \in \mathbb{Z}$$

and therefore strictly stationary and strongly mixing with exponential mixing rates (see Example (i) on page 8). Let

$$m(x) = 0.5x$$

be the conditional mean function and consider first the following conditional variance function

$$\sigma_t^2(x) = \begin{cases} e^{0.2x}, & t = 1,\dots,\lfloor ns_0\rfloor \\ e^{2\Delta_0+0.2x}, & t = \lfloor ns_0\rfloor + 1,\dots,n \end{cases}$$

with different break sizes $\Delta_0 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$. For $\Delta_0 = 0$ the model satisfies the null hypothesis of no change in the variance function. The stationarity and mixing properties are inherited to $(Y_t)_{t\in\mathbb{Z}}$. For $\Delta_0 \neq 0$ the model satisfies the alternative hypothesis of one changepoint in k_0. Table 6.12 shows the rejection frequencies for break ratios $\Delta_0 \in \{0, 0.2, 0.4, 0.6\}$ for the tests based on $\tilde{T}_{n1}$ to $\tilde{T}_{n4}$, as well as the tests based on the KS and CM test statistics. Additionally, Figure 6.10 shows visualizations of the performances of $\tilde{T}_{n1}$, $\tilde{T}_{n2}$, KS and CM for $n = 100$ and $n = 300$. It can be seen that the rejection frequencies for all tests are near the level under the null, where $\tilde{T}_{n2}$, $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ overestimate the level more often than $\tilde{T}_{n1}$, KS and CM. Furthermore, under the alternative with increasing break sizes, the rejection frequencies increase for all tests. For $n = 100$ the KS and CM tests perform a bit better than their modifications. For $n = 500$ all tests perform comparably well and the rejection frequencies increase much steeper than for smaller sample sizes. Hence, for larger sample sizes the tests can already detect small break sizes reasonably well. In conclusion, in this model the new tests perform in both level and power simulations fairly well. Note that the CUSUM tests are also consistent procedures in detecting these kind of alternatives.

		$\hat{\tilde{T}}_n(s,z)$				$\hat{\tilde{T}}_n(s,\infty)$	
n	Δ_0	$\tilde{T}_{n1}$	$\tilde{T}_{n2}$	$\tilde{T}_{n3}$	$\tilde{T}_{n4}$	KS	CM
100	0	0.034	0.078	0.068	0.072	0.042	0.058
	0.2	0.164	0.206	0.162	0.202	0.178	0.188
	0.4	0.474	0.542	0.484	0.520	0.540	0.516
	0.6	0.752	0.774	0.732	0.772	0.780	0.780
	0.8	0.858	0.878	0.868	0.888	0.896	0.890
200	0	0.048	0.088	0.080	0.086	0.064	0.068
	0.2	0.352	0.392	0.358	0.340	0.386	0.358
	0.4	0.874	0.890	0.888	0.878	0.900	0.886
	0.6	0.950	0.962	0.980	0.976	0.952	0.962
	0.8	0.948	0.968	0.970	0.970	0.950	0.962
300	0	0.066	0.108	0.088	0.088	0.068	0.072
	0.2	0.566	0.576	0.540	0.528	0.586	0.544
	0.4	0.954	0.966	0.970	0.964	0.964	0.962
	0.6	0.974	0.984	0.988	0.986	0.974	0.982
	0.8	0.964	0.984	0.976	0.974	0.968	0.980
500	0	0.060	0.078	0.072	0.072	0.060	0.052
	0.2	0.778	0.782	0.738	0.744	0.780	0.732
	0.4	0.986	0.992	0.998	0.998	0.988	0.992
	0.6	0.982	0.986	0.990	0.992	0.984	0.986
	0.8	0.972	0.984	0.984	0.984	0.980	0.984

Table 6.12: Regression with AR exogenous variables

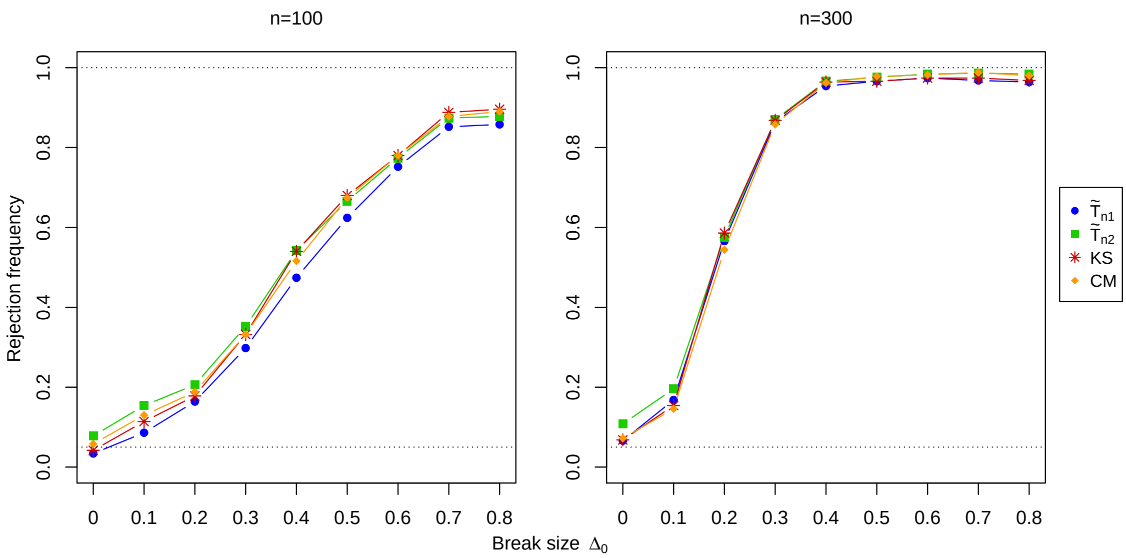

Figure 6.10: Regression with AR exogenous variables

Additionally, consider model (6.12) with $m(x) = 0.5x$ and the follwoing two choices for the conditional variance function

$$\sigma_t^2(x) = 0.25e^{-0.4x}, \; t = 1, \ldots, n, \tag{6.13}$$

and

$$\sigma_t^2(x) = \begin{cases} 0.25e^{-0.4x}, & t = 1, \ldots, \lfloor ns_0 \rfloor \\ 0.25e^{0.4x}, & t = \lfloor ns_0 \rfloor + 1, \ldots, n \end{cases}. \tag{6.14}$$

Note that model (6.13) satisfies the null hypothesis. Furthermore, model (6.14) satisfies the alternative hypothesis and in this case it holds for all $s \in [0,1]$ and $z \in \mathbb{R}$

$$\hat{\tilde{T}}_n(s,z) = \begin{cases} \sqrt{n}s(1-s_0)0.25 \int\limits_{-\infty}^{z} \left(e^{-0.4u} - e^{0.4u}\right) \tilde{\varphi}(u)du + o_P(\sqrt{n}), & s \leq s_0 \\ \sqrt{n}s_0(1-s)0.25 \int\limits_{-\infty}^{z} \left(e^{-0.4u} - e^{0.4u}\right) \tilde{\varphi}(u)du + o_P(\sqrt{n}), & s > s_0 \end{cases}$$

where $\tilde{\varphi}$ is the density function of X_t, which is a centered normal random variable. The integral over the whole real line again vanishes, as $\tilde{\varphi}$ is an even and

$$x \mapsto e^{-0.4x} - e^{0.4x}$$

an odd function. The KS and CM tests are thus not consistent in theory. Tables 6.13 and 6.14 show the rejection frequencies of the $\tilde{T}_{n1}$ to $\tilde{T}_{n4}$, as well as the KS and CM tests under both the $\tilde{H}_0$ and $\tilde{H}_1$ respectively. Additionally, Figure 6.11 is a visualization of the performance of $\tilde{T}_{n1}$, $\tilde{T}_{n2}$, KS and CM under $\tilde{H}_1$. All level simulations show reasonably good results. Furthermore, the CUSUM type tests do not recognize the alternative even under large sample sizes. The tests based on $\tilde{T}_{n1}$ to $\tilde{T}_{n4}$, however, show nice consistency properties, as they reject the null more frequently with increasing sample size. Also note that T_{n2} and T_{n4} are more powerful than T_{n1} and T_{n3}. In conclusion, in this model the new tests perform as good as the CUSUM tests under the null and much better under the alternative, as the CUSUM tests are not consistent.

	$\hat{\tilde{T}}_n(s,z)$				$\hat{\tilde{T}}_n(s,\infty)$	
n	$\tilde{T}_{n1}$	$\tilde{T}_{n2}$	$\tilde{T}_{n3}$	$\tilde{T}_{n4}$	KS	CM
100	0.056	0.076	0.058	0.084	0.062	0.046
200	0.058	0.080	0.066	0.066	0.076	0.066
300	0.070	0.092	0.064	0.076	0.070	0.068
500	0.070	0.082	0.056	0.066	0.072	0.062

Table 6.13: Regression with AR exogenous variables under $\tilde{H}_0$

	$\hat{\tilde{T}}_n(s,z)$				$\hat{\tilde{T}}_n(s,\infty)$	
n	$\tilde{T}_{n1}$	$\tilde{T}_{n2}$	$\tilde{T}_{n3}$	$\tilde{T}_{n4}$	KS	CM
100	0.078	0.138	0.154	0.166	0.062	0.074
200	0.144	0.268	0.216	0.268	0.076	0.074
300	0.222	0.338	0.290	0.348	0.072	0.068
500	0.476	0.618	0.536	0.574	0.056	0.066

Table 6.14: Regression with AR exogenous variables under $\tilde{H}_1$

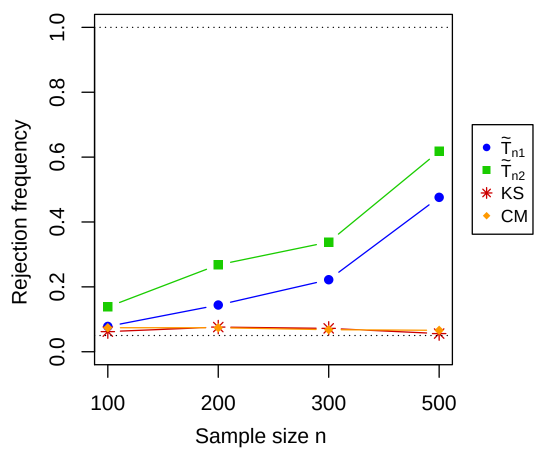

Figure 6.11: Regression with AR exogenous variables under $\tilde{H}_1$

6.2.2 Heteroscedastic autoregression models

Finally, consider the following AR(1)-ARCH(1) model

$$Y_t = m(Y_{t-1}) + \sigma_t(Y_{t-1})\varepsilon_t,\ t = 1, \dots, n$$

with innovations $(\varepsilon_t)_{t\in\mathbb{Z}} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0,1)$ such that ε_t is independent of Y_j for all $j \le t-1$. Consider the conditional variance function

$$\sigma_t^2(x) = \begin{cases} 1 + 0.1x^2, & t = 1, \dots, \lfloor ns_0 \rfloor \\ 1 + (0.1 + \Delta_0)x^2, & t = \lfloor ns_0 \rfloor + 1, \dots, n \end{cases}$$

with different break sizes $\Delta_0 \in \{0, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8\}$ and two different conditional mean functions, namely

$$m(x) = -0.5x \tag{6.15}$$

for a negative influence of the mean and

$$m(x) = 0.5x \tag{6.16}$$

for a positive influence of the mean. For $\Delta_0 = 0$ the model satisfies the null hypothesis of no changepoint. Notice that $(Y_t)_{t\in\mathbb{Z}}$ is strongly mixing with exponential mixing rates as $(0.5)^2 + 0.1 < 1$ (see Example (iv) on page 9). For $\Delta_0 \neq 0$ the model satisfies the alternative hypothesis of one changepoint in k_0. Tables 6.15 and 6.16 show the rejection frequencies for break ratios $\Delta_0 \in \{0, 0.2, 0.4, 0.6\}$ using the tests based on $\tilde{T}_{n1}$ to $\tilde{T}_{n4}$, KS and CM for both models (6.15) and (6.16) respectively. Figures 6.12 and 6.13 show visualizations of the performance of the $\tilde{T}_{n1}$, $\tilde{T}_{n2}$, KS and CM in both models. First, let model (6.15) be considered. Under the null all tests perform reasonably well, where $\tilde{T}_{n1}$ as well as KS and CM tend to approximately hold the level of 5%, while $\tilde{T}_{n2}$, $\tilde{T}_{n3}$ and $\tilde{T}_{n4}$ rather overestimate it a little. In general, it can be noted that for all sample sizes the tests based on KS and CM are more powerful than the new testing procedures. For the small sample size of $n = 100$, the rejections slightly increase with increasing break size, but do not exceed 25% for all tests. For $n = 200$ and $0.3 \le \Delta_0 \le 0.7$ the rejection frequency increases strictly for all tests, reaching a value of 25% for $\tilde{T}_{n1}$, but rather stays constant for $0.7 \le \Delta_0 \le 0.8$. This effect is even more extreme for larger sample sizes. For $n = 500$ and $\tilde{T}_{n1}$ for instance, the rejections increase rapidly for $0.1 \le \Delta_0 \le 0.6$ reaching values of 70%, but for an even larger break size $\Delta_0 \in \{0.7, 0.8\}$ it decreases again fairly dramatically down to a value of 52%. This effect can be observed for all tests, including the CUSUM tests. To find a possible explanation, the behavior of $\hat{\tilde{T}}_n(s,z)$ under different alternatives, meaning for different break sizes Δ_0 needs to be investigated a bit more detailed. As was suggested in Section 5.3, it can be shown that for all fixed $z \in \mathbb{R}$

$$\left|\hat{\tilde{T}}_n(s_0,z)\right| = \sqrt{n}\Delta_0 \int\limits_{-\infty}^{z} u^2 \frac{1}{n}\sum_{i=1}^{\lfloor ns_0 \rfloor} f_{n,i}(u)\underbrace{\left(1 - \frac{\sum\limits_{i=1}^{\lfloor ns_0 \rfloor} f_{n,i}(u)}{\sum\limits_{i=1}^{n} f_{n,i}(u)}\right)}_{(*)} du + o_P(\sqrt{n}),$$

holds, where $f_{n,t}$ is the density function of Y_{t-1}, that depends on n although this is not reflected in the notation used here. For more extreme alternatives, meaning for increasing break sizes, not only Δ_0 changes, but also the density functions after the break, i.e. $f_{n,i}$ for $i \in \{\lfloor ns_0 \rfloor + 2, \dots, n\}$ do. Hence, the fraction $(*)$ in the integral needs to be investigated. It holds, that

$$0 \leq \frac{\sum_{i=1}^{\lfloor ns_0 \rfloor} f_{n,i}(u)}{\sum_{i=1}^{n} f_{n,i}(u)} \leq 1, \ \forall \ u \in \mathbb{R}.$$

Now for very large Δ_0 the variance of Y_{i-1} is extremely high and the density function $f_{n,i}$ is rather flat for all $i \in \{\lfloor ns_0 \rfloor + 2, \dots, n\}$. As a result, the fraction in $(*)$ can be larger and the integral therefore smaller for more extreme alternatives. To be more precise for a break size of $\Delta_0 = 0.8$ the test statistic can be smaller, causing the test to reject the null less often, than in the case of $\Delta_0 = 0.6$. This explains the non-monotonically behavior of the rejection frequency with increasing break sizes. Finally, it can be seen that all tests in model (6.16) show similar results, indicating that the shape of the conditional mean function does hardly have any influence on the performance of the tests. In conclusion, both level and power simulations show fairly good results for the new tests based on the marked empirical process of residuals. However in these models they are not an improvement as the CUSUM tests show slightly better results.

		$\hat{\tilde{T}}_n(s,z)$				$\hat{\tilde{T}}_n(s,\infty)$	
n	Δ_0	$\tilde{T}_{n1}$	$\tilde{T}_{n2}$	$\tilde{T}_{n3}$	$\tilde{T}_{n4}$	KS	CM
100	0	0.046	0.080	0.086	0.098	0.040	0.046
	0.2	0.048	0.092	0.074	0.098	0.064	0.080
	0.4	0.086	0.152	0.134	0.156	0.124	0.148
	0.6	0.098	0.174	0.188	0.204	0.136	0.170
	0.8	0.122	0.160	0.220	0.214	0.160	0.160
200	0	0.044	0.088	0.060	0.072	0.058	0.066
	0.2	0.056	0.112	0.106	0.124	0.092	0.104
	0.4	0.166	0.248	0.210	0.240	0.230	0.242
	0.6	0.228	0.310	0.302	0.314	0.318	0.324
	0.8	0.254	0.296	0.350	0.380	0.330	0.344
300	0	0.058	0.068	0.076	0.082	0.060	0.042
	0.2	0.128	0.172	0.118	0.134	0.190	0.176
	0.4	0.288	0.346	0.302	0.340	0.368	0.372
	0.6	0.436	0.436	0.428	0.444	0.474	0.466
	0.8	0.428	0.428	0.464	0.450	0.460	0.438
500	0	0.060	0.074	0.072	0.060	0.058	0.050
	0.2	0.256	0.258	0.232	0.214	0.302	0.268
	0.4	0.554	0.570	0.546	0.550	0.648	0.606
	0.6	0.706	0.718	0.684	0.696	0.766	0.730
	0.8	0.522	0.570	0.578	0.602	0.616	0.584

Table 6.15: AR(1)-ARCH(1) with $m(x) = -0.5x$

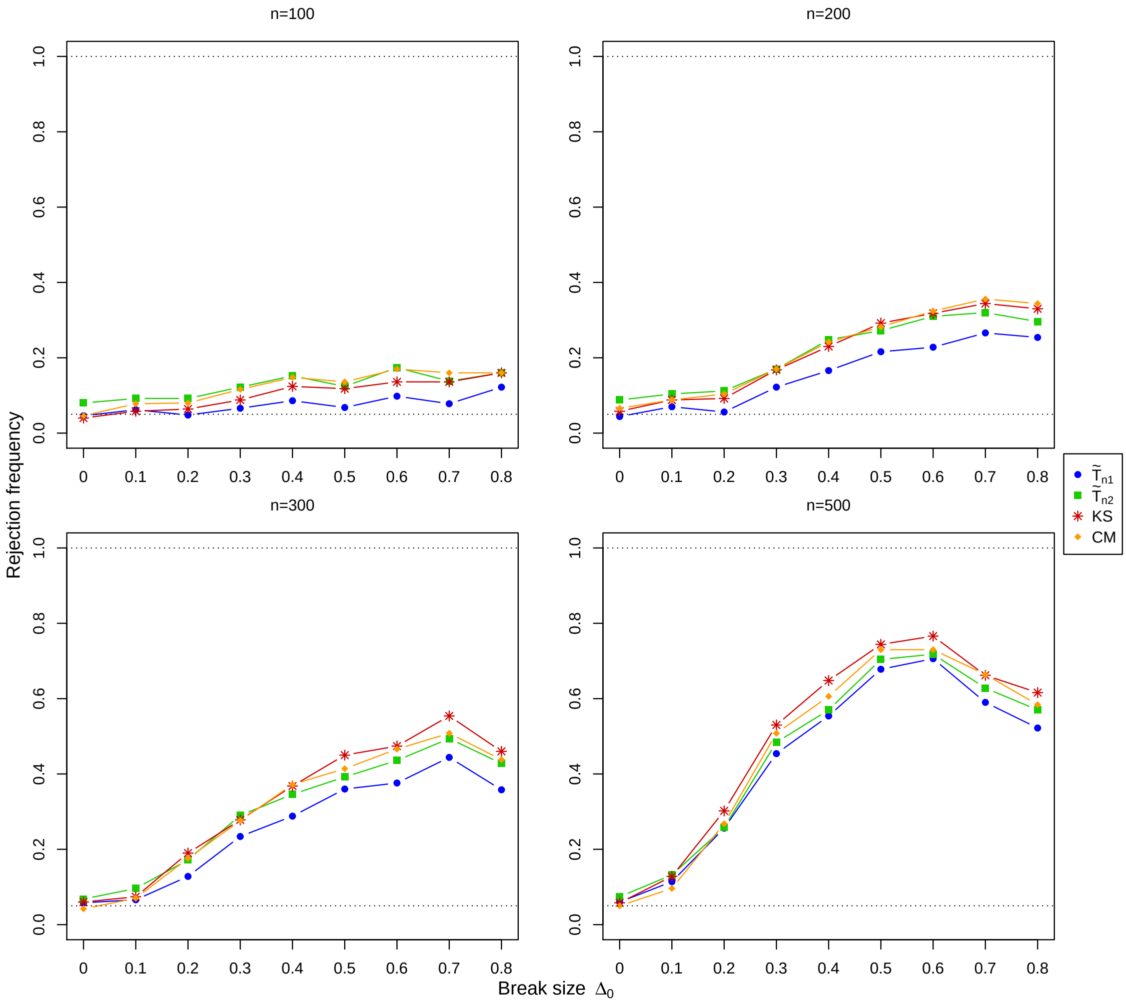

Figure 6.12: AR(1)-ARCH(1) with $m(x) = -0.5x$

		$\hat{\tilde{T}}_n(s,z)$				$\hat{\tilde{T}}_n(s,\infty)$	
n	Δ_0	$\tilde{T}_{n1}$	$\tilde{T}_{n2}$	$\tilde{T}_{n3}$	$\tilde{T}_{n4}$	KS	CM
100	0	0.050	0.090	0.078	0.089	0.060	0.050
	0.2	0.040	0.078	0.078	0.110	0.052	0.066
	0.4	0.058	0.112	0.112	0.136	0.092	0.108
	0.6	0.068	0.128	0.160	0.188	0.102	0.130
	0.8	0.094	0.118	0.182	0.200	0.144	0.138
200	0	0.064	0.087	0.060	0.056	0.058	0.058
	0.2	0.098	0.136	0.140	0.138	0.124	0.124
	0.4	0.188	0.252	0.230	0.252	0.262	0.250
	0.6	0.202	0.258	0.274	0.292	0.286	0.280
	0.8	0.270	0.288	0.360	0.358	0.366	0.328
300	0	0.074	0.084	0.078	0.074	0.080	0.072
	0.2	0.136	0.154	0.148	0.158	0.172	0.146
	0.4	0.276	0.346	0.320	0.336	0.390	0.388
	0.6	0.376	0.436	0.420	0.462	0.486	0.474
	0.8	0.368	0.434	0.474	0.470	0.458	0.442
500	0	0.066	0.084	0.080	0.072	0.064	0.052
	0.2	0.246	0.256	0.216	0.216	0.292	0.266
	0.4	0.542	0.542	0.510	0.516	0.638	0.598
	0.6	0.632	0.632	0.650	0.672	0.720	0.684
	0.8	0.522	0.522	0.592	0.596	0.582	0.562

Table 6.16: AR(1)-ARCH(1) with $m(x) = 0.5x$

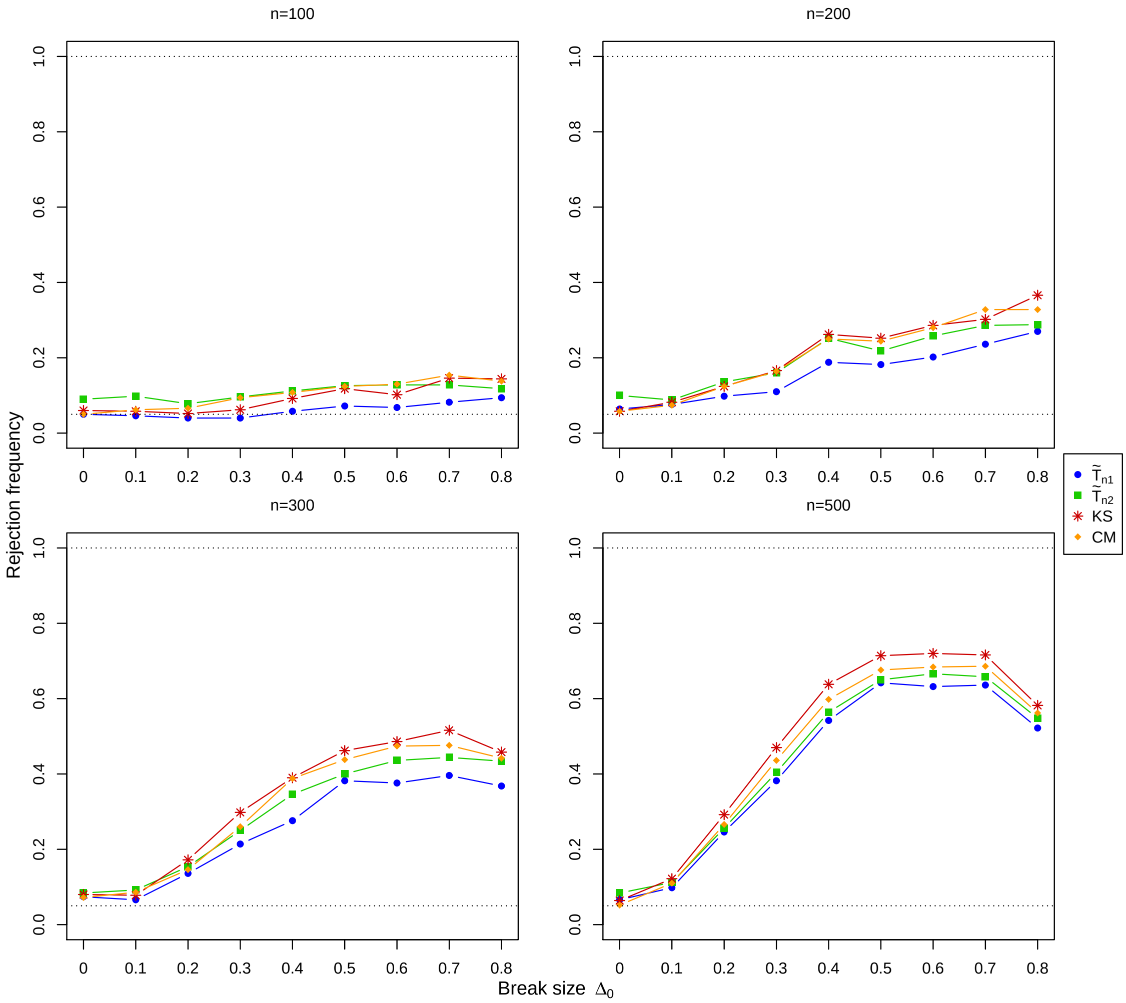

Figure 6.13: AR(1)-ARCH(1) with $m(x) = 0.5x$

6.3 Simulation techniques

In this section, it will be discussed shortly, how the simulations were conduced. In particular, the generation of the time series data, the construction of the test statistics, as well as the simulated critical values of their limiting distributions under the null will be presented. Furthermore, the choice of bandwidth and the bootstrap procedure will be explained. All simulations have been conduced with the statistics software R.

To generate a time series data set of n observations, in each case $1000+n$ random variables were generated according to the model under consideration, starting with a standard normal distributed one. To ensure the stationarity up to some possible changepoint, only the last n observations were used to construct the test statistic.

In the case of $d = 1$, for the construction of the test statistics, the so called *Epanechnikov kernel* of order $r = 4$ was used. The definition of the following version can for example be found in [71]

$$k : \mathbb{R} \to \mathbb{R},\ x \mapsto \frac{3}{4\sqrt{5}} \left(\frac{15}{8} - \frac{7}{8}x^2 \right) \left(1 - \frac{1}{5}x^2 \right) I \left\{ |x| \leq \sqrt{5} \right\}. \tag{6.17}$$

In the models where $d = 2$ the following product kernel is used

$$K : \mathbb{R}^2 \to \mathbb{R}, (x_1, x_2) \mapsto k(x_1)k(x_2),$$

with $k : \mathbb{R} \to \mathbb{R}$ from equation (6.17). Furthermore, the weighting function ω_n was chosen to be one, meaning that $\boldsymbol{J}_n = \mathbb{R}^d$ for all $n \in \mathbb{N}$. Note that this does not satisfy the condition in assumption **(J)** on page 16. In particular, in theory uniform rates of convergence for the kernel estimators can not be obtained on the whole $\mathbb{R}^d$. However, as the simulations show good results, it can be assumed that the performance of the tests is not particularly sensitive to this choice. For the standardization of all tests, $\hat{c}_n$ from (3.29) on page 58 was used. Furthermore, for the tests based on T_{n3} and T_{n4} the modified versions from (3.30) and (3.31) on page 59 were used respectively.

The choice of the bandwidth is often a difficult but also very important task as it has a large influence on the performance of the kernel estimators. In the simulations the R-function regCVBwSelC was used, which is included in the R-package locpol. It uses a cross-validation procedure to obtain an in some sense optimal bandwidth, for more information see Chapter 3 in [78].

In the case of $d = 1$ and strict stationarity, the critical values of the limiting distributions are used. Table 6.17 gives the simulated critical values of the limiting distributions of the standardized tests using T_{n1}, T_{n2}, T_{n3} and T_{n4}, namely

$$K_1 := \sup_{s\in[0,1]} \sup_{t\in[0,1]} |K_0(s,t)|,$$

$$K_2 := \sup_{t\in[0,1]} \int_0^1 |K_0(s,t)|^2 ds,$$

$$K_3 := \sup_{s\in[0,1]} \int_0^1 |K_0\left(s,t\right)|^2 dt,$$

$$K_4 := \int_0^1 \int_0^1 |K_0(s,t)|^2 dtds,$$

where $K_0 = \{K_0(s,t) : s,t \in [0,1]\}$ is the Kiefer-Müller process. Furthermore, the critical values of the standardized KS and CM tests are the corresponding functionals of a Brownian bridge process and can for instance be found in [71].

	Kiefer-Müller				Brownian bridge	
level	K_1	K_2	K_3	K_4	KS	CM
0.05	1.3866	0.5445	0.7058	0.2033	1.2620	0.4614

Table 6.17: Critical values for T_{n1},T_{n2}, T_{n3}, T_{n4}, KS and CM

Finally, some notes will be made concerning the implementation of the bootstrap test. As suggested in Section 4.3, a Wild bootstrap procedure is used in the case of $d = 2$ or if non-stationary variances occur. As for instance was done in [71], the sequence of i.i.d. random variables $\{\eta_t : 1 \leq t \leq n\}$ was generated according to a two-point distribution with masses $\frac{1+\sqrt{5}}{2\sqrt{5}}$ and $1 - \frac{1+\sqrt{5}}{2\sqrt{5}}$ in the points $\frac{1-\sqrt{5}}{2}$ and $\frac{1+\sqrt{5}}{2}$ respectively. The construction of the bootstrap data was already described briefly in Section 4.3. However, the choice of bandwidth plays a special role and will therefore be discussed a bit more detailed. As was also done by Su and Xiao [71] and many other authors, two different bandwidths were used in the procedure. To emphasis the role of the bandwidth, the Nadaraya-Watson estimator will be equipped with a second subscript h_n, if h_n is the used bandwidth. The bootstrap data is produced by

$$\begin{aligned}
&U_t^* := \hat{U}_t \eta_t \text{ with } \hat{U}_t := (Y_t - \hat{m}_{n,h_n}(\boldsymbol{X}_t))\\
&Y_t^* := \hat{m}_{n,\tilde{h}_n}(\boldsymbol{X}_t) + U_t^*,\\
&\hat{U}_n^* := Y_t^* - \hat{m}_{n,h_n}^*(\boldsymbol{X}_t).
\end{aligned}$$

Now, $\tilde{h}_n$ needs to converge to zero at a slower rate than h_n, see Härdle and Marron [27] for an heuristic explanation. The bandwidth $\tilde{h}_n$ is chosen via a cross-validation method, using the R-function regCVBwSelC and then h_n is chosen by a rule of thumb (see for instance [71] and [44]), namely $h_n := \tilde{h}_n n^{\frac{1}{9}} n^{-\frac{1}{5}}$. The residuals $\hat{U}_t$ are used to construct T_{n1} and the bootstrap version $\hat{U}_t^*$ are used to construct in each bootstrap replication $b \in \{1, \ldots, B\}$ a bootstrap version $T_{n1,b}^*$. Note that for the bootstrap test, T_{n1} was standardized with $\hat{c}_n$, while $T_{n1,b}^*$ was standardized using the bootstrap version of it, namely $\hat{c}_n^* := \frac{1}{n}\sum_{i=1}^n \hat{U}_i^{*2} \omega_n(\boldsymbol{X}_i)$. Analogous constructions lead to the bootstrap versions of T_{n2}, T_{n3} and T_{n4}.

6.4 Real data application

In this section, the tests obtained in this thesis will be applied to two real data sets, that have been used frequently in the context of changepoint analysis. The first one is the flow of the river Nile in Aswan, recorded annually between 1871 and 1970. The second data set is the DJIA index, which was collected weekly between July 1st

1971 and August 2nd 1974. Both tests in the conditional mean function and in the conditional variance function will be applied. The results will be compared briefly with existing literature.

The first example to consider is the Nile data set, obtained by the R-package datasets. It is the set of 100 measurements of the annual flow of the river Nile at Aswan in the time interval of 1871-1970 scaled by 10^8 and measured in m^3. Let Y_t be the measurement at time t for all $t \in \{1, \dots, 100\}$. Figure 6.14 shows the raw data Y_t plotted against the time.

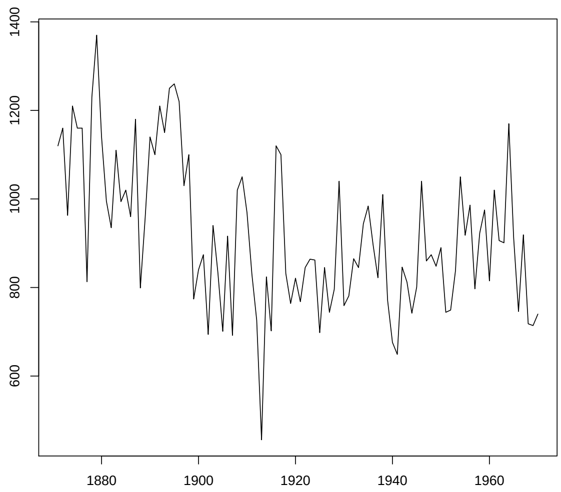

Figure 6.14: Annual flow of the Nile in Aswan

Set $X_t = Y_{t-1}$ and consider the sample $\{(Y_t, X_t) : t = 2, \dots, 100\}$ of size $n = 99$. The bootstrap test with level 5% applied to this sample rejects the null of no changepoint in the conditional mean function. The possible changepoint can be estimated by

$$\hat{s} := \arg\max_{s\in[0,1]} \sup_{z\in\mathbb{R}} |\hat{T}_n(s,z)|. \tag{6.18}$$

Figure 6.15 shows the corresponding cumulative sum, namely $\sup_{z\in\mathbb{R}} |\hat{T}_n(s,z)|$ for $s = \frac{t}{n}$ and data points $t = 1, \dots, n$. Also the critical value (red dashed horizontal line), estimated by the bootstrap procedure, as well as the changepoint, estimated by (6.18) at $\hat{s} = 27$ (green dashed vertical line), can be seen. The corresponding year of estimated changepoint is 1898. An additional application of the test to the sub data sets before and after the estimated break did not indicate the existence of a second changepoint in the mean.

