Cumulated sum processes of residuals for goodness-of-fit tests in linear regression models

Zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

bei der Mathematisch-Geographischen Fakultät
der Katholischen Universität Eichstätt-Ingolstadt
eingereichte und genehmigte

Dissertation

von

M. Sc.
Thomas Heindl

Katholische Universität
Eichstätt-Ingolstadt

Eichstätt im August 2022

Referent:	Prof. Dr. Wolfgang Bischoff
Korreferent:	Prof. Dr. Frank Miller
Tag der mündlichen Prüfung:	19. Juli 2022

Danksagung

An erster Stelle gilt mein Dank meinem Doktorvater, Herrn Prof. Dr. Wolfgang Bischoff, für die wissenschaftliche Betreuung dieser Arbeit. Er zeigte mir das Thema auf und beriet mich mit seinem Fachwissen und seiner langjährigen wissenschaftlichen Erfahrung.

Bei Herrn Prof. Dr. Frank Miller bedanke ich mich für die bereitwillige Übernahme des Korreferats.

Mein besonderer Dank gilt all jenen, die auf die eine oder andere Weise zum Gelingen dieser Doktorarbeit beigetragen haben. Angefangen bei allen Lehrkräften, Kommilitonen und Kommilitoninnen, Übungsleitern und Übungsleiterinnen und Professoren die an meiner mathematischen Ausbildung beteiligt waren, über all jene, die durch ihren fachlichen Rat oder in Form von Verbesserungsvorschlägen direkt auf die Arbeit eingewirkt haben, bis hin zu meiner Lebensgefährtin, meiner Familie und meinen Freundinnen und Freunden, die mich stets unterstützt und motiviert haben.

Diese Doktorarbeit entstand im Laufe meiner Lehr- und Forschungstätigkeit als wissenschaftlicher Mitarbeiter am Lehrstuhl für Mathematik – Statistik der Katholischen Universität Eichstätt-Ingolstadt. Daher möchte ich zuletzt meinen Kolleginnen und Kollegen sowohl des wissenschaftlichen als auch des wissenschaftsunterstützenden Bereichs für die gute Arbeitsatmosphäre und ihre offenen Ohren bei Fragen und Problemen danken.

Eichstätt, im April 2022

Contents

List of Figures

List of abbreviations

a.a.	**a**lmost **a**ll
a.e.	**a**lmost **e**verywhere
a.s.	**a**lmost **s**urely
càdlàg	**c**ontinue **à** **d**roite, **l**imite **à** **g**auche (French), continuous on the right, limit on the left
CUSUM	**cu**mulative **sum**
fidis	**fi**nite-**di**mensional **dis**tributions
iid	**i**ndependent and **i**dentically **d**istributed
LSE	**l**east-**s**quares **e**stimator
RKHS	**r**eproducing **k**ernel **H**ilbert **s**pace

List of symbols

Symbol	Meaning
$B = (B(t))_{t\in[0,1]}$	Standard Brownian motion, p. 30
$B^0 = (B^0(t))_{t\in[0,1]}$	Brownian bridge on $[0,1]$, p. 78
$B_{f,F}$	Residual partial sum limit process with regression function f and limit design distribution F, p. 32
$Ber(p)$	Bernoulli distribution with success probability $p \in [0,1]$, p. 71
$BV_D(\mathcal{H}_B)$	Subspace of functions in $\mathcal{H}_B$, whose derivatives are of bounded variation, p. 65
$Cov(X)$	Covariance of a random element X, p. 17
$C(I)$	Set of continuous functions on I, p. 30
$D(I)$	Skorokhod space on a set I, p. 15
F^-	Quantile function associated to a distribution function F, p. 21
F_n	Empirical distribution function, p. 30
F_z	Distribution function F, constant from value $z \in [0,1]$ onwards, p. 31
$\overline{F}_n$	Empirical distribution function of a random sample $U_1, \ldots, U_n \overset{iid}{\sim} U[0,1]$, p. 77
$\mathrm{GL}(n, \mathbb{R})$	General linear group over $\mathbb{R}$ of degree n, p. 118
H_0	Null hypothesis, p. 2
H_1	Alternative hypothesis, p. 2
$H(R)$	RKHS with respect to the reproducing kernel R, p. 58
I_n	Identity matrix of size $n \times n$, p. 85
$Im(f)$	Image of a mapping f, p. 143
$N_n(\mu, \Sigma)$	n-dimensional normal distribution with mean vector $\mu \in \mathbb{R}^n$ and positive definite covariance matrix $\Sigma \in \mathbb{R}^{n\times n}$, p. 17
R_n	Marked empirical process, where the marks are the regression errors, p. 13
$\overline{R}_n$	Uniform empirical process of residuals, p. 99
R_n^1	Marked empirical process, where the marks are the residuals, p. 13
$\overline{R}_n^1$	Uniform marked empirical process, p. 24
$\widetilde{R}_n^1$	Marked empirical process of the vector of backwards recursive

	least-squares residuals r_n, p. 95
R_∞	Limit process of the marked empirical process R_n, p. 15
R^1_∞	Limit process of the marked empirical process R^1_n, p. 18
$\overline{R}^1_\infty$	Limit process of the uniform marked empirical process $\overline{R}^1_n$, p. 24
$S(R)$	Space of finite linear combinations of a covariance kernel R, p. 57
$\overline{S(R)}$	Closure of the set $S(R)$, p. 60
T_n	Partial sum operator, p. 30
T^*	Khmaladze transformation in regression models, p. 90
$U(I)$	Uniform distribution on an interval I, p. 9
V_n	Empirical process, p. 77
$\overline{V}_n$	Uniform empirical process, p. 77
$V_{n,\hat{\theta}_n}$	Parametric empirical process, p. 78
$\overline{V}_{n,\hat{\theta}_n}$	Uniform parametric empirical process, p. 79
$\overline{V}_{f,\theta_0}$	Limit process of the uniform parametric empirical process $\overline{V}_{n,\hat{\theta}_n}$, p. 80
$W_{\mathcal{H}_B}$	Subspace of $\mathcal{H}_B$ that is spanned by the integrated regression functions, p. 64
$W^\perp_{\mathcal{H}_B}$	Orthogonal complement of the space $W_{\mathcal{H}_B}$, p. 64

$adj(M)$	Adjugate matrix of a square matrix M, p. 195
$\det(M)$	Determinant of a square matrix M, p. 184
id_I	Identity function on a set I, p. 22
$int(S)$	Interior of a set S, p. 17
log	Natural logarithm, p. 87
$m(x), m(x,\theta)$	Regression function of a general regression model, p. 1
$o_{\mathbb{P}}(1)$	Stochastic Laundau notation small o, p. 17
pr_S	Orthogonal projector onto the space S, p. 65
r_n	Vector of backwards recursive least-squares residuals, p. 94
$rank(J)$	Rank of a matrix J, p. 20
$span(S)$	Linear span of the set of vectors S, p. 2
$supp(F)$	Support of a distribution function F, p. 31
$supp(X)$	Support of a random variable X, p. 31
$supp(g), supp(\mu_g)$	Support of a function g that is right continuous and of bounded variation or the signed measure μ_g associated to g respectively, p. 34

$\hat{\beta}_n$	Least-squares estimator for the regression coefficients β in a linear model with fixed design, p. 29
$\delta_{\{x\}}$	Dirac measure at point x, p. 187
$\varepsilon(x)$	Random error component of a regression model, p. 1
$\hat{\varepsilon}_n$	Vector of least-squares residuals in a linear regression model with fixed design, p. 30
Θ	$\subseteq \mathbb{R}^p$, Set of regression parameters, p.12
θ_0	$\in \mathbb{R}^p$, true but unknown regression coefficients in a regression model with random design (existent under H_0), p. 1
$\hat{\theta}_n$	Estimator for θ_0, p. 1
$\lambda\!\!\lambda$	Lebesgue measure on the real line, p. 22
$\lambda\!\!\lambda_I$	Lebesgue measure on an interval I, p. 22
$\sigma(f)$	σ-algebra generated by a mapping f, p. 22
$\mathcal{X}_{\min}(A)$	Minimal eigenvalue of a symmetric matrix S, p. 186
$\mathcal{X}_{\max}(A)$	Maximal eigenvalue of a symmetric matrix S, p. 186
$\mathbb{E}(X)$	Expected value of a random element X, p. 2
$\mathbb{E}(\varepsilon\|X)$	Conditional expectation of a random element ε given a random element X, p. 12
$\mathbb{E}(Y\|X=x)$	Factorisation of the conditional expectation of a random element Y given a random element X at point $X=x$, p. 12
$\mathbb{N}$	Set of natural numbers $\{1,2,3,\dots\}$, p. 1
$\mathbb{P}$	Probability measure, unless otherwise specified, defined on a measurable space $(\Omega,\mathcal{A})$, p. 17
$\mathbb{P}^X$	Distribution of a random element X or pushforward measure of the measure $\mathbb{P}$ using the measurable mapping X, p. 22
$\mathbb{R}$ resp. $\mathbb{R}^n$	Set of real numbers, resp. n-dimensional real vectors, p. 3
$\mathbb{R}^{m\times n}$	Vector space of real $m\times n$ matrices, p. 24
$\overline{\mathbb{R}}$	$\mathbb{R}\cup\{-\infty,\infty\}$, p. 28
$\mathbb{V}ar(X)$	Variance of a random element X, p. 2
$\mathbb{V}ar(\varepsilon\|X=x)$	Factorisation of the conditional variance of a random element ε given a random element X at point $X=x$, p. 12
$\mathbb{Z}$	The set of integers $\{\dots,-2,-1,0,1,2,\dots\}$, p. 30
$\mathcal{B}$	Borel-σ-Algebra on $[0,1]$, p. 22
$\mathcal{B}_I$	Borel-σ-Algebra on the measurable set I, p. 22

$\mathcal{E}$	Experimental region of a linear regression model with fixed design, p. 29
$\mathcal{F}$	Hypothesised family of distribution functions, p. 76
$\mathcal{H}_B$	RKHS of the standard Brownian motion, p. 63
$\mathcal{L}^2(I, f)$	Space of square-integrable functions on the interval I with respect to the integrator f, p. 24
$\mathcal{L}^2(\mathbb{P})$	Space of square-integrable random elements with respect to a probability measure $\mathbb{P}$, p. 197
$\mathcal{M}$	Hypothesised set of potential regression functions, p. 2
$f \circ g$	Composition of the mappings f and g, p. 22
$\xrightarrow[n\to\infty]{a.s.}$	Almost sure convergence, p. 118
$\xrightarrow[n\to\infty]{\mathcal{D}}$	Convergence in distribution resp. weak convergence, p. 17
$\xrightarrow[n\to\infty]{\mathcal{L}^2(\mathbb{P})}$	Convergence in $\mathcal{L}^2(\mathbb{P})$, p. 197
$\xrightarrow[n\to\infty]{\mathbb{P}}$	Stochastic convergence, p. 17
$\sim$	"is distributed like", p. 9
$\overset{\mathcal{D}}{=}$	Equality in distribution, p. 159
$\mathbb{1}_S$	Indicator function of a set S, p. 13
$\langle \cdot, \cdot \rangle_S$	Inner product. The corresponding vector space S is either given or unambiguous from the context, p. 24
$\int_I g \, dF$	Measure integral of a measurable function g over interval I with respect to the measure associated to the distribution function F, p. 15
$\overset{(R)}{\int_I} f \, dg$	Riemann-Stieltjes integral of a function f over an interval I with respect to a function g, p. 33
$\neg P$	Negation of a proposition P, p. 126
$\|f\|_\infty$	Supremum norm of a function f, p. 33
$\|v\|_S$	Norm of the vector $v \in S$. S is either given or unambiguous from the context, p. 60
$\|f\|_V$	Total variation of a function f, p. 65
$\frac{\partial f}{\partial x}$	Partial derivative of a function f with respect to one of its variables x, p. 17
$\lfloor x \rfloor$	Largest integer that is smaller than or equal to x, p. 30
$\lceil x \rceil$	Smallest integer that is bigger than or equal to x, p. 108
$\prec$	Smaller with respect to the Loewner order, p. 201

1 Introduction

Collecting and analysing data plays a crucial role in almost all areas of life. In mathematical statistics, mathematical models are used to deduce knowledge from the collected data. In this thesis we shall study aspects of regression models, which are some of the most commonly used statistical models. In particular, we will mainly deal with linear regression models, which are widely used and popular in practice as they allow a relatively intuitive interpretation.

We start with an introduction to what we mean by a regression model. Consider a model of the form

$$Y(x) = m(x) + \varepsilon(x),$$

where x is called the independent or input or design variable and $Y(x)$ is the dependent or outcome variable. The model assumes that there is a functional deterministic relationship $x \mapsto m(x)$ between the input x and the dependent variable $Y(x)$. Therefore, Y is a function of x. However, the functional deterministic relationship $m(x)$ is unknown, since the outcome $Y(x)$ depends not only on $m(x)$, but also on a random error component $\varepsilon(x)$, which is in general unobservable. Thus, the above model explains the outcome $Y(x)$ as the sum of a deterministic functional relationship $m(x)$ and an unobservable random error $\varepsilon(x)$ on the basis of an explanatory variable x.
Although the deterministic functional relationship $x \mapsto m(x)$ cannot be specified according to the model above, it is precisely this relationship that is of interest in practice because it describes how the explanatory variable x influences the outcome $Y(x)$ in a systematic manner. In order to make well-founded statements about the functional relationship $m(x)$, data about $m(x)$ is collected indirectly, by looking at $Y(x)$. To be more precise, one measures or fixes various design points $x_1 \leq \cdots \leq x_n$ and measures the corresponding outcome values $Y(x_1), \ldots, Y(x_n)$. This experiment can be described as the regression model

$$Y_i = m(x_i) + \varepsilon_i, \quad i = 1, \ldots, n,\ n \in \mathbb{N}, \tag{1.1}$$

where we use the notations $Y_i := Y(x_i)$, $\varepsilon_i := \varepsilon(x_i)$, $i = 1, \ldots, n$. In this thesis, for simplicity, we assume that the design variable x is univariate, i.e. $x \in \mathbb{R}$. In order to be able to examine this regression model for the functional relationship $x \mapsto m(x)$, the following assumptions are made:

- The unknown deterministic functional relation $x \mapsto m(x)$, which is called the regression function m, is assumed to belong to a given class $\mathcal{M}$ of functions. If one additionally assumes that m is a linear combination of finitely many known functions, i.e. $\mathcal{M} = span\{f_1, \ldots, f_p\}$, then one speaks of a linear regression model.
- The random regression errors $\varepsilon_1, \ldots, \varepsilon_n$ are usually assumed to be independent and identically distributed (iid). In this case, one speaks of a homoscedastic regression model. If, in contrast, the random error component $\varepsilon(x)$ is such that $\mathbb{V}ar(\varepsilon(x))$ is not constant, one speaks of a heteroscedastic regression model. Furthermore, we assume without loss of generality that m is chosen such that $\mathbb{E}(\varepsilon_1) = \cdots = \mathbb{E}(\varepsilon_n) = 0$.

In this thesis, we will study regression models that are slightly generalised compared to model (1.1). To be more specific, we will assume that the design points are indexed by $n1, \ldots, nn$. That is, we assume a triangular array $((x_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of design points with $x_{n1} \leq \cdots \leq x_{nn}$ for $n \in \mathbb{N}$. The design points $x_1 \leq \cdots \leq x_n$ of the regression model (1.1) are thus considered to constitute the n-th row of $((x_{ni})_{i=1}^n)_{n\in\mathbb{N}}$. This notation is convenient because in this thesis we will conduct asymptotic investigations, i.e. we will consider sequences of designs with an increasing number of observation points. We are therefore concerned with regression models of the form

$$Y_{ni} = m(x_{ni}) + \varepsilon_{ni}, \quad i = 1, \ldots, n, \ n \in \mathbb{N}, \tag{1.2}$$

or, more specifically, primarily with linear regression models

$$Y_{ni} = \sum_{j=1}^{p} f_j(x_{ni})\theta_j + \varepsilon_{ni}, \quad i = 1, \ldots, n, \ n \in \mathbb{N}, \ \theta_1, \ldots, \theta_p \in \mathbb{R}, \ p \in \mathbb{N}. \tag{1.3}$$

Regression models (1.2) and (1.3) constitute the focus of numerous research works, both classical and contemporary. Some of the problems considered are, for example, how to estimate the regression function m, how accurate these estimations are, how to obtain reliable predictions for future observations and which designs are advantageous for answering certain questions, to name but a few. For references and elaborations on these (and many other) problems see, for e.g. the textbooks Searle (1971) and Hocking (2013) on linear regression models and the literature cited there.
However, one essential question in this context is whether the assumed class of potential regression functions $\mathcal{M}$ is adequately chosen, i.e. it must be checked whether m actually belongs to the hypothesis class $\mathcal{M}$. If this were not the case, all further analyses (such as estimating the regression function $m \in \mathcal{M}$, for example) would be error-ridden from the very start. Stute (1997) states in this context that "... in order to prevent wrong conclusions, every statistical inference that is based on a model $\mathcal{M}$ should be accompanied by a proper model check, that is, by a test for $H_0 : m \in \mathcal{M}$ versus $H_1 : m \notin \mathcal{M}$". Some early research in this context includes, for example Eubank and Spiegelman (1990), Firth, Glosup, and Hinkley (1991), Eubank and Hart (1992) and Müller (1992). Therein, such

model checks are referred to as full model checks or goodness-of-fit tests for regression models as they try to investigate the question whether the assumed class of potential regression functions is suited to fit the data or not. This type of statistical question, in the context of (primarily linear) regression models, is the core area of study of this thesis.

The following example gives a first glimpse of how a goodness-of-fit analysis is carried out. Figure 1.1 illustrates a fictional scenario of an investigation of the braking distance of a certain car dependent on the car's speed at the moment when deceleration starts. In this example, twenty pairs of data were collected and two different linear regression models were fitted. In a first attempt, the blue regression function was estimated, assuming the hypothesis class of potential regression functions to be $\mathcal{M}_1 := \{f(x) = ax + b \,|\, a, b \in \mathbb{R}\}$, while in the second attempt, the red regression function was estimated using the hypothesis class of potential regression functions $\mathcal{M}_2 := \{f(x) = ax^2 \,|\, a \in \mathbb{R}\}$. One notices that the red parabola fits the given data better than the blue straight line.

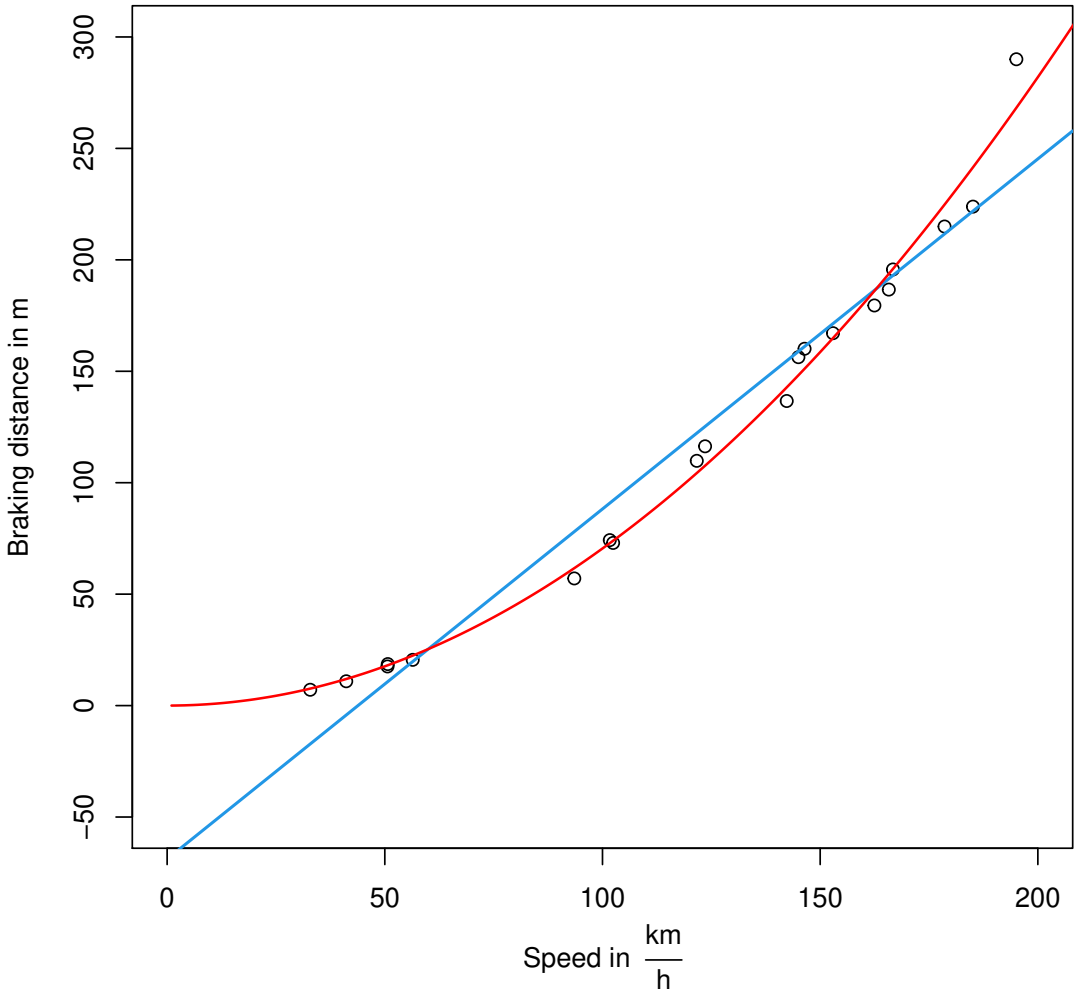


Figure 1.1: Fictional data set of 20 braking distances and speeds as well as the graphs of the estimated regression line or regression parabola.

A technique commonly used to investigate whether the hypothesis class of potential regression functions $\mathcal{M}$ is adequately specified is the examination of the so-called cumulated sum (CUSUM) processes of regression residuals. Figure 1.2 shows such a residual CUSUM process for each of the two models in the example above. One notices that the (blue) CUSUM process belonging to the straight line regression model takes on more extreme values than the (red) CUSUM process belonging to the fitted parabola.

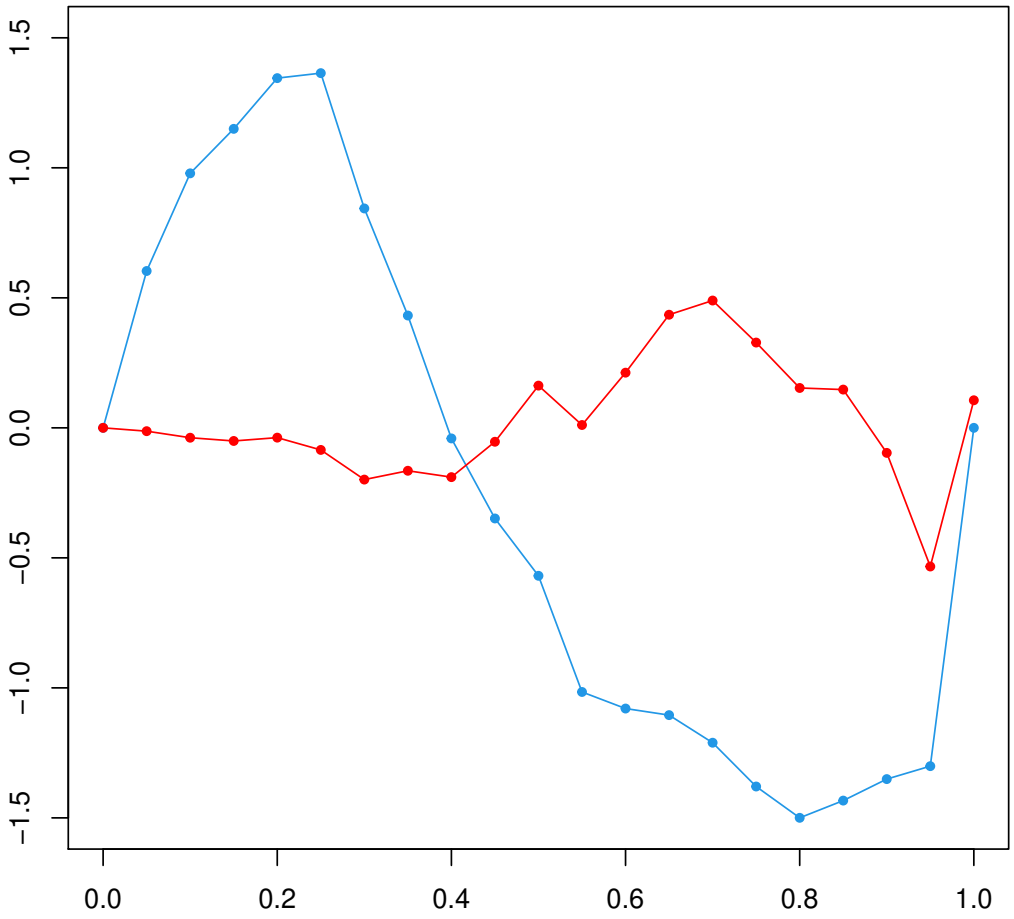


Figure 1.2: CUSUM processes belonging to the regression functions fitted in Figure 1.1.

In the field of artificial intelligence, machine learning has seen ever expanding use over the last decade due to overwhelming practical success in applications. This has happened even beyond provable theoretical expectations. Examples of this can be found in Krizhevsky, Sutskever, and Hinton (2012), Mnih et al. (2013) and He et al. (2016).
It is worth pointing out that the theoretical expressiveness of "large" hypothesis classes $\mathcal{M}$ can be inappropriate in some practical applications. Such "large" hypothesis classes can be found, for example, in neural networks: For instance, Leshno et al. (1993) prove that a (sufficiently large) standard feedforward network with a continuous activation function can approximate any continuous function to any degree of accuracy (in the uniform norm on compact sets) if and only if the network's activation function is not a polynomial. Among others, the following reasons are arguments why "large" hypothesis classes $\mathcal{M}$ may be

inappropriate in practice.

i) A simple interpretation of the relation between independent and dependent variables is desired.

ii) Processing power and storage space are prohibitively expensive.

iii) There is not enough training data available to successfully train such large models.

Therefore, ideally, one would like to optimize within a hypothesis class $\mathcal{M}$ that is relevant from a practical perspective and that is "just large enough" for the task at hand. This strict dependency on the concrete problem to be solved necessitates the use of simpler models and goodness-of-fit tests to decide whether $H_0 : m \in \mathcal{M}$ or $H_1 : m \notin \mathcal{M}$ is true. Linear regression models in particular are a fundamental part of machine learning and popular with practitioners as they are resource-efficient and relatively easy to interpret. A recent review of linear regression in machine learning can be found, for example, in Maulud and Abdulazeez (2020). For some concrete examples of the application of linear regression models in machine learning, see e.g. Schuld, Sinayskiy, and Petruccione (2016) or Kim et al. (2020).
We can therefore conclude that goodness-of-fit tests — especially for linear regression models — play an important role in checking machine learning algorithms for applicability or in determining whether the structure of the data has changed.

Structure and results of this work

Chapter 2 At the beginning of Chapter 2, we address the fundamental difference between random and fixed experimental designs and illustrate this difference with a fictional example of two linear regression models.
On the one hand, we study in Section 2.1.1 marked empirical processes of regression residuals in regression models with random designs on the example of a paper by Winfried Stute, see Stute (1997). We consider well-known results on the limit process R^1_∞ of the marked empirical process R^1_n of the regression residuals with respect to goodness-of-fit tests. Section 2.1.2 provides a formal proof of the result that linear regression models with random designs using the least-squares estimator (LSE) are special cases of general regression models with random designs. Finally, in Theorem 2.1.12 of Section 2.1.3, we make precise the statement that in a linear regression model with random design, one can assume (to a certain extent) a uniform design on $[0, 1]$.
On the other hand, we study in Section 2.2.1 residual partial sum processes of the regression residuals in regression models with fixed designs on the example of a paper by Wolfgang Bischoff, see Bischoff (1998). We consider well-known results on the limit process $B_{f,F}$ of the residual partial sum process $\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)$ of the regression residuals with respect to goodness-of-fit tests. Section 2.2.2 contains several results previously unknown. The main result is Theorem 2.2.9, which states that when only the distribution of the

residual partial sum limit process is of interest, one can always assume the design of a linear regression model with fixed design to be the equidistant design on the unit interval $[0, 1]$. On the way to this theorem, we state and prove the technical Lemma 2.2.7, which provides a new characterisation of bounded variation functions. This allows us to state in Theorem 2.2.11 a generalisation of Theorem 2.2 from Bischoff (1998).

Chapter 3 In this chapter, we examine common properties of marked empirical processes of regression residuals in linear regression models with random designs and residual partial sums processes in linear regression models with fixed designs, as well as differences between them. Particular attention is paid to the limit distributions of the marked empirical process R_n^1 as well as the residual partial sum process $\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)$ of the regression residuals — both are collectively referred to as residual CUSUM processes — as it is common to base test statistics of asymptotic goodness-of-fit tests on them.
The main results of Section 3.1 are Theorem 3.1.1 and Corollary 3.1.2, which are new in the literature and state the equality of the respective residual CUSUM processes under certain conditions.
In Section 3.2, we then proceed by identifying the underlying common properties between both residual CUSUM processes and their respective regression designs. From this analysis, we derive the so-called generic linear regression model in Definition 3.2.1 and establish the central statement of this work. This statement is new in the literature and reads as follows:

> When it comes to goodness-of-fit tests in linear regression models based on the asymptotic distribution of residual CUSUM processes, one can assume without loss of generality a generic linear regression model.

Therefore, henceforth any result about goodness-of-fit tests applicable to residual partial sum processes in linear models with fixed designs can also be used for marked empirical processes in linear regression models with random designs, and vice versa.

Chapter 4 This chapter is mainly concerned with a geometric interpretation of residual CUSUM limit processes as projections onto certain reproducing kernel Hilbert spaces (RKHS). We start in Section 4.1 with some general and well-known definitions and results on RKHS of stochastic processes. We then focus on Brownian motion RKHS. In Section 4.2, we use these concepts to identify the residual partial sum limit processes $B_{f,\lambda_{[0,1]}}$, in the case of a generic linear regression model, as the orthogonal projection $pr_{W_{\mathcal{H}_B}^{\perp}}(B(\cdot))$ of a standard Brownian motion onto the orthogonal complement of the space spanned by the integrated regression functions within the RKHS associated to the standard Brownian motion. This result is stated in Theorem 4.2.3, which is new in the literature as it generalises Theorem 3.2 in Bischoff (2002).

Chapter 5 This chapter builds on the previous findings of this thesis and addresses the work of MacNeill, Mao, and Xie (1994). It contains two theorems. In Theorem 5.0.2, we state and prove a functional central limit theorem for the partial sum processes of

heteroscedastic and independent regression errors. In Theorem 5.0.3, the same is done for residual partial sum processes in heteroscedastic regression models. Although functional central limit theorems in the given situation are already known and intuitive to some degree, the proof of Theorem 5.0.2 that we present is new and utilises theorems of Prokhorov and Rubin. Furthermore, we prove this theorem for triangular arrays of random variables and under slightly weaker assumptions than those implicitly made in MacNeill, Mao, and Xie (1994) about sequences of random variables. The proof of Theorem 5.0.3 generalises Theorem 3.2 in Bischoff (2002).

Chapter 6 In this chapter, we discuss a fundamental problem of goodness-of-fit tests when the estimation of parameters is necessary. This is often referred to as the "Durbin problem". This problem did not initially arise in the context of goodness-of-fit tests in regression models, but in the context of testing whether or not a sample of random variables was taken from a distribution belonging to a particular class of distribution functions. In this context, tests based on the empirical distribution functions of random samples are considered and therefore, empirical processes are studied.
We will start our exploration of the Durbin problem in Section 6.1.1 in the context of empirical processes. In Section 6.1.2, we will then — also in the context of empirical processes — briefly recall a well-known way of dealing with such problems. This method goes back to Khmaladze (1980) and Khmaladze (1982). In Section 6.2 we show how Stute, Thies, and Zhu (1998) have applied this solution in the context of goodness-of-fit tests in regression models.
Finally, in Section 6.3, in the situation of linear regression models, we state and prove a theorem that makes precise the interpretation of the Khmaladze transformation as a continuous-time backwards recursive least-squares method. We note that, while this theorem and its proof are new in the literature, Bai (2003) already gives a non-rigorous argument for this interpretation of the Khmaladze transformation in the context of time series data.

Chapter 7 In order to facilitate the reading of this thesis, extensive or particularly technical proofs from the above-mentioned chapters are deferred to the appendix (Chapter 7). We emphasise this at the appropriate places in the respective chapters. In each case, a reference to the corresponding place in the appendix is provided.

Used software

All graphics contained in this thesis were created by the author using the statistical programming language R (R Core Team (2020)).

2 Cumulated sum processes of residuals in regression models

In statistics, the study of the dependency of a variable Y on a variable X can be classified based on the way in which the observations $x_1, \ldots, x_n$ of the independent variable X are obtained.
If the observations $x_1, \ldots, x_n$ are realisations of iid random variables $X_1, \ldots, X_n \sim X$, one speaks of a **random design**. An example for such a design is obtained by taking the realisations of the random variables $U_1, U_2, \ldots, U_n \overset{iid}{\sim} U[0,1]$ as observation points. Alternatively, if the observations $x_1, \ldots, x_n$ are non-stochastic (deterministic), one speaks of a **fixed design**. The most common example for a fixed design is the equidistant design, in which the observations are sampled at equidistant points, for example, $\frac{1}{n}, \frac{2}{n}, \ldots, 1$ in the experimental region $[0,1]$.

In particular, this fundamental distinction of study designs applies to the special case of regression models. The literature on full model checks for regression models via analysis of regression residuals distinguishes between two design types and two residual CUSUM processes. Namely, there are

- regression models with random designs in which the design points themselves are supposed to be realisations of independent and identically distributed random variables of a known or unknown distribution. When collecting data, both the realisations of the design points and the corresponding realisations of outcome values are observed and recorded. In the dataset obtained, a regression of the outcome on the design points is performed and the fit is analysed using marked empirical processes of the regression residuals.
- regression models with fixed designs in which the design points are non-stochastic but purposefully planned. Here, the outcome values are observed and recorded at these fixed design points. In the obtained dataset, a regression of the outcome on the design points is performed and the fit is analysed using residual partial sum processes.

As a first example, we shall consider regression models with both designs in the fictional scenario of an investigation of the braking distance of a certain car dependent on the car's speed at the moment when deceleration is started.

Under the random design paradigm, a study with the aim of investigating this dependency could be conducted as follows. A certain number of times, e.g. 40 times, the same driver is asked to evenly accelerate in a straight line from 0 km/h onwards until he or she freely decides to start an emergency braking. Then, the speed of the car at the moment the deceleration begins and the braking distance (to standstill) are measured and recorded. In the end, one gets a dataset of 40 realisations of random speeds and braking distances and a regression of the braking distance on the speed is performed. Figure 2.1 shows a possible data set obtained in this way.

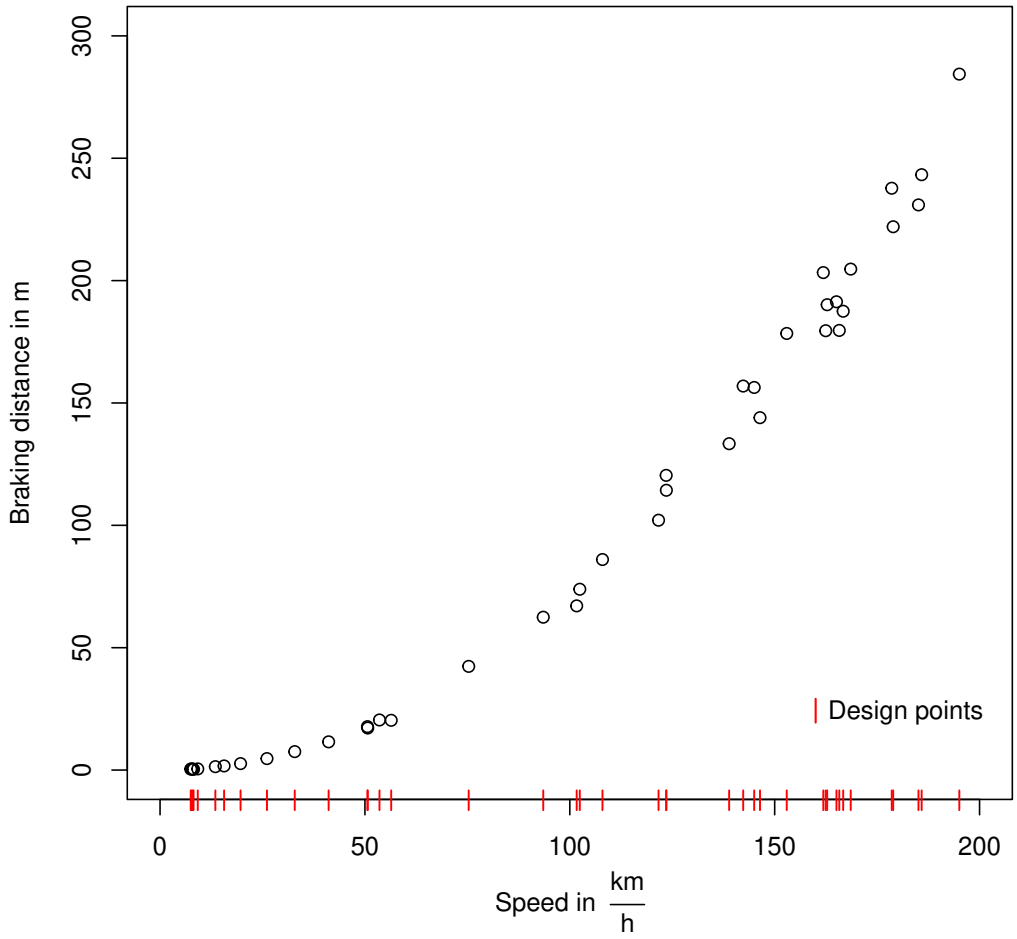


Figure 2.1: Fictional data set of 40 braking distances and speeds obtained under the random design paradigm with a uniform design distribution.

Under the fixed design paradigm, a certain number of fixed speeds, e.g. 40 speeds, between 0 km/h and 200 km/h would be pre-determined in advance. Then, for each of these 40 speeds, the same driver would accelerate in a straight line from 0 km/h onwards until he or she reaches the fixed speed and then would immediately start an emergency braking. Again the speed of the car at the moment the car started decelerating and the braking distance are measured and recorded and a regression of the braking distance on the speed is performed. See Figure 2.2 for a hypothetical data set obtained in this way.

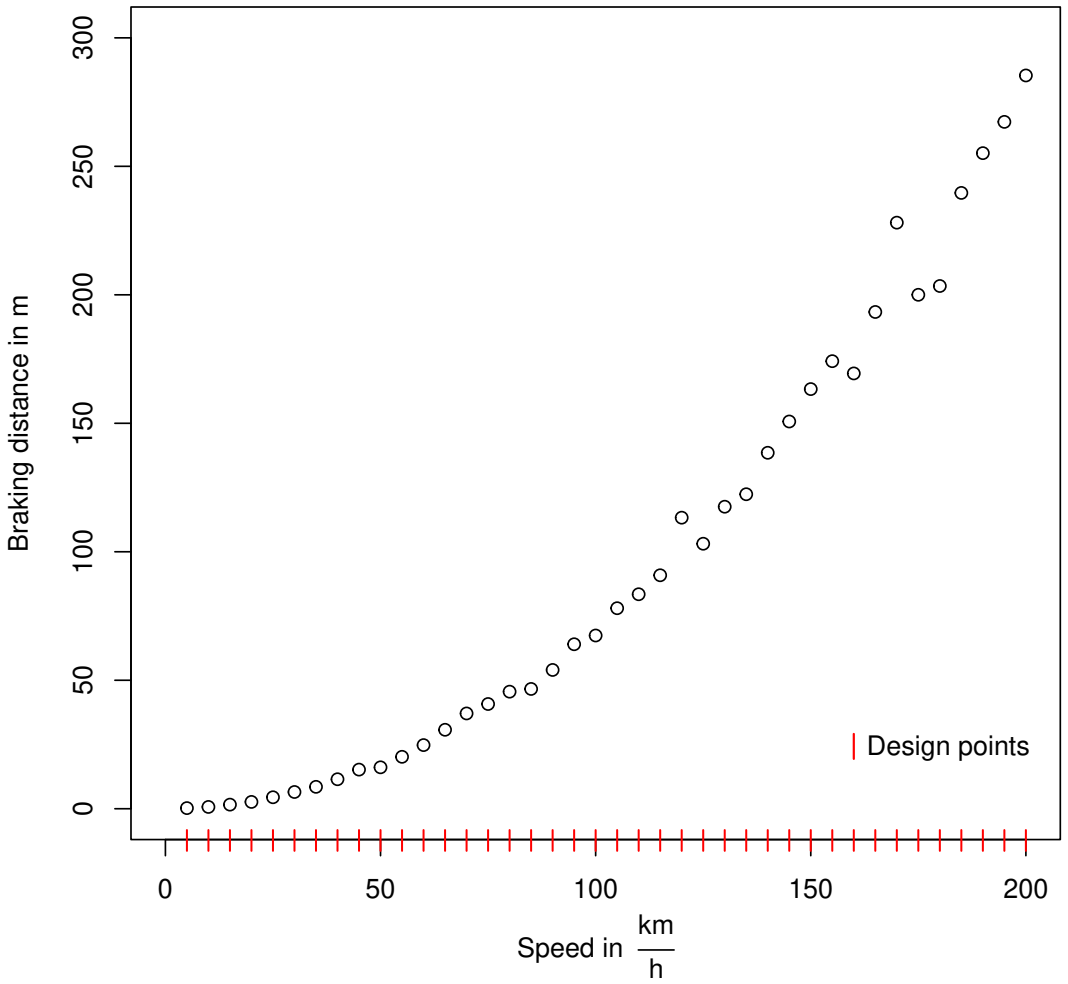


Figure 2.2: Fictional data set of 40 braking distances and speeds obtained under the equidistant fixed design paradigm.

Note that we assume that the repeated emergency brakings do not influence the condition of the car or the driver and that the reaction times of the driver do not matter in this case, since we always note the speed and the breaking distance starting with the point at which the car starts decelerating.
Note furthermore, that since in this example only one explanatory variable (the spped of the car) is given, a regression of the outcome (the braking distance) on this explanatory variable is called univariate.

2.1 Marked empirical processes and random designs

This section consists of three parts: In the first part, we study marked empirical processes of regression residuals in regression models with random designs on the example of a paper by Winfried Stute, see Stute (1997), and state their well known limit residual processes. In the second part, we shall prove that linear regression by means of the LSE on a random design is a special case of a regression model with random design. Finally, in the third part, we specify and prove the statement that in a regression model with random design, one can assume to a certain extent that the design is the uniform distribution on $[0, 1]$.

2.1.1 Results of Stute (1997)

A regression model with random design has the form

$$Y = m(X) + \varepsilon.$$

Here,

$$m(x) := \mathbb{E}(Y \,|\, X = x)$$

constitutes the true but unknown regression function that depends on the input of a random variable X and consequently

$$\mathbb{E}(\varepsilon \,|\, X) = 0\,.$$

Since we always assume Y to be integrable,

$$m(x) = \mathbb{E}(Y \,|\, X = x)$$

exists. The output or outcome Y is the sum of the regression function m and the regression error ε with

$$\mathbb{V}ar(\varepsilon \,|\, X = x) = \sigma^2(x) > 0\,.$$

To simplify the analysis, we only consider univariate X and Y. Hence, we concern ourselves with univariate regression models.
In order to obtain full model checks it is assumed that m belongs to a function class $\mathcal{M}$ whose elements are uniquely identified by a parameter θ. Hence, the model hypothesis is

$$\mathrm{H}_0 : \; m \in \mathcal{M} := \{\, m(\cdot, \theta) \,|\, \theta \in \Theta\}, \quad \Theta \subseteq \mathbb{R}^p,\, p \in \mathbb{N},$$

where Θ is a suitable set of regression parameters.

Now, for a given data set

$$(x_1, y_1), \ldots, (x_n, y_n) \in \mathbb{R}^2, \quad n \in \mathbb{N},$$

the observed outcomes $y_1, \ldots, y_n$ are realisations of the random variables $Y_1, \ldots, Y_n$ which correspond to the design variables $x_1, \ldots, x_n$, which in turn are realisations of the random variables $X_1, \ldots, X_n$. The data set $(x_1, y_1), \ldots, (x_n, y_n)$ is therefore a realisation of the random sample

$$(X_1, Y_1), \ldots, (X_n, Y_n) \overset{iid}{\sim} (X, Y).$$

Another notation is that the distribution function of the random variable X is denoted by F, in short:

$$X \sim F\,.$$

Under the model hypothesis $\mathrm{H}_0 : \; m \in \mathcal{M} = \{\, m(\cdot, \theta) \,|\, \theta \in \Theta\}$, there exists a true but unknown parameter

$$\theta_0 \in \Theta$$

for which

$$m(x) = m(x, \theta_0)$$

holds true. Moreover, let $\hat{\theta}_n$ be any reasonable estimator for θ_0 — for example, the LSE.

In the given situation the so-called *marked empirical process* is defined as follows:

Definition 2.1.1.

Let I be any interval containing all $x \in \mathbb{R}$ with $0 < F(x) < 1$. Then, for $n \in \mathbb{N}$, we call

i) $$R_n(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(X_i) \cdot (Y_i - m(X_i)), \quad x \in I,$$

the marked empirical process where the marks are the true but unknown regression errors $\varepsilon_i = Y_i - m(X_i),\ 1 \le i \le n$.

ii) $$R_n^1(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, x]}(X_i) \cdot (Y_i - m(X_i, \hat{\theta}_n)), \quad x \in I,$$

the marked empirical process where the marks are the so-called (regression) residuals

$$\hat{\varepsilon}_{ni} := Y_i - m(X_i, \hat{\theta}_n), \quad 1 \le i \le n.$$

Figure 2.3 shows the graph of $R_n^1(x)$ for some random sample $X_1, \ldots, X_n$ and some residuals $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{nn}$.

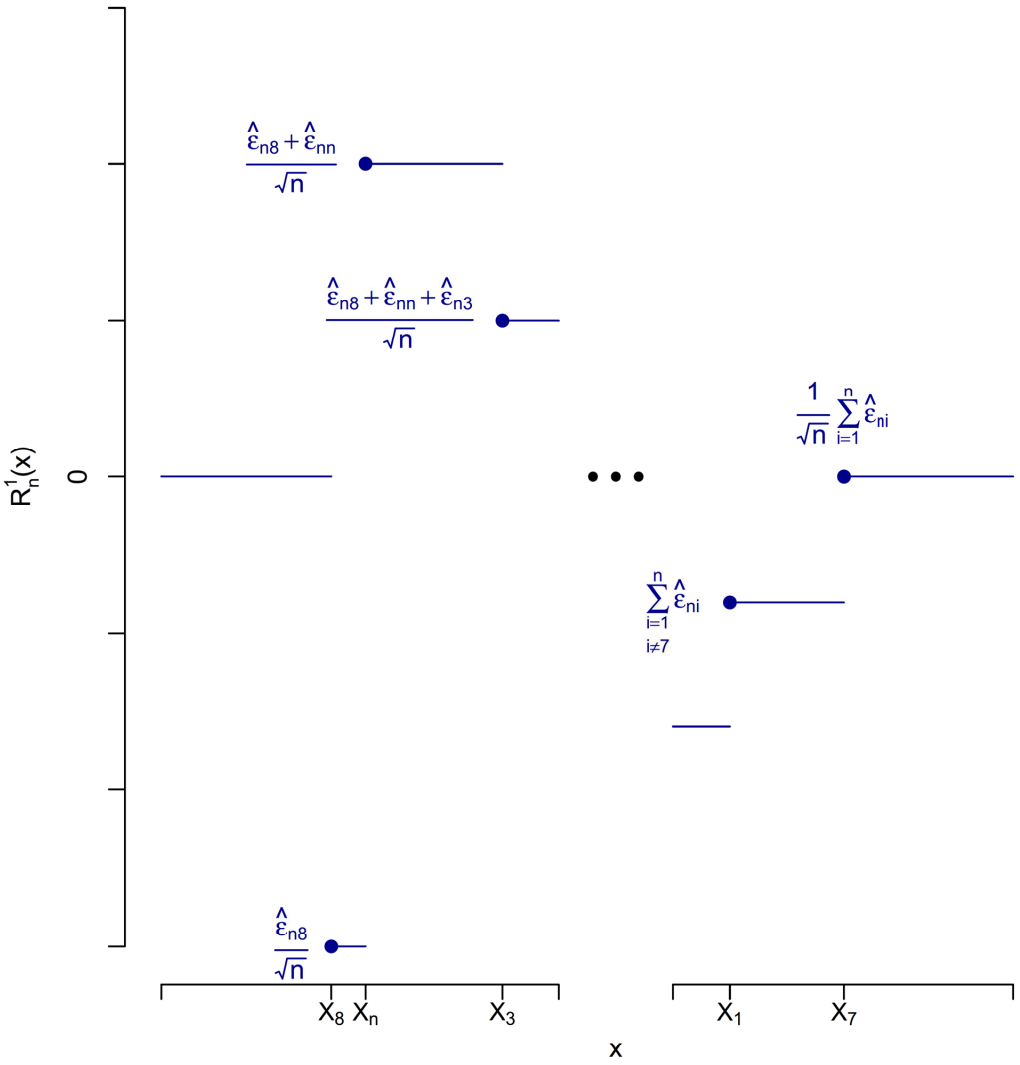


Figure 2.3: Graph of $R_n^1(x)$ for the random sample $X_1, \ldots, X_n$ and residuals $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{nn}$.

Remark 2.1.2. **[Stute (1997)]**

The paths of the processes R_n and R_n^1 are càdlàg (continuous on the right, limit on the left) functions defined on $\mathbb{R}$. By continuously extending the processes R_n and R_n^1 to $-\infty$ and ∞ by putting

$$R_n(-\infty) := 0 \quad \text{and} \quad R_n(\infty) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - m(X_i))$$

as well as

$$R_n^1(-\infty) := 0 \quad \text{and} \quad R_n^1(\infty) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (Y_i - m(X_i, \hat{\theta}_n)),$$

both process paths become elements of $D[-\infty,\infty]$. Here, $D(I)$ denotes the Skorokhod space of càdlàg functions defined on a set I, which, unless otherwise specified, always is endowed with the Skorokhod metric. Note that $D[-\infty,\infty]$ is isometric to $D[0,1]$. Such an isometry is given, for example, by

$$\psi : [-\infty,\infty] \longrightarrow [0,1],\ x \longmapsto \begin{cases} 0, & x = -\infty, \\ \frac{1}{2} \cdot \left(1 + \frac{x}{1+|x|}\right), & x \in \mathbb{R}, \\ 1, & x = \infty. \end{cases}$$

Remark 2.1.3.

In the course of this work, we consider measure integrals of measurable functions g defined on $\mathbb{R}$ with respect to the distribution function F of X. Note that here, and throughout the remainder of this work, we shall identify the measure corresponding to a distribution function F with the distribution function F itself. In such integrals, and independent of g, we make the natural notations

$$\int\limits_{[-\infty,a]} g(x)\,dF(x) := \int\limits_{(-\infty,a]} g(x)\,dF(x), \quad a \in \mathbb{R},$$

$$\int\limits_{[a,\infty]} g(x)\,dF(x) := \int\limits_{[a,\infty)} g(x)\,dF(x), \quad a \in \mathbb{R},$$

$$\int\limits_{[-\infty,\infty]} g(x)\,dF(x) := \int\limits_{(-\infty,\infty)} g(x)\,dF(x).$$

Furthermore, we use the natural notations $F(-\infty) := 0$, $F(\infty) := 1$.

For the process $R_n(x)$ the following limit theorem can be given:

Theorem 2.1.4. **[Stute (1997), Theorem 1.1]**

Let $\mathbb{E}(Y^2) < \infty$. Then,

$$R_n(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} R_\infty(\cdot) \quad \text{in } D[-\infty,\infty],$$

where the limit process $R_\infty(\cdot)$ is a centered Gaussian process with the covariance function

$$K(s,t) = \int\limits_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y|X=x)\,dF(x), \quad s,t \in [-\infty,\infty].$$

Proof: We reproduce a more detailed version of the proof from Stute (1997) in Section 7.1 of the appendix, since the proof in Stute (1997) is kept very brief and some of the techniques will be used in subsequent parts of this work. □

Remark 2.1.5.

The practical scope of the above theorem is rather limited, as it only allows one to construct tests of simple hypotheses like

$$H_0 : \ m \in \mathcal{M} \quad \text{v.s.} \quad H_1 : m \notin \mathcal{M},$$

where the class of functions $\mathcal{M}$ consists of a single element, i.e.

$$\mathcal{M} = \{m(\cdot, \theta) \,|\, \theta \in \Theta = \{\theta_1\}\}.$$

To see this, note that in order to test H_0 with the marked empirical process $R_n(\cdot)$, an explicit regression function m is needed. That is, we need an m with known instead of estimated regression parameters. To put it differently, only for such simple hypothesis do we know that under H_0 the residuals and the regression errors coincide.

In goodness-of-fit tests, hypotheses of the form

$$H_0 : \ m \in \mathcal{M} = \{\, m(\cdot, \theta) \,|\, \theta \in \Theta\}, \quad |\Theta| > 1,$$

are tested, which means that the regression function m is specified up to a parameter $\theta \in \Theta$. Note that we always assume that $m(\cdot, \theta) \neq m(\cdot, \widetilde{\theta})$ for all $\theta \neq \widetilde{\theta} \in \Theta$. In this situation, the estimation of the unknown parameter θ becomes necessary, and thus Stute (1997) studied the marked empirical process

$$R_n^1(\cdot) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty, \cdot]}(X_i) \cdot (Y_i - m(X_i, \hat{\theta}_n)).$$

Here, the marks are the residuals

$$\hat{\varepsilon}_{ni} = Y_i - m(X_i, \hat{\theta}_n), \quad 1 \leq i \leq n.$$

Note that, unlike the regression errors $\varepsilon_1, \ldots, \varepsilon_n$, the residuals $\hat{\varepsilon}_{n1}, \ldots, \hat{\varepsilon}_{n1}$, $n \in \mathbb{N}$, need not be independent nor uncorrelated, which makes the analysis of $R_n^1(\cdot)$ much more difficult as compared to $R_n(\cdot)$.

In order to be able to determine the limit process of $R_n^1(\cdot)$, we need to assume regularity of $\hat{\theta}_n$ and smoothness of the functions in $\mathcal{M}$. More precisely, we make the following two assumptions:

Assumption 1

Under the model hypothesis H_0, $\hat{\theta}_n$ admits the expansion

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1) \quad \text{as } n \to \infty,$$

where l denotes an $\mathbb{R}^p$-valued function satisfying

i) $\mathbb{E}(l(X, Y, \theta_0)) = 0$,

ii) $Cov(l(X, Y, \theta_0))$ exists.

Concerning Assumption 1, note the following:

i) For a series of random variables $(X_n)_{n\in\mathbb{N}}$ and a sequence of real numbers $(a_n)_{n\in\mathbb{N}}$, we have

$$(X_n)_{n\in\mathbb{N}} = o_{\mathbb{P}}(a_n) \;:\Longleftrightarrow\; \frac{X_n}{a_n} \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

ii) The multivariate central limit theorem is applicable, i.e.,

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) \xrightarrow[n\to\infty]{\mathcal{D}} N_p(0, Cov(l(X, Y, \theta_0))). \tag{2.1}$$

Assumption 2

i) The regression function $m(x, \theta)$ is continuously differentiable with respect to θ for all $\theta \in int(\Theta)$ and for all $x \in \mathbb{R}$. Here $int(\Theta)$ denotes the interior of the set Θ and it is assumed that Θ convex. Therefore, we can define

$$\begin{aligned} g(x, \theta) &:= (g_1(x, \theta), \ldots, g_p(x, \theta))^T \\ &:= \left(\frac{\partial m(x, \theta)}{\partial \theta_1}, \ldots, \frac{\partial m(x, \theta)}{\partial \theta_p} \right)^T \\ &= \frac{\partial m(x, \theta)}{\partial \theta}. \end{aligned} \tag{2.2}$$

ii) There is an $F-$integrable function M such that

$$\forall\, 1 \le j \le p \quad \forall\, x \in \mathbb{R} \quad \forall\, \theta \in int(\Theta): \quad |g_j(x, \theta)| \le M(x).$$

Note that according to Assumption 2 the function

$$G(x, \theta) := \begin{pmatrix} G_1(x, \theta) \\ \vdots \\ G_p(x, \theta) \end{pmatrix},$$

with

$$G_i(x,\theta) := \int\limits_{[-\infty,x]} g_i(u,\theta)\, dF(u) = \mathbb{E}\left(\mathbb{1}_{[-\infty,x]}(X)\cdot g_i(X,\theta)\right), \quad i=1,\dots,p, \tag{2.3}$$

is well defined and continuous for $\theta \in int(\Theta)$.

Under Assumption 1 and Assumption 2, the following holds true for the limit process $R^1_\infty(\cdot)$ of $R^1_n(\cdot)$.

Theorem 2.1.6. [Stute (1997), Theorem 1.2, Corollary 1.3]

In a regression model with random design as introduced in this section let

- $\mathbb{E}(Y^2) < \infty$,
- Assumption 1 and Assumption 2 be fulfilled,
- $X \sim F$.

Then, under the hypothesis

$$\mathrm{H}_0: \; m \in \mathcal{M} = \{\, m(\cdot,\theta) \,|\, \theta \in \Theta\}, \;\; \Theta \subseteq \mathbb{R}^p,$$

we have:

i)
$$\sup_{x\in[-\infty,\infty]} \left| R^1_n(x) - \left(R_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x,\theta_0)\, l(X_i,Y_i,\theta_0) \right) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0,$$

ii)
$$R^1_n(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} R^1_\infty(\cdot) \quad \text{in } D[-\infty,\infty],$$

where $R^1_\infty(\cdot)$ is a centred Gaussian process with the covariance function

$$\begin{aligned} K^1(s,t) = & \int\limits_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y|X=x)\, dF(x) + G^T(s,\theta_0)\, Cov(l(X,Y,\theta_0))\, G(t,\theta_0) \\ & - G^T(s,\theta_0)\, \mathbb{E}\left[\mathbb{1}_{[-\infty,t]}(X)(Y-m(X,\theta_0))\, l(X,Y,\theta_0)\right] \\ & - G^T(t,\theta_0)\, \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X)(Y-m(X,\theta_0))\, l(X,Y,\theta_0)\right], \; s,t \in [-\infty,\infty]. \end{aligned}$$

Proof: A sketch of the proof is given in Stute (1997) on page 638. Note that we could not find a formal argument for the uniform convergence in i) in Stute (1997). Keeping this in mind, and for the sake of clarity, we present a complete (and partially new) proof of Theorem 2.1.6 in Section 7.1 of the appendix starting on page 102. □

2.1.2 Linear regression models and arbitrary random designs

In order to simplify the following analysis, we restrict ourselves here to a homoscedastic univariate linear regression model under the usage of the LSE. Specifically, we consider:

i) the univariate regression model with random design

$$Y = m(X) + \varepsilon,$$

with

$$m(x) = \mathbb{E}(Y|X = x),$$

where

$$H_0 : m \in \mathcal{M} = \{f^T(\cdot)\,\theta \,|\, \theta \in \mathbb{R}^p\}$$

and call the known function

$$f = (f_1, \ldots, f_p)^T, \quad p \in \mathbb{N},$$

with $f_1, \ldots, f_p : \mathbb{R} \longrightarrow \mathbb{R}$ the regression function.

ii) the LSE $\hat{\theta}_n$ for estimating the regression coefficients.

iii) homoscedastic regression errors ε. That is

$$\mathbb{V}ar(\varepsilon|X = x) = \sigma^2(x) = \sigma^2, \quad x \in \mathbb{R}.$$

In order to simplify the notation even further, we assume without loss of generality that

$$\sigma^2 = 1,$$

since one can divide $R_n(\cdot)$ and $R_n^1(\cdot)$ by σ to normalise the error. If σ^2 is unknown, one can estimate it by means of a consistent estimator $\hat{\sigma}_n^2$ like the one based on the sum of squares of residuals. This procedure does, according to the continuous mapping theorem, not alter asymptotic distributional properties of the processes involved.

The following theorem is a version of Theorem 2.1.6 adapted to this particular scenario.

Theorem 2.1.7. **[Stute (1997)]**

Consider a univariate linear regression model with random design and homoscedastic regression errors of variance 1 under the usage of the LSE. Furthermore, let F be the unknown distribution function of X, assume that

$$J := \int\limits_{[-\infty,\infty]} f(x)\, f^T(x)\, dF(x)$$

exist, and assume that

$$rank(J) = p.$$

Then, under the hypothesis

$$\mathrm{H}_0 : \; m \in \mathcal{M} = \{f^T(\cdot)\,\theta \,|\, \theta \in \mathbb{R}^p\},$$

we have

$$R_n^1(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} R_\infty^1(\cdot) \quad \text{in } D[-\infty,\infty],$$

where $R_\infty^1(\cdot)$ is a centred Gaussian process with the covariance function

$$K^1(s,t) = F(\min\{s,t\}) - \left(\int\limits_{[-\infty,s]} f(x)\, dF(x)\right)^T J^{-1} \int\limits_{[-\infty,t]} f(x)\, dF(x), \quad s,t \in [-\infty,\infty].$$

Proof: In Stute (1997) page 620 and 621, a guideline for the proof can be found. We present a detailed proof of Theorem 2.1.7 according to this guideline starting on page 116 in Section 7.1 of the appendix onwards. □

2.1.3 Linear regression models and the uniform random design

First, we start with the formal definition of a quantile function F^- associated to a distribution function F. See, for example, Chaper 21 in Vaart (1998).

Definition 2.1.8.

Let F be a distribution function on $\mathbb{R}$. By

$$F^- : (0,1) \longrightarrow \mathbb{R},\ t \longmapsto \inf\{x \in \mathbb{R} : F(x) \geq t\},$$

we denote the quantile function associated to F. F^- is continuous from the left and monotonically increasing.

Remark 2.1.9.

Let F be a distribution function on $\mathbb{R}$ and define

$$a := \inf\{x \in \mathbb{R} | F(x) > 0\} \in [-\infty, \infty)$$

and

$$b := \sup\{x \in \mathbb{R} | F(x) < 1\} \in (-\infty, \infty].$$

If necessary, we may extend F^- to $[0,1]$ by defining

$$F^-(0) := a \ \text{ and } \ F^-(1) := b.$$

In Stute (1997), the approach of applying the quantile transformation $X \stackrel{\mathcal{D}}{=} F^-(U)$, where $X \sim F$ and $U \sim U(0,1)$, has been used to simplify the proofs of Theorem 2.1.4 and Theorem 2.1.6. In this regard, it is mentioned on page 637 that "... we may and do assume in the following that F is the uniform distribution on $[0,1]$".
In this section, however, we consider a different transformation, the so-called probability integral transformation to transform a regression model with random design into a regression model with uniform random design. To be more precise, the following theorem specifies the statement that for a regression model with random design, there is an a.s. equivalent corresponding regression model whose design is, to a certain extent, the uniform distribution on the unit interval. To our knowledge, this fact has not been stated or proven in the literature so far.

Theorem 2.1.10.

Consider a regression model with random design

$$Y = m(X) + \varepsilon,$$

where $X \sim F$, $m(x) = \mathbb{E}(Y \,|\, X = x)$, $\mathbb{E}(\varepsilon \,|\, X = x) = 0$, $\mathbb{V}ar(\varepsilon \,|\, X = x) = \sigma^2(x)$ and where Assumption 2 of Section 2.1.1 is satisfied.
Then, this regression model is almost surely equal to the corresponding regression model with uniform random design on the unit interval. More precisely, this regression model has the form

$$Y = (m \circ F^-)(U) + \varepsilon,$$

where

- $U := F(X)$ is a real-valued random variable with

$$F_U(t) := \mathbb{P}(U \leq t) = \begin{cases} 0, & t \leq 0, \\ \lim\limits_{x \uparrow F^-(t)} F(x), & t \notin F(\mathbb{R}) \text{ and } t \in (0,1), \\ t, & t \in F(\mathbb{R}) \text{ and } t \in (0,1), \\ 1, & t \geq 1 \end{cases}$$

  and, in particular, $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$,

- $(m \circ F^-)(u) = \mathbb{E}(Y|U = u)$,

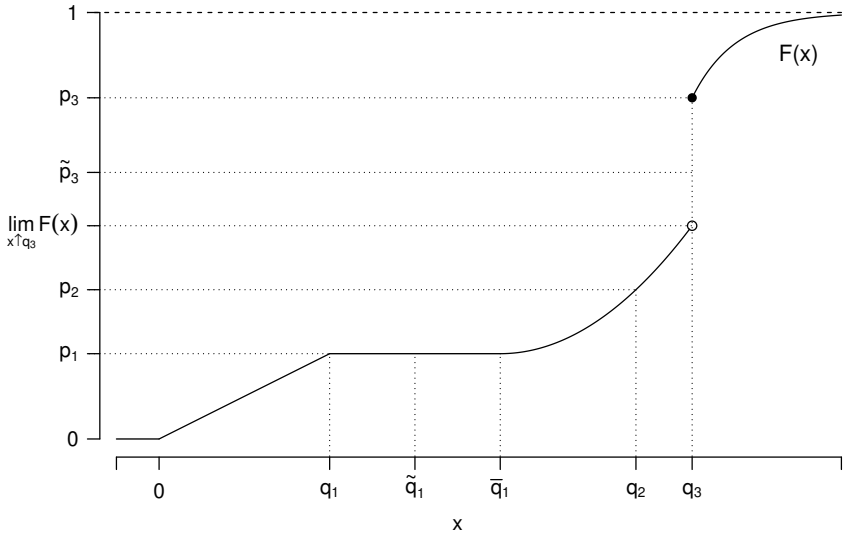
- $\mathbb{E}(\varepsilon \,|\, U = u) = 0$ and $\mathbb{V}ar(\varepsilon \,|\, U = u) = \sigma^2(F^-(u))$,

- Assumption 2 of Section 2.1.1 is satisfied.

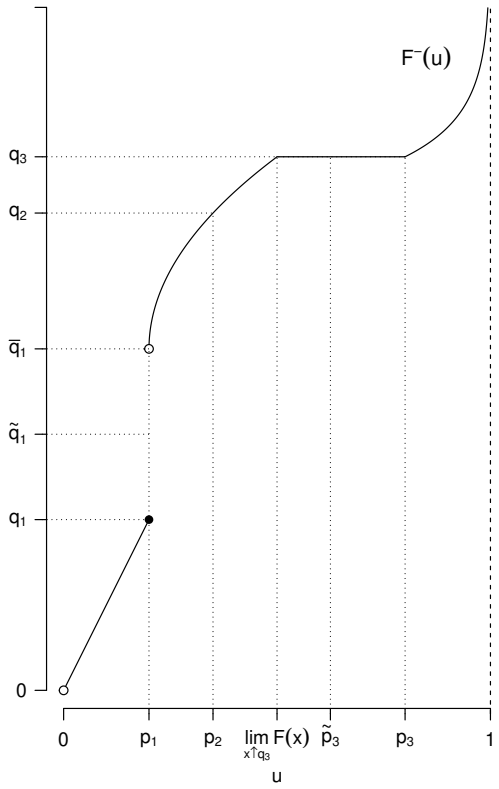
Proof: The proof of this theorem is deferred to Section 7.1 of the appendix, starting on page 122. □

According to Theorem 2.1.10, the distribution $\mathbb{P}^U$ of the design variable U of the transformed model $Y = (m \circ F^-)(U) + \varepsilon$ is the uniform distribution on the Sub-σ-Algebra $\sigma(F^-)$ of the Borel-σ-Algebra $\mathcal{B}$. Therefore, we refer to the transformed model as the regression model with uniform random design. In the special situation that the distribution function F of the original design variable X is continuous, we have $(0,1) \subseteq F(\mathbb{R})$, $F_U|_{[0,1]} = id_{[0,1]}$ and $\sigma(F^-) = \mathcal{B}$. Thus, $P^U = \lambda\!\!\lambda_{[0,1]}$ holds true.

As an example of a possible distribution function F of the design variable X in the original model $Y = m(X) + \varepsilon$, see Figure 2.4a. The corresponding quantile function F^- is given in Figure 2.4b and for the corresponding distribution function F_U of the design variable $U := F(X)$ in the transformed model $Y = (m \circ F^-)(U) + \varepsilon$, see Figure 2.5a. The quantile function F_U^- belonging to F_U can be found in Figure 2.5b.



(a) Distribution function F of X.



(b) Quantile function F^-.

Figure 2.4: Example graphs of a distribution function F and its quantile function F^-.

Remark 2.1.11.

Consider the transformed model

$$Y = (m \circ F^-)(U) + \varepsilon.$$

If we additionally assume $\mathbb{E}(Y^2) < \infty$ and that Assumption 1 of Section 2.1.1 is satisfied, under $H_0 : m \in \mathcal{M} = \{m(\cdot,\theta)|\theta \in \Theta\}$, $\Theta \subseteq \mathbb{R}^p$, the following convergence of the uniform marked empirical process

$$\overline{R}_n^1(x) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[0,x]}(U_i) \cdot (Y_i - m(F^-(U_i), \hat{\theta}_n)), \quad x \in [0,1],$$

holds true according to Theorem 2.1.6:

$$\overline{R}_n^1(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} \overline{R}_\infty^1(\cdot) \quad \text{in } D[0,1].$$

Here, $\overline{R}_\infty^1(\cdot)$ is a centred Gaussian process with the covariance function

$$\begin{aligned} K^1(s,t) = & \int\limits_{[0,\min\{s,t\}]} \mathbb{V}ar(Y|U=u)\, dF_U(u) + \overline{G}^T(s,\theta_0)\, Cov(l(U,Y,\theta_0))\, \overline{G}(t,\theta_0) \\ & - \overline{G}^T(s,\theta_0)\, \mathbb{E}\left[\mathbb{1}_{[0,t]}(U) \cdot (Y - m(F^-(U),\theta_0))\, l(U,Y,\theta_0)\right] \\ & - \overline{G}^T(t,\theta_0)\, \mathbb{E}\left[\mathbb{1}_{[0,s]}(U) \cdot (Y - m(F^-(U),\theta_0))\, l(U,Y,\theta_0)\right], \quad s,t \in [0,1]. \end{aligned}$$

For the definition of $\overline{G}(u,\theta)$, see the end of the proof of Theorem 2.1.10.

In the remainder of this section, we consider the special situation of a univariate linear regression model with random design and homoscedastic regression errors of variance 1 under the usage of the LSE $\hat{\theta}_n$ (see Section 2.1.2). Additionally, let us suppose that the matrix

$$J = \int\limits_{[-\infty,\infty]} f(x) f^T(x)\, dF(x) \in \mathbb{R}^{p\times p}$$

of the original regression functions exists and has full rank. For the interpretation of integrals over $[-\infty,\infty]$, we refer to Remark 2.1.3. On $\mathcal{L}^2([-\infty,\infty],F)$ the inner product is defined by

$$\langle\cdot,\cdot\rangle : \mathcal{L}^2([-\infty,\infty],F) \times \mathcal{L}^2([-\infty,\infty],F) \longrightarrow \mathbb{R}, \;\; (g,h) \longmapsto \int\limits_{[-\infty,\infty]} g(t)h(t)\, dF(t),$$

thus J is the Gram matrix of $f_1,\ldots,f_p$, which has full rank if, and only if, the functions $f_1,\ldots,f_p$ are linearly independent in $\mathcal{L}^2([-\infty,\infty],F)$. Here, as usual, $\mathcal{L}^2([-\infty,\infty],F)$

denotes the Hilbert space of square integrable functions defined on $[-\infty,\infty]$ with respect to F.
Furthermore, the regression functions $f_1,\dots,f_p$ are linearly independent in $\mathcal{L}^2([-\infty,\infty],F)$ if, and only if, the transformed regression functions $f_1\circ F^-,\dots,f_p\circ F^-$ are linearly independent in $\mathcal{L}^2([0,1],\lambda\!\!\!\lambda_{[0,1]})$. This, in turn, is precisely the case when the transformed regression functions $f_1\circ F^-,\dots,f_p\circ F^-$ are linearly independent in $\mathcal{L}^2([0,1],F_U)$. The first assertion can be verified similarly to the proof of assertion i) β) on page 149 and from there, the second assertion is true since $[0,1]\in\sigma(F^-)$ and $\mathbb{P}^U|_{\sigma(F^-)}=\lambda\!\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$. The following assertions are thus equivalent:

i) The Matrix $J=\int\limits_{[-\infty,\infty]} f(x)f^T(x)\,dF(x)$ exists and has full rank.

ii) The Matrix $\int\limits_{[0,1]}(f\circ F^-)(u)(f\circ F^-)^T(u)\,dF_U(u)$ exists and has full rank.

iii) The Matrix $\int\limits_{[0,1]}(f\circ F^-)(u)(f\circ F^-)^T(u)\,d\lambda\!\!\!\lambda_{[0,1]}(u)$ exists and has full rank.

iv) The original regression functions $f_1,\dots,f_p$ are linearly independent in $\mathcal{L}^2([-\infty,\infty],F)$.

v) The transformed regression functions $f_1\circ F^-,\dots,f_p\circ F^-$ are linearly independent in $\mathcal{L}^2([0,1],F_U)$.

vi) The transformed regression functions $f_1\circ F^-,\dots,f_p\circ F^-$ are linearly independent in $\mathcal{L}^2([0,1],\lambda\!\!\!\lambda_{[0,1]})$.

In the present situation, Theorem 2.1.7 now yields that the uniform marked empirical process

$$\overline{R}_n^1(x)=\frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{1}_{[0,x]}(U_i)\cdot(Y_i-(f\circ F^-)^T(U_i)\,\hat{\theta}_n),\quad x\in[0,1],$$

converges weakly, that is

$$\overline{R}_n^1(\cdot)\xrightarrow[n\to\infty]{\mathcal{D}}\overline{R}_\infty^1(\cdot)\quad\text{in } D[0,1].$$

Here, $\overline{R}_\infty^1(\cdot)$ is a centred Gaussian process with the covariance function

$$K^1(s,t)=F_U(\min\{s,t\})-\left(\int\limits_{[0,s]} f(F^-(x))\,dF_U(x)\right)^T\times \tag{2.4}$$

$$\times\left(\int\limits_{[0,1]} f(F^-(x))f^T(F^-(x))\,dF_U(x)\right)^{-1}\int\limits_{[0,t]} f(F^-(x))\,dF_U(x),\quad s,t\in[0,1].$$

Since $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$ holds true, we have

$$\int\limits_{[0,1]} f(F^-(x))f^T(F^-(x))\,dF_U(x) = \int\limits_{[0,1]} f(F^-(x))f^T(F^-(x))\,d\lambda\!\!\!\lambda_{[0,1]}(x),$$

but the same need not to be true for the two other integrals in (2.4), since $[0,s]$ or $[0,t]$ might not be in the Sub-σ-Algebra $\sigma(F^-)$. Nevertheless, one can apply the change of variables formula (see Lemma 7.1.3) to the first integral in (2.4), obtaining

$$\begin{aligned}\int\limits_{[0,s]} f(F^-(x))\,dF_U(x) &= \int\limits_{[0,F_U(s)]} f(F^-(F_U^-(x)))\,d\lambda\!\!\!\lambda_{[0,1]}(x)\\ &= \int\limits_{[0,F_U(s)]} f(F^-(x))\,d\lambda\!\!\!\lambda_{[0,1]}(x) \qquad (2.5)\end{aligned}$$

and proceed similarly with the last integral in (2.4). Equation (2.5) above is true, since

$$F^-(F_U^-(t)) = F^-(t), \quad t \in [0,1]. \qquad (2.6)$$

To see this, first note that, since $F_U^-(0) = 0$ and $F_U^-(1) = 1$, (2.6) holds for $t = 0$ and $t = 1$. If $t \in (0,1) \cap F(\mathbb{R})$ we have $F_U^-(t) = t$ and thus, (2.6) holds as well. In the remaining case we have $t \in (0,1)$ and $t \notin F(\mathbb{R})$. Thus, according to Theorem 2.1.10, F_U takes the constant value $\lim\limits_{x\uparrow F^-(t)} F(x)$ in the interval $\left[\lim\limits_{x\uparrow F^-(t)} F(x), F(F^-(t))\right)$, which includes t. As a consequence, one can conclude $F_U^-(u) = F(F^-(t))$ for all $u \in \left(\lim\limits_{x\uparrow F^-(t)} F(x), F(F^-(t))\right)$ and thus

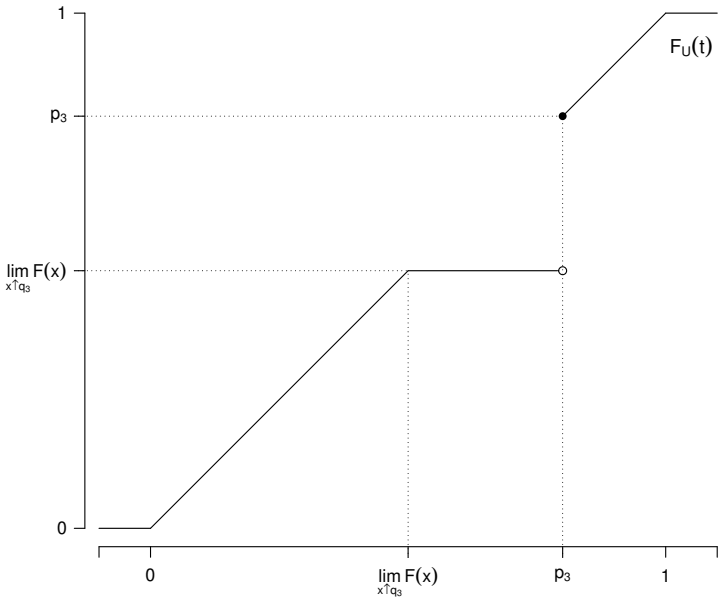
$$F^-(F_U^-(t)) = F^-(F(F^-(t))) = F^-(t),$$

if $t \in \left(\lim\limits_{x\uparrow F^-(t)} F(x), F(F^-(t))\right)$, where we used (7.39) of Lemma 7.1.1. If, in contrast, $t = \lim\limits_{x\uparrow F^-(t)} F(x)$, we get

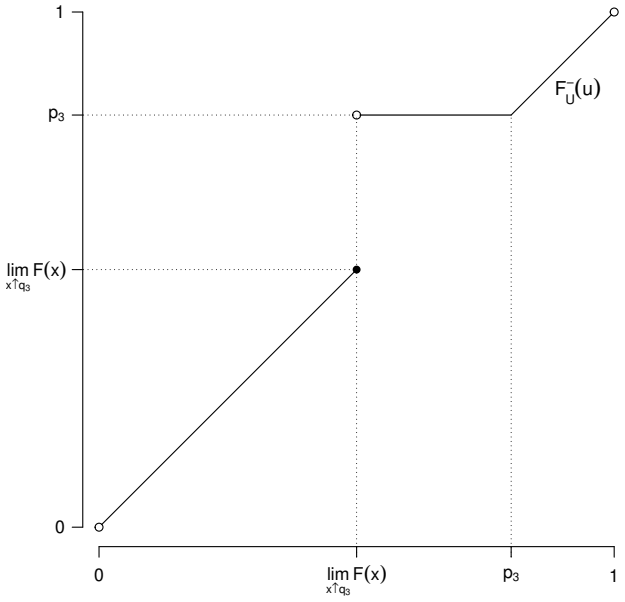
$$F_U^-\left(\lim_{x\uparrow F^-(t)} F(x)\right) = \lim_{x\uparrow F^-(t)} F(x),$$

since F_U^- is continuous from the left and $F(x) \in F(\mathbb{R})$, for all $x \in \mathbb{R}$, which implies $F_U^-(F(x)) = F(x)$.

As an example of a possible distribution function F_U of the design variable U in the transformed model and the corresponding quantile function F_U^-, see Figure 2.5a and Figure 2.5b respectively.



(a) Distribution function F_U of $U = F(X)$.



(b) Quantile function F_U^-.

Figure 2.5: Example graphs of the distribution function F_U of $U = F(X)$ and its quantile function F_U^- corresponding to the distribution function F of X from Figure 2.4a.

Taking the above facts into consideration, we conclude the following.

Theorem 2.1.12.

Consider a homoscedastic linear regression model with random design

$$Y = f^T(X)\,\theta + \varepsilon,$$

where $f = (f_1, \ldots, f_p)^T$, $p \in \mathbb{N}$ with $f_1, \ldots, f_p : \overline{\mathbb{R}} \longrightarrow \mathbb{R}$ that are linearly independent in $\mathcal{L}^2([-\infty, \infty], F)$, $X \sim F$, $\mathbb{E}(\varepsilon \,|\, X = x) = 0$ and $\mathbb{V}ar(\varepsilon \,|\, X = x) = 1$.

Then, this regression model is almost surely equal to the corresponding linear regression model with uniform random design on the unit interval. More precisely, this regression model has the form

$$Y = (f \circ F^-)^T(U)\,\theta + \varepsilon,$$

- where $U := F(X)$ is a real-valued random variable with

$$F_U(t) := \mathbb{P}(U \leq t) = \begin{cases} 0, & t \leq 0, \\ \lim\limits_{x \uparrow F^-(t)} F(x), & t \notin F(\mathbb{R}) \text{ and } t \in (0,1), \\ t, & t \in F(\mathbb{R}) \text{ and } t \in (0,1), \\ 1, & t \geq 1 \end{cases}$$

  and, in particular, $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$,

- where $f \circ F^- = (f_1 \circ F^-, \ldots, f_p \circ F^-)^T$ are the regression functions with $f_1 \circ F^-, \ldots, f_p \circ F^- : [0,1] \longrightarrow \mathbb{R}$ that are linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\lambda_{[0,1]})$,
- where $\mathbb{E}(\varepsilon \,|\, U) = 0$ and $\mathbb{V}ar(\varepsilon \,|\, U = u) = 1$ for all $u \in [0,1]$.

In this model, under the usage of the LSE $\hat{\theta}_n$, the uniform marked empirical process reads

$$\overline{R}_n^1(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[0,x]}(U_i) \cdot (Y_i - (f \circ F^-)^T(U_i)\,\hat{\theta}_n), \quad x \in [0,1],$$

and the weak convergence

$$\overline{R}_n^1(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} \overline{R}_\infty^1(\cdot) \quad \text{in } D[0,1]$$

holds true. Here, $\overline{R}_\infty^1(\cdot)$ is a centred Gaussian process whose covariance function, for $s, t \in [0,1]$, is

$$K^1(s,t) = F_U(\min\{s,t\}) - \left(\int\limits_{[0, F_U(s)]} f(F^-(x))\, d\lambda\!\!\lambda_{[0,1]}(x) \right)^T \times$$

$$\times \left(\int\limits_{[0,1]} f(F^-(x)) f^T(F^-(x))\, d\lambda\!\!\lambda_{[0,1]}(x) \right)^{-1} \int\limits_{[0, F_U(t)]} f(F^-(x))\, d\lambda\!\!\lambda_{[0,1]}(x).$$

2.2 Partial sum processes and fixed designs

This section consists of two parts. In the first part, we study residual partial sum processes of regression residuals in linear regression models with fixed designs on the example of Bischoff (1998). Furthermore, the limit process of the respective residual partial sum processes is specified and its covariance structure is investigated. In the second part, we prove that in the case where only the distribution of the residual partial sum limit process is of interest in a linear regression model with a fixed design, one can always assume the design to be the equidistant design on the unit interval $[0,1]$. To our knowledge this fact has not been explicitly stated or proven in the literature so far.

2.2.1 Results of Bischoff (1998)

A linear regression model with fixed design has the form

$$Y_n = X_n\beta + \varepsilon_n, \qquad n \in \mathbb{N},$$

where Y_n is the n-dimensional output vector composed of a deterministic part $X_n\beta$ and a vector-valued random variable $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$. Here, $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ are real-valued iid random variables with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = \sigma^2$. Thus, we are in a non-parametric and homoscedastic scenario.
Regarding the deterministic part of the model,

$$\beta = (\beta_1, \ldots, \beta_p) \in \mathbb{R}^p$$

are the unknown regression coefficients, which, in a model with $n \geq p$ observations, are to be estimated via the LSE $\hat{\beta}_n$. We consider a triangular array of design points,

$$\left((t_{nj})_{j=1}^n\right)_{n\in\mathbb{N}} \subseteq [a,b] =: \mathcal{E}, \quad a < b \in \mathbb{R},$$

whose n-th row is called an (exact) design (for n observations). We refer to $\mathcal{E}$ as the experimental region. Furthermore, without loss of generality, we may assume that

$$a \leq t_{n1} \leq t_{n2} \leq \cdots \leq t_{nn-1} \leq t_{nn} \leq b, \quad n \in \mathbb{N}.$$

The vector-valued regression function

$$f = (f_1, \ldots, f_p)^T : [a,b] \longrightarrow \mathbb{R}^p$$

consists of real valued, known and continuous functions of bounded variation and, the design matrix reads

$$X_n := \left(f_r\left(t_{ns}\right)\right)_{s=1,r=1}^{n\ \ p} \in \mathbb{R}^{n\times p}.$$

We are therefore dealing with univariate linear regression models, since there is only one explanatory variable. Moving on, the ordinary least-squares residuals are defined as

$$\hat{\varepsilon}_n := Y_n - X_n\hat{\beta}_n \in \mathbb{R}^n,$$

where, if X_n has full Rank,

$$\hat{\beta}_n := (X_n^T X_n)^{-1} X_n^T Y_n$$

is the LSE approximation for β. For following applications, we define for $t_{n1}, \ldots, t_{nn}$, that is the n-th row of a given design,

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(t_{ni}), \quad x \in \mathbb{R},$$

as the empirical distribution function uniquely corresponding to this design.

Two additional assumptions are made in order to determine a limit process for the least squares residual process:

Assumption 1
There is a limit design distribution function F for which

$$\sup_{t\in[a,b]} |F_n(t) - F(t)| \xrightarrow[n\to\infty]{} 0 \quad \text{is satisfied.}$$

Assumption 2
The matrix

$$J := \int_{[a,b]} f(t) f^T(t)\, dF(t) \in \mathbb{R}^{p\times p} \quad \text{has full rank.} \tag{2.7}$$

As we already know (see the preliminary remarks on Theorem 2.1.12), this is equivalent to the linear independence of the regression functions $f_1, \ldots, f_p$ in $\mathcal{L}^2([a,b], F)$.

For the remainder of this work, by $B = (B(t))_{t\in[0,1]}$, we denote a standard Brownian motion. Furthermore, we define the following operator, which will enable us to generate a continuous stochastic process out of the regression residuals.

Definition 2.2.1.

For $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$ and $\lfloor x \rfloor := \max\{k \in \mathbb{Z} \,|\, k \le x\}$, one defines the partial sum operator

$$T_n : \mathbb{R}^n \longrightarrow C[0,1]$$

by

$$T_n(a)(t) = \begin{cases} \sum_{i=1}^{\lfloor nt \rfloor} a_i + (nt - \lfloor nt \rfloor)\, a_{\lfloor nt \rfloor + 1}\,, & t \in [0,1), \\ \sum_{i=1}^{n} a_i\,, & t = 1. \end{cases}$$

Figure 2.6 shows the graph of $T_n(a)(t)$ for $a = (a_1, \ldots, a_n)^T \in \mathbb{R}^n$, $n \in \mathbb{N}$.

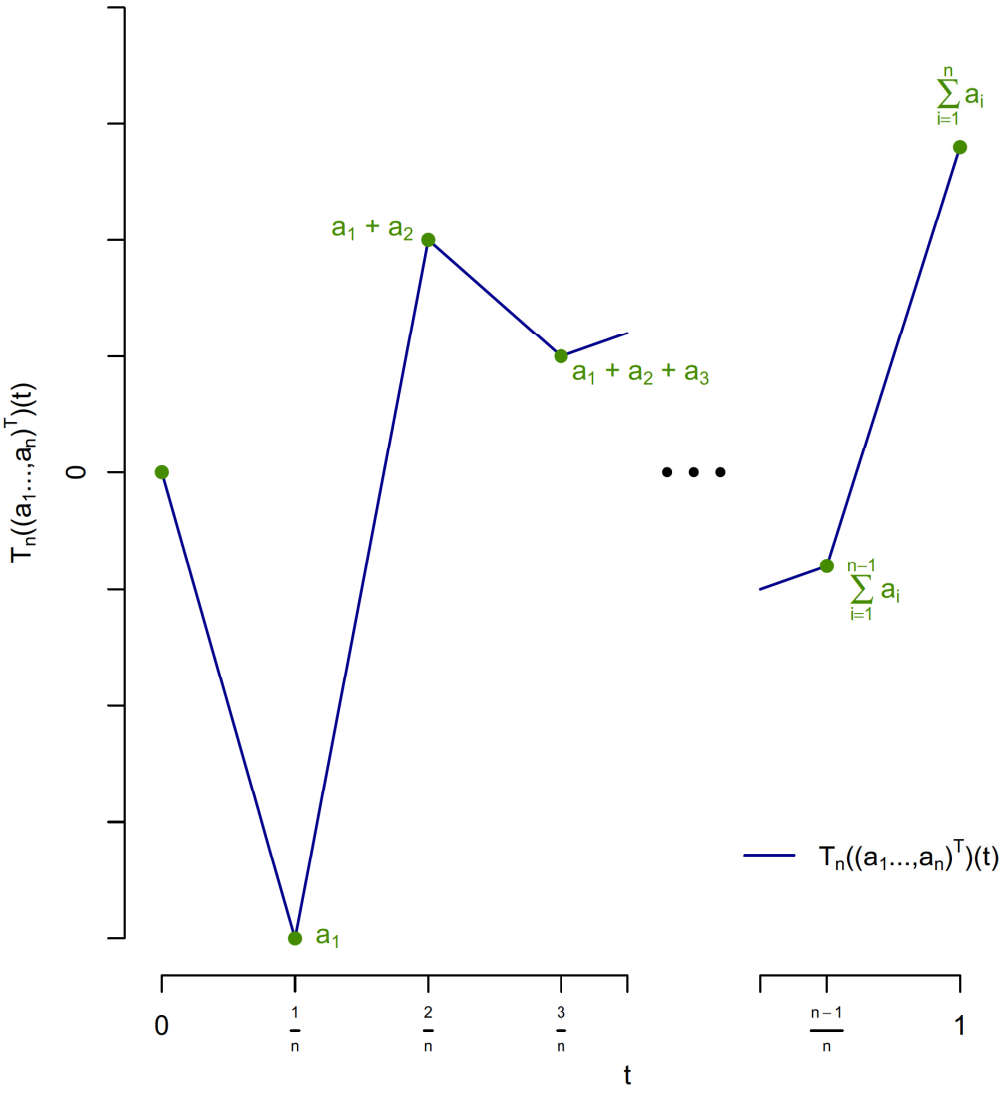


Figure 2.6: Operating principle of the partial sum operator T_n.