The result is consistent with existing studies of this data set. Possibly the first investigation and publication of the data was done by Cobb [11]. Several other authors have analyzed this data set, including Wu [82] and Kirch and Kamgaing [35] just to mention a few. All of the results indicate a changepoint in 1989. In fact, as pointed out by all mentioned authors, that was the year, when the first damn was built in Aswan.

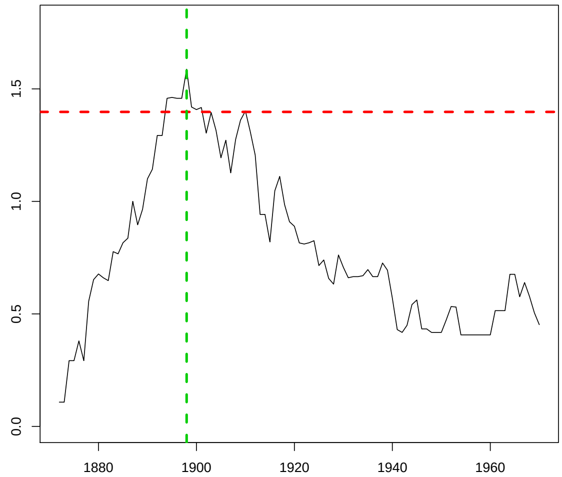

Figure 6.15: CUSUM for Nile data

Secondly, the DJIA data set obtained from the R-package strucchange will be investigated. It contains 162 return values of the DJIA index, recorded weekly between July 1st 1971 and August 2nd 1974. The differences of log-returns will be considered, namely

$$Y_t := \log(P_t) - \log(P_{t-1}),\ t = 1 \ldots, 161,$$

where P_{t-1} is the return at time t. This is a common approach when dealing with returns, as pointed out for instance by Kreiß and Neuhaus in [41]. Figure 6.16 shows the transformed raw data Y_t plotted against the time.

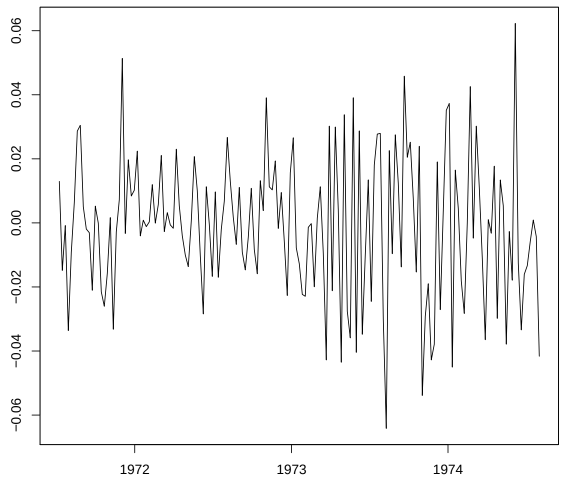

Figure 6.16: Differences of log-returns of the DJIA

Set $X_t = Y_{t-1}$ and consider the sample $\{(Y_t, X_t) : t = 2, \ldots, 161\}$ of size $n = 160$. An application of the bootstrap test for change in the conditional mean function does not reject the null of no change in mean. An additional application of the test in change in conditional variance function, using the critical value of the limiting distribution with a level of 5%, rejects the null of no change in variance. Figure 6.17 shows the cumulative sum for the test in mean (left) and in variance (right), namely $\sup_{z\in\mathbb{R}^d} |\hat{T}_n(s,z)|$ and $\sup_{z\in\mathbb{R}^d} |\hat{\hat{T}}_n(s,z)|$ respectively for $s = \frac{t}{n}$ and data points $t = 1,\ldots,n$. Furthermore, the critical values (red dashed horizontal line) and the estimated changepoint in case of existence (green dashed vertical line, right figure) can be seen as well. The estimated changepoint in the conditional variance function is $\hat{s} = 88$, which corresponds to the date of March 3rd of 1973. An additional application of the test to the sub data sets before and after the estimated break did not indicate an additional changepoint in the variance.

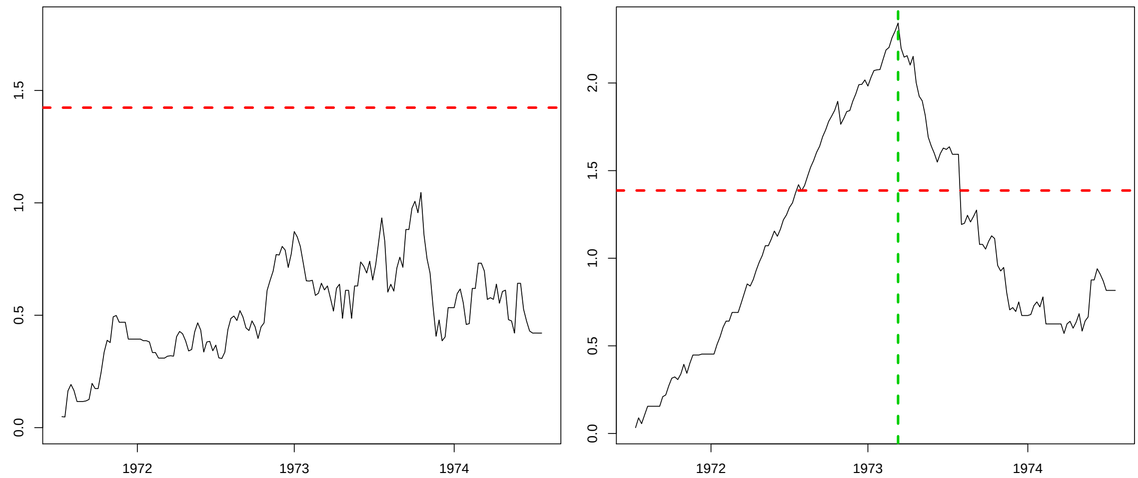

Figure 6.17: CUSUM for DJIA data - test in mean (left) and variance (right)

The results are consistent with existing studies of this data set. Possibly the first researcher that investigated this data set is Hsu [33], who detected a changepoint for the third week of March in 1973. Several other authors have suggested the existence of a changepoint in the variance in March 1973 using different kind of tests, for instance Chen and Gupta [9] and Zeileis and Honik [84], just to mention a few. As also mentioned by Hsu [33] possible reasons for a changepoint are the Watergate affair and steadily increasing prime interest rates in the U.S. during the first part of 1973.

A Proofs

In this chapter, technical and auxiliary lemmata are proven. They are needed for the decomposition of $\hat{T}_n$, stated in Theorem 3.1. Together with Theorem 3.2 it implies the weak convergence of $\hat{T}_n$ under the null and regularity assumptions.

A.1 Technical lemmata

The first section includes the proofs of Lemmata A.1, A.2, A.3 and A.4. They are all needed to proof the decomposition of $\hat{T}_n$ stated in Theorem 3.1.

Lemma A.1. *Under the assumptions of Theorem 3.1 and under H_0*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}(m(\boldsymbol{X}_i)-\hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq \boldsymbol{z}\}$$
$$= s\sqrt{n}\int_{\mathbb{R}^d}(m(\boldsymbol{x})-\hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\leq \boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x}+o_P(1)$$

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

Remark. For the proof of Lemma A.1 the difference $m-\hat{m}_n$ will be embedded in some smooth function class, using the uniform rates of convergence from Lemma 2.2. This class then will be partitioned using the bracketing notion from Definition 1.5. A second function class will be partitioned using $L_2(P)$-brackets (where $\boldsymbol{X}_t\sim P$) and bounds for the corresponding bracketing number will be obtained in Lemma A.5. Eventually the main step will be an application of Theorem 2.1 in [46], which is an exponential inequality for the sum of strongly mixing random variables.

Proof. To begin with, some notation will be introduced. Let

$$\hat{h}_n:\mathbb{R}^d\to\mathbb{R}$$
$$\boldsymbol{x}\mapsto(m(\boldsymbol{x})-\hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x}),$$

for all $n\in\mathbb{N}$ and $\mathcal{F}:=\{\boldsymbol{x}\mapsto I\{\boldsymbol{x}\leq\boldsymbol{z}\}:\boldsymbol{z}\in\mathbb{R}^d\}$. Note that $\hat{h}_n(\boldsymbol{x})=0$ for all $\boldsymbol{x}\notin\boldsymbol{J}_n$. Then

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

where $\boldsymbol{X}_t \sim P$, will be shown. Together with

$$\underbrace{\sqrt{n}\sup_{s\in[0,1]}\left|\frac{\lfloor ns\rfloor}{n}-s\right|}_{=O\left(\frac{1}{n}\right)}\underbrace{\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{x})-\hat{m}_n(\boldsymbol{x}))\omega_n(\boldsymbol{x})I\{\boldsymbol{x}\leq\boldsymbol{z}\}f(\boldsymbol{x})d\boldsymbol{x}\right|}_{=O_P(1)}=o_P(1),$$

it implies the assertion of Lemma A.1.

The proof of (A.1) consists of three main steps. In the first step, it will be shown that the function $\hat{h}_n$ lies in some function class $\mathcal{H}$ with probability converging to 1. Considering the supremum over all possible functions $h \in \mathcal{H}$ instead of $\hat{h}_n$ then simplifies the problem as the functions $h \in \mathcal{H}$ do not depend on the observations anymore (while $\hat{h}_n$ does). Moreover $\mathcal{H}$ is well understood and controlled in the sense of metric entropy properties. Secondly covering $[0,1]$ by finitely many intervals, and $\mathcal{F}$ and $\mathcal{H}$ by finitely many brackets respectively, the suprema will be bounded by the maxima over finitely many objects. Note that, while the centers of the intervalls used to cover $[0,1]$ are again elements of $[0,1]$, the lower and upper bounds of the brackets do not need to lie in $\mathcal{F}$ and $\mathcal{H}$ respectively. However they do possess some main properties of these function classes. The third and last step will be an application of Liebscher's Theorem 2.1 in [46] which gives an exponential inequality for strongly mixing processes.

Step 1: First, note that by assumption (3.5) in **(B3)** on page 39, it holds that

$$h_n^r p_n = o\left(\sqrt{\frac{\log(n)}{nh_n^d}}\right)$$

and therefore

$$\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right)q_n\delta_n = O\left(\sqrt{\frac{\log(n)}{nh_n^d}}q_n\delta_n\right). \tag{A.2}$$

Defining $z_n := \sqrt{\frac{\log(n)}{nh_n^d}}q_n\delta_n$, Lemma 2.2 (iii) thus implies that

$$\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|\hat{h}_n(\boldsymbol{x})\right| = O_P(z_n) = o_P(z_n\sqrt{\log(n)}),$$

$$\max_{\substack{\boldsymbol{k}\in\mathbb{N}_0^d\\1\leq|\boldsymbol{k}|\leq l}}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}\left|D^{\boldsymbol{k}}\hat{h}_n(\boldsymbol{x})\right| = o_P(1)$$

and

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

Therefore, defining for some l-times differentiable function $h : \boldsymbol{J}_n \to \mathbb{R}$ 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 the function classes

$$\mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n) := \left\{h : \boldsymbol{J}_n \to \mathbb{R} : \|h\|_{l+\eta} \le 1\right\},$$

and

$$\mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n) := \left\{h : \boldsymbol{J}_n \to \mathbb{R} : \|h\|_{l+\eta} \le 1, \sup_{\boldsymbol{x}\in\boldsymbol{J}_n} |h(\boldsymbol{x})| \le z_n\sqrt{\log(n)}\right\},$$

it holds that

$$P\left(\hat{h}_n \in \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)\right) \underset{n\to\infty}{\to} 1.$$

Note that the factor $\sqrt{\log(n)}$ in the bound on the functions in $\mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)$ is arbitrary, in the sense that any positive sequence that diverges to infinity is possible. However, it will have an influence on the bandwidth assumptions in **(B3)** which is why a slow log-rate was chosen. For notational simplicity this functions class will be denoted by $\mathcal{H} := \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)$. Then (A.1) is implied by

$$\sup_{s\in[0,1]} \sup_{\varphi\in\mathcal{F}} \sup_{h\in\mathcal{H}} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)\right| = o_P(1). \tag{A.3}$$

Step 2: To bound the suprema over infinitely many objects by the maxima over finitely many ones the notion of covering and bracketing numbers will be used. Let first

$$\varepsilon_{n1} := \frac{1}{\sqrt{n}},\ \varepsilon_{n2} := \frac{1}{\sqrt{n}}, \varepsilon_{n3} := \frac{1}{\sqrt{n}\log n},\ \forall\, n \in \mathbb{N}.$$

Note that for the choice of ε_{n3} it is necessary that $\varepsilon_{n3} = o(n^{-\frac{1}{2}})$ holds. Next, for all $n \in \mathbb{N}$ let $K_n \in \mathbb{N}$, such that $0 = s_1 < \cdots < s_{K_n} = 1$ are the centers of K_n intervals of length $2\varepsilon_{n1}$ that cover the interval $[0,1]$. Furthermore, for all $n \in \mathbb{N}$ 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 $\|\varphi\|_{L_2(P)} := (\int \varphi^2 dP)^{\frac{1}{2}}$ for all $\varphi \in \mathcal{F}$ and $\|h\|_\infty := \sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|h(\boldsymbol{x})|$ for all $h \in \mathcal{H}$. The function class $\mathcal{F}$ will be covered with J_n-many brackets which are denoted by $[\varphi_1^l, \varphi_1^u], \ldots, [\varphi_{J_n}^l, \varphi_{J_n}^u]$. To cover $\mathcal{H}$, M_n-many brackets will be used which are denoted by $[h_1^l, h_1^u], \ldots, [h_{M_n}^l, h_{M_n}^u]$. Bounds on K_n, J_n and M_n can be found in the following way. It is clear that

$$K_n = O\left(\varepsilon_{n1}^{-1}\right) = O(\sqrt{n}).$$

In Lemma A.5 on page 149 in Section A.2 it will be shown that

$$J_n = O\left(\varepsilon_{n2}^{-2d}\right) = O\left(n^d\right).$$

For the bound on M_n Theorem 2.7.1 of Van der Vaart & Wellner ([75], p. 155) can be used. It implies the following bound for the covering number (see Definition 2.1.5 in [75])

$$N\left(\varepsilon_{n3}, \mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n), \|\cdot\|_\infty\right) = O\left(\exp\left(c_n^d \varepsilon_{n3}^{-\frac{d}{l+\eta}}\right)\right).$$

As

$$N\left(\varepsilon_{n3}, \mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n), \|\cdot\|_\infty\right) = N_{[\,]}\left(2\varepsilon_{n3}, \mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n), \|\cdot\|_\infty\right),$$

which can be seen for example in [75], p. 84, the bracketing number of $\mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n)$ possesses the same bound. As $\mathcal{H} := \mathcal{C}_{1,n}^{l+\eta}(\boldsymbol{J}_n)$ is a subset of $\mathcal{C}_1^{l+\eta}(\boldsymbol{J}_n)$ the same bound can be found for M_n, namely

$$M_n = O\left(\exp\left(c_n^d \varepsilon_{n3}^{-\frac{d}{l+\eta}}\right)\right) = O\left(\exp\left(c_n^d (\sqrt{n}\log n)^{\frac{d}{l+\eta}}\right)\right).$$

Now it can be obtained that

$$\begin{aligned}
&\sup_{s\in[0,1]}\sup_{\varphi\in\mathcal{F}}\sup_{h\in\mathcal{F}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)\right| \\
&= \max_{\substack{1\le k\le K_n\\ 1\le j\le J_n\\ 1\le m\le M_n}} \sup_{\substack{s\in[0,1]\\ |s-s_k|\le\varepsilon_{n1}}} \sup_{\varphi\in[\varphi_j^l,\varphi_j^u]} \sup_{h\in[h_m^l,h_m^u]} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)\right| \\
&\le \max_{\substack{1\le k\le K_n\\ 1\le j\le J_n\\ 1\le m\le M_n}} \sup_{\substack{s\in[0,1]\\ |s-s_k|\le\varepsilon_{n1}}} \sup_{\varphi\in[\varphi_j^l,\varphi_j^u]} \sup_{h\in[h_m^l,h_m^u]} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)\cdot\right. \\
&\qquad\qquad \left.\cdot\left(I\left\{\frac{i}{n}\le s\right\} - I\left\{\frac{i}{n}\le s_k\right\}\right)\right| \qquad (A.4) \\
&+ \max_{\substack{1\le k\le K_n\\ 1\le j\le J_n\\ 1\le m\le M_n}} \sup_{\varphi\in[\varphi_j^l,\varphi_j^u]} \sup_{h\in[h_m^l,h_m^u]} \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)\right|. \qquad (A.5)
\end{aligned}$$

First, it will be shown that (A.4) tends to zero as n tends to infinity. Using

$$\left|h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right| \le 2\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|h(\boldsymbol{x})\varphi(\boldsymbol{x})| = 2\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|h(\boldsymbol{x})| \le 2z_n\sqrt{\log(n)}$$

for all $h\in\mathcal{H}$, $\varphi\in\mathcal{F}$ and for all $i = 1,\dots,n$ and $n\in\mathbb{N}$, it can be obtained that (A.4) is bounded by

$$2z_n\sqrt{\log(n)}\max_{1\le k\le K_n}\sup_{\substack{s\in[0,1]\\ |s-s_k|\le\varepsilon_{n1}}}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left|I\left\{\frac{i}{n}\le s\right\} - I\left\{\frac{i}{n}\le s_k\right\}\right|$$

$$
\begin{aligned}
&\overset{(*)}{\leq} 2z_n\sqrt{\log(n)} \max_{1\leq k\leq K_n} \sup_{\substack{s\in[0,1] \\ |s-s_k|\leq\varepsilon_{n1}}} \left(\sqrt{n}|s-s_k| + \frac{1}{\sqrt{n}}\right) \\
&\leq 2z_n\sqrt{\log(n)}\sqrt{n}\varepsilon_{n1} + 2z_n\frac{\sqrt{\log(n)}}{\sqrt{n}} \\
&= 2z_n\sqrt{\log(n)} + 2z_n\frac{\sqrt{\log(n)}}{\sqrt{n}} \\
&\overset{(**)}{=} o(1),
\end{aligned}
$$

where $(*)$ holds because

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n} |I\{i\leq\lfloor ns\rfloor\} - I\{i\leq\lfloor ns_k\rfloor\}| \\
&= \frac{1}{n}\sum_{i=1}^{n} I\{\min(\lfloor ns_k\rfloor, \lfloor ns\rfloor) < i \leq \max(\lfloor ns_k\rfloor, \lfloor ns\rfloor)\} \\
&= \frac{1}{n}(\max(\lfloor ns_k\rfloor, \lfloor ns\rfloor) - \min(\lfloor ns_k\rfloor, \lfloor ns\rfloor)) \\
&\leq \frac{1}{n}(\max(ns_k, ns) - \min(ns_k, ns) + 1) \\
&= \max(s_k, s) - \min(s_k, s) + \frac{1}{n} \\
&= |s-s_k| + \frac{1}{n}.
\end{aligned}
$$

Additionally, $(**)$ is implied by (2.9) in **(B2)** on page 16 and (3.6) in **(B3)** on page 39 as

$$
z_n\sqrt{\log(n)} = \frac{\log(n)}{\sqrt{nh_n^d}}q_n\delta_n = \underbrace{\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}}q_n\delta_n}_{\overset{(2.9)}{=}o(1)} \underbrace{\sqrt{\log(n)h_n^{2(l+1)}}}_{\overset{(3.6)}{=}o(1)} = o(1).
$$

Secondly, the term in (A.5) will be considered. Using the brackets of $\mathcal{F}$ and $\mathcal{H}$, a lower and an upper bound for

$$
\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns_k\rfloor}\left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP\right)
$$

will be obtained leading to a bound for its absolute value. For a given $m \in \{1,\ldots,M_n\}$ and $j \in \{1,\ldots,J_n\}$ let $h \in [h_m^l, h_m^u]$ and $\varphi \in [\varphi_j^l, \varphi_j^u]$ hold. Then $h_m^l \leq h \leq h_m^u$ and $\varphi_j^l \leq \varphi \leq \varphi_j^u$, as well as $\|h_m^u - h_m^l\|_\infty \leq \varepsilon_{n3}$ and $\|\varphi_j^u - \varphi_j^l\|_{L_2(P)} \leq \varepsilon_{n2}$ hold. Furthermore, φ_j^l, φ_j^u can be chosen to be indicator functions. In particular, φ_j^l, φ_j^u are then non-negative. Then it holds that

$$
\begin{aligned}
h\varphi &= hI\{h\geq 0\}\varphi + hI\{h<0\}\varphi \\
&\leq hI\{h\geq 0\}\varphi_j^u + hI\{h<0\}\varphi_j^l
\end{aligned}
$$

$$\leq h_m^u I\{h_m^u \geq 0\}\varphi_j^u + h_m^u I\{h_m^u < 0\}\varphi_j^l$$

and

$$\begin{aligned}
h\varphi &= hI\{h \geq 0\}\varphi + hI\{h < 0\}\varphi \\
&\geq hI\{h \geq 0\}\varphi_j^l + hI\{h < 0\}\varphi_j^u \\
&\geq h_m^l I\{h_m^l \geq 0\}\varphi_j^l + h_m^l I\{h_m^l < 0\}\varphi_j^u.
\end{aligned}$$

For an upper bound it therefore can be obtained that

$$\begin{aligned}
&h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP \\
&\leq h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) \geq 0\}\varphi_j^u(\boldsymbol{X}_i) + h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) < 0\}\varphi_j^l(\boldsymbol{X}_i) \\
&\quad - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l dP - \int h_m^l I\{h_m^l < 0\}\varphi_j^u dP \\
&= h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) \geq 0\}\varphi_j^u(\boldsymbol{X}_i) + h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) < 0\}\varphi_j^l(\boldsymbol{X}_i) \\
&\quad - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l dP - \int h_m^l I\{h_m^l < 0\}\varphi_j^u dP \\
&\quad \pm \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP \pm \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&\quad \pm \int h_m^u I\{h_m^u < 0\}\varphi_j^u dP \pm \int h_m^l I\{h_m^l \geq 0\}\varphi_j^u dP \\
&= h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) \geq 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP \\
&\quad + h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) < 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&\quad + \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP + \int h_m^u I\{h_m^u < 0\}\varphi_j^u dP \\
&\quad - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^u dP - \int h_m^l I\{h_m^l < 0\}\varphi_j^u dP \\
&\quad + \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP - \int h_m^u I\{h_m^u < 0\}\varphi_j^u dP \\
&\quad + \int h_m^l I\{h_m^l \geq 0\}\varphi_j^u dP - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l dP \\
&= h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) \geq 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP \\
&\quad + h_m^u(\boldsymbol{X}_i)I\{h_m^u(\boldsymbol{X}_i) < 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&\quad + \int (h_m^u - h_m^l)\varphi_j^u dP \qquad \text{(A.6)} \\
&\quad + \int h_m^u I\{h_m^u < 0\}(\varphi_j^l - \varphi_j^u) dP \qquad \text{(A.7)} \\
&\quad + \int h_m^l I\{h_m^l \geq 0\}(\varphi_j^u - \varphi_j^l) dP. \qquad \text{(A.8)}
\end{aligned}$$

It holds that (A.6) is bounded by

$$\|h_m^u - h_m^l\|_\infty \|\varphi_j^u\|_\infty = \|h_m^u - h_m^l\|_\infty \leq \varepsilon_{n3} = o\left(\frac{1}{\sqrt{n}}\right).$$

As $\|h_m^u\|_\infty \leq \|h_m^u - h_m^l\|_\infty + \|h\|_\infty \leq \varepsilon_{n3} + z_n\sqrt{\log(n)}$, (A.7) is bounded by

$$\|h_m^u I\{h_m^u < 0\}\|_\infty \|\varphi_j^l - \varphi_j^u\|_{L_2(P)} \leq (\varepsilon_{n3} + z_n\sqrt{\log(n)})\varepsilon_{n2} = o\left(\frac{1}{\sqrt{n}}\right).$$

Similarly, $\|h_m^l\|_\infty \leq \varepsilon_{n3} + z_n$ and therefore (A.8) is bounded by

$$\|h_m^l I\{h_m^l \geq 0\}\|_\infty \|\varphi_j^u - \varphi_j^l\|_{L_2(P)} \leq (\varepsilon_{n3} + z_n\sqrt{\log(n)})\varepsilon_{n2} = o\left(\frac{1}{\sqrt{n}}\right).$$

To calculate a lower bound it can be obtained that

$$\begin{aligned}
&h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP \\
&\geq h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) \geq 0\}\varphi_j^l(\boldsymbol{X}_i) + h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) < 0\}\varphi_j^u(\boldsymbol{X}_i) \\
&\quad - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP - \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&= h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) \geq 0\}\varphi_j^l(\boldsymbol{X}_i) + h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) < 0\}\varphi_j^u(\boldsymbol{X}_i) \\
&\quad - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP - \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&\quad \pm \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l dP \pm \int h_m^l I\{h_m^l < 0\}\varphi_j^u dP \\
&\quad \pm \int h_m^l I\{h_m^l < 0\}\varphi_j^l dP \pm \int h_m^u I\{h_m^u \geq 0\}\varphi_j^l dP \\
&= h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) \geq 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l \\
&\quad + h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) < 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l < 0\}\varphi_j^u \\
&\quad + \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l dP + \int h_m^l I\{h_m^l < 0\}\varphi_j^l dP \\
&\quad - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^l dP - \int h_m^u I\{h_m^u < 0\}\varphi_j^l dP \\
&\quad + \int h_m^l I\{h_m^l < 0\}\varphi_j^u dP - \int h_m^l I\{h_m^l < 0\}\varphi_j^l dP \\
&\quad + \int h_m^u I\{h_m^u \geq 0\}\varphi_j^l dP - \int h_m^u I\{h_m^u \geq 0\}\varphi_j^u dP \\
&= h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) \geq 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l \geq 0\}\varphi_j^l \\
&\quad + h_m^l(\boldsymbol{X}_i)I\{h_m^l(\boldsymbol{X}_i) < 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l < 0\}\varphi_j^u \\
&\quad - \int (h_m^u - h_m^l)\varphi_j^l dP && \text{(A.9)} \\
&\quad - \int h_m^l I\{h_m^l < 0\}(\varphi_j^l - \varphi_j^u) dP && \text{(A.10)} \\
&\quad - \int h_m^u I\{h_m^u \geq 0\}(\varphi_j^u - \varphi_j^l) dP. && \text{(A.11)}
\end{aligned}$$

Similar to before, it holds that (A.9) is bounded from below by $-\varepsilon_{n3}$, while (A.10) and (A.11) are each bounded from below by $-(\varepsilon_{n3} + z_n\sqrt{\log(n)})\varepsilon_{n2}$. Both these bounds converge again to zero at a rate of $o\big(\frac{1}{\sqrt{n}}\big)$. Coming back to bounding (A.5), it therefore holds that

$$\begin{aligned}
&\max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \sup_{\varphi\in[\varphi_j^l,\varphi_j^u]} \sup_{h\in[h_m^l,h_m^u]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h(\boldsymbol{X}_i)\varphi(\boldsymbol{X}_i) - \int h\varphi dP \right) \right| \\
&\leq \max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \max \Bigg\{ \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i)\geq 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u\geq 0\}\varphi_j^u dP \right) \right| \\
&\qquad + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i)< 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u< 0\}\varphi_j^l dP \right) \right|, \\
&\qquad \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^l(\boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)\geq 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l\geq 0\}\varphi_j^l dP \right) \right| \\
&\qquad + \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^l(\boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)< 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l< 0\}\varphi_j^u dP \right) \right| \Bigg\} \\
&+ o(1).
\end{aligned}$$

Step 3: Finally, the assertion of (A.3) follows by the following four assertions

$$\max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i)\geq 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u\geq 0\}\varphi_j^u dP \right) \right| = o_P(1), \tag{A.12}$$

$$\max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i)< 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u< 0\}\varphi_j^l dP \right) \right| = o_P(1), \tag{A.13}$$

$$\max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^l(\boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)\geq 0\}\varphi_j^l(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l\geq 0\}\varphi_j^l dP \right) \right| = o_P(1), \tag{A.14}$$

$$\max_{\substack{1\leq k\leq K_n\\ 1\leq j\leq J_n\\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k\rfloor} \left(h_m^l(\boldsymbol{X}_i) I\{h_m^l(\boldsymbol{X}_i)< 0\}\varphi_j^u(\boldsymbol{X}_i) - \int h_m^l I\{h_m^l< 0\}\varphi_j^u dP \right) \right| = o_P(1). \tag{A.15}$$

Within the proof it will become clear that it is sufficient to only show (A.12) as the other ones work analogous. As mentioned before, Liebscher's Theorem 2.1 in [46] will be used. Following the notation in [46]

$$Z_i := \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right) I\left\{ \frac{i}{n} \leq s_k \right\},$$

is defined for all $1 \leq i \leq n$ and $n \in \mathbb{N}$. Note that Liebscher's result is an inequality for fixed $n \in \mathbb{N}$ and the dependency of Z_i on n is not reflected in the notation. Then

$$\begin{aligned} &P\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k \rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right) \right| > \epsilon \right) \\ &= P\left(\left| \sum_{i=1}^{n} Z_i \right| > \sqrt{n}\epsilon \right) \end{aligned}$$

holds. To bound the last probability, the conditions on Z_i of Theorem 2.1 in [46] need verification. First, note that Z_i is centered and

$$\begin{aligned} |Z_i| &\leq \left| h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right| \\ &\leq 2 \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| h_m^u(\boldsymbol{x}) I\{h_m^u(\boldsymbol{x}) \geq 0\} \varphi_j^u(\boldsymbol{x}) \right| \\ &= 2 \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} |h_m^u(\boldsymbol{x})| \\ &\leq 2(\varepsilon_{n3} + z_n \sqrt{\log(n)}) =: S(n). \end{aligned}$$

Furthermore, using

$$\begin{aligned} &E\left[Z_i^2\right] \\ &= E\left[\left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right)^2 I\left\{ \frac{i}{n} \leq s_k \right\} \right] \\ &\leq 4 \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} \left| h_m^u(\boldsymbol{x}) I\{h_m^u(\boldsymbol{x}) \geq 0\} \varphi_j^u(\boldsymbol{x}) \right|^2 \\ &= 4 \sup_{\boldsymbol{x} \in \boldsymbol{J}_n} |h_m^u(\boldsymbol{x})|^2 \\ &\leq 4(\varepsilon_{n3} + z_n \sqrt{\log(n)})^2 \end{aligned}$$

results in

$$\begin{aligned} E\left[\left(\sum_{i=T+1}^{(T+N)\wedge n} Z_i \right)^2 \right] &= \sum_{i_1=T+1}^{(T+N)\wedge n} \sum_{i_2=T+1}^{(T+N)\wedge n} E[Z_{i_1} Z_{i_2}] \\ &\leq N^2 E[Z_{i_1} Z_{i_2}] \\ &\leq N^2 E\left[Z_i^2\right] \\ &\leq 4N^2(\varepsilon_{n3} + z_n \sqrt{\log(n)})^2 =: D(n, N) \end{aligned}$$

for all $T = 0, \dots, n-1$ and $1 \leq N \leq n$, $n \in \mathbb{N}$. As measurable functions maintain mixing properties (see for instance [21] Subsection 2.6.1 (ii), p. 69), it holds that

$$\left(h_m^u(\boldsymbol{X}_t) I\{h_m^u(\boldsymbol{X}_t) \geq 0\} \varphi_j^u(\boldsymbol{X}_t) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right)_{t \in \mathbb{Z}}$$

is strongly mixing with the same mixing coefficients $\alpha(\cdot)$ as $(Y_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$. Let $\{\tilde{\alpha}_n(t) : t \in \mathbb{N}\}$ be the sequence of coefficients of $\{Z_t : 1 \leq t \leq n, n \in \mathbb{N}\}$ defined as in (1.4) in Definition 1.1 on page 7. For fixed $n \in \mathbb{N}$ they can be bounded by the mixing coefficients $\{\alpha(t) : t \in \mathbb{N}\}$ of $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{N}\}$ (see for instance [4], Section 2, remark (iv)). An application of Theorem 2.1 in [46] leads to

$$P\left(\left| \sum_{i=1}^n Z_i \right| > \sqrt{n}\epsilon \right) \leq 4 \exp\left(- \frac{n\epsilon^2}{64 \frac{n}{N} D(n,N) + \frac{8}{3}\sqrt{n}\epsilon N S(n)} \right) + 4\frac{n}{N}\alpha(N),$$

for all $\epsilon > 0$ with $\sqrt{n}\epsilon > 4NS(n)$ and for all $1 \leq N \leq n$, $n \in \mathbb{N}$. Note here that for (A.13), (A.14) and (A.15) the same terms for $S(n)$ and $D(n,N)$ will be obtained and that the proofs for these assertions therefore work analogously. Next it is clear that $\varepsilon_{n3} := \frac{1}{\sqrt{n}\log n} = o(z_n\sqrt{\log(n)})$ and therefore

$$S(n) = 2(\varepsilon_{n3} + z_n\sqrt{\log(n)}) = O(z_n\sqrt{\log(n)}) = O\left(\frac{\log(n)}{\sqrt{nh_n^d}} q_n \delta_n \right)$$

and

$$D(n,N) = 4N^2(\varepsilon_{n3} + z_n\sqrt{\log(n)})^2 = N^2 O(z_n^2 \log(n)) = N^2 O\left(\frac{\log(n)^2}{nh_n^d} q_n^2 \delta_n^2 \right)$$

hold. Now let $N := \left\lfloor \sqrt{nh_n^d} \right\rfloor \overset{n\to\infty}{\to} \infty$, then it holds that $1 \leq N \leq n$ and for all $\epsilon > 0$

$$\begin{aligned} 4NS(n) &= 8 \left\lfloor \sqrt{nh_n^d} \right\rfloor (\varepsilon_{n3} + z_n\sqrt{\log(n)}) \\ &= 8 \left\lfloor \sqrt{nh_n^d} \right\rfloor O\left(\frac{\log(n)}{\sqrt{nh_n^d}} q_n \delta_n \right) \\ &= 8 \left\lfloor \sqrt{nh_n^d} \right\rfloor \sqrt{h_n^{-d}} \underbrace{O\left(\frac{\log(n)}{\sqrt{n}} q_n \delta_n \right)}_{\overset{(3.3)}{=} o(1)} < \sqrt{n}\varepsilon, \end{aligned}$$

for n large enough. Therefore, it holds that for all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough

$$\begin{aligned} &P\left(\max_{\substack{1\leq k\leq K_n \\ 1\leq j\leq J_n \\ 1\leq m\leq M_n}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k \rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right) \right| > \epsilon \right) \\ &\leq \sum_{\substack{1\leq k\leq K_n \\ 1\leq j\leq J_n \\ 1\leq m\leq M_n}} P\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns_k \rfloor} \left(h_m^u(\boldsymbol{X}_i) I\{h_m^u(\boldsymbol{X}_i) \geq 0\} \varphi_j^u(\boldsymbol{X}_i) - \int h_m^u I\{h_m^u \geq 0\} \varphi_j^u dP \right) \right| > \epsilon \right) \end{aligned}$$

$$
\begin{aligned}
&\leq \sum_{\substack{1\leq k\leq K_n\\1\leq j\leq J_n\\1\leq m\leq M_n}} \left(4\exp\left(-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}\right)+4\frac{n}{N}\alpha(N)\right)\\
&= 4K_nJ_nM_n\left(\exp\left(-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}\right)+\frac{n}{N}\alpha(N)\right)\\
&= 4\exp\left(\log(K_n)+\log(J_n)+\log(M_n)-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}\right) \qquad \text{(A.16)}\\
&\quad +4\exp\left(\log(K_n)+\log(J_n)+\log(M_n)+\log\left(\frac{n}{N}\right)+\log\left(\alpha(N)\right)\right). \qquad \text{(A.17)}
\end{aligned}
$$

It is left to show, that the exponents in (A.16) and (A.17) diverge to $-\infty$ as $n\to\infty$. Starting with (A.16), it is to show that

$$
\frac{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}N\epsilon S(n)}{n\epsilon^2} = o\left(\frac{1}{\log\left(K_n\right)+\log(J_n)+\log(M_n)}\right).
$$

Replacing first $K_n, J_n, M_n, N, D(n,N)$ and $S(n)$ by their rates, and finally using the rate of c_n in **(J)** on page 16, it can be obtained that

$$
\begin{aligned}
&\frac{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}N\epsilon S(n)}{n\epsilon^2}\left(\log\left(K_n\right)+\log(J_n)+\log(M_n)\right)\\
&= \left(\frac{64}{\epsilon^2}\frac{1}{N}D(n,N)+\frac{1}{\sqrt{n}\epsilon}\frac{8}{3}NS(n)\right)\left(\log\left(K_n\right)+\log(J_n)+\log(M_n)\right)\\
&= O\left(\left(\sqrt{nh_n^d}\frac{\log(n)^2}{nh_n^d}q_n^2\delta_n^2+\frac{1}{\sqrt{n}}\sqrt{nh_n^d}\frac{\log(n)}{\sqrt{nh_n^d}}q_n\delta_n\right)\left(\log(n)+c_n^d\left(\sqrt{n}\log(n)\right)^{\frac{d}{l+\eta}}\right)\right)\\
&= O\left(\left(\frac{\log\left(n\right)^2}{\sqrt{nh_n^d}}q_n^2\delta_n^2+\frac{\log\left(n\right)}{\sqrt{n}}q_n\delta_n\right)\left(\log(n)+c_n^d\left(\sqrt{n}\log(n)\right)^{\frac{d}{l+\eta}}\right)\right)\\
&= O\left(\frac{\log(n)^3}{\sqrt{nh_n^d}}q_n^2\delta_n^2+c_n^d\frac{\log(n)^{2+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}}q_n^2\delta_n^2+\frac{\log(n)^2}{\sqrt{n}}q_n\delta_n+c_n^d\frac{\log(n)^{1+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}}}q_n\delta_n\right)\\
&= O\left(\frac{\log(n)^3}{\sqrt{nh_n^d}}q_n^2\delta_n^2+\frac{\log(n)^{3+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}}q_n^2\delta_n^2+\frac{\log(n)^2}{\sqrt{n}}q_n\delta_n+\frac{\log(n)^{2+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}}}q_n\delta_n\right)\\
&= o(1),
\end{aligned}
$$

by assumption (3.3) in **(B3)** on page 39. Concerning (A.17) it is to show that

$$
\frac{1}{|\log\left(\alpha(N)\right)|} = o\left(\frac{1}{\log\left(K_n\right)+\log(J_n)+\log(M_n)+\log\left(\frac{n}{N}\right)}\right).
$$

Using again all bounds and the exponential rates of convergence of the mixing coefficient in **(G)**, namely $|\log\left(\alpha(t)\right)|^{-1} = O(t^{-1})$, $(t\to\infty)$, it can be obtained, that

$$
\frac{1}{|\log\left(\alpha(N)\right)|}\left(\log\left(K_n\right)+\log(J_n)+\log(M_n)+\log\left(\frac{n}{N}\right)\right)
$$

$$
\begin{aligned}
&= O\left(\frac{1}{\sqrt{nh_n^d}}\left(\log(n) + c_n^d\left(\sqrt{n}\log(n)\right)^{\frac{d}{l+\eta}} + \log\left(\frac{n}{\sqrt{nh_n^d}}\right)\right)\right)\\
&= O\left(\frac{\log(n)}{\sqrt{nh_n^d}} + c_n^d\frac{\log(n)^{\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}} + \frac{\log\left(\sqrt{\frac{n}{h_n^d}}\right)}{\sqrt{nh_n^d}}\right)\\
&= O\left(\frac{\log(n)}{\sqrt{nh_n^d}} + \frac{\log(n)^{1+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}} + \frac{\log\left(\frac{n}{h_n^d}\right)}{\sqrt{nh_n^d}}\right)\\
&= O\left(\frac{\log(n)}{\sqrt{nh_n^d}} + \frac{\log(n)^{1+\frac{d}{l+\eta}}}{\sqrt{n^{1-\frac{d}{l+\eta}}h_n^d}} + \frac{|\log(h_n)|}{\sqrt{nh_n^d}}\right)\\
&= o(1),
\end{aligned}
$$

by assumption (3.3) and (3.4) in **(B3)**. This finally proves (A.3) and therefore the assertion of Lemma A.1.

□

Lemma A.2. *Under the assumptions of Theorem 3.1*

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

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

Remark. The proof of Lemma A.2 uses a similar technique as before. An appropriate function class will be defined and partitioned using $L_1(P)$-brackets (where $\boldsymbol{X}_t \sim P$) according to Definition 1.5. The bound for the corresponding bracketing number will be proven in Lemma A.6. Finally, the exponential inequality for strongly mixing processes in [46] will be applied again.

Proof. The proof of Lemma A.2 consists of three main steps. First, it will be shown that $\hat{f}_n$ can be replaced by f. Then defining the function class

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

it will be shown in the second step that

$$\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi(\boldsymbol{X}_i) - \int\varphi dP\right)\right| = o_P(1), \tag{A.18}$$

where $\boldsymbol{X}_t \sim P$. In the third step it will be obtained that

$$\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\int\varphi dP\right| = o\left(\frac{1}{\sqrt{n}}\right). \tag{A.19}$$

Step 1: Let $\tilde{Y}_t = m(\boldsymbol{X}_t)$ and $\tilde{m}_n(\cdot)$ be the Nadaraya-Watson estimator for $m(\cdot) = E[\tilde{Y}_t|\boldsymbol{X}_t = \cdot]$. Note that $(\tilde{Y}_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a strictly stationary process and the uniform rates of convergence for kernel estimators obtained in Chapter 2 (see Lemma 2.2 (i)(a) and (iii)(a) on page 17) can be applied. Thus, it can be obtained that

$$
\begin{aligned}
&\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\int_{\mathbb{R}^d}\left(m(\boldsymbol{y})-m(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}\left(\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}-1\right)d\boldsymbol{y}\right|\\
&=\sqrt{n}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^n\left(m(\boldsymbol{y})-m(\boldsymbol{X}_i)\right)K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}\frac{f(\boldsymbol{y})-\hat{f}_n(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}d\boldsymbol{y}\right|\\
&\le\sqrt{n}\left(\int_{\mathbb{R}^d}\left(m(\boldsymbol{y})-\frac{1}{n}\sum_{i=1}^n m(\boldsymbol{X}_i)K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\frac{1}{\hat{f}_n(\boldsymbol{y})}\right)^2\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)^{\frac{1}{2}}\\
&\qquad\qquad\qquad\qquad\cdot\left(\int_{\mathbb{R}^d}\left(f(\boldsymbol{y})-\hat{f}_n(\boldsymbol{y})\right)^2\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)^{\frac{1}{2}}\\
&\le\sqrt{n}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|m(\boldsymbol{x})-\tilde{m}_n(\boldsymbol{x})|\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|f(\boldsymbol{x})-\hat{f}_n(\boldsymbol{x})|\\
&=\sqrt{n}O_P\left(\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right)q_n\delta_n\right)O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right)\\
&=O_P\left(\frac{\log(n)}{\sqrt{n}h_n^d}q_n\delta_n\right),
\end{aligned}
$$

where the last equality holds due to (A.2) on page 114 in the proof of Lemma A.1. Finally, by using the bandwidth assumptions (3.3) (which implies that $l+1>d$) and (3.6) in **(B3)** on page 39 as well as (2.9) in **(B2)** on page 16, it holds that

$$
\frac{\log(n)}{\sqrt{n}h_n^d}q_n\delta_n=\underbrace{\sqrt{\frac{\log(n)}{nh_n^{d+2(l+1)}}}q_n\delta_n}_{\overset{(2.9)}{=}o(1)}\underbrace{\sqrt{h_n^{l+1-d}}}_{\overset{(3.3)}{=}o(1)}\underbrace{\sqrt{\log(n)h_n^{l+1}}}_{\overset{(3.6)}{=}o(1)}=o(1). \tag{A.20}
$$

Step 2: The function class $\mathcal{F}_{n,1}$ will be covered with finitely many brackets. Then the supremum can by bounded by a maximum, which then can be bounded using the exponential inequality in [46]. The partition of $\mathcal{F}_{n,1}$, that is given in Lemma A.6, will be used. Let therefore $\boldsymbol{z}_0,\ldots,\boldsymbol{z}_{J_n}$ be the partition of $\mathbb{R}^d$ and $[\varphi_1^l,\varphi_1^u],\ldots,[\varphi_{J_n}^l,\varphi_{J_n}^u]$ be the corresponding brackets from Lemma A.6, where

$$
J_n := N_{[\,]}\left(\varepsilon_n,\mathcal{F}_{n,1},\|\cdot\|_{L_1(P)}\right),
$$

namely for $j\in\{1,\ldots,J_n\}$

$$
\begin{aligned}
\varphi_j^u(\boldsymbol{x}) &:= \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]}(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\,I\{(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&+\int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]}(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\,I\{(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})<0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}
\end{aligned}
$$

and

$$\varphi_j^l(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} (m(\boldsymbol{y}) - m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \geq 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}$$
$$+ \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} (m(\boldsymbol{y}) - m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) < 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}.$$

Then for all $\varphi \in \mathcal{F}_{n,1}$ there exists a $j \in \{1, \ldots, J_n\}$ such that $\varphi_j^l \leq \varphi \leq \varphi_j^u$ and $\|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \leq \varepsilon_n$. In Lemma A.6 it will been shown that

$$J_n = O\left(\varepsilon_n^{-d}\right).$$

Let $\varepsilon_n := \frac{1}{\sqrt{n}\log(n)}$. Note that the choice of ε_n needs to imply $\varepsilon_n = o(n^{-\frac{1}{2}})$. Then for $\varphi \in [\varphi_j^l, \varphi_j^u]$ for some $j \in \{1, \ldots, J_n\}$, it holds that

$$\begin{aligned}\varphi(\boldsymbol{X}_i) - \int \varphi dP &\leq \varphi_j^u(\boldsymbol{X}_i) - \int \varphi_j^l dP \pm \int \varphi_j^u dP\\ &= \varphi_j^u(\boldsymbol{X}_i) - \int \varphi_j^u dP + \int (\varphi_j^u - \varphi_j^l) dP\end{aligned}$$

and

$$\int (\varphi_j^u - \varphi_j^l) dP = \|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \leq \varepsilon_n = o\left(\frac{1}{\sqrt{n}}\right).$$

Similarly, it can be obtained that

$$\begin{aligned}\varphi(\boldsymbol{X}_i) - \int \varphi dP &\geq \varphi_j^l(\boldsymbol{X}_i) - \int \varphi_j^u dP \pm \int \varphi_j^l dP\\ &= \varphi_j^l(\boldsymbol{X}_i) - \int \varphi_j^l dP - \int (\varphi_j^u - \varphi_j^l) dP\end{aligned}$$

and

$$-\int (\varphi_j^u - \varphi_j^l) dP = -\|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \geq -\varepsilon_n = o\left(\frac{1}{\sqrt{n}}\right).$$

Therefore, it holds that

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

The proof of (A.18) therefore reduces to the proofs of

$$\max_{1 \leq j \leq J_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_j^u(\boldsymbol{X}_i) - \int \varphi_j^u dP\right)\right| = o_P(1) \tag{A.21}$$

and

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_j^l(\boldsymbol{X}_i)-\int\varphi_j^l dP\right)\right| = o_P(1). \tag{A.22}$$

By defining

$$\varphi_{j,1}^u(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} (m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\, I\{(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}$$

and

$$\varphi_{j,2}^u(\boldsymbol{x}) := \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z}_{j-1}]} (m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\, I\{(m(\boldsymbol{y})-m(\boldsymbol{x}))K_{h_n}(\boldsymbol{y}-\boldsymbol{x})< 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}$$

it holds that

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

Similarly, functions $\varphi_{j,1}^l$ and $\varphi_{j,2}^l$ can be defined such that $\varphi_j^l \equiv \varphi_{j,1}^l + \varphi_{j,2}^l$. Therefore, the validity of (A.21) and (A.22) is implied by

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{j,1}^u(\boldsymbol{X}_i)-\int\varphi_{j,1}^u dP\right)\right| = o_P(1), \tag{A.23}$$

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{j,2}^u(\boldsymbol{X}_i)-\int\varphi_{j,2}^u dP\right)\right| = o_P(1),$$

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

$$\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{j,2}^l(\boldsymbol{X}_i)-\int\varphi_{j,2}^l dP\right)\right| = o_P(1).$$

It is only necessary to show (A.23) as will become clear within the proof. Following the notation in [46]

$$Z_i := \varphi_{j,1}^u(\boldsymbol{X}_i) - \int \varphi_{j,1}^u dP$$

is defined for all $1\le i\le n$ and $n\in\mathbb{N}$. Note that again the dependency on $n\in\mathbb{N}$ is not reflected in the notation. Let $\{\tilde{\alpha}_n(t): t\in\mathbb{N}\}$ be the sequence of coefficients of $\{Z_t : 1\le t\le n, n\in\mathbb{N}\}$ defined as in (1.4) in Definition 1.1 on page 7. For fixed $n\in\mathbb{N}$ they can be bounded by the mixing coefficients $\{\alpha(t): t\in\mathbb{N}\}$ of $\{(Y_t,\boldsymbol{X}_t): t\in\mathbb{N}\}$ (see for instance [4], Section 2, remark (iv)). Also Z_i is centered and

$$|Z_i| \le 2\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\varphi_{j,1}^u(\boldsymbol{x})| =: S(n).$$

Furthermore, for all $T=0,\dots,n-1$ and $1\le N\le n$, $n\in\mathbb{N}$

$$E\left[\left(\sum_{i=T+1}^{(T+N)\wedge n} Z_i\right)^2\right] \le N^2 E\left[Z_i^2\right] \le 4N^2\sup_{\boldsymbol{x}\in\mathbb{R}^d}|\varphi_{j,1}^u(\boldsymbol{x})|^2 =: D(n,N).$$

To apply Theorem 2.1 in [46] bounds for $S(n)$ and $D(n,N)$ are required. Using integration by substitution and Taylor's expansion of m in $\boldsymbol{x}$ up to order 1, it can be obtained that for all $j \in \{1,\dots,J_n\}$

$$
\begin{aligned}
&\sup_{\boldsymbol{x}\in\mathbb{R}^d} |\varphi^u_{j,1}(\boldsymbol{x})| \\
&= \sup_{\boldsymbol{x}\in\mathbb{R}^d} \left| \int_{(-\boldsymbol{\infty},\boldsymbol{z}_j]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \geq 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \right| \\
&\leq \sup_{\boldsymbol{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} |m(\boldsymbol{y}) - m(\boldsymbol{x})|\, |K_{h_n}(\boldsymbol{y}-\boldsymbol{x})|\, \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\
&= \sup_{\boldsymbol{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} |m(\boldsymbol{x}+\boldsymbol{t}h_n) - m(\boldsymbol{x})|\, |K(\boldsymbol{t})|\, \omega_n(\boldsymbol{x}+\boldsymbol{t}h_n) d\boldsymbol{t} \\
&\leq \sup_{\boldsymbol{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} \left| \sum_{|\boldsymbol{i}|=1} \frac{D^{\boldsymbol{i}} m(\boldsymbol{x}) (\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right| |K(\boldsymbol{t})|\, \omega_n(\boldsymbol{x}+\boldsymbol{t}h_n) d\boldsymbol{t} && \text{(A.24)} \\
&+ \sup_{\boldsymbol{x}\in\mathbb{R}^d} \int_{\mathbb{R}^d} \left| \sum_{|\boldsymbol{i}|=2} \frac{D^{\boldsymbol{i}} m(\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{t}}) (\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right| |K(\boldsymbol{t})|\, \omega_n(\boldsymbol{x}+\boldsymbol{t}h_n) d\boldsymbol{t}, && \text{(A.25)}
\end{aligned}
$$

for some $\boldsymbol{\xi}_{\boldsymbol{x},\boldsymbol{t}}$ on the line segment between $\boldsymbol{x}+\boldsymbol{t}h_n$ and $\boldsymbol{x}$. Using

$$\omega_n(\boldsymbol{x}+\boldsymbol{t}h_n) I\{\boldsymbol{t} \in [-C,C]^d\} = I\{\boldsymbol{x} \in \boldsymbol{I}_n\} I\{\boldsymbol{t} \in [-C,C]^d\},$$

the term in (A.24) can be bounded by

$$h_n \sum_{|\boldsymbol{i}|=1} \frac{1}{\boldsymbol{i}!} \underbrace{\sup_{\boldsymbol{x}\in\boldsymbol{I}_n} |D^{\boldsymbol{i}} m(\boldsymbol{x})|}_{=O(q_n)} \underbrace{\int_{\mathbb{R}^d} |\boldsymbol{t}|^{\boldsymbol{i}} |K(\boldsymbol{t})| d\boldsymbol{t}}_{<\infty} = O(h_n q_n).$$

By similar calculation the term in (A.25) can be bounded by

$$h_n^2 \sum_{|\boldsymbol{i}|=2} \frac{1}{\boldsymbol{i}!} \underbrace{\sup_{\boldsymbol{x}\in[-c_n-2h_nC,\, c_n+2h_nC]} |D^{\boldsymbol{i}} m(\boldsymbol{x})|}_{\overset{\textbf{(F2)}}{=} O(q_n)} \underbrace{\int_{\mathbb{R}^d} |\boldsymbol{t}|^{\boldsymbol{i}} |K(\boldsymbol{t})| d\boldsymbol{t}}_{<\infty} = O(h_n^2 q_n) = O(h_n q_n).$$

Using these results it holds that

$$S(n) = O(h_n q_n)$$

and

$$D(n,N) = N^2 O(h_n^2 q_n^2).$$

Note here that the corresponding bounds for $S(n)$ and $D(n,N)$ for the other three cases will be the same which is the reason to only look at the first case (A.23). By choosing for all $n \in \mathbb{N}$,

$$N := \lfloor \log(n)^2 \rfloor + 1,$$

it holds that $1 \leq N \leq n$ and

$$4NS(n) = 4(\lfloor \log(n)^2 \rfloor + 1) S(n) = \underbrace{O\left(\log(n)^2 h_n q_n\right)}_{\overset{(3.6)}{=} o(1)} < \sqrt{n}\epsilon,$$

for all $\epsilon > 0$ and for n large enough. Hence, applying Liebscher's Theorem 2.1 in [46] it holds for all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough that

$$\begin{aligned}
&P\left(\max_{1\leq j\leq J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi_{j,1}^{u}(\boldsymbol{X}_i)-\int\varphi_{j,1}^{u}dP\right)\right|>\epsilon\right)\\
&\leq\sum_{j=1}^{J_n}P\left(\left|\sum_{i=1}^{n}\left(\varphi_{j,1}^{u}(\boldsymbol{X}_i)-\int\varphi_{j,1}^{u}dP\right)\right|>\sqrt{n}\epsilon\right)\\
&\leq J_n\left(4\exp\left(-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}\right)+4\frac{n}{N}\alpha(N)\right)\\
&=4\exp\left(\log(J_n)-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}\right) \qquad \text{(A.26)}\\
&\quad+4\exp\left(\log(J_n)+\log\left(\frac{n}{N}\right)+\log\left(\alpha(N)\right)\right). \qquad \text{(A.27)}
\end{aligned}$$

It is left to show that the exponents in (A.26) and (A.27) diverge to $-\infty$ as $n\to\infty$. Starting with (A.26), it is to show that

$$\frac{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}{n\epsilon^2}=o\left(\frac{1}{\log(J_n)}\right).$$

Using

$$J_n=O\left(\varepsilon_n^{-d}\right)=O\left((\sqrt{n}\log(n))^d\right)=O\left(n^{\frac{d}{2}}\log(n)^d\right)$$

and therefore

$$\log(J_n)=O(\log(n)),$$

and replacing N, $D(n,N)$ and $S(n)$ by their bounds, it holds that

$$\begin{aligned}
\frac{64\frac{n}{N}D(n,N)+\frac{8}{3}\sqrt{n}\epsilon NS(n)}{n\epsilon^2}\log(J_n)&=\left(\frac{64}{\epsilon^2}\frac{1}{N}D(n,N)+\frac{1}{\sqrt{n}\epsilon}\frac{8}{3}NS(n)\right)\log(J_n)\\
&=O\left(\left(\log(n)^2h_n^2q_n^2+\frac{1}{\sqrt{n}}\log(n)^2h_nq_n\right)\log(n)\right)\\
&=O\left(\log(n)^3h_n^2q_n^2+\frac{1}{\sqrt{n}}\log(n)^3h_nq_n\right)\\
&=o(1),
\end{aligned}$$

by assumption (3.6) in **(B3)**. Concerning (A.27), it is to show that

$$\frac{1}{|\log\left(\alpha(N)\right)|}=o\left(\frac{1}{\log(J_n)+\log\left(\frac{n}{N}\right)}\right).$$

Again inserting the bounds and the exponential rates of convergence of the mixing coefficients in **(G)** yields

$$\begin{aligned}
\frac{1}{|\log\left(\alpha(N)\right)|}\left(\log(J_n)+\log\left(\frac{n}{N}\right)\right)&=O\left(\frac{1}{N}\log(n)+\frac{1}{N}(\log(n)+\log(N))\right)\\
&=O\left(\frac{1}{\log(n)}\right)=o(1).
\end{aligned}$$

Step 3: It holds that

$$
\begin{aligned}
&\sup_{\varphi\in\mathcal{F}_{n,1}}\left|\int \varphi dP\right| \\
&= \sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\int_{\mathbb{R}^d}(m(\boldsymbol{y})-m(\boldsymbol{x}))\frac{1}{h_n^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}d\boldsymbol{y}f(\boldsymbol{x})d\boldsymbol{x}\right| \\
&\le \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y})-m(\boldsymbol{x}))\frac{1}{h_n^d}K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\omega_n(\boldsymbol{y})f(\boldsymbol{x})d\boldsymbol{x}\right|d\boldsymbol{y} \\
&= \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y})-m(\boldsymbol{y}-\boldsymbol{t}h_n))K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y}-\boldsymbol{t}h_n)d\boldsymbol{t}\right|d\boldsymbol{y} \\
&\le \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y}-\boldsymbol{t}h_n)-m(\boldsymbol{y}))K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{t}\right|d\boldsymbol{y} \qquad \text{(A.28)} \\
&+ \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y}-\boldsymbol{t}h_n)-m(\boldsymbol{y}))K(\boldsymbol{t})\omega_n(\boldsymbol{y})\left(f(\boldsymbol{y}-\boldsymbol{t}h_n)-f(\boldsymbol{y})\right)d\boldsymbol{t}\right|d\boldsymbol{y} \qquad \text{(A.29)}
\end{aligned}
$$

Concerning (A.28), Taylor's expansion of m in $\boldsymbol{y}$ up to order $r-1$ with Lagrange remainder term, namely

$$
m(\boldsymbol{y}-\boldsymbol{t}h_n)-m(\boldsymbol{y}) = \sum_{|\boldsymbol{i}|=1}^{r-1}\frac{(-1)^{|\boldsymbol{i}|}D^{\boldsymbol{i}}m(\boldsymbol{y})(\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} + \sum_{|\boldsymbol{i}|=r}\frac{(-1)^{|\boldsymbol{i}|}D^{\boldsymbol{i}}m(\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{t}})(\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!},
$$

for some $\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{t}}$ on the line segment between $\boldsymbol{y}$ and $\boldsymbol{y}-\boldsymbol{t}h_n$ is used. Furthermore, the conditions on the kernel function K in **(K)** and on the partial derivatives of m in **(F1)** are used to obtain

$$
\begin{aligned}
&\int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}(m(\boldsymbol{y}-\boldsymbol{t}h_n)-m(\boldsymbol{y}))K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{t}\right|d\boldsymbol{y} \\
&= \int_{\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\sum_{|\boldsymbol{i}|=1}^{r-1}\frac{(-1)^{|\boldsymbol{i}|}D^{\boldsymbol{i}}m(\boldsymbol{y})(\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{t}\right. \\
&\qquad\qquad \left. + \int_{\mathbb{R}^d}\sum_{|\boldsymbol{i}|=r}\frac{(-1)^{|\boldsymbol{i}|}D^{\boldsymbol{i}}m(\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{t}})(\boldsymbol{t}h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!}K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{t}\right|d\boldsymbol{y} \\
&= \int_{\mathbb{R}^d}\left|\sum_{|\boldsymbol{i}|=1}^{r-1}h_n^{|\boldsymbol{i}|}\frac{(-1)^{|\boldsymbol{i}|}D^{\boldsymbol{i}}m(\boldsymbol{y})}{\boldsymbol{i}!}\omega_n(\boldsymbol{y})f(\boldsymbol{y})\underbrace{\int_{\mathbb{R}^d}\boldsymbol{t}^{\boldsymbol{i}}K(\boldsymbol{t})d\boldsymbol{t}}_{=0\ \forall|\boldsymbol{i}|=1,\dots,r-1}\right. \\
&\qquad\qquad \left. + h_n^r\sum_{|\boldsymbol{i}|=r}\frac{(-1)^{|\boldsymbol{i}|}}{\boldsymbol{i}!}\int_{\mathbb{R}^d}D^{\boldsymbol{i}}m(\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{t}})\boldsymbol{t}^{\boldsymbol{i}}K(\boldsymbol{t})\omega_n(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{t}\right|d\boldsymbol{y} \\
&\le h_n^r\int_{\mathbb{R}^d}\sum_{|\boldsymbol{i}|=r}\frac{1}{\boldsymbol{i}!}\int_{\mathbb{R}^d}\underbrace{\left|D^{\boldsymbol{i}}m(\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{t}})\right|}_{\le\sup_{\boldsymbol{x}\in\boldsymbol{I}_n}\left|D^{\boldsymbol{i}}m(\boldsymbol{x})\right|}\left|\boldsymbol{t}^{\boldsymbol{i}}K(\boldsymbol{t})\right|\omega_n(\boldsymbol{y})d\boldsymbol{t}f(\boldsymbol{y})d\boldsymbol{y}
\end{aligned}
$$

$$
\begin{aligned}
&\leq h_n^r \sum_{|\boldsymbol{i}|=r} \underbrace{\sup_{\boldsymbol{x}\in \boldsymbol{I}_n} \left|D^{\boldsymbol{i}} m(\boldsymbol{x})\right|}_{=O(q_n)\ \forall |\boldsymbol{i}|=r} \underbrace{\int_{\mathbb{R}^d} |\boldsymbol{t}^{\boldsymbol{i}} K(\boldsymbol{t})| d\boldsymbol{t}}_{<\infty} \int_{\mathbb{R}^d} f(\boldsymbol{y}) d\boldsymbol{y} \\
&= O\left(h_n^r q_n\right) \\
&= o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where the last equality holds by assumption (3.5) in **(B3)**. Concerning (A.29), Taylor's expansion for both m and f is used. By analogues calculations as above it can be obtained that

$$
\begin{aligned}
\int_{\mathbb{R}^d} \left| \int_{\mathbb{R}^d} (m(\boldsymbol{y} - \boldsymbol{t} h_n) - m(\boldsymbol{y})) K(\boldsymbol{t}) \omega_n(\boldsymbol{y}) \left(f(\boldsymbol{y} - \boldsymbol{t} h_n) - f(\boldsymbol{y})\right) d\boldsymbol{t} \right| d\boldsymbol{y} &= O(h_n^r p_n q_n) \\
&= o\left(\frac{1}{\sqrt{n}}\right),
\end{aligned}
$$

where the last equality again holds by assumption (3.5) in **(B3)**. This completes (A.19) in Step 3 and therefore finally the proof of Lemma A.2.

□

Lemma A.3. *Under the same assumptions of Theorem 3.1*

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

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

Remark. The idea for the proof of Lemma A.3 is the same as before, where the bound for the corresponding bracketing number will be proven in Lemma A.7.

Proof. The proof of Lemma A.3 consists of three steps. First, it will be shown that $\hat{f}_n$ can be replaced by f. Then considering

$$
\begin{aligned}
&\frac{1}{\sqrt{n}} \sum_{i=1}^n U_i \left(\int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \right) \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \Bigg(U_i I\left\{|U_i| > n^{\frac{1}{q}}\right\} \left(\int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \right) \\
&\qquad - E\left[U_i I\left\{|U_i| > n^{\frac{1}{q}}\right\} \left(\int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \right) \right] \Bigg) \qquad \text{(A.30)} \\
&+ \frac{1}{\sqrt{n}} \sum_{i=1}^n \Bigg(U_i I\left\{|U_i| \leq n^{\frac{1}{q}}\right\} \left(\int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \right) \\
&\qquad - E\left[U_i I\left\{|U_i| \leq n^{\frac{1}{q}}\right\} \left(\int_{(-\infty, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} \right) \right] \Bigg), \qquad \text{(A.31)}
\end{aligned}
$$

where $q := Q\frac{2+\gamma}{2}$, in the second and third step it will be shown that the terms in (A.30) and (A.31) converge to zero in probability uniformly in $\boldsymbol{z} \in \mathbb{R}^d$ respectively.

Step 1: Let $\tilde{Y}_t = h(\boldsymbol{X}_t) + U_t$ where $h \equiv 0$ and $\hat{h}_n(\cdot)$ be the Nadaraya-Watson estimator for $E[\tilde{Y}_t|\boldsymbol{X}_t = \cdot] = h(\cdot)$. Note that $(\tilde{Y}_t, \boldsymbol{X}_t)_{t\in\mathbb{Z}}$ is a strictly stationary process and the uniform rates of convergence for kernel estimators obtained in Chapter 2 (see Lemma 2.2 (i)(a) on page 17 and (2.12) on page 26) can be applied. Thus it holds that

$$
\begin{aligned}
&\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i \int_{\mathbb{R}^d} K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}\left(\frac{f(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}-1\right)d\boldsymbol{y}\right| \\
&= \sqrt{n}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int_{\mathbb{R}^d}\frac{1}{n}\sum_{i=1}^n U_i K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}\frac{f(\boldsymbol{y})-\hat{f}_n(\boldsymbol{y})}{\hat{f}_n(\boldsymbol{y})}d\boldsymbol{y}\right| \\
&\le \sqrt{n}\left(\int_{\mathbb{R}^d}\left(\frac{1}{n}\sum_{i=1}^n U_i K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\frac{1}{\hat{f}_n(\boldsymbol{y})}\right)^2\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)^{\frac{1}{2}} \\
&\qquad\qquad\qquad\qquad\cdot\left(\int_{\mathbb{R}^d}\left(f(\boldsymbol{y})-\hat{f}_n(\boldsymbol{y})\right)^2\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)^{\frac{1}{2}} \\
&\le \sqrt{n}\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|h(\boldsymbol{x})-\hat{h}_n(\boldsymbol{x})|\sup_{\boldsymbol{x}\in\boldsymbol{J}_n}|f(\boldsymbol{x})-\hat{f}_n(\boldsymbol{x})| \\
&= \sqrt{n}O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}\delta_n\right)O_P\left(\sqrt{\frac{\log(n)}{nh_n^d}}+h_n^r p_n\right) \\
&= O_P\left(\frac{\log(n)}{\sqrt{n}h_n^d}\delta_n\right),
\end{aligned}
$$

where the last equality is due to (A.2) on page 114 in the proof of Lemma A.1. Finally,

$$\frac{\log(n)}{\sqrt{n}h_n^d}\delta_n = o(1)$$

follows directly by (A.20) on page 125 in the proof of Lemma A.2.