Given a real-valued random variable X with distribution function F, we define

- for $z \in [0,1]$,
$$F_z(x) := \min\{F(x), z\},\ x \in \mathbb{R}.$$
- the support of X respectively F by
$$supp(F) := supp(X) := \{x \in \mathbb{R} \,|\, \mathbb{P}(X \in (x-r, x+r)) > 0, \text{ for all } r > 0\}.$$

We finally consider the residual partial sum process of the least-squares residuals

$$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot),$$

a stochastic process that — as it linearly interpolates the residuals $\hat{\varepsilon}_n$ over the interval $[0,1]$ — has paths in $C[0,1]$. Functional central limit theorems for residual partial sum processes can be found, for example, in MacNeill (1978b), MacNeill (1978a), as well as Bischoff (1998) and read as follows:

Theorem 2.2.2. [Bischoff (1998), Theorem 2.2, Theorem 3.1]

Let the regression functions

$$f_1, \ldots, f_p, \quad p \in \mathbb{N},$$

be continuous, known and real-valued functions of bounded variation that are linearly independent in $\mathcal{L}^2([a,b], F)$. Furthermore, let

$$F_n \xrightarrow[n\to\infty]{} F \quad \text{uniformly},$$

where F is a distribution function with $supp(F) \subseteq [a,b]$. Then,

i) $$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(\cdot) \quad \text{in } C[0,1],$$

where $B_{f,F}(\cdot)$ is a Gaussian process that for $z \in [0,1]$ is given by

$$B_{f,F}(z) := B(z) + \left(\int_{[a,b]} f(t)\, dF_z(t) \right)^T J^{-1} \left(\int_{[a,b]}^{(R)} B(F(t))\, df(t) - B(1) f(F^-(1)) \right).$$

ii) $$\frac{1}{\sigma\sqrt{n}} (T_n(\hat{\varepsilon}_n) \circ F_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (B_{f,F} \circ F)(\cdot) \quad \text{in } (D[a,b], \|\cdot\|_\infty),$$

where $(B_{f,F} \circ F)(\cdot)$ is a Gaussian process that, for $s \in [a,b]$, is given by

$$(B_{f,F} \circ F)(s) = B(F(s)) + \left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1} \left(\int_{[a,b]}^{(R)} B(F(t))\, df(t) - B(1) f(F^-(1)) \right).$$

With regard to Theorem 2.2.2, note the following:

- In i), as usual, we consider the uniform topology on $C[0,1]$.
- The residual partial sum limit process $B_{f,F}$ does depend on the experimental region $[a,b]$, the regression functions $f_1, \ldots, f_p$ and the limit distribution F of the design points. However, it does not depend on the choice of the design distributions F_n, $n \in$

$\mathbb{N}$, converging uniformly to F. Therefore, Theorem 2.2.2 constitutes an "invariance principle".

- Up to this theorem, all integrals considered are measure integrals. However, the integral $\int_{[a,b]}^{(R)} B(F(t))\,df(t)$ is obtained as the limit of Riemann-Stieltjes sums. In general, we denote the Riemann-Stieltjes integral of a function f over an interval I with respect to a function g by $\int_{I}^{(R)} f(x)\,dg(x)$. For some theory on Riemann-Stieltjes integrals see, for example, Strook (1994).

- The residual partial sum limit process $B_{f,F}(\cdot)$ has paths in $C[0,1]$, but since the experimental region was given as the interval $[a,b]$, it seems more natural to consider stochastic processes with paths defined on this interval rather than on $[0,1]$. Part ii) of Theorem 2.2.2 provides a process with this convenient property. Note that here weak convergence takes place in $D[a,b]$, rather than in $C[a,b]$, since the distribution functions F_n, $n \in \mathbb{N}$, and F need not be continuous and therefore are generally elements of $D[a,b]$. We emphasise that here, in contrast to Section 2.1, we endow $D[a,b]$ with the uniform topology, i.e., the topology given by the uniform metric

$$\varrho(x,y) := \|x-y\|_\infty = \sup\{t \in [a,b]|\,|x(t)-y(t)|\}, \quad x,y \in D[a,b].$$

  That is, we consider weak convergence in $(D[a,b], \|\cdot\|_\infty)$.

Remark 2.2.3.

Note that in Theorem 2.2.2 a mass point is permitted at the left endpoint of the experimental region $[a,b]$. That is, it is possible that $F(a) =: c > 0$. In this case, this mass point is noticed by measure integrals like $\int_{[a,s]} f(t)\,dF(t)$, since it holds true that

$$\int_{[a,s]} f(t)\,dF(t) = c \cdot f(a) + \int_{(a,s]} f(t)\,dF(t), \quad s \in (a,b].$$

However, such a mass point at the left endpoint of the area of integration is not noticed by a Riemann-Stieltjes integral as

$$\int_{[a,b]}^{(R)} B(F(t))\,df(t) = \int_{(a,b]} B(F(t))\,d\mu_f(t). \tag{2.8}$$

For a proof of equation (2.8), see Theorem 5.1.2 in Strook (1994). Note that here, μ_f refers to the signed measure associated to f by the following procedure: Since $f = (f_1,\ldots,f_p)^T$ and $f_i : [a,b] \to \mathbb{R}$, $i = 1,\ldots,p$, are continuous functions of bounded variation for $i \in \{1,\ldots,p\}$, there is a pair of continuous and monotonically increasing functions f_i^+ and f_i^- such that $f_i(x) = f_i^+(x) - f_i^-(x)$, $x \in [a,b]$ (see

for example Aufgabe 1.10 of Kapitel VII in Elstrodt (2009)). Therefore, for $i = 1, \ldots, p$, and $* \in \{+, -\}$ there is a finite Borel measure $\mu_{f_i^*}$ on $((a,b], \mathcal{B}_{(a,b]})$ with $\mu_{f_i^*}((l,m]) = f_i^*(m) - f_i^*(l)$, $l \leq m \in [a,b]$ (see for example Bauer (1992), p. 36). Hence, we define $\mu_{f_i} = \mu_{f_i^+} - \mu_{f_i^-}$, $i = 1, \ldots, p$ and $\mu_f = (\mu_{f_1}, \ldots, \mu_{f_p})^T$.
The above theorem is therefore adequate for a fixed design linear regression model with the experimental region $[a,b]$ without a mass point at the left endpoint (i.e. $F(a) = 0$). In this regard, see part i) of Corollary 3.2 in Bischoff (1998) and note that there is a typographical error (there should be a $-$ instead of a $+$ before the integral) and that in addition it is implicitly assumed that there is no mass point at the left end point of the experimental region.
The aim now is to state a theorem that is suitable for a fixed-design linear model that can also have a mass point at the left endpoint of $[a,b]$, i.e. that perceives it. For this purpose, we make the following definition of the Riemann-Stieltjes integral, that deviates sightly from the original definition in the sense that it recognizes mass points at the left endpoint of the area of integration: Let $f : [a,b] \to \mathbb{R}$ be continuous and $g : \mathbb{R} \to \mathbb{R}$ be of bounded variation and right continuous. Let μ_g denote the signed measure associated to g and let $supp(\mu_g) \subseteq [a,b]$. In this situation,

$$\begin{aligned} supp(g) &:= supp(\mu_g) \\ &:= \{x \in \mathbb{R} \,|\, \mu_{g^+}((x-r, x+r]) > 0 \text{ or } \mu_{g^-}((x-r, x+r]) > 0, \text{ for all } r > 0\} \end{aligned}$$

denotes the support of g or μ_g respectively. Furthermore, assume that the Riemann-Stieltjes integral $\int\limits_{[a,b]}^{(R)} f(t)\, dg(t)$ exists. We then define

$$\widetilde{g}(x) := \begin{cases} g(x), & x \in [a,b], \\ 0, & x \in [-\infty, a) \end{cases} \quad \text{and} \quad \overline{f}(x) := \begin{cases} f(x), & x \in [a,b], \\ f(a), & x \in [-\infty, a) \end{cases}$$

and set

$$\int\limits_{[a,b]}^{(R^*)} f(x)\, dg(x) := \int\limits_{[-\infty,b]}^{(R)} \overline{f}(x)\, d\tilde{g}(x).$$

If we denote with $\mu_{\widetilde{g}}$ the measure associated to $\widetilde{g}$, similar to (2.8), it follows that

$$\int\limits_{[a,b]} f(x)\, d\mu_g(x) = \int\limits_{(-\infty,b]} \overline{f}(x)\, d\mu_{\tilde{g}}(x) = \int\limits_{[-\infty,b]}^{(R)} \overline{f}(x)\, d\tilde{g}(x) = \int\limits_{[a,b]}^{(R^*)} f(x)\, dg(x) \tag{2.9}$$

holds true. The above equations are true, since $\overline{f}(x) = f(x)$, $x \in [a,b]$, $\widetilde{g}(x) = 0$, $x \in [-\infty, a)$ and $supp(\mu_{\tilde{g}}) = supp(\mu_g) \subseteq [a,b]$.

The following theorem is strongly based on Theorem 2.2.2, but is new in the literature. Note that Theorem 3.1 in Bischoff and Miller (2000) is very similar to the following theorem, however no proof is given there.

Theorem 2.2.4.

Let the regression functions $f_1, \ldots, f_p$, $p \in \mathbb{N}$, be continuous, real-valued and known functions of bounded variation that are linearly independent in $\mathcal{L}^2([a,b], F)$. Furthermore, let

$$F_n \xrightarrow[n\to\infty]{} F \quad \text{uniformly,}$$

where F is a distribution function with $supp(F) \subseteq [a,b]$ and $F^-(1) = b$. Then,

i) $$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(\cdot) \quad \text{in } C[0,1],$$

where $B_{f,F}(\cdot)$ is a Gaussian process that is given by

$$B_{f,F}(z) = B(z) - \left(\int_{[a,b]} f(t)\, dF_z(t) \right)^T J^{-1} \int_{[a,b]} f(t)\, dB(F(t)), \quad z \in [0,1].$$

ii) $$\frac{1}{\sigma\sqrt{n}} (T_n(\hat{\varepsilon}_n) \circ F_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (B_{f,F} \circ F)(\cdot) \quad \text{in } (D[a,b], \|\cdot\|_\infty),$$

where $(B_{f,F} \circ F)(\cdot)$ is a Gaussian process that is given by

$$(B_{f,F} \circ F)(s) = B(F(s)) - \left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1} \int_{[a,b]} f(t)\, dB(F(t)), \quad s \in [a,b].$$

With regard to the above theorem, the following two facts should be noted.

- Although the last integrals in i) and ii) look as if they should be Riemann-Stieltjes integrals, we can consider them to be measure integrals. See the explanation of equation (2.15) in the proof of Theorem 2.2.4 in this regard.

- The prerequisite, $F^-(1) = b$, is a purely formal one. Since we are in the situation of a linear regression model with fixed design, the distribution function F of the limit design is known. Therefore any experimental region $[a,b]$ that includes $supp(F)$ is sufficient for our analysis. Consequently, we can chose the "non-superfluous" experimental region where $F^-(1) = b$. Therefore, in the further course of this work, whenever Theorem 2.2.4 is applied, we will not explicitly address the condition $F^-(1) = b$, but implicitly assume that the experimental region $[a,b]$ was chosen such that the condition is satisfied.

Proof: Since the difference between Theorem 2.2.2 and Theorem 2.2.4 consists exclusively in the representation of the last factor of the residual partial sum limit processes, it is sufficient to prove only statement ii). The proof of statement i) is exactly the same.
Let $\varepsilon > 0$. Consider the fixed-design linear regression model that can be associated to the given model by enlarging the experimental region by ε to the left. To be more specific, we consider the linear regression model with the experimental region $[a-\varepsilon, b]$ together with the empirical distribution functions F_n, $n \in \mathbb{N}$, that were originally given. That is, $supp(F_n) \subseteq [a,b]$ for all $n \in \mathbb{N}$. Furthermore, in the new model, we consider the regression functions $f^* := (f_1^*, \ldots, f_p^*)^T$, $p \in \mathbb{N}$, where

$$f_i^*(x) := \begin{cases} f_i(x), & x \in [a,b], \\ f_i(a), & x \in [a-\varepsilon, a) \end{cases}, \quad i = 1, \ldots, p.$$

It holds true that $F_n \xrightarrow[n\to\infty]{} F$ uniformly and $supp(F) \subseteq [a-\varepsilon, b]$. Furthermore, $f_1^*, \ldots, f_p^*$ are continuous, real-valued and known functions of bounded variation that are linearly independent in $\mathcal{L}^2([a-\varepsilon, b], F)$. Thus, Theorem 2.2.2 is applicable and states that

$$\frac{1}{\sigma\sqrt{n}}(T_n(\hat{\varepsilon}_n) \circ F_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (B_{f^*,F} \circ F)(\cdot) \quad \text{in } (D[a-\varepsilon, b], \|\cdot\|_\infty).$$

Here, $(B_{f^*,F} \circ F)(\cdot)$ is a Gaussian process that for $s \in [a-\varepsilon, b]$ is given by

$$(B_{f^*,F} \circ F)(s) = B(F(s)) + \tag{2.10}$$

$$+ \left(\int_{[a-\varepsilon,s]} f^*(t)\, dF(t) \right)^T J^{-1} \left(\int_{[a-\varepsilon,b]}^{(R)} B(F(t))\, df^*(t) - B(1) f^*(F^-(1)) \right). \tag{2.11}$$

Since $supp(F) \subseteq [a,b]$ and $F(t) = 0$ for $t \in [a-\varepsilon, a)$ it holds true that

$$\int_{[a-\varepsilon,s]} f^*(t)\, dF(t) = \int_{[a,s]} f^*(t)\, dF(t) = \int_{[a,s]} f(t)\, dF(t), \quad s \in [a,b]. \tag{2.12}$$

Furthermore,

$$\int_{[a-\varepsilon,b]}^{(R)} B(F(t))\, df^*(t) - B(1) f^*(F^-(1))$$

$$= -f^*(b)B(F(b)) + f^*(a-\varepsilon)B(F(a-\varepsilon)) + \int_{[a-\varepsilon,b]}^{(R)} B(F(t))\, df^*(t) \tag{2.13}$$

$$= -\int_{[a-\varepsilon,b]}^{(R)} f^*(t)\, dB(F(t)) = -\int_{[a,b]}^{(R^*)} f(t)\, dB(F(t)) \tag{2.14}$$

$$= - \int\limits_{[a,b]} f(t)\, dB(F(t)). \tag{2.15}$$

The following should be noted with regard to the above equations:

Regarding (2.13) This equation holds true because of the following two facts. Firstly, $B(F(a-\varepsilon)) = B(0) = 0$, as $supp(F) \subseteq [a,b]$, and secondly, we assumed that $F^-(1) = b$ and thus $F(b) = 1$ (see Lemma 7.1.1).

Regarding (2.14) First, integration by parts was used. See for example Theorem 1.2.7 in Strook (1994). Then, since $supp(F) \subseteq [a,b]$ and $\bar{f}(x) = \widetilde{f}(x)$ for all $x \in [a-\varepsilon, b]$, it holds true that

$$\int\limits_{[a-\varepsilon,b]}^{(R)} f^*(t)\, dB(F(t)) = \int\limits_{[-\infty,b]}^{(R)} \overline{f}(t)\, d(\widetilde{B \circ F})(t) = \int\limits_{[a,b]}^{(R^*)} f(t)\, dB(F(t)).$$

Regarding (2.15) To see that this equation is valid, see equation (2.9) in Remark 2.2.3. It should be noted that writing $\int\limits_{[a,b]} f(t)\, dB(F(t))$ — i.e. considering it a measure integral — is a slight abuse of notation that will simplify many of the notations that follow. That is, at first glance $\int\limits_{[a,b]} f(t)\, dB(F(t))$ can only be understood as a Riemann-Stieltjes integral, since $B(F(t))$ might not be of bounded variation and therefore no measure must exist that can be associated to it. But, since

$$\begin{aligned}
\int\limits_{[a,b]} f(t)\, dB(F(t)) &= \int\limits_{[a,b]}^{(R^*)} f(t)\, dB(F(t)) = \int\limits_{[-\infty,b]}^{(R)} \overline{f}(t)\, d(\widetilde{B \circ F})(t) \\
&= f(a)B(F(a)) + \int\limits_{[a,b]}^{(R)} f(t)\, dB(F(t)) \\
&= f(a)B(F(a)) + f(b)B(F(b)) - f(a)B(F(a)) - \int\limits_{[a,b]}^{(R)} B(F(t))\, df(t) \\
&= f(b)B(F(b)) - \int\limits_{[a,b]} B(F(t))\, d\mu_f(t)
\end{aligned}$$

holds true, where μ_f denotes the measure associated to f, $\int\limits_{[a,b]} f(t)\, dB(F(t))$ can be considered as a measure Integral.

Finally, by inserting (2.12) and (2.15) in (2.10), we get

$$(B_{f,F} \circ F)(s) = B(F(s)) - \left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1} \int_{[a,b]} f(t)\, dB(F(t)), \quad s \in [a,b],$$

which completes the proof. □

In the next corollary, we compute the mean and covariance function of $(B_{f,F} \circ F)(\cdot)$ in the situation of Theorem 2.2.4.

Corollary 2.2.5.

Retain the assumptions of Theorem 2.2.4. Then the mean function m and the covariance function K of the stochastic process $(B_{f,F} \circ F)(\cdot)$ introduced in Theorem 2.2.4 are given by

$$m(s) = 0, \quad s \in [a,b],$$

and, for $s, t \in [a,b]$, by

$$K(s,t) = F(\min\{s,t\}) - \left(\int_{[a,s]} f(x)\, dF(x) \right)^T \left(\int_{[a,b]} f(t) f^T(t)\, dF(t) \right)^{-1} \int_{[a,t]} f(x)\, dF(x).$$

Proof: This corollary is similar to Lemma 3.1. in Bischoff and Miller (2000) and its proof therefore is similar to the one stated therein on pages 674 and 675. Nevertheless, since the present situation is not exactly the same, we will give a detailed proof for Corollary 2.2.5 on page 129 and the following pages in Section 7.1 of the appendix. □

2.2.2 Linear regression models and the uniform fixed design

The main result of this section is Theorem 2.2.9, which states that when only the distribution of the residual partial sum limit process is of interest, one can always assume the design of a linear regression model with fixed design to be the equidistant design on the unit interval $[0, 1]$. We start our way to this theorem with the definition of so-called generalised inverse functions.

Definition 2.2.6.

A function $G : \mathbb{R} \to \mathbb{R}$ is called a measure generating function if it is monotonically increasing and continuous from the right. For a given measure generating function G with values in $[c, d]$, $c \leq d \in [-\infty, \infty]$,

$$G^- : (c, d) \longrightarrow \mathbb{R},\ u \longmapsto \inf\{x \in \mathbb{R} | \, G(x) \geq u\}$$

denotes the generalised inverse function associated to G. If necessary, we may extend G^- to $[c, d]$ by defining

$$G^-(c) := \inf\{x \in \mathbb{R} | G(x) > c\} \in [-\infty, \infty)$$

and

$$G^-(d) := \sup\{x \in \mathbb{R} | G(x) < d\} \in (-\infty, \infty].$$

The following lemma provides a characterization of bounded variation functions and constitutes an important ingredient for proving the main result of this section. However, the lemma is an interesting result in itself that, to our knowledge, does not appear in the literature so far.

Lemma 2.2.7.

Let F be a distribution function with $supp(F) \subseteq [c, d]$, $c < d$, and let $f_1, \ldots, f_p : [c, d] \longrightarrow \mathbb{R}$, $p \in \mathbb{N}$, be functions of bounded variation that are continuous from the left and right continuous in c. Furthermore, let $a < b$ be real numbers.
Then, there is a measure generating function $G : \mathbb{R} \longrightarrow [c, d]$ with $supp(G) \subseteq [a, b]$ and functions $g_1, \ldots, g_p : [a, b] \longrightarrow \mathbb{R}$ that are Lipschitz continuous, such that

$$f_i \ = \ g_i \circ G^-, \quad i = 1, \ldots, p.$$

Furthermore, if the functions $f_1, \ldots, f_p$ are linearly independent in $\mathcal{L}^2([c, d], F)$, then the functions $g_1, \ldots, g_p$ are linearly independent in $\mathcal{L}^2([a, b], F \circ G)$.
If, in particular, $c = 0$ and $d = 1$, G is a distribution function and G^- is its associated quantile function.

Proof: The proof of this Lemma is deferred to Section 7.1 of the appendix, starting with page 135. □

The following theorem is new in the literature and describes a simple way of transforming a linear regression model with fixed design, as introduced in Section 2.2.1, into a linear regression model that has an equidistant design on the unit interval. However, the applicability of the theorem is somewhat limited, since the regression functions in the transformed model depend on the dimension $n \in \mathbb{N}$ of the outcome vector Y_n, i.e. the number of design points $t_{n1}, \ldots, t_{nn}$.

Theorem 2.2.8.

Consider a linear regression model with fixed design

$$Y_n = X_n\beta + \varepsilon_n, \quad \beta \in \mathbb{R}^p,\, p \in \mathbb{N},\, n \in \mathbb{N}, \tag{2.16}$$

where $[a,b]$, $a < b$, is the experimental region, $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [a,b]$ is the triangular array of design points whose corresponding sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$ converges uniformly to a distribution function F, $X_n = (f_r(t_{ni}))_{i=1,r=1}^{n\quad p}$ is the design matrix where the regression functions $f_1, \ldots, f_p : [a,b] \to \mathbb{R}$ are continuous, of bounded variation and linearly independent in $\mathcal{L}^2([a,b], F)$ and $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$ are the real-valued iid regression errors with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

Then, this regression model is equal to the regression model

$$Y_n = \left((f_r \circ F_n^-)\left(\frac{i}{n}\right) \right)_{i=1,r=1}^{n\quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{2.17}$$

where $[0,1]$ is the experimental region, $\left(\left(\frac{i}{n}\right)_{i=1}^n\right)_{n\in\mathbb{N}}$ is the triangular array of equidistant design points, F_n^- denotes the quantile function corresponding to the empirical distribution function F_n, which in turn corresponds to the n-th row of the triangular array $((t_{ni})_{i=1}^n)$ of original design points, and $f_1 \circ F_n^-, \ldots, f_p \circ F_n^- : [0,1] \to \mathbb{R}$ are the regression functions that are of bounded variation and, for n large enough, linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\lambda_{[0,1]})$.
For the residual partial sum process of model (2.17), the weak convergence

$$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(\cdot) \quad \text{in } C[0,1]$$

holds true.

Proof:
For $n \in \mathbb{N}$ and $i = 1, \ldots, n$, we have $F_n^-\left(\frac{i}{n}\right) = t_{ni}$, since

$$F_n^-\left(\frac{i}{n}\right) = \inf\left\{ x \in \mathbb{R} \,\middle|\, \frac{1}{n}\sum_{j=1}^n \mathbb{1}_{(-\infty,x]}(t_{nj}) \geq \frac{i}{n} \right\} = t_{ni}$$

holds true, as we have assumed that without loss of generality $t_{n1} \leq t_{n2} \leq \cdots \leq t_{nn}$ holds. Therefore, models (2.16) and (2.17) are equal.

Regarding the properties of the regression functions in the transformed regression model, we know that, since the regression functions $f_1, \ldots, f_p$ in model (2.16) are of bounded variation and F_n^- are monotonically increasing for all $n \in \mathbb{N}$, the transformed regression functions $f_1 \circ F_n^-, \ldots, f_p \circ F_n^-$ in model (2.17) are of bounded variation, too. Note that unlike all the regression models considered so far, the regression functions $f_1 \circ F_n^-, \ldots, f_p \circ F_n^-$ in model (2.17) depend on $n \in \mathbb{N}$. Furthermore, since the regression functions $f_1, \ldots, f_p$ in model (2.16) are linearly independent in $\mathcal{L}^2([a,b], F)$, we can conclude, as we will establish thoroughly in the proof of Theorem 2.2.9, that the functions $f_1 \circ F^-, \ldots, f_p \circ F^-$ are linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$.
From this, we can infer that the transformed regression functions $f_1 \circ F_n^-, \ldots, f_p \circ F_n^-$ in model (2.17) are linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$ for $n \in \mathbb{N}$ large enough, too. This is indeed the case, since we assumed that F_n converges to F uniformly, which implies that $F_n(t)$ converges to $F(t)$ at every t at which F is continuous, and this in turn implies, according to Lemma 21.2 in Vaart (1998), that $F_n^-(u)$ converges to $F^-(u)$ at every u in which F^- is continuous. This holds everywhere, except on an at most countable set.
Finally, since the models (2.16) and (2.17) are equal, their respective residual partial sum processes coincide and, according to Theorem 2.2.2, converge weakly to $B_{f,F}$ in $C[0,1]$. □

Similar to the case of a linear regression model with random design, it is now our aim to prove that in the situation of linear regression models with fixed designs one can assume a more specific situation. To be more precise, we are going to investigate the relation between a linear regression model with fixed design

$$Y_n = X_n\beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{2.18}$$

where

- $[a,b]$, $a < b$, is the experimental region,
- $\beta = (\beta_1, \ldots, \beta_p)^T \in \mathbb{R}^p$, $p \in \mathbb{N}$, are the unknown regression coefficients,
- $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [a,b]$ is the triangular array of design points whose corresponding sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$ converges uniformly to a distribution function F,
- $X_n = (f_r(t_{ni}))_{i=1,r=1}^{n\ p}$ is the design matrix where the regression functions $f_1, \ldots, f_p : [a,b] \to \mathbb{R}$ are continuous, of bounded variation and linearly independent in $\mathcal{L}^2([a,b], F)$,
- $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$ are the real-valued iid regression errors with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$,

and a corresponding linear regression model with *uniform* fixed design

$$\widetilde{Y}_n = \widetilde{X}_n\beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{2.19}$$

where

- $[0,1]$ is the experimental region,
- $\beta = (\beta_1, \ldots, \beta_p)^T \in \mathbb{R}^p$, $p \in \mathbb{N}$, are the unknown regression coefficients (exactly those of model (2.18)),
- $((\frac{i}{n})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [0,1]$ is the triangular array of equidistant design points whose corresponding sequence of empirical distribution functions converges uniformly to the distribution function of the uniform distribution on the unit interval, i.e. the identity function on $[0,1]$,
- $\widetilde{X}_n := (\widetilde{f}_r\left(\frac{i}{n}\right))_{i=1,r=1}^{n\quad p} := \left((f_r \circ F^-)\left(\frac{i}{n}\right)\right)_{i=1,r=1}^{n\quad p}$ is the design matrix where the regression functions $\widetilde{f}_1 := f_1 \circ F^-, \ldots, \widetilde{f}_p := f_p \circ F^- : [0,1] \to \mathbb{R}$ are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$,
- $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$ are the real-valued iid regression errors (exactly those of model (2.18)) with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

The following assertion is the main theorem of this section and, to our knowledge, new in the literature.

Theorem 2.2.9.

Let a linear regression model with fixed design be given (cf. model (2.18)). Then, the distribution of the residual partial sum limit processes of this model is equal to that of the corresponding linear regression model with uniform fixed design (cf. model (2.19)). In particular, in model both models, the weak convergence

$$\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(z) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(z) \quad \text{in } C[0,1]$$

holds true.

Proof: The proof of this theorem is deferred to the pages 148 to 151 in Section 7.1 of the appendix. □

Note that without loss of generality we assume in the above theorem that the regression errors $\varepsilon_{ni}, i \in \{1, \ldots, n\}, n \in \mathbb{N}$, are of variance 1. Therefore, the normalizing factor $\frac{1}{\sigma}$ in front of the partial sum operator $\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot)$ is not needed in this case.
Observe, furthermore, the following subtle but crucial difference to Theorem 2.2.8. A linear regression model with fixed design and the corresponding model with uniform fixed design are not entirely equal because their design matrices X_n and $\widetilde{X}_n$ may differ. More precisely, a design point t_{ni} in the general case need not be equal to $F^-\left(\frac{i}{n}\right)$ in the uniform case, since F is already the *limit* design distribution, whereas the sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$ corresponding to the triangular array $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ in the

general fixed design setup converges uniformly to F. Therefore, in a certain sense, the equality of both models only holds in the limit.

The following lemma contains a transformation of the residual partial sum limit process $B_{f,F}$, which will be of importance for the subsequent theorem.

Lemma 2.2.10.

Let the residual partial sum process $B_{f,F}(\cdot)$ of a linear regression model with fixed design or uniformly fixed design be given. Then, it holds true that

$$B_{f,F}(z) = B_{f\circ F^-,\lambda_{[0,1]}}(z), \quad z \in [0,1].$$

Proof: The proof of this Lemma is deferred to Section 7.1 of the appendix, starting with page 152. □

In Theorem 2.2.2, the weak convergence of the residual partial sum process to the limit process $B_{f,F}$ was stated for a linear regression model with fixed design. In particular, it was assumed that the regression functions $f_1, \ldots, f_p : [a,b] \longrightarrow \mathbb{R}$ are continuous and of bounded variation. Lemma 2.2.7 now gives us the possibility to drop the assumption of continuity of the regression functions. To our knowledge, this result is new in the literature.

Theorem 2.2.11.

Consider a linear regression model with fixed design

$$Y_n = X_n\beta + \varepsilon_n, \quad \beta \in \mathbb{R}^p,\ p \in \mathbb{N},\ n \in \mathbb{N}, \tag{2.20}$$

where $[a,b]$, $a < b$, is the experimental region, $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [a,b]$ is the triangular array of design points whose corresponding sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$ converges uniformly to a distribution function F, $X_n = (f_r(t_{ni}))_{i=1,r=1}^{n\ \ p}$ is the design matrix where the regression functions $f_1, \ldots, f_p : [a,b] \to \mathbb{R}$ are of bounded variation and linearly independent in $\mathcal{L}^2([a,b],F)$ and $\varepsilon_n = (\varepsilon_{n1}, \ldots, \varepsilon_{nn})^T$ are the real-valued iid regression errors with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$. Then,

$$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(\cdot) \quad \text{in } C[0,1]$$

holds true.

Proof: The proof of this theorem is deferred to Section 7.1 of the appendix, starting with page 156. □

3 Commonalities of the two cumulated sum limit processes

In this chapter, we will investigate commonalities and disparities between random design and fixed design linear regression models as introduced in Section 2.1 and Section 2.2 respectively, with regard to their residual CUSUM processes.
In the first section we will have a closer look at Theorem 2.1.7 on page 20 as well as Theorem 2.2.4 on page 35, and give a theorem which states the equality of distributions of the limit of the marked empirical process and the residual partial process under certain circumstances. This result can be seen as the intuitive building block for discussing the similarities between random and fixed design linear regression models when it comes to their respective residual CUSUM limit processes.
In the second section, we identify the underlying common properties between random design and fixed design linear regression models that account for their similarities. Finally, we state a generic linear regression model under which both linear regression models with random and fixed designs can be subsumed and give a functional central limit theorem for the CUSUM process of regression residuals in this generic situation.

3.1 Equality of the two cumulated sum limit processes

In this section, we prove a theorem which, simply put, states that in the situation of a univariate linear regression model, the CUSUM limit process of the regression residuals in the case of a model with random design has the same distribution as the one we obtain under a model with fixed design. Thus, results about the structure of such a limit process based on one of the two designs will always apply to the other.
Note that in the context of a linear model with random design the CUSUM process of the regression residuals has been referred to as the marked empirical process and in case of a model with fixed design the CUSUM process of the regression residuals was referred to as the residual partial sum process.

The announced theorem is new in the literature to our knowledge and reads as follows.

Theorem 3.1.1.

Let the assumptions of Theorem 2.1.7 and Theorem 2.2.4 be satisfied. Therefore, in particular, we are given the regression functions

$$f = (f_1, \ldots, f_p)^T, \quad p \in \mathbb{N},$$

that are continuous, of bounded variation, real-valued and known functions on the experimental region $[a, b] \subseteq \mathbb{R}$ that are linearly independent in $\mathcal{L}^2([a, b], F)$.

Then it holds true, that

$$(B_{f,F} \circ F)(\cdot) \sim R^1_\infty(\cdot).$$

Proof: If the conditions of both Theorem 2.1.7 and Theorem 2.2.4 are satisfied, the following holds true:

i) According to Theorem 2.1.7 from page 20, $R^1_\infty(\cdot)$ is a Gaussian process with the mean function

$$m(s) = 0, \quad s \in \mathbb{R},$$

and the covariance function

$$K^1(s,t) = F(\min\{s,t\}) - \left(\int_{[-\infty,s]} f(x)\, dF(x) \right)^T J^{-1} \int_{[-\infty,t]} f(x)\, dF(x), \ s,t \in [-\infty,\infty]. \tag{3.1}$$

Since, in the present theorem, the experimental region is given as the compact interval $[a, b]$, which means that the design distribution function F has its support in $[a, b]$ and the regression functions $f_1, \ldots, f_n$ are functions with domain $[a, b]$, equation (3.1), for $s, t \in [a, b]$, reduces to

$$K^1(s,t) = F(\min\{s,t\}) - \left(\int_{[a,s]} f(x)\, dF(x) \right)^T \left(\int_{[a,b]} f(x) f^T(x)\, dF(x) \right)^{-1} \int_{[a,t]} f(x)\, dF(x).$$

ii) According to Corollary 2.2.5 on page 38, $(B_{f,F} \circ F)(\cdot)$ is a Gaussian process with the mean function

$$m(s) = 0, \quad s \in [a, b].$$

Furthermore, for $s, t \in [a, b]$, the covariance function reads

$$K(s,t) = F(\min\{s,t\}) - \left(\int_{[a,s]} f(x)\, dF(x) \right)^T \left(\int_{[a,b]} f(x) f^T(x)\, dF(x) \right)^{-1} \int_{[a,t]} f(x)\, dF(x)$$

To sum up, we therefore know that both processes are centred Gaussian processes with the same covariance function. Since it is well-known that the distribution of a Gaussian process is uniquely identified through its mean function and its covariance function (see for example Kapitel 7.6 in Gänsler and Stute (1977)), we can conclude that

$$(B_{f,F} \circ F)(\cdot) \sim R^1_\infty(\cdot).$$

as Gaussian processes. □

Note that in the situation of a linear regression model with random design, no analytical formula for the limit process $R^1_\infty(\cdot)$ of the marked empirical process $R^1_n(\cdot)$ is available in the literature. In contrast, in a linear regression model with fixed design, there is an analytical formula for the residual partial sum limit process $B_{f,F}$.

The result below is a corollary to Theorem 3.1.1 in the special situation that a linear regression model with uniform random design (see Theorem 2.1.12) as well as a linear regression model with uniform fixed design (as introduced directly before Theorem 2.2.9) with the same regression functions are given.

Corollary 3.1.2.

Let a linear regression model with uniform random design as well as a linear regression model with uniform fixed design with the same regression functions be given. Thus, in particular, $\lambda\!\!\lambda_{[0,1]}$ is the design distribution on the unit interval and

$$f = (f_1, \ldots, f_p)^T : [0,1] \longrightarrow \mathbb{R}^p, \quad p \in \mathbb{N},$$

are the regression functions that are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\lambda_{[0,1]})$.

Then it holds true that

$$B_{f,\lambda\!\!\lambda_{[0,1]}}(\cdot) \sim \bar{R}^1_\infty(\cdot).$$

3.2 A generic linear regression model for goodness-of-fit tests

It is a standard approach in the literature to base analyses of the goodness-of-fit of regression models on their residual CUSUM processes. See for example Stute, Thies, and Zhu (1998), who stated that "there is good reason to base model diagnostics on the CUSUM process of the residuals [...]".
Therefore, although linear regression models with random designs and linear regression models with fixed designs, as well as their respective residual processes, appear to be different, Theorem 3.1.1 and Corollary 3.1.2 are arguments in favour of the viewpoint that CUSUM based model diagnostics do not differ for them. The three facts below shed light on this in more detail.

i) Based on the result of Theorem 2.1.12, we can restrict our analysis of marked empirical processes of regression residuals in linear regression models with random designs to their corresponding uniform marked empirical process in the linear regression model with uniform random design. Such a corresponding linear regression model with uniform random design has the form

$$Y = (f \circ F^-)^T(U)\theta + \varepsilon,$$

where

- F is a distribution function, $f_1, \ldots, f_p$ are known real-valued functions that are linearly independent in $\mathcal{L}^2([-\infty,\infty], F)$ and therefore the regression functions $(f \circ F^-) = (f_1 \circ F^-, ..., f_p \circ F^-)^T : [0,1] \longrightarrow \mathbb{R}^p$ are linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$,
- the design is sampled according to a random variable U with the distribution function

$$F_U(t) = \begin{cases} 0, & t \leq 0, \\ \lim\limits_{x \uparrow F^-(t)} F(x), & t \notin F(\mathbb{R}) \text{ and } t \in (0,1), \\ t, & t \in F(\mathbb{R}) \text{ and } t \in (0,1), \\ 1, & t \geq 1 \end{cases}$$

  and, in particular, $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$,
- $\theta \in \mathbb{R}^p$ are the regression coefficients estimated using the LSE,
- $\mathbb{E}(\varepsilon|U) = 0$ and $\mathbb{V}ar(\varepsilon|U = u) = 1$ for all $u \in [0,1]$.

In this situation, the CUSUM process of the regression residuals converges weakly to $\overline{R}^1_\infty(\cdot)$ in $D[0,1]$. Here, $\overline{R}^1_\infty(\cdot)$ is a centred Gaussian process whose covariance function, for $s, t \in [0,1]$, is

$$K^1(s,t) = F_U(\min\{s,t\}) - \left(\int\limits_{[0,F_U(s)]} (f \circ F^-)(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) \right)^T \times \tag{3.2}$$

$$\times \left(\int\limits_{[0,1]} (f \circ F^-)(x)(f \circ F^-)^T(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) \right)^{-1} \int\limits_{[0,F_U(t)]} (f \circ F^-)(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x).$$

ii) Based on the result of Theorem 2.2.9, we can restrict our analysis of residual partial sum processes of regression residuals in linear regression models with fixed designs to the case of their corresponding linear regression model with uniform fixed design and the respective residual partial sum process. Such a corresponding linear regression model with uniform fixed design has the form

$$Y_n = \left((f_r \circ F^-) \left(\frac{i}{n} \right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N},$$

where

- F is a distribution function, $f_1, \ldots, f_p$ are known real-valued functions that are continuous, of bounded variation and linear independent in $\mathcal{L}^2([a,b], F)$ and therefore the regression functions $f \circ F^- = (f_1 \circ F^-, \ldots, f_p \circ F^-)^T : [0,1] \to \mathbb{R}^p$ are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$,
- $\beta \in \mathbb{R}^p$ are the regression coefficients estimated using the LSE,
- $\left(\left(\frac{i}{n} \right)_{i=1}^n \right)_{n \in \mathbb{N}} \subseteq [0,1]$ is the triangular array of equidistant design points,
- $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ are iid random variables with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

In this situation, the CUSUM process of the regression residuals converges in distribution to $B_{f,F}$ in $C[0,1]$. Furthermore, according to Lemma 2.2.10, $B_{f,F}(z) = B_{f \circ F^-, \lambda\!\!\!\lambda_{[0,1]}}(z)$ for all $z \in [0,1]$ and thus, according to Corollary 2.2.5, the residual CUSUM limit process is a centred Gaussian process whose covariance function, for $s,t \in [0,1]$, is

$$K(s,t) = \min\{s,t\} - \left(\int\limits_{[0,s]} (f \circ F^-)(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) \right)^T \times \tag{3.3}$$

$$\times \left(\int\limits_{[0,1]} (f \circ F^-)(x)\, (f \circ F^-)^T(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) \right)^{-1} \int\limits_{[0,t]} (f \circ F^-)(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x).$$

iii) According to Corollary 3.1.2, we know that if $\lambda\!\!\lambda_{[0,1]}$ is the design distribution, the distribution of the limit process $\overline{R}^1_\infty(\cdot)$ of the marked empirical process in the situation of a linear regression model with uniform random design coincides with the distribution of the residual partial sum limit process $B_{f,\lambda\!\!\lambda_{[0,1]}}(\cdot)$ in the situation of the same linear regression model, but with a uniform fixed design.
For general linear regression models with random design and fixed design that have the same regression functions $f_1, \ldots, f_p$ and the same design distribution F, we considered their respective corresponding uniform designs in points i) and ii) of this enumeration. In particular, we noted that under both designs the respective residual CUSUM limit processes are centred Gaussian processes with the covariance functions given in (3.2) and (3.3). Therefore, the distributions of their respective residual CUSUM limit processes $\overline{R}^1_\infty(u)$ and $B_{f,F}(u)$ coincide for all $u \in F(\mathbb{R}) \cup \{0,1\}$, since in this case $F_U(u) = u$ holds true.
Otherwise, if $u \in (0,1)$ is not in $F(\mathbb{R})$ (that is, F has a jump at $F^-(u)$), $\overline{R}^1_\infty$ has the constant value

$$\overline{R}^1_\infty\left(\lim_{x\uparrow F^-(u)} F(x)\right) \quad \text{in the interval} \quad \left[\lim_{x\uparrow F^-(u)} F(x), F(F^-(u))\right)$$

and then jumps to the value $\overline{R}^1_\infty\left(F(F^-(u))\right)$. In contrast, since the function $f \circ F^-$ is constant in the intervals that are not in $F(\mathbb{R})$, $B_{f,F} = B_{f\circ F^-,\lambda\!\!\lambda_{[0,1]}}$ interpolates these two values linearly. These jumps of F originate in design points with positive probability. The magnitude of such a jump's probability determines the size of the interval on which the residual CUSUM limit processes differ. It can thus be noted that $\overline{R}^1_\infty$ and $B_{f,F}$ differ only in the representation of those intervals of their domain belonging to design points with positive probabilities. See Figure 2.4 and Figure 2.5 and in particular the discrepancy between the graph of F_U (see Figure 2.5a) and the distribution function of $\lambda\!\!\lambda_{[0,1]}$, which originates from the jump of F at q_3 (see Figure 2.4a).
To summarize, the distributions of residual CUSUM limit processes of linear regression models with the same regression functions $f_1, \ldots, f_p$ and the same design distribution F coincide under uniform fixed and uniform random designs on $F(\mathbb{R}) \cup \{0,1\}$. For this reason, test statistics in goodness-of-fit tests based on $\overline{R}^1_n$ in the random design case and $\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)$ in the fixed design case agree on $\sigma(F^-)$, as their asymptotic distributions are equal.
Note that the difference between $\overline{R}^1_\infty$ and $B_{f,F}$ is not in content but only in the representation of design points with positive probability. In the literature on random designs, this discreteness is passed from the design to the marked empirical process and, from there, to limit process $\overline{R}^1_\infty$. In contrast, in the literature on fixed designs, the residual partial sum process is continuous, piecewise linear and equidistantly embedded in $[0,1]$. Thus, in the limit process $B_{f,F}$, the positive probability of such a design point is uniformly distributed over the corresponding interval that is not in $F(\mathbb{R})$ and therefore, $B_{f,F}$ linearly interpolates the jump in $\overline{R}^1_\infty$.

The three facts outlined above therefore justify the following concluding statement:

When it comes to goodness-of-fit tests in linear regression models based on the asymptotic distribution of residual CUSUM processes, one can assume without loss of generality the following generic linear regression model.

Definition 3.2.1.

A generic linear regression model has the form

$$\begin{pmatrix} Y_{n1} \\ \vdots \\ Y_{nn} \end{pmatrix} = \begin{pmatrix} f_1\left(\frac{1}{n}\right) & \dots & f_p\left(\frac{1}{n}\right) \\ \vdots & \ddots & \vdots \\ f_1(1) & \dots & f_p(1) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nn} \end{pmatrix}, \quad n \in \mathbb{N},$$

where

- $f = (f_1, \dots, f_p)^T : [0,1] \to \mathbb{R}^p, p \in \mathbb{N}$ are regression functions that are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\lambda_{[0,1]})$,
- $X_n := \left(f_r\left(\frac{i}{n}\right)\right)_{i=1,r=1}^{n \quad p}$ is refereed to as the n-th design matrix,
- $\beta \in \mathbb{R}^p$ are the regression coefficients,
- $\frac{1}{n}, \frac{2}{n}, \dots, 1$ is the n-th row of the triangular array $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ of equidistant design points,
- $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ are iid random variables with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$,
- $Y_n = (Y_{n1}, \dots, Y_{nn})^T$ is the vector of outcomes.

With regard to the generic linear regression model and the fact that one can assume a generic linear regression model when it comes to goodness-of-fit tests in linear regression models based on the asymptotic distribution of residual CUSUM processes is should be noted that henceforth, any result about goodness-of-fit tests applicable to residual partial sum processes in linear models with fixed designs can also be used for marked empirical processes in linear regression models with random designs, and vice versa.

In the situation of a generic linear regression model, we can immediately state the following corollary about the convergence of the residual CUSUM processes.

Corollary 3.2.2.

Let a generic linear regression according to Definition 3.2.1 be given. Then, the marked empirical process

$$R_n^1(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}\left(\frac{i}{n}\right) \hat{\varepsilon}_{ni}$$

as well as the partial sum process

$$\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \hat{\varepsilon}_{ni} + (nx - \lfloor nx \rfloor)\, \hat{\varepsilon}_{n,\lfloor nx \rfloor + 1}$$

of the least-squares residuals converge to the limit process

$$B_{f,\lambda_{[0,1]}}(s) = B(s) - \left(\int_{[0,s]} f(t)\, d\lambda_{[0,1]}(t) \right)^T \left(\int_{[0,1]} f(t)\, f^T(t)\, d\lambda_{[0,1]}(t) \right)^{-1} \int_{[0,1]} f(t)\, dB(t),$$

$s \in [0,1]$, weakly in $C[0,1]$ as $n \to \infty$.

From another point of view, this means that both random design models and fixed design models can be interpreted as examples of a generic linear regression model. Thus, only the following two disparities between linear regression models with random and fixed designs remain.

Disparity i) Given a linear regression model with random design, the corresponding linear regression model with uniform random design reads

$$Y = (f \circ F^-)^T(U)\theta + \varepsilon$$

where U is a uniformly distributed random variable on $\sigma(F^-)$. In contrast, in a linear regression model with fixed design, the corresponding linear regression model with uniform fixed design reads

$$Y_n = \left((f_r \circ F^-)\left(\frac{i}{n}\right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N},$$

where $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ is the triangular array of equidistant design points.
Similarly, in the uniform random case, the regression error is conditionally centred and standardised with respect to the design variable U, i.e. $\mathbb{E}(\varepsilon|U = u) = 0$, and $\mathbb{V}ar(\varepsilon|U = u) = 1$ for all $u \in [0,1]$, whereas, in the uniform fixed design case, the regression errors $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ are iid random variables with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

Explanation :
In the corresponding linear regression model with uniform random design, one considers an n-sample of random variables $U_1, \dots, U_n \overset{iid}{\sim} U$ where $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$ and the model in detail then reads

$$\begin{pmatrix} Y_{n1} \\ \vdots \\ Y_{nn} \end{pmatrix} = \begin{pmatrix} (f_1 \circ F^-)(U_1) & \dots & (f_p \circ F^-)(U_1) \\ \vdots & \ddots & \vdots \\ (f_1 \circ F^-)(U_n) & \dots & (f_p \circ F^-)(U_n) \end{pmatrix} \begin{pmatrix} \theta_1 \\ \vdots \\ \theta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{n1} \\ \vdots \\ \varepsilon_{np} \end{pmatrix}, \quad n \in \mathbb{N}.$$

In contrast, in the corresponding linear model with uniform fixed design one considers an n-sample of equidistant design points $\frac{1}{n}, \frac{2}{n}, \dots, 1$ and the model in detail reads

$$\begin{pmatrix} Y_{n1} \\ \vdots \\ Y_{nn} \end{pmatrix} = \begin{pmatrix} (f_1 \circ F^-)(\frac{1}{n}) & \dots & (f_p \circ F^-)(\frac{1}{n}) \\ \vdots & \ddots & \vdots \\ (f_1 \circ F^-)(1) & \dots & (f_p \circ F^-)(1) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{n1} \\ \vdots \\ \varepsilon_{np} \end{pmatrix}, \quad n \in \mathbb{N}.$$

The difference between both scenarios is that a uniform random design consists of an iid sample of random variables whose distribution is the uniform distribution on $\sigma(F^-)$, whereas in a uniform fixed design, the design consists of a concrete array of data points that are row wise equidistant in $[0,1]$. Therefore, we can see that the two designs are structurally identical with the only difference being the sampling points where the regression functions are computed.
To be more specific, under a uniform random design the model is applied to an n-sample of design variables $U_1, \dots, U_n \overset{iid}{\sim} U$ with $U \sim \lambda\!\!\!\lambda_{[0,1]}|_{\sigma(F^-)}$ — or a sequence of design variables $U_1, U_2, \dots \overset{iid}{\sim} U$ respectively. In contrast, under a uniform fixed design the model is applied to the "typical" realisation $\frac{1}{n}, \frac{2}{n}, \dots, 1$ of design variables $U_1, \dots, U_n \overset{iid}{\sim} U[0,1]$ — or to the triangular array of "typical" realisations $\left(\left(\frac{i}{n}\right)_{i=1}^n\right)_{n\in\mathbb{N}}$ of the sequence of design variables $U_1, U_2, \dots \overset{iid}{\sim} U[0,1]$ respectively. Note that if F has no jumps $U \sim U[0,1]$. If F has a jump, $\sigma(F^-) \neq \mathcal{B}_{[0,1]}$ and thus in the uniform random design case, $U_1, \dots, U_n \overset{iid}{\sim} U[0,1]$ does not hold true. In contrast, in the uniform fixed design case the triangular array of "typical" realisations still is $\left(\left(\frac{i}{n}\right)_{i=1}^n\right)_{n\in\mathbb{N}}$. This is merely a difference in presentation and not in content, since, if F indeed has such a jump, the regression functions $f_1 \circ F^-, \dots, f_p \circ F^-$ are constant in the corresponding interval that is not in $F(\mathbb{R})$. Thus, the design matrices

$$\begin{pmatrix} (f_1 \circ F^-)(U_1) & \dots & (f_p \circ F^-)(U_1) \\ \vdots & \ddots & \vdots \\ (f_1 \circ F^-)(U_n) & \dots & (f_p \circ F^-)(U_n) \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} (f_1 \circ F^-)(\frac{1}{n}) & \dots & (f_p \circ F^-)(\frac{1}{n}) \\ \vdots & \ddots & \vdots \\ (f_1 \circ F^-)(1) & \dots & (f_p \circ F^-)(1) \end{pmatrix}$$

do not differ.
In the above paragraph, we used the word "typical" in quotation marks since the probability for $U_1, \dots, U_n \overset{iid}{\sim} U$ or even $U_{1:n}, \dots, U_{n:n}$, where $U_{i:n}$ denotes the i-th order statistic of the sample, to realise as $\frac{1}{n}, \frac{2}{n}, \dots, 1$ is zero for a non-discrete design

distribution. However, the sequence of distribution functions

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,t]}\left(\frac{i}{n}\right)\right)_{n\in\mathbb{N}}, \quad t\in[0,1],$$

related to the triangular array $((t_{ni})_{i=1}^{n})_{n\in\mathbb{N}} = \left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ of this design points, converges to the distribution function of $U[0,1]$ which, on $\sigma(F^{-})$, coincides with the distribution function F_U of U, from which the iid sample $U_1,\ldots,U_n$ was taken. Thus, taking the design points $\frac{1}{n},\frac{2}{n},\ldots,1$ or the triangular array $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ respectively can simply be understood as a generic way to simulate the realisation of the n-sample $U_1,\ldots,U_n \overset{iid}{\sim} U$ or the sequence of design variables $U_1,U_2,\ldots \overset{iid}{\sim} U$ respectively.
From another point of view, according to the Glivenko-Cantelli theorem (see page 158 in Section 7.2 of the appendix for the formulation of the theorem and its proof), sampling iid variables $U_1,U_2,\ldots$ according to a random variable U with $P^U|_{\sigma(F^-)} = \lambda_{[0,1]}|_{\sigma(F^-)}$ is one way to get a sequence of empirical distribution functions

$$F_n(t) = \frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,t]}(U_i), \quad n\in\mathbb{N},\ t\in[0,1],$$

that converges uniformly a.s. to the given limit distribution function F_U.
Note that choosing the triangular array

$$((t_{ni})_{i=1}^{n})_{n\in\mathbb{N}} = \left(\left(\frac{i}{n}\right)^{n}_{i=1}\right)_{n\in\mathbb{N}}$$

as design points is just one way of getting a sequence of distribution functions

$$\left(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,t]}\left(\frac{i}{n}\right)\right)_{n\in\mathbb{N}}, \quad t\in[0,1],$$

that converges uniformly to the distribution function of $U[0,1]$. Every other choice of the triangular array $((t_{ni})_{i=1}^{n})_{n\in\mathbb{N}}$ as design points which fulfills

$$\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{(-\infty,t]}(t_{ni}) \xrightarrow[n\to\infty]{} id_{[0,1]}(t), \quad t\in[0,1], \tag{3.4}$$

uniformly is fine, too, as the residual partial sum limit process $B_{(f\circ F^-),F}$ (see Theorem 2.2.2 or Theorem 2.2.4) does depend on the limit distribution function F, but does not depend on the choice of the design distributions F_n converging uniformly to F. Therefore, without loss of generality, the triangular array of design points can be chosen to be $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$, as in our case $F_U(t) = id_{[0,1]}(t)$ holds true for all $t\in F(\mathbb{R})\cup\{0,1\}$.
Thus, the uniform random design and the uniform fixed design start at different places (that are closely related with each other). However, their asymptotic properties regarding the distribution of the limit process of the residual CUSUM precesses do not depend on this disparity (as Corollary 3.1.2 shows), but only depends on the

(limit) distribution function of the design. This, in both cases, is the identity on the unit interval.

Similarly, in the uniform random design case, the error variable ε is conditionally standardised, i.e, $\mathbb{E}(\varepsilon|U = u) = 0$ and $\mathbb{V}ar(\varepsilon|U = u) = 1$ for all $u \in [0,1]$ with respect to the design variable $U \sim U[0,1]$, of which the design random variables $U_1, \ldots, U_n$ are iid samples. Thus, if we pick the n-th row of the triangular array of design points $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ as the realisation of the n-th sample of these random variables — as it is done in a uniform fixed design — we end up with iid regression errors $\varepsilon_{n1}, \ldots, \varepsilon_{nn}$ with $\mathbb{E}(\varepsilon_{n1}) = 0$ and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

Disparity ii) In both, a corresponding linear regression model with uniform random design and in a corresponding linear regression model with uniform fixed design, the regression functions $(f_1 \circ F^-), \ldots, (f_p \circ F^-)$ are supposed to be linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$. However, in the uniform fixed design case, the regression functions are additionally assumed to be of bounded variation.

Explanation :

Note that a function of bounded variation is bounded and therefore square Lebesgue integrable on the unit interval. Therefore, assuming that the regression functions are of bounded variation is one way of ensuring their square integrability. However, not all functions in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$ are necessarily of bounded variation, for example Brownian motion paths. Thus, the conditions on the regression functions in a linear regression model with uniform fixed design are stricter than the conditions on the regression functions in a linear model with uniform random design.

However, if we take a closer look at the results on linear regression models with fixed designs (see Section 2.2) and their proofs, it becomes obvious, that the assumption of bounded variation of the regression functions in a linear regression model with fixed design is a proof artefact. To be more precise, it is a technical requirement necessary to guarantee the existence of certain Riemann-Stieltjes integrals. In practice, regression functions can only be evaluated at finitely many design points and therefore, one only chooses regression functions whose variability between two such design points is limited. This means that one can restrict oneself to regression functions that are of bounded variation.

4 Projection techniques

The main point of this chapter is a geometric interpretation of residual CUSUM limit processes as projections onto certain reproducing kernel Hilbert spaces (RKHS). We start with some general definitions and some theory on RKHS of stochastic processes with covariance function R and then focus on Brownian motion RKHS.
We emphasise that, with the exception of Theorem 4.2.3, the results of this chapter are well known in the literature (see for example Aronszajn (1950), Parzen (1959) and Bischoff (2002). So, it is the aim of this chapter to succinctly present and prove these results in order to make them available for later use in this work.

4.1 Reproducing kernel Hilbert space of the Brownian motion

In this section we are going to define, in analogy to Parzen (1959) and Adler (1990), what a RKHS of a stochastic process with covariance function R is and what the RKHS of a Brownian motion looks like. First, however, the following definitions shall be given:

Definition 4.1.1.

Let $T \subseteq \mathbb{R}$ and $X_t := (X(t))_{t\in T}$ be a stochastic process with $\mathbb{E}(X_s X_t) < \infty$ for all $s, t \in T$.

i) With
$$R : T \times T \longrightarrow \mathbb{R},\ (s,t) \longmapsto Cov(X_s, X_t),$$
we denote the covariance kernel or covariance function of the stochastic process X_t.

ii) We denote with
$$S(R) := \left\{ f : T \to \mathbb{R} \;\middle|\; f(\cdot) = \sum_{i=1}^{n} a_i R(s_i, \cdot),\ n \in \mathbb{N},\ a_i \in \mathbb{R},\ s_i \in T,\ i \in \{1, \dots, n\} \right\}$$
the space of finite linear combinations of the covariance kernel.

Note that since the covariance kernel is the covariance function of a stochastic process, it is symmetric and positive semidefinite and that $S(R)$ is a subspace of $\{f : T \to \mathbb{R}\}$. We now proceed with the definition of a RKHS with respect to the reproducing kernel of a covariance function of a stochastic process.

Definition 4.1.2.

Let $T \subseteq \mathbb{R}$ and $X_t := (X(t))_{t\in T}$ be a stochastic process with covariance kernel R. By $H(R)$ we denote the RKHS (of the stochastic process X_t) with respect to the reproducing kernel R. $H(R)$ is defined by the following three properties:

i) $H(R) \subseteq \{f : T \to \mathbb{R}\}$ endowed with some inner product $\langle\cdot,\cdot\rangle_{H(R)}$ is a Hilbert space with the induced norm $\|\cdot\|_{H(R)}$,

ii) $R(\cdot, t) \in H(R)$ for every $t \in T$,

iii) $\langle g(\cdot), R(\cdot, t)\rangle_{H(R)} = g(t)$ for all $t \in T$ and all $g \in H(R)$.

Note that property iii) of Definition 4.1.2 is often referred to as the "reproducing kernel" property.

We now proceed in five steps, to construct the RKHS $H(R)$ with respect to the reproducing kernel R of a stochastic process X_t as the completion of the space $S(R)$ under a natural inner product.

Step 1

We define a bilinear form $\langle\cdot,\cdot\rangle_{S(R)}$ on $S(R)$ via

$$\langle\cdot,\cdot\rangle_{S(R)} : S(R) \times S(R) \longrightarrow \mathbb{R},\ (f, g) \longmapsto \langle f, g\rangle_{S(R)} = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\, b_j\, R(s_i, t_j).$$

Here, since f and g are functions in $S(R)$, there are $n, m \in \mathbb{N}$, $a_1, \ldots, a_n, b_1, \ldots, b_m \in \mathbb{R}$ and $s_1, \ldots, s_n, t_1, \ldots, t_m \in T$ such that f and g admit the representations

$$f(\cdot) = \sum_{i=1}^{n} a_i\, R(s_i, \cdot) \quad \text{and} \quad g(\cdot) = \sum_{j=1}^{m} b_j\, R(t_j, \cdot).$$

Note that this mapping is well-defined, as the value of the mapping $\langle f, g\rangle_{S(R)}$ does not depend on the representations of the functions $f, g \in S(R)$. To demonstrate this, let different representations for f and g be given, for example

$$\sum_{i=1}^{n} a_i\, R(s_i, \cdot) = f(\cdot) = \sum_{i=1}^{\widetilde{n}} \widetilde{a}_i\, R(\widetilde{s}_i, \cdot)$$

and

$$\sum_{j=1}^{m} b_j\, R(t_j, \cdot) = g(\cdot) = \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j\, R(\widetilde{t}_j, \cdot).$$

Then, the following equation holds true:

$$\begin{aligned}
\langle f, g\rangle_{S(R)} &= \left\langle \sum_{i=1}^{n} a_i\, R(s_i, \cdot), \sum_{j=1}^{m} b_j\, R(t_j, \cdot) \right\rangle_{S(R)} \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\, b_j\, R(s_i, t_j) = \sum_{i=1}^{n} a_i \sum_{j=1}^{m} b_j\, R(s_i, t_j) \\
&= \sum_{i=1}^{n} a_i\, g(s_i) = \sum_{i=1}^{n} a_i \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j\, R(\widetilde{t}_j, s_i) \\
&= \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j \sum_{i=1}^{n} a_i\, R(s_i, \widetilde{t}_j) = \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j\, f(\widetilde{t}_j) \\
&= \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j \sum_{i=1}^{\widetilde{n}} \widetilde{a}_i R(\widetilde{s}_i, \widetilde{t}_j) = \sum_{j=1}^{\widetilde{m}}\sum_{i=1}^{\widetilde{n}} \widetilde{a}_i \widetilde{b}_j\, R(\widetilde{s}_i, \widetilde{t}_j) \\
&= \left\langle \sum_{i=1}^{\widetilde{n}} \widetilde{a}_i R(\widetilde{s}_i, \cdot), \sum_{j=1}^{\widetilde{m}} \widetilde{b}_j R(\widetilde{t}_j, \cdot) \right\rangle_{S(R)}.
\end{aligned}$$

Step 2

The bilinear form $\langle \cdot, \cdot \rangle_{S(R)}$ satisfies the reproducing kernel property, since for all $f \in S(R)$ and $t \in T$, we have

$$\begin{aligned}
\langle f(\cdot), R(t, \cdot)\rangle_{S(R)} &= \left\langle \sum_{i=1}^{n} a_i R(s_i, \cdot), R(t, \cdot) \right\rangle_{S(R)} \\
&= \sum_{i=1}^{n} a_i R(s_i, t) \\
&= f(t).
\end{aligned}$$

Step 3

The bilinear form $\langle \cdot, \cdot \rangle_{S(R)}$ defines an inner product on $S(R)$, since

- firstly $\langle \cdot, \cdot \rangle_{S(R)}$ is positive semidefinite because $R(\cdot, \cdot)$ is a covariance kernel and therefore

  $$\langle f, f\rangle_{S(R)} = \sum_{i=1}^{n}\sum_{j=1}^{n} a_i\, a_j\, R(s_i, s_j) = (a_1, \ldots, a_n)\, (R(s_i, s_j))_{i,j=1}^{n}\, (a_1, \ldots, a_n)^T \geq 0,$$

  holds true. Note that, since $\langle \cdot, \cdot \rangle_{S(R)}$ is a non-negative bilinear form, the Cauchy-Schwarz inequality holds, stating that

  $$\langle f, g\rangle^2 \leq \langle f, f\rangle \cdot \langle g, g\rangle \quad \text{for all } f, g \in S(R).$$

- secondly we have
$$\langle f, f\rangle_{S(R)} = 0 \iff f \equiv 0,$$
since the necessary condition is trivially satisfied and the inequality
$$f^2(t) = \langle f(\cdot), R(t,\cdot)\rangle^2_{S(R)} \leq \langle f, f\rangle_{S(R)} \langle R(t,\cdot), R(t,\cdot)\rangle_{S(R)} \quad \forall\, f \in S(R)\ \forall\, t \in T,$$
which holds true due to the reproducing kernel property and the applicability of the Cauchy-Schwarz inequality, implies sufficiency.

Step 4

By means of the inner product $\langle\cdot,\cdot\rangle_{S(R)}$, we are now able to define a norm on $S(R)$ via
$$||f||_{S(R)} := \sqrt{\langle f, f\rangle_{S(R)}}, \quad f \in S(R).$$
As a consequence, for a given Cauchy sequence $(f_n)_{n\in\mathbb{N}} \subseteq S(R)$ we can conclude that $(f_n)_{n\in\mathbb{N}}$ is pointwise Cauchy since, for all $t \in T$,
$$\begin{aligned} |f_n(t) - f_m(t)|^2 &= \langle f_n(\cdot) - f_m(\cdot), R(t,\cdot)\rangle^2_{S(R)} \\ &\leq \langle f_n - f_m, f_n - f_m\rangle_{S(R)} \langle R(t,\cdot), R(t,\cdot)\rangle_{S(R)} \\ &= ||f_n - f_m||^2_{S(R)} \langle R(\cdot,t), R(t,\cdot)\rangle_{S(R)} \\ &= ||f_n - f_m||^2_{S(R)} R(t,t) \end{aligned}$$
holds true and therefore, $(f_n)_{n\in\mathbb{N}}$ is pointwise convergent. In the equation above, the first and the last equality are true because of the reproducing kernel property and the inequality is the Cauchy-Schwarz inequality.

Step 5

Define $H(R)$, that is, the RKHS with respect to the kernel R, as the completion of $S(R)$. To be more specific, define

$$H(R) := \overline{S(R)}$$
$$= \left\{ f : T \to \mathbb{R} \;\middle|\; \exists\, (f_n(\cdot))_{n\in\mathbb{N}} \subseteq S(R) \text{ Cauchy sequence with } f(\cdot) = \lim_{n\to\infty} f_n(\cdot) \text{ pointwise} \right\}.$$

This means that $H(R)$ contains all functions in $S(R)$ as well as all functions which are pointwise limits of Cauchy sequences in $S(R)$.
Note that $S(R)$ does not have to be complete with respect to the norm $||\cdot||_{S(R)}$. In order to see this, we consider the process
$$(X_t)_{t\in\mathbb{R}} = \left(\sum_{i\in\mathbb{Z}} \varepsilon_i \cdot \mathbb{1}_{[i,i+1)}(t) \right)_{t\in\mathbb{R}},$$

where $\varepsilon_1, \varepsilon_2, \dots$ are iid random variables with $\mathbb{P}(\varepsilon_1 = -1) = \frac{1}{2} = \mathbb{P}(\varepsilon_1 = 1)$. Then

$$R(s,t) = Cov(X_s, X_t) = \mathbb{1}_{[\lfloor s \rfloor, \lfloor s \rfloor + 1)}(t), \quad s,t \in \mathbb{R}.$$

and thus $R(s,\cdot) = \mathbb{1}_{[\lfloor s \rfloor, \lfloor s \rfloor + 1)}(\cdot)$ holds true. Therefore,

$$S(R) = \left\{ \sum_{i=1}^{n} a_i R(s_i, \cdot) \,\middle|\, n \in \mathbb{N},\, a_i, s_i \in \mathbb{R},\, i \in \{1, \dots, n\} \right\} = span\left\{ \mathbb{1}_{[i,i+1)} \,|\, i \in \mathbb{Z} \right\}$$

with $\langle \mathbb{1}_{[n,n+1)}, \mathbb{1}_{[m,m+1)} \rangle_{S(R)} = \mathbb{1}_{\{m\}}(n)$, $n, m \in \mathbb{Z}$. Now consider

$$f_n(x) := \sum_{i=1}^{n} \frac{1}{i} \mathbb{1}_{[i,i+1)}(x),\ n \in \mathbb{N}, \text{ and } \quad f_\infty(x) := \sum_{i=1}^{\infty} \frac{1}{i} \mathbb{1}_{[i,i+1)}(x), \quad x \in \mathbb{R}.$$

Obviously, $f_n \in S(R)$, $n \in \mathbb{N}$ and $f_\infty \notin S(R)$, but $f_\infty \in \overline{S(R)}$ holds true, since

$$\|f_\infty - f_n\|_{S(R)} = \left\| \sum_{i=n+1}^{\infty} \frac{1}{i} \mathbb{1}_{[i,i+1)} \right\|_{S(R)} = \sqrt{\sum_{i=n+1}^{\infty} \frac{1}{i^2}} \xrightarrow[n \to \infty]{} 0.$$

Hence, $S(R)$ is not complete in this example.

In many text books containing RKHS techniques, the inner product space $S(R)$ is constructed, then it is shown that $S(R)$ fullfills properties ii) and iii) of a RKHS space, and lastly it is claimed that the completion $H(R)$ of $S(R)$ is the RKHS with respect to the kernel R. This is usually done in a similar manner to what we have done in our construction in Step 1 through Step 5. To conclude Step 5, we now prove that the completion of $S(R)$ is an inner product space. We will do this thoroughly in the proof of the theorem below which is based on Aronszajn (1950):

Theorem 4.1.3. [Aronszajn (1950)]

Let $T \subseteq \mathbb{R}$, $X_t := (X(t))_{t \in T}$ be a stochastic process with covariance kernel R and let $H(R) := \overline{S(R)}$ be as constructed above. Then $H(R)$ is the RKHS with respect to the reproducing kernel R.

Proof: We present a full proof based on Aronszajn (1950) starting from page 160 in Section 7.3 of the appendix. □

In the remainder of this section, we consider the RKHS with respect to the reproducing covariance kernel R of the standard Brownian motion — or the RKHS of the standard Brownian motion for short — as an example.

Example 4.1.4.

Let $T = [0,1]$ and $B = B_t = (B_t)_{t\in[0,1]}$ be the standard Brownian motion on the unit interval. Then

i) $R(s,t) = Cov(B_s, B_t) = \min(s,t), \quad s,t \in [0,1].$

ii) $R(s,\cdot) : [0,1] \to \mathbb{R},\ t \mapsto \min(s,t) = t \cdot \mathbb{1}_{[0,s[}(t) + s \cdot \mathbb{1}_{[s,1]}(t), \quad s \in [0,1].$

iii) $S(R) :=$

$$= \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f(t) = \sum_{i=1}^{n} a_i \min(s_i,t),\ n \in \mathbb{N},\ a_i \in \mathbb{R},\ s_i \in [0,1], i \in \{1,\dots,n\} \right\}$$

$$= \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f(t) = \sum_{i=1}^{n} a_i\, t \cdot \mathbb{1}_{[0,s_i[}(t) + a_i\, s_i \cdot \mathbb{1}_{[s_i,1]}(t),\ n \in \mathbb{N},\ a_i \in \mathbb{R},\ s_i \in [0,1] \right\}.$$

iv) for $f, g \in S(R)$ with

$$f(\cdot) = \sum_{i=1}^{n} a_i\ \min(s_i,\cdot) \quad \text{and} \quad g(\cdot) = \sum_{j=1}^{m} b_j\ \min(t_j,\cdot)$$

we have

$$\begin{aligned}
\langle f, g\rangle_{S(R)} &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\, b_j\ \min(s_i,t_j) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{m} a_i\, b_j \int_{[0,1]} \mathbb{1}_{[0,\min(s_i,t_j)]}(t)\, d\lambda\!\!\lambda_{[0,1]}(t) \\
&= \int_{[0,1]} \sum_{i=1}^{n} a_i \cdot \mathbb{1}_{[0,s_i]}(t) \sum_{j=1}^{m} b_j \cdot \mathbb{1}_{[0,t_j]}(t)\, d\lambda\!\!\lambda_{[0,1]}(t) \\
&= \int_{[0,1]} f'(t) g'(t)\, d\lambda\!\!\lambda_{[0,1]}(t) \\
&= \langle f', g'\rangle_{\mathcal{L}^2([0,1],\lambda\!\!\lambda_{[0,1]})},
\end{aligned}$$

since for f in $S(R)$

$$f'(t) = \frac{d}{dt}\sum_{i=1}^{n} a_i\ \min(s_i,t) = \sum_{i=1}^{n} a_i\, \frac{d}{dt}\min(s_i,t) = \sum_{i=1}^{n} a_i \cdot \mathbb{1}_{[0,s_i]}(t) \quad \lambda\!\!\lambda_{[0,1]}\text{-a.s.}$$

We now introduce the RKHS of the standard Brownian motion. This theorem and its proof are well-known in the literature and stated here for the sake of completeness.

Theorem 4.1.5.

Let $B = (B_t)_{t\in[0,1]}$ be the standard Brownian motion and define

$$\mathcal{H}_B := \left\{ f : [0,1] \to \mathbb{R} \,\middle|\, f(t) = \int_{[0,t]} f'(s)\, d\lambda(s),\ f' \in \mathcal{L}^2([0,1], \lambda) \right\}$$

as well as the bilinear mapping

$$\langle \cdot, \cdot \rangle_{\mathcal{H}_B} : \mathcal{H}_B \times \mathcal{H}_B \longrightarrow \mathbb{R},\ (f,g) \longmapsto \int_{[0,1]} f'\, g'\, d\lambda\,.$$

Then,

$$\mathcal{H}_B = H(R)$$

holds true, where $H(R)$ is the RKHS with respect to the reproducing kernel R of the standard Brownian motion.

Note that all functions $f \in \mathcal{H}_B$ are absolutely continuous functions.

Proof: We have to prove that $\mathcal{H}_B$ satisfies all three conditions from Definition 4.1.2. Firstly, due to Theorem 4.1.3, we know that $H(R) = \overline{S(R)}$ is a Hilbert space of functions $f : [0,1] \to \mathbb{R}$. Here, $S(R)$ is the space we considered in Example 4.1.4. Moreover, since $\{a \cdot \mathbb{1}_{[0,s]} \,|\, a \in \mathbb{R},\ s \in [0,1]\}$ is dense in $\{f' \,|\, f \in \mathcal{H}_B\}$ it follows that $S(R)$ is dense in $\mathcal{H}_B$ and therefore, $\mathcal{H}_B$ is a Hilbert space of functions $f : [0,1] \to \mathbb{R}$.
Secondly, since for all $t \in [0,1]$

$$R(\cdot, t) = \min(\cdot, t) = \int_{[0,\cdot]} \mathbb{1}_{[0,t]}(x)\, d\lambda(x) = \int_{[0,\cdot]} \frac{\partial}{\partial x} \min(t,x)\, d\lambda(x) \in \mathcal{H}_B$$

holds true, property ii) is satisfied since, furthermore,

$$\int_{[0,1]} \left(\frac{\partial}{\partial x} \min(t,x) \right)^2 d\lambda(x) = \int_{[0,1]} \mathbb{1}_{[0,t]}(x)\, d\lambda(x) = t < \infty.$$

Thirdly, because for all $t \in [0,1]$ and $f \in \mathcal{H}_B$

$$\langle f(\cdot), R(\cdot,t) \rangle_{\mathcal{H}_B} = \int_{[0,1]} f'(x) \cdot \frac{\partial}{\partial x} \min(t,x)\, d\lambda(x) = \int_{[0,t]} f'(x)\, d\lambda(x) = f(t)$$

is valid, the reproducing kernel property is satisfied, which completes the proof. □

4.2 Structure of the residual partial sum limit process

The main result of this chapter is Theorem 4.2.3, which sheds some light on the geometric structure of the residual partial sum limit process of a generic linear regression model. Before stating the theorem, we need the following definition.

Definition 4.2.1.

Consider a generic linear regression according to Definition 3.2.1. We then define

i) $f_{i,n} = \left(f_i\left(\frac{1}{n}\right), f_i\left(\frac{2}{n}\right), \ldots, f_i\left(1\right)\right)^T, \quad i = 1, \ldots, p,\ n \in \mathbb{N},$

ii) $W_n := Im(X_n) = span\left\{f_{1,n}, \ldots, f_{p,n}\right\}, \quad n \in \mathbb{N},$

iii) $h_{f_i}(u) := \int\limits_{[0,u]} f_i(t)\, d\lambda\!\!\lambda_{[0,1]}(t), \quad u \in [0,1],\ i = 1, \ldots, p,$

iv) $W_{\mathcal{H}_B} := span\{h_{f_1}(\cdot), \ldots, h_{f_p}(\cdot)\}.$

Note that for $n \in \mathbb{N}$, W_n is a subspace of $\mathbb{R}^n$ and $W_{\mathcal{H}_B}$ is a subspace of $\mathcal{H}_B$. Furthermore, as usual, we denote by $W_n^{\perp}$ the orthogonal complement of W_n with respect to the Euclidean inner product, and by $W_{\mathcal{H}_B}^{\perp}$, we denote the orthogonal complement of $W_{\mathcal{H}_B}$ with respect to the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_B}$.

The following lemma is easy to verify and gives a representation of projections onto finite dimensional subspaces of $\mathcal{H}_B$.

Lemma 4.2.2.

Let $g_1, \ldots, g_p$, $p \in \mathbb{N}$, be linearly independent functions in $\mathcal{H}_B$. For $i \in \{1, \ldots, p\}$, let g_i' be a function in $\mathcal{L}^2([0,1], \lambda\!\!\lambda_{[0,1]})$ that satisfies

$$g_i(t) = \int\limits_{[0,t]} g_i'(s)\, d\lambda\!\!\lambda_{[0,1]}(s), \quad t \in [0,1].$$

Then the mapping

$$pr_{span\{g_1,\ldots,g_p\}} : \mathcal{H}_B \longrightarrow \mathcal{H}_B,$$

defined by

$$pr_{span\{g_1,\ldots,g_p\}}(h)(\cdot) = (g_1, \ldots, g_p)(\cdot) \left(\int\limits_{[0,1]} \begin{pmatrix} g_1' \\ \vdots \\ g_p' \end{pmatrix} (g_1', \ldots, g_p')\, d\lambda\!\!\lambda_{[0,1]} \right)^{-1} \overset{(R)}{\int\limits_{[0,1]}} \begin{pmatrix} g_1' \\ \vdots \\ g_p' \end{pmatrix} dh,$$

is the orthogonal projection onto $span\{g_1, \ldots, g_p\}$ in $\mathcal{H}_B$.

As a consequence of the lemma above, the orthogonal projection $pr_{W_{\mathcal{H}_B}}(h)$ of a function $h \in \mathcal{H}_B$ onto $W_{\mathcal{H}_B}$ can be represented as

$$pr_{W_{\mathcal{H}_B}}(h)(\cdot) = (h_{f_1}, \ldots, h_{f_p})(\cdot) \left(\int_{[0,1]} \begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix} (f_1, \ldots, f_p)\, d\lambda\!\!\!\lambda_{[0,1]} \right)^{-1} \int_{[0,1]}^{(R)} \begin{pmatrix} f_1 \\ \vdots \\ f_p \end{pmatrix} dh$$

$$= \left(\int_{[0,\cdot]} f(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^T \left(\int_{[0,1]} f(t)\, f^T(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^{-1} \int_{[0,1]}^{(R)} f(t)\, dh(t).$$

Based on this, we now extend $pr_{W_{\mathcal{H}_B}} \colon \mathcal{H}_B \longrightarrow W_{\mathcal{H}_B}$ to $pr_{W_{\mathcal{H}_B}} \colon C[0,1] \longrightarrow W_{\mathcal{H}_B}$, according to the construction by Bischoff (2002), by extending the inner product $\langle \cdot, \cdot \rangle_{\mathcal{H}_B}$ of $\mathcal{H}_B$ to a bilinear mapping

$$\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B) \times C[0,1]} \colon BV_D(\mathcal{H}_B) \times C[0,1] \longrightarrow \mathbb{R},\ (h, u) \longmapsto \int_{[0,1]}^{(R)} h'\, du.$$

Here,

$$BV_D(\mathcal{H}_B) := \{h \in \mathcal{H}_B |\, h' \text{ is of bounded variation}\}$$

and due to partial integration (see for example Satz 6.3 in Walter (2002))

$$\int_{[0,1]}^{(R)} h' du = u(1)h'(1) - u(0)h'(0) - \int_{[0,1]}^{(R)} u\, dh'$$

holds true. Thus, one obtains

$$|\langle h, u \rangle_{BV_D(\mathcal{H}_B) \times C[0,1]}| \leq \|u\|_\infty (|h'(1)| + |h'(0)| + \|h'\|_V) < \infty,$$

where $\|h'\|_V$ denotes the total variation of h', which is finite as $h \in BV_D(\mathcal{H}_B)$.
Due to the technicality that $W_{\mathcal{H}_B} = span\{h_{f_1}(\cdot), \ldots, h_{f_p}(\cdot)\}$ and $h_{f_i} \in BV_D(\mathcal{H}_B), i = 1, \ldots, p$ (since the regression functions $f_1, \ldots, f_p$ are of bounded variation), it suffices to show that

$$\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B) \times C[0,1]} \colon BV_D(\mathcal{H}_B) \times C[0,1] \longrightarrow \mathbb{R}$$

is an extension of the original inner product

$$\langle \cdot, \cdot \rangle_{\mathcal{H}_B} \colon \mathcal{H}_B \times \mathcal{H}_B \longrightarrow \mathbb{R},$$

restricted to $BV_D(\mathcal{H}_B) \times \mathcal{H}_B$ in order to extend the domain of the projection $pr_{W_{\mathcal{H}_B}} : \mathcal{H}_B \longrightarrow W_{\mathcal{H}_B}$ to $C[0,1]$. We thus only have to show that

$$\langle \cdot, \cdot \rangle_{\mathcal{H}_B}|_{BV_D(\mathcal{H}_B) \times \mathcal{H}_B} = \langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B) \times C[0,1]}|_{BV_D(\mathcal{H}_B) \times \mathcal{H}_B}$$

holds true. The above equation indeed holds true, since for $h \in BV_D(\mathcal{H}_B)$ and $u \in \mathcal{H}_B$ one gets

$$\langle h, u\rangle_{\mathcal{H}_B} = \int\limits_{[0,1]} h'(t)u'(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) = \int\limits_{(0,1]} h'(t)u'(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t)$$

$$= \int\limits_{[0,1]}^{(R)} h'(t)u'(t)\, dt = \int\limits_{[0,1]}^{(R)} h'(t)\, du(t) = \langle h, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}.$$

Note that in the above chain of equations, we used Theorem 5.1.2 from Strook (1994) to prove the equality between the Riemann-Stieltjes integral and the measure integral and then, we applied Theorem 6.4 from Walter (2002).
Now, using the bilinear mapping $\langle\cdot,\cdot\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}$ instead of $\langle\cdot,\cdot\rangle_{\mathcal{H}_B}$ enables us to define the operation $pr_{W_{\mathcal{H}_B}}\colon C[0,1] \longrightarrow W_{\mathcal{H}_B}$ similarly to Lemma 4.2.2.

We can now identify the residual partial sum limit processes, in a generic linear regression model, with the orthogonal projection of a standard Brownian motion onto the orthogonal complement of the space spanned by the integrated regression functions within the RKHS associated to the standard Brownian motion. This limit process, for $n \to \infty$, is obtained as the weak limit of the partial sum operator evaluated at the orthogonal projection of the regression outcome onto the orthogonal complement of the space which is spanned by the columns of the design matrix.
The following theorem makes this statement mathematically precise:

Theorem 4.2.3.

Let a generic linear regression model according to Definition 3.2.1 be given. Then, for the residual partial sum process, it holds true, that

$$\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) = \frac{1}{\sqrt{n}} T_n(pr_{W_n^{\perp}}(Y_n))(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} pr_{W_{\mathcal{H}_B}^{\perp}}(B(\cdot)) \quad \text{in } C[0,1].$$

Proof: As usual, $pr_{W_n^{\perp}} = id - pr_{W_n}$ and $pr_{W_{\mathcal{H}_B}^{\perp}} = id - pr_{W_{\mathcal{H}_B}}$ Thus, $\frac{1}{\sqrt{n}} T_n(pr_{W_n^{\perp}}(Y_n))(\cdot)$ is in fact the residual partial sum process $\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot)$, which has been introduced in Section 2.2.1, since

$$\hat{\varepsilon}_n = Y_n - X_n\hat{\beta}_n = Y_n - X_n(X_n^T X_n)^{-1} X_n^T Y_n = Y_n - pr_{W_n}(Y_n) = pr_{W_n^{\perp}}(Y_n)$$

holds true. Therefore, Corollary 3.2.2 applies and yields the convergence

$$\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,\lambda\!\!\!^{-}_{[0,1]}}(\cdot) \quad \text{in } C[0,1].$$

Furthermore, the equation

$$
\begin{aligned}
& pr_{W_{\mathcal{H}_B}^{\perp}}(B(\cdot)) \\
& \quad = B(s) - \left(\int_{[0,s]} f(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) \right)^T \left(\int_{[0,1]} f(t)\, f^T(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) \right)^{-1} \int_{[0,1]}^{(R)} f(t)\, dB(t) \qquad (4.1) \\
& \quad = B(s) - \left(\int_{[0,s]} f(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) \right)^T \left(\int_{[0,1]} f(t)\, f^T(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) \right)^{-1} \int_{[0,1]} f(t)\, dB(t) \qquad (4.2) \\
& \quad = B_{f,\lambda\!\!\!^{-}_{[0,1]}}(\cdot)
\end{aligned}
$$

is satisfied. Here, equation (4.1) is valid, according to Lemma 4.2.2 and the extension of the projection $pr_{W_{\mathcal{H}_B}}$ from $\mathcal{H}_B$ to $C[0,1]$. To see that equation (4.2) is satisfied, we refer to equation (2.15) in Section 2.2.1 and the fact that

$$
\int_{[0,1]} f(t)\, dB(t) \;=\; f(0) \underbrace{B(0)}_{=0} + \int_{[0,1]}^{(R)} f(t)\, dB(t)
$$

holds true. This completes the proof. □

Note that although Theorem 4.2.3 looks similar to Theorem 3.2 in Bischoff (2002), they differ in their assumptions. To be more specific, in Theorem 3.2 in Bischoff (2002), in addition to a generic linear regression model (that is a linear regression model with uniform fixed design), it is assumed that the regression functions are continuous.

5 Heteroscedastic linear regression models

The effects on residual CUSUM processes and goodness-of-fit tests of heteroscedastic errors in linear regression models have already been studied in the literature.
The main results in the literature show that CUSUM processes of heteroscedastic regression errors converge in distribution to time-transformed Brownian motions and that residual CUSUM processes of heteroscedastic linear regression models converge to projections of time-transformed Brownian motions onto RKHS. See for example Remark 4.5 in Bischoff and Miller (2000) or MacNeill, Mao, and Xie (1994), where limit processes for independent heteroscedastic errors are given in one-dimensional and spatial regression time series models and limits of residual partial sum processes are computed in some special cases. More recently, in heteroscedastic spatial regression models, limit processes have been determined for partial sum processes of independent errors as well as for residual partial sum processes in Somayasa (2011).

This chapter builds on the previous findings of this thesis and addresses the work of MacNeill, Mao, and Xie (1994). We will state and prove functional central limit theorems for the partial sum processes of heteroscedastic and independent regression errors (see Theorem 5.0.2) as well as for residual partial sum processes of heteroscedastic regression models (see Theorem 5.0.3). Although functional central limit theorems in the given situation are already known and intuitive to some extent, the proof of Theorem 5.0.2 that we give in this chapter is novel and utilises theorems of Prokhorov and Rubin. Furthermore, we prove Theorem 5.0.2 for triangular arrays of random variables and under slightly weaker assumptions than those implicitly made in MacNeill, Mao, and Xie (1994) about sequences of random variables. See Remark 5.0.1 in this regard. On the one hand, the proof of Theorem 5.0.3 relies on Theorem 5.0.2. On the other hand, it generalises Theorem 3.2 and its proof in Bischoff (2002).

To be more specific, we consider a heteroscedastic generic linear regression model. This model has the form

$$Y_n = X_n\beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{5.1}$$

where

i) $[0,1]$ is the experimental region,

ii) $\beta = (\beta_1, \ldots, \beta_p)^T$ are the unknown regression coefficients,

iii) $((\frac{i}{n})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [0,1]$ is the triangular array of equidistant design points whose cor-

responding sequence of empirical distribution functions converges uniformly to the distribution function of the uniform distribution on the unit interval, i.e., the identity function on $[0,1]$,

iv) $X_n = \left(f_r\left(\frac{i}{n}\right)\right)_{i=1,r=1}^{n\quad p}$ is the design matrix where the regression functions $f_1, \ldots, f_p : [0,1] \to \mathbb{R}$ are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$,

v) $(\varepsilon(t))_{t\in[0,1]}$ is the random process of regression errors with $\mathbb{E}(\varepsilon(t)) = 0$ and $\mathbb{V}ar(\varepsilon(t)) = \sigma^2(t)$ for all $t \in [0,1]$. Here, $\sigma : [0,1] \to (0,\infty)$ is a function of bounded variation. Furthermore, for all $n \in \mathbb{N}$ and $i = 1, \ldots, n$, the regression errors $\varepsilon\left(\frac{1}{n}\right), \varepsilon\left(\frac{2}{n}\right), \ldots, \varepsilon(1)$ are assumed to be independent and we define $\varepsilon_n := \left(\varepsilon\left(\frac{1}{n}\right), \varepsilon\left(\frac{2}{n}\right), \ldots, \varepsilon(1)\right)^T$,

vi) we additionally suppose that for some $\delta > 0$ there is a $C > 0$ with $\mathbb{E}(|\varepsilon(t)|^{2+\delta}) \leq C$ for all $t \in [0,1]$. This additional assumption is further discussed in the following remark and is needed to ensure that the Lindeberg condition is satisfied.

Remark 5.0.1.

In Chapter 3 of MacNeill, Mao, and Xie (1994), the following assumptions are made: For $X_1, X_2, \ldots, X_n$ a sequence of independent observations with

$$\mathbb{E}(X_i) = \mu \in \mathbb{R} \quad \text{and} \quad \mathbb{V}ar(X_i) = \sigma_i^2 \in (0,\infty),$$

the partial sums are defined as

$$S_k = \sum_{i=1}^{k} (X_i - \mu), \quad k = 1, 2, \ldots, n,$$

and heteroscedasticity is characterized via

$$\sigma_i^2 = h\left(\frac{i}{n}\right),$$

where the positive function $h(\cdot)$ is of bounded variation on the interval $[0,1]$. Under these assumptions, Theorem 1 of MacNeill, Mao, and Xie (1994) claims that the partial sum process $\{\Theta_{hn}(t), t \in [0,1]\}$, whose paths for $k \in 0, 1, \ldots, n$ have the values $\frac{1}{\sqrt{n}} S_k$ at the point $\frac{k}{n}$ and are linear in between these points, converges weakly to a time-transformed centred Brownian motion

$$\left[B\left(\int_{[0,t]}^{(R)} h(x)\, dx \right) \right]_{t\in[0,1]}.$$

To prove this theorem, it is argued that

- firstly, the Lindeberg theorem and the Cramer-Wold device can be used to demonstrate asymptotic multivariate normality of the finite-dimensional distributions (fidis) of Θ_{hn}, (5.2)

- secondly, if one lets $K \in \mathbb{R}$ be a bound for $h(\cdot)$, then one can show that

$$\mathbb{E}\left(\frac{1}{\sqrt{n\sigma}}\{S(k) - S(j)\}\right)^4 \leq K^2\left|\frac{k}{n} - \frac{j}{n}\right|^2 \tag{5.3}$$

  holds true for $k, j \in \{0, 1, \ldots, n\}$, which implies that the partial-sum process is tight.

Ultimately, this proves weak convergence.

In contrast to MacNeill's assumptions, we additionally supposed for model (5.1) in Assumption vi) that for some $\delta > 0$ there is a $C > 0$ with $\mathbb{E}(|\varepsilon(t)|^{2+\delta}) \leq C$ for all $t \in [0, 1]$. Therefore, at first glance, it seems that the assumptions we made in model (5.1) are stronger than the ones made in MacNeill, Mao, and Xie (1994) or Somayasa (2011) (who, for the proof of Theorem 2.1, refers to the proof of MacNeill, Mao, and Xie (1994).
However, we will now argue that the implicit conditions of MacNeill, Mao, and Xie (1994) are in fact more stringent.

- Concerning (5.2), the Lindeberg theorem does not have to apply:

  The Lindeberg theorem states that the central limit theorem applies to a sequence of independent random variables $(X_n)_{n\in\mathbb{N}}$ (see for example Satz 4.1.8 in Gänsler and Stute (1977)) or a triangular array of random variables $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ (see for example Satz 15.43 in Klenke (2013)) that are centred and normalised, if the condition

$$\frac{1}{s_n^2}\sum_{i=1}^{n}\int \mathbb{1}_{[\varepsilon s_n,\infty)}(|X_i|)\cdot X_i^2\, dP \xrightarrow[n\to\infty]{} 0,$$

  known as Lindeberg's condition, is satisfied for every $\varepsilon > 0$.
  So let $h \equiv 1$, $n \in \mathbb{N}$, $n \geq 2$ and $U_1, \ldots, U_n, V_1, \ldots, V_n$ be stochastically independent with

$$U_1, V_1 \sim Ber\left(\frac{1}{2}\right) \quad \text{and} \quad U_i, V_i \sim Ber\left(\frac{1}{i}\right), \quad i = 2, \ldots, n.$$

For $i = 2, \ldots, n$, we then define

$$X_i := \lambda_i \, (U_i - V_i),$$

where

$$\lambda_1 := \sqrt{2} \quad \text{and} \quad \lambda_i := \left(2 \cdot \frac{1}{i} \cdot \left(1 - \frac{1}{i} \right) \right)^{-1/2}.$$

It now holds true that $\mathbb{E}(X_i) = 0, \; i \in \{1, \ldots, n\}$,

$$\mathbb{V}ar(X_1) = 2 \cdot 2 \cdot \left(\frac{1}{2} \right)^2 = 1 = h(1),$$

and due to independence

$$\begin{aligned} \mathbb{V}ar(X_i) &= \lambda_i^2 \cdot \mathbb{V}ar \, (U_i - V_i) = \lambda_i^2 \cdot 2 \cdot \mathbb{V}ar \, (U_i) \\ &= \left(2 \cdot \frac{1}{i} \cdot \left(1 - \frac{1}{i} \right) \right)^{-1} \cdot 2 \cdot \left(\frac{1}{i} \left(1 - \frac{1}{i} \right) \right) \\ &= 1 = h \left(\frac{i}{n} \right), \quad i = 2, \ldots, n. \end{aligned}$$

We thus have

$$s_n^2 := \sum_{i=1}^{n} \mathbb{V}ar(X_i) = n$$

and therefore, the random variables $X_1, X_2, \ldots, X_n$ satisfy the conditions MacNeill, Mao, and Xie (1994) made. However, if we define for $n \in \mathbb{N}$ and $\varepsilon > 0$

$$I_{n,\varepsilon} := \{ i \in \{1, \ldots, n\} \, : \, X_i^2 \cdot \mathbb{1}_{[\varepsilon s_n, \infty)}(|X_i|) = X_i^2 \}$$

and pick $\varepsilon \in \left(0, \frac{1}{2}\right]$, it holds true that, since $|X_i| = 0$ or $|X_i| = \lambda_i$, $\lambda_i \geq \varepsilon s_n$ is sufficient for $i \in I_{n,\varepsilon}$. Therefore, for $n \geq 2$ and $\varepsilon \in \left(0, \frac{1}{2}\right]$,

$$\begin{aligned} |I_{n,\varepsilon}| &\geq |\{ i \in \{1, \ldots, n\} \, : \, \lambda_i \geq \varepsilon s_n \}| \\ &\geq \left| \left\{ i \in \{1, \ldots, n\} \, : \, \left(2 \cdot \frac{1}{i} \cdot \left(1 - \frac{1}{i} \right) \right)^{-1/2} \geq \frac{\sqrt{n}}{2} \right\} \right| \\ &= \left| \left\{ i \in \{1, \ldots, n\} \, : \, i \geq \frac{n-1}{2} \right\} \right| \\ &\geq \frac{n}{2} \end{aligned}$$

is valid. We thus get

$$\frac{1}{s_n^2} \sum_{i=1}^{n} \int X_i^2 \cdot \mathbb{1}_{[\frac{1}{2} s_n, \infty)}(|X_i|) \, dP \; =$$

$$= \frac{1}{s_n^2} \sum_{i \in I_{n,\varepsilon}} \int X_i^2 \cdot \mathbb{1}_{[\frac{1}{2} s_n, \infty)}(|X_i|)\, dP + \frac{1}{s_n^2} \sum_{i \notin I_{n,\varepsilon}}^{n} \int X_i^2 \cdot \mathbb{1}_{[\frac{1}{2} s_n, \infty)}(|X_i|)\, dP$$

$$\geq \frac{1}{s_n^2} \sum_{i \in I_{n,\varepsilon}} \int X_i^2\, dP = \frac{1}{s_n^2} \sum_{i \in I_{n,\varepsilon}} \mathbb{V}ar(X_i)$$

$$\geq \frac{1}{n} \cdot \frac{n}{2} \cdot 1 = \frac{1}{2} \xrightarrow[n \to \infty]{\not\longrightarrow} 0,$$

which means that Lindeberg's condition is not satisfied.