Step 2: To verify that (A.30) is negligible in probability uniformly in $\boldsymbol{z} \in \mathbb{R}^d$, the following is considered

$$
\begin{aligned}
&\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n U_i I\left\{|U_i|>n^{\frac{1}{q}}\right\}\left(\int K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}d\boldsymbol{y}-\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\right)\right| \\
&\le \frac{1}{\sqrt{n}}\sum_{i=1}^n |U_i| I\left\{|U_i|>n^{\frac{1}{q}}\right\}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}d\boldsymbol{y}-\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}\right| \\
&\le \frac{1}{\sqrt{n}}\sum_{i=1}^n |U_i| I\left\{|U_i|>n^{\frac{1}{q}}\right\}\sup_{\boldsymbol{z}\in\mathbb{R}^d}\left|\int K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})I\{\boldsymbol{y}\le\boldsymbol{z}\}d\boldsymbol{y}\right| \\
&+ \frac{1}{\sqrt{n}}\sum_{i=1}^n |U_i| I\left\{|U_i|>n^{\frac{1}{q}}\right\}\sup_{\boldsymbol{z}\in\mathbb{R}^d}|\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\le\boldsymbol{z}\}|
\end{aligned}
$$

$$\leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}|U_i|I\left\{|U_i|>n^{\frac{1}{q}}\right\}\int|K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)|d\boldsymbol{y}+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|U_i|I\left\{|U_i|>n^{\frac{1}{q}}\right\}$$
$$=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|U_i|I\left\{|U_i|>n^{\frac{1}{q}}\right\}\left(\int|K(\boldsymbol{t})|d\boldsymbol{t}+1\right)$$
$$\leq \bar{C}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}|U_i|I\left\{|U_i|>n^{\frac{1}{q}}\right\},$$

for some $\bar{C}<\infty$. With

$$\begin{aligned}E\left[|U_i|I\left\{|U_i|>n^{\frac{1}{q}}\right\}\right]&=E\left[|U_i|^q|U_i|^{-(q-1)}I\left\{|U_i|>n^{\frac{1}{q}}\right\}\right]\\&\leq n^{-\frac{q-1}{q}}E\left[|U_i|^q\right]\\&=n^{-\frac{q-1}{q}}E\Bigg[\underbrace{E\left[|U_i|^q|\boldsymbol{X}_i\right]}_{\overset{(\mathbf{U})}{\leq} c(\boldsymbol{X}_i)^Q\ \text{a.s.}}\Bigg]\\&\leq n^{-\frac{q-1}{q}}\underbrace{E\left[c(\boldsymbol{X}_i)^Q\right]}_{<\infty}\\&=O\left(n^{-\frac{q-1}{q}}\right)\end{aligned}$$

it therefore holds, that

$$\begin{aligned}&\sup_{\boldsymbol{z}\in\mathbb{R}^d}\Bigg|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\Bigg(U_iI\left\{|U_i|>n^{\frac{1}{q}}\right\}\left(\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})d\boldsymbol{y}-\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}\right)\\&\qquad-E\left[U_iI\left\{|U_i|>n^{\frac{1}{q}}\right\}\left(\int_{(-\boldsymbol{\infty},\boldsymbol{z}]}K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\omega_n(\boldsymbol{y})d\boldsymbol{y}-\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i\leq\boldsymbol{z}\}\right)\right]\Bigg)\Bigg|\\&=O_P\left(n^{\frac{1}{2}}n^{-\frac{q-1}{q}}\right)\\&=O_P\left(n^{-\frac{q-2}{2q}}\right)\\&=o_P(1),\end{aligned}$$

where the last equality holds because $q=Q\frac{2+\gamma}{2}>2$ (as $Q\geq 2,\gamma>0$).

Step 3: To show that the term in (A.31) is negligible in probability uniformly in $\boldsymbol{z}\in\mathbb{R}^d$ the function class

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

is defined. The assertion then follows by

$$\sup_{\varphi\in\mathcal{F}_{n,2}}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\varphi(U_i,\boldsymbol{X}_i)-\int\varphi dP\right)\right|=o_P(1),\tag{A.32}$$

where $(U_t, \boldsymbol{X}_t) \sim P$. To do that the class $\mathcal{F}_{n,2}$ will be covered with finitely many brackets. Then the supremum will be bounded by a maximum which then will be bounded using the exponential inequality in [46].

The partition of $\mathcal{F}_{n,2}$, that is given in Lemma A.7, will be used. Let therefore $\boldsymbol{z}_0, \ldots, \boldsymbol{z}_{J_n}$ be the partition of $\mathbb{R}^d$ and $[\varphi_1^l, \varphi_1^u], \ldots, [\varphi_{J_n}^l, \varphi_{J_n}^u]$ be the corresponding brackets of $\mathcal{F}_{n,2}$ from Lemma A.7 where

$$J_n := N_{[\,]}\left(\varepsilon_n, \mathcal{F}_{n,2}, \|\cdot\|_{L_1(P)}\right).$$

Then for all $\varphi \in \mathcal{F}_{n,2}$ there exists a $j \in \{1, \ldots, J_n\}$ such that $\varphi_j^l \le \varphi \le \varphi_j^u$ and $\|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \le \varepsilon_n$. In Lemma A.7 it will be shown that

$$J_n = O\left(\varepsilon_n^{-d}\right).$$

It holds that

$$\varphi - \int \varphi dP \le \varphi_j^u - \int \varphi_j^u dP + \int \left(\varphi_j^u - \varphi_j^l\right) dP$$

and

$$\varphi - \int \varphi dP \ge \varphi_j^l - \int \varphi_j^l dP - \int \left(\varphi_j^u - \varphi_j^l\right) dP.$$

Let $\varepsilon_n := \frac{1}{\sqrt{n}\log(n)}$. As $\int \left(\varphi_j^u - \varphi_j^l\right) dP = \|\varphi_j^u - \varphi_j^l\|_{L_1(P)} \le \varepsilon_n = o\left(\frac{1}{\sqrt{n}}\right)$ for all $j \in \{1, \ldots, J_n\}$, it holds that

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

Therefore, (A.32) is implied by

$$\max_{1 \le j \le J_n} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\varphi_j^u(U_i, \boldsymbol{X}_i) - \int \varphi_j^u dP\right) \right| = o_P(1) \tag{A.33}$$

and

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

where only (A.33) will be shown in more detail as the second assertion works analogously. In Lemma A.7, φ_j^u is defined as

$$\varphi_j^u(u, \boldsymbol{x})$$

$$
\begin{aligned}
&= uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u \ge 0\} \int\limits_{(-\infty,\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \\
&+ uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u < 0\} \int\limits_{(-\infty,\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \\
&+ uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u \ge 0\} \int\limits_{(-\infty,\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) < 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \\
&+ uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u < 0\} \int\limits_{(-\infty,\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) < 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \\
&- uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u \ge 0\}\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_{j-1}\} \\
&- uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u < 0\}\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\} \\
&= \varphi^u_{j,1}(u,\boldsymbol{x}) + \varphi^u_{j,2}(u,\boldsymbol{x}) + \varphi^u_{j,3}(u,\boldsymbol{x}),
\end{aligned}
$$

where

$$
\begin{aligned}
\varphi^u_{j,1}(u,\boldsymbol{x}) &:= uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u < 0\}\left(\int\limits_{(-\infty,\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\omega_n(\boldsymbol{y})d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\}\right), \\
\varphi^u_{j,2}(u,\boldsymbol{x}) &:= uI\left\{|u| \le n^{\frac{1}{q}}\right\} I\{u \ge 0\}\left(\int\limits_{(-\infty,\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x})\omega_n(\boldsymbol{y})d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_{j-1}\}\right), \\
\varphi^u_{j,3}(u,\boldsymbol{x}) &:= |u|I\left\{|u| \le n^{\frac{1}{q}}\right\}\left(\int_{(-\infty,\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \right. \\
&\qquad \left. - \int_{(-\infty,\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \ge 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\right).
\end{aligned}
$$

for all $j \in \{1,\dots,J_n\}$. Then (A.33) is implied by

$$
\max_{1\le j\le J_n}\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n\left(\varphi^u_{j,k}(U_i,\boldsymbol{X}_i) - \int \varphi^u_{j,k}dP\right)\right| = o_P(1),\ \forall\ k=1,2,3. \tag{A.34}
$$

Equation (A.34) will be shown for $k=1$ in detail. The case $k=2$ is analogues and for $k=3$ only the parts that differ from the first case will be discussed. Following the notation of Liebscher [46]

$$
Z_i := \varphi^u_{j,1}(U_i,\boldsymbol{X}_i) - \int \varphi^u_{j,1}dP
$$

is defined for all $1 \le i \le n$ and $n \in \mathbb{N}$. Note that again the dependency on $n \in \mathbb{N}$ is not reflected in the notation. Let $\{\tilde{\alpha}_n(t) : t \in \mathbb{N}\}$ be the sequence of coefficients of $\{Z_t : 1 \le t \le n, n \in \mathbb{N}\}$ defined as in (1.4) in Definition 1.1 on page 7. For fixed $n \in \mathbb{N}$ they can be bounded by the mixing coefficients of $\{(U_t,\boldsymbol{X}_t) : t \in \mathbb{N}\}$ (see for instance [4], Section 2, remark (iv)). They in turn have the same properties as the mixing coefficients $\{\alpha(t) : t \in \mathbb{N}\}$ of $\{(Y_t, \boldsymbol{X}_t) : t \in \mathbb{N}\}$. Furthermore, Z_i is centered and because

$$\begin{aligned}
&|\varphi^u_{j,1}(U_i, \boldsymbol{X}_i)| \\
&= |U_i| I\left\{|U_i| \leq n^{\frac{1}{q}}\right\} I\{U_i < 0\} \left| \int\limits_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y} - \boldsymbol{X}_i)\, \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\}\right| \\
&\leq n^{\frac{1}{q}} \left| \int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} \frac{1}{h_n^d} K\left(\frac{\boldsymbol{y} - \boldsymbol{X}_i}{h_n}\right) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\}\right| \\
&\leq n^{\frac{1}{q}} \left(\int \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y} - \boldsymbol{X}_i}{h_n}\right)\right| d\boldsymbol{y} + 1\right) \\
&= n^{\frac{1}{q}} \Bigg(\underbrace{\int |K(\boldsymbol{t})| d\boldsymbol{t}}_{<\infty} + 1\Bigg) \\
&= O\left(n^{\frac{1}{q}}\right),
\end{aligned}$$

it holds that $|Z_i| \leq S(n)$ for some $S(n) = O\left(n^{\frac{1}{q}}\right)$. Next it will be shown that

$$E\left[\varphi^u_{j,1}(U_i, \boldsymbol{X}_i)^2\right] = O(h_n) \tag{A.35}$$

and therefore for all $T = 0, \ldots, n-1$ and $1 \leq N \leq n, n \in \mathbb{N}$

$$E\left[\left(\sum_{i=T+1}^{(T+N)\wedge n} Z_i\right)^2\right] \leq N^2 E[Z_i^2] =: D(n, N)$$

with $D(n, N) = N^2 O(h_n)$. To show (A.35) consider

$$\begin{aligned}
&E\left[\varphi^u_{j,1}(U_i, \boldsymbol{X}_i)^2\right] \\
&= E\Bigg[U_i^2 I\left\{U_i \leq n^{\frac{1}{q}}\right\} I\{U_i < 0\} \\
&\qquad \cdot \left(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} \frac{1}{h_n^d} K\left(\frac{\boldsymbol{y} - \boldsymbol{x}}{h_n}\right) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{X_i}) I\{\boldsymbol{X}_i \leq \boldsymbol{z}_j\}\right)^2\Bigg] \\
&\leq E\left[\sigma^2(\boldsymbol{X}_i)\left(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} \frac{1}{h_n^d} K\left(\frac{\boldsymbol{y} - \boldsymbol{x}}{h_n}\right) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\}\right)^2\right] \\
&= \int \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int_{(-\boldsymbol{\infty}, \boldsymbol{z}_j]} \frac{1}{h_n^d} K\left(\frac{\boldsymbol{y} - \boldsymbol{x}}{h_n}\right) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\}\right)^2 d\boldsymbol{x} \\
&= \int \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\}\right) d\boldsymbol{t}\right)^2 d\boldsymbol{x}.
\end{aligned}$$

The later integral over $\mathbb{R}^d$ will be observed in more detail. For $\boldsymbol{x} = (x_1, \ldots, x_d)$, $\boldsymbol{t} = (t_1, \ldots, t_d)$ and $\boldsymbol{z}_j = (z_{j,1}, \ldots, z_{j,d})$ and with $\omega_n(\cdot) = I\{\cdot \in [-c_n, c_n]^d\}$ it holds that

$$\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\}$$

$$
\begin{aligned}
&= \prod_{i=1}^{d} I\{-c_n - t_i h_n \leq x_i \leq (c_n \wedge z_{j,i}) - t_i h_n\} \\
&= \begin{cases} 1, & \text{if } x_i \in [-c_n - t_i h_n, (c_n \wedge z_{j,i}) - t_i h_n],\ \forall\ i = 1, \ldots, d \\ 0, & \text{else} \end{cases}
\end{aligned}
$$

and

$$
\begin{aligned}
\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} &= \prod_{i=1}^{d} I\{-c_n \leq x_i \leq c_n \wedge z_{j,i}\} \\
&= \begin{cases} 1 & \text{if } x_i \in [-c_n, c_n \wedge z_{j,i}]\ \forall\ i = 1, \ldots, d \\ 0 & \text{else} \end{cases}.
\end{aligned}
$$

The integral over $\mathbb{R}^d$ will now be partitioned in the following way

$$
\mathbb{R}^d = \mathop{\times}_{i=1}^{d} \bigcup_{k=1}^{5} I_k^{(i)},
$$

where for $i = 1, \ldots, d$

$$
\begin{aligned}
I_1^{(i)} &:= (-\infty, -c_n - Ch_n) \\
I_2^{(i)} &:= [-c_n - Ch_n, -c_n + Ch_n] \\
I_3^{(i)} &:= (-c_n + Ch_n, (c_n \wedge z_{j,i}) - Ch_n) \\
I_4^{(i)} &:= [(c_n \wedge z_{j,i}) - Ch_n, (c_n \wedge z_{j,i}) + Ch_n] \\
I_5^{(i)} &:= ((c_n \wedge z_{j,i}) + Ch_n, \infty).
\end{aligned}
$$

Note that the kernel function K has compact support $[-C, C]^d$. Because

$$
\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} = \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} = 0
$$

for all $\boldsymbol{t} \in [-C, C]^d$ and $\boldsymbol{x} \in I_1^{(1)} \times \mathbb{R}^{d-1}$ or $\boldsymbol{x} \in I_5^{(1)} \times \mathbb{R}^{d-1}$, it holds that

$$
\begin{aligned}
&\int\limits_{I_1^{(1)}} \int\limits_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right. \\
&\qquad\qquad\qquad \left.\left. - \, \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \right) d\boldsymbol{t} \right)^2 dx_d \ldots dx_2 dx_1 \\
&= \int\limits_{I_5^{(1)}} \int\limits_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right. \\
&\qquad\qquad\qquad \left.\left. - \, \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \right) d\boldsymbol{t} \right)^2 dx_d \ldots dx_2 dx_1 \\
&= 0.
\end{aligned}
$$

On $I_2^{(1)} \times \mathbb{R}^{d-1}$ it holds that

$$
\int\limits_{I_2^{(1)}} \int\limits_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right.
$$

$$- \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\}) \, d\boldsymbol{t} \Big)^2 dx_d \dots dx_2 dx_1$$

$$\le \int_{I_2^{(1)}} \int_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \Big(\int |K(\boldsymbol{t})| \, |(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \le \boldsymbol{z}_j\}$$

$$- \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\})| \, d\boldsymbol{t} \Big)^2 dx_d \dots dx_2 dx_1$$

$$\le \int_{I_2^{(1)}} \int_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \underbrace{\Big(\int |K(\boldsymbol{t})| d\boldsymbol{t} \Big)}_{\le \bar{C}, \text{ for some } \bar{C} < \infty}^2 dx_d \dots dx_2 dx_1$$

$$\le \bar{C}^2 \int_{I_2^{(1)}} \underbrace{\int_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) dx_d \dots dx_2}_{=:\Sigma_1(x_1)} \, dx_1$$

$$= \bar{C}^2 \int_{[-c_n - Ch_n, -c_n + Ch_n]} \Sigma_1(x_1) dx_1$$

$$\le \bar{C}^2 2Ch_n \underbrace{\sup_{x_1 \in \mathbb{R}} \Sigma_1(x_1)}_{< \infty}$$

$$= O(h_n).$$

Similar calculations conclude that

$$\int_{I_4^{(1)}} \int_{\mathbb{R}^{d-1}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \Big(\int K(\boldsymbol{t}) \, (\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \le \boldsymbol{z}_j\}$$

$$- \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\}) \, d\boldsymbol{t} \Big)^2 dx_d \dots dx_2 dx_1$$

$$= O(h_n).$$

What is left to consider is the integral over $I_3^{(1)} \times \mathbb{R}^{d-1} = I_3^{(1)} \times \left(\bigcup_{k=1}^5 I_k^{(2)} \right) \times \mathbb{R}^{d-2}$. Then again

$$\int_{I_3^{(1)}} \int_{I_1^{(2)}} \int_{\mathbb{R}^{d-2}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \Big(\int K(\boldsymbol{t}) \, (\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \le \boldsymbol{z}_j\}$$

$$- \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\}) \, d\boldsymbol{t} \Big)^2 dx_d \dots dx_3 dx_2 dx_1$$

$$= \int_{I_3^{(1)}} \int_{I_5^{(2)}} \int_{\mathbb{R}^{d-2}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \Big(\int K(\boldsymbol{t}) \, (\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \le \boldsymbol{z}_j\}$$

$$- \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \le \boldsymbol{z}_j\}) \, d\boldsymbol{t} \Big)^2 dx_d \dots dx_3 dx_2 dx_1$$

$$= 0.$$

Furthermore,

$$\int\limits_{I_3^{(1)}} \int\limits_{I_2^{(2)}} \int\limits_{\mathbb{R}^{d-2}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right.$$

$$\left.\left. - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \right) d\boldsymbol{t} \right)^2 dx_d \dots dx_3 dx_2 dx_1$$

$$= O(h_n)$$

and

$$\int\limits_{I_3^{(1)}} \int\limits_{I_4^{(2)}} \int\limits_{\mathbb{R}^{d-2}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right.$$

$$\left.\left. - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \right) d\boldsymbol{t} \right)^2 dx_d \dots dx_3 dx_2 dx_1$$

$$= O(h_n).$$

Continuing in this manner all subsets of $\mathbb{R}^d$ except for $\times_{i=1}^d I_3^{(i)}$ have been considered and the integrals either vanish or are of order $O(h_n)$. Finally, the integral over $\times_{i=1}^d I_3^{(i)}$, namely

$$\int\limits_{I_3^{(1)}} \int\limits_{I_3^{(2)}} \cdots \int\limits_{I_3^{(d)}} \sigma^2(\boldsymbol{x}) f(\boldsymbol{x}) \left(\int K(\boldsymbol{t}) \left(\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} \right.\right.$$

$$\left.\left. - \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} \right) d\boldsymbol{t} \right)^2 dx_d \dots dx_2 dx_1$$

vanishes as well because

$$\omega_n(\boldsymbol{x} + \boldsymbol{t}h_n) I\{\boldsymbol{x} + \boldsymbol{t}h_n \leq \boldsymbol{z}_j\} = \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}_j\} = 1$$

for all $\boldsymbol{t} \in [-C, C]^d$ and $\boldsymbol{x} \in \times_{i=1}^d I_3^{(i)}$. Finally (A.35) is proven.

Thus, Liebscher's Theorem 2.1 in [46] can now be applied. By choosing

$$N := \left\lfloor \log(n)^2 \right\rfloor + 1, \ \forall n \in \mathbb{N}$$

it holds that $1 \leq N \leq n$ and

$$4NS(n) = O\left(\log(n)^2 n^{\frac{1}{q}}\right) = \sqrt{n}\, \underbrace{O\left(\log(n)^2 n^{-\frac{q-2}{2q}}\right)}_{\overset{q \geq 2}{=} o(1)} < \sqrt{n}\epsilon,$$

for all $\epsilon > 0$ and for $n \in \mathbb{N}$ large enough. Hence, applying Liebscher's Theorem 2.1 in [46] it holds for all $\epsilon > 0$ and $n \in \mathbb{N}$ large enough that

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

$$
\begin{aligned}
&\leq \sum_{j=1}^{J_n} P\left(\left|\sum_{i=1}^{n}\left(\varphi^u_{j,1}(U_i, \boldsymbol{X}_i) - \int \varphi^u_{j,1} dP\right)\right| > \sqrt{n}\epsilon\right) \\
&\leq J_n \left(4\exp\left(-\frac{n\epsilon^2}{64\frac{n}{N}D(n,N) + \frac{8}{3}\sqrt{n}\epsilon N S(n)}\right) + 4\frac{n}{N}\alpha(N)\right) \\
&= 4\exp\left(\log(J_n) - \frac{n\epsilon^2}{64\frac{n}{N}D(n,N) + \frac{8}{3}\sqrt{n}\epsilon N S(n)}\right) \qquad \text{(A.36)}\\
&+ 4\exp\left(\log(J_n) + \log\left(\frac{n}{N}\right) + \log(\alpha(N))\right). \qquad \text{(A.37)}
\end{aligned}
$$

It is left to show that the exponents in (A.36) and (A.37) diverge to $-\infty$ as $n \to \infty$. Starting with (A.37), it is to show that

$$
\frac{64\frac{n}{N}D(n,N) + \frac{8}{3}\sqrt{n}\epsilon N S(n)}{n\epsilon^2} = o\left(\frac{1}{\log(J_n)}\right).
$$

Using $\log(J_n) = O(\log(n))$ and the bounds for $S(n)$, $D(n,N)$ and N it holds that

$$
\begin{aligned}
\frac{64\frac{n}{N}D(n,N) + \frac{8}{3}\sqrt{n}\epsilon N S(n)}{n\epsilon^2}\log(J_n) &= \left(\frac{64}{\epsilon^2}\frac{1}{N}D(n,N) + \frac{1}{\sqrt{n}\epsilon}\frac{8}{3}NS(n)\right)\log(J_n) \\
&= O\left(\left(\log(n)^2 h_n + \log(n)^2 n^{-\frac{1}{2}} n^{\frac{1}{q}}\right)\log(n)\right) \\
&= O\left(\log(n)^3 h_n + \frac{\log(n)^3}{n^{\frac{q-2}{2q}}}\right) \\
&= o(1),
\end{aligned}
$$

due to equation (3.6) in **(B3)** and because $q > 2$. Concerning (A.37), it is to show that

$$
\frac{1}{|\log(\alpha(N))|} = o\left(\frac{1}{\log(J_n) + \log\left(\frac{n}{N}\right)}\right).
$$

Again inserting the bounds and the exponential rates of convergence of the mixing coefficients in **(G)** yields

$$
\begin{aligned}
\frac{1}{|\log(\alpha(N))|}\left(\log(J_n) + \log\left(\frac{n}{N}\right)\right) &= O\left(\frac{1}{N}\log(n) + \frac{1}{N}(\log(n) + \log(N))\right) \\
&= O\left(\frac{1}{\log(n)}\right) \\
&= o(1),
\end{aligned}
$$

which finally proves (A.34) for $k = 1$. The proof of the case $k = 3$ works similar. Only the parts that differ from the first case will be discussed in more detail. Again

$$
Z_i := \varphi^u_{j,3}(U_i, \boldsymbol{X}_i) - \int \varphi^u_{j,3} dP
$$

is defined for all $1 \leq i \leq n$ and $n \in \mathbb{N}$. Then Z_i is centered and $|Z_i| \leq S(n)$ for some $S(n) = O\left(n^{\frac{1}{q}}\right)$. It will be shown that

$$
E[Z_i^2] = O\left(\log(n)^{-1} n^{-\frac{q-2}{2q}}\right), \qquad \text{(A.38)}
$$

and therefore $D(n,N) = N^2 O\left(\log(n)^{-1} n^{-\frac{q-2}{2q}}\right)$. It holds that

$$E[Z_i^2] \leq S(n)E[|Z_i|] = O\left(n^{\frac{1}{q}}\right) E\left[\left|\varphi_{j,3}^u(U_i,\boldsymbol{X}_i)\right|\right].$$

Furthermore, it can be obtained that

$$\begin{aligned}
&E\left[\left|\varphi_{j,3}^u(U_i,\boldsymbol{X}_i)\right|\right] \\
&= E\Bigg[|U_i| I\left\{|U_i| \leq n^{\frac{1}{q}}\right\}\Bigg(\int_{(-\infty,\boldsymbol{z}_j]} K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\geq 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y} \\
&\qquad\qquad - \int_{(-\infty,\boldsymbol{z}_{j-1}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{X}_i)\geq 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\Bigg)\Bigg] \\
&\leq E\left[|U_i|\left(\int_{(-\infty,\boldsymbol{z}_j]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{X}_i\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y} - \int_{(-\infty,\boldsymbol{z}_{j-1}]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{X}_i\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)\right] \\
&\leq E\left[\sigma(\boldsymbol{X}_i)\left(\int_{(-\infty,\boldsymbol{z}_j]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{X}_i\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y} - \int_{(-\infty,\boldsymbol{z}_{j-1}]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{X}_i\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\right)\right] \\
&= \int \sigma(\boldsymbol{x})\left(\int_{(-\infty,\boldsymbol{z}_j]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{x}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y} - \int_{(-\infty,\boldsymbol{z}_{j-1}]} \left|K_{h_n}\left(\boldsymbol{y}-\boldsymbol{x}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\right) f(\boldsymbol{x})d\boldsymbol{x} \\
&= \int_{(-\infty,\boldsymbol{z}_j]}\int \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})\sigma(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{y} \\
&\qquad\qquad - \int_{(-\infty,\boldsymbol{z}_{j-1}]}\int \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})\sigma(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{y} \\
&= Q_n(\boldsymbol{z}_j) - Q_n(\boldsymbol{z}_{j-1}) \\
&\leq \frac{\varepsilon_n}{2},
\end{aligned}$$

with $Q_n : \mathbb{R}^d \to \mathbb{R}$ defined in (A.50) on page 154 and where the last inequality holds because of (A.51) on page 155 (see proof of Lemma A.7). As $\varepsilon_n = \frac{1}{\sqrt{n}\log(n)}$, it holds that

$$E[Z_i^2] = O\left(n^{\frac{1}{q}} n^{-\frac{1}{2}}\log(n)^{-1}\right) = O\left(\log(n)^{-1} n^{-\frac{q-2}{2q}}\right),$$

and thus equation (A.38) is valid. After applying again Liebscher's inequality with $N = \lfloor \log(n)^2 \rfloor + 1$, the upper bound converges to zero as $n \to \infty$, because in this case it holds that

$$\begin{aligned}
\left(\frac{64}{\epsilon^2}\frac{1}{N}D(n,N) + \frac{1}{\sqrt{n}\epsilon}\frac{8}{3}NS(n)\right)\log(J_n) &= O\left(\left(\log(n) n^{-\frac{q-2}{2q}} + \log(n)^2 n^{-\frac{1}{2}} n^{\frac{1}{q}}\right)\log(n)\right) \\
&= O\left(\frac{\log(n)^2}{n^{\frac{q-2}{2q}}} + \frac{\log(n)^3}{n^{\frac{q-2}{2q}}}\right) \\
&= o(1),
\end{aligned}$$

as $q > 2$. $\square$

Lemma A.4. *Under the same assumptions of Theorem 3.1*

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i\omega_n(\boldsymbol{X}_i)I\{\boldsymbol{X}_i \le \boldsymbol{z}\} = T_n(s,\boldsymbol{z}) + o_P(1),$$

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

Remark. For the proof of Lemma A.4 a different technique will be used. A sequential empirical process indexed in some function class will be defined and it will be shown that it satisfies an asymptotic equicontinuity condition by an application of Corollary B.3 from Appendix B. Note that here the different bracketing notion introduced in Definition 1.6 is needed. The final argument uses the fact that all functions in that class converge to zero with respect to the considered norm.

For the sake of understanding it is to mention that the method using Liebscher's inequality does not lead to a proof of Lemma A.4. The reason is that no rates of convergence for the variance of the random variables, that are to define, can be obtained (using the notations in [46] the problematic term is $D(n,N)$). The alternative approach does not need rates of convergence, but only convergence to zero.

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

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{\lfloor ns\rfloor} U_i I\{\boldsymbol{X}_i \le \boldsymbol{z}\} I\{\boldsymbol{X}_i \notin [-c_n,c_n]^d\} = o_P(1). \tag{A.39}$$

To do that, define the function class

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

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 $\int \varphi dP = 0$. First it will be shown, by an application of Corollary B.3 from Chapter B, that for all $\delta_n \searrow 0$ and with $d(\varphi,\psi) := \|\varphi-\psi\|_{L_{Q\frac{2+\gamma}{2}}(P)}$,

$$\sup_{\substack{\{s,t\in[0,1],\varphi,\psi\in\mathcal{F}:\\ |s-t|+d(\varphi,\psi)<\delta_n\}}} |G_n(s,\varphi) - G_n(t,\psi)| = o_P(1). \tag{A.40}$$

Hence, conditions **(A1)**, **(A2)** and **(A3)** from Theorem B.1 and Corollary B.3 from Chapter B will be shown. Condition **(A1)** is implied by the stronger assumption in **(G)** with $Q > (d+1)(2+\gamma)$ from assumption **(U)**. To show the validity of **(A2)** the choice of approximating functions and bounding functions of the function class $\mathcal{F}$ will be discussed in more detail.

Let the notations of h, h_i and H, H_i on page 42 in the proof of Theorem 3.2 be used. Again note that for all $i = 1, \dots, d$, H_i is continuous, monotonically increasing and $H_i(-\infty) = H(-\boldsymbol{\infty}) = 0$, as well as $H_i(\infty) = H(\boldsymbol{\infty}) \leq M$, for $M < \infty$ from assumption **(U)**. Let additionally the following notations be introduced

$$\tilde{H} : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int_{[\boldsymbol{x},\boldsymbol{\infty})} h(\boldsymbol{t}) d\boldsymbol{t},$$

and for all $i = 1, \dots, d$

$$\tilde{H}_i : \mathbb{R} \to \mathbb{R}, x \mapsto \int_x^\infty h_i(t) dt,$$

where for all $i = 1, \dots, d$, $\tilde{H}_i$ is continuous, monotonically decreasing and $\tilde{H}_i(\infty) = \tilde{H}(\boldsymbol{\infty}) = 0$, as well as $\tilde{H}_i(-\infty) = \tilde{H}(-\boldsymbol{\infty}) = H(\boldsymbol{\infty}) \leq M$.

Now let for $\varepsilon > 0$, $N_i \in \mathbb{N}$, $i = 1, \dots, d$ and $K \in \mathbb{N}$

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

be a partition of $\mathbb{R}$ for all $i = 1, \dots, d$ such that

$$H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right) \leq \frac{\varepsilon^2}{3d}\ \forall\ j_i = 1, \dots, N_i,\ i = 1, \dots, d \tag{A.41}$$

and

$$0 = a_0 < \cdots < a_K = \infty$$

be a partition of $\mathbb{R}_+$ such that

$$H_i\left(a_k\right) - H_i\left(a_{k-1}\right) \leq \frac{\varepsilon^2}{3d}\ \forall\ k = 1, \dots, K,\ i = 1, \dots, d \tag{A.42}$$

and

$$\tilde{H}_i\left(-a_k\right) - \tilde{H}_i\left(-a_{k-1}\right) \leq \frac{\varepsilon^2}{3d}\ \forall\ k = 1, \dots, K,\ i = 1, \dots, d. \tag{A.43}$$

Due to aforementioned properties of H_i and $\tilde{H}_i$, K and N_i can be chosen to be smaller than $6dM\varepsilon^{-2}$ for all $i = 1, \dots, d$. For simplicity reasons the following notation will be used. For $\boldsymbol{j} = (j_1, \dots, j_d) \in \mathbb{N}^d$ let

$$\boldsymbol{z_j} := \left(z_{j_1,1}, \dots, z_{j_d,d}\right),$$

and $\boldsymbol{j-1} := (j_1 - 1, \dots, j_d - 1) \in \mathbb{N}^d$. For $\boldsymbol{z} := (z_1, \dots, z_d)$ and $\underline{a} := (a, \dots, a) \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\varphi(u, \boldsymbol{x}) &:= uI\left\{\boldsymbol{x} \leq \boldsymbol{z}\right\} I\{\boldsymbol{x} \notin [-\underline{a}, \underline{a}]\} \\
&= uI\left\{\boldsymbol{x} \leq \boldsymbol{z}\right\}(1 - I\{\boldsymbol{x} \in [-\underline{a}, \underline{a}]\}) \\
&= u\left(I\left\{\boldsymbol{x} \leq \boldsymbol{z}\right\} - I\left\{-\underline{a} \leq \boldsymbol{x} \leq \min(\boldsymbol{z}, \underline{a})\right\}\right) \\
&= uI\Bigg\{\boldsymbol{x} \in \underbrace{(-\boldsymbol{\infty}, \boldsymbol{z}] \setminus [-\underline{a}, \min(\boldsymbol{z}, \underline{a})]}_{=:A_{\boldsymbol{z},a}}\Bigg\}.
\end{aligned}$$

Note that $I\{\boldsymbol{a} < \boldsymbol{x} \leq \boldsymbol{b}\} := 0$ for $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^d$ with $b_i \leq a_i$ for some $i \in \{1, \ldots, d\}$. Define the following sets

$$\begin{aligned} B_{\boldsymbol{j},k} &:= (-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}] \setminus [-\underline{a}_k, \min(\boldsymbol{z_{j-1}}, \underline{a}_k)] \subset \mathbb{R}^d, \\ C_{\boldsymbol{j},k} &:= (-\boldsymbol{\infty}, \boldsymbol{z_j}] \setminus [-\underline{a}_{k-1}, \min(\boldsymbol{z_j}, \underline{a}_{k-1})] \subset \mathbb{R}^d. \end{aligned}$$

Choose for $\boldsymbol{z} \in \mathbb{R}^d$, $\boldsymbol{j} \in \times_{i=1}^d \{1, \ldots, N_i\}$ such that $\boldsymbol{z} \in (\boldsymbol{z_{j-1}}, \boldsymbol{z_j}]$ and for $a \in \mathbb{R}_+$, $k \in \{1, \ldots, K\}$ such that $a \in (a_{k-1}, a_k] \subset \mathbb{R}$ and therefore $\underline{a} \in (\underline{a}_{k-1}, \underline{a}_k] \subset \mathbb{R}^d$. Then it holds that

$$B_{\boldsymbol{j},k} \subseteq A_{\boldsymbol{z},a} \subseteq C_{\boldsymbol{j},k}$$

and by defining

$$a_{\boldsymbol{j},k}(u, \boldsymbol{x}) := uI\left\{\boldsymbol{x} \in B_{\boldsymbol{j},k}\right\}$$

and

$$b_{\boldsymbol{j},k}(u, \boldsymbol{x}) := |u| I\left\{\boldsymbol{x} \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\}.$$

it holds for all $u \in \mathbb{R}$ and $\boldsymbol{x} \in \mathbb{R}^d$ that

$$\begin{aligned} |\varphi(u, \boldsymbol{x}) - a_{\boldsymbol{j},k}(u, \boldsymbol{x})| &= |uI\{\boldsymbol{x} \in A_{\boldsymbol{z},a}\} - uI\{\boldsymbol{x} \in B_{\boldsymbol{j},k}\}| \\ &= |u| I\{\boldsymbol{x} \in A_{\boldsymbol{z},a} \setminus B_{\boldsymbol{j},k}\} \\ &\leq |u| I\left\{\boldsymbol{x} \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\} \\ &= b_{\boldsymbol{j},k}(u, \boldsymbol{x}). \end{aligned}$$

Furthermore, for all $\boldsymbol{j} \in \times_{i=1}^d \{1, \ldots, N_i\}$ and $k \in \{1, \ldots, K\}$ it holds that

$$\|b_{\boldsymbol{j},k}\|_{L_2(P)} \leq \varepsilon \text{ and } \max_{2 \leq i \leq Q} \left(\int |b_{\boldsymbol{j},k}|^{i\frac{2+\gamma}{2}} dP \right)^{\frac{1}{2}} \leq \varepsilon. \tag{A.44}$$

To show (A.44), it can first be obtained that

$$\begin{aligned} \|b_{\boldsymbol{j},k}\|_{L_2(P)}^2 &= E\left[|U_t|^2 I\left\{\boldsymbol{X}_t \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\}\right] \\ &= E\left[\sigma^2(\boldsymbol{X}_t) I\left\{\boldsymbol{X}_t \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\}\right] \\ &= \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} \sigma^2(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u} \\ &\leq \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} \bar{c}(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u} \\ &= \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} h(\boldsymbol{u}) d\boldsymbol{u}, \end{aligned}$$

and for all $i = 2, \ldots, Q$

$$\begin{aligned} \int |b_{\boldsymbol{j},k}|^{i\frac{2+\gamma}{2}} dP &= E\left[|U_t|^{i\frac{2+\gamma}{2}} I\left\{\boldsymbol{X}_t \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\}\right] \\ &\leq E\left[c(\boldsymbol{X}_t)^i I\left\{\boldsymbol{X}_t \in C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}\right\}\right] \\ &= \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} c(\boldsymbol{u})^i f(\boldsymbol{u}) d\boldsymbol{u} \end{aligned}$$

$$\begin{aligned}
&\leq \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} \bar{c}(\boldsymbol{u}) f(\boldsymbol{u}) d\boldsymbol{u} \\
&= \int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} h(\boldsymbol{u}) d\boldsymbol{u}.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\int_{C_{\boldsymbol{j},k} \setminus B_{\boldsymbol{j},k}} h(\boldsymbol{u}) d\boldsymbol{u} &= \int_{C_{\boldsymbol{j},k}} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{B_{\boldsymbol{j},k}} h(\boldsymbol{u}) d\boldsymbol{u} \\
&= \left(\int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{[-\underline{a}_{k-1}, \min(\boldsymbol{z_j}, \underline{a}_{k-1})]} h(\boldsymbol{u}) \right) \\
&- \left(\int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{[-\underline{a}_{k}, \min(\boldsymbol{z_{j-1}}, \underline{a}_{k})]} h(\boldsymbol{u}) d\boldsymbol{u} \right) \\
&= H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}) \\
&+ \int_{[-\underline{a}_{k}, \min(\boldsymbol{z_{j-1}}, \underline{a}_{k})]} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{[-\underline{a}_{k-1}, \min(\boldsymbol{z_j}, \underline{a}_{k-1})]} h(\boldsymbol{u}) d\boldsymbol{u}
\end{aligned}$$

holds. Now it can be shown that

$$\begin{aligned}
&\int_{[-\underline{a}_{k}, \min(\boldsymbol{z_{j-1}}, \underline{a}_{k})]} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{[-\underline{a}_{k-1}, \min(\boldsymbol{z_j}, \underline{a}_{k-1})]} h(\boldsymbol{u}) d\boldsymbol{u} \\
&\leq \left(\tilde{H}(-\underline{a}_k) - \tilde{H}(-\underline{a}_{k-1}) \right) + \left(H(\underline{a}_k) - H(\underline{a}_{k-1}) \right) \qquad \text{(A.45)}
\end{aligned}$$

To argue the validity of (A.45), the following three cases will be considered.