Note that, since Feller's condition

$$\max_{i \in \{1,\ldots,n\}} \frac{\mathbb{V}ar(X_i)}{s_n^2} = \frac{1}{n} \xrightarrow[n \to \infty]{} 0$$

(see for example Satz 4.1.20 in Gänsler and Stute (1977)) is satisfied, Lindeberg's condition is sufficient and necessary for the applicability of the central limit theorem. Hence we know that in this situation, the central limit theorem also does not apply.

- Concerning (5.3), at least $\mathbb{E}(X_i^4) < \infty$ for all $i \in \mathbb{N}$ was assumed implicitly:

  For $n \geq 2$ and $j < k \in \{1, \ldots, n\}$,

$$\mathbb{E}\left(\frac{1}{\sqrt{n\sigma}} \{S(k) - S(j)\} \right)^4 = \mathbb{E}\left(\frac{1}{n^2\sigma^2} \left(\sum_{l=j+1}^{k} X_l \right)^4 \right)$$

$$= \frac{1}{n^2\sigma^2} \mathbb{E}\left(\sum_{l=j+1}^{k} X_l^4 + \sum_{l=j+1}^{k} \sum_{\substack{m=j+1 \\ m \neq l}}^{k} X_l^2 X_m^2 \right)$$

$$= \frac{1}{n^2\sigma^2} \sum_{l=j+1}^{k} \mathbb{E}\left(X_l^4\right) + \frac{1}{n^2\sigma^2} \sum_{l=j+1}^{k} \sum_{\substack{m=j+1 \\ m \neq l}}^{k} \mathbb{E}\left(X_l^2\right) \mathbb{E}\left(X_m^2\right)$$

  holds true. Thus, for (5.3) to be satisfied — as stated in MacNeill, Mao, and Xie (1994) — it at least has to be assumed that $\mathbb{E}(X_i^4) < \infty$ for all $i = 1, \ldots, n$. This condition is obviously stronger than our Assumption vi) in model (5.1). As we will see in (7.92) each of these additional assumptions will lead to the fulfilment of the Lindeberg condition.

The following theorem is an invariance principle for the partial sum processes of independent heteroscedastic regression errors. We present a new proof based on results of Prokhorov and Rubin.

Theorem 5.0.2.

Let a heteroscedastic generic linear regression model as described in (5.1) be given. Then,

$$\frac{1}{\sqrt{n}} T_n\left(\varepsilon_n\right)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B\left(\int_{[0,\cdot]} \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t)\right) \quad \text{in } C[0,1].$$

Proof: The proof of this theorem is deferred to Section 7.4 of the appendix starting at page 164. □

The following theorem is a functional central limit theorem for the residual partial sum processes in a heteroscedastic generic linear regression model. The theorem is new in the literature, as it generalises Theorem 3.2 of Bischoff (2002) for regression functions that no longer have to be continuous.

Theorem 5.0.3.

Let a heteroscedastic generic linear regression model as described in (5.1) be given. Let $W_{\mathcal{H}_B}$ be as defined in Definition 4.2.1. Then,

$$\begin{aligned}
&\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} pr_{W_{\mathcal{H}_B}^{\perp}}\left(B\left(\int_{[0,\cdot]} \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t)\right)\right) \\
&\quad = B\left(\int_{[0,\cdot]} \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t)\right) - \left(\int_{[0,\cdot]} f(t) d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^T \left(\int_{[0,1]} f(t) f^T(t) d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^{-1} \times \\
&\qquad \times \int_{[0,1]} f(t) dB\left(\int_{[0,t]} \sigma^2(u)\, d\lambda\!\!\!^{-}_{[0,1]}(u)\right) \quad \text{in } C[0,1].
\end{aligned}$$

Proof: The proof of this theorem is deferred to Section 7.4 of the appendix starting at page 173. □

6 Goodness-of-fit test in linear regression

So far, we have mainly occupied ourselves with the task of determining residual CUSUM limit processes in regression models with either fixed or random designs. However, little has been said about the application for this work, which is the desire to construct goodness-of-fit tests for linear regression models in order to decide whether a given family of functions is appropriate to model the relation between the covariables and an outcome. In this chapter, we are going to discuss a fundamental problem of such goodness-of-fit tests when the estimation of parameters is necessary. This is often referred to as the "Durbin problem" (see for example Koenker and Xiao (2002)). This problem did not initially arise in the context of goodness-of-fit tests in regression models, but in the context of testing whether or not a sample of random variables was taken from a distribution belonging to a particular class of distribution functions.
In the later context, tests based on the empirical distribution functions of random samples are considered and therefore, empirical processes are studied. As a consequence, we will start our exploration of the Durbin problem in Section 6.1.1 in the context of empirical processes. We will then — also in the context of empirical processes — briefly introduce a well-known way of dealing with such problems that goes back to Khmaladze (1980) and Khmaladze (1982) (see Section 6.1.2) before, in Section 6.2, we show how Stute, Thies, and Zhu (1998) have applied this solution to the context of goodness-of-fit tests in regression models. Finally, in Section 6.3, in the situation of linear regression models, we state and prove a theorem that specifies the interpretation of the Khmaladze transformation as a continuous-time backwards recursive least-squares method.

6.1 A brief history of goodness-of-fit tests

The task of conducting goodness-of-fit tests initially arose in the context of testing whether or not a sample of random variables was taken from a distribution belonging to a particular class of distribution functions. In this context, tests are usually based on the empirical distribution function of a random sample and empirical processes are therefore examined. Consequently, we will begin our exploration of goodness-of-fit tests in the situation described above by drawing attention to a classical problem in goodness-of-fit tests, the so-called "Durbin problem", before introducing a famous solution to this problem, the Khmaladze transformation.

6.1.1 Durbin problem

Consider the following problem. We are given a random sample

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} F, \quad n \in \mathbb{N},$$

where F is an unknown distribution function. The task is to construct a distribution free test statistic in order to perform a goodness-of-fit test for the hypothesis

$$H_0 : F \in \mathcal{F} \quad \text{v.s.} \quad H_1 : F \notin \mathcal{F}.$$

Here,

$$\mathcal{F} := \{F(\cdot, \theta) : \mathbb{R} \to [0,1] \,|\, \theta \in \Theta\}, \quad \Theta \subseteq \mathbb{R}^p \text{ open},$$

is a family of distribution functions, where $F(\cdot, \theta)$ is a distribution function that is indexed and uniquely identified by a parameter $\theta \in \Theta$.
Test statistics used to perform the above test require an estimation of the unknown parameter θ. See for example the test statistic $V_{n,\hat{\theta}_n}(x)$ in equation (6.4) on page 78. It is therefore natural that the distribution of the test statistic may depend on true parameter $\theta_0 \in \Theta$ (which exists under H_0). Furthermore, although only of minor consequence for certain practitioners, test statistics may also depend on model characteristics, namely the family of distribution functions $\mathcal{F}$.
Nevertheless, it is desirable that the distribution of the test statistic (or at least its asymptotic distribution) does not depend on θ_0 or, depending on the concrete application, does not depend on model characteristics. If this were not the case, critical values would not be calculable or would vary from one concrete test to another and therefore would have to be computed separately for each concrete test, which is rather impractical. Tests or test statistics are referred to as "distribution free" — or asymptotically "distribution free" — if estimated parameters or model characteristics do not influence the test statistics (asymptotic) distribution.

Throughout Section 6.1 we assume that the following regularity condition on $\mathcal{F}$ holds:

(A1) $F(\cdot, \theta)$ is absolutely continuous with respect to the Lebesgue-measure for all $\theta \in \Theta$ and $F(\cdot, \theta_1)$ and $F(\cdot, \theta_2)$ are mutually absolutely continuous for all $\theta_1, \theta_2 \in \Theta$.

Example 6.1.1.

Let us consider the most simple special case of the test scenario from above. That is, let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F$, $n \in \mathbb{N}$, and $\mathcal{F} = \{F_0\}$, where $F_0 : \mathbb{R} \to [0,1]$ is a known distribution function. We call such a hypothesis as a "simple" hypothesis, as in this case $\mathcal{F} = \{F_0\}$ for a known distribution function F_0. We want a vehicle to decide

whether

$$H_0 : F = F_0 \quad \text{or} \quad H_1 : F \neq F_0 .$$

In order to quantify a deviation from H_0, we consider the empirical process

$$V_n(x) := \sqrt{n}\,(F_n(x) - F_0(x))\,, \quad x \in \mathbb{R}, \tag{6.1}$$

where F_n is the empirical distribution function of the random sample $X_1, \ldots, X_n$ and use functionals of it to generate test statistics. According to Assumption (A1), F_0 is absolutely continuous. Thus, by carrying out the substitution

$$t = F_0(x), \quad x \in \mathbb{R},$$

we obtain on the one hand, under H_0, i.e., under $F = F_0$, the transformed sample

$$U_1, \ldots, U_n \overset{\text{iid}}{\sim} U[0,1]$$

belonging to the initial sample $X_1, \ldots, X_n \overset{\text{iid}}{\sim} F_0$ by the relation

$$U_i := F_0(X_i), \quad i = 1, \ldots, n,$$

(see Lemma 7.1.2 in this regard) and, on the other hand, the relation $F_0(F_0^-(t)) = t$ holds true for all $t \in (0,1)$, according to (7.40) of Lemma 7.1.1. We can thus define the uniform empirical process

$$\overline{V}_n(t) = \sqrt{n}\,\big(\overline{F}_n(t) - t\big)\,, \quad t \in [0,1], \tag{6.2}$$

where $\overline{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(U_i)$, $x \in \mathbb{R}$, is the empirical distribution function corresponding to the random sample $U_1, \ldots, U_n$. Note that under H_0, we have

$$\begin{aligned}
\overline{V}_n(t) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(U_i) - t\right) \;=\; \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(F_0(X_i)) - t\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,F_0^-(t)]}(X_i) - t\right) \\
&= \sqrt{n}\,\big(F_n(F_0^-(t)) - F_0(F_0^-(t))\big) \;=\; V_n(F_0^-(t)) \\
&= V_n(x),
\end{aligned} \tag{6.3}$$

where $x = F_0^-(t) \in \mathbb{R}$. Therefore, $\overline{V}_n$ is suitable to detect deviations from H_0. Note furthermore, that the distribution of $\overline{V}_n(\cdot)$ does no longer depend on the hypothesised distribution function F_0, that is, on the model, since F_0 is not a part of $\overline{V}_n$ any more. As a consequence, $\overline{V}_n$ is suitable to construct distribution free test

statistics like

$$T_{KS} := \sup_{t\in[0,1]} \left|\overline{V}_n(t)\right|,$$

used in the Kolmgorow-Smirnov test or

$$T_{CvM} := \int_{[0,1]}^{(R)} \left(\overline{V}_n(t)\right)^2 \, dt,$$

used in the Cramer-von Mieses test. To see how such a test is constructed, remember that the uniform empirical process $\overline{V}_n$ converges on $C[0,1]$ in distribution to a Brownian bridge B^0 (see for example Billingsley (1968) Theorem 13.1, p. 105). Thus, continuous functionals of $\overline{V}_n$ converge to the corresponding functionals of B^0 due to the continuous mapping theorem. Therefore,

$$T_{KS} \xrightarrow[n\to\infty]{C[0,1]} \sup_{t\in[0,1]} \left|B^0(t)\right|$$

as well as

$$T_{CvM} \xrightarrow[n\to\infty]{C[0,1]} \int_{[0,1]}^{(R)} \left(B^0(t)\right)^2 \, dt,$$

and since both limit distributions are well-known, critical values are already available.

In the general case, the class of hypothesised distribution functions was $\mathcal{F} := \{F(\cdot,\theta) : \mathbb{R} \to [0,1] \,|\, \theta \in \Theta\}$. Thus, given the hypothesis being true, there is a parameter $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ with $F(x) = F(x,\theta_0)$. Since the value of θ_0 is unknown, it has to be estimated from the given data, which complicates the investigation substantially.
In order to quantify a deviation from H_0, empirical processes are again considered. However, now there is no hypothesised distribution function F_0 to compare with the empirical distribution function F_n of the random sample $X_1, \ldots, X_n$. Thus, within $\mathcal{F}$ we estimate a distribution function $F(x,\hat{\theta}_n)$ and compare it with the empirical distribution function F_n. In doing so, we get the so-called parametric empirical process

$$V_{n,\hat{\theta}_n}(x) := \sqrt{n}\left(F_n(x) - F(x,\hat{\theta}_n)\right), \quad x \in \mathbb{R}, \tag{6.4}$$

where $\hat{\theta}_n$ is an estimator with values in Θ, which we are going to specify in Assumption **(A2)** in a moment and again use functionals of this empirical process to generate test statistics. Analogously to the simple hypothesis case, it would be convenient to have a uniform version of the parametric empirical process (see equation (6.2)), which is defined on $[0,1]$ and whose distribution does not depend on the hypothesised family of distribution functions $\mathcal{F}$. To achieve this, in the simple hypothesis case, the substitution $t = F_0(x)$, $x \in \mathbb{R}$, was performed, but now there is no hypothesised distribution function

F_0, but a whole hypothesised family of distribution functions $\mathcal{F}$. Therefore, we take the estimated distribution function $F(x, \hat{\theta}_n)$ and conduct the substitution

$$t = F(x, \hat{\theta}_n), \quad x \in \mathbb{R}.$$

In doing so, we define the uniform parametric empirical process

$$\overline{V}_{n,\hat{\theta}_n}(t) := \sqrt{n}\left(\hat{F}_n(t) - t\right), \quad t \in [0,1], \tag{6.5}$$

where $\hat{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(F(X_i, \hat{\theta}_n))$, $x \in \mathbb{R}$, is the empirical distribution function of the random sample

$$F(X_1, \hat{\theta}_n), \ldots, F(X_n, \hat{\theta}_n).$$

Note that different to the simple hypothesis case, where the distribution of the uniform empirical process $\overline{V}_n(t)$ did not depend on model characteristics, the distribution of $\overline{V}_{n,\hat{\theta}_n}(\cdot)$ depends on the distribution of the random sample $F(X_1, \hat{\theta}_n), \ldots, F(X_n, \hat{\theta}_n)$. Similar to equation (6.3), we again have for $t \in [0,1]$ and $x = F^-(t, \hat{\theta}_n)$

$$\begin{aligned}
\overline{V}_{n,\hat{\theta}_n}(t) &= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,t]}(F(X_i, \hat{\theta}_n)) - t\right) \\
&= \sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,F^-(t,\hat{\theta}_n)]}(X_i) - F(F^-(t, \hat{\theta}_n), \hat{\theta}_n)\right) \\
&= \sqrt{n}\left(F_n(x) - F(x, \hat{\theta}_n)\right) \\
&= V_{n,\hat{\theta}_n}(x).
\end{aligned}$$

In order to determine the asymptotic distribution of $\overline{V}_{n,\hat{\theta}_n}$, the following two assumptions are made:

(A2)

Under the model hypothesis H_0, an estimator $\hat{\theta}_n$ exists for the true but unknown θ_0, which, for $n \to \infty$, admits the expansion

$$\sqrt{n}\,(\hat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} h(X_i, \theta_0) + o_{\mathrm{P}}(1).$$

Here, $h(x, \theta)$ is an $\mathbb{R}^p$-valued function with

i) $\mathbb{E}(h(X, \theta_0)) = 0$,

ii) $Cov(h(X, \theta_0))$ is existent,

wherein $X \sim F$.

(A3)

$F(x,\theta)$ is differentiable with respect to θ for all x for all $\theta \in K$ and the vector-valued function

$$f(t,\theta) := \frac{\partial F(x,\theta)}{\partial \theta}\Big|_{x=F^-(t,\theta)}$$

is continuous in (t,θ) for all $\theta \in K$ and all $t \in [0,1]$, where $K \subseteq \Theta$ is the closure of a neighbourhood of $\theta_0 \in \Theta$. Note that it was assumed that $\Theta \subseteq \mathbb{R}^p$ is an open set.

Under the Assumptions **(A1)** - **(A3)**, it can be shown that under H_0,

$$\overline{V}_{n,\hat{\theta}_n}(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} \overline{V}_{f,\theta_0}(\cdot) \quad \text{in } D[0,1], \tag{6.6}$$

where $\overline{V}_{f,\theta_0}$ is a zero-mean Gaussian process with covariance function

$$Cov(\overline{V}_{f,\theta_0}(t_1), \overline{V}_{f,\theta_0}(t_2)) = \min(t_1,t_2) - t_1 t_2 - f^T(t_1,\theta_0) \int_{[0,t_2]}^{(R)} h(F^-(s,\theta_0),\theta_0) ds$$

$$- f^T(t_2,\theta_0) \int_{[0,t_1]}^{(R)} h(F^-(s,\theta_0),\theta_0)\, ds + f^T(t_1,\theta_0)\, Cov(l(X,\theta_0))\, f(t_2,\theta_0)\,. \tag{6.7}$$

A first proof of (6.6) and (6.7) is due to James Durbin, see Durbin (1973b) Theorem 1, page 281, and the following corollaries on page 282.

Considering the limit process $\left(\overline{V}_{f,\theta_0}(t)\right)_{t\in[0,1]}$ in (6.6) and in particular its covariance structure (see (6.7)), it becomes apparent that its distribution depends on the family of hypothesized distribution functions $\mathcal{F}$. Furthermore, the true but unknown parameter θ_0 also occurs in (6.7). It follows that $\left(\overline{V}_{f,\theta_0}(t)\right)_{t\in[0,1]}$, and functionals thereof that may be used as test statistics are not asymptotically distribution free. This leads to the already mentioned problems in testing such parametric hypothesis, which we referred to as the "Durbin problem".

Remark 6.1.2.

Note that **Assumption 1** and **Assumption 2** from Section 2.1 are very similar to Assumptions **(A1)** and **(A2)** from above. From a historical perspective, this is not at all surprising, since the goodness-of-fit tests for regression models developed by Stute, Thies, and Zhu (1998) were derived from the goodness-of-fit tests for the distribution function of random samples.

6.1.2 Khmaladze transformation

A first proposal to get around the Durbin problem was made by Durbin (1973a), see p. 59, who suggested to randomly split the sample in halves and take one half to estimate the unknown parameter θ_0 and, then use this estimate — which now is handled as if it were known — together with the second half to construct a then distribution free test for H_0. Apart from the loss of power, the downside to this procedure is the fact that test results will rely on which particular half of the sample was used for estimation.

In order to find a way to circumvent the Durbin problem, in a far-reaching paper, Khmaladze (1982) suggested a martingale transformation of a slightly different time-transformation of the parametric empirical process

$$V_{n,\hat{\theta}_n}(x) = \sqrt{n}\left(F_n(x) - F(x,\hat{\theta}_n)\right), \quad x \in \mathbb{R},$$

which he had already examined in a preceding paper. To be more precise, in Khmaladze (1980), he investigated the process

$$\overline{U}_{n,\hat{\theta}_n}(t) := \sqrt{n}\left(\overline{F}_n(t) - F(F^-(t,\theta_0),\hat{\theta}_n)\right), \quad t \in [0,1], \tag{6.8}$$

arising from $V_{n,\hat{\theta}_n}(x)$ by inserting $t = F(x,\theta_0)$ under the assumption that H_0 holds true, since then there is $\theta_0 \in \Theta$ such that $F(x) = F(x,\theta_0)$ holds true. See Example 6.1.1 for the definition of $\overline{F}_n$.
In contrast, Durbin (1973b) investigated the uniform parametric empirical process

$$\overline{V}_{n,\hat{\theta}_n}(t) = \sqrt{n}\left(\hat{F}_n(t) - t\right), \quad t \in [0,1],$$

which originates from $V_{n,\hat{\theta}_n}(x)$ and the substitution $t = F(x,\hat{\theta}_n)$. Note that again, and similar to (6.3), there is

$$\overline{U}_{n,\hat{\theta}_n}(t) = V_{n,\hat{\theta}_n}(x), \quad \text{for } x = F^-(t,\theta_0) \text{ and } t \in [0,1].$$

Therefore, $\overline{U}_{n,\hat{\theta}_n}$ is suitable to detect deviations from H_0, too. Note furthermore that, though $\overline{U}_{n,\hat{\theta}_n}$ is, like $\overline{V}_{n,\hat{\theta}_n}$, a uniform parametric empirical process originating from $V_{n,\hat{\theta}_n}(x)$ via a substitution, $\overline{U}_{n,\hat{\theta}_n}$ is not (in practice) computable, unlike $V_{n,\hat{\theta}_n}(x)$, since θ_0 is unknown.

Making the Assumptions **(A1)** - **(A3)** (see Section 6.1.1), Khmaladze then shows that under H_0

$$\overline{U}_{n,\hat{\theta}_n}(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B^0(\cdot) - f^T(\cdot,\theta_0)\int_0^1 h(s,\theta_0)\,dB^0(s) =: \overline{U}_{f,\theta_0}(\cdot) \quad \text{in } \mathcal{L}^2([0,1],\lambda\!\!\!\lambda), \tag{6.9}$$

where

- $\theta_0 \in \Theta$ is the true but unknown parameter for which $F(x, \theta_0) = F(x)$ holds true,
- $\hat{\theta}_n$ is an estimator for θ_0 satisfying **(A2)** from Section 6.1.1,
- $B^0 = (B^0(t))_{t\in[0,1]}$ is the Brownian bridge on $[0, 1]$,
- $f(t, \theta) := \frac{\partial F(F^-(t,\theta_0),\theta)}{\partial\theta}$ is a continuous vector-valued function (see **(A3)** in Section 6.1.1),
- $h(x, \theta)$ is vector-valued function (see **(A2)** in Section 6.1.1),
- $\int_0^1 h(s, \theta_0)\, dB^0(s)$ denotes the stochastic integral with respect to the Brownian bridge B^0.

Additionally, Khmaladze shows that $\overline{U}_{f,\theta_0}$ is a zero-mean Gaussian process with covariance function

$$Cov(\overline{U}_{f,\theta_0}(t_1), \overline{U}_{f,\theta_0}(t_2)) = \min(t_1, t_2) - t_1 t_2 - f^T(t_1, \theta_0) \int_{[0,t_2]}^{(R)} h(F^-(s, \theta_0), \theta_0)\, ds$$

$$- f^T(t_2, \theta_0) \times \int_{[0,t_1]}^{(R)} h(F^-(s, \theta_0), \theta_0)\, ds + f^T(t_1, \theta_0)\, Cov(l(X, \theta_0))\, f(t_2, \theta_0)\,. \quad (6.10)$$

For proofs of (6.9) and (6.10), see Khmaladze (1980), pages 287 to 289.

Considering (6.6), (6.7), (6.9) and (6.10), it becomes apparent that

$$\overline{V}_{f,\theta_0}(\cdot) = \overline{U}_{f,\theta_0}(\cdot) = B^0(\cdot) - f^T(\cdot, \theta_0) \int_0^1 h(s, \theta_0)\, dB^0(s) \quad (6.11)$$

holds true, since both are zero-mean Gaussian processes with the same covariance function. We thus have

$$\overline{U}_{n,\hat{\theta}_n}(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} \overline{V}_{f,\theta_0}(\cdot) \quad \text{in } D[0, 1], \quad (6.12)$$

too. Test statistics based on $\overline{U}_{n,\hat{\theta}_n}$ are not asymptotically distribution free, as the distribution of $\overline{U}_{f,\theta_0}$ depends on model characteristics. Note that unlike Khmaladze states in Khmaladze (1980), page 288, Durbin (1973b) has not shown (6.12) but has shown (6.6), which initially is a different statement.

Khmaladze (1982) proposed to transform the uniform parametric empirical process $\overline{U}_{n,\hat{\theta}_n}$ into a martingale in such a way that (under H_0) the distribution of the resulting martingale converges weakly to the distribution of a standard Brownian motion and thus does not depend on these model characteristics. This transformation therefore is called "Khmaladze martingale transformation" — or briefly just "Khmaladze transformation"— and denoted by K in this section.
On the following pages, we will approach the Khmaladze transformation in two major steps and a variety of smaller steps, where, for the sake of brevity, we mostly give references to proofs. In the first step, the Doob-Meyer-Decomposition of the uniform empirical process $\overline{V}_n(t)$ in the simple hypothesis scenario (see Example 6.1.1) will be given and used as a blueprint to determine the martingale part of the uniform parametric empirical process $\overline{U}_{n,\hat{\theta}_n}$ in the general situation in Step 2.

Step 1

i) The paths of the empirical distribution function $\overline{F}_n$ are bounded and monotonically increasing. Consequently, $\left(\overline{F}_n(t)\right)_{t\in[0,1]}$ is a submartingale with respect to the natural flow of $\sigma-$algebras $\left(\mathcal{A}_{\overline{F}_n,t}\right)_{t\in[0,1]}$, where $\mathcal{A}_{\overline{F}_n,t} := \sigma\left(\overline{F}_n(s)|\, s \leq t\right)$. Therefore, the Doob-Meyer decomposition yields a representation

$$\overline{F}_n(t) = A_n(t) + M_n(t),$$

where M_n is a martingale and $A_n(t)$ is a predictable (i.e., in our case left-continuous), monotonically increasing process starting at 0 and often referred to as the compensator. See for example Beiglboeck, Schachermayer, and Veliyev (2012). To be more specific, $\overline{F}_n$ admits the following representation:

$$\overline{F}_n(t) = \int_{[0,t]}^{(R)} \frac{1-\overline{F}_n(s)}{1-s}\, ds + M_n(t), \quad t \in [0,1],$$

where

$$\left(\int_{[0,t]}^{(R)} \frac{1-\overline{F}_n(s)}{1-s}\, ds\right)_{t\in[0,1]}$$

is the left continuous increasing compensator and

$$M_n(t) := \overline{F}_n(t) - \int_{[0,t]}^{(R)} \frac{1-\overline{F}_n(s)}{1-s}\, ds$$

is a martingale with respect to $\left(\mathcal{A}_{\overline{F}_n,t}\right)_{t\in[0,1]}$. See Khmaladze (1982), p. 242 for a proof.

ii) As Khmaladze (1982), p. 243, has shown, the uniform empirical process $\overline{V}_n(t)$ (see (6.2)) admits the representation

$$\overline{V}_n(t) = -\int\limits_{[0,t]}^{(R)} \frac{\overline{V}_n(s)}{1-s}\, ds + W_n(t), \tag{6.13}$$

where $W_n(t) := \sqrt{n}\, M_n(t)$ is a martingale with respect to the filtration $\left(\mathcal{A}_{\overline{V}_n,t}\right)_{t\in[0,1]}$ with $\mathcal{A}_{\overline{V}_n,t} := \sigma\left(\overline{V}_n(s)|\, s \le t\right) = \sigma\left(\overline{F}_n(s)|\, s \le t\right)$. Therefore, $\left(\mathcal{A}_{\overline{V}_n,t}\right)_{t\in[0,1]} = \left(\mathcal{A}_{\overline{F}_n,t}\right)_{t\in[0,1]}$.

iii) For $B = (B(t))_{t\in[0,1]}$, a standard Brownian motion, the process

$$B^0(t) := (1-t)\int\limits_0^t \frac{1}{1-s}\, dB(s), \quad t \in [0,1)$$

is a Brownian bridge. Here, $\int\limits_0^t \frac{1}{1-s}\, dB(s)$ denotes the stochastic integral with respect to the standard Brownian motion B. Moreover, B^0 is the solution of the stochastic differential equation

$$dX_t = -\frac{X_t}{1-t}\, dt + dB(t), \quad t \in [0,1) \text{ with } X_0 = 0.$$

Thus,

$$B^0(t) = -\int\limits_{[0,t]}^{(R)} \frac{B^0(s)}{1-s}\, ds + B(t), \quad t \in [0,1), \tag{6.14}$$

is satisfied, which is the Doob-Meyer decomposition of the Brownian bridge with the predictable compensator $A(t) = -\int\limits_{[0,t]}^{(R)} \frac{B^0(s)}{1-s}\, ds$ and the Brownian motion $B(t)$ as the martingale part.

iv) Comparing equation (6.13) and equation (6.14) and taking into account that $\overline{V}_n$ converges in $C[0,1]$ (and therefore also in $D[0,1]$) in distribution to B^0 (see Billingsley (1968) Theorem 13.1, p. 105) it is natural to guess that

$$W_n(t) \xrightarrow[n\to\infty]{\mathcal{D}} B(t) \quad \text{in } D[0,1]. \tag{6.15}$$

Indeed, equation (6.15) holds true, which has been proven, for example, by Aki (1986), pages 2 - 5.

Using this knowledge in the situation of Example 6.1.1, page 76, we now know that

under $H_0 : F = F_0$,

$$\left(\widetilde{K}(\overline{V}_n)\right)(t) := \overline{V}_n(t) + \int\limits_{[0,t]}^{(R)} \frac{\overline{V}_n(s)}{1-s}\, ds = W_n(t) \xrightarrow[n\to\infty]{\mathcal{D}} B(t) \quad \text{in } D[0,1].$$

Therefore, using the continuous mapping theorem, test statistics based on $\widetilde{K}(\overline{V}_n)$ can be generated which are asymptotically distribution free and of known distribution in order to test the simple hypothesis

$$H_0 : F \in \mathcal{F} \quad \text{v.s.} \quad H_1 : F \notin \mathcal{F}.$$

Step 2

i) In the general situation, it would be desirable to find a decomposition for the limit process $\overline{V}_{f,\theta_0}(t)$ of the uniform parametric empirical process $\overline{U}_{n,\hat{\theta}_n}(t)$ similar to the decomposition we found in Step 1, equation (6.14) for the Brownian bridge B^0, which is the limit process of uniform empirical process $\overline{V}_n(t)$. Khmaladze (1982), p. 250 f., showed that under **(A1)** - **(A3)** from Section 6.1.1 with

$$h(t,\theta_0) = f'(t,\theta_0) := \frac{\partial f(t,\theta_0)}{\partial t} \quad \text{and} \quad \int\limits_{[0,1]}^{(R)} f'(t,\theta_0)\, f'^{T}(t,\theta_0)\, dt = I_p, \tag{6.16}$$

where I_p is the $p \times p$ unit matrix and under the additional assumption

(A4) The functions $f'_1(t,\theta_0), \ldots, f'_p(t,\theta_0) \in \mathcal{L}^2([0,1]\lambda_{[0,1]})$ exist and are linearly independent in a neighbourhood of $t = 1$.

the decomposition

$$\overline{V}_{f,\theta_0}(t) = B(t) + \int\limits_{[0,t]}^{(R)} f'^{T}(s,\theta_0) \int\limits_0^s \left(\int\limits_{[\tau,1]}^{(R)} f'(v,\theta_0)\, f'^{T}(v,\theta_0)\, dv \right)^{-1} f'(\tau,\theta_0)\, dB(\tau)\, ds,$$

holds true for $t \in [0,1)$. Thus, $\overline{V}_{f,\theta_0}(\cdot)$ has the martingale part

$$M(t) = B(t)$$

and the compensator

$$A(t) = \int\limits_{[0,t]}^{(R)} f'^{T}(s,\theta_0) \int\limits_0^s \left(\int\limits_{[\tau,1]}^{(R)} f'(v,\theta_0)\, f'^{T}(v,\theta_0)\, dv \right)^{-1} f'(\tau,\theta_0)\, dB(\tau)\, ds.$$

In the same theorem, Khmaladze states that this relation is invertible, so that for $t \in [0,1)$, also

$$B(t) = \overline{V}_{f,\theta_0}(t) + \int\limits_{[0,t]}^{(R)} f'^{\,T}(s,\theta_0) \left(\int\limits_{[s,1]}^{(R)} f'(v,\theta_0)\, f'^{\,T}(v,\theta_0)\, dv \right)^{-1} \int\limits_0^s f'(\tau,\theta_0)\, d\overline{V}_{f,\theta_0}(\tau)\, ds, \tag{6.17}$$

is valid. Here, $\int\limits_0^s f'(\tau,\theta_0)\, d\overline{V}_{f,\theta_0}(\tau)$ is a stochastic integral and $B \equiv (B(t))_{t\in[0,1]}$ is a Brownian motion with respect to the filtration $\left(\mathcal{A}_{\overline{V}_{f,\theta_0},t}\right)_{t\in[0,1]}$, where $\mathcal{A}_{\overline{V}_{f,\theta_0},t} := \sigma\left(\overline{V}_{f,\theta_0}(s) |\, s \le t\right)$.

ii) Because of the identity (6.11) and the addition of (6.16) to **(A3)** in Step 2 part i), we have

$$\begin{aligned}
\int\limits_0^1 f'(t,\theta_0)\, d\overline{V}_{f,\theta_0}(t) &= \int\limits_0^1 f'(t,\theta_0)\, d\left(B^0(\cdot) - f^T(\cdot,\theta_0) \int\limits_0^1 f'(s,\theta_0)\, dB^0(s) \right)(t) \\
&= \int\limits_0^1 f'(t,\theta_0)\, dB^0(t) - \int\limits_{[0,1]}^{(R)} f'(t,\theta_0)\, d\left(f^T(\cdot,\theta_0) \int\limits_0^1 f'(s,\theta_0)\, dB^0(s) \right)(t) \\
&= \int\limits_0^1 f'(t,\theta_0)\, dB^0(t) - I_p \int\limits_0^1 f'(t,\theta_0)\, dB^0(t) \\
&= 0.
\end{aligned}$$

Therefore,

$$\int\limits_0^s f'(t,\theta_0)\, d\overline{V}_{f,\theta_0}(t) = -\int\limits_s^1 f'(t,\theta_0)\, d\overline{V}_{f,\theta_0}(t), \quad s \in [0,1],$$

holds true. As a consequence, for $t \in [0,1)$, equation (6.17) is equivalent to

$$B(t) = \overline{V}_{f,\theta_0}(t) - \int\limits_{[0,t]}^{(R)} f'^{\,T}(s,\theta_0) \left(\int\limits_{[s,1]}^{(R)} f'(v,\theta_0)\, f'^{\,T}(v,\theta_0)\, dv \right)^{-1} \int\limits_s^1 f'(\tau,\theta_0)\, d\overline{V}_{f,\theta_0}(\tau)\, ds. \tag{6.18}$$

iii) Considering equation (6.18), the right hand side can be interpreted as a transformation of $\overline{V}_{f,\theta_0}(t)$, which is the limit process of the uniform parametric empirical process $\overline{U}_{n,\hat{\theta}_n}(t)$. It is natural to guess that the same kind of transformation applied to the uniform parametric empirical process $\overline{U}_{n,\hat{\theta}_n}(t)$ itself is also (asymptotically) distributed like a Brownian motion. To investigate this closer, Khmaladze (1982),

page 253, defined for $t \in [0,1)$,

$$W_n(t) := \overline{U}_{n,\hat{\theta}_n}(t) - \int\limits_{[0,t]}^{(R)} {g'}^T(s,\theta_0) \left(\int\limits_{[s,1]}^{(R)} g'(v,\theta_0) {g'}^T(v,\theta_0)\, dv \right)^{-1} \int\limits_s^1 g'(\tau,\theta_0)\, d\overline{U}_{n,\hat{\theta}_n}(\tau)\, ds. \tag{6.19}$$

Here, $\int\limits_s^1 g'(\tau,\theta_0)\, d\overline{U}_{n,\hat{\theta}_n}(\tau)$ is a stochastic integral, $g(t,\theta_0) := \Gamma^{-1/2} \begin{pmatrix} t \\ f(t,\theta_0) \end{pmatrix}$ with $\Gamma := \begin{pmatrix} 1 & 0 \\ 0 & \gamma \end{pmatrix}$, $\gamma := (\gamma_{i,j})_{i,j=1}^p$ and $\gamma_{i,j} := \int\limits_0^1 f_i'(t,\theta_0)\, {f_j'}^T(t,\theta_0)\, dt$, $i,j = 1,\dots,p$.

Furthermore, the following additional assumptions are made:

(A4') The functions $f_i'(t,\theta_0) := \frac{\partial f_i(t,\theta_0)}{\partial t}$, $i = 1,\dots,p$, exist, are elements of $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$ and $1, f_1(t,\theta_0), \dots, f_p(t,\theta_0)$ are linearly independent in the neighbourhood of $t = 1$.

(A5) $\int\limits_{[0,1]}^{(R)} \sup \left\{ |f'(t,\theta) - f'(t,\theta_0)|^2 \;\middle|\; \theta \in \Theta,\, \sum_{i=1}^m |\theta_i - \theta_{0,i}|^2 < \varepsilon \right\} dt \xrightarrow[\varepsilon \to 0]{} 0.$

(A2') The estimator is square root consistent, which means that $\sqrt{n}(\hat{\theta}_n - \theta_0) = O_{\mathbb{P}}(1)$ for $n \to \infty$. This is a weaker assumption than **(A2)** from Section 6.1.1, but strong enough to ensure — together with **(A5)** — that

$$\overline{U}_{n,\hat{\theta}_n}(t) = \overline{V}_n(t) - f^T(t,\theta_0)\, \sqrt{n}\, (\hat{\theta}_n - \theta_0) + r_n(t),$$

with $\int\limits_0^1 r_n'^2(t)\, dt = o_{\mathbb{P}}(1)$, holds true for $n \to \infty$.

Note that due to

$$f'(t,\theta_0) = \left(\frac{\partial}{\partial \theta} \log \frac{\partial}{\partial t} F \right) (F^-(t,\theta_0), \theta_0), \tag{6.20}$$

$f'(t,\theta_0)$ is the score function of the model family $\mathcal{F}$. Thus γ is an augmented version of the Fisher-information matrix of the model family $\mathcal{F}$. For a proof of (6.20), see Lemma 7.5.1 in Section 7.5 of the appendix. For some theory on score functions and the Fisher-information matrix, see for example Ly et al. (2017). Note further that $W_n(t)$ as defined above is different from the $W_n(t)$ defined in Step 1. Nevertheless, we use the same notation to emphasise the analogy between the two.

iv) In the present situation it can be shown that under H_0 and the Assumptions **(A1)**, **(A2')**, **(A3)**, **(A4')** and **(A5)**

$$W_n(t) = \tag{6.21}$$

$$= \underbrace{\sqrt{n}\left(\overline{F}_n(t) - \int\limits_{[0,1]}\int\limits_{[0,\min\{t,\tau\}]}^{(R)} g'^T(s,\hat{\theta}_n)\left(\int\limits_{[s,1]}^{(R)} g'(v,\hat{\theta}_n)g'^T(v,\hat{\theta}_n)\,dv\right)^{-1} ds\; g'(\tau,\hat{\theta}_n)\,d\overline{F}_n(\tau)\right)}_{=:(K(\overline{F}_n))(t)} + o_{\mathbb{P}}(1)$$

for $n \to \infty$ and $t \in [0,1)$ as well as

$$W_n(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B(\cdot) \quad \text{in } \mathcal{L}^2([0,1],\lambda\!\!\lambda)$$

hold true, meaning that $W_n(\cdot)$ is asymptotically distributed like a standard Brownian motion. Thus, $W_n(\cdot)$ is asymptotically distribution free. Again, appropriate functionals of W_n and therefore $K(\overline{F}_n)$ respectively can be used to generate asymptotically distribution free test statistics of a known distribution.
Note that K is essentially the same transformation as the one conducted on the right hand side of equation (6.19), but with the estimate $\hat{\theta}_n$ instead of θ_0, and applied to $\overline{F}_n$ instead of $\overline{U}_{n,\hat{\theta}_n}$. Both changes are asymptotically negligible. For a proof of the first statement, see Khmaladze (1982) pages 253 f., 255 and 257. For a proof of the second one, see Khmaladze (1982) pages 256 f.

v) Considering equation (6.21), it becomes apparent that for $t \in [0,1)$

$$A_n(t,\overline{F}_n,\hat{\theta}_n) := \int\limits_{[0,1]}\int\limits_{[0,\min\{t,\tau\}]}^{(R)} g'^T(s,\hat{\theta}_n)\left(\int\limits_{[s,1]}^{(R)} g'(v,\hat{\theta}_n)\,g'^T(v,\hat{\theta}_n)\,dv\right)^{-1} ds\, g'(\tau,\hat{\theta}_n)\,d\overline{F}_n(\tau)$$

manipulates $\overline{F}_n(t)$ in such a way that the (appropriately scaled) result

$$\sqrt{n}\left(\overline{F}_n(\cdot) - A_n(\cdot,\overline{F}_n,\hat{\theta}_n)\right)$$

is asymptotically a martingale. Therefore, $A_n(t,\overline{F}_n,\hat{\theta}_n)$ is called the compensator of $\overline{F}_n(t)$. Utilising Fubini's theorem, we additionally see that

$$\begin{aligned} A_n(t,\overline{F}_n,\hat{\theta}_n) &= \int\limits_{[0,1]}\int\limits_{[0,\min\{t,\tau\}]}^{(R)} g'^T(s,\hat{\theta}_n)\left(\int\limits_{[s,1]}^{(R)} g'(v,\hat{\theta}_n)\,g'^T(v,\hat{\theta}_n)\,dv\right)^{-1} ds\, g'(\tau,\hat{\theta}_n)\,d\overline{F}_n(\tau) \\ &= \int\limits_{[0,t]}^{(R)} g'^T(s,\hat{\theta}_n)\left(\int\limits_{[s,1]}^{(R)} g'(v,\hat{\theta}_n)\,g'^T(v,\hat{\theta}_n)\,dv\right)^{-1} \int\limits_{[s,1]} g'(\tau,\hat{\theta}_n)\,d\overline{F}_n(\tau)\,ds \end{aligned} \tag{6.22}$$

holds true, which opens the door for geometrical interpretations of the compensator as a projection operator. The discussion of such interpretations takes place in the next two sections in the situation of Khmaladze transformations in goodness-of-fit tests for linear regression models.

6.2 Khmaladze transformation in linear regression models

Returning to the main subject of study in this work, we will again consider regression models, their respective CUSUM processes and goodness-of-fit tests based on them.
The Khmaladze transformation was proposed in Khmaladze (1982) as a remedy to the Durbin problem in a different scenario of goodness-of-fit tests — see Section 6.1.2. Stute, Thies, and Zhu (1998) have adapted the Khmaladze transformation to the context of goodness-of-fit tests via marked empirical processes of regression residuals in (not necessary linear) regression models with random designs. The result of this adaptation is shown in the following theorem.

Theorem 6.2.1. **[Stute, Thies, and Zhu (1998), Theorem 1.2]**

Let the prerequisites of a regression model with random design (see Theorem 2.1.6) be fulfilled and assume homoscedastic regression errors with variance $\sigma^2 = 1$. Furthermore, let

$$\int\limits_{(y,\infty)} g(u)\, g^T(u)\, dF(u)$$

be invertible for all $y \in \mathbb{R}$ and define the so-called Khmaladze transformation T^* by

$$(T^*(h))(x) := h(x) - \int\limits_{(-\infty,x]} g^T(y) \left(\int\limits_{(y,\infty)} g(u)\, g^T(u)\, dF(u) \right)^{-1} \left(\int\limits_{(y,\infty)} g(z)\, dh(z) \right) dF(y)$$

for functions h either of bounded variation or time-transformed Brownian motions $B \circ F$. Then, under H_0,

$$\left(T^* \circ R_n^1\right)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (B \circ F)(\cdot) \quad \text{in } D[-\infty, \infty).$$

With regard to the above theorem, note the following.

- In the case that T^* is applied to a function h that is of bounded variation, the integral $\int\limits_{(y,\infty)} g(z)\, dh(z)$ can be understood as a measure integral. Whereas in the case that h is a time-transformed Brownian motion $B \circ F$, the integral $\int\limits_{(y,\infty)} g(z)\, dh(z)$ must be understood as a stochastic integral.

- The fact that T^* is a direct adaptation of the Khmaladze transformation K from Section 6.1.2 to the situation of a linear regression model with random design becomes apparent when looking at equation (6.21) on page 88 and equation (6.22) on page 88.

Considering the limit process $(B \circ F)(\cdot)$ in the above theorem, it is striking that this process still depends on model characteristics, namely the distribution F of the design. In

the special case of linear regression models, it is possible to overcome this dependency as follows: According to Section 3.2, in goodness-of-fit tests in linear regression models based on the asymptotic distribution of residual CUSUM processes, one can assume without loss of generality a generic linear regression model. Therefore, in the following corollary, we can specify a version of Theorem 6.2.1 in this scenario.

Corollary 6.2.2.

Let a generic linear regression model (see Definition 3.2.1) be given. Additionally assume that

$$\int\limits_{[y,1]} f(u)\, f^T(u)\, d\lambda\!\!\!\lambda_{[0,1]}(u)$$

is invertible for all $y \in [0,1)$. Define the Khmaladze transformation T^* via

$$(T^*(h))(x) := h(x) - \int\limits_{[0,x]} f^T(y) \left(\int\limits_{[y,1]} f(t)\, f^T(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^{-1} \left(\int\limits_{[y,1]} f(t)\, dh(t) \right) d\lambda\!\!\!\lambda_{[0,1]}(y)$$

for $x \in [0,1)$ and functions h that are either of bounded variation or a Brownian motion. Then, under H_0, it holds true that

$$\left(T^* \circ \frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n) \right)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B(\cdot) \quad \text{in } C[0,1)$$

and

$$\left(T^* \circ R_n^1 \right)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B(\cdot) \quad \text{in } C[0,1).$$

Remark 6.2.3.

In the above corollary, for suitable functions g, the Khmaladze transformation

$$(T^*(g))(x) = g(x) - \int\limits_{[0,x]} f^T(y) \left(\int\limits_{[y,1]} f(t) f^T(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^{-1} \left(\int\limits_{[y,1]} f(t)\, dg(t) \right) d\lambda\!\!\!\lambda_{[0,1]}(y)$$

of g looks similar to the orthogonal projection

$$pr_{W_{\mathcal{H}_B}^{\perp}}(g)(x) = g(x) - \left(\int\limits_{[0,x]} f(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^T \left(\int\limits_{[0,1]} f(t) f^T(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t) \right)^{-1} \int\limits_{[0,1]} f(t)\, dg(t)$$

of g onto the orthogonal complement of $W_{\mathcal{H}_B} = span\{h_{f_1} \ldots, h_{f_p}\}$, which we discussed in Section 4.2. This similarity motivates a further analysis of the connection between the Khmaladze transformation and projections onto RKHS of integrated regression functions and thus residual CUSUM processes. This analysis will be the main topic of the next section.

6.3 Khmaladze transformation as a recursive residual transformation

Recursive (least-squares) residuals have already been used in a landmark paper by Brown, Durbin, and Evans (1975) to investigate the stability over time of regression relationships using CUSUM processes. Therein, recursive residuals are defined as standardised differences between current observations and their forecasts that were made with the previously sampled observations. We note that the notion of recursive residuals has been in use since 1891, as Farebrother (1978) points out.
Brown, Durbin, and Evans (1975) develop their results on recursive residuals for time series and Sen (1982) provides an invariance principle for CUSUM processes of recursive residuals in this setting. Investigations of the CUSUM tests of Brown, Durbin, and Evans (1975) under more general conditions can be found in Krämer, Ploberger, and Alt (1988). Evers (2022) proves a result similar to that in Sen (1982) in the situation of linear regression models with triangular arrays of design points and uses this result to calculate the limit under local alternatives in a simple manner.
In this work, however, we consider so-called backwards recursive (least-squares) residuals (see Definition 6.3.2). Although it is already remarked in Brown, Durbin, and Evans (1975) on page 155 that "it is often informative to look at the set of plots [of the CUSUM processes] which are obtained by running the analysis backwards through time as well as forwards", no specific definitions or results are given regarding backwards recursive residuals. In the context of ordinary least-squares residuals, Bischoff et al. (2005) have already shown that tests based on partial sums can have a larger power when the partial sums are taken from the time reversed data.
Explicit definitions of backwards recursive residuals as well as examples of their use in change point analysis can be found in Hawkins (1987), Pesaran and Timmermann (2002) and more recently in Otto and Breitung (2020).

Let us consider the univariate linear regression model with random design

$$Y = (f_1, \dots, f_p)(X)\beta + \varepsilon,$$

where $X \sim F$, F is a distribution function with $supp(F) \subseteq [a,b]$, $a < b \in \mathbb{R}$, $f_1, \dots, f_p$, $p \in \mathbb{N}$, are the regression functions that are of bounded variation, $\beta = (\beta_1, \dots, \beta_p)^T$ are the regression coefficients and ε is the regression error with $\mathbb{E}(\varepsilon \,|\, X) = 0$ and $\mathbb{V}ar(\varepsilon \,|\, X) \equiv 1$. Moreover, suppose that there is a $c \in [a,b)$ with $F(c) < 1$ and such that $f_1 \cdot \mathbb{1}_{(c,b]}, \dots, f_p \cdot \mathbb{1}_{(c,b]}$ are linearly independent in $\mathcal{L}^2([a,b], F)$.
Consider an n-sample $(\widetilde{Y}_{n1}, \widetilde{X}_{n1}, \widetilde{\varepsilon}_{n1}), \dots, (\widetilde{Y}_{nn}, \widetilde{X}_{nn}, \widetilde{\varepsilon}_{nn}) \overset{iid}{\sim} (Y, X, \varepsilon)$, $n \in \mathbb{N}$. We denote with $X_{ni} := X_{i:n}$, $i \in \{1, \dots, n\}$, the i-th order statistic of $\{\widetilde{X}_{n1}, \dots, \widetilde{X}_{nn}\}$ and with

$$Y_{ni} := Y_{[i:n]} \quad \text{and} \quad \varepsilon_{ni} := \varepsilon_{[i:n]}, \quad i = 1, \dots, n,$$

we denote the associated concomitants. Then, the above model has the form

$$\begin{pmatrix} Y_{n1} \\ \vdots \\ Y_{nn} \end{pmatrix} = \begin{pmatrix} f_1(X_{n1}) & \dots & f_p(X_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(X_{nn}) & \dots & f_p(X_{nn}) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nn} \end{pmatrix}, \quad n \in \mathbb{N}.$$

Let a realisation $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ be given. We thus consider the linear regression model with fixed design

$$\begin{pmatrix} Y_{n1} \\ \vdots \\ Y_{nn} \end{pmatrix} = \begin{pmatrix} f_1(t_{n1}) & \dots & f_p(t_{n1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \dots & f_p(t_{nn}) \end{pmatrix} \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix} + \begin{pmatrix} \varepsilon_{n1} \\ \vdots \\ \varepsilon_{nn} \end{pmatrix}, \quad n \in \mathbb{N}, \tag{6.23}$$

where

- $Y_n = (Y_{n1}, \dots, Y_{nn})^T$ is the vector of outcomes,
- $[a, b]$, $a < b$, is the experimental region,
- $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} \subseteq [a, b]$ with $t_{n1} \leq \dots \leq t_{nn}$, $n \in \mathbb{N}$, is the triangular array of design points. Note that for almost all realisations $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$, the corresponding sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$ converges uniformly to the distribution function F with $supp(F) \subseteq [a, b]$,
- $f = (f_1, \dots, f_p)^T : [a, b] \to \mathbb{R}^p$, $p \in \mathbb{N}$, are the regression functions that are of bounded variation and there is a $c \in [a, b)$ with $F(c) < 1$ and such that $f_1 \cdot \mathbb{1}_{(c,b]}, \dots, f_p \cdot \mathbb{1}_{(c,b]}$ are linearly independent in $\mathcal{L}^2([a, b], F)$,
- $\beta = (\beta_1, \dots, \beta_p)^T$ are the regression coefficients,
- $\varepsilon_{n1}, \dots, \varepsilon_{nn}$ are the iid regression errors with $\mathbb{E}(\varepsilon_{n1}) = 0$, and $\mathbb{V}ar(\varepsilon_{n1}) = 1$.

In order to proceed, we need the following designations.

- For $n \in \mathbb{N}$ and $i \in \{1, 2, \dots, n\}$, we define the notations

$$\underbrace{\begin{pmatrix} Y_{ni} \\ \vdots \\ Y_{nn} \end{pmatrix}}_{=:Y_n^{(i)} \in \mathbb{R}^{n-i+1}} = \underbrace{\begin{pmatrix} f_1(t_{ni}) & \dots & f_p(t_{ni}) \\ \vdots & \ddots & \vdots \\ f_1(t_{nn}) & \dots & f_p(t_{nn}) \end{pmatrix}}_{=:X_n^{(i)} \in \mathbb{R}^{(n-i+1)\times p}} \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}}_{=\beta\in\mathbb{R}^p} + \underbrace{\begin{pmatrix} \varepsilon_{ni} \\ \vdots \\ \varepsilon_{nn} \end{pmatrix}}_{=:\varepsilon_n^{(i)}},$$

  referring to the regression model (6.23) excluding the first $i - 1$ rows.

- For all $i \in \{1, \dots, n - p + 1\}$ such that $X_n^{(i)}$ has full rank, we denote with

$$\hat{\beta}_n^{(i)} := \underbrace{((X_n^{(i)})^T X_n^{(i)})}_{\in\mathbb{R}^{p\times p}}^{-1} (X_n^{(i)})^T Y_n^{(i)}$$

the ordinary least-squares estimates for β, using the last $n-i+1$ observations $Y_{ni}, Y_{ni+1}, \ldots, Y_{nn}$ corresponding to the design points $t_{ni}, t_{ni+1}, \ldots, t_{nn}$.

- As usual, for $i \in \{1, \ldots, n\}$, the ordinary least-squares residuals are denoted by $\hat{\varepsilon}_{ni} = Y_{ni} - f^T(t_{ni})\hat{\beta}_n$, where $\hat{\beta}_n := \hat{\beta}_n^{(1)}$. We denote the vector of the last $n-i+1$ ordinary least-squares residuals by $\hat{\varepsilon}_n^{(i)} := (\hat{\varepsilon}_{ni}, \ldots, \hat{\varepsilon}_{nn})^T$ and $\hat{\varepsilon}_n := \hat{\varepsilon}_n^{(1)}$.
- For $t_{n1} \leq \cdots \leq t_{nn}$, $n \in \mathbb{N}$, we define

$$j_n : \mathbb{R} \longrightarrow \{0, 1, \ldots, n\},\ y \longmapsto \max\{0, \sup\{i \in \{1, \ldots, n\} \mid t_{ni} \leq y\}\}.$$

For the remainder of this section, we will only consider triangular arrays $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ that satisfy the following assumption.

$$\text{The triangular array } ((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} \text{ is such that there exists an } n_0 \in \mathbb{N} \text{ such that for all } n \geq n_0, rank\left((X_n^{(i+1)})^T X_n^{(i+1)}\right) = p,\ i = 0, \ldots, j_n(c), \text{ is satisfied.} \tag{6.24}$$

Remark 6.3.1.

The above assumption about the triangular array $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ is quite natural, as it describes a generic realisation of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$. It can be verified that almost all realisations $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ satisfy Assumption (6.24). A proof of this assertion can be found on page 184 of the appendix.

We can now proceed to define the backwards recursive residuals.

Definition 6.3.2.

Consider the linear regression model (6.23). Let Assumption (6.24) be satisfied and let $n \geq n_0$. Then the backwards recursive (least-squares) residuals r_{ni} are defined by

$$r_{ni} := \frac{Y_{ni} - f^T(t_{ni})\hat{\beta}_n^{(i+1)}}{\sqrt{1 + f^T(t_{ni})\left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1} f(t_{ni})}}, \quad i = 1, \ldots, j_n(c).$$

Furthermore, we denote the vector of backwards recursive residuals by $r_n := \left(r_{n1}, r_{n2}, \ldots, r_{nj_n(c)}\right)^T \in \mathbb{R}^{j_n(c)}$.

Note that the above definition is only one of many ways to define recursive residuals. We have chosen this particular definition in order to make the most suggestive connection to the Khmaladze transformation (see Theorem 6.3.3).
In general, in a regression model with $n \in \mathbb{N}$ observations and p regression functions, $\frac{n!}{p!}$ different sets of recursive residuals can be defined (given that all data is already available

and assuming that each submatrix of the design matrix with p rows has full rank). For example, on page 151 Brown, Durbin, and Evans (1975) define recursive residuals ordered by time, taking the first p observations to compute an initial estimate for β, and proceeding stepwise from there to compute one recursive residual at a time. In contrast, it is only possible to compute the i-th backwards recursive residuum r_{ni} if all the outcomes $Y_{ni}, \ldots, Y_{nn}$ have been observed. This is the case, since for the computation of r_{ni} the estimate $\hat{\beta}_{i+1}$ has to be calculated, which requires the knowledge of $Y_{ni+1}, \ldots, Y_{nn}$. Backwards recursive residuals, similar to "normal" ones, i.e. non-backwards recursive residuals as defined, for example, in Brown, Durbin, and Evans (1975), have some convenient properties. For example, they are homoscedastic and uncorrelated and can be considered as the transformation of the heteroscedastic and correlated ordinary least squares residuals. See Lemma 7.5.2 and Lemma 7.5.3 in the appendix for these and other properties of backwards recursive residuals.

We now move on to the main theorem of this section. To our knowledge this theorem is new in the literature.

Theorem 6.3.3.

Consider the linear regression model (6.23). Let Assumption (6.24) be satisfied and let $n \geq n_0$. Furthermore, let

- T^* be the Khmaladze transformation introduced in Theorem 6.2.1,
- $R_n^1 := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(t_{ni}) \cdot \hat{\varepsilon}_{ni}$ be the marked empirical process of the vector of ordinary least-squares residuals $\hat{\varepsilon}_n$,
- $\widetilde{R}_n^1(x) := \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(c)} \mathbb{1}_{(-\infty,x]}(t_{ni}) \cdot r_{ni}$ be the marked empirical process of the vector of backwards recursive residuals r_n.

Then, the statement

$$(T^*(R_n^1))(x) = \widetilde{R}_n^1(x) + o_{\mathbb{P}}(1)$$

holds true uniformly in $x \in (-\infty, c]$ for $n \longrightarrow \infty$.

Note that, in view of Section 3.2, if we wanted to show a distributional result in Theorem 6.3.3, it would have sufficed to only consider the generic linear regression model. However, since the result addresses equality in probability, and is thus more general, we cannot without loss of generality make additional assumptions on the design.

Proof of Theorem 6.3.3: See the pages 187 and the following pages in Section 7.5 of the appendix. □

The proof of the above theorem shows that for $x \in (-\infty, c]$, $T^*(R_n^1)$ and $\widetilde{R}_n^1$ can be represented as follows:

$$(T^*(R_n^1))(x) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \hat{\varepsilon}_{ni} - \int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u) f^T(u) dF(u) \right)^{-1} \left(\int\limits_{(y,\infty)} f(z) dR_n^1(z) \right) dF(y)$$

and

$$\widetilde{R}_n^1(x) = \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \hat{\varepsilon}_{ni} - \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} f^T(t_{ni}) \Big((X_n^{(i+1)})^T X_n^{(i+1)} \Big)^{-1} (X_n^{(i+1)})^T \hat{\varepsilon}_n^{(i+1)} + o_{\mathbb{P}}(1).$$

Comparing these two representations, it becomes apparent that the application of the Khmaladze transformation to the marked empirical process R_n^1 of the least-squares residuals is quite similar to the marked empirical process $\widetilde{R}_n^1$ of the backwards recursive residuals. In fact, the only difference is that

$$\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} f^T(t_{ni}) \left((X_n^{(i+1)})^T X_n^{(i+1)} \right)^{-1} (X_n^{(i+1)})^T \hat{\varepsilon}_n^{(i+1)}$$

is the discrete time approximation for

$$\int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u) f^T(u) dF(u) \right)^{-1} \left(\int\limits_{(y,\infty)} f(z) dR_n^1(z) \right) dF(y).$$

Note that in the proof of Theorem 6.3.3 and the integrals above, we slightly abuse notation because the function f is only defined on $[a, b]$ and not on $(-\infty, \infty)$. However, this is not problematic here, since $supp(F) \subseteq [a, b]$ holds true.

The Khmaladze transformation thus is a continuous-time backwards recursive least-squares method that, as we know from Theorem 6.2.1, asymptotically leads to Brownian motion processes. In the context of time series data, this interpretation has already been made by Bai (2003) p. 544. Therein, it is also mentioned that in the context of time series, partial sums of recursive residuals lead to Brownian motion processes — a result that can be traced back to Sen (1982). According to the results of Evers (2022), this also applies to recursive residuals in linear regression models with triangular arrays of designs. Note that the interpretation of the Khmaladze transformation as a continuous-time recursive least-squares method is far from formally proved in Bai (2003). In fact, only an intuitive argument is given. Moreover, this intuitive argument is valid only up to a normalizing constant (as Bai mentions) and it only covers the case of time series data.

7 Appendix

7.1 Proofs for Chapter 2

Proof of Theorem 2.1.4
We divide the proof into two steps:

Step I : Let $n \in \mathbb{N}$, $i \in \{1, \dots, n\}$ and $x \in [-\infty, \infty]$ be given. Then

$$\begin{aligned}
&\mathbb{E}\left(\mathbb{1}_{[-\infty,\infty]}(X_i) \cdot (Y_i - m(X_i)) \,|\, X_i\right) \\
&\qquad = \mathbb{E}\left(\mathbb{1}_{[-\infty,\infty]}(X_i) \cdot Y_i \,|\, X_i\right) - \mathbb{E}\left(\mathbb{1}_{[-\infty,\infty]}(X_i) \cdot m(X_i) \,|\, X_i\right) \\
&\qquad = \mathbb{1}_{[-\infty,\infty]}(X_i) \cdot \mathbb{E}\left(Y_i \,|\, X_i\right) - \mathbb{E}\left(\mathbb{1}_{[-\infty,\infty]}(X_i) \cdot \mathbb{E}(Y_i \,|\, X_i) \,|\, X_i\right) \\
&\qquad = \mathbb{1}_{[-\infty,\infty]}(X_i) \cdot \mathbb{E}\left(Y_i \,|\, X_i\right) - \mathbb{1}_{[-\infty,\infty]}(X_i) \cdot \mathbb{E}(Y_i \,|\, X_i) \\
&\qquad = 0
\end{aligned}$$

holds true. Thus, the i-th summand of $R_n(x)$ is conditionally centred and therefore the whole process $R_n(\cdot)$ is centred. Furthermore, for $n \in \mathbb{N}$ and $s, t \in [-\infty, \infty]$, the covariance function of $R_n(\cdot)$ can be calculated in the following way:

$$\begin{aligned}
K(s,t) &:= Cov(R_n(s), R_n(t)) \\
&= \mathbb{E}(R_n(s) \cdot R_n(t)) - \mathbb{E}(R_n(s))\mathbb{E}(R_n(t)) \\
&= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i))\right) \cdot \left(\frac{1}{\sqrt{n}} \sum_{j=1}^{n} \mathbb{1}_{[-\infty,t]}(X_j) \cdot (Y_j - m(X_j))\right)\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i)) \cdot \mathbb{1}_{[-\infty,t]}(X_j) \cdot (Y_j - m(X_j))\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{1}_{[-\infty,\min\{s,t\}]}(X_i) \cdot (Y_i - m(X_i))^2\right] + \\
&\quad + \frac{1}{n} \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i))\right] \mathbb{E}\left[\mathbb{1}_{[-\infty,t]}(X_j) \cdot (Y_j - m(X_j))\right]
\end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}\left(\mathbb{1}_{[-\infty,\min\{s,t\}]}(X)\cdot(Y-m(X))^2\right)\\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{[-\infty,\min\{s,t\}]}(X)\cdot(Y-m(X))^2\,|\,X\right)\right]\\
&= \mathbb{E}\left[\mathbb{1}_{[-\infty,\min\{s,t\}]}(X)\cdot\mathbb{E}\left((Y-m(X))^2\,|\,X\right)\right]\\
&= \int\limits_{[-\infty,\min\{s,t\}]} \mathbb{E}((Y-\mathbb{E}(Y\,|\,X))^2|X=u)\,dF(u)\\
&= \int\limits_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y\,|\,X=u)\,dF(u).
\end{aligned}
$$

Note, that in the given situation (X,Y) is the random vector whose iid realisations form the random sample $(X_1,Y_1),\ldots,(X_n,Y_n)$.

Therefore, since for $n\in\mathbb{N}$ and all $x,x_1,x_2\in[-\infty,\infty]$ the summands $\mathbb{1}_{[-\infty,x]}(X_i)\cdot(Y_i-m(X_i))$ of $R_n(x)$ are conditionally centred and

$$
K(x_1,x_2)=Cov(R_n(x_1),R_n(x_2))=\int\limits_{[-\infty,\min\{x_1,x_2\}]}\mathbb{V}ar(Y\,|\,X=u)\,dF(u),
$$

the multivariate central limit theorem yields for all $-\infty\le x_1\le\cdots\le x_k\le\infty$ and $k\in\mathbb{N}$

$$
\begin{pmatrix}R_n(x_1)\\ \vdots\\ R_n(x_k)\end{pmatrix}\xrightarrow[n\to\infty]{\mathcal{D}} N_k\left(\begin{pmatrix}0\\ \vdots\\ 0\end{pmatrix},\begin{pmatrix}K(x_1,x_1) & \ldots & K(x_1,x_k)\\ \vdots & \ddots & \vdots\\ K(x_k,x_1) & \ldots & K(x_k,x_k)\end{pmatrix}\right).
$$

We thus define

$$
R_\infty(x):=N(0,K(x,x)),\quad x\in[-\infty,\infty], \tag{7.1}
$$

with the covariance function

$$
Cov(R_\infty(x_1),R_\infty(x_2)):=K(x_1,x_2),\quad x_1,x_2\in[-\infty,\infty]. \tag{7.2}
$$

Note that (7.1) and (7.2) define R_∞ on $D[-\infty,\infty]$, since Gaussian processes are uniquely determined by their mean function and covariance function. Note, furthermore, that the sequence of random processes $(R_n(\cdot))_{n\in\mathbb{N}}$ has at most the accumulation point $R_\infty(\cdot)$ and no other, since the distribution of a random process is uniquely determined by its fidis.

Step II : In order to prove the weak convergence of $R_n(\cdot)$, it remains to show the tightness of the process $R_n(\cdot)$. To achieve this, Stute (1997) proposes to use Theorem 15.6 from Billingsley (1968) and a quantile transformation (see for example Chapter 21 in Vaart (1998)). That is, for the random sample

$$
X_1,\ldots,X_n\overset{iid}{\sim}F
$$

there is a random sample

$$U_1, \ldots, U_n \overset{iid}{\sim} U(0,1),$$

such that

$$X_i \overset{\mathcal{D}}{=} F^-(U_i), \quad i = 1, \ldots, n,$$

holds true. Here, F^- denotes the quantile function associated to F (see Def. 2.1.8). For this uniformly distributed random sample we define the uniform empirical process

$$\overline{R}_n(u) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[0,u]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i)),$$

which has paths in $D[0,1]$ and note that in this situation

$$R_n(x) = \overline{R}_n(F(x)), \quad x \in [-\infty, \infty],$$

holds true. As Stute (1997) notes, it can be observed that

$$\mathbb{E}(Y_i \,|\, U = u) = (m \circ F^-)(u), \quad u \in [0,1], \tag{7.3}$$

so one can, without loss of generality, assume that F is the uniform distribution on $[0,1]$. With regard to (7.3), the following should be noted:

i) U should be replaced by U_i.

ii) Stute (1997) does not contain a proof of equation (7.3) and, contrary to the idea of conducting a quantile transformation, we get a similar result in Section 2.1.3 by means of a probability integral transformation.

Considering the above statements and our finding in Section 2.1.3, we can assume a regression on Y according to the design $U_1, \ldots, U_n \overset{iid}{\sim} U(0,1)$ instead of $X_1, \ldots, X_n \overset{iid}{\sim} F$, where the regression functions are now not m but $m \circ F^-$ in the case where only distributional properties of the model matter. Theorem 15.6 from Billingsley (1968) now states that $\overline{R}_n(\cdot)$ is tight if

$$\mathbb{E}\left[|\overline{R}_n(u_2) - \overline{R}_n(u)|^\gamma |\overline{R}_n(u) - \overline{R}_n(u_1)|^\gamma\right] \leq [H(u_2) - H(u_1)]^{2\alpha} \tag{7.4}$$

holds true for $0 \leq u_1 \leq u \leq u_2 \leq 1$ and $n \geq 1$ for some $\gamma \geq 0$, $\alpha > \frac{1}{2}$ and H a non-decreasing continuous function on $[0,1]$. We choose $\gamma = 2$ and $\alpha = 1$ and obtain

$$\mathbb{E}\left[\left(\overline{R}_n(u_2) - \overline{R}_n(u)\right)^2 \left(\overline{R}_n(u) - \overline{R}_n(u_1)\right)^2\right] = \frac{1}{n^2}\mathbb{E}\left(\left[\sum_{i=1}^{n} \alpha_i\right]^2 \left[\sum_{j=1}^{n} \beta_j\right]^2\right) \tag{7.5}$$

$$= \frac{1}{n^2}\left[n\mathbb{E}\left(\alpha_1^2\beta_1^2\right) + n(n-1)\mathbb{E}\left(\alpha_1^2\right)\mathbb{E}\left(\beta_1^2\right)\right] \tag{7.6}$$

$$\leq \mathbb{E}(\alpha_1^2)\mathbb{E}(\beta_1^2) \tag{7.7}$$

$$= \int\limits_{(u_1,u]} \mathbb{V}ar(Y_1|U_1 = v)\, d\lambda\!\!\lambda_{[0,1]}(v) \int\limits_{(u,u_2]} \mathbb{V}ar(Y_1|U_1 = v)\, d\lambda\!\!\lambda_{[0,1]}(v) \tag{7.8}$$

$$\leq \left[\int\limits_{(u_1,u_2]} \mathbb{V}ar(Y_1|U_1 = v)\, d\lambda\!\!\lambda_{[0,1]}(v) \right]^2$$

$$= \left(H(u_2) - H(u_1)\right)^2,$$

where

$$H(u) := \int\limits_{[0,u]} \mathbb{V}ar(Y|U = v)\, d\lambda\!\!\lambda_{[0,1]}(v)$$

is a non-decreasing continuous function. Therefore, condition (7.4) is satisfied and thus $\overline{R}_n(\cdot)$ is tight. Note that $\mathbb{V}ar(Y|U = v)$ need not be continuous, but since it is integrated with respect to the Lebesgue-measure, H is continuous. In the equation above the following relations were used:

(7.5) We define

$$\alpha_i := \mathbb{1}_{(u_1,u]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i))$$

and

$$\beta_i := \mathbb{1}_{(u,u_2]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i))$$

and therefore we have

$$\overline{R}_n(u) - \overline{R}_n(u_1) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(u_1,u]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \alpha_i,$$

as well as

$$\overline{R}_n(u_2) - \overline{R}_n(u) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \beta_i.$$

(7.6) (α_i, β_i), $i = 1, \ldots, n$, are square integrable bivariate random vectors that are independent and identically distributed with

$$\mathbb{E}(\alpha_i) = 0 = \mathbb{E}(\beta_i),$$

as can be verified similarly to the beginning of **Step I**. We first determine

$$\left[\sum_{i=1}^{n} \alpha_i\right]^2 \left[\sum_{j=1}^{n} \beta_j\right]^2 = \left[\sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \alpha_{i_1} \alpha_{i_2}\right] \left[\sum_{j_1=1}^{n} \sum_{j_2=1}^{n} \beta_{j_1} \beta_{j_2}\right]$$

$$= \left[\sum_{i=1}^{n} \alpha_i^2 + \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \alpha_{i_1}\alpha_{i_2}\right] \left[\sum_{j=1}^{n} \beta_j^2 + \sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \beta_{j_1}\beta_{j_2}\right]$$

$$= \left[\sum_{i=1}^{n} \alpha_i^2\right] \left[\sum_{j=1}^{n} \beta_j^2\right] + \left[\sum_{i=1}^{n} \alpha_i^2\right] \left[\sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \beta_{j_1}\beta_{j_2}\right] +$$

$$+ \left[\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \alpha_{i_1}\alpha_{i_2}\right] \left[\sum_{j=1}^{n} \beta_j^2\right] + \left[\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \alpha_{i_1}\alpha_{i_2}\right] \left[\sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \beta_{j_1}\beta_{j_2}\right]$$

and then calculate the expectation of each of the four summands individually:

First summand:

$$\mathbb{E}\left(\left[\sum_{i=1}^{n} \alpha_i^2\right] \left[\sum_{j=1}^{n} \beta_j^2\right]\right) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mathbb{E}\left(\alpha_i^2\beta_j^2\right)$$

$$= \sum_{i=1}^{n} \mathbb{E}\left(\alpha_i^2\beta_i^2\right) + \sum_{i=1}^{n} \sum_{\substack{j=1 \\ j \neq i}}^{n} \mathbb{E}\left(\alpha_i^2\beta_j^2\right)$$

$$= n\mathbb{E}\left(\alpha_1^2\beta_1^2\right) + n(n-1)\mathbb{E}\left(\alpha_1^2\right)\mathbb{E}\left(\beta_1^2\right).$$

Second summand:

$$\mathbb{E}\left(\left[\sum_{i=1}^{n} \alpha_i^2\right] \left[\sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \beta_{j_1}\beta_{j_2}\right]\right) = \sum_{i=1}^{n}\sum_{j_1=1}^{n}\sum_{\substack{j_2=1 \\ j_2 \neq j_2}}^{n} \mathbb{E}(\alpha_i^2\beta_{j_1}\beta_{j_2}) = 0.$$

Third summand:

$$\mathbb{E}\left(\left[\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \alpha_{i_1}\alpha_{i_2}\right] \left[\sum_{j=1}^{n} \beta_j^2\right]\right) = \sum_{j=1}^{n}\sum_{i_1=1}^{n}\sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \mathbb{E}\left(\beta_j^2\alpha_{i_1}\alpha_{i_2}\right) = 0.$$

Forth summand:

$$\mathbb{E}\left(\left[\sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \alpha_{i_1}\alpha_{i_2}\right] \left[\sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \beta_{j_1}\beta_{j_2}\right]\right) = \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{j_1=1}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \mathbb{E}(\alpha_{i_1}\alpha_{i_2}\beta_{j_1}\beta_{j_2})$$

$$= \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq i_1}}^{n} \mathbb{E}(\alpha_{i_1}\alpha_{i_2}\beta_{i_1}\beta_{j_2}) + \sum_{i_1=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{\substack{j_1=1 \\ j_1 \neq i_1}}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \mathbb{E}(\alpha_{i_1}\alpha_{i_2}\beta_{j_1}\beta_{j_2})$$

$$= \sum_{i=1}^{n} \sum_{\substack{i_2=1 \\ i_2 \neq i}}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq i}}^{n} \mathbb{E}(\underbrace{(\alpha_i \beta_i)}_{=0} \alpha_{i_2} \beta_{j_2}) + \sum_{i_1=1}^{n} \underbrace{\mathbb{E}(\alpha_{i_1})}_{=0} \sum_{\substack{i_2=1 \\ i_2 \neq i_1}}^{n} \sum_{\substack{j_1=1 \\ j_1 \neq i_1}}^{n} \sum_{\substack{j_2=1 \\ j_2 \neq j_1}}^{n} \mathbb{E}(\alpha_{i_2} \beta_{j_1} \beta_{j_2})$$

$$= 0.$$

(7.7) $\alpha_i^2 \cdot \beta_i^2 \equiv (\alpha_i \beta_i)(\alpha_i \beta_i) \equiv 0.$

(7.8)
$$\begin{aligned}
\mathbb{E}(\alpha_i^2) &= \mathbb{E}\left[\mathbb{1}_{(u_1,u]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i))^2\right] \\
&= \mathbb{E}\left[\mathbb{E}\left(\mathbb{1}_{(u_1,u]}(U_i) \cdot (Y_i - (m \circ F^-)(U_i))^2 | U_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{(u_1,u]}(U_i) \mathbb{E}\left((Y_i - (m \circ F^-)(U_i))^2 | U_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{(u_1,u]}(U_i) \cdot \mathbb{E}\left((Y_i - \mathbb{E}(Y_i|U_i))^2 | U_i\right)\right] \\
&= \mathbb{E}\left[\mathbb{1}_{(u_1,u]}(U_i) \cdot \mathbb{V}ar(Y_i|U_i)\right] \\
&= \int_{(u_1,u]} \mathbb{V}ar(Y_i|U_i = u)\, d\lambda\!\!\!\lambda_{[0,1]}(u)
\end{aligned}$$

and similarly it can be shown that

$$\mathbb{E}(\beta_i^2) = \int_{(u,u_2]} \mathbb{V}ar(Y_i|U_i = u)\, d\lambda\!\!\!\lambda_{[0,1]}(u).$$

Altogether, since the relation

$$R_n(x) = \overline{R}_n(F(x))$$

holds true, we know that the process $R_n(\cdot)$ is also tight, which completes the proof. □

Proof of Theorem 2.1.6

Proof of i) Since we operate under H_0, for $x \in [-\infty, \infty]$, we have

$$R_n^1(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (Y_i - m(X_i, \hat{\theta}_n))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (Y_i - m(X_i, \theta_0)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (m(X_i, \hat{\theta}_n) - m(X_i, \theta_0))$$

$$= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (Y_i - m(X_i)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (m(X_i, \hat{\theta}_n) - m(X_i, \theta_0))$$

$$= R_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (m(X_i, \hat{\theta}_n) - m(X_i, \theta_0)).$$

Now, because of part i) of Assumption 2, the multivariate mean value theorem is applicable, which means that for $\hat{\theta}_n, \theta_0 \in \mathbb{R}^p$ and every $i \in \{1, \dots, n\}$ there is a $\theta_{ni} \in \mathbb{R}^p$ on the line between $\hat{\theta}_n$ and θ_0 for which

$$m(X_i, \hat{\theta}_n) - m(X_i, \theta_0) = (\hat{\theta}_n - \theta_0)^T g(X_i, \theta_{ni})$$

holds true. We therefore can conclude that

$$\begin{aligned}
&\sup_{x \in [-\infty,\infty]} \left| R_n^1(x) - \left(R_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x, \theta_0)\, l(X_i, Y_i, \theta_0) \right) \right| \\
&= \sup_{x \in [-\infty,\infty]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x, \theta_0)\, l(X_i, Y_i, \theta_0) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (\hat{\theta}_n - \theta_0)^T g(X_i, \theta_{ni}) \right| \\
&= \sup_{x \in [-\infty,\infty]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x, \theta_0)\, l(X_i, Y_i, \theta_0) - \sqrt{n}\,(\hat{\theta}_n - \theta_0)^T G(x, \theta_0) + \sqrt{n}\,(\hat{\theta}_n - \theta_0)^T G(x, \theta_0) \right. \\
&\qquad \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (\hat{\theta}_n - \theta_0)^T \left[g(X_i, \theta_{ni}) - g(X_i, \theta_0) + g(X_i, \theta_0) \right] \right| \\
&\leq \underbrace{\sup_{x \in [-\infty,\infty]} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x, \theta_0)\, l(X_i, Y_i, \theta_0) - G^T(x, \theta_0) \sqrt{n} (\hat{\theta}_n - \theta_0) \right|}_{=:\mathrm{I}} \\
&\quad + \underbrace{\sup_{x \in [-\infty,\infty]} \left| \sqrt{n}\,(\hat{\theta}_n - \theta_0)^T \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot \left[g(X_i, \theta_{ni}) - g(X_i, \theta_0) \right] \right|}_{=:\mathrm{II}} \\
&\quad + \underbrace{\sup_{x \in [-\infty,\infty]} \left| \sqrt{n}\,(\hat{\theta}_n - \theta_0)^T \left[\left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g(X_i, \theta_0) \right) - G(x, \theta_0) \right] \right|}_{=:\mathrm{III}}.
\end{aligned}$$

For the proof of part i), it is therefore sufficient to verify that the summands I, II and III converge stochastically towards zero.

Convergence of summand I

According to Assumption 1, we have

$$\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0) - \sqrt{n}(\hat{\theta}_n - \theta_0) \right\| \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

Furthermore, according to Assumption 2, for $i = 1, \dots, p$,

$$|G_i(x, \theta_0)| = \left| \int\limits_{[-\infty, x]} g_i(t, \theta_0)\, dF(t) \right| \leq \int\limits_{[-\infty, \infty]} |g_i(t, \theta_0)|\, dF(t) \leq \int\limits_{[-\infty, \infty]} M(t)\, dF(t)$$

holds true and thus

$$\sup_{x \in [-\infty, \infty]} |G_i(x, \theta_0)| \leq \int\limits_{[-\infty, \infty]} M(t)\, dF(t) < \infty, \quad i = 1, \dots, p.$$

Hence, we can infer that

$$\begin{aligned} \mathrm{I} &= \sup_{x \in [-\infty, \infty]} \left| G^T(x, \theta_0) \left[\frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0) - \sqrt{n}(\hat{\theta}_n - \theta_0) \right] \right| \\ &\leq \left(\sup_{x \in [-\infty, \infty]} |G_1(x, \theta_0)|, \dots, \sup_{x \in [-\infty, \infty]} |G_p(x, \theta_0)| \right) \left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l(X_i, Y_i, \theta_0) - \sqrt{n}(\hat{\theta}_n - \theta_0) \right| \\ &\xrightarrow[n \to \infty]{\mathbb{P}} 0 \end{aligned}$$

is satisfied.

Convergence of summand II

According to Assumption 1 it holds true that

$$\sqrt{n}\, (\hat{\theta}_n - \theta_0) \xrightarrow[n \to \infty]{\mathcal{D}} N_p(0, Cov(l(X, Y, \theta_0))).$$

Thus, according to the theorem of Prohorov (see for example Satz 13.29 in Klenke (2013)), we can infer that $\left(\sqrt{n}(\hat{\theta}_n - \theta_0) \right)_{n \in \mathbb{N}}$ is tight. That is,

$$\left(\sqrt{n}(\hat{\theta}_n - \theta_0) \right)_{n \in \mathbb{N}} \text{ is bounded stochastically as a sequence of random vectors in } \mathbb{R}^p, \tag{7.9}$$

which means that, for every $\delta > 0$ there is an $n_0 \in \mathbb{N}$ and a $K \in \mathbb{R}$ such that $\mathbb{P}(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\| > K) < \delta$ holds true for all $n \geq n_0$. Note that, as a consequence, for every $\delta > 0$ and every $\varepsilon > 0$ there is an $n_0 \in \mathbb{N}$ and a $K \in \mathbb{R}$ such that

$$\mathbb{P}(\|\hat{\theta}_n - \theta_0\| \geq \varepsilon) \leq \mathbb{P}\left(\|\hat{\theta}_n - \theta_0\| \geq \frac{K}{\sqrt{n}} \right) = \mathbb{P}(\|\sqrt{n}(\hat{\theta}_n - \theta_0)\| \geq K) < \delta$$

holds true for all $n \geq n_0$. Hence,

$$\hat{\theta}_n \xrightarrow[n \to \infty]{\mathbb{P}} \theta_0 \tag{7.10}$$

holds true. Moving on, since

$$\text{II} \;\leq |\sqrt{n}\,(\hat{\theta}_n - \theta_0)^T| \begin{pmatrix} \sup\limits_{x\in[-\infty,\infty]} \left| \frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_1(X_i,\theta_{ni}) - g_1(X_i,\theta_0)] \right| \\ \vdots \\ \sup\limits_{x\in[-\infty,\infty]} \left| \frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_p(X_i,\theta_{ni}) - g_p(X_i,\theta_0)] \right| \end{pmatrix}$$

and (7.9) hold true, it is sufficient to prove that

$$\sup_{x\in[-\infty,\infty]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_j(X_i,\theta_{ni}) - g_j(X_i,\theta_0)] \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad j = 1,\ldots,p, \tag{7.11}$$

is satisfied. Let $j \in \{1,\ldots,p\}$ be fixed. Since θ_{ni} is on the line between $\hat{\theta}_n$ and θ_0, it holds true that

$$\begin{aligned} &\sup_{x\in[-\infty,\infty]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_j(X_i,\theta_{ni}) - g_j(X_i,\theta_0)] \right| \\ &\qquad \leq \frac{1}{n} \sum_{i=1}^{n} \sup_{\substack{||\theta-\theta_0|| \leq ||\hat{\theta}_n - \theta_0|| \\ \theta\in int(\Theta)}} |g_j(X_i,\theta) - g_j(X_i,\theta_0)| \\ &\qquad = \frac{1}{n} \sum_{i=1}^{n} h_{||\hat{\theta}_n-\theta_0||}(X_i) \end{aligned} \tag{7.12}$$

where

$$h_\delta : \mathbb{R} \longrightarrow \mathbb{R},\; t \longmapsto \sup_{\substack{||\theta-\theta_0|| \leq \delta \\ \theta\in int(\Theta)}} |g_j(t,\theta) - g_j(t,\theta_0)|.$$

Regarding $h_\delta(t)$, note the following two facts.

- $h_\delta(t) \xrightarrow[\delta\downarrow 0]{} 0$ for all $t \in \mathbb{R}$, since $g(t,\theta)$ is continuous in θ for all $t \in \mathbb{R}$ and thus, for all $\varepsilon > 0$ there is a $\delta_0 > 0$ such that $|g_j(t,\theta) - g_j(t,\theta_0)| < \frac{\varepsilon}{2}$ for all $||\theta - \theta_0|| < \delta_0$. Therefore, if $0 < \delta < \delta_0$,

$$h_\delta(t) \;= \sup_{\substack{||\theta-\theta_0|| \leq \delta \\ \theta\in int(\Theta)}} |g_j(t,\theta) - g_j(t,\theta_0)| \leq \frac{\varepsilon}{2} \leq \varepsilon.$$

- $0 \leq h_\delta(t) \leq 2 \cdot M(t) \in \mathcal{L}([-\infty,\infty], F)$ since, according to Assumption 2, $|g_j(t,\theta)| \leq M(t)$ for all $j = 1\ldots,p$, all $t \in \mathbb{R}$ and all $\theta \in int(\Theta)$.

We can therefore infer, that $\int\limits_{[-\infty,\infty]} h_\delta(t)\, dF(t) \xrightarrow[\delta\downarrow 0]{} 0$ holds true, due to the theorem of Lebesgue.

Now, let $\varepsilon > 0$ be given. Since $\int\limits_{[-\infty,\infty]} h_\delta(t)\, dF(t) \xrightarrow[\delta \downarrow 0]{} 0$, there is a $\delta_0 > 0$ such that

$$\int\limits_{[-\infty,\infty]} h_\delta(t)\, dF(t) < \frac{\varepsilon}{2} \quad \text{for all } 0 \leq \delta \leq \delta_0.$$

First, the strong law of large numbers yields that

$$\frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) \xrightarrow[n\to\infty]{f.s.} \mathbb{E}(h_{\delta_0}(X_1)) = \int\limits_{[-\infty,\infty]} h_{\delta_0}(t)\, dF(t) < \frac{\varepsilon}{2}.$$

Then, restricting to the event where $||\hat{\theta}_n - \theta_0|| \leq \delta_0$, we see by (7.12) that

$$\sup_{x\in[-\infty,\infty]} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_j(X_i,\theta_{ni}) - g_j(X_i,\theta_0)] \right|$$
$$\leq \frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) \xrightarrow[n\to\infty]{f.s.} \int\limits_{[-\infty,\infty]} h_{\delta_0}(t)\, dF(t) < \frac{\varepsilon}{2}$$

is satisfied. Since almost sure convergence implies convergence in probability and since $\int\limits_{[-\infty,\infty]} h_{\delta_0}(t)\, dF(t) < \frac{\varepsilon}{2}$, we have

$$\mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) < \varepsilon \right) \geq \mathbb{P}\left(\left| \frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) - \int\limits_{[-\infty,\infty]} h_{\delta_0}(t)\, dF(t) \right| < \frac{\varepsilon}{2} \right) \xrightarrow[n\to\infty]{} 1. \tag{7.13}$$

Thus,

$$\mathbb{P}\left(\sup_{x\in[-\infty,\infty]} \left| \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot [g_j(X_i,\theta_{ni}) - g_j(X_i,\theta_0)] \right| \geq \varepsilon \right)$$
$$\leq \mathbb{P}\left(\left\{ \frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) \geq \varepsilon \right\} \cup \left\{ ||\hat{\theta}_n - \theta_0|| > \delta_0 \right\} \right)$$
$$\leq \mathbb{P}\left(\left\{ \frac{1}{n}\sum_{i=1}^{n} h_{\delta_0}(X_i) \geq \varepsilon \right\} \right) + \mathbb{P}\left(\left\{ ||\hat{\theta}_n - \theta_0|| > \delta_0 \right\} \right) \xrightarrow[n\to\infty]{} 0$$

thanks to (7.13) and (7.10). Since $j \in \{1, \ldots, p\}$ was arbitrary, we thus have proven (7.11).

Convergence of summand III

Since

$$\text{III} \leq |\sqrt{n}\,(\hat{\theta}_n - \theta_0)^T| \begin{pmatrix} \sup\limits_{x\in[-\infty,\infty]} \left| \left(\frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g_1(X_i,\theta_0) \right) - G_1(x,\theta_0) \right| \\ \vdots \\ \sup\limits_{x\in[-\infty,\infty]} \left| \left(\frac{1}{n} \sum\limits_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g_p(X_i,\theta_0) \right) - G_p(x,\theta_0) \right| \end{pmatrix}$$

and (7.9) hold true, it is sufficient to prove that

$$\sup_{x\in[-\infty,\infty]} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g_j(X_i,\theta_0) \right) - G_j(x,\theta_0) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad j = 1,\ldots,p, \tag{7.14}$$

is satisfied. So let $j \in \{1,\ldots,p\}$ and

$$\mathcal{M} := \{ f_x : \mathbb{R} \longrightarrow \mathbb{R},\, t \longmapsto \mathbb{1}_{[-\infty,x]}(t) \cdot g_j(t,\theta_0) \,|\, x \in [-\infty,\infty] \}.$$

Then

$$\begin{aligned} &\sup_{x\in[-\infty,\infty]} \left| \left(\frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g_j(X_i,\theta_0) \right) - G_j(x,\theta_0) \right| \\ &\quad = \sup_{x\in[-\infty,\infty]} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot g_j(X_i,\theta_0) - \int\limits_{[-\infty,\infty]} \mathbb{1}_{[-\infty,x]}(t) \cdot g_j(t,\theta_0)\, dF(t) \right| \\ &\quad = \sup_{f\in\mathcal{M}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \int\limits_{[-\infty,\infty]} f(t)\, dF(t) \right| \end{aligned}$$

holds true. According to Lemma 1 in Dehardt (1971),

$$\sup_{f\in\mathcal{M}} \left| \frac{1}{n} \sum_{i=1}^{n} f(X_i) - \int\limits_{[-\infty,\infty]} f(t)\, dF(t) \right| \xrightarrow[n\to\infty]{a.s.} 0$$

holds true, if for every $\varepsilon > 0$ there is a finite class of functions $\mathcal{M}_\varepsilon$ such that for each $f \in \mathcal{M}$ there are f_1, f_2 in $\mathcal{M}_\varepsilon$ with $f_1 \leq f \leq f_2$ and $\int f_2\, dF - \int f_1\, dF < \varepsilon$. Furthermore, since almost sure convergence implies stochastic convergence, for (7.14) to be true, it is thus sufficient to find for any $\varepsilon > 0$ a suitable class of functions $\mathcal{M}_\varepsilon$ with these desired properties. The remaining proof of the convergence of summand III consists in the construction of $\mathcal{M}_\varepsilon$ and the proof of the required properties.