Case 1: It holds that

$$\min(\boldsymbol{z_{j-1}}, \underline{a}_k) > \min(\boldsymbol{z_j}, \underline{a}_{k-1}).$$

Case 2: It holds that

$$\min(\boldsymbol{z_{j-1}}, \underline{a}_k) \leq \min(\boldsymbol{z_j}, \underline{a}_{k-1}).$$

Case 3: There exists an $I \subset \{1, \ldots, d\}$ with $|I| \in \{1, \ldots, d-1\}$ such that

$$\min(\boldsymbol{z}_{j_i - 1}, a_k) > \min(\boldsymbol{z}_{j_i}, a_{k-1}), \forall i \in I$$

and the inequality does not hold for $i \notin I$.

Note that if $\min(\boldsymbol{z}_{j_i - 1}, a_k) > \min(\boldsymbol{z}_{j_i}, a_{k-1})$ holds for all $i \in \{1, \ldots, d\}$, this is case 1. If it does not hold for any $i \in \{1, \ldots, d\}$, this is case 2.

Case 1: If case 1 holds, then $\boldsymbol{z}_j > \boldsymbol{z}_{j-1} \geq \min(\boldsymbol{z}_{j-1}, \underline{a}_k) > \min(\boldsymbol{z}_j, \underline{a}_{k-1})$ and therefore $\min(\boldsymbol{z}_j, \underline{a}_{k-1}) = \underline{a}_{k-1}$. Hence, it holds that

$$\int_{\left[-\underline{a}_k, \underbrace{\min(\boldsymbol{z_{j-1}}, \underline{a}_k)}_{\leq \underline{a}_k}\right]} h(\boldsymbol{u}) d\boldsymbol{u} - \int_{\left[-\underline{a}_{k-1}, \underbrace{\min(\boldsymbol{z_j}, \underline{a}_{k-1})}_{= \underline{a}_{k-1}}\right]} h(\boldsymbol{u}) d\boldsymbol{u}$$

$$\leq \underbrace{\int_{[-\underline{a}_k,\underline{a}_k]} h(\boldsymbol{u})d\boldsymbol{u} - \int_{[-\underline{a}_{k-1},\underline{a}_k]} h(\boldsymbol{u})d\boldsymbol{u}}_{\text{left hand side (l.h.s.)}}$$

$$\leq \underbrace{\left(\tilde{H}(-\underline{a}_k) - \tilde{H}(-\underline{a}_{k-1})\right) + (H(\underline{a}_k) - H(\underline{a}_{k-1})),}_{\text{right hand side (r.h.s.)}}$$

which proves the assertion in (A.45). Figure A.1 shows a visualization of the last inequality for $d = 2$.

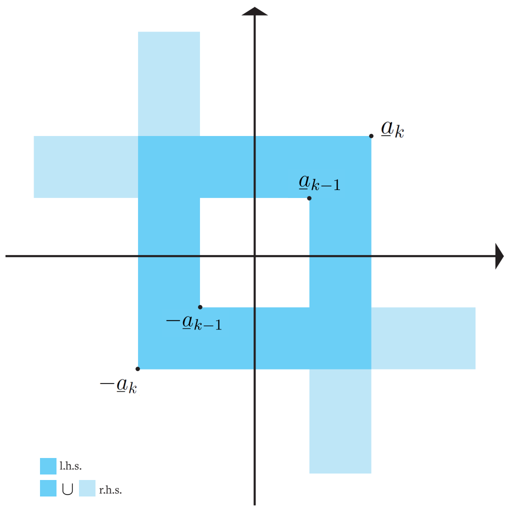

Figure A.1: Case 1 for $d = 2$

Case 2: If case 2 holds, then

$$\underbrace{\int_{[-\underline{a}_k,\min(\boldsymbol{z_{j-1}},\underline{a}_k)]} h(\boldsymbol{u})d\boldsymbol{u} - \int_{[-\underline{a}_{k-1},\min(\boldsymbol{z_j},\underline{a}_{k-1})]} h(\boldsymbol{u})d\boldsymbol{u}}_{\text{l.h.s.}}$$

$$\leq \underbrace{\left(\tilde{H}(-\underline{a}_k) - \tilde{H}(-\underline{a}_{k-1})\right) - \int_S h(\boldsymbol{u})d\boldsymbol{u},}_{\text{r.h.s.}}$$

for some $S \subset \mathbb{R}^d$. Figure A.2 shows a visualization of the inequality for $d = 2$. As

$$\int_S h(\boldsymbol{u})d\boldsymbol{u} \geq 0,$$

the assertion in (A.45) follows.

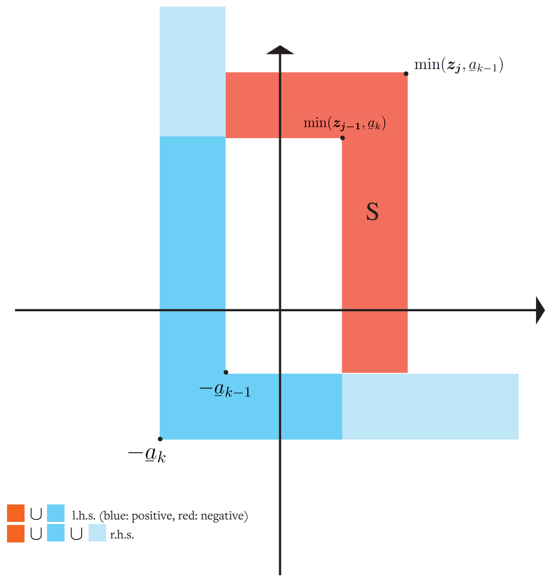

Figure A.2: Case 2 for $d = 2$

Case 3: If case 3 holds, then

$$\underbrace{\int_{[-\underline{a}_k, \min(\boldsymbol{z}_{j-1}, \underline{a}_k)]} h(\boldsymbol{u})d\boldsymbol{u} - \int_{[-\underline{a}_{k-1}, \min(\boldsymbol{z}_j, \underline{a}_{k-1})]} h(\boldsymbol{u})d\boldsymbol{u}}_{\text{l.h.s.}}$$
$$\leq \underbrace{\left(\tilde{H}(-\underline{a}_k) - \tilde{H}(-\underline{a}_{k-1})\right) + (H(\underline{a}_k) - H(\underline{a}_{k-1})) - \int_S h(\boldsymbol{u})d\boldsymbol{u},}_{\text{r.h.s.}}$$

for some $S \subset \mathbb{R}^d$. Figure A.3 shows a visualization of the inequality for $d = 2$. As again $\int_S h(\boldsymbol{u})d\boldsymbol{u} \geq 0$, the assertion in (A.45) follows.

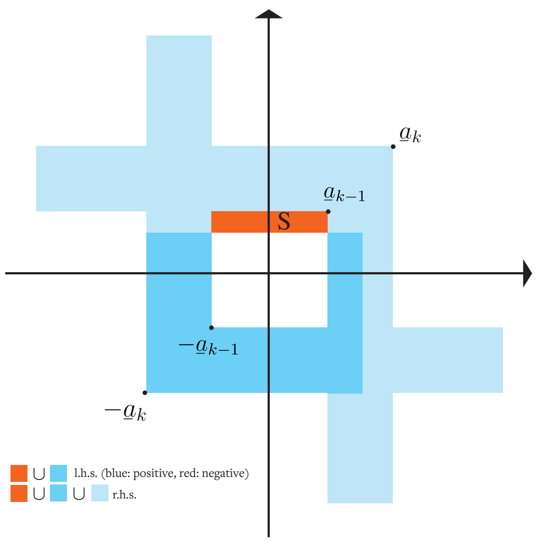

Figure A.3: Case 3 for $d = 2$

Hence, for all three cases, the assertion in (A.45) holds. Furthermore, as it has been shown in the proof of Theorem 3.2, it holds that

$$H(\boldsymbol{z_j}) - H(\boldsymbol{z_{j-1}}) \le \sum_{i=1}^{d} \left(H_i\left(z_{j_i,i}\right) - H_i\left(z_{j_i-1,i}\right)\right) \overset{(A.41)}{\le} d\frac{\varepsilon^2}{3d} = \frac{\varepsilon^2}{3}.$$

Similarly, it holds that

$$H(\underline{a}_k) - H(\underline{a}_{k-1}) \le \sum_{i=1}^{d} \left(H_i\left(a_k\right) - H_i\left(a_{k-1}\right)\right) \overset{(A.42)}{\le} d\frac{\varepsilon^2}{3d} = \frac{\varepsilon^2}{3}$$

and

$$\tilde{H}(-\underline{a}_k) - \tilde{H}(-\underline{a}_{k-1}) \le \sum_{i=1}^{d} \left(\tilde{H}_i\left(-a_k\right) - \tilde{H}_i\left(-a_{k-1}\right)\right) \overset{(A.43)}{\le} d\frac{\varepsilon^2}{3d} = \frac{\varepsilon^2}{3},$$

which proves (A.44). As $N_i = O\left(\varepsilon^{-2}\right)$ for all $i = 1, \ldots, d$ and $K = O\left(\varepsilon^{-2}\right)$, it holds that

$$\tilde{N}_{[\ ]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}\right) \le \left| \bigtimes_{i=1}^{d} \{1, \ldots, N_i\} \times \{1, \ldots, K\} \right| = \prod_{i=1}^{d} N_i \cdot K = O\left(\varepsilon^{-2(d+1)}\right).$$

Hence, the assumptions on the bracketing number and bounding functions in **(A2)** are satisfied as $Q > (d+1)(2+\gamma)$. Additionally, **(A3)** is also satisfied as

$$\begin{aligned}
\sup_{\varphi \in \mathcal{F}} \int |\varphi|^{Q\frac{2+\gamma}{2}} dP &= \sup_{\boldsymbol{z} \in \mathbb{R}^d} \sup_{a \in \mathbb{R}_+} E\left[|U_t|^{Q\frac{2+\gamma}{2}} I\{\boldsymbol{X}_t \le \boldsymbol{z}\} I\{\boldsymbol{X}_t \notin (-\underline{a}, \underline{a}]\}\right] \\
&= E\left[|U_t|^{Q\frac{2+\gamma}{2}}\right] \\
&\le \int c(\boldsymbol{u})^Q dF(\boldsymbol{u}) \\
&< \infty.
\end{aligned}$$

and because $\bar{F} : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R},\ (u, \boldsymbol{x}) \mapsto u$ is an envelope function of $\mathcal{F}$ that fulfills

$$\begin{aligned}
\int |\bar{F}|^Q dP &= E\left[|U_t|^Q\right] \\
&\le E\left[|U_t|^{Q\frac{2+\gamma}{2}}\right]^{\frac{2}{2+\gamma}} \\
&= E\left[E\left[|U_t|^{Q\frac{2+\gamma}{2}} | X_t\right]\right]^{\frac{2}{2+\gamma}} \\
&\le E\left[c(X_t)^Q\right]^{\frac{2}{2+\gamma}} \\
&= \left(\int c(\boldsymbol{u})^Q dF(\boldsymbol{u})\right)^{\frac{2}{2+\gamma}} \\
&< \infty.
\end{aligned}$$

An application of the first part of Corollary B.3 concludes the proof of (A.40).

Next for some fixed $\boldsymbol{z} \in \mathbb{R}^d$ defining

$$\varphi_n(u, \boldsymbol{x}) := uI\{\boldsymbol{x} \leq \boldsymbol{z}\}I\{\boldsymbol{x} \notin [-c_n, c_n]^d\},$$

it holds that $\varphi_n \in \mathcal{F}$ for all $n \in \mathbb{N}$ and

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

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

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

With (A.40) the last term is $o_P(1)$ which proves the assertion in (A.39) and therefore the assertion of the lemma.

□

A.2 Auxiliary lemmata

In this section, bounds for the bracketing numbers used in Lemmata A.1, A.2 and A.3 will be proven. They are stated in Lemmata A.5, A.6 and A.7 respectively.

Lemma A.5. *For the bracketing number $N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}\right)$ where $\boldsymbol{X}_t \sim P$ and*

$$\mathcal{F} := \left\{\boldsymbol{x} \mapsto I\{\boldsymbol{x} \leq \boldsymbol{z}\} : \boldsymbol{z} \in \mathbb{R}^d\right\}$$

defined on page 115 in the proof of Lemma A.1, it holds that

$$N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}\right) = O\left(\varepsilon^{-2d}\right).$$

Proof. Let F_i denote the one-dimensional distribution function of the i-th component of $\boldsymbol{X}_1 = (X_{1,1}, \ldots, X_{1,d})$, for all $i = 1, \ldots, d$, and F the d-dimensional distribution function of $\boldsymbol{X}_1$. Let now for $\varepsilon > 0$ and some $N_i \in \mathbb{N}$, $i = 1, \ldots, d$

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

be a partition of $\mathbb{R}$ for all $i = 1, \ldots, d$ such that

$$F_i\left(z_{j_i,i}\right) - F_i\left(z_{j_i-1,i}\right) \leq \frac{\varepsilon^2}{d} \;\forall\; j_i = 1, \ldots, N_i,\; i = 1, \ldots, d. \tag{A.46}$$

Since F_i is continuous and $F_i(-\infty) = 0, F_i(\infty) = 1$ holds, N_i can be chosen to be smaller than $2d\varepsilon^{-2}$ 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

$$\boldsymbol{z_j} := \left(z_{j_1,1}, \ldots, z_{j_d,d}\right),$$

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

$$\varphi^u_{\boldsymbol{j}}(\boldsymbol{x}) := I\left\{\boldsymbol{x} \leq \boldsymbol{z_j}\right\}$$

and

$$\varphi^l_{\boldsymbol{j}}(\boldsymbol{x}) := I\left\{\boldsymbol{x} \leq \boldsymbol{z_{j-1}}\right\}.$$

With these notations it holds that for all $\boldsymbol{z} \in \mathbb{R}^d$ 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 all $\varphi \in \mathcal{F}$ there exists a $\boldsymbol{j} \in \times_{i=1}^d \{1, \ldots, N_i\}$ such that $\varphi \in [\varphi^l_{\boldsymbol{j}}, \varphi^u_{\boldsymbol{j}}]$. Furthermore, it holds that

$$\begin{aligned}
\left\|\varphi^u - \varphi^l\right\|^2_{L_2(P)} &= E\left[\left|I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\right\}\right|^2\right] \\
&= E\left[I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_j}\right\} - I\left\{\boldsymbol{X}_t \leq \boldsymbol{z_{j-1}}\right\}\right] \\
&= F(\boldsymbol{z_j}) - F(\boldsymbol{z_{j-1}}) \\
&\overset{(*)}{\leq} \sum_{i=1}^d \left(F_i(z_{j_i,i}) - F_i(z_{j_{i-1},i})\right) \\
&\overset{(A.46)}{\leq} d\frac{\varepsilon^2}{d} = \varepsilon^2,
\end{aligned}$$

where the inequality in $(*)$ holds by similar calculations as in the proof of Theorem 3.2 on page 44. Since $N_i = O(\varepsilon^{-2})$ for all $i = 1, \ldots, d$, it holds that

$$N_{[\,]}\left(\varepsilon, \mathcal{F}, \|\cdot\|_{L_2(P)}\right) \leq \left|\mathop{\times}_{i=1}^d \{1, \ldots, N_i\}\right| = \prod_{i=1}^d N_i = O(\varepsilon^{-2d}).$$

□

Lemma A.6. *For the bracketing number $N_{[\,]}\left(\varepsilon, \mathcal{F}_{n,1}, \|\cdot\|_{L_1(P)}\right)$ where $\boldsymbol{X}_t \sim P$ and*

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

defined on page 125 in the proof of Lemma A.2, it holds that

$$N_{[\,]}\left(\varepsilon, \mathcal{F}_{n,1}, \|\cdot\|_{L_1(P)}\right) = O\left(\varepsilon^{-d}\right),$$

independent of $n \in \mathbb{N}$.

Proof. First some notations are necessary. Let for all $n \in \mathbb{N}$

$$\begin{aligned} p_n &: \mathbb{R}^d \to \mathbb{R} \\ &\boldsymbol{x} \mapsto \int |m(\boldsymbol{x}) - m(\boldsymbol{y})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_n}\right) \right| \omega_n(\boldsymbol{x}) f(\boldsymbol{y}) d\boldsymbol{y} \end{aligned}$$

and

$$P_n : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty},\boldsymbol{x}]} p_n(\boldsymbol{t}) d\boldsymbol{t},$$

as well as for all $n \in \mathbb{N}$ and $i = 1, \ldots, d$

$$\begin{aligned} p_{n,i} &: \mathbb{R} \to \mathbb{R} \\ &x \mapsto \int \cdots \int p_n(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_d \end{aligned}$$

and

$$P_{n,i} : \mathbb{R} \to \mathbb{R},\ x \mapsto \int_{-\infty}^{x} p_{n,i}(t) dt.$$

Note that for all $n \in \mathbb{N}$ and $i = 1, \ldots, d$, $P_{n,i}$ and P_n are monotonically increasing and it holds that

$$\begin{aligned} P_{n,i}(\infty) &= P_n(\boldsymbol{\infty}) \\ &= \int \int |m(\boldsymbol{t}) - m(\boldsymbol{y})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{t}-\boldsymbol{y}}{h_n}\right) \right| \omega_n(\boldsymbol{t}) f(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{t} \\ &= \int \int |m(\boldsymbol{y} + \boldsymbol{x} h_n) - m(\boldsymbol{y})| |K(\boldsymbol{x})| \omega_n(\boldsymbol{y} + \boldsymbol{x} h_n) f(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x} \\ &\leq \int \int \left| \sum_{|\boldsymbol{i}|=1} \frac{D^{\boldsymbol{i}} m(\boldsymbol{y}) (\boldsymbol{x} h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right| |K(\boldsymbol{x})| \omega_n(\boldsymbol{y} + \boldsymbol{x} h_n) f(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x} \qquad \text{(A.47)} \\ &+ \int \int \left| \sum_{|\boldsymbol{i}|=2} \frac{D^{\boldsymbol{i}} m(\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{x}}) (\boldsymbol{x} h_n)^{\boldsymbol{i}}}{\boldsymbol{i}!} \right| |K(\boldsymbol{x})| \omega_n(\boldsymbol{y} + \boldsymbol{x} h_n) f(\boldsymbol{y}) d\boldsymbol{y} d\boldsymbol{x}, \qquad \text{(A.48)} \end{aligned}$$

for some $\boldsymbol{\xi}_{\boldsymbol{y},\boldsymbol{x}}$ on the line segment between $\boldsymbol{y} + \boldsymbol{x} h_n$ and $\boldsymbol{y}$. The term in (A.47) can be bounded by

$$h_n \sum_{|\boldsymbol{i}|=1} \frac{1}{\boldsymbol{i}!} \underbrace{\sup_{\boldsymbol{z} \in \boldsymbol{I}_n} |D^{\boldsymbol{i}} m(\boldsymbol{z})|}_{=O(q_n)} \underbrace{\int |\boldsymbol{x}|^{\boldsymbol{i}} |K(\boldsymbol{x})| dx}_{<\infty} \underbrace{\int f(\boldsymbol{y}) d\boldsymbol{y}}_{=1} = O(h_n q_n)$$

and similarly the term in (A.48) can be bounded by

$$h_n^2 \sum_{|\boldsymbol{i}|=2} \frac{1}{\boldsymbol{i}!} \underbrace{\sup_{\boldsymbol{z} \in [-c_n - 2h_n C, c_n + 2h_n C]} |D^{\boldsymbol{i}} m(\boldsymbol{z})|}_{\overset{\textbf{(F2)}}{=} O(q_n)} \underbrace{\int |\boldsymbol{x}|^{\boldsymbol{i}} |K(\boldsymbol{x})| dx}_{<\infty} \underbrace{\int f(\boldsymbol{y}) d\boldsymbol{y}}_{=1} = O(h_n^2 q_n).$$

Both bounds converge to zero as n tends to infinity. Hence, all above integrals are finite and p_n, P_n, as well as $p_{n,i}$ and $P_{n,i}$ are well defined for all $i = 1, \ldots, d$ and for all fixed $n \in \mathbb{N}$. Additionally, this implies that there exists a constant $\bar{P} < \infty$ (independent of $n \in \mathbb{N}$), such that $P_{n,i}(\infty) < \bar{P}$ for all $i = 1, \ldots, d$ and $n \in \mathbb{N}$. Furthermore, $P_{n,i}$ is continuous and it holds that $P_{n,i}(-\infty) = P_n(-\boldsymbol{\infty}) = 0$ for all $i = 1, \ldots, d$ and $n \in \mathbb{N}$. Hence for all $i = 1, \ldots, d$ and $\varepsilon > 0$, some $N_i \in \mathbb{N}$ and a partition

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

of $\mathbb{R}$ can be chosen, such that for all $n \in \mathbb{N}$

$$P_{n,i}(z_{j_i,i}) - P_{n,i}(z_{j_i-1,i}) \le \frac{\varepsilon}{d}, \ \forall\, j_i = 1, \ldots, N_i, \ i = 1, \ldots, d, \tag{A.49}$$

and where N_i can be chosen to be smaller than $2d\bar{P}\varepsilon^{-1}$ 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 $\boldsymbol{j-1} := (j_1 - 1, \ldots, j_d - 1)$. Then for all $\boldsymbol{z} \in \mathbb{R}^d$ a $\boldsymbol{j} \in \times_{i=1}^d \{1, \ldots, N_i\}$ can be found such that $\boldsymbol{z} \in (\boldsymbol{z_{j-1}}, \boldsymbol{z_j}]$. Let for $\boldsymbol{z} \in (\boldsymbol{z_{j-1}}, \boldsymbol{z_j}]$

$$\varphi(\boldsymbol{x}) := \int_{(-\boldsymbol{\infty},\boldsymbol{z}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y}$$

and define

$$\begin{aligned}\varphi_{\boldsymbol{j}}^u(\boldsymbol{x}) :=& \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \ge 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\ &+ \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) < 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y}\end{aligned}$$

and

$$\begin{aligned}\varphi_{\boldsymbol{j}}^l(\boldsymbol{x}) :=& \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \ge 0\} \omega_n(\boldsymbol{y})\} d\boldsymbol{y} \\ &+ \int\limits_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} (m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x})\, I\{(m(\boldsymbol{y}) - m(\boldsymbol{x})) K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) < 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y}.\end{aligned}$$

Then it holds that $\varphi_{\boldsymbol{j}}^l \le \varphi \le \varphi_{\boldsymbol{j}}^u$. Furthermore, it holds that

$$\begin{aligned}\left|\varphi_{\boldsymbol{j}}^u(\boldsymbol{x}) - \varphi_{\boldsymbol{j}}^l(\boldsymbol{x})\right| = \Bigg| & \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y} - \boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\ & - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y} - \boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y} \Bigg|\end{aligned}$$

$$= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y}$$
$$- \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y}$$

and therefore

$$\begin{aligned}
E\left[\left|\varphi_{\boldsymbol{j}}^u(\boldsymbol{X}_i) - \varphi_{\boldsymbol{j}}^l(\boldsymbol{X}_i)\right|\right] &= \int \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y} f(\boldsymbol{x}) d\boldsymbol{x} \\
&\quad - \int \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) d\boldsymbol{y} f(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \int |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) f(\boldsymbol{x}) d\boldsymbol{x} d\boldsymbol{y} \\
&\quad - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \int |m(\boldsymbol{y}) - m(\boldsymbol{x})| \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right| \omega_n(\boldsymbol{y}) f(\boldsymbol{x}) d\boldsymbol{x} d\boldsymbol{y} \\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} p_n(\boldsymbol{y}) d\boldsymbol{y} - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} p_n(\boldsymbol{y}) d\boldsymbol{y} \\
&= P_n(\boldsymbol{z_j}) - P_n(\boldsymbol{z_{j-1}}) \\
&\overset{(*)}{\leq} \sum_{i=1}^{d} (P_{n,i}(z_{j_i,i}) - P_{n,i}(z_{j_i-1,i})) \\
&\overset{\text{(A.49)}}{\leq} d\frac{\varepsilon}{d} = \varepsilon,
\end{aligned}$$

where the inequality in $(*)$ holds by similar calculations as in the proof of Theorem 3.2 on page 44. As $N_i = O(\varepsilon^{-1})$ for all $i = 1, \ldots, d$, it holds that

$$N_{[\,]}\left(\varepsilon, \mathcal{F}_{n,1}, \|\cdot\|_{L_1(P)}\right) \leq \left| \bigtimes_{i=1}^{d} \{1, \ldots, N_i\} \right| = \prod_{i=1}^{d} N_i = O(\varepsilon^{-d}).$$

□

Lemma A.7. *For the bracketing number* $N_{[\,]}\left(\varepsilon, \mathcal{F}_{n,2}, \|\cdot\|_{L_1(P)}\right)$ *where* $(U_t, \boldsymbol{X}_t) \sim P$ *and*

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

defined on page 134 in the proof of Lemma A.3, it holds that

$$N_{[\,]}\left(\varepsilon, \mathcal{F}_{n,2}, \|\cdot\|_{L_1(P)}\right) = O\left(\varepsilon^{-d}\right).$$

Proof. First some notations are necessary. Let

$$q_n : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{x}-\boldsymbol{y}}{h_n}\right)\right| \omega_n(\boldsymbol{x})\sigma(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y},$$

where $\sigma(\boldsymbol{y})$ is the positive square root of $\sigma^2(\boldsymbol{y})$ for $\boldsymbol{y} \in \mathbb{R}^d$,

$$Q_n : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty},\boldsymbol{x}]} q_n(\boldsymbol{t})d\boldsymbol{t} \tag{A.50}$$

and for all $i = 1, \ldots, d$ let

$$\begin{aligned} q_{n,i} &: \mathbb{R} \to \mathbb{R} \\ &x \mapsto \int \cdots \int q_n(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_d, \end{aligned}$$

$$Q_{n,i} : \mathbb{R} \to \mathbb{R},\ x \mapsto \int_{-\infty}^{x} q_{n,i}(t)dt.$$

Furthermore, let

$$r_n : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \sigma^2(\boldsymbol{x})f(\boldsymbol{x})\omega_n(\boldsymbol{x}),$$

$$R_n : \mathbb{R}^d \to \mathbb{R},\ \boldsymbol{x} \mapsto \int_{(-\boldsymbol{\infty},\boldsymbol{x}]} r_n(\boldsymbol{t})d\boldsymbol{t},$$

and for all $i = 1, \ldots, d$ let

$$\begin{aligned} r_{n,i} &: \mathbb{R} \to \mathbb{R} \\ &x \mapsto \int \cdots \int r_n(x_1, \ldots, x_{i-1}, x, x_{i+1}, \ldots, x_d)dx_1 \ldots dx_{i-1}dx_{i+1} \ldots dx_d, \end{aligned}$$

$$R_{n,i} : \mathbb{R} \to \mathbb{R},\ x \mapsto \int_{-\infty}^{x} r_{n,i}(t)dt.$$

Note that for all $n \in \mathbb{N}$ and $i = 1, \ldots, d$, $Q_{n,i}$ and $R_{n,i}$ are continuous. Additionally, it holds that $Q_{n,i}(-\infty) = Q_n(-\boldsymbol{\infty}) = 0$ and $R_{n,i}(-\infty) = R_n(-\boldsymbol{\infty}) = 0$, as well as

$$\begin{aligned} Q_{n,i}(\infty) &= Q_n(\boldsymbol{\infty}) \\ &= \int \int \frac{1}{h_n^d} \left| K\left(\frac{\boldsymbol{t}-\boldsymbol{y}}{h_n}\right)\right| \omega_n(\boldsymbol{t})\sigma(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}d\boldsymbol{t} \\ &\le \int \int |K(\boldsymbol{x})| \, \sigma(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}d\boldsymbol{x} \\ &= \underbrace{\int |K(\boldsymbol{x})| \, d\boldsymbol{x}}_{<\infty} \underbrace{\int \sigma(\boldsymbol{y})f(\boldsymbol{y})d\boldsymbol{y}}_{<\infty} \\ &< \bar{Q}, \end{aligned}$$

and

$$\begin{aligned}
R_{n,i}(\infty) &= R_n(\boldsymbol{\infty}) \\
&= \int \sigma(\boldsymbol{t}) f(\boldsymbol{t}) \omega_n(\boldsymbol{t}) d\boldsymbol{t} \\
&\leq \int \sigma(\boldsymbol{t}) f(\boldsymbol{t}) d\boldsymbol{t} \\
&< \bar{R},
\end{aligned}$$

for some $\bar{Q}, \bar{R} < \infty$ that do not depend on $n \in \mathbb{N}$. Hence, for all $i = 1, \ldots, d$ and $\varepsilon > 0$, some $N_i \in \mathbb{N}$ and a partition

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

of $\mathbb{R}$ can be chosen such that for all $n \in \mathbb{N}$

$$Q_{n,i}(z_{j_i,i}) - Q_{n,i}(z_{j_i-1,i}) \leq \frac{\varepsilon}{2d}, \ \forall \ j_i = 1, \ldots, N_i, \ i = 1, \ldots, d, \tag{A.51}$$

and

$$R_{n,i}(z_{j_i,i}) - R_{n,i}(z_{j_i-1,i}) \leq \frac{\varepsilon}{2d}, \ \forall \ j_i = 1, \ldots, N_i, \ i = 1, \ldots, d, \tag{A.52}$$

and where N_i can be chosen to be smaller than $4d \max(\bar{Q}, \bar{R})\varepsilon^{-1}$ 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 $\boldsymbol{j-1} := (j_1-1, \ldots, j_d-1)$. Then for all $\boldsymbol{z} \in \mathbb{R}^d$ there exists a $\boldsymbol{j} \in \times_{i=1}^d \{1, \ldots, N_i\}$ such that $\boldsymbol{z} \in (\boldsymbol{z_{j-1}}, \boldsymbol{z_j}]$. For $\boldsymbol{z} \in (\boldsymbol{z_{j-1}}, \boldsymbol{z_j}]$ define

$$\varphi(u, \boldsymbol{x}) := uI\left\{|u| \leq n^{\frac{1}{q}}\right\} \int_{(-\boldsymbol{\infty}, \boldsymbol{z}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \omega_n(\boldsymbol{y}) d\boldsymbol{y} - uI\left\{|u| \leq n^{\frac{1}{q}}\right\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z}\}.$$

Then it holds that

$$\begin{aligned}
\varphi(u, \boldsymbol{x}) \leq\ & uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\} \int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \geq 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\} \int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) \geq 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\} \int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) < 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\} \int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y} - \boldsymbol{x}) < 0\} \omega_n(\boldsymbol{y}) d\boldsymbol{y} \\
&- uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z_{j-1}}\} \\
&- uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\} \omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z_j}\} \\
&=: \varphi_{\boldsymbol{j}}^u(u, \boldsymbol{x}).
\end{aligned}$$

And similarly,

$$\begin{aligned}
\varphi(u,\boldsymbol{x}) \geq{}& uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\} \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \geq 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\} \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) \geq 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\} \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) < 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&+ uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\} \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) I\{K_{h_n}(\boldsymbol{y}-\boldsymbol{x}) < 0\}\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&- uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u \geq 0\}\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z_j}\}\\
&- uI\left\{|u| \leq n^{\frac{1}{q}}\right\} I\{u < 0\}\omega_n(\boldsymbol{x}) I\{\boldsymbol{x} \leq \boldsymbol{z_{j-1}}\}\\
&=: \varphi^l_{\boldsymbol{j}}(u,\boldsymbol{x}).
\end{aligned}$$

It holds that

$$\begin{aligned}
\left|\varphi^u_{\boldsymbol{j}}(u,\boldsymbol{x}) - \varphi^l_{\boldsymbol{j}}(u,\boldsymbol{x})\right| = |u| I\left\{|u| \leq n^{\frac{1}{q}}\right\} \Bigg(& \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&- \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&+ I\{\boldsymbol{x} \leq \boldsymbol{z_j}\}\omega_n(\boldsymbol{x}) - I\{\boldsymbol{x} \leq \boldsymbol{z_{j-1}}\}\omega_n(\boldsymbol{x})\Bigg)
\end{aligned}$$

and thus using $E[|U_i||\boldsymbol{X}_i] \leq \left(E\left[U_i^2|\boldsymbol{X}_i\right]\right)^{\frac{1}{2}} = \sigma(\boldsymbol{X}_i)$ a.s., it can be obtained that

$$\begin{aligned}
&E\left[\left|\varphi^u_{\boldsymbol{j}}(U_i,\boldsymbol{X}_i) - \varphi^l_{\boldsymbol{j}}(U_i,\boldsymbol{X}_i)\right|\right]\\
&\leq E\Bigg[\sigma(\boldsymbol{X}_i)\Bigg(\int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{X}_i}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&\qquad - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{X}_i}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}\\
&\qquad + I\{\boldsymbol{X}_i \leq \boldsymbol{z_j}\}\omega_n(\boldsymbol{X}_i) - I\{\boldsymbol{X}_i \leq \boldsymbol{z_{j-1}}\}\omega_n(\boldsymbol{X}_i)\Bigg)\Bigg]\\
&= \int \sigma(\boldsymbol{x}) \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}f(\boldsymbol{x})d\boldsymbol{x}\\
&- \int \sigma(\boldsymbol{x}) \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})d\boldsymbol{y}f(\boldsymbol{x})d\boldsymbol{x}\\
&+ \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \sigma(\boldsymbol{x})f(\boldsymbol{x})\omega_n(\boldsymbol{x})d\boldsymbol{x} - \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \sigma(\boldsymbol{x})f(\boldsymbol{x})\omega_n(\boldsymbol{x})d\boldsymbol{x}\\
&= \int_{(-\boldsymbol{\infty},\boldsymbol{z_j}]} \int \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})\sigma(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{y}\\
&- \int_{(-\boldsymbol{\infty},\boldsymbol{z_{j-1}}]} \int \frac{1}{h_n^d}\left|K\left(\frac{\boldsymbol{y}-\boldsymbol{x}}{h_n}\right)\right|\omega_n(\boldsymbol{y})\sigma(\boldsymbol{x})f(\boldsymbol{x})d\boldsymbol{x}d\boldsymbol{y}
\end{aligned}$$

$$
\begin{aligned}
&+ \int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} \sigma(\boldsymbol{x}) f(\boldsymbol{x}) \omega_n(\boldsymbol{x}) d\boldsymbol{x} - \int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} \sigma(\boldsymbol{x}) f(\boldsymbol{x}) \omega_n(\boldsymbol{x}) d\boldsymbol{x} \\
&= \int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} q_n(\boldsymbol{y}) d\boldsymbol{y} - \int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} q_n(\boldsymbol{y}) d\boldsymbol{y} \\
&+ \int_{(-\boldsymbol{\infty}, \boldsymbol{z_j}]} r_n(\boldsymbol{y}) d\boldsymbol{y} - \int_{(-\boldsymbol{\infty}, \boldsymbol{z_{j-1}}]} r_n(\boldsymbol{y}) d\boldsymbol{y} \\
&= Q_n(\boldsymbol{z_j}) - Q_n(\boldsymbol{z_{j-1}}) + R_n(\boldsymbol{z_j}) - R_n(\boldsymbol{z_{j-1}}) \\
&\overset{(*)}{\leq} \sum_{i=1}^{d} \underbrace{(Q_{n,i}(z_{j_i,i}) - Q_{n,i}(z_{j_i-1,i}))}_{\overset{(A.51)}{\leq} \frac{\varepsilon}{2d}} + \sum_{i=1}^{d} \underbrace{(R_{n,i}(z_{j_i,i}) - R_{n,i}(z_{j_i-1,i}))}_{\overset{(A.52)}{\leq} \frac{\varepsilon}{2d}} \\
&\leq \varepsilon,
\end{aligned}
$$

where $(*)$ holds due to similar calculations as in the proof of Theorem 3.2 on page 44. As $N_i = O(\varepsilon^{-1})$ for all $i = 1, \ldots, d$, it holds that

$$
N_{[\,]}(\varepsilon, \mathcal{F}_{n,2}, \|\cdot\|_{L_1(P)}) \leq \left| \bigtimes_{i=1}^{d} \{1, \ldots, N_i\} \right| = \prod_{i=1}^{d} N_i = O(\varepsilon^{-d}).
$$

□

B A weak convergence result for sequential empirical processes under weak dependence

The purpose of this Chapter is to prove a weak convergence result for empirical processes indexed in general classes of functions and with an underlying α-mixing triangular array process. In particular, the uniformly boundedness assumption on the function class, which is required in most of the existing literature, is spared. Furthermore, under strict stationarity a weak convergence result for the sequential empirical process indexed in function classes is obtained, which is directly applicable to the process T_n from Chapter 3. A short literature review on weak convergence of empirical processes will be followed by the main results and their proofs.

B.1 Literature review

The asymptotic behavior of empirical processes has been studied for decades. Inspired by the study of the empirical distribution function, more generally empirical processes indexed in function classes gained a lot of attention. In particular, central limit results, i.e. weak convergence of the sequence of the stochastic processes to a Gaussian process, are of interest. Such results are sometimes referred to as a uniform central limit theorem (CLT) for the empirical process indexed in function classes and as a uniform functional central limit theorem (FCLT) for the partial sum process indexed in function classes, also referred to as the sequential empirical process indexed in function classes.

The most simple case is given if the underlying process is a family of i.i.d. random variables. In this situation many results are available. Ossiander [57] showed a uniform CLT under a metric entropy condition on the function class. The uniform FCLT follows directly by the non-functional one. For example van der Vaart and Wellner [75] state this result in Section 2.12 of their book and also give a great overview on empirical processes for the i.i.d. case in general. For dependent data much less is known. There are several results, concerning the non-sequential case. Doukhan, Massart and Rio [19] showed a uniform CLT under a metric entropy condition on the function class and β-mixing, strictly stationary data. Dedecker and Louhichi [13] generalized this result, imposing a condition on suitable maximal inequalities for the empirical process indexed in finite sets of functions. Their result is applicable to β-mixing and non-uniform ϕ-mixing sequences. Andrews and Pollard [2] showed a uniform CLT for α-mixing arrays and uniformly bounded function classes, fulfilling a metric entropy condition. Massart [51] showed a uniform CLT for uniformly bounded function classes and strictly stationary, α-mixing sequences, when the mixing coefficient decays exponentially fast. Given uniformly bounded function classes, Hariz [28] gave more general conditions in terms of bracketing numbers with respect to a norm resulting from a moment inequality satisfied by the underlying process. He particularly improves among others the results in [51] and [2]. Hansen [25] proved a uniform CLT for mixingale arrays and classes of Lipschitz-continuous functions. More recent results use alternative dependence conditions. Hagemann [24] uses an alternative short-range dependence notion, applicable to non-linear time series models, and uniformly bounded classes of functions. Dehling, Durieu and Tusche [15] showed a uniform CLT for multiple mixing and strictly stationary data, and uniformly bounded function classes. In the dependent setup the convergence of the sequential process does not follow directly by the convergence of the non-sequential one, but requires additional conditions. Dehling, Durieu and Tusche [14] extended their aforementioned uniform CLT to a functional version. Volgushev and Shao [77] established more general assumptions, under which a uniform FCLT holds for strictly stationary data. The result particularly requires a strong version of asymptotic equicontinuity for the non-sequential process.

An intensive study of the literature led to two main findings. First, most uniform central limit results for dependent data impose the condition of uniformly bounded classes of functions or strong smoothness conditions. And secondly, very few results are available regarding the uniform FCLT. The result in this chapter is a uniform CLT for empirical processes with an α-mixing triangular array process indexed by a function class, that fulfills a metric entropy condition. It is a generalization of the result of Andrews and Pollard [2] to unbounded function classes. The result particularly implies the strong version of asymptotic equicontinuity, needed in [77]. In the case of strict stationarity a uniform FCLT can therefore be obtained simultaneously.

B.2 Main results

Theorem B.1 gives conditions on the underlying triangular array process $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ and on the function class $\mathcal{F}$, under which the (non-sequential) empirical process $\{G_n(1,\varphi) : \varphi \in \mathcal{F}\}$, defined in Definition 1.4, on page 10 satisfies a strong form of asymptotic equicontinuity.

Theorem B.1 (Equicontinuity). *Let $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ be a triangular array 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 $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ be strongly mixing with mixing coefficient $\alpha(\cdot)$, such that*

$$\sum_{t=1}^{\infty} t^{Q-2} \alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty,$$

for some $\gamma > 0$ and some even $Q \geq 2$.

(A2) *For* Q *and* γ *from assumption* ***(A1)*** *and*

$$\rho(\varphi) := \sup_{n\in\mathbb{N}} \sup_{1\le t\le n} E[|\varphi(X_{n,t})|^2]^{\frac{1}{2}},$$

for all measurable functions $\varphi : \mathcal{X} \to \mathbb{R}$ *and* $\tilde{N}_{[\,]}(\cdot, \mathcal{F}, \rho)$ *from Definition 1.6 on page 11, let*

$$\int_0^1 x^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}(x, \mathcal{F}, \rho)\right)^{\frac{1}{Q}} dx < \infty.$$

Furthermore, assume that each $\varepsilon > 0$ *allows a choice of bounding class* $\mathcal{B}$*, such that for all* $i = 2, \ldots, Q$

$$\sup_{n\in\mathbb{N}} \sup_{1\le t\le n} E\left[|b(X_{n,t})|^{i\frac{2+\gamma}{2}}\right]^{\frac{1}{2}} \le \varepsilon, \ \forall\, b \in \mathcal{B}. \tag{B.1}$$

Then with $d(\varphi, \psi) := \sup_{n\in\mathbb{N}} \sup_{1\le t\le n} E\left[|\varphi(X_{n,t}) - \psi(X_{n,t})|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}\frac{2}{2+\gamma}}$*, it holds that*

$$\lim_{\delta\searrow 0} \limsup_{n\to\infty} E^*\left[\sup_{\{\varphi,\psi\in\mathcal{F}: d(\varphi,\psi)<\delta\}} |G_n(1,\varphi) - G_n(1,\psi)|^Q\right]^{\frac{1}{Q}} = 0.$$

The proof is given in Section B.3. A uniform CLT as direct consequence is obtained and stated in Corollary B.2. In contrast to most results for strongly mixing processes in the literature, it does not require uniformly bounded function classes.

Corollary B.2 (Uniform CLT)**.** *Let the assumptions of Theorem B.1 hold and let additionally for all* $K \in \mathbb{N}$ *and all finite collections* $\varphi_k \in \mathcal{F}$*,* $k = 1, \ldots, K$*,*

$$(G_n(1, \varphi_k))_{k=1,\ldots,K} \xrightarrow[n\to\infty]{\mathcal{D}} (G(1, \varphi_k))_{k=1,\ldots,K}\,,$$

where $G := \{G(1, \varphi) : \varphi \in \mathcal{F}\}$ *is a centered Gaussian process. Then*

$$\{G_n(1, \varphi) : \varphi \in \mathcal{F}\} \underset{n\to\infty}{\rightsquigarrow} G,$$

in $l^\infty(\mathcal{F})$*.*

The proof is given in Section B.3. Note that Corollary B.2 is a result for triangular arrays, which are more powerful than their analogues for a single sequence. For strictly stationary sequences a uniform FCLT can be obtained, which is stated in the following.

Corollary B.3 (Uniform FCLT)**.** *Let* $(X_t)_{t\in\mathbb{Z}}$ *be a strictly stationary sequence of random variables. Let the assumptions of Theorem B.1 be satisfied by* $X_{n,t} := X_t$*, for all* $1 \le t \le n, n \in \mathbb{N}$ *and by* $Q > 2$*. Additionally, let*

(A3) $\mathcal{F}$ *possess an envelope function (see [75], p. 84 for the definition)* F*, with* $E[|F(X_1)|^Q] < \infty$ *and let there exist a constant* $L < \infty$*, such that*

$$\sup_{\varphi\in\mathcal{F}} E\left[|\varphi(X_1)|^{Q\frac{2+\gamma}{2}}\right] \le L.$$

Then with $d(\varphi,\psi) := E\left[|\varphi(X_1)-\psi(X_1)|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}\frac{2}{2+\gamma}}$, *it holds that*

$$\forall\ \varepsilon > 0: \lim_{\delta\searrow 0}\limsup_{n\to\infty} P^*\left(\sup_{\substack{\{s,t\in[0,1],\varphi,\psi\in\mathcal{F}:\\ |s-t|+d(\varphi,\psi)<\delta\}}} |G_n(s,\varphi)-G_n(t,\psi)| > \varepsilon\right) = 0.$$

If additionally for all $K \in \mathbb{N}$ *and all finite collections* $\varphi_k \in \mathcal{F}$, $s_k \in [0,1]$, $k = 1,\ldots,K$,

$$(G_n(s_k,\varphi_k))_{k=1,\ldots,K} \xrightarrow[n\to\infty]{\mathcal{D}} (G(s_k,\varphi_k))_{k=1,\ldots,K}\,,$$

where $G := \{G(s,\varphi) : s\in[0,1],\varphi\in\mathcal{F}\}$ *is a centered Gaussian process, then it can be concluded that*

$$\{G_n(s,\varphi) : s\in[0,1],\varphi\in\mathcal{F}\} \underset{n\to\infty}{\rightsquigarrow} G,$$

in $l^\infty([0,1]\times\mathcal{F})$.

The proof is given in Section B.3. Note that due to the strict stationarity assumption, the semi norm ρ on $\mathcal{F}$ simplifies to the $L_2(P)$ norm and the semi metric d is the $L_{Q\frac{2+\gamma}{2}}(P)$ distance, where $X_1 \sim P$. Furthermore, condition (B.1) in Theorem B.1 simplifies to

$$E\left[|b(X_1)|^{i\frac{2+\gamma}{2}}\right]^{\frac{1}{2}} \le \varepsilon,\ \forall\ b\in\mathcal{B},\ \forall\ i = 2,\ldots,Q.$$

Remark. Corollary B.3 is applied to show the weak convergence of the process T_n in Theorem 3.2 on page 41 in Chapter 3. Furthermore, it is used in the proof of Lemma A.4 on page 142 in Appendix A. Certainly the uniform CLT and FCLT are of independent interest as they can be a powerful tool for proofs of asymptotic results in mathematical statistics in different situations.

B.3 Proofs

The key tool in proving Theorem B.1 is a moment inequality for $G_n(1,\varphi)$ for fixed $n\in\mathbb{N}$, which is stated in the following lemma. It is a generalization of Lemma 3.1 of Andrews and Pollard [2], who proved a moment inequality for bounded and strongly mixing random variables. Extending this result to unbounded random variables makes it possible to extend the uniform CLT to unbounded function classes. Nevertheless, it comes at the cost of moment constraints. Note that similar results are available, for example Theorem 2 on page 26 in [18] or Corollary A.0.1 on page 319 in [63].

Lemma B.4. *Let* $\{Z_{n,t} : 1\le t\le n, n\in\mathbb{N}\}$ *be a strongly mixing triangular array of random variables with values in* $\mathbb{R}$ *and with mixing coefficient* $\alpha(\cdot)$. *Let furthermore for some even* $Q \ge 2$ *and some* $\gamma > 0$, $\tau > 0$

(i) $\sum_{t=1}^{\infty} t^{Q-2}\alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty$ *and*

(ii) *$E[Z_{n,t}] = 0$, $E\left[|Z_{n,t}|^{i\frac{2+\gamma}{2}}\right] \leq \tau^{2+\gamma}$, for all $i = 2, \ldots, Q$ and $1 \leq t \leq n, n \in \mathbb{N}$.*

Then

$$E\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{n,i}\right|^{Q}\right]^{\frac{1}{Q}} \leq C \max\left(n^{-\frac{1}{2}}, \tau\right), \ \forall\, n \in \mathbb{N}, \tag{B.2}$$

for some constant C only depending on Q, γ and the mixing coefficient.

For the proof of Lemma B.4 the following covariance inequality for strongly mixing triangular arrays is used. It was stated by Sun and Chiang [72] (see Lemma 2.1) for α-mixing sequences of real valued random variables. Su and Xiao [71] extended it to α-mixing sequences of multivariate random variables (see Lemma D.1). As Su and Ullah [69] argued, the result is also valid for α-mixing triangular arrays of random variables (see Lemma A.2 in the supplement material to [69]).

Lemma B.5. *Let $\{\xi_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ be a strongly mixing triangular array of random variables with values in $\mathbb{R}^l$ and with mixing coefficient $\alpha(\cdot)$. Let $F_{n,i_1\ldots i_m}$ denote the distribution function of $(\xi_{n,i_1}, \ldots, \xi_{n,i_m})$. For some $m > 1$ and some integers $(i_1, \ldots, i_m)$ such that $1 \leq i_1 < \cdots < i_m \leq n$, let g be a Borel measurable function such that*

$$\int |g(x_1, \ldots, x_m)|^{1+\delta} dF_{n,i_1\ldots i_m}(x_1, \ldots, x_m) \leq M_n$$

and

$$\int |g(x_1, \ldots, x_m)|^{1+\delta} dF_{n,i_1\ldots i_j}(x_1, \ldots, x_j) dF_{n,i_{j+1}\ldots i_m}(x_{j+1}, \ldots, x_m) \leq M_n,$$

for some $\delta > 0$ and some $1 \leq j < m$. Then it holds that

$$\begin{aligned}&\bigg|\int g(x_1, \ldots, x_m) dF_{n,i_1\ldots i_m}(x_1, \ldots, x_m)\\ &\qquad - \int g(x_1, \ldots, x_m) dF_{n,i_1\ldots i_j}(x_1, \ldots, x_j) dF_{n,i_{j+1}\ldots i_m}(x_{j+1}, \ldots, x_m)\bigg|\\ &\leq 4M_n^{\frac{1}{1+\delta}} \alpha(i_{j+1} - i_j)^{\frac{\delta}{1+\delta}}.\end{aligned}$$

Sun and Chiang gave a proof of their version Lemma 2.1 in [72]. The proof of the generalized result Lemma B.5 works analogously and is therefore omitted.

Remark. Note that for the proof of Lemma B.4 the assertion in Lemma B.5 will only be needed for one-dimensional triangular arrays as it will be applied to $\{Z_{n,t} \in \mathbb{R} : 1 \leq t \leq n, n \in \mathbb{N}\}$. The more general version, however, is used for the proof of the consistency result in Theorem 3.4 on page 50.

Proof of Lemma B.4. The proof is closely related to the proof of Lemma 3.1 by Andrews and Pollard [2], but uses the covariance inequality for not necessarily bounded and strongly mixing random variables in Lemma B.5.

To begin with, it will be proven via induction that for all $Q \geq 2$ (not necessarily even) satisfying assumptions **(i)** and **(ii)**, there exists a constant C' only depending on Q, γ and the mixing coefficient, such that

$$\sum_{\boldsymbol{i} \in \boldsymbol{I}^Q} \left| E\left[Z_{n,i_1} \cdots Z_{n,i_Q}\right] \right| \leq C' \left(n\tau^2 + \cdots + (n\tau^2)^{\lfloor \frac{Q}{2} \rfloor} \right), \ \forall\, n \in \mathbb{N}, \tag{B.3}$$

where $\boldsymbol{I}^Q := \{\boldsymbol{i} = (i_1, \ldots, i_Q) \in \{1, \ldots, n\}^Q : i_1 \leq \cdots \leq i_Q\}$.

Let for $Q = 2$ the assumptions of Lemma B.4 hold for some $\gamma, \tau > 0$. Let furthermore $\{\tilde{Z}_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ be an independent copy of $\{Z_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$. Then applying **(ii)**

$$\begin{aligned} E\left[|Z_{n,i_1} \tilde{Z}_{n,i_2}|^{\frac{2+\gamma}{2}} \right] &\leq E\left[|Z_{n,i_1}|^{2\frac{2+\gamma}{2}} \right]^{\frac{1}{2}} E\left[|\tilde{Z}_{n,i_2}|^{2\frac{2+\gamma}{2}} \right]^{\frac{1}{2}} \\ &\leq \tau^{2+\gamma} \end{aligned}$$

and similarly

$$E\left[|Z_{n,i_1} Z_{n,i_2}|^{\frac{2+\gamma}{2}} \right] \leq \tau^{2+\gamma}$$

holds for all $i_1, i_2 \in \{1, \ldots, n\}$ with $i_1 \neq i_2$ and $n \in \mathbb{N}$. Lemma B.5 can therefore be applied with $g(x_1, x_2) := x_1 x_2$, $\delta := \frac{\gamma}{2}$ and $M_n := \tau^{2+\gamma}$ for all $n \in \mathbb{N}$. It implies that

$$\begin{aligned} \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \left| E\left[Z_{n,i_1} Z_{n,i_2}\right] \right| &= \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \left| E\left[Z_{n,i_1} Z_{n,i_2}\right] - E\left[Z_{i_1}\right] E\left[Z_{i_2}\right] \right| \\ &\leq \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \alpha(i_2 - i_1)^{\frac{\gamma}{2+\gamma}} 4 \left(\tau^{2+\gamma}\right)^{\frac{2}{2+\gamma}} \\ &\leq n\tau^2 4 \sum_{t=1}^{\infty} \alpha(t)^{\frac{\gamma}{2+\gamma}} \\ &= C'' n\tau^2, \end{aligned}$$

for $C'' := 4 \sum_{t=1}^{\infty} \alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty$ by assumption **(i)**, a constant therefore only depending on γ and the mixing coefficient. Furthermore, this and

$$E\left[|Z_{n,i_1}|^2 \right] \leq E\left[|Z_{n,i_1}|^{2\frac{2+\gamma}{2}} \right]^{\frac{2}{2+\gamma}} \leq \left(\tau^{2+\gamma}\right)^{\frac{2}{2+\gamma}} = \tau^2,$$

for all $i_1 \in \{1, \ldots, n\}$ and $n \in \mathbb{N}$ by assumption **(ii)**, leads to

$$\begin{aligned} \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1 \leq i_2}}^{n} \left| E\left[Z_{n,i_1} Z_{n,i_2}\right] \right| &= \sum_{i_1=1}^{n} \left| E\left[Z_{n,i_1}^2\right] \right| + \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_1<i_2}}^{n} \left| E\left[Z_{n,i_1} Z_{n,i_2}\right] \right| \\ &\leq n\tau^2 + C'' n\tau^2 \\ &= C' n\tau^2, \end{aligned}$$

with $C' := (1 + C'')$, which is the assertion of (B.3) for $Q = 2$. For the inductive step, let now $Q > 2$ be arbitrary, but fixed and let the assertion in (B.3) be true for all integers in $\{2, \ldots, Q-1\}$. Then it is to show, that (B.3) holds for Q as well. Let therefore assumptions **(i)** and **(ii)** be satisfied for this arbitrary, but fixed $Q > 2$ and for some $\gamma > 0$ and $\tau > 0$. Note that then the assumptions are satisfied for all integers in $\{2, \ldots, Q-1\}$ as well, such that the inductive hypothesis can be applied.

The idea of the proof is the following. First, the smallest index where the gap between two succeeding indices is largest and positive (to exclude the case, where all indices are equal) is identified. The random variables after this time point will then be replaced by independent copies of themselves. For the new term, the induction hypothesis can be used as there will be less than Q indices left. The remainder term can be bounded using Lemma B.5. Following the notation of Andrews and Pollard [2], let for all $\boldsymbol{i} \in \boldsymbol{I}^Q$

$$G(\boldsymbol{i}) := \max\{(i_{j+1} - i_j) : (i_{j+1} - i_j) > 0, 1 \leq j \leq Q-1\}$$

and

$$m(\boldsymbol{i}) := \min\{j \in \{1, \ldots, Q-1\} : (i_{j+1} - i_j) = G(\boldsymbol{i})\}.$$

Then it can be obtained that

$$\begin{aligned}
&\sum_{\boldsymbol{i} \in \boldsymbol{I}^Q} \left| E\left[Z_{n,i_1} \cdots Z_{n,i_Q}\right]\right| \\
&= \sum_{\substack{\boldsymbol{i} \in \boldsymbol{I}^Q \\ i_1 = \cdots = i_Q}} \left|E[Z_{n,i_1} \cdots Z_{n,i_Q}]\right| + \sum_{m=1}^{Q-1} \sum_{\substack{\boldsymbol{i} \in \boldsymbol{I}^Q \\ m(\boldsymbol{i}) = m}} \left|E\left[Z_{n,i_1} \cdots Z_{n,i_Q}\right]\right| \\
&\leq \sum_{i_1=1}^{n} \left|E\left[Z_{n,i_1}^Q\right]\right| && \text{(B.4)} \\
&+ \sum_{m=1}^{Q-1} \sum_{\substack{\boldsymbol{i} \in \boldsymbol{I}^Q \\ m(\boldsymbol{i}) = m}} \left|E\left[Z_{n,i_1} \cdots Z_{n,i_Q}\right] - E\left[Z_{n,i_1} \cdots Z_{n,i_m}\right] E\left[Z_{n,i_{m+1}} \cdots Z_{n,i_Q}\right]\right| && \text{(B.5)} \\
&+ \sum_{m=1}^{Q-1} \sum_{\substack{\boldsymbol{i} \in \boldsymbol{I}^Q \\ m(\boldsymbol{i}) = m}} \left|E\left[Z_{n,i_1} \cdots Z_{n,i_m}\right] E\left[Z_{n,i_{m+1}} \cdots Z_{n,i_Q}\right]\right|. && \text{(B.6)}
\end{aligned}$$

Let first (B.4) be considered. Using assumption **(ii)**, it holds that

$$\sum_{i_1=1}^{n} \left|E\left[Z_{n,i_1}^Q\right]\right| \leq \sum_{i_1=1}^{n} E\left[|Z_{n,i_1}|^{Q\frac{2+\gamma}{2}}\right]^{\frac{2}{2+\gamma}} \leq \sum_{i_1=1}^{n} \left(\tau^{2+\gamma}\right)^{\frac{2}{2+\gamma}} = n\tau^2.$$

Let next (B.5) be considered. Using Hölder's inequality and assumption **(ii)**,

$$\begin{aligned}
&E\left[|Z_{n,i_1} \cdots Z_{n,i_k} \tilde{Z}_{n,i_{k+1}} \cdots \tilde{Z}_{n,i_Q}|^{\frac{2+\gamma}{2}}\right] \\
&\leq E\left[|Z_{n,i_1}|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}} \cdots E\left[|Z_{n,i_k}|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}} E\left[|\tilde{Z}_{n,i_{k+1}}|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}} \cdots E\left[|\tilde{Z}_{n,i_Q}|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}} \\
&\leq \tau^{2+\gamma}
\end{aligned}$$

and similarly

$$E\left[|Z_{n,i_1} \cdots Z_{n,i_Q}|^{\frac{2+\gamma}{2}}\right] \leq \tau^{2+\gamma}$$

holds, for all $k \in \{1, \dots, Q-1\}$. Hence, applying Lemma B.5 with $g(x_1, \dots, x_Q) := x_1 \cdots x_Q$, $\delta := \frac{\gamma}{2}$ and $M_n := \tau^{2+\gamma}$ for all $n \in \mathbb{N}$, implies for all $k \in \{1, \dots, Q-1\}$,

$$\left| E\left[Z_{n,i_1} \cdots Z_{n,i_Q}\right] - E\left[Z_{n,i_1} \cdots Z_{i_k}\right] E\left[Z_{n,i_{k+1}} \cdots Z_{n,i_Q}\right] \right| \leq 4\tau^2 \alpha(i_{k+1} - i_k)^{\frac{\gamma}{2+\gamma}}.$$

Using this and distinguishing the indices furthermore by location of the gap $l \in \{1, \dots, n\}$ and size of the gap $g \in \{1, \dots, n\}$, (B.5) can be bounded by

$$4\tau^2 \sum_{m=1}^{Q-1} \sum_{\substack{\boldsymbol{i} \in \boldsymbol{I}^Q \\ m(\boldsymbol{i})=m}} \alpha(i_{m+1} - i_m)^{\frac{\gamma}{2+\gamma}} = 4\tau^2 \sum_{m=1}^{Q-1} \sum_{l=1}^{n} \sum_{g=1}^{n} \sum_{\boldsymbol{i} \in \boldsymbol{I}^Q_{m,g,l}} \alpha(g)^{\frac{\gamma}{2+\gamma}}, \tag{B.7}$$

where $\boldsymbol{I}^Q_{m,g,l} := \{\boldsymbol{i} \in \boldsymbol{I}^Q : m(\boldsymbol{i}) = m, G(\boldsymbol{i}) = g, i_m = l\}$. A more detailed study of the set of indices $\boldsymbol{I}^Q_{m,g,l}$ will lead to a suitable bound for (B.7). For fixed m, l, g, it will be investigated, how many elements the set $\boldsymbol{I}^Q_{m,g,l}$ contains at most. For all $\boldsymbol{i} = (i_1, \dots, i_Q) \in \boldsymbol{I}^Q_{m,g,l}$ the first $m-1$ indices $i_1, \dots, i_{m-1}$ satisfy

- $1 \leq i_1 \leq \cdots \leq i_{m-1} \leq i_m = l$ and
- $i_{j+1} - i_j < g$ for all $j = 1, \dots, m-1$.

With these restrictions, for fixed $i_2, \dots, i_m$, the sum over i_1 ranges over at most g elements. For fixed $i_3, \dots, i_m$ the sum over i_2, again ranges over at most g elements, with the above restrictions. Continuing in this way, there are at most g^{m-1} choices for the first $m-1$ indices $i_1, \dots, i_{m-1}$. Because of $m(\boldsymbol{i}) = m, i_m = l$ and $G(\boldsymbol{i}) = g$, it holds that $i_{m+1} = l + g$ and therefore the last $Q-m-1$ indices $i_{m+2}, \dots, i_Q$ satisfy the following restrictions

- $l + g = i_{m+1} \leq i_{m+2} \leq \cdots \leq i_Q \leq n$ and
- $i_{j+1} - i_j < g + 1$ for all $j = m+1, \dots, Q$.

Hence, following the same arguments as above, there are at most $(g+1)^{Q-m-1}$ choices for the last $Q-m-1$ indices $i_{m+2}, \dots, i_Q$. Therefore, (B.7) can further be bounded by

$$\begin{aligned}
&4\tau^2 \sum_{m=1}^{Q-1} \sum_{l=1}^{n} \sum_{g=1}^{n} g^{m-1}(g+1)^{Q-m-1} \alpha(g)^{\frac{\gamma}{2+\gamma}} \\
&\leq 4\tau^2 (Q-1) n \sum_{g=1}^{\infty} (g+1)^{Q-2} \alpha(g)^{\frac{\gamma}{2+\gamma}} \\
&= C'' n \tau^2,
\end{aligned}$$

for $C'' := 4(Q-1) \sum_{t=1}^{\infty} (t+1)^{Q-2} \alpha(t)^{\frac{\gamma}{2+\gamma}} < \infty$ by assumption **(i)**, a constant only depending on Q, γ and the mixing coefficient.

It is left to consider (B.6). Introducing the following notation

$$B_n(i) := n\tau^2 + \left(n\tau^2\right)^2 + \cdots + \left(n\tau^2\right)^{\lfloor \frac{i}{2} \rfloor} \ \forall\, i = 1, \dots, Q-1, \ \forall\, n \in \mathbb{N}$$

and applying the induction hypothesis, it holds that

$$\begin{aligned}
&\sum_{m=1}^{Q-1} \sum_{\substack{\boldsymbol{i}\in \boldsymbol{I}^Q \\ m(\boldsymbol{i})=m}} \left|E\left[Z_{n,i_1}\cdots Z_{n,i_m}\right] E\left[Z_{n,i_{m+1}}\cdots Z_{n,i_Q}\right]\right| \\
&\leq \sum_{m=1}^{Q-1} \sum_{\boldsymbol{i}\in \boldsymbol{I}^m} \left|E\left[Z_{n,i_1}\cdots Z_{n,i_m}\right]\right| \sum_{\boldsymbol{i}\in \boldsymbol{I}^{Q-m}} \left|E\left[Z_{n,i_{m+1}}\cdots Z_{n,i_Q}\right]\right| \\
&\leq \sum_{m=1}^{Q-1} C_m B_n(m) C_{Q-m} B_n(Q-m),
\end{aligned}$$

for some constants C_i only depending on i, γ and the mixing coefficient for all $i = 1, \dots, Q-1$. As $B_n(m)B_n(Q-m)$ is a polynomial in $n\tau^2$ of degree

$$\left\lfloor \frac{m}{2} \right\rfloor + \left\lfloor \frac{Q-m}{2} \right\rfloor \leq \left\lfloor \frac{Q}{2} \right\rfloor,$$

there exists a constant $C_{m,Q}$, such that $B_n(m)B_n(Q-m) \leq C_{m,Q}B_n(Q)$. Hence, the above sum can be bounded by

$$\sum_{m=1}^{Q-1} C_m C_{Q-m} C_{m,Q} B_n(Q) = C''' B_n(Q),$$

for $C''' := \sum_{m=1}^{Q-1} C_m C_{Q-m} C_{m,Q} < \infty$. Putting the results for (B.4), (B.5) and (B.6) together, it can be obtained that

$$\begin{aligned}
\sum_{\boldsymbol{i}\in \boldsymbol{I}^Q} \left|E\left[Z_{n,i_1}\cdots Z_{n,i_Q}\right]\right| &\leq n\tau^2 + C'' n\tau^2 + C''' B_n(Q) \\
&\leq C'\left(n\tau^2 + \cdots + (n\tau^2)^{\lfloor \frac{Q}{2} \rfloor}\right),
\end{aligned}$$

for $C' = 1 + C'' + C'''$ only depending on Q, γ and the mixing coefficient, which completes the induction and therefore the proof of (B.3) for all $Q \geq 2$ satisfying the assumptions.