So let $\varepsilon > 0$ be given. Then the following assertions are satisfied:

- Due to the theorem of Lebesgue

$$\exists\, x \in \mathbb{R} \text{ such that } 2 \cdot \int\limits_{[-\infty,x]} |g_j(t,\theta_0)|\, dF(t) < \frac{\varepsilon}{3}, \tag{7.15}$$

since $2 \cdot \mathbb{1}_{[-\infty,-n]}(t) \cdot |g_j(t,\theta_0)| \xrightarrow[n\to\infty]{a.s.} 0$, $|2 \cdot \mathbb{1}_{[-\infty,-n]}(t) \cdot |g_j(t,\theta_0)|| \leq 2 \cdot M(t)$, and $M \in \mathcal{L}^1([-\infty,\infty], F)$.

- Due to the theorem of Lebesgue

$$\exists\, y \in \mathbb{R} \text{ such that } 2 \cdot \int_{(y,\infty]} |g_j(t,\theta_0)|\, dF(t) < \frac{\varepsilon}{3}, \tag{7.16}$$

since $2 \cdot \mathbb{1}_{(m,\infty]}(t) \cdot |g_j(t,\theta_0)| \xrightarrow[m\to\infty]{a.s.} 0$, $|2 \cdot \mathbb{1}_{(m,\infty]}(t) \cdot |g_j(t,\theta_0)|| \leq 2 \cdot M(t)$, and $M \in \mathcal{L}^1([-\infty,\infty], F)$. Without loss of generality, we assume that $y > x$.

- Let $D := \{x \in \mathbb{R} \,|\, F(x) - F(x-) > 0\}$, where $F(x-) := \lim\limits_{\substack{y_n \uparrow x \\ (y_n)_{n\in\mathbb{N}} \subseteq \mathbb{R}}} F(y_n)$. Then, since

$$\sum_{x\in D} M(x) \cdot (F(x) - F(x-)) \leq \int_{[-\infty,\infty]} M(t)\, dF(t) < \infty,$$

we can infer that

$$\exists\, D_\varepsilon \subseteq D \text{ with } |D_\varepsilon| < \infty \text{ and } 2 \cdot \sum_{x\in D\setminus D_\varepsilon} M(x) \cdot (F(x) - F(x-)) < \frac{\varepsilon}{3}. \tag{7.17}$$

- Let μ_F denote the measure associated to F. For $E \in \mathcal{B}$ define $\mu_{F,d}(E) := \mu_F(M \cap D)$ and $\mu_{F,c}(E) := \mu_F(E \cap (\mathbb{R} \setminus D))$. Now consider $\varphi_c(t) := \int\limits_{[-\infty,t)} M(t)\, d\mu_{F,c}(t)$, $t \in [-\infty,\infty]$. Since φ_c is continuous, monotonically increasing and the limits $\lim\limits_{t\to\pm\infty} \varphi_c(t)$ do exist, φ_c is uniformly continuous. Therefore,

$$\begin{gathered} \exists\, \delta > 0 \text{ such that for all } x < y \in \mathbb{R} \text{ with } |x-y| < \delta \\ 2 \cdot \int_{[x,y)} M(t)\, d\mu_{F,c}(t) = \varphi_c(y) - \varphi_c(x) < \frac{\varepsilon}{3}. \end{gathered} \tag{7.18}$$

Moving on, according to (7.17), there is an $\widetilde{n} \in \mathbb{N}$ and $\widetilde{x}_1 < \cdots < \widetilde{x}_{\widetilde{n}} \in \mathbb{R}$ such that $D_\varepsilon = \{\widetilde{x}_1, \ldots, \widetilde{x}_{\widetilde{n}}\}$. Next, choose x according to (7.15) and y according to (7.16). Then, for $\lceil\cdot\rceil : \mathbb{R} \to \{1,2,\ldots\}$, $x \mapsto \min\{k \in \{1,2,\ldots\} \,|\, k \geq x\}$ and given the choice of δ in (7.18), we define

$$t_i := x + i \cdot \frac{y-x}{\lceil \frac{y-x}{\delta} \rceil}, \quad i = 1, 2, \ldots, \left\lceil \frac{y-x}{\delta} \right\rceil - 1.$$

We thus have constructed the set $\left\{x, \widetilde{x}_1, \ldots, \widetilde{x}_{\widetilde{n}}, t_1, \ldots, t_{\lceil \frac{y-x}{\delta} \rceil - 1}, y\right\}$. Therefore, there is an $n \in \mathbb{N}$ and $x_1 < x_2 < \cdots < x_n$ such that

$$\{x_1, \ldots, x_n\} = \left\{x, \widetilde{x}_1, \ldots, \widetilde{x}_{\widetilde{n}}, t_1, \ldots, t_{\lceil \frac{y-x}{\delta} \rceil - 1}, y\right\}.$$

Note that by the construction of $\{x_1, \dots, x_n\}$,

$$2 \cdot \int\limits_{[-\infty, x_1]} |g_j(t, \theta_0)| \, dF(t) < \frac{\varepsilon}{3} \quad \text{and} \quad 2 \cdot \int\limits_{(x_n, \infty]} |g_j(t, \theta_0)| \, dF(t) < \frac{\varepsilon}{3} \tag{7.19}$$

as well as

$$2 \cdot \int\limits_{(x_i, x_{i+1})} M(t) \, d\mu_{F,d}(t) \leq 2 \cdot \sum_{x \in D \setminus D_\varepsilon} M(x) \cdot (F(x) - F(x-)) < \frac{\varepsilon}{3}, \quad i = 1, \dots, n-1, \tag{7.20}$$

and moreover $x_{i+1} - x_i < \delta$ and thus

$$2 \cdot \int\limits_{(x_i, x_{i+1})} M(t) \, d\mu_{F,c}(t) < \frac{\varepsilon}{3}, \quad i = 1, \dots, n-1, \tag{7.21}$$

hold true. We then proceed defining

$$\begin{aligned}
\mathcal{M}_\varepsilon := \big\{ & \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot |g_j(\cdot, \theta_0)| \,, \; \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) \,, \\
& \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot |g_j(\cdot, \theta_0)| + \mathbb{1}_{(x_1, x_n]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_n, \infty]}(\cdot) \cdot |g_j(\cdot, \theta_0)|, \\
& \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) + \mathbb{1}_{(x_1, x_n]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_n, \infty]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) \big\} \\
\cup \; \big\{ & \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot |g_j(\cdot, \theta_0)| + \mathbb{1}_{(x_1, x_i]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_i, x_{i+1})}(\cdot) \cdot |g_j(\cdot, \theta_0)|, \\
& \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) + \mathbb{1}_{(x_1, x_i]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_i, x_{i+1})}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) \mid i = 1, \dots, n-1 \big\}.
\end{aligned}$$

Now let $f_x(\cdot) = \mathbb{1}_{[-\infty, x]}(\cdot) \cdot g_j(\cdot, \theta_0) \in \mathcal{M}$ be given. We distinguish the following three cases.

Case 1 : $x < x_1$.
In this case, for $f_1(\cdot) := \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) \in \mathcal{M}_\varepsilon$ and $f_2(\cdot) := \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot |g_j(\cdot, \theta_0)| \in \mathcal{M}_\varepsilon$, $f_1(t) \leq f_x(t) \leq f_2(t)$, $t \in [-\infty, \infty]$, holds true and

$$\int\limits_{[-\infty, \infty]} f_2(t) \, dF(t) - \int\limits_{[-\infty, \infty]} f_1(t) \, dF(t) = 2 \cdot \int\limits_{[-\infty, x_1]} |g_j(t, \theta_0)| \, dF(t) < \frac{\varepsilon}{3} < \varepsilon$$

is satisfied according to (7.19).

Case 2 : $x \in [x_i, x_{i+1})$ for some $i \in \{1, \dots, n-1\}$.
In this case, for $f_1(\cdot) := \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) + \mathbb{1}_{(x_1, x_i]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_i, x_{i+1})}(\cdot) \cdot (-|g_j(\cdot, \theta_0)|) \in \mathcal{M}_\varepsilon$ and $f_2(\cdot) := \mathbb{1}_{[-\infty, x_1]}(\cdot) \cdot |g_j(\cdot, \theta_0)| + \mathbb{1}_{(x_1, x_i]}(\cdot) \cdot g_j(\cdot, \theta_0) + \mathbb{1}_{(x_i, x_{i+1})}(\cdot) \cdot |g_j(\cdot, \theta_0)| \in \mathcal{M}_\varepsilon$, $f_1(t) \leq f_x(t) \leq f_2(t)$, $t \in [-\infty, \infty]$, holds true and

$$\int\limits_{[-\infty, \infty]} f_2(t) \, dF(t) - \int\limits_{[-\infty, \infty]} f_1(t) \, dF(t) = 2 \cdot \int\limits_{[-\infty, x_1]} |g_j(t, \theta_0)| \, dF(t) + 2 \cdot \int\limits_{(x_i, x_{i+1})} |g_j(t, \theta_0)| \, dF(t)$$

$$= 2 \cdot \int\limits_{[-\infty,x_1]} |g_j(t,\theta_0)|\, dF(t) + 2 \cdot \int\limits_{(x_i,x_{i+1})} |g_j(t,\theta_0)|\, d\mu_{F,d}(t) + 2 \cdot \int\limits_{(x_i,x_{i+1})} |g_j(t,\theta_0)|\, d\mu_{F,c}(t)$$

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$$

is satisfied according to (7.19), (7.20) and (7.21).

Case 3 : $x \geq x_n$.
In this case, for $f_1(\cdot) := \mathbb{1}_{[-\infty,x_1]}(\cdot) \cdot (-|g_j(\cdot,\theta_0)|) + \mathbb{1}_{(x_1,x_n]}(\cdot) \cdot g_j(\cdot,\theta_0) + \mathbb{1}_{(x_n,\infty]}(\cdot) \cdot (-|g_j(\cdot,\theta_0)|) \in \mathcal{M}_\varepsilon$ and $f_2(\cdot) := \mathbb{1}_{[-\infty,x_1]}(\cdot) \cdot |g_j(\cdot,\theta_0)| + \mathbb{1}_{(x_1,x_n]}(\cdot) \cdot g_j(\cdot,\theta_0) + \mathbb{1}_{(x_n,\infty]}(\cdot) \cdot |g_j(\cdot,\theta_0)| \in \mathcal{M}_\varepsilon$, $f_1(t) \leq f_x(t) \leq f_2(t)$, $t \in [-\infty,\infty]$, holds true and

$$\int\limits_{[-\infty,\infty]} f_2(t)\, dF(t) - \int\limits_{[-\infty,\infty]} f_1(t)\, dF(t) = 2 \cdot \int\limits_{[-\infty,x_1]} |g_j(t,\theta_0)|\, dF(t) + 2 \cdot \int\limits_{(x_n,\infty]} |g_j(t,\theta_0)|\, dF(t)$$

$$< \frac{\varepsilon}{3} + \frac{\varepsilon}{3} < \varepsilon$$

is satisfied according to (7.19). This completes the proof.

Proof of ii) According i)

$$\begin{aligned} R_n^1(x) &= R_n(x) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x,\theta_0)\, l(X_i,Y_i,\theta_0) + o_{\mathbb{P}}(1) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{[-\infty,x]}(X_i) \cdot (Y_i - m(X_i)) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x,\theta_0)\, l(X_i,Y_i,\theta_0) + o_{\mathbb{P}}(1) \\ &= \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mathbb{1}_{[-\infty,x]}(X_i) \cdot (Y_i - m(X_i)) - G^T(x,\theta_0)\, l(X_i,Y_i,\theta_0)\right]}_{=:\widetilde{R}_n^1(x)} + o_{\mathbb{P}}(1). \end{aligned} \tag{7.22}$$

In the remainder of this proof, we show that

$$\widetilde{R}_n^1(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} R_\infty^1(\cdot) \quad \text{in } D[-\infty,\infty].$$

This would indeed complete the proof, since in this case,

$$R_n^1(x) = \widetilde{R}_n^1(\cdot) + o_{\mathbb{P}}(1) \xrightarrow[n\to\infty]{\mathcal{D}} R_\infty^1(\cdot) \quad \text{in } D[-\infty,\infty] \tag{7.23}$$

could be inferred. Equation (7.23) holds true, since for random elements X_n and X with $X_n \xrightarrow[n\to\infty]{\mathcal{D}} X$ and a random element Y_n with $Y_n \xrightarrow[n\to\infty]{\mathbb{P}} c$, where c is a constant, $X_n + Y_n \xrightarrow[n\to\infty]{\mathcal{D}} X + c$ holds. See, for example, Theorem 2.7 and Lemma 2.8 in Vaart (1998) for a prove of this statement.

Moving on, for $x \in [-\infty, \infty]$,

$$\mathbb{E}\left(\widetilde{R}_n^1(x)\right) = \mathbb{E}\left(R_n(x)\right) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G^T(x, \theta_0)\, \mathbb{E}\left(l(X_i, Y_i, \theta_0)\right) = 0 \tag{7.24}$$

holds true, hence $\widetilde{R}_n^1(\cdot)$ is a centred process. Because of Assumption 1 and the independence and identical distribution of $(X_1, Y_1), \ldots, (X_n, Y_n)$, we can calculate the covariance function $\widetilde{K}^1(\cdot, \cdot)$ of $\widetilde{R}_n^1$ under H_0 for $s, t \in [-\infty, \infty]$ as follows:

$$\begin{aligned}
&\widetilde{K}^1(s,t) = Cov(\widetilde{R}_n^1(s), \widetilde{R}_n^1(t)) \\
&= \mathbb{E}(\widetilde{R}_n^1(s)\widetilde{R}_n^1(t)) - \mathbb{E}(\widetilde{R}_n^1(s))\mathbb{E}(\widetilde{R}_n^1(t)) \\
&= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i)) - G^T(s, \theta_0)\, l(X_i, Y_i, \theta_0)\right] + o_{\mathbb{P}}(1)\right) \times \right. \\
&\qquad \left. \times \left(\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\mathbb{1}_{[-\infty,t]}(X_i) \cdot (Y_i - m(X_i)) - G^T(t, \theta_0)\, l(X_i, Y_i, \theta_0)\right] + o_{\mathbb{P}}(1)\right)\right] \\
&= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i)) \mathbb{1}_{[-\infty,t]}(X_j)(Y_j - m(X_j))\right] - \\
&\qquad - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X_i) \cdot (Y_i - m(X_i)) G^T(t, \theta_0)\, l(X_j, Y_j, \theta_0)\right] \\
&\qquad - \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[G^T(s, \theta_0)\, l(X_i, Y_i, \theta_0) \mathbb{1}_{[-\infty,t]}(X_j) \cdot (Y_j - m(X_j))\right] \\
&\qquad + \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} \mathbb{E}\left[\left(G^T(s, \theta_0)\, l(X_i, Y_i, \theta_0)\right)\left(G^T(t, \theta_0)\, l(X_j, Y_j, \theta_0)\right)\right] \\
&= \mathbb{E}\Big[\mathbb{1}_{[-\infty,\min\{s,t\}]}(X) \cdot (Y - m(X))^2\Big] - \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X) \cdot (Y - m(X))\, G^T(t, \theta_0)\, l(X, Y, \theta_0)\right] \\
&\quad - \mathbb{E}\left[G^T(s, \theta_0)\, l(X, Y, \theta_0) \mathbb{1}_{[-\infty,t]}(X)\, (Y - m(X))\right] \\
&\quad + \mathbb{E}\left[\left(G^T(s, \theta_0)\, l(X, Y, \theta_0)\right)\left(G^T(t, \theta_0)\, l(X, Y, \theta_0)\right)\right] \\
&= \int_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y|X = x)\, dF(x) - G^T(t, \theta_0)\mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X) \cdot (Y - m(X))\; l(X, Y, \theta_0)\right] - \\
&\quad - G^T(s, \theta_0)\mathbb{E}\big[\mathbb{1}_{[-\infty,t]}(X) \cdot (Y - m(X))\, l(X, Y, \theta_0)\big] + G^T(s, \theta_0)\, Cov(l(X, Y, \theta_0))\, G(t, \theta_0) \\
&= K^1(s,t).
\end{aligned}$$

We now proceed with the actual proof of the invariance principle. The multivariate central limit theorem applies to the fidis of $\widetilde{R}_n^1(\cdot)$ and states that for $k \in \mathbb{N}$ and $-\infty \leq x_1 \leq \cdots \leq$

$x_k \leq \infty$

$$\begin{pmatrix} \widetilde{R}_n^1(x_1) \\ \vdots \\ \widetilde{R}_n^1(x_k) \end{pmatrix} \xrightarrow[n\to\infty]{\mathcal{D}} N_k \left(\begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix}, \begin{pmatrix} K^1(x_1,x_1) & \dots & K^1(x_1,x_k) \\ \vdots & \ddots & \vdots \\ K^1(x_k,x_1) & \dots & K^1(x_k,x_k) \end{pmatrix} \right)$$

holds true. Therefore, we define

$$R_\infty^1(x) := N(0, K^1(x,x)), \quad x \in [-\infty, \infty],$$

with the covariance function

$$Cov(R_\infty^1(x_1), R_\infty^1(x_2)) := K^1(x_1, x_2), \quad x \in [-\infty, \infty].$$

In order to complete the proof, it remains to prove the tightness of $\widetilde{R}_n^1(\cdot)$. To this regard, as proposed in Stute (1997), we shall use Theorem 15.6 from Billingsley (1968) again, but before this, we perform the standard quantile transformation as it has been done in the proof of Theorem 2.1.4. To be more specific, to the random sample $X_1, \dots, X_n \overset{iid}{\sim} F$ a random sample $U_1, \dots, U_n \overset{iid}{\sim} U(0,1)$ with $X_i \overset{\mathcal{D}}{=} F^-(U_i),\ i = 1, \dots, n$, can be associated. Then, for $u \in (0,1)$ we define

$$\overline{g}(u,\theta) := \frac{\partial \overline{m}(u,\theta)}{\partial \theta} \in \mathbb{R}^p,$$

where

$$\overline{m}(u,\theta) := m(F^-(u),\theta).$$

Note that because g is $F-$integrable, $\overline{g}$ is $\lambda\!\!\!\lambda_{[0,1]}-$integrable and therefore

$$\forall\, u \in [0,1] \quad \forall\, i \in \{1,\dots,p\}: \quad \overline{G}_i(u,\theta) := \int\limits_{[0,u]} \overline{g}_i(v,\theta)\, d\lambda\!\!\!\lambda_{[0,1]}(v)$$

exists and is continuous. For the uniform random sample $U_1, \dots, U_n$ we define the uniform marked empirical process

$$\overline{R}_n^1(u) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{1}_{[0,u]}(U_i) \cdot (Y_i - m(F^-(U_i), \hat{\theta}_n))$$

that has paths in $D[0,1]$. Due to part i), we get

$$\overline{R}_n^1(u) = \overline{R}_n(u) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \overline{G}^T(u,\theta_0)\, l(F^-(U_i), Y_i, \theta_0) + o_{\mathbb{P}}(1) \ \text{ for } n \to \infty, \text{ uniformly in } u \in [0,1],$$

and note that in this situation

$$\widetilde{R}_n^1(x) = \overline{R}_n^1(F(x)), \quad x \in [-\infty, \infty], \tag{7.25}$$

holds true. Thus, again, we can assume a regression of Y on the design $U_1, \ldots, U_n \overset{iid}{\sim} U(0,1)$ instead of $X_1, \ldots, X_n \overset{iid}{\sim} F$, where the regression function is now not

$$m : [-\infty, \infty] \times \Theta \to \mathbb{R}, \;\; (x, \theta) \mapsto m(x, \theta),$$

but

$$\overline{m} : (0,1) \times \Theta \to \mathbb{R}, \;\; (u, \theta) \mapsto \overline{m}(u, \theta) := m(F^-(u), \theta),$$

since for proving Part ii) of Theorem 2.1.6, only distributional properties matter. Theorem 15.6 from Billingsley (1968) now states that $\overline{R}_n^1(\cdot)$ is tight if

$$\mathbb{E}\left[|\overline{R}_n^1(u_2) - \overline{R}_n^1(u)|^\gamma |\overline{R}_n^1(u) - \overline{R}_n^1(u_1)|^\gamma\right] \leq [H(u_2) - H(u_1)]^{2\alpha} \tag{7.26}$$

holds true for $0 \leq u_1 \leq u \leq u_2 \leq 1$, $n \geq 1$ for some $\gamma \geq 0$, $\alpha > \frac{1}{2}$ and H a non-decreasing continuous function on $[0,1]$. Because of

$$\begin{aligned}
\overline{R}_n^1(u_2) - \overline{R}_n^1(u) &= \overline{R}_n(u_2) - \frac{1}{\sqrt{n}} \sum_{i=1}^n \overline{G}^T(u_2, \theta_0)\, l(F^-(U_i), Y_i, \theta_0) + o_{\mathbb{P}}(1) \\
&\quad - \overline{R}_n(u) + \frac{1}{\sqrt{n}} \sum_{i=1}^n \overline{G}^T(u, \theta_0)\, l(F^-(U_i), Y_i, \theta_0) + o_{\mathbb{P}}(1) \\
&= \overline{R}_n(u_2) - \overline{R}_n(u) - \left[\overline{G}^T(u_2, \theta_0) - \overline{G}^T(u, \theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n l(F^-(U_i), Y_i, \theta_0) + o_{\mathbb{P}}(1)
\end{aligned}$$

and the fact that (similar to our reasoning regarding equation (7.23)) we can drop the $o_{\mathbb{P}}(1)$ term, we can conclude that for $\gamma = 1$,

$$\begin{aligned}
&\mathbb{E}\left[|\overline{R}_n^1(u_2) - \overline{R}_n^1(u)|^\gamma |\overline{R}_n^1(u) - \overline{R}_n^1(u_1)|^\gamma\right] = \mathbb{E}\left[|\overline{R}_n^1(u_2) - \overline{R}_n^1(u)||\overline{R}_n^1(u) - \overline{R}_n^1(u_1)|\right] \\
&\leq \mathbb{E}\left[|\overline{R}_n(u_2) - \overline{R}_n(u)||\overline{R}_n(u) - \overline{R}_n(u_1)|\right] + \qquad (7.27) \\
&\quad + \mathbb{E}\left[|\overline{R}_n(u_2) - \overline{R}_n(u)| \left|\left[\overline{G}^T(u, \theta_0) - \overline{G}^T(u_1, \theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n l(F^-(U_i), Y_i, \theta_0)\right|\right] + \qquad (7.28) \\
&\quad + \mathbb{E}\left[\left|\left[\overline{G}^T(u_2, \theta_0) - \overline{G}^T(u, \theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n l(F^-(U_i), Y_i, \theta_0)\right| |\overline{R}_n(u) - \overline{R}_n(u_1)|^\gamma\right] + \qquad (7.29) \\
&\quad + \mathbb{E}\left[\left|\left[\overline{G}^T(u_2, \theta_0) - \overline{G}^T(u, \theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n l(F^-(U_i), Y_i, \theta_0)\right| \times\right. \qquad (7.30) \\
&\qquad \left. \times \left|\left[\overline{G}^T(u, \theta_0) - \overline{G}^T(u_1, \theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^n l(F^-(U_i), Y_i, \theta_0)\right|\right]
\end{aligned}$$

holds true.

In order to prove tightness of $\overline{R}_n^1(\cdot)$, it therefore suffices to prove tightness of the four summands from above independently with the help of Theorem 15.6 from Billingsley (1968). Regarding the first summand (7.27), the first factor in the expectation (7.28) and the last factor in the expectation (7.29), we have already shown that $\overline{R}_n(\cdot)$ is tight in Step II of the proof of Theorem 2.1.4 and therefore the fluctuation of the differences $\overline{R}_n(k) - \overline{R}_n(l),\ 0 \leq k \leq l \leq 1$, is under control. Regarding the remaining factors in the three summands (7.28), (7.29) and (7.30), we note that their fluctuation is determined by the fluctuation of the deterministic function $\overline{G}$, whereas the only stochastic part is in the function l, whose expectation is controllable due to Assumption 1. To be more specific, we will analyse the fourth summand (7.30) in detail. The calculations for the second and third summands are similar. For the fourth summand (7.30), we consider at one of the p summands of the scalar product. So let $1 \leq j, k \leq p$. We get

$$
\begin{aligned}
&\mathbb{E}\left[\left|\left[\,\overline{G}_j(u_2,\theta_0) - \overline{G}_j(u,\theta_0)\right]\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_j(F^-(U_i),Y_i,\theta_0)\right| \times \right. \\
&\qquad \left. \times \left|\left[\,\overline{G}_k(u,\theta_0) - \overline{G}_k(u_1,\theta_0)\right]\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_k(F^-(U_i),Y_i,\theta_0)\right|\right] \\
&= \left|\int_{(u,u_2]} \overline{g}_j(v,\theta_0)\, d\lambda\!\!\!\lambda_{[0,1]}(v)\right| \cdot \left|\int_{(u_1,u]} \overline{g}_k(v,\theta_0)\, d\lambda\!\!\!\lambda_{[0,1]}(v)\right| \times \\
&\qquad \times\ \mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_j(F^-(U_i),Y_i,\theta_0)\right| \cdot \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_k(F^-(U_i),Y_i,\theta_0)\right|\right] \\
&\leq \int_{(u,u_2]} |\overline{g}_j(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \cdot \int_{(u_1,u]} |\overline{g}_k(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \times \\
&\qquad \times\ \mathbb{E}\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_j(F^-(U_i),Y_i,\theta_0)\right| \cdot \left|\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_k(F^-(U_i),Y_i,\theta_0)\right|\right] \\
&\overset{i)}{\leq} \int_{(u,u_2]} |\overline{g}_j(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \int_{(u_1,u]} |\overline{g}_k(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \times \\
&\qquad \times\ \sqrt{\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_j(F^-(U_i),Y_i,\theta_0)\right]^2} \cdot \sqrt{\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} l_k(F^-(U_i),Y_i,\theta_0)\right]^2} \\
&\overset{ii)}{=} \int_{(u,u_2]} |\overline{g}_j(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \cdot \int_{(u_1,u]} |\overline{g}_k(v,\theta_0)|\, d\lambda\!\!\!\lambda_{[0,1]}(v) \times \\
&\qquad \times\ \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[l_j(F^-(U_i),Y_i,\theta_0)\right]^2} \cdot \sqrt{\frac{1}{n}\sum_{i=1}^{n} \mathbb{E}\left[l_k(F^-(U_i),Y_i,\theta_0)\right]^2}
\end{aligned}
$$

$$
\begin{aligned}
&\overset{iii)}{=} \int\limits_{(u,u_2]} |\overline{g}_j(v,\theta_0)|\, d\lambda\!\!\!^{-}_{[0,1]}(v) \cdot \int\limits_{(u_1,u]} |\overline{g}_k(v,\theta_0)|\, d\lambda\!\!\!^{-}_{[0,1]}(v) \times \\
&\qquad \times \sqrt{\mathbb{E}\left[l_j(F^-(U_i),Y_i,\theta_0)\right]^2} \cdot \sqrt{\mathbb{E}\left[l_k(F^-(U_i),Y_i,\theta_0)\right]^2} \\
&= \int\limits_{(u,u_2]} |\overline{g}_j(v,\theta_0)|\, d\lambda\!\!\!^{-}_{[0,1]}(v) \cdot \int\limits_{(u_1,u])} |\overline{g}_k(v,\theta_0)|\, d\lambda\!\!\!^{-}_{[0,1]}(v) \times \\
&\qquad \times \sqrt{Cov(l(X,Y,\theta_0))_{jj} \cdot Cov(l(X,Y,\theta_0))_{kk}} \\
&\overset{iv)}{\leq} \left[\int\limits_{(u_1,u_2]} M\left(F^-(v)\right)\, d\lambda\!\!\!^{-}_{[0,1]}(v)\right]^2 \cdot \sqrt{Cov(l(X,Y,\theta_0))_{jj} \cdot Cov(l(X,Y,\theta_0))_{kk}} \\
&\overset{v)}{=} [J(u_2) - J(u_1)]^2 \cdot \sqrt{Cov(l(X,Y,\theta_0))_{jj} \cdot Cov(l(X,Y,\theta_0))_{kk}}.
\end{aligned}
$$

In the sequence of equations above

i) is true due to the Cauchy-Schwarz inequality,

ii) is true due to $\mathbb{E}(l_j(F^-(U_i),Y_i,\theta_0)) = 0$ and the independence of $(U_1,Y_1),\ldots,(U_n,Y_n)$,

iii) is true because $(U_1,Y_1),\ldots,(U_n,Y_n)$ are identically distributed,

iv) can be deduced from Assumption 2 part ii) and

v) is true for

$$J(u) := \int\limits_{[0,u]} M(F^-(v))\, d\lambda\!\!\!^{-}_{[0,1]}(v),$$

which is a non-decreasing and continuous function on $[0,1]$.

As a consequence of the boundedness of the fluctuation of the summands

$$\left| \left[\, \overline{G}_j(u_2,\theta_0) - \overline{G}_j(u,\theta_0)\right] \frac{1}{\sqrt{n}} \sum_{i=1}^{n} l_j(F^-(U_i),Y_i,\theta_0) \right|, \quad 1 \leq j \leq p,$$

the criterion of Theorem 15.6 from Billingsley (1968) applies and therefore the boundedness of their sum, i.e., the scalar product

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left[\overline{G}^T(u_2,\theta_0) - \overline{G}^T(u,\theta_0)\right] l(F^-(U_i),Y_i,\theta_0)$$

(see (7.30)) follows. Altogether, we have proven that $\overline{R}_n^1(\cdot)$ is tight and since the relation

$$\widetilde{R}_n^1(x) = \overline{R}_n^1(F(x))$$

holds true, we know that the process $\widetilde{R}_n^1(\cdot)$ is tight, too, which completes the proof. $\square$

Proof of Theorem 2.1.7

The proof is divided in two parts.

Part 1

We are in the situation of a univariate linear regression model

- with the design variable X, which has the distribution function F,
- where the regression errors are assumed to be homoscedastic with variance 1,
- where parameter estimation is performed using the LSE.

In this situation, we are going to show that all conditions under which Theorem 2.1.6 provides a limit process for the marked empirical process of the residuals are fulfilled. For that, we have to prove that

1) $\mathbb{E}(Y^2) < \infty$,
2) Assumption 1 is satisfied,
3) Assumption 2 is satisfied.

As to 1:

$$\begin{aligned}
\mathbb{E}(Y^2) &= \mathbb{E}((m(X)+\varepsilon)^2) \\
&= \mathbb{E}(m(X)^2) + 2\mathbb{E}(m(X))\mathbb{E}(\varepsilon) + \mathbb{E}(\varepsilon^2) \\
&= \mathbb{E}(m(X)^2) + 2\mathbb{E}(m(X))\mathbb{E}(\mathbb{E}(\varepsilon|X=x)) + \mathbb{E}(\mathbb{E}(\varepsilon^2|X=x)) \\
&= \mathbb{E}(m(X)^2) + \mathbb{E}(\mathbb{V}ar(\varepsilon|X=x)) \\
&= \mathbb{E}((f^T(X)\theta)^2) + 1 \\
&= \int\limits_{[-\infty,\infty]} \left(\sum_{i=1}^{p} f_i(x)\theta_i\right)^2 dF(x) + 1 \\
&= \sum_{i=1}^{p}\sum_{j=1}^{p} \theta_i\theta_j \int\limits_{[-\infty,\infty]} f_i(x)f_j(x)\, dF(x) + 1 \\
&< \infty,
\end{aligned}$$

as it is assumed that the matrix J exists. Thus,

$$\int\limits_{[-\infty,\infty]} f_i(x)f_j(x)\, dF(x) < \infty, \quad i,j \in \{1,\ldots,p\}.$$

As to 2: For $n \in \mathbb{N}$, we denote by

$$Y(n) := (Y_1, \ldots, Y_n)^T$$

the outcome vector to the random sample $X_1, \ldots, X_n \overset{iid}{\sim} X$ and by

$$X(n) := \begin{pmatrix} f_1(X_1) & \cdots & f_p(X_1) \\ \vdots & \ddots & \vdots \\ f_1(X_n) & \cdots & f_p(X_n) \end{pmatrix} \in \mathbb{R}^{n \times p}$$

the associated random design matrix. We assume that there exists $n_0 \in \mathbb{N}$ such that $rank(X(n)) = p$ for all $n \geq n_0$ and in the sequel, we only consider the case $n \geq n_0$. Hence, the LSE $\hat{\theta}_n$ equals

$$\hat{\theta}_n = (X^T(n)X(n))^{-1} X^T(n)Y(n).$$

This assumption is no restriction as in the theorem $rank(J) = p$ is assumed, which is equivalent to the linear independence of $f_1, \ldots, f_p$ in $\mathcal{L}^2([-\infty, \infty], F)$ and thus such a n_0 exists (a.s.) and, since we consider an asymptotic result, for $n \to \infty$, we can assume that $n \geq n_0$. Note, that n_0 is the realisation of a random variable. For

$$J_n := \frac{1}{n} X^T(n)X(n) = \frac{1}{n} \begin{pmatrix} f_1(X_1) & \cdots & f_1(X_n) \\ \vdots & \ddots & \vdots \\ f_p(X_1) & \cdots & f_p(X_n) \end{pmatrix} \begin{pmatrix} f_1(X_1) & \cdots & f_p(X_1) \\ \vdots & \ddots & \vdots \\ f_1(X_n) & \cdots & f_p(X_n) \end{pmatrix}$$

$$= \left(\frac{1}{n} \sum_{i=1}^{n} f_l(X_i) f_m(X_i) \right)_{1 \leq l,m \leq p} \in \mathbb{R}^{p \times p},$$

we again only consider $n \geq n_0$ such that

$$rank(J_n) = p. \tag{7.31}$$

Then, under $H_0 : m(\cdot) = f^T(\cdot)\theta_0$, the following equations holds true:

$$J_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i = J_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)(Y_i - m(X_i))$$

$$= \left(\frac{1}{n} X^T(n)X(n) \right)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)(Y_i - f^T(X_i)\theta_0)$$

$$= \sqrt{n}(X^T(n)X(n))^{-1} \sum_{i=1}^{n} f(X_i)\, Y_i - \sqrt{n}(X^T(n)X(n))^{-1} \sum_{i=1}^{n} f(X_i)\, f^T(X_i)\theta_0$$

$$= \sqrt{n}(X^T(n)X(n))^{-1} X^T(n)Y(n) - \sqrt{n}(X^T(n)X(n))^{-1} X^T(n)X(n)\theta_0$$

$$= \sqrt{n} \left(\hat{\theta}_n - \theta_0 \right).$$

Hence, under H_0, we have

$$\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) = J_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\,\varepsilon_i. \tag{7.32}$$

Because of

$$X_1, \ldots, X_n \overset{\text{iid}}{\sim} X \sim F,$$

$$\overline{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow[n\to\infty]{a.s.} \mathbb{E}(X) = \int X\, dP = \int\limits_{[-\infty,\infty]} x\, dF(x)$$

results from the strong law of large numbers. By applying the strong low of large numbers again, for $1 \leq l, m \leq p$,

$$(J_n)_{lm} = \frac{1}{n}\sum_{i=1}^{n} f_l(X_i) f_m(X_i) \xrightarrow[n\to\infty]{a.s.} \mathbb{E}(f_l(X) f_m(X)) = \int\limits_{[-\infty,\infty]} f_l(x) f_m(x)\, dF(x) = J_{lm}$$

can be concluded and therefore, we finally get

$$J_n \xrightarrow[n\to\infty]{a.s.} J.$$

Note that it was assumed in the present theorem that $rank(J) = p$, hence the result from above justifies our assumption in (7.31).
Under the assumption that J is regular and for n large enough such that J_n is regular and because of the continuity of matrix inversion

$$\cdot^{-1} : \mathrm{GL}(n, \mathbb{R}) \to \mathrm{GL}(n, \mathbb{R}),\ A \mapsto A^{-1}$$

we can conclude

$$J_n^{-1} \xrightarrow[n\to\infty]{a.s.} J^{-1}.$$

By further taking (7.32) into account, we get

$$\begin{aligned}
\sqrt{n}\left(\hat{\theta}_n - \theta_0\right) &= J_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i \\
&= J_n^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i + J^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i - J^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i \\
&= J^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i + \frac{1}{\sqrt{n}} \left(J_n^{-1} - J^{-1}\right) \sum_{i=1}^{n} f(X_i)\varepsilon_i \\
&= J^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} f(X_i)\varepsilon_i + o_{\mathbb{P}}(1),
\end{aligned}$$

since $J_n^{-1} - J^{-1}$ converges to zero a.s. and $\frac{1}{\sqrt{n}} \sum\limits_{i=1}^{n} f(X_i)\varepsilon_i$ converges in distribution and is hence a.s. bounded.

In order to satisfy Assumption 1, the following equality must hold

$$J^{-1}\,\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_i)\varepsilon_i + o_{\mathbb{P}}(1) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} l(X_i, Y_i, \theta_0) + o_{\mathbb{P}}(1),$$

where $\mathbb{E}(l(X, Y, \theta_0)) = 0$ and $Cov(l(X, Y, \theta_0))$ exists. So we simply define

$$l(X_i, Y_i, \theta_0) \coloneqq J^{-1} f(X_i)\varepsilon_i = J^{-1} f(X_i)(Y_i - m(X_i, \theta_0)). \tag{7.33}$$

As a consequence, the first part of Assumption 1 is satisfied. It remains to prove properties i) and ii) of Assumption 1.

i)
$$\begin{aligned}
\mathbb{E}(l(X, Y, \theta_0)) &= \mathbb{E}\left(J^{-1} f(X)(Y - m(X, \theta_0)\right) \\
&= J^{-1}\left(\mathbb{E}(f(X)Y) - \mathbb{E}(f(X)m(X, \theta_0))\right) \\
&= J^{-1}\left(\mathbb{E}(f(X)Y) - \mathbb{E}(f(X)\mathbb{E}(Y|X))\right) \\
&= J^{-1}\left(\mathbb{E}(f(X)Y) - \mathbb{E}(\mathbb{E}(f(X)Y|X))\right) \\
&= J^{-1}\left(\mathbb{E}(f(X)Y) - \mathbb{E}(f(X)Y)\right) \\
&= 0.
\end{aligned}$$

ii)
$$\begin{aligned}
Cov(l(X, Y, \theta_0)) &= \mathbb{E}(l(X, Y, \theta_0)\, l^T(X, Y, \theta_0)) \\
&= \mathbb{E}\left(J^{-1} f(X)(Y - m(X, \theta_0))\, \left[J^{-1} f(X)(Y - m(X, \theta_0))\right]^T\right) \\
&= \mathbb{E}\left(J^{-1} f(X)(Y - m(X, \theta_0))\,(Y - m(X, \theta_0)) f^T(X) J^{-1}\right) \\
&= \int\limits_{[-\infty,\infty]} J^{-1} f(x)\mathbb{E}\left[(Y - m(X, \theta_0))^2 \middle| X = x\right] f^T(x) J^{-1}\, dF(x) \\
&\overset{\text{under}}{\underset{H_0}{=}} \int\limits_{[-\infty,\infty]} J^{-1} f(x)\mathbb{E}\left[(m(X, \theta_0) + \varepsilon - m(X, \theta_0))^2 | X = x\right] f^T(x) J^{-1}\, dF(x) \\
&= \int\limits_{[-\infty,\infty]} J^{-1} f(x)\mathbb{E}\left[\varepsilon^2 \middle| X = x\right] f^T(x) J^{-1}\, dF(x) \\
&= \int\limits_{[-\infty,\infty]} J^{-1} f(x)\mathbb{E}\left[\left(\varepsilon - \underbrace{\mathbb{E}(\varepsilon|X = x)}_{=0}\right)^2 \middle| X = x\right] f^T(x) J^{-1}\, dF(x) \\
&= \int\limits_{[-\infty,\infty]} J^{-1} f(x)\mathbb{V}ar(\varepsilon|X = x) f^T(x) J^{-1}\, dF(x) \\
&= \int\limits_{[-\infty,\infty]} J^{-1} f(x)\sigma^2(x) f^T(x) J^{-1}\, dF(x)
\end{aligned}$$

$$= J^{-1} \int\limits_{[-\infty,\infty]} f(x) f^T(x)\, dF(x)\; J^{-1}$$

$$= J^{-1}\, J\, J^{-1}$$

$$= J^{-1} \tag{7.34}$$

and thus, $Cov(l(X, Y, \theta_0))$ exists.

As to 3:
As we are in a linear regression scenario, under H_0, we have

$$m(x) = m(x, \theta) = f^T(x)\, \theta$$

and therefore

$$g(x, \theta) = \frac{\partial m(x, \theta)}{\partial \theta} = \frac{\partial\, f^T(x)\, \theta}{\partial \theta} = f(x) = (f_1(x), \ldots, f_p(x))^T$$

holds true. Thus, $m(x, \theta)$ is continuously differentiable in θ for all $\theta \in \mathbb{R}^p$ and all $x \in \mathbb{R}$. Hence,

$$G_i(x, \theta) = \int\limits_{[-\infty,x]} g_i(u, \theta)\, dF(u) = \int\limits_{[-\infty,x]} f_i(u)\, dF(u), \quad 1 \le i \le p,$$

exists, since $f_i, \ldots, f_p$ are supposed to be (square) integrable with respect to F. Thus, ii) of Assumption 2 holds. Note that in a general regression model with random design, g (and therefore also G) may depend on θ (see Section 2.1.1), whereas in the present situation of a *linear* regression model with random design, g (and therefore also G) do not depend on θ any more.

Part 2

Due to Theorem 2.1.6, we already know the mean function and the covariance function of the process $R^1_\infty(\cdot)$ in the general situation of a regression model with random design:

$$m(s) \coloneqq \mathbb{E}(R^1_\infty(s)) = 0, \quad s \in [-\infty, \infty],$$

and

$$K^1(s,t) = Cov(R^1_\infty(s), R^1_\infty(t)) \tag{7.35}$$

$$= \int\limits_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y|X = u)\, dF(u) + G^T(s, \theta_0)\, Cov(l(X, Y, \theta_0))\, G(t, \theta_0)$$

$$- G^T(s, \theta_0)\, \mathbb{E}\left[\mathbb{1}_{[-\infty,t]}(X) \cdot (Y - m(X, \theta_0))\; l(X, Y, \theta_0)\right] \tag{7.36}$$

$$- G^T(t, \theta_0)\, \mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X) \cdot (Y - m(X, \theta_0))\; l(X, Y, \theta_0)\right], \quad s, t \in [-\infty, \infty].$$

In the present situation of Theorem 2.1.7, we obtain for the first summand

$$\begin{aligned}\int\limits_{[-\infty,\min\{s,t\}]} \mathbb{V}ar(Y|\,X=u)\,dF(u) &= \int\limits_{[-\infty,\min\{s,t\}]} \sigma^2\,dF(u)\\ &= \int\limits_{[-\infty,\min\{s,t\}]} 1\,dF(u)\\ &= F(\min\{s,t\})\,.\end{aligned}$$

Because of the representation in equation (7.34), we have $Cov(l(X,Y,\theta_0)) = J^{-1}$. Thus, for the second summand, we get

$$\begin{aligned}G^T(s,\theta_0)\,Cov(l(X,Y,\theta_0))\,G(t,\theta_0) &= G^T(s,\theta_0)\,J^{-1}\,G(t,\theta_0)\\ &\overset{H_0}{=} G^T(s)\,J^{-1}\,G(t)\,.\end{aligned}$$

For the third summand, we get with (7.33)

$$\begin{aligned}&G^T(s,\theta_0)\,\mathbb{E}\left[\mathbb{1}_{[-\infty,t]}(X)\cdot(Y-m(X,\theta_0))\,l(X,Y,\theta_0)\right]\\ &\quad= G^T(s,\theta_0)\,J^{-1}\int\limits_{[-\infty,t]}\mathbb{E}\left[(Y-m(X,\theta_0))^2 f(X)\Big|X=x\right]dF(x)\\ &\quad= G^T(s,\theta_0)\,J^{-1}\int\limits_{[-\infty,t]}\mathbb{E}\left[(Y-m(x,\theta_0))^2\,|\,X=x\right]f(x)\,dF(x)\\ &\quad= G^T(s,\theta_0)\,J^{-1}\int\limits_{[-\infty,t]}\mathbb{V}ar\,(Y\,|\,X=x)\,f(x)\,dF(x)\\ &\quad= G^T(s,\theta_0)\,J^{-1}\int\limits_{[-\infty,t]}\sigma^2(x)f(x)\,dF(x)\\ &\quad= G^T(s,\theta_0)\,J^{-1}\int\limits_{[-\infty,t]}f(x)\,dF(x)\\ &\quad= G^T(s,\theta_0)\,J^{-1}G(t,\theta_0)\\ &\quad\overset{H_0}{=} G^T(s)\,J^{-1}G(t)\end{aligned}$$

and similarly for the fourth summand, we obtain

$$G^T(t,\theta_0)\,\mathbb{E}\left[\mathbb{1}_{[-\infty,s]}(X)\cdot(Y-m(X,\theta_0))\,l(X,Y,\theta_0)\right] = G^T(t)\,J^{-1}G(s)\,.$$

Inserting all four summands in the covariance function from equation (7.36), for $s,t \in [-\infty,\infty]$, we get

$$\begin{aligned} K^1(s,t) &= F(\min\{s,t\}) + G^T(s)\, J^{-1}\, G(t) - G^T(s)\, J^{-1} G(t) - G^T(t)\, J^{-1} G(s) \\ &= F(\min\{s,t\}) - G^T(t)\, J^{-1} G(s) \\ &= F(\min\{s,t\}) - \left(\int_{[-\infty,s]} f(x)\, dF(x) \right)^T \left(\int_{[-\infty,\infty]} f(x)\, f^T(x)\, dF(x) \right)^{-1} \int_{[-\infty,t]} f(x)\, dF(x), \end{aligned}$$

which was what had to be proven. □

Proof of Theorem 2.1.10

We begin the proof with two technical lemmas, the first of which is well known and addresses general observations regarding quantile functions.