Using

$$(n\tau^2)^i \leq \max\left(1, (n\tau^2)^{\lfloor \frac{Q}{2} \rfloor}\right) \;\forall\; 1 \leq i \leq \left\lfloor \frac{Q}{2} \right\rfloor, \tag{B.8}$$

it can be obtained that for Q being even

$$\begin{aligned}
E\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} Z_{n,i}\right|^Q\right]^{\frac{1}{Q}} &= n^{-\frac{1}{2}} E\left[\sum_{i_1=1}^{n}\cdots\sum_{i_Q=1}^{n} Z_{n,i_1}\cdots Z_{n,i_Q}\right]^{\frac{1}{Q}} \\
&\leq n^{-\frac{1}{2}} (Q!)^{\frac{1}{Q}} \left(\sum_{\boldsymbol{i}\in \boldsymbol{I}^Q} \left|E\left[Z_{n,i_1}\cdots Z_{n,i_Q}\right]\right|\right)^{\frac{1}{Q}} \\
&\overset{(B.3)}{\leq} n^{-\frac{1}{2}} (Q!)^{\frac{1}{Q}} C'^{\frac{1}{Q}} \left((n\tau^2) + \cdots + (n\tau^2)^{\frac{Q}{2}}\right)^{\frac{1}{Q}}
\end{aligned}$$

$$\overset{(B.8)}{\leq} n^{-\frac{1}{2}}(Q!)^{\frac{1}{Q}} C'^{\frac{1}{Q}} \left(\frac{Q}{2}\right)^{\frac{1}{Q}} \left(\max\left(1, (n\tau^2)^{\frac{Q}{2}}\right)\right)^{\frac{1}{Q}}$$
$$= C \max\left(n^{-\frac{1}{2}}, \tau\right),$$

for C' from inequality (B.3) and $C := \left(C'Q!\frac{Q}{2}\right)^{\frac{1}{Q}}$ only depending on Q, γ and the mixing coefficient, which proves the assertion of Lemma B.4.

□

Within the proof of Theorem B.1, let the following simplifying notation hold. Denote $G_n(\varphi) := G_n(1, \varphi)$ for measurable functions $\varphi : \mathcal{X} \to \mathbb{R}$ and

$$\{G_n(\varphi) : \varphi \in \mathcal{F}\} := \{G_n(1, \varphi) : \varphi \in \mathcal{F}\}.$$

Proof of Theorem B.1. The proof is closely related to the proof of Theorem 2.2 of Andrews and Pollard [2]. It will be shown that for all $\epsilon > 0$, there exists a $\delta = \delta(\epsilon) > 0$ and an $n_0 = n_0(\epsilon)$, such that for all $n \geq n_0$,

$$E^* \left[\sup_{\{\varphi, \psi \in \mathcal{F} : d(\varphi, \psi) < \delta\}} |G_n(\varphi) - G_n(\psi)|^Q \right]^{\frac{1}{Q}} < \epsilon. \tag{B.9}$$

Let therefore be $\epsilon > 0$. Let for $k \in \mathbb{N}$, $\delta_k := 2^{-k}$, $\tau_k := \delta_k^{\frac{2}{2+\gamma}}$ and $N_k := \tilde{N}_{[\ ]}(\delta_k, \mathcal{F}, \rho)$ and let $\mathcal{A}_k$ be the approximating class and $\mathcal{B}_k$ the bounding class from Definition 1.6, that are chosen such that assumption (B.1) in **(A2)** holds. In particular, it holds that for all $\varphi \in \mathcal{F}$, there exist an $a_k^* \in \mathcal{A}_k$ and a $b_k^* \in \mathcal{B}_k$, such that

$$|\varphi - a_k^*| \leq b_k^*, \tag{B.10}$$

and for all $b \in \mathcal{B}_k$

$$\sup_{n \in \mathbb{N}} \sup_{1 \leq t \leq n} E\left[|b(X_{n,t})|^2\right]^{\frac{1}{2}} \leq \delta_k, \ \sup_{n \in \mathbb{N}} \sup_{1 \leq t \leq n} E\left[|b(X_{n,t})|^{i\frac{2+\gamma}{2}}\right]^{\frac{1}{2}} \leq \delta_k, \ \forall\, i = 2, \ldots, Q. \tag{B.11}$$

The proof splits into two steps. First it will be shown that there exist an $m = m(\epsilon)$ and for each $\varphi \in \mathcal{F}$ a function $a_m^{(\varphi)} \in \mathcal{A}_m$ and an $n_1 = n_1(\epsilon)$, such that for all $n \geq n_1$,

$$E^* \left[\sup_{\varphi \in \mathcal{F}} \left| G_n(\varphi) - G_n\left(a_m^{(\varphi)}\right) \right|^Q \right]^{\frac{1}{Q}} < \frac{\epsilon}{8}. \tag{B.12}$$

Note that $a_m^{(\varphi)}$ is not necessarily the corresponding approximating function, denoted by $a_m^* \in \mathcal{A}_m$, from Definition 1.6 on page 11, but rather results from a constructive argument such that (B.12) holds.

Secondly, for this fixed $m \in \mathbb{N}$, $\mathcal{F}$ will be partitioned into N_m many classes, each class containing all functions φ in $\mathcal{F}$, that lead to the same $a_m^{(\varphi)} \in \mathcal{A}_m$ in step one. Within each class inequality (B.12) will be applied. By a right choice of functions the gap between two different classes can also be bridged suitably.

Step 1: The proof of (B.12) is again divided into two parts. First, a sequence $k(n) \to \infty$ and an $n_2 = n_2(\epsilon)$ are chosen, such that for all $n \geq n_2$,

$$E^* \left[\sup_{\varphi \in \mathcal{F}} \left| G_n(\varphi) - G_n\left(a^*_{k(n)}\right) \right|^Q \right]^{\frac{1}{Q}} < \frac{\epsilon}{16}, \tag{B.13}$$

where for each $\varphi \in \mathcal{F}$ and $k(n) \in \mathbb{N}$, $a^*_{k(n)}$ is the corresponding approximating function in $\mathcal{A}_{k(n)}$, as in (B.10) for $k = k(n)$.

Secondly, $m = m(\epsilon)$ and for each $\varphi \in \mathcal{F}$, $a_m^{(\varphi)} \in \mathcal{A}_m$ and $n_3 = n_3(\epsilon)$ are chosen, such that for all $n \geq n_3$ with $k(n) > m$,

$$E^* \left[\sup_{\varphi \in \mathcal{F}} \left| G_n\left(a^*_{k(n)}\right) - G_n\left(a_m^{(\varphi)}\right) \right|^Q \right]^{\frac{1}{Q}} < \frac{\epsilon}{16}. \tag{B.14}$$

Here, for each $\varphi \in \mathcal{F}$, $a^*_{k(n)}$ is the corresponding approximating function in $\mathcal{A}_{k(n)}$ from (B.10), while $a_m^{(\varphi)} \in \mathcal{A}_m$ not necessarily is. It rather results from an iterative choice of functions $a_k \in \mathcal{A}_k$ to $a_{k-1} \in \mathcal{A}_{k-1}$ for $k = k(n), \ldots, m+1$. The choice of $a_m := a_m^{(\varphi)}$ then depends on φ and n, as it is the last link in the chain, that starts with $a_{k(n)} := a^*_{k(n)}$ (which depends on φ by Definition 1.6, despite the fact, that this is not reflected in the notation). Nevertheless, the choice of m does only depend on ϵ eventually. Both (B.13) and (B.14) together imply (B.12) by choosing $n_1 = \max(n_2, n_3)$.

Proof of (B.13): Let $k(n)$ be the largest value of $k \in \mathbb{N}$, such that

$$2^{-k\frac{2}{2+\gamma}} = \tau_k \geq n^{-\frac{1}{2}}. \tag{B.15}$$

Note that then

$$\sqrt{n}\tau_{k(n)+1}^{\frac{2+\gamma}{2}} \leq \sqrt{n}\left(n^{-\frac{1}{2}}\right)^{\frac{2+\gamma}{2}} = n^{-\frac{\gamma}{4}} \overset{n\to\infty}{\longrightarrow} 0$$

holds. It follows that

$$\sqrt{n} \sup_{m \in \mathbb{N}} \sup_{1 \leq t \leq m} E\left[|b(X_{m,t})|^2\right]^{\frac{1}{2}} \overset{(B.11)}{\leq} \sqrt{n}\delta_{k(n)} = \sqrt{n}2\delta_{k(n)+1} = \sqrt{n}2\tau_{k(n)+1}^{\frac{2+\gamma}{2}} = o(1),$$

for all $b \in \mathcal{B}_{k(n)}$. Thus, there exists an $n_2' = n_2'(\epsilon)$, such that

$$2\sqrt{n} \sup_{m \in \mathbb{N}} \sup_{1 \leq t \leq m} E\left[|b(X_{m,t})|^2\right]^{\frac{1}{2}} < \frac{\epsilon}{32}, \tag{B.16}$$

for all $b \in \mathcal{B}_{k(n)}$ and $n \geq n_2'$. Hence, for $\varphi \in \mathcal{F}$ and corresponding approximation function $a^*_{k(n)} \in \mathcal{A}_{k(n)}$, applying (B.10), it holds that

$$\left| G_n(\varphi) - G_n\left(a^*_{k(n)}\right) \right| = \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(\varphi(X_{n,i}) - a^*_{k(n)}(X_{n,i}) - E\left[\varphi(X_{n,i}) - a^*_{k(n)}(X_{n,i})\right] \right) \right|$$

$$
\begin{aligned}
&\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(|\varphi(X_{n,i}) - a^*_{k(n)}(X_{n,i})| + E\left[|\varphi(X_{n,i}) - a^*_{k(n)}(X_{n,i})| \right] \right) \\
&\leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} b^*_{k(n)}(X_{n,i}) + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left[b^*_{k(n)}(X_{n,i}) \right] \\
&= G_n\left(b^*_{k(n)}\right) + 2\frac{1}{\sqrt{n}} \sum_{i=1}^{n} E\left[b^*_{k(n)}(X_{n,i}) \right] \\
&\leq G_n\left(b^*_{k(n)}\right) + 2\sqrt{n} \sup_{m\in\mathbb{N}} \sup_{1\leq t\leq m} E\left[|b^*_{k(n)}(X_{m,t})|^2 \right]^{\frac{1}{2}} \\
&\overset{(B.16)}{<} G_n\left(b^*_{k(n)}\right) + \frac{\epsilon}{32},
\end{aligned} \tag{B.17}
$$

for all $n \geq n_2'$. Next, the moment inequality from Lemma B.4 will be applied to $G_n(b)$ for $b \in \mathcal{B}_{k(n)}$. To do that, set for fixed $n \in \mathbb{N}$ and $b \in \mathcal{B}_{k(n)}$,

$$
Z_{m,t} := b(X_{m,t}) - E[b(X_{m,t})], \ \forall\ 1 \leq t \leq m, m \in \mathbb{N}.
$$

Then assumption **(ii)** of Lemma B.4 is satisfied as for all $1 \leq t \leq m, m \in \mathbb{N}$ and $i = 2, \ldots, Q$, it holds that $E[Z_{m,t}] = 0$ and

$$
\begin{aligned}
E\left[|Z_{m,t}|^{i\frac{2+\gamma}{2}} \right] &= E\left[|b(X_{m,t}) - E\left[b(X_{m,t})\right]|^{i\frac{2+\gamma}{2}} \right] \\
&\overset{(*)}{\leq} 2^{i\frac{2+\gamma}{2}} E\left[|b(X_{m,t})|^{i\frac{2+\gamma}{2}} \right] \\
&\leq 2^{Q\frac{2+\gamma}{2}} \sup_{m\in\mathbb{N}} \sup_{1\leq t\leq m} E\left[|b(X_{m,t})|^{i\frac{2+\gamma}{2}} \right] \\
&\overset{(B.11)}{\leq} 2^{Q\frac{2+\gamma}{2}} \delta^2_{k(n)} \\
&= \left(2^{\frac{Q}{2}} \tau_{k(n)} \right)^{2+\gamma},
\end{aligned}
$$

where $(*)$ holds because

$$
|x+y|^p = 2^p \left| \frac{1}{2}x + \frac{1}{2}y \right|^p \leq 2^p \left(\frac{1}{2}|x|^p + \frac{1}{2}|y|^p \right) = 2^{p-1}(|x|^p + |y|^p)
$$

for all $x, y \in \mathbb{R}$ and for all $p \in [1, \infty)$, as $x \mapsto |x|^p$ is a convex function.

Assumption **(i)** of Lemma B.4 is also satisfied by **(A1)** for $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ and inherited to $\{Z_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$. Applying Lemma B.4 to $Z_{n,t}$, it follows that for all $b \in \mathcal{B}_{k(n)}$, there exists some constant C, only depending on Q, γ and the mixing coefficient, such that

$$
E\left[|G_n(b)|^Q \right]^{\frac{1}{Q}} \leq C \max\left(n^{-\frac{1}{2}}, 2^{\frac{Q}{2}} \tau_{k(n)} \right) \overset{(B.15)}{=} C' \tau_{k(n)}, \tag{B.18}
$$

with $C' := C2^{\frac{Q}{2}}$. Finally, it can be concluded that for all $n \geq n_2'$

$$
E^*\left[\sup_{\varphi\in\mathcal{F}} \left| G_n(\varphi) - G_n\left(a^*_{k(n)}\right) \right|^Q \right]^{\frac{1}{Q}} \overset{(B.17)}{\leq} E\left[\max_{b\in\mathcal{B}_{k(n)}} |G_n(b)|^Q \right]^{\frac{1}{Q}} + \frac{\varepsilon}{32}
$$

$$
\begin{aligned}
&\leq N_{k(n)}^{\frac{1}{Q}} \max_{b \in \mathcal{B}_{k(n)}} E\left[|G_n(b)|^Q\right]^{\frac{1}{Q}} + \frac{\varepsilon}{32} \\
&\overset{(B.18)}{\leq} C' N_{k(n)}^{\frac{1}{Q}} \tau_{k(n)} + \frac{\varepsilon}{32} \\
&= C' \left(\tilde{N}_{[\,]}\left(\delta_{k(n)}, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} \delta_{k(n)}^{\frac{2}{2+\gamma}} + \frac{\varepsilon}{32} \\
&= C' \delta_{k(n)}^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}\left(\delta_{k(n)}, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} \delta_{k(n)} + \frac{\varepsilon}{32} \\
&\leq C' \int_0^{\delta_{k(n)}} x^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}(x, \mathcal{F}, \rho)\right)^{\frac{1}{Q}} dx + \frac{\varepsilon}{32},
\end{aligned}
$$

where the last inequality uses the fact that $x \mapsto x^{-\frac{\gamma}{2+\gamma}}\left(\tilde{N}_{[\,]}(x, \mathcal{F}, \rho)\right)^{\frac{1}{Q}}$ is decreasing and that the integral exists by assumption **(A2)**. As $\delta_{k(n)} \searrow 0$, there exists a $n_2'' = n_2''(\epsilon)$ such that

$$
C' \int_0^{\delta_{k(n)}} x^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}(x, \mathcal{F}, \rho)\right)^{\frac{1}{Q}} dx < \frac{\varepsilon}{32},
$$

for all $n \geq n_2''$. By choosing $n_2 = \max(n_2', n_2'')$, the assertion in (B.13) follows.

Proof of (B.14): The aim is to choose an $m = m(\epsilon)$ fixed (dependent only on ϵ eventually) and for each $\varphi \in \mathcal{F}$ the corresponding approximating function $a_{k(n)}^* \in \mathcal{A}_{k(n)}$. Then a chain from $a_{k(n)} := a_{k(n)}^* \in \mathcal{A}_{k(n)}$ to $a_m := a_m^{(\varphi)} \in \mathcal{A}_m$ for all $k(n) > m$ is built. In what follows, the iterative choice of functions from one chain link a_k to the next one a_{k-1} will be illustrated. For an already chosen $a_k \in \mathcal{A}_k$, choose $a_{k-1} \in \mathcal{A}_{k-1}$, such that

$$
a_{k-1} \in \left\{ a \in \mathcal{A}_{k-1} : \max_{2 \leq i \leq Q} \sup_{n \in \mathbb{N}} \sup_{1 \leq t \leq n} E\left[|a_k(X_{n,t}) - a(X_{n,t})|^{i\frac{2+\gamma}{2}}\right] \text{ is minimal.} \right\}
\tag{B.19}
$$

Such an object exists as the considered term is bounded from below by zero. If there is more than one minimizer, then one of them is chosen randomly. By doing so, while φ ranges over $\mathcal{F}$, each difference $(a_k - a_{k-1})$ ranges over at most N_k functions because a_k ranges over at most $|\mathcal{A}_k| = N_k$ functions and for each a_k, a_{k-1} is chosen according to the procedure above. Then it holds that

$$
\begin{aligned}
E\left[\sup_{\varphi \in \mathcal{F}} |G_n(a_k) - G_n(a_{k-1})|^Q\right]^{\frac{1}{Q}} &\leq N_k^{\frac{1}{Q}} \sup_{\varphi \in \mathcal{F}} E\left[|G_n(a_k) - G_n(a_{k-1})|^Q\right]^{\frac{1}{Q}} \\
&= N_k^{\frac{1}{Q}} \sup_{\varphi \in \mathcal{F}} E\left[|G_n(a_k - a_{k-1})|^Q\right]^{\frac{1}{Q}}. \qquad \text{(B.20)}
\end{aligned}
$$

Notice again that unlike the supremum suggests, the possible difference ranges over finitely many functions and therefore the inequality used in (B.20) is valid and the outer expectation simplifies to the usual expectation. For the expected value of the last term, the moment inequality from Lemma B.4 will be used again. Defining

$$Z_{n,t} := a_k(X_{n,t}) - a_{k-1}(X_{n,t}) - E[a_k(X_{n,t}) - a_{k-1}(X_{n,t})],\ \forall\ 1 \leq t \leq n, n \in \mathbb{N},$$

it can be obtained that $E[Z_{n,t}] = 0$ for all $1 \leq t \leq n, n \in \mathbb{N}$. Furthermore, by assumption **(A2)**, for $a_k \in \mathcal{A}_k \subset \mathcal{F}$, there exist an $\tilde{a}^*_{k-1} \in \mathcal{A}_{k-1}$ and a $\tilde{b}^*_{k-1} \in \mathcal{B}_{k-1}$, such that

$$\left|a_k - \tilde{a}^*_{k-1}\right| \leq \tilde{b}^*_{k-1} \text{ and } \sup_{n\in\mathbb{N}} \sup_{1\leq t\leq n} E\left[\left|b(X_{n,t})\right|^{i\frac{2+\gamma}{2}}\right]^{\frac{1}{2}} \leq \delta_{k-1},\ \forall\, i = 2, \ldots, Q,\ \forall\, b \in \mathcal{B}_{k-1}.$$

Using (B.19), it thus holds that

$$\begin{aligned}
\max_{2\leq i\leq Q} E\left[\left|Z_{n,t}\right|^{i\frac{2+\gamma}{2}}\right] &= \max_{2\leq i\leq Q} E\left[\left|a_k(X_{n,t}) - a_{k-1}(X_{n,t}) - E[a_k(X_{n,t}) - a_{k-1}(X_{n,t})]\right|^{i\frac{2+\gamma}{2}}\right] \\
&\leq 2^{Q\frac{2+\gamma}{2}} \max_{2\leq i\leq Q} \sup_{n\in\mathbb{N}} \sup_{1\leq t\leq n} E\left[\left|a_k(X_{n,t}) - a_{k-1}(X_{n,t})\right|^{i\frac{2+\gamma}{2}}\right] \\
&\overset{(B.19)}{\leq} 2^{Q\frac{2+\gamma}{2}} \max_{2\leq i\leq Q} \sup_{n\in\mathbb{N}} \sup_{1\leq t\leq n} E\left[\left|a_k(X_{n,t}) - \tilde{a}^*_{k-1}(X_{n,t})\right|^{i\frac{2+\gamma}{2}}\right] \\
&\leq 2^{Q\frac{2+\gamma}{2}} \max_{2\leq i\leq Q} \sup_{n\in\mathbb{N}} \sup_{1\leq t\leq n} E\left[\left|\tilde{b}^*_{k-1}(X_{n,t})\right|^{i\frac{2+\gamma}{2}}\right] \\
&\leq 2^{Q\frac{2+\gamma}{2}} \delta^2_{k-1} = \left(2^{\frac{Q}{2}}\tau_{k-1}\right)^{2+\gamma},
\end{aligned}$$

and thus assumption **(ii)** of Lemma B.4 is satisfied. Condition **(i)** holds by **(A1)** for $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ and is inherited to $\{Z_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$. Then applying Lemma B.4 and using $\tau_k \geq n^{-\frac{1}{2}}$ for all $1 \leq k \leq k(n)$ by (B.15), yields

$$E\left[\left|G_n\left(a_k - a_{k-1}\right)\right|^Q\right]^{\frac{1}{Q}} \leq C\tau_{k-1}, \tag{B.21}$$

for some constant C only depending on γ, Q and the mixing coefficient. Now all tools to build the bridge between $a_{k(n)} := a^*_{k(n)}$ and $a_m := a^{(\varphi)}_m$ are obtained and it holds that

$$\begin{aligned}
&E\left[\sup_{\varphi\in\mathcal{F}} \left|G_n\left(a^*_{k(n)}\right) - G_n\left(a^{(\varphi)}_m\right)\right|^Q\right]^{\frac{1}{Q}} \\
&= E\left[\sup_{\varphi\in\mathcal{F}} \left|\sum_{k=m+1}^{k(n)} \left(G_n\left(a_k\right) - G_n\left(a_{k-1}\right)\right)\right|^Q\right]^{\frac{1}{Q}} \\
&\leq \sum_{k=m+1}^{k(n)} E\left[\sup_{\varphi\in\mathcal{F}} \left|G_n\left(a_k\right) - G_n\left(a_{k-1}\right)\right|^Q\right]^{\frac{1}{Q}} \\
&\overset{(B.20)}{\leq} \sum_{k=m+1}^{k(n)} N_k^{\frac{1}{Q}} \sup_{\varphi\in\mathcal{F}} E\left[\left|G_n\left(a_k - a_{k-1}\right)\right|^Q\right]^{\frac{1}{Q}} \\
&\overset{(B.21)}{\leq} C \sum_{k=m+1}^{k(n)} N_k^{\frac{1}{Q}} \tau_{k-1}
\end{aligned}$$

$$
\begin{aligned}
&\leq 2^{\frac{2}{2+\gamma}} C \sum_{k=m+1}^{\infty} \left(\tilde{N}_{[\,]}\left(\delta_k, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} \delta_k^{\frac{2}{2+\gamma}} \\
&= 2^{\frac{2}{2+\gamma}} C \sum_{k=m+1}^{\infty} \delta_k^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}\left(\delta_k, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} \delta_k \\
&= 2^{\frac{2}{2+\gamma}+1} C \sum_{k=m+1}^{\infty} \delta_k^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}\left(\delta_k, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} \left(\delta_k - \delta_{k-1}\right) \\
&\leq 2^{\frac{2}{2+\gamma}+1} C \int_0^{\delta_m} x^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}\left(x, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}} dx,
\end{aligned}
$$

for all $k(n) > m$. The last equality holds because $\delta_k - \delta_{k-1} = 2^{-1}\delta_k$. The last inequality again holds as $x \mapsto x^{-\frac{\gamma}{2+\gamma}} \left(\tilde{N}_{[\,]}\left(x, \mathcal{F}, \rho\right)\right)^{\frac{1}{Q}}$ is decreasing and the integral exists due to assumption **(A2)**. Furthermore, $\delta_m \searrow 0$ for $m \to \infty$. Hence, for a given $\epsilon > 0$, $m = m(\epsilon)$ and $n_3 = n_3(\epsilon)$ can be chosen large enough, such that

$$
E\left[\sup_{\varphi \in \mathcal{F}} \left|G_n\left(a^*_{k(n)}\right) - G_n\left(a_m^{(\varphi)}\right)\right|^Q\right]^{\frac{1}{Q}} < \frac{\epsilon}{16}
$$

for all $n \geq n_3$ with $k(n) > m$, which proves inequality (B.14).

Step 2: In the second and last step of the proof, the comparison of infinitely many functions in $\mathcal{F}$ will be reduced to finitely many functions, making use of inequality (B.12). To do that, let $m \in \mathbb{N}$ be the integer fixed in step one and refer with $a_m^{(\varphi)}$ to the element in $\mathcal{A}_m$, that is chosen dependent on $\varphi \in \mathcal{F}$, according to the procedure in step one. Let the following relation on $\mathcal{F}$ (dependent on m) be introduced

$$
\varphi \sim_m \psi :\Leftrightarrow a_m^{(\varphi)} = a_m^{(\psi)}.
$$

This relation is obviously an equivalence relation and, as $|\mathcal{A}_m| = N_m$, partitions $\mathcal{F}$ into N_m many equivalence classes, denoted by

$$
\mathcal{E}^{(m)}[1], \ldots, \mathcal{E}^{(m)}[N_m].
$$

Each class thus contains all φ in $\mathcal{F}$, that have the same $a_m^{(\varphi)}$ in $\mathcal{A}_m$, that has been chosen in step one. Within one equivalence class, inequality (B.12) can be applied twice, leading to

$$
\begin{aligned}
&E^*\left[\sup_{\{\varphi,\psi \in \mathcal{F} \mid \varphi \sim_m \psi\}} \left|G_n(\varphi) - G_n(\psi)\right|^Q\right]^{\frac{1}{Q}} \\
&= E^*\left[\sup_{\{\varphi,\psi \in \mathcal{F} \mid \varphi \sim_m \psi\}} \left|\left(G_n(\varphi) - G_n(a_m^{(\varphi)})\right) - \left(G_n(\psi) - G_n(a_m^{(\psi)})\right)\right|^Q\right]^{\frac{1}{Q}} \\
&\leq 2E^*\left[\sup_{\varphi \in \mathcal{F}} \left|G_n(\varphi) - G_n(a_m^{(\varphi)})\right|^Q\right]^{\frac{1}{Q}} \\
&\overset{(B.12)}{<} \frac{\epsilon}{4},
\end{aligned} \tag{B.22}
$$

for all $n \geq n_1$. To bridge the gap between the N_m classes, let

$$d(\mathcal{E}^{(m)}[k], \mathcal{E}^{(m)}[j]) := \inf\left\{d(\varphi,\psi) : \varphi \in \mathcal{E}^{(m)}[k], \psi \in \mathcal{E}^{(m)}[j]\right\}$$

define a distance between two classes $\mathcal{E}^{(m)}[k]$ and $\mathcal{E}^{(m)}[j]$ for $k, j \in \{1, \ldots, N_m\}$. For fixed $\delta > 0$, that will be specified later, and fixed $k, j \in \{1, \ldots, N_m\}$, choose functions $\varphi'_{kj} \in \mathcal{E}^{(m)}[k]$ and $\psi'_{jk} \in \mathcal{E}^{(m)}[j]$, such that

$$d(\varphi'_{kj}, \psi'_{jk}) < d(\mathcal{E}^{(m)}[k], \mathcal{E}^{(m)}[j]) + \delta.$$

Note that for $\varphi \in \mathcal{E}^{(m)}[k]$ and $\psi \in \mathcal{E}^{(m)}[j]$ with $d(\varphi,\psi) < \delta$, it holds that $d(\varphi'_{kj}, \psi'_{jk}) < 2\delta$ for all $k, j \in \{1, \ldots, N_m\}$. Then applying (B.22) for all $n \geq n_1$, it can be obtained that

$$\begin{aligned}
&E^*\left[\sup_{\{\varphi,\psi\in\mathcal{F}: d(\varphi,\psi)<\delta\}} |G_n(\varphi) - G_n(\psi)|^Q\right]^{\frac{1}{Q}} \\
&= E^*\left[\max_{\substack{1\leq k\leq N_m\\ 1\leq j\leq N_m}} \sup_{\substack{\{\varphi\in\mathcal{E}^{(m)}[k],\psi\in\mathcal{E}^{(m)}[j]:\\ d(\varphi,\psi)<\delta\}}} |G_n(\varphi) - G_n(\psi)|^Q\right]^{\frac{1}{Q}} \\
&= E^*\left[\max_{\substack{1\leq k\leq N_m\\ 1\leq j\leq N_m}} \sup_{\substack{\{\varphi\in\mathcal{E}^{(m)}[k],\psi\in\mathcal{E}^{(m)}[j]:\\ d(\varphi,\psi)<\delta\}}} |G_n(\varphi) - G_n(\psi) \pm G_n(\varphi'_{kj}) \pm G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}} \\
&\leq 2E^*\left[\max_{1\leq k\leq N_m} \sup_{\{\varphi,\varphi'\in\mathcal{E}^{(m)}[k]\}} |G_n(\varphi) - G_n(\varphi')|^Q\right]^{\frac{1}{Q}} + E\left[\max_{\substack{1\leq k\leq N_m\\ 1\leq j\leq N_m}} |G_n(\varphi'_{kj}) - G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}} \\
&\overset{(B.22)}{<} \frac{\epsilon}{2} + E\left[\max_{\substack{1\leq k\leq N_m\\ 1\leq j\leq N_m}} |G_n(\varphi'_{kj}) - G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}} \\
&\leq \frac{\epsilon}{2} + N_m^{\frac{2}{Q}} \max_{\substack{1\leq k\leq N_m\\ 1\leq j\leq N_m}} E\left[|G_n(\varphi'_{kj}) - G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}},
\end{aligned}$$

where $d(\varphi'_{kj}, \psi'_{jk}) < 2\delta$ holds for all $k, j \in \{1, \ldots, N_m\}$. To find a bound on $E\left[|G_n(\varphi'_{kj}) - G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}}$, the moment inequality of Lemma B.4 will be used. Let therefore $k, j \in \{1, \ldots, N_m\}$ be fixed and let $\varphi'_{kj} \in \mathcal{E}^{(m)}[k]$ and $\psi'_{jk} \in \mathcal{E}^{(m)}[j]$ with

$$d(\varphi'_{kj}, \psi'_{jk}) = \sup_{n\in\mathbb{N}} \sup_{1\leq t\leq n} E\left[\left|\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t})\right|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}\frac{2}{2+\gamma}} < 2\delta \qquad \text{(B.23)}$$

and set $Z_{n,t} := \varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t}) - E[\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t})]$ for all $1 \leq t \leq n$ and $n \in \mathbb{N}$. Then assumption **(ii)** of Lemma B.4 is satisfied as for all $1 \leq t \leq n, n \in \mathbb{N}$ and $i = 2, \ldots, Q$ it holds that $E[Z_{n,t}] = 0$ and

$$E\left[\left|\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t}) - E[\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t})]\right|^{i\frac{2+\gamma}{2}}\right]$$

$$
\begin{aligned}
&\leq 2^{Q\frac{2+\gamma}{2}} E\left[|\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t})|^{i\frac{2+\gamma}{2}}\right] \\
&\leq 2^{Q\frac{2+\gamma}{2}} E\left[|\varphi'_{kj}(X_{n,t}) - \psi'_{jk}(X_{n,t})|^{Q\frac{2+\gamma}{2}}\right]^{\frac{i}{Q}} \\
&\leq 2^{Q\frac{2+\gamma}{2}} d(\varphi'_{kj}, \psi'_{jk})^{i\frac{2+\gamma}{2}} \\
&\overset{(B.23)}{\leq} 2^{Q\frac{2+\gamma}{2}} (2\delta)^{i\frac{2+\gamma}{2}} \\
&\leq 2^{Q\frac{2+\gamma}{2}} (2\delta)^{2+\gamma}, \quad \text{for } \delta \leq \frac{1}{2} \\
&= \left(2^{\frac{Q}{2}+1}\delta\right)^{2+\gamma}.
\end{aligned}
$$

Condition **(i)** holds due to **(A1)** for $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ and is inherited to $\{Z_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$. Applying Lemma B.4 yields

$$
E\left[|G_n(\varphi'_{kj} - \psi'_{jk})|^Q\right]^{\frac{1}{Q}} \leq C \max\left(n^{-\frac{1}{2}}, 2^{\frac{Q}{2}+1}\delta\right),
$$

for some constant C only depending on γ, Q and the mixing coefficient and for $\delta \leq \frac{1}{2}$. Therefore, it holds that

$$
\begin{aligned}
E^*\left[\sup_{\{\varphi,\psi\in\mathcal{F}:d(\varphi,\psi)<\delta\}} |G_n(\varphi) - G_n(\psi)|^Q\right]^{\frac{1}{Q}} &< \frac{\epsilon}{2} + N_m^{\frac{2}{Q}} \max_{\substack{1\leq k\leq N_m \\ 1\leq j\leq N_m}} E\left[|G_n(\varphi'_{kj}) - G_n(\psi'_{jk})|^Q\right]^{\frac{1}{Q}} \\
&< \frac{\epsilon}{2} + N_m^{\frac{2}{Q}} C \max\left(n^{-\frac{1}{2}}, 2^{\frac{Q}{2}+1}\delta\right),
\end{aligned}
$$

for all $n \geq n_1$ and for $\delta \leq \frac{1}{2}$. Choose $\delta = \delta(\epsilon)$ small enough, such that

$$
N_m^{\frac{2}{Q}} C 2^{\frac{Q}{2}+1}\delta < \frac{\epsilon}{2} \text{ and } \delta \leq \frac{1}{2}
$$

and for this fixed δ, let $n_4 = n_4(\epsilon)$, such that

$$
\max\left(n^{-\frac{1}{2}}, 2^{\frac{Q}{2}+1}\delta\right) = 2^{\frac{Q}{2}+1}\delta,
$$

for all $n \geq n_4$. By finally choosing $n_0 := \max(n_1, n_4)$, the assertion in (B.9) is proven.

□

Proof of Corollary B.2. This is a direct consequence of Theorem B.1 and Markov's inequality.

□

Proof of Corollary B.3. To prove Corollary B.3, Theorem 4.10 by Volgushev and Shao [77] will be used. Note that it particularly requires a strictly stationary sequence of random variables. Applying Theorem B.1, it follows that with

$$
d(\varphi, \psi) := E\left[|\varphi(X_1) - \psi(X_1)|^{Q\frac{2+\gamma}{2}}\right]^{\frac{1}{Q}\frac{2}{2+\gamma}}
$$

there exists a semi metric d on $\mathcal{F}$, such that $(\mathcal{F}, d)$ is totally bounded and there exists a $Q > 2$, such that

$$\lim_{\delta \searrow 0} \limsup_{n\to\infty} E^* \left[\sup_{\{\varphi,\psi\in\mathcal{F}:d(\varphi,\psi)<\delta\}} |G_n(1, \varphi - \psi)|^Q \right] = 0,$$

which is condition (9) of Theorem 4.10 of Volgushev and Shao [77]. Furthermore, condition (10) of Theorem 4.10 in [77], namely

$$\sup_{n\in\mathbb{N}} \sup_{\varphi\in\mathcal{F}} E\left[|G_n(1,\varphi)|^Q\right] < \infty$$

is also satisfied. To see this, define $Z_t := \varphi(X_t) - E[\varphi(X_t)]$ for all $t \in \mathbb{Z}$. Applying **(A3)**, it then holds that $E[Z_1] = 0$ and for all $i = 2, \ldots, Q$

$$\begin{aligned} E\left[|Z_1|^{i\frac{2+\gamma}{2}}\right] &= E\left[|\varphi(X_1) - E[\varphi(X_1)]|^{i\frac{2+\gamma}{2}}\right] \\ &\leq 2^{Q\frac{2+\gamma}{2}} E\left[|\varphi(X_1)|^{i\frac{2+\gamma}{2}}\right] \\ &\leq 2^{Q\frac{2+\gamma}{2}} E\left[|\varphi(X_1)|^{Q\frac{2+\gamma}{2}}\right]^{\frac{i}{Q}} \\ &\leq 2^{Q\frac{2+\gamma}{2}} \max\left(L^{\frac{2}{Q}}, L\right) \\ &=: \tau^{2+\gamma}, \end{aligned}$$

for the constant $L < \infty$ from assumption **(A3)**. Applying Lemma B.4, it holds that for all $n \in \mathbb{N}$ and $\varphi \in \mathcal{F}$

$$E\left[|G_n(1,\varphi)|^Q\right]^{\frac{1}{Q}} \leq C \max\left(n^{-\frac{1}{2}}, \tau\right), \tag{B.24}$$

for some constant C, only depending on Q, γ and the mixing coefficient. As the inequality in **(A3)** holds uniformly in $\varphi \in \mathcal{F}$, the constant $\tau < \infty$ does not depend on φ. Therefore, (B.24) implies condition (10) of Theorem 4.10 in [77]. By assumption **(A3)** the function class $\mathcal{F}$ possesses an envelope function with finite Q-th moment. As can be seen within the proof of Theorem 4.10 of Volgushev and Shao [77], the first assertion of Corollary B.3 follows by these assumptions.

If additionally all finite dimensional marginal distributions converge, as assumed in the second part of Corollary B.3, the weak convergence assertion follows.

□

Outlook

Finally, some possible further research topics related to this thesis shall be presented in the following.

As noted in Chapter 4, changepoint detection in the conditional mean function, when also changes in the variance occur, is an important but also difficult task. To the best of our knowledge, it has only been investigated in the literature for simple cases, that do not allow for autoregression models, by assuming stationarity of the covariates under the null, or that do not include heteroscedastic effects. The model in Section 4.2 covers both autoregressive and heteroscedastic effect, where changes may occur in the conditional variance function. It is suggested that in this case the process $\hat{T}_n$ still converges to some centered Gaussian process under the null of no change in the conditional mean function. Such a result, however, requires a uniform central limit theorem for sequential empirical processes indexed in general function classes and with an underlying triangular array process of weakly dependent and possibly non-stationary random variables. To be a bit more precise, for some given weakly dependent triangular array process $\{X_{n,t} : 1 \leq t \leq n, n \in \mathbb{N}\}$ with values in some measure space $\mathcal{X}$ and some class $\mathcal{F}$ of measurable functions $\mathcal{X} \to \mathbb{R}$, the following sequential empirical process is of interest

$$\left\{ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} \left(\varphi(X_{n,i}) - E[\varphi(X_{n,i})]\right) : s \in [0,1], \varphi \in \mathcal{F} \right\}.$$

Appendix B gives some literature review on results concerning sequential empirical processes. Basically all existing results allowing for dependent data, including the new Corollary B.3 on page 161, impose a strict stationarity assumption on the underlying process. The extension to non-stationary data would not only be applicable to the process $\hat{T}_n$ under non-stationary variances as in Chapter 4, it would in fact be a powerful tool for proving limiting results in all kind of models in mathematical statistics.

Additionally, bootstrap procedures are important and very useful tools for testing hypothesis in mathematical statistics. They are used to approximate the distribution of the test statistic under the null, without knowing its limiting distribution. This is particularly useful if the limiting distribution is not known. For instance, the test for change in the conditional mean function in Chapter 3 acquires for multidimensional covariates, i.e. for $d > 1$, a limiting distribution that contains unknown quantities. Hence, critical values using the asymptotic distribution can not be computed. The potential limiting distribution of the process allowing for non-stationary variances look even more complicated. However, even if the limiting distribution is known, it is still only an approximation, a possibly poor one for small sample sizes. For these reasons bootstrap procedures are of undeniable interest. The conjecture that the wild bootstrap test, proposed in Section 4.3, is a valid testing procedure, should be proved. Note that apart from the weak convergence in probability of $\hat{T}_n^*$ under both the null and the alternative, this also requires the the weak convergence of $\hat{T}_n$ under the null.

Tests for change in the conditional variance function in heteroscedastic models can be of particular interest as well. In financial time series and other econometric data, they can be particularly useful, when models are for instance used for hedging strategies and risk management. However, not that many results consider nonparametric models. Thus, the test proposal from Chapter 5 should be formulated more detailed and a formal proof for the limiting distribution should be given. As already suggested in Chapter 5 the methods will be similar, however the proof can possibly be even more technical.

Finally, it is to mention that apart from changepoint detection, which this thesis focuses on, it is also of great interest to estimate both size and location of possible changes. Such estimators can possibly be constructed based on the sequential marked empirical process of residuals, obtained in this thesis. Possible estimators for a changepoint in the conditional mean function using $\hat{T}_n$ from Chapter 3 are

$$\operatorname*{arg\,max}_{s\in[0,1]} \sup_{\boldsymbol{z}\in\mathbb{R}^d} |\hat{T}_n(s,\boldsymbol{z})|,$$

and

$$\operatorname*{arg\,max}_{s\in[0,1]} \int\limits_{\mathbb{R}^d} |\hat{T}_n(s,\boldsymbol{z})|^2 w(\boldsymbol{z}) d\boldsymbol{z},$$

where $w : \mathbb{R}^d \to \mathbb{R}$ is some weighting function such that the integral exists. These estimators can be investigated with respect to properties, such as consistency and asymptotic distribution.

Bibliography

[1] M. G. Akritas and I. Van Keilegom. Non-Parametric Estimation of the Residual Distribution. *Scand. J. Statist.*, 28:549–567, 2001.

[2] D. W. K. Andrews and D. Pollard. An Introduction to Functional Central Limit Theorems for Dependent Stochastic Processes. *Int. Stat. Rev.*, 62:549–567, 1994.

[3] J. Bai. On the partial sums of residuals in autoregressive and moving average models. *J. Time Ser. Anal.*, 14:247–260, 1993.

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

[5] R. L. Brown, J. Durbin, and J. M. Evans. Techniques for Testing the Constancy of Regression Relationships over Time. *J. R. Stat. Soc.*, 37:149–192, 1975.

[6] M. D. Burke and G. Bewa. Change-Point Detection for General Nonparametric Regression Models. *Open Journal of Statistics*, 3:261–267, 2013.

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

[8] G. Chen, Y. K. Choi, and Y. Zhou. Nonparametric estimation of structural change points in volatility models for time series. *J. Econom.*, 126:79–114, 2005.

[9] J. Chen and A. K. Gupta. Testing and Locating Variance Changepoints with Application to Stock Prices. *J. Amer. Statist. Assoc.*, 92:739–747, 1997.

[10] Z. Chen and Z. Tian. Ratio tests for variance change in nonparametric regression. *Statistics*, 48:1–16, 2014.

[11] G. W. Cobb. The Problem of the Nile: Conditional Solution to a Changepoint Problem. *Biometrika*, 65:243–251, 1978.

[12] M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley, Chichester, 1997.

[13] J. Dedecker and S. Louhichi. Maximal Inequalities and Empirical Central Limit Theorems. In H. Dehling, T. Mikosch, and M. Sørensen, editors, *Empirical Process Techniques for Dependent Data*, pages 137–159. Birkhäuser, Boston, 2002.

[14] H. Dehling, O. Durieu, and M. Tusche. A sequential empirical CLT for multiple mixing processes with application to $\mathcal{B}$-geometrically ergodic Markov chains. *Electron. J. Probab.*, 19:1–26, 2014.

[15] H. Dehling, O. Durieu, and M. Tusche. Approximating class approach for empirical processes of dependent sequences indexed by functions. *Bernoulli*, 20:1372–1403, 2014.

[16] M. A. Delgado and J. Hidalgo. Nonparametric inference on structural breaks. *J. Econom.*, 96:113–144, 2000.

[17] H. Dette, J. C. Pardo-Fernández, and I. Van Keilegom. Goodness-of-Fit Tests for Multiplicative Models with Dependent Data. *Scand. J. Statist.*, 36:782–799, 2009.

[18] P. Doukhan. *Mixing: Properties and Examples.* Springer, New York, 1995.

[19] P. Doukhan, P. Massart, and E. Rio. Invariance principles for absolutely regular empirical processes. *Ann. Inst. H. Poincaré*, 31:393–427, 1995.

[20] L. C. Evans. *Partial differential equations.* American Math. Soc., Providence, RI, 2. edition, 2010.

[21] J. Fan and Q. Yao. *Nonlinear Time Series.* Springer, New York, 2003.

[22] E. Giné and J. Zinn. Bootstrapping general empirical measures. *Ann. Probab.*, 18:851–869, 1990.

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

[24] A. Hagemann. Stochastic equicontinuity in nonlinear times series models. *Econom. J.*, 17:188–196, 2014.

[25] B. E. Hansen. Stochastic equicontinuity for unbounded dependent heterogeneous arrays. *Econom. Theory*, 12:347–359, 1996.

[26] B. E. Hansen. Uniform convergence rates for kernel estimation with dependent data. *Econom. Theory*, 24:726–748, 2008.

[27] W. Härdle and J. S. Marron. Bootstrap simultaneous error bars for nonparametric regression. *Ann. Statist.*, 19:778–796, 1991.

[28] S. B. Hariz. Uniform CLT for empirical process. *Stochastic Process. Appl.*, 115:339–358, 2005.

[29] J. Hidalgo. A Nonparametric Conditional Moment Test for Structural Stability. *Econom. Theory*, 1:671–698, 1995.

[30] J. Hidalgo and V. Dalla. Testing for Breaks in Regression Models with Dependent Data. In R. Cao, W. G. Manteiga, and J. Romo, editors, *Nonparametric Statistics*, pages 19–44. Springer, Berlin, 2016.

[31] T. Honda. The CUSUM tests with nonparametric regression residuals. *J. Japan Statist. Soc.*, 27:45–63, 1997.

[32] L. Horváth. Change in autoregressive processes. *Stochastic Process. Appl.*, 44:221–242, 1993.

[33] D. A. Hsu. Detecting Shifts of Parameter in Gamma Sequences with Applications to Stock Price and Air Traffic Flow Analysis. *J. Amer. Statist. Assoc.*, 74:31–40, 1979.

[34] T. Juhl and Z. Xiao. Nonparametric tests of moment condition stability. *Econom. Theory*, 29:90–114, 2013.

[35] C. Kirch and J. T. Kamgaing. Testing for parameter stability in nonlinear autoregressive models. *J. Time Ser. Anal.*, 33:365–385, 2012.

[36] K. Königsberger. *Analysis 2*. Springer, Berlin, 5. edition, 2003.

[37] H. L. Koul. Asymptotics of some estimators and sequential residual empiricals in nonlinear time series. *Ann. Statist.*, 24:380–404, 1996.

[38] W. Krämer, W. Ploberger, and R. Alt. Testing for Structural Change in Dynamic Models. *Econometrica*, 56:1355–1369, 1988.

[39] J.-P. Kreiß. Manuscript: Asymptotic properties of residual bootstrap for autoregressions, 1997.

[40] J.-P. Kreiß and S. N. Lahiri. Bootstrap Methods for Time Series. In S. Subba Rao T. Subba Rao and C.R. Rao, editors, *Handbook of Statistics*, pages 3–26. Elsevier, Amsterdam, 2012.

[41] J.-P. Kreiß and G. Neuhaus. *Einführung in die Zeitreihenanalyse*. Springer, Berlin, 2006.

[42] D. Kristensen. Uniform convergence rates of kernel estimators with heterogeneous dependent data. *Econom. Theory*, 25:1433–1445, 2009.

[43] D. Kristensen. Non-parametric detection and estimation of structural change. *Econom. J.*, 15:420–461, 2012.

[44] S. Lee. Efficient semiparametric estimation of partially linear quantile regression models. *Econom. Theory*, 19:1–31, 2003.

[45] S. Lee, O. Na, and S. Na. On the CUSUM of squares test for the variance change in nonstationary and nonparametric time series models. *Ann. Inst. Statist. Math.*, 55:467–485, 2003.

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

[47] E. Liebscher. Towards a Unified Approach for Proving Geometric Ergodicity and Mixing Properties of Nonlinear Autoregressive Processes . *J. Time Ser. Anal.*, 26:669–689, 2005.

[48] S. Ling. Weak convergence of the sequential empirical processes of residuals in nonstationary autoregressive models. *Ann. Statist.*, 26:741–754, 1998.

[49] R. Y. Liu. Bootstrap procedure under some non-i.i.d. models. *Ann. Statist.*, 16:1696–1708, 1988.

[50] Z. Lu. On the geometric ergodicity of a non-linear autoregressive model with an autoregressive conditional heteroscedastic term. *Stat. Sin.*, 8:1205–1217, 1998.

[51] P. Massart. Invariance principles for empirical processes: the weakly dependent case. In *Chapter 1B of Ph.D. dissertation*, pages 58–102. University of Paris-South, Orsay, 1987.

[52] M. Mohr. Weak Convergence of Sequential Empirical Processes under Weak Dependence. preprint on arxiv at `https://arxiv.org/abs/1711.05112`, November 2017.

[53] É. A. Nadaraya. On non-parametric estimates of density functions and regression curves. *Theory Probab. Appl.*, 10:186–190, 1964.

[54] N. Neumeyer and I. Van Keilegom. Change-Point Tests for the Error Distribution in Non-parametric Regression. *Scand. J. Statist.*, 36:518–541, 2009.

[55] N. Neumeyer and I. Van Keilegom. Estimating the error distribution in non-parametric multiple regression with applications to model testing. *J. Multivariate Anal.*, 101:1067–1078, 2010.

[56] N. Neumeyer, M. Omelka, and S. Hudecova. A copula approach for dependence modeling in multivariate nonparametric time series. preprint on arxiv at `https://arxiv.org/abs/1705.07605`, May 2017.

[57] M. Ossiander. A central limit theorem under metric entropy with L_2-bracketing. *Ann. Probab.*, 15:897–919, 1987.

[58] E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41:100–115, 1954.

[59] E. S. Page. Control Charts with Warning Lines. *Biometrika*, 42:243–257, 1955.

[60] E. Parzen. On the estimation of a probability density function and mode. *Ann. Math. Stat.*, 33:1065–1076, 1955.

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

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

[63] D. N. Politis, J. P. Romano, and M. Wolf. *Subsampling.* Springer, New York, 1999.

[64] E. Rio. About the Lindeberg method for strongly mixing sequences. *ESAIM Probab. Statist.*, 1:35–61, 1995.

[65] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *Ann. Math. Stat.*, 27:832–837, 1956.

[66] L. Selk and N. Neumeyer. Testing for a Change of the Innovation Distribution in Nonparametric Autoregression: The Sequential Empirical Process Approach. *Scand. J. Statist.*, 40:770–788, 2013.

[67] X. Shao. The Dependent Wild Bootstrap. *J. Am. Stat. Assoc.*, 105:218–235, 2010.

[68] W. Stute. Nonparametric model checks for regression. *Ann. Statist.*, 25:613–641, 1997.

[69] L. Su and A. Ullah. A nonparametric Goodness-of-fit-based test for conditional heteroscedasticity. *Econom. Theory*, 29:187–212, 2013.

[70] L. Su and H. White. Testing structural change in partially linear models. *Econom. Theory*, 26:1761–1806, 2010.

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

[72] S. Su and C.-Y. Chiang. Limiting behavior of the perturbed empirical distribution functions evaluated at U-statistics for strongly mixing sequences of random variables. *J. Appl. Math. Stoch. Anal.*, 10:3–20, 1997.

[73] A. B. Tsybakov. *Introduction to Nonparametric Estimation.* Springer, New York, 2009.

[74] A. W. van der Vaart. *Asymptotic Statistics.* Cambridge University Press, Cambridge, 2000.

[75] A. W. van der Vaart and J. A. Wellner. *Weak convergence and empirical processes.* Springer, New York, 1996.

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

[77] S. Volgushev and X. Shao. A general approach to the joint asymptotic analysis of statistics from sub-samples. *Electron. J. Stat.*, 8:390–431, 2014.

[78] M. P. Wand and M. C. Jones. *Kernel Smooting.* Chapman & Hall, London, 1995.

[79] G. S. Watson. Smooth Regression Analysis. *Sankhya Ser. A*, 26:359–372, 1964.

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

[81] J. Wu. Detecting structural changes under nonstationary volatility. *Econom. Lett.*, 146:151–154, 2016.

[82] Y. Wu. *Inference for Change-Point and Post-Change Means After a CUSUM Test*. Springer, New York, 2005.

[83] K.-L. Xu. Testing for structural change under non-stationary variances. *Econom. J.*, 18:274–305, 2015.

[84] A. Zeileis and K. Hornik. Generalized M-fluctuation tests for parameter instability. *Statist. Neerlandica*, 61:488–508, 2007.

Abstract

When observing data over a given period of time, a typical assumption is structural stability in some form or another. However, in all possible areas, as for instance climate and weather data, finance and econometric time series, as well as data collected in biology or medicine, instabilities do occur and most likely lead to false inference if disregarded. Hence, change detection and estimation procedures are of undeniable interest and have justifiably gained extensive attention in literature. This thesis makes a contribution in the field of changepoint analysis in nonparametric heteroscedastic time series regression models, focusing on the detection of possible changes rather than their estimation.

More specifically, a weakly dependent stochastic process $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t \in \mathbb{Z}\}$ is considered such that

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

holds, where the innovations $(U_t)_{t\in\mathbb{Z}}$ satisfy

$$E[U_t|\mathcal{F}^t] = 0,\ t \in \mathbb{Z}$$

almost surely with $\mathcal{F}^t := \sigma\{U_{j-1}, \boldsymbol{X}_j : j \leq t\}$. Thus $\mathcal{F}^t$ contains the whole past information of U_j up to time $t-1$ and of $\boldsymbol{X}_j$ up to time t. As a direct consequence, the regression function $m_t(\cdot)$ appears as the conditional mean function, meaning

$$m_t(\cdot) = E[Y_t|\boldsymbol{X}_t = \cdot],\ t \in \mathbb{Z}.$$

The subscript t suggests that it may not be stable in t, but depend on the time of observation. The main part of this thesis is the construction of a test for change in the regression function in Chapter 3. Given $n \in \mathbb{N}$ observations $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ the null hypothesis

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

for some $m : \mathbb{R}^d \to \mathbb{R}$ not depending on t, is tested against the changepoint alternative

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

for some $m_{(1)}, m_{(2)} : \mathbb{R}^d \to \mathbb{R}$ with $m_{(1)} \not\equiv m_{(2)}$. Note that the procedure allows in particular for autoregressive models by allowing $\boldsymbol{X}_t$ to include finitely many lagged values of Y_t, for instance $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-d})^T$. Furthermore, heteroscedastic models are covered, as the second moments of the innovation U_t conditioned on $\boldsymbol{X}_t$ may depend on $\boldsymbol{X}_t$. The considered test statistic is based on the sequential marked empirical process of residuals, namely

$$\hat{T}_n := \left\{ \hat{T}_n(s, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_i - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d \right\},$$

where $\omega_n : \mathbb{R}^d \to \mathbb{R}$ is a weighting function that ensures uniform consistency properties of the Nadaraya-Watson estimator $\hat{m}_n$. The use of nonparametric estimators such as $\hat{m}_n$ avoids additional model misspecification. New uniform rates of convergence for the nonparametric kernel estimators are given in Chapter 2. Note that $\hat{T}_n$ is a modification of the test statistic of Su and Xiao [71], who constructed a CUSUM test to detect changepoints in the regression function in a time series regression model with nonparametric methods. A major drawback of their procedure is the non-consistency against some alternatives. In other words, the test does not detect structural breaks in some situations. The idea of the marked version steamed from Burke and Bewa [6], who used a similar technique in a simple i.i.d. nonparametric regression model to test for changepoints in the regression function. Using this method, however, requires the powerful tool of empirical processes, which have been studied quite intensively in the i.i.d. setting, but are much more difficult to handle in the dependent case. The main result of this thesis is Corollary 3.3, which gives the asymptotic behavior of $\hat{T}_n$ under the null and a strict stationarity assumption. In contrast to the result of Burke and Bewa [6], the limiting distribution of the new testing procedure acquires a very simple structure and for one-dimensional covariates even results in a distribution free limiting distribution. The proof requires some sophisticated techniques, as well as a new weak convergence result for sequential empirical processes indexed by general function classes. This result is stated in Corollary B.3 in Appendix B and might be of independent interest. Furthermore, a bootstrap version of the test is constructed in Chapter 4, and it is suggested that it is a valid testing procedure even without the validity of the strict stationarity assumption under the null. This particularly allows for changes in the variance when testing for changes in the regression function.

By imposing

$$E[U_t^2|\boldsymbol{X}_t] = \sigma_t^2(\boldsymbol{X}_t),\ t \in \mathbb{Z}$$

almost surely, it holds that σ_t^2 is the conditional variance function, meaning

$$\sigma_t^2(\cdot) = \mathrm{Var}(Y_t|\boldsymbol{X}_t = \cdot),\ t \in \mathbb{Z}.$$

In Chapter 5, a test for change in σ_t^2 is suggested, under the assumption that there is no change in the regression function. Note that the reasonableness of this assumption can be investigated using the bootstrap test.

Finally, a simulation study is conducted and an application to real data sets is given in Chapter 6. The Monte-Carlo simulations consider different kinds of models, such as i.i.d. models, regression models with time series covariates, as well as homoscedastic and heteroscedastic autoregressive models. Both performances of the tests in conditional mean and variance function, as well as the bootstrap test are investigated. It turns out that they perform reasonably well. In particular, under some simple alternatives they behave significantly better than the CUSUM test in [71] as the theory suggests. The real data sets under consideration are the annual flow of the river Nile in Aswan between 1871 and 1970 and the weekly closing values of the Dow Jones Industrial Average (DJIA) index between 1971 and 1974, which are both popular choices in changepoint analysis. The tests indicate that there is a change in mean in the Nile data set and a change in variance in the DJIA data set. These results are consistent with existing literature.

Zusammenfassung

Bei der Betrachtung zeitlich geordneter Daten gehört die Stabilität, in der einen oder anderen Form, zu den typischen Annahmen. Strukturelle Veränderungen treten allerdings in allen Anwendungsbereichen, wie beispielsweise in Klima- und Wetterdaten, Finanz- und Ökonometriezeitreihen, aber auch in Datenerhebungen in der Biologie oder Medizin, auf. Sie bei der Modellbildung unberücksichtigt zu lassen, führt zu falschen Analysen und Schlussfolgerungen. Tests auf strukturelle Veränderungen, sogenannte Changepoints, sind daher von großer Bedeutung und Gegenstand aktueller Forschung. Die vorliegende Arbeit leistet einen Beitrag im Bereich der Changepointanalyse in nichtparametrischen heteroskedastischen Zeitreihenmodellen, wobei die Herleitung von Hypothesentests auf mögliche Changepoints im Mittelpunkt steht.

Betrachtet wird ein schwach abhängiger stochastischer Prozess $\{(Y_t, \boldsymbol{X}_t) \in \mathbb{R} \times \mathbb{R}^d : t \in \mathbb{Z}\}$, für den

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

gilt. Die Innovationen $(U_t)_{t\in\mathbb{Z}}$ erfüllen

$$E[U_t|\mathcal{F}^t] = 0,\ t \in \mathbb{Z}$$

fast sicher. Hierbei ist $\mathcal{F}^t := \sigma\{U_{j-1}, \boldsymbol{X}_j : j \leq t\}$, also die σ-Algebra, die alle vergangenen Informationen von U_j bis zum Zeitpunkt $t-1$ und von $\boldsymbol{X}_j$ bis zum Zeitpunkt t enthält. Damit ergibt sich für die unbekannte Regressionsfunktion $m_t(\cdot)$ gerade

$$m_t(\cdot) = E[Y_t|\boldsymbol{X}_t = \cdot],\ t \in \mathbb{Z}.$$

Sie wird daher auch als bedingte Erwartungswertfunktion bezeichnet. Der Index t legt nahe, dass diese strukturell vom Zeitpunkt der Beobachtung t abhängen kann. Der Hauptteil dieser Arbeit ist die Konstruktion eines Hypothesentests auf strukturelle Stabilität der Regressionsfunktion. Bei Beobachtung einer Stichprobe $(Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)$ der Größe $n \in \mathbb{N}$ soll die Nullhypothese

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

für ein $m : \mathbb{R}^d \to \mathbb{R}$, unabhängig von t, gegen die Alternative

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

für $m_{(1)}, m_{(2)} : \mathbb{R}^d \to \mathbb{R}$ mit $m_{(1)} \not\equiv m_{(2)}$, getestet werden. Die Regressoren $\boldsymbol{X}_t$ dürfen insbesondere endlich viele vergangene Werte von Y_t enthalten, sodass beispielsweise $\boldsymbol{X}_t = (Y_{t-1}, \ldots, Y_{t-d})^T$ möglich ist. Das Verfahren ist somit auf autoregressive Modelle anwendbar. Heteroskedastische Effekte sind ebenfalls erlaubt, da die zweiten Momente von U_t bedingt auf $\boldsymbol{X}_t$ von $\boldsymbol{X}_t$ abhängen dürfen. Die Teststatistik basiert auf dem markierten sequentiellen empirischen Prozess der Residuen

$$\hat{T}_n := \left\{ \hat{T}_n(s, \boldsymbol{z}) := \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor ns \rfloor} (Y_i - \hat{m}_n(\boldsymbol{X}_i))\omega_n(\boldsymbol{X}_i) I\{\boldsymbol{X}_i \leq \boldsymbol{z}\} : s \in [0,1], \boldsymbol{z} \in \mathbb{R}^d \right\}.$$

Die Gewichtsfunktion $\omega_n : \mathbb{R}^d \to \mathbb{R}$ ermöglicht es hierbei, gleichmäßige Konsistenzeigenschaften für den Nadaraya-Watson Schätzer $\hat{m}_n$ zu benutzen. Durch Verwendung nichtparametrischer Schätzmethoden werden zusätzliche fehlerhafte Modellspezifikationen vermieden. In Kapitel 2 werden neue gleichmäßige Konvergenzraten für die verwendeten nichtparametrischen Kernschätzer hergeleitet. Bei $\hat{T}_n$ handelt es sich um eine Modifikation der Teststatistik von Su und Xiao [71]. Sie konstruierten einen Changepointtest für die Regressionsfunktion in nichtparametrischen Zeitreihenmodellen unter Verwendung eines CUSUM Tests. Dieser ist unter bestimmten Alternativen jedoch nicht konsistent. Der Ansatz der markierten Residuen stammt von Burke und Bewa [6]. Sie benutzten ein ähnliches Verfahren in einem einfachen nichtparametrischen i.i.d. Regressionsmodell, um auf Strukturbrüche in der Regressionsfunktion zu testen. Diese Herangehensweise erfordert jedoch Methoden empirischer Prozesstheorie, welche im Zusammenhang mit abhängigen Daten schwerer zu handhaben sind. Das Hauptresultat dieser Arbeit ist Korollar 3.3. Es macht eine Aussage über das Grenzverhalten von $\hat{T}_n$ unter H_0 und strikter Stationarität. Im Gegensatz zum Resultat von Burke und Bewa [6] ergibt sich hier eine sehr einfache Grenzverteilung. Im Falle eindimensionaler Regressoren resultiert dies sogar in einer Grenzverteilung, die keinerlei unbekannte Parameter enthält. Für den Beweis werden einige technische Resultate, sowie eine schwache Konvergenzaussage für sequentielle empirische Prozesse induziert in Funktionenklassen und mit abhängigen Daten, benötigt. Letzteres ist in Anhang B als Korollar B.3 zu finden und ist in einer Allgemeinheit formuliert, die anderweitige Anwendungen ermöglicht. Desweiteren wird in Kapitel 4 ein Bootstraptest konstruiert, welcher auch bei fehlender Stationarität unter H_0 anwendbar ist. Insbesondere erlaubt dies das Testen auf Strukturbrüche in der Regressionsfunktion, auch bei auftretenden Varianzschwankungen.

Unter der Bedingung, dass es $\sigma_t^2 : \mathbb{R}^d \to \mathbb{R}$ gibt, sodass

$$E[U_t^2|\boldsymbol{X}_t] = \sigma_t^2(\boldsymbol{X}_t),\ t \in \mathbb{Z}$$

fast sicher gilt, ist σ_t^2 die bedingte Varianzfunktion, das heißt

$$\sigma_t^2(\cdot) = \mathrm{Var}(Y_t|\boldsymbol{X}_t = \cdot),\ t \in \mathbb{Z}.$$

In Kapitel 5 wird ein Test auf Changepoints in σ_t^2 beschrieben, welcher under der Annahme valide ist, dass es keine Strukturbrüche in der Regressionsfunktion gibt. Wie angemessen solch eine Annahme ist, kann mit Hilfe des Bootstraptests untersucht werden.

Kapitel 6 beinhaltet schließlich eine Simulationsstudie und die Anwendung der Tests auf zwei Datensätze. In den Monte-Carlo Simualtionen werden i.i.d. Regressionsmodelle, homoskedastische und heteroskedastische Autoregressionsmodelle, sowie weitere Zeitreihen betrachtet. Es werden sowohl die Changepointtests für die bedingte Erwartungswert- und Varianzfunktion, als auch der Bootstraptest, durchgeführt. Es stellt sich heraus, dass sie akzeptable Ergebnisse liefern und darüber hinaus einige einfache Alternativen deutlich besser erkennen als der CUSUM Test von Su and Xiao [71]. Für die Anwendung der Tests auf echte Daten wurde zum einen die jährliche Durchflussmenge des Nil in Aswan zwischen 1871 und 1970, und zum anderen die wöchentlichen Schlusskurse des Dow Jones Industrial Average (DJIA) Indexes zwischen 1971 und 1974, betrachtet. Beide Datensätze wurden bereits oftmals in der wissenschaftlichen Literatur untersucht. Die neuen Tests zeigen einen Changepoint im bedinten Erwartungswert im Nildatensatz und einen Changepoint in der bedingten Varianz im DJIA-Datensatz an. Diese Ergebnisse sind konsistent mit existierenden Studien.

Publications derived from this dissertation

The weak convergence result stated in Appendix B can be found as a preprint on arxiv (see [52]).

Danksagung

An dieser Stelle möchte ich den Menschen danken, die mich bei der Erstellung dieser Arbeit unterstützt haben. Zuallererst möchte ich Prof. Dr. Natalie Neumeyer für die hervorragende Betreuung danken. Neben der Anregung zum Thema hat sie mit vielen neuen Ideen und hilfreichen Diskussionen maßgeblich zum Gelingen meines Promotionsvorhabens beigetragen und darüber hinaus stets eine angenehme Arbeitsatmosphäre geschaffen. Dafür danke ich ihr sehr. Allen Mitarbeitern des Bereichs Mathematische Statistik und Stochastische Prozesse der Universität Hamburg danke ich für ihre Hilfsbereitschaft und die vielen anregenden Gespräche. Sie alle haben meinen Arbeitsalltag bereichert. Prof. Dr. Claudia Kirch möchte ich herzlich danken, für die Erstellung des Zweitgutachtens und für die Einladung ins Oberseminar zur Stochastik der Universität Magdeburg. Ihr und allen Mitarbeitern des Instituts für Mathematische Stochastik der Universität Magdeburg möchte ich in diesem Zusammenhang für die vielen hilfreichen Kommentare zu meinem Vortrag danken. Des Weiteren danke ich meinen Kollegen Daniel Ebel, Nick Kloodt, Miran Knežević, Dr. Sonja Otten, Tobias Sohr, Rui Wu und meiner Freundin Teresa Kraft für die Mühe, die sie sich beim Korrekturlesen dieser Arbeit gemacht haben. Weiter möchte ich meinen Eltern Uta und Ferdinand Mohr und meinen Großeltern Rita und Lothar Mohr danken, die mich nicht nur finanziell während des Studiums unterstützt haben, sondern mir auch danach stets mit Rat und Tat zur Seite standen. Zu guter Letzt möchte ich meinen zwei liebsten Menschen danken. Mein Freund Simon Kufeld hat mich trotz der damit verbundenen Entbehrungen stets in meinen Vorhaben unterstützt und mich motiviert. Meine Schwester Judith Mohr ist mit ihrem unermüdlichen Zuspruch und ihrer Fürsorge auch in Zeiten der Zweifel immer eine mentale Stütze gewesen. Dafür bin ich ihnen von Herzen dankbar.