Lemma 7.1.1.

Let F be a distribution function and F^- be the associated quantile function. Let $a, b, F^-(0)$ and $F^-(1)$ be defined as in Remark 2.1.9. Then, for all $t \in (0,1)$ and all $x \in (a,b)$, the following relations hold true:

$$i)\ \ F(F^-(t)) \geq t \quad \text{and} \quad F^-(F(x)) \leq x. \tag{7.37}$$

$$ii)\ \ F(x) \geq t \iff F^-(t) \leq x. \tag{7.38}$$

$$iii)\ \ F(F^-(F(x))) = F(x) \quad \text{and } F^-(F(F^-(t))) = F^-(t). \tag{7.39}$$

$$iv)\ \ F(F^-(t)) = t \iff t \in F((a,b] \cap \mathbb{R}). \tag{7.40}$$

Note that, if $0 < F(x) < 1$ for all $x \in \mathbb{R}$, we have $(a,b) = \mathbb{R}$ and note that it can happen that $(a,b) = \emptyset$.

Proof of Lemma 7.1.1 : Since the distribution function F is right-continuous,

$$F^-(t) = \min\{x \in \mathbb{R} : F(x) \geq t\}, \quad t \in (0,1),$$

holds true. Therefore, for all $t \in (0,1)$ and $x \in (a,b)$,

$$F(F^-(t)) = F(\min\{x \in \mathbb{R} : F(x) \geq t\}) \geq t$$

and

$$F^-(F(x)) = \min\{s \in \mathbb{R} : F(s) \geq F(x)\} \leq x$$

can be deduced. This proves (7.37).

Regarding the second assertion, if $F(x) \geq t$ holds true, due to the fact that F^- is monotonically increasing and (7.37) holds true, we can infer that $F^-(t) \leq F^-(F(x)) \leq x$. Otherwise, if $F^-(t) \leq x$ holds true, since F is monotonically increasing and (7.37) holds true, $F(x) \geq F(F^-(t)) \geq t$ can be inferred, which proves the other implication.

Regarding (7.39), the first part of (7.37) (with $t = F(x)$) shows that $F(F^-(F(x))) \geq F(x)$, and applying F to the second part of (7.37) yields $F(F^-(F(x))) \leq F(x)$. Hence, $F(F^-(F(x))) = F(x)$ for all $x \in (a, b)$. Similarly, the second part of (7.37) applied to $x = F^-(t)$ shows that $F^-(F(F^-(t))) \leq F^-(t)$, while applying F^- to the first property yields $F^-(F(F^-(t))) \geq F^-(t)$, and hence, we have $F^-(F(F^-(t))) = F^-(t)$ for all $t \in (0, 1)$. We have thus shown (7.39).

Regarding the fourth assertion, since $F^-(t) \in (a, b] \cap \mathbb{R}$ for all $t \in (0, 1)$, we can infer from $t = F(F^-(t))$ that $t \in F((a, b] \cap \mathbb{R})$ holds true. To see the reverse implication, let $t \in F((a, b] \cap \mathbb{R})$ be true. Then, there is an $x_0 \in (a, b] \cap \mathbb{R}$ with $F(x_0) = t$ and thus $\{x \in (a, b] \cap \mathbb{R} \,|\, F(x) = t\}$ is not empty. Therefore, because F is monotonically increasing, we can conclude that

$$F(F^-(t)) = F(\min\{x \in (a, b] \cap \mathbb{R} | F(x) \geq t\}) = t$$

holds true. □

The second lemma is concerned with the main technical identities necessary for the proof of Theorem 2.1.10.

Lemma 7.1.2.

Let X be a random variable with distribution function F. Let F^- denote the quantile function associated to F. Define

$$U := F(X) \quad \text{and} \quad \widetilde{X} := F^-(U).$$

Then, the following assertions hold true:

i) $\mathbb{P}^U|_{\sigma(F^-)} = \lambda|_{\sigma(F^-)}$.

ii)
$$F_U(t) = \mathbb{P}(U \leq t) = \begin{cases} 0, & t \leq 0, \\ \lim\limits_{x \uparrow F^-(t)} F(x), & t \notin F(\mathbb{R}) \text{ and } t \in (0, 1), \\ t, & t \in F(\mathbb{R}) \text{ and } t \in (0, 1), \\ 1, & t \geq 1. \end{cases}$$

iii) $X = \widetilde{X}$ a.s.

iv) $F(\widetilde{X}) = U$.

Proof of Lemma 7.1.2

We claim that in the given situation

$$\mathbb{P}\left(F^-(F(X)) \in (-\infty, d]\right) = F(d), \quad d \in (a, b). \tag{7.41}$$

Here, as defined in 2.1.9, $a = \inf\{x \in \mathbb{R} | F(x) > 0\}$, $b = \sup\{x \in \mathbb{R} | F(x) < 1\}$ and $F^-(1) = b$. Note that $X > a$ a.s. and therefore, in this proof, we can assume $X > a$.
To see that equation (7.41) holds true, first note that if $X \leq d$, then $F^-(F(X)) \leq F^-(F(d)) \leq d$ due to (7.37). This proves

$$\mathbb{P}\left(F^-(F(X)) \in (-\infty, d]\right) \geq \mathbb{P}(X \leq d) = F(d).$$

To prove the converse inequality, define $c := \sup\{y \in \mathbb{R} | F(y) \leq F(d)\}$ and note that $c \in [d, b) \subset (a, b) \subseteq \mathbb{R}$ is well defined, since $F(d) < 1$. Now, there are two cases:

Case 1; $F(c) \leq F(d)$:
In this case, if $F^-(F(X)) \leq d$, (7.39) shows that $F(X) = F(F^-(F(X))) \leq F(d)$ and hence $X \leq c$. Therefore,

$$\mathbb{P}\left(F^-(F(X)) \in (-\infty, d]\right) \leq \mathbb{P}(X \leq c) = F(c) \leq F(d).$$

Case 2; $F(c) > F(d)$:
In this case, if $F^-(F(X)) \leq d$, then (7.39) shows $F(X) = F(F^-(F(X))) \leq F(d)$ and hence $X < c$ by the definition of c and because $F(c) > F(d)$. Therefore,

$$\begin{aligned}
\mathbb{P}(F^-(F(X)) \in (-\infty, d]) &\leq \mathbb{P}(X < c) \\
&= \lim_{n\to\infty} \mathbb{P}\left(X \leq c - \frac{1}{n}\right) \\
&= \lim_{n\to\infty} F\left(c - \frac{1}{n}\right) \\
&\leq F(d)
\end{aligned}$$

holds true by the definition of c and the fact that F is monotonically increasing.

Altogether, (7.41) is proved.

To complete the proof of Lemma 7.1.2, we shall use the probability space $((0,1), \mathcal{B}, \lambda\!\!\lambda_{[0,1]})$, on which we define the random variable $V(x) = x$, meaning that $V \sim U(0,1)$. It is well known that $F^-(V) \sim F \sim X$. Furthermore, we define

$$P := \{(F^-)^{-1}((-\infty, d]) \,|\, d \in (a, b)\}$$

and note that $\sigma(F^-) = \sigma(P)$ and that P is closed under non-empty finite intersections. We will now prove the assertions i) - iv) of Lemma 7.1.2 one by one.

Regarding i): It is sufficient to prove that $\mathbb{P}(U \in M) = \lambda\!\!\!^{-}(M)$ for all $M \in P$. But with $M = (F^-)^{-1}((-\infty, d])$ for some $d \in (a, b)$, we see in view of (7.41) and because of $F^-(V) \sim X$, that

$$\begin{aligned}\lambda\!\!\!^{-}_{[0,1]}(M) &= \lambda\!\!\!^{-}_{[0,1]}((F^-)^{-1}((-\infty, d])) = \lambda\!\!\!^{-}_{[0,1]}(F^-(V) \in (-\infty, d]) \\ &= \mathbb{P}(X \in (-\infty, d]) = F(d) \\ &= \mathbb{P}(F^-(F(X)) \in (-\infty, d]) = \mathbb{P}(U \in (F^-)^{-1}((-\infty, d])) \\ &= \mathbb{P}(U \in M).\end{aligned}$$

Regarding ii): According to the definition of $U := F(X)$, U takes values in $[0, 1]$. Therefore, $F_U(t) = 0$ for $t < 0$, $F_U(t) = 1$ for $t \geq 1$. For $t = 0$, we get

$$F_U(0) = \mathbb{P}(F(X) \leq 0) = \mathbb{P}(F(X) = 0) = 0, \tag{7.42}$$

since, either $F(x) > 0$ for all $x \in \mathbb{R}$, or there is an $x_0 \in \mathbb{R}$ such that $F(x) > 0$ for all $x > x_0$ and $F(x) = 0$ for all $x < x_0$. In the first case, the last equation in (7.42) is obviously valid, and in the second case

$$\mathbb{P}(F(X) = 0) = \mathbb{P}(X < x_0) = \lim_{n\to\infty} F\left(x_0 - \frac{1}{n}\right) = 0$$

holds, stating that the last equation in (7.42) is valid, too. The following two cases remain:

Case 1; $t \in (0, 1)$ and $t \in F(\mathbb{R})$:
Since $t \in F(\mathbb{R})$, there is an $x \in \mathbb{R}$ with $F(x) = t$ and therefore, according to (7.38),

$$\begin{aligned}(0, t] = (0, F(x)] &= \{s \in (0, 1) | s \leq F(x)\} = \{s \in (0, 1) | F^-(s) \leq x\} \\ &= (F^-)^{-1}((-\infty, x]) \in \sigma(F^-)\end{aligned}$$

holds true. Thus, since we have $\mathbb{P}^U|_{\sigma(F^-)} = \lambda\!\!\!^{-}|_{\sigma(F^-)}$, we can conclude

$$F_U(t) = \mathbb{P}(U \leq t) = \mathbb{P}(U \in (0, t]) = \lambda\!\!\!^{-}_{[0,1]}((0, t]) = t.$$

Case 2; $t \in (0, 1)$ and $t \notin F(\mathbb{R})$:
In this case, as claimed,

$$\begin{aligned}\mathbb{P}(F(X) \leq t) &= \mathbb{P}(X < F^-(t)) \\ &= \mathbb{P}^X\left(\bigcup_{n=1}^{\infty}\left(-\infty, F^-(t) - \frac{1}{n}\right]\right)\end{aligned} \tag{7.43}$$

$$= \lim_{n\to\infty} \mathbb{P}^X \left(\left(-\infty, F^-(t) - \frac{1}{n} \right] \right)$$

$$= \lim_{n\to\infty} F\left(F^-(t) - \frac{1}{n} \right)$$

$$= \lim_{x \uparrow F^-(t)} F(x).$$

is valid. In the above equation, (7.43) holdes true, since for $x \in \mathbb{R}$ and $t \in (0,1)$ the following equivalences apply:

$$\begin{aligned} F(x) \leq t \text{ and } t \notin F(\mathbb{R}) &\iff F(x) < t \\ &\iff \neg\left(F(x) \geq t\right) \\ &\overset{(7.38)}{\iff} \neg\left(F^-(t) \leq x\right) \\ &\iff x < F^-(t). \end{aligned}$$

Here $\neg P$ denotes the negation of a proposition P.

Regarding iii): (7.37) implies that $\widetilde{X} = F^-(U) = F^-(F(X)) \leq X$. Assertion i), however, implies that for any Borel set $M \subseteq \mathbb{R}$,

$$\begin{aligned} \mathbb{P}(\widetilde{X} \in M) &= \mathbb{P}\left(U \in (F^-)^{-1}(M)\right) = \lambda\!\!\!\lambda_{[0,1]}\left((F^-)^{-1}(M)\right) \\ &= \lambda\!\!\!\lambda_{[0,1]}\left(V \in (F^-)^{-1}(M)\right) = \lambda\!\!\!\lambda_{[0,1]}\left(F^-(V) \in M\right) \\ &= \mathbb{P}(X \in M), \end{aligned} \tag{7.44}$$

holds true, since $F^-(V) \sim F \sim X$. Hence, $\widetilde{X} \sim X$ and $\widetilde{X} \leq X$ almost surely, which is only possible if $\widetilde{X} = X$ a.s. To see this, consider $Y = \arctan(X)$ and $\widetilde{Y} = \arctan(\widetilde{X})$. Then, $\widetilde{Y} \leq Y$ and $\widetilde{Y} \sim Y$ hold true and thus

$$\mathbb{E}\left(|Y - \widetilde{Y}|\right) = \mathbb{E}\left(Y - \widetilde{Y}\right) = \mathbb{E}(Y) - \mathbb{E}(\widetilde{Y}) = 0.$$

Thus, $|Y - \widetilde{Y}| = 0$ a.s., hence $Y = \widetilde{Y}$ a.s. and therefore $X = \widetilde{X}$ a.s. Note that we have used the arctan transformation to ensure that $\mathbb{E}(Y)$ exists.

Regarding iv): From (7.39), we can infer that

$$F(\widetilde{X}) = F(F^-(U)) = F(F^-(F(X))) = F(X) = U.$$

□

We can now proceed with the actual proof of Theorem 2.1.10. The application of Lemma 7.1.2 now guarantees the existence of a random variable U that has the desired distributional properties and which we use as the design variable of the transformed regression model with random design

$$Y = (m \circ F^-)(U) + \varepsilon.$$

It remains to prove the following assertions:

i) In the transformed model, the relation

$$(m \circ F^-)(u) = \mathbb{E}(Y|U = u), \quad u \in (0,1),$$

holds true, which states that $m \circ F^-$ is indeed the new regression function.

ii) The relations $\mathbb{E}(\varepsilon \,|\, U = u) = 0$ and $\mathbb{V}ar(\varepsilon \,|\, U = u) = \sigma^2(F^-(u))$ hold true in the transformed model.

iii) Assumption 2 of Section 2.1.1 is satisfied in the situation of the transformed model.

In the following, we prove these assertions one by one.

Regarding i) Since $U = F(\widetilde{X})$ and $\widetilde{X} = F^-(U)$, we have $\sigma(\widetilde{X}) = \sigma(U)$ and therefore

$$\mathbb{E}(Y|\widetilde{X}) = \mathbb{E}(Y|U) \quad \text{a.s.}$$

holds true. Furthermore, we already know that $\widetilde{X} = X$ a.s. and, since $\sigma(\widetilde{X}) = \sigma(U) \subseteq \sigma(X)$, we additionally know that $\mathbb{E}(Y|\widetilde{X})$ is $\sigma(X)$ measurable. We therefore get for all $B \in \mathcal{B}$

$$\mathbb{E}(\mathbb{E}(Y|\widetilde{X}) \cdot \mathbb{1}_B(X)) = \mathbb{E}(\mathbb{E}(Y|\widetilde{X}) \cdot \mathbb{1}_B(\widetilde{X})) = \mathbb{E}(Y \cdot \mathbb{1}_B(\widetilde{X})) = E(Y \cdot \mathbb{1}_B(X)) \text{ a.s.}, \tag{7.45}$$

which means that $\mathbb{E}(Y|\widetilde{X}) = \mathbb{E}(Y|X)$ a.s. Altogether, we have

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|U) \quad \text{a.s.}$$

Now, since in the original model the factorisation $m(x) = \mathbb{E}(Y|X = x)$ of the conditional expectation $\mathbb{E}(Y|X)$ was given as the regression function and

$$(m \circ F^-)(U) = m(F^-(U)) = m(\widetilde{X}) = m(X) = \mathbb{E}(Y|X) = \mathbb{E}(Y|U) \quad \text{a.s.}$$

holds, we know that $m \circ F^-$ is a factorisation of $\mathbb{E}(Y|U)$ and thus get

$$(m \circ F^-)(u) = \mathbb{E}(Y|U = u) \quad \mathbb{P}^U\text{-a.e. for } u \in (0,1).$$

Regarding ii) According to assertion i), in the transformed model we have

$$\mathbb{E}(\varepsilon|U) = \mathbb{E}(Y - m(F^-(U))|U) = \mathbb{E}(Y|U) - m(F^-(U)) = 0$$

and therefore

$$\mathbb{E}(\varepsilon|U=u)=0 \quad \mathbb{P}^U\text{-a.e. for } u\in(0,1).$$

With the same arguments as used in the proof of (7.45), one can see that $\mathbb{E}(Z|\widetilde{X}) = \mathbb{E}(Z|X)$ a.s. for every integrable random variable Z and therefore one has $\mathbb{E}(Y^2|\widetilde{X}) = \mathbb{E}(Y^2|X)$. Again, since $\sigma(\widetilde{X}) = \sigma(U)$, we can infer that $\mathbb{E}(Y^2|U) = \mathbb{E}(Y^2|X)$ holds true a.s. for a square integrable Y and thus, we have

$$\mathbb{E}(Y^2|X=F^-(u))=\mathbb{E}(Y^2|U=u) \quad \mathbb{P}^U\text{-a.e. for } u\in(0,1).$$

Using this, we conclude that for $\mathbb{P}^U$-a.a. $u\in(0,1)$,

$$\begin{aligned}
\sigma^2(F^-(u)) &= \mathbb{V}ar(\varepsilon\,|\,X=F^-(u)) \\
&= \mathbb{V}ar(Y\,|\,X=F^-(u)) \\
&= \mathbb{E}(Y^2\,|\,X=F^-(u)) - [\mathbb{E}(Y\,|\,X=F^-(u))]^2 \\
&= \mathbb{E}(Y^2\,|\,U=u) - [\mathbb{E}(Y\,|\,U=u)]^2 \\
&= \mathbb{V}ar(Y\,|\,U=u) \\
&= \mathbb{V}ar(\varepsilon\,|\,U=u)
\end{aligned}$$

holds true.

Regarding iii) Under the null hypothesis $\mathrm{H}_0: m\in\mathcal{M}:=\{m(\cdot,\theta)|\theta\in\Theta\}$, $\Theta\subseteq\mathbb{R}^p$, for $u\in(0,1)$, similar to the proof of Theorem 2.1.6, we define

$$\begin{aligned}
\overline{g}(u,\theta) &:= (\overline{g}_1(u,\theta),\ldots,\overline{g}_p(u,\theta))^T \\
&:= \left(\frac{\partial m(F^-(u),\theta)}{\partial\theta_1},\ldots,\frac{\partial m(F^-(u),\theta)}{\partial\theta_p}\right)^T \\
&= \frac{\partial m(F^-(u),\theta)}{\partial\theta},
\end{aligned}$$

which is well defined since Assumption 2 is satisfied in the original model. Note that, since Assumption 2 in the original model is satisfied, there is the $\lambda\!\!\lambda_{[0,1]}-$integrable function $\overline{M}(x) := (M\circ F^-)(x)$, $x\in\mathbb{R}$, such that

$$\forall\, 1\le j\le p \quad \forall\, x\in\mathbb{R} \quad \forall\,\theta\in int(\Theta): \quad |\overline{g}_i(x,\theta)|\le\overline{M}(x).$$

Therefore, Assumption 2 is also fulfilled in the transformed model. In particular, $\overline{g}_i(\cdot,\theta)$ is $\lambda\!\!\lambda_{[0,1]}-$integrable for every $i=1,\ldots,p$ and therefore,

$$\overline{G}(u,\theta) := (\overline{G}_1(u,\theta),\ldots,\overline{G}_p(u,\theta))^T$$

$$:= \left(\int_{[0,u]} \overline{g}_1(v,\theta)\, d\lambda\!\!\!^{-}_{[0,1]}(v), \ldots, \int_{[0,u]} \overline{g}_p(v,\theta)\, d\lambda\!\!\!^{-}_{[0,1]}(v) \right)^T, \quad u \in (0,1),$$

exists and its components are continuous functions with respect to θ. □

Proof of Corollary 2.2.5

First, we compute the mean function $m(s)$ for $s \in [a,b]$.

$$\begin{aligned} m(s) &= \mathbb{E}((B_{f,F} \circ F)(s)) \\ &= \mathbb{E}\left(B(F(s)) - \left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1} \int_{[a,b]} f(t)\, dB(F)(t) \right) \\ &= \underbrace{\mathbb{E}(B(F(s)))}_{=0} - \mathbb{E}\left(\left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1} \int_{[a,b]} f(t)\, dB(F)(t) \right) \\ &= -\left(\int_{[a,s]} f(t)\, dF(t) \right)^T J^{-1}\, \mathbb{E}\left[\int_{[a,b]} f(t)\, dB(F)(t) \right], \quad s \in [a,b]. \end{aligned} \tag{7.46}$$

According to Theorem 2.2.4 and its proof

$$\int_{[a,b]} f(t)\, dB(F(t)) = \overset{(R^*)}{\int_{[a,b]}} f(t)\, dB(F(t)) = \overset{(R)}{\int_{[a-\varepsilon,b]}} \overline{f}(t)\, d\widetilde{(B \circ F)}(t) = \overset{(R)}{\int_{[a-\varepsilon,b]}} \overline{f}(t)\, dB(F(t)) \tag{7.47}$$

holds true for every $\varepsilon > 0$, where $\overline{f}(x) = f(x)$, $x \in [a,b]$, $\overline{f}(x) = f(a)$, $x \in [-\infty, a)$ and $\widetilde{(B \circ F)}(t) = B(0) = 0 = B(0) = B(F(t))$, $t \in [a-\varepsilon, a)$. Let $\varepsilon > 0$ be given. For $m \in \mathbb{N}$, let $a - \varepsilon = x_{m1} \leq \cdots \leq x_{mm} = b$ be a partition of the interval $[a-\varepsilon, b]$ with $\sup_{j=2,\ldots,m} |x_{mj} - x_{mj-1}| \xrightarrow[m\to\infty]{} 0$. Due to the theorem of Lebesgue, we get

$$\begin{aligned} \mathbb{E}\left[\int_{[a,b]} f(t)\, dB(F)(t) \right] &= \mathbb{E}\left[\overset{(R)}{\int_{[a-\varepsilon,b]}} \overline{f}(t)\, dB(F(t)) \right] \\ &= \mathbb{E}\left[\lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi}) \left(B(F(x_{mi})) - B(F(x_{mi-1})) \right) \right] \end{aligned}$$

$$= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi}) \left[\underbrace{\mathbb{E}\left(B(F(x_{mi}))\right)}_{=0} - \underbrace{\mathbb{E}\left(B(F(x_{mi-1}))\right)}_{=0} \right]$$

$$= 0.$$

Inserting this in equation (7.46) yields

$$m(s) = 0, \quad s \in [a,b],$$

which had to be proven. Secondly, we have to show that for $s,t \in [a,b]$,

$$K(s,t) = F(\min\{s,t\}) - \left(\int_{[a,s]} f(x)\, dF(x) \right)^T \left(\int_{[a,b]} f(t) f^T(t)\, dF(t) \right)^{-1} \int_{[a,t]} f(x)\, dF(x)$$

is valid. Therefore, for $s,t \in [a,b]$, we calculate

$$K(s,t) = Cov\left[(B_{f,F} \circ F)(s), (B_{f,F} \circ F)(t)\right]$$

$$= Cov\left[B(F(s)) - \left(\int_{[a,s]} f(x) dF(x) \right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x), \right. \tag{7.48}$$

$$\left. B(F(t)) - \left(\int_{[a,t]} f(x) dF(x) \right)^T J^{-1} \int_{[a,b]} f(x) dB(F)(x) \right]$$

$$= \underbrace{Cov(B(F(s)), B(F(t)))}_{=:\,\mathrm{I}} - \underbrace{Cov\left(B(F(s)), \left(\int_{[a,t]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x) \right)}_{=:\,\mathrm{II}}$$

$$\underbrace{- Cov\left(B(F(t)), \left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x) \right)}_{=:\,\mathrm{III}}$$

$$\underbrace{+ Cov\left(\left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x), \left(\int_{[a,t]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x) \right)}_{=:\,\mathrm{IV}}. \tag{7.49}$$

Now, the task is to compute the terms **I** to **IV**.

As for I:

$$Cov(B(F(s)), B(F(t))) = \min\{F(s), F(t)\} = F(\min\{s,t\}), \quad s,t \in [a,b],$$

since B is a standard Brownian motion.

As for II: Let $\varepsilon > 0$ be given. For $m \in \mathbb{N}$, let $a - \varepsilon = x_{m1} \leq \cdots \leq x_{mm} = b$ be a partition of the interval $[a-\varepsilon, b]$ with $\sup\limits_{j=2,\ldots,m} |x_{mj} - x_{mj-1}| \xrightarrow[m\to\infty]{} 0$. According to (7.47), we then obtain

$$\begin{aligned}
Cov\left(B(F(s)), \int\limits_{[a,b]} f(x)\, dB(F)(x)\right) &= Cov\left(B(F(s)), \int\limits_{[a-\varepsilon,b]}^{(R)} \overline{f}(x)\, dB(F)(x)\right) \\
&= Cov\left(B(F(s)), \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\,(B(F(x_{mi})) - B(F(x_{mi-1})))\right) \\
&= \mathbb{E}\left(B(F(s)) \cdot \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\,(B(F(x_{mi})) - B(F(x_{mi-1})))\right), \quad s \in [a,b],
\end{aligned}$$

as the expected value of both summands is zero. Due to the theorem of Lebesgue, we get

$$\begin{aligned}
&\mathbb{E}\left(B(F(s)) \cdot \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\,(B(F(x_{mi})) - B(F(x_{mi-1})))\right) \\
&\quad= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\, \mathbb{E}\,(B(F(s)) \cdot (B(F(x_{mi})) - B(F(x_{mi-1})))) \\
&\quad= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\, Cov\,(B(F(s)), (B(F(x_{mi})) - B(F(x_{mi-1})))) \\
&\quad= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\ (\min\{F(s), F(x_{mi})\} - \min\{F(s), F(x_{mi-1})\}) \\
&\quad= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})\ \left(F_{F(s)}(x_{mi}) - F_{F(s)}(x_{mi-1})\right) \\
&\quad= \int\limits_{[a-\varepsilon,b]}^{(R)} \overline{f}(x)\, dF_{F(s)}(x) \ = \int\limits_{[a,b]} f(x)\, dF_{F(s)}(x) \\
&\quad= \int\limits_{[a,s]} f(x)\, dF(x), \quad s \in [a,b]. \qquad (7.50)
\end{aligned}$$

Note that equation (7.50) holds true, since the signed measure $\mu_{F_{F(s)}}$ associated to $F_{F(s)}$ has its support in $[a,s]$ and $\mu_{F_{F(s)}} = \mu_F|_{[a,s]}$, where, as usual, μ_F denotes the measure associated to F. Taking the above into account, we obtain

$$\begin{aligned} &Cov\left(B(F(s)), \left(\int_{[a,t]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\,dB(F)(x)\right) \\ &\qquad = \left(\int_{[a,t]} f(x)\,dF(x)\right)^T J^{-1}\, Cov\left(B(F(s)), \int_{[a,b]} f(x)\,dB(F)(x)\right) \\ &\qquad = \left(\int_{[a,t]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,s]} f(x)\,dF(x), \quad s,t \in [a,b]. \end{aligned}$$

As for III: Similar to II, we obtain

$$\begin{aligned} &Cov\left(B(F(t)), \left(\int_{[a,s]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\,dB(F)(x)\right) \\ &\qquad = \left(\int_{[a,s]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,t]} f(x)\,dF(x), \quad s,t \in [a,b]. \end{aligned}$$

As for IV: Again, let $\varepsilon > 0$. For $m \in \mathbb{N}$ let $a - \varepsilon = x_{m1} \leq \cdots \leq x_{mm} = b$ be a partition of the interval $[a-\varepsilon, b]$ with $\sup_{j=2,\ldots,m} |x_{mj} - x_{mj-1}| \xrightarrow[m\to\infty]{} 0$. Thus, we obtain

$$\begin{aligned} &Cov\left(\left(\int_{[a,s]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\,dB(F)(x), \left(\int_{[a,t]} f(x)\,dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\,dB(F)(x)\right) \\ &= Cov\left(\left(\int_{[a,s]} f(x)\,dF(x)\right)^T J^{-1} \overset{(R)}{\int_{[a-\varepsilon,b]}} \overline{f}(x)\,dB(F)(x), \right. \\ &\qquad \left. \left(\int_{[a,t]} f(x)\,dF(x)\right)^T J^{-1} \overset{(R)}{\int_{[a-\varepsilon,b]}} \overline{f}(x)\,dB(F)(x)\right) \end{aligned}$$

$$= \left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} Cov \left(\int_{[a-\varepsilon,b]}^{(R)} \overline{f}(x)\, dB(F)(x), \int_{[a-\varepsilon,b]}^{(R)} \overline{f}(x)\, dB(F)(x) \right) \times$$

$$\times\ J^{-1} \int_{[a,t]} f(x)\, dF(x)$$

$$= \left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} \times$$

$$\times\ Cov \left(\lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})(B(F(x_{mi})) - B(F(x_{mi-1}))), \right.$$

$$\left. \lim_{m\to\infty} \sum_{j=2}^{m} \overline{f}(x_{mj})(B(F(x_{mj})) - B(F(x_{mj-1}))) \right) \times$$

$$\times\ J^{-1} \int_{[a,t]} f(x)\, dF(x), \qquad s,t \in [a,b]. \tag{7.51}$$

Due to the theorem of Lebesgue, we can reshape the covariance from the term above as follows:

$$Cov \left(\lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi}) \left(B(F(x_{mi})) - B(F(x_{mi-1}))\right), \right.$$

$$\left. \lim_{m\to\infty} \sum_{j=2}^{m} \overline{f}(x_{mj})(B(F(x_{mj})) - B(F(x_{mj-1}))) \right)$$

$$= \lim_{m\to\infty} \sum_{i=2}^{m} \sum_{j=2}^{m} \overline{f}(x_{mi}) \overline{f}^T(x_{mj}) \left[Cov(B(F(x_{mi})), B(F(x_{mj}))) - \right.$$

$$Cov(B(F(x_{mi-1})), B(F(x_{mj}))) -$$

$$Cov(B(F(x_{mi})), B(F(x_{mj-1}))) +$$

$$\left. Cov(B(F(x_{mi-1})), B(F(x_{mj-1}))) \right]$$

$$= \lim_{m\to\infty} \left[\sum_{i=2}^{m} \sum_{j=2}^{i-1} \overline{f}(x_{mi}) \overline{f}^T(x_{mj}) \left(F(x_{mj}) - F(x_{mj}) - F(x_{mj-1}) + F(x_{mj-1})\right) + \right.$$

$$\sum_{i=2}^{m} \sum_{j=i+1}^{m} \overline{f}(x_{mi}) \overline{f}^T(x_{mj}) \left(F(x_{mi}) - F(x_{mi-1}) - F(x_{mi}) + F(x_{mi-1})\right)$$

$$\left. \sum_{i=2}^{m} \overline{f}(x_{mi}) \overline{f}^T(x_{mi}) \left(F(x_{mi}) - F(x_{mi-1}) - F(x_{mi-1}) + F(x_{mi-1})\right) \right]$$

$$= \lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi}) \overline{f}^{\,T}(x_{mi}) \left(F(x_{mi}) - F(x_{mi-1})\right)$$

$$= \int_{[a-\varepsilon,b]}^{(R)} \overline{f}(x) \overline{f}^{\,T}(x)\, dF$$

$$= \int_{[a,b]} f(x) f^T(x)\, dF$$

$$= J. \tag{7.52}$$

The last equation holds true, due to the definition of J in (2.7) on page 30. Finally, inserting (7.52) in (7.51) leads to

$$Cov\left(\left(\int_{[a,s]} f(x)\, dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x), \left(\int_{[a,t]} f(x)\, dF(x)\right)^T J^{-1} \int_{[a,b]} f(x)\, dB(F)(x)\right)$$

$$= \left(\int_{[a,s]} f(x)\, dF(x)\right)^T J^{-1} \times$$

$$\times\ Cov\left(\lim_{m\to\infty} \sum_{i=2}^{m} \overline{f}(x_{mi})(B(F(x_{mi})) - B(F(x_{mi-1}))),\right.$$

$$\left.\lim_{m\to\infty} \sum_{j=2}^{m} \overline{f}(x_{mj})(B(F(x_{mj})) - B(F(x_{mj-1})))\right) \times$$

$$\times\ J^{-1} \int_{[a,t]} f(x)\, dF(x)$$

$$= \left(\int_{[a,s]} f(x)\, dF(x)\right)^T J^{-1}\, J\, J^{-1} \int_{[a,t]} f(x)\, dF(x)$$

$$= \left(\int_{[a,s]} f(x)\, dF(x)\right)^T J^{-1} \int_{[a,t]} f(x)\, dF(x), \quad s,t \in [a,b].$$

We insert **I**, **II**, **III** and **IV** in equation (7.49) on page 130 to determine the covariance function of $B_{f,F} \circ F$, for $s,t \in [a,b]$, as follows:

$$K(s,t) = F(\min\{s,t\}) - \left(\int_{[a,t]} f(x)\, dF(x)\right)^T J^{-1} \int_{[a,s]} f(x)\, dF(x)$$

$$
\begin{aligned}
&- \left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,t]} f(x)\, dF(x) + \left(\int_{[a,s]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,t]} f(x)\, dF(x) \\
&= F(\min\{s,t\}) - \left(\int_{[a,t]} f(x)\, dF(x) \right)^T J^{-1} \int_{[a,s]} f(x)\, dF(x) \\
&= F(\min\{s,t\}) - \left(\int_{[a,s]} f(x)\, dF(x) \right)^T \left(\int_{[a,b]} f(t) f^T(t)\, dF(t) \right)^{-1} \int_{[a,t]} f(x)\, dF(x).
\end{aligned}
$$

This completes the proof. □

Proof of Lemma 2.2.7

The proof is divided into four lemmas and one corollary, which together proof Lemma 2.2.7. These are:

Lemma 7.1.3: This lemma states a variant of the well-known change of variables formula for integrals in the situation of an integral with respect to a distribution function. Since this transformation is used in a number of places in this work, we state a proof here for the sake of completeness.

Lemma 7.1.4: This lemma addresses a duality between distribution functions and quantile functions.

Lemma 7.1.5: This lemma deals with the special case of Lemma 2.2.7 in which the distribution function F has support in $[0,1]$ (i.e. $c = 0$ and $d = 1$) and $a = 0$ and $b = 1$ hold true. Note that Evers (2022) contains a different proof of a result that is quite similar to Lemma 7.1.5. The difference is that the proof in Evers (2022) works for infinitely many functions f_i, but yields only continuous functions g_i, instead of Lipschitz continuous functions.

Corollary 7.1.6: This corollary contains the main assertion of Lemma 2.2.7 and is a generalisation of Lemma 7.1.5. This corollary can also be found in Evers (2022), as it was developed together with the author.

Lemma 7.1.7: This lemma proves the additional assertion made in Lemma 2.2.7, according to which the linear independence of the functions $f_1, \ldots, f_p$ and $g_1, \ldots, g_p$ is equivalent.

Lemma 7.1.3.

Let $F : \mathbb{R} \longrightarrow \mathbb{R}$ be a distribution function, let $a \leq b \in \mathbb{R}$ and let $g : \mathbb{R} \to \mathbb{R}$ be measurable and F-integrable. Then

$$\int\limits_{(a,b]} g(x)\, dF(x) = \int\limits_{(F(a),F(b)]} (g \circ F^-)(u))\, d\lambda\!\!\lambda_{[0,1]}(u)$$

holds true. If, in addition $a \leq \inf(supp(F))$ is valid, it holds true that

$$\int\limits_{[a,b]} g(x)\, dF(x) = \int\limits_{[0,F(b)]} (g \circ F^-)(u))\, d\lambda\!\!\lambda_{[0,1]}(u).$$

Proof of Lemma 7.1.3:

First, note that $\mathbb{1}_{(a,b]}(x) \cdot g(x)$ is measurable and F-integrable, since g is and that the quantile function $F^- : (0,1) \to \mathbb{R}$ is continuous from the left and therefore measurable, hence $g \circ F^-$ is measurable. Additionally, we already know from equation (7.44), page 126, that $\lambda\!\!\lambda_{[0,1]}^{F^-} \sim X \sim F$. That is, $\mathbb{P}^X$ for $X \sim F$ is the pushforward measure of $\lambda\!\!\lambda_{[0,1]}$ with respect to the measurable mapping F^-. We will use this fact in equation (7.53) of the following series of equations. Thus,

$$\begin{aligned}
\int\limits_{(a,b]} g(x)\, dF(x) &= \int\limits_{(a,b]} g(x)\, d\lambda\!\!\lambda_{[0,1]}^{F^-}(x) && (7.53)\\
&= \int\limits_{\mathbb{R}} \mathbb{1}_{(a,b]}(x) \cdot g(x)\, d\lambda\!\!\lambda_{[0,1]}^{F^-}(x)\\
&= \int\limits_{(F^-)^{-1}(\mathbb{R})} \big((g \cdot \mathbb{1}_{(a,b]}) \circ F^-\big)(u)\, d\lambda\!\!\lambda_{[0,1]}(u) && (7.54)\\
&= \int\limits_{[0,1]} \mathbb{1}_{(a,b]}(F^-(u)) \cdot g(F^-(u))\, d\lambda\!\!\lambda_{[0,1]}(u)\\
&= \int\limits_{[0,1]} \mathbb{1}_{(F(a),F(b)]}(u) \cdot (g \circ F^-)(u)\, d\lambda\!\!\lambda_{[0,1]}(u) && (7.55)\\
&= \int\limits_{(F(a),F(b)]} (g \circ F^-)(u)\, d\lambda\!\!\lambda_{[0,1]}(u).
\end{aligned}$$

holds true.

In the equations above, (7.54) holds true, according to the standard change of variables formula for pushforward measures (see for example Paragraph 19 in Bauer (1992) in this regard). Furthermore, in equation (7.55) we used the equivalence

$$a < F^-(u) \leq b \iff F(a) < u \leq F(b), \quad u \in [0,1],$$

which holds true according to (7.38) of Lemma 7.1.1 and the fact that

$$a < F^-(u) \iff \neg(a \geq F^-(u)) \iff \neg(F(a) \geq u) \iff F(a) < u.$$

If, in addition $a \leq \inf(supp(F))$ is valid, then either $F(a) = 0$ and F is continuous in a or $F(x) = 0$ for all $x \in (-\infty, a)$ and F has a jump at a. In the first case,

$$\begin{aligned}
\int_{[a,b]} g(x)\, dF(x) &= \overbrace{\int_{\{a\}} g(x)\, dF(x)}^{=0} + \int_{(a,b]} g(x)\, dF(x) \\
&= \overbrace{\int_{[0,F(a)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u)}^{=0} + \int_{(F(a),F(b)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u) \\
&= \int_{[0,F(b)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u)
\end{aligned}$$

holds true. In the second case $F^-(u) = a$ holds true for all $u \in [0, F(a)]$. Thus,

$$\begin{aligned}
\int_{[a,b]} g(x)\, dF(x) &= \int_{\{a\}} g(x)\, dF(x) + \int_{(a,b]} g(x)\, dF(x) \\
&= F(a) \cdot g(a) + \int_{(F(a),F(b)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u) \\
&= \int_{[0,F(a)]} (g \circ F^-)(u)\, \lambda\!\!\!\lambda_{[0,1]}(u) + \int_{(F(a),F(b)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u) \\
&= \int_{[0,F(b)]} (g \circ F^-)(u))\, \lambda\!\!\!\lambda_{[0,1]}(u).
\end{aligned}$$

is satisfied. □

Lemma 7.1.4.

i) Let $Q : [0,1] \longrightarrow [Q(0), Q(1)]$ be a quantile function, i.e. Q is continuous from the left and monotonically increasing. Define $Q^+ : \mathbb{R} \longrightarrow [0,1]$ by

$$Q^+(x) := \begin{cases} 0, & x < Q(0), \\ \sup\{u \in [0,1] \,|\, Q(u) \leq x\}, & x \in [Q(0), Q(1)), \\ 1, & x \geq Q(1). \end{cases}$$

Then Q^+ is a distribution function and $(Q^+)^-(u) = Q(u)$, $u \in (0,1]$. Furthermore, if Q is right continuous in zero, then $(Q^+)^-(0) = Q(0)$ holds true, too and therefore $(Q^+)^- = Q$ is satisfied.

ii) Let $F : \mathbb{R} \to [0,1]$ be a distribution function with $supp(F) \subseteq [c,d]$, $c < d \in \mathbb{R}$. Then F^- is a quantile function and $(F^-)^+ = F$.

Proof of Lemma 7.1.4

Part i): From the definition it is clear that Q^+ is monotonically increasing. Note that, since Q is continuous from the left, it is valid that

$$Q^+(x) = \sup\{u \in [0,1] \,|\, Q(u) \leq x\} = \max\{u \in [0,1] \,|\, Q(u) \leq x\} \text{ for } x \in [Q(0), Q(1)).$$

Next, we show that Q^+ is continuous from the right. To see this, let $x_0 \in [Q(0), Q(1))$ and $(x_n)_{n\in\mathbb{N}} \subseteq [Q(0), Q(1))$ with $x_n \downarrow x_0$ be given. Due to $x_{n+1} \leq x_n$, $n \in \mathbb{N}$, and the fact that Q^+ is monotonically increasing, it can be deduced that $(Q^+(x_n))_{n\in\mathbb{N}}$ is monotonically decreasing. Since, furthermore, $(Q^+(x_n))_{n\in\mathbb{N}}$ is bounded from below by zero, there is a lower bound $u_0 \in [0,1]$ with $Q^+(x_n) \downarrow u_0$. Furthermore, due to $Q^+(x_0) \leq Q^+(x_n)$ for all $n \in \mathbb{N}$, we have $Q^+(x_0) \leq u_0$. Now let us suppose towards a contradiction that $Q^+(x_0) < u_0$. Then, we could infer that

$$Q(u_0) \leq Q(Q^+(x_n)) = Q(\max\{u \in [0,1] \,|\, Q(u) \leq x_n\}) \leq x_n \text{ holds true for all } n \in \mathbb{N}.$$

Hence, $Q(u_0) \leq x_0$, respectively $u_0 \in \{u \in [0,1] \,|\, Q(u) \leq x_0\}$ could be deduced. But this is a contradiction to the fact that $Q^+(x_0) = \max\{u \in [0,1] \,|\, Q(u) \leq x_0\}$ is the biggest number for which $Q(Q^+(x_0)) \leq x_0$ holds true. Thus $Q^+(x_0) = u_0$ has to be true and therefore, we have $Q^+(x_0) = u_0 = \lim_{n\to\infty} Q^+(x_n)$, that is, Q^+ is continuous from the right. Together with the fact that Q^+ is monotonically increasing, we thus know that Q^+ is a distribution function.
It remains to prove that $(Q^+)^-(u) = Q(u)$ for all $u \in (0,1]$. We first notice that for $t \in (0,1)$,

$$Q^+(Q(t)) = \max\{u \in [0,1] \,|\, Q(u) \leq Q(t)\} \geq t$$

is satisfied and thus, $Q(t) \in \{x \in \mathbb{R} \,|\, t \leq Q^+(x)\}$. It can be inferred that

$$(Q^+)^-(t) = \inf\{x \in \mathbb{R} \,|\, t \leq Q^+(x)\} \leq Q(t).$$

Let us suppose towards a contradiction that $x := (Q^+)^-(t) < Q(t)$. Then, since Q^+ is a distribution function, we can infer from equation (7.38) in Lemma 7.1.1 that

$$x < Q(t) \iff \neg(x \geq Q(t)) \iff \neg(Q^+(x) \geq t) \iff Q^+(x) < t$$

is satisfied, i.e. it holds true that $Q^+((Q^+)^-(t)) < t$. This contradicts the fact that the distribution function Q^+, according to equation (7.37) in Lemma 7.1.1, satisfies the inequality $Q^+((Q^+)^-(t)) \geq t$. Therefore, the assumption $(Q^+)^-(t) < Q(t)$ was false and due to $(Q^+)^-(t) \leq Q(t)$, it follows that $(Q^+)^-(t) = Q(t)$. Since $t \in (0,1)$ was arbitrary, $(Q^+)^-(u) = Q(u)$ is satisfied for all $u \in (0,1)$. Furthermore, since Q and $(Q^+)^-$ are quantile functions, they are continuous from the left and thus

$$(Q^+)^-(1) = \lim_{\substack{t \uparrow 1 \\ t \in (0,1)}} (Q^+)^-(t) = \lim_{\substack{t \uparrow 1 \\ t \in (0,1)}} Q(t) = Q(1)$$

is satisfied, too.
If, in addition, Q is right continuous in zero, then

$$\begin{aligned} Q(0) = \lim_{\substack{t \downarrow 0 \\ t \in (0,1)}} Q(t) = \lim_{\substack{t \downarrow 0 \\ t \in (0,1)}} (Q^+)^-(t) &= \lim_{\substack{t \downarrow 0 \\ t \in (0,1)}} \inf\{x \in \mathbb{R} \,|\, Q^+(x) \geq t\} \\ &= \inf\{x \in \mathbb{R} \,|\, Q^+(x) > 0\} \qquad (7.56) \\ &= (Q^+)^-(0) \qquad (7.57) \end{aligned}$$

can be inferred. Here, equation (7.57) is true by the definition of the quantile function, see Remark 2.1.9. It remains to prove that equation (7.56) is satisfied. For this, we define

$$M_t := \{x \in \mathbb{R} \,|\, Q^+(x) \geq t\},\ t \in (0,1] \quad \text{and} \quad M := \{x \in \mathbb{R} \,|\, Q^+(x) > 0\}.$$

Note that M is bounded below, since $Q^+(x) = 0$ for all $x < Q(0)$. We prove that

$$\inf M = \lim_{\substack{t \downarrow 0 \\ t \in (0,1)}} \inf M_t$$

is satisfied. First, we note that for $t \geq t' \in (0,1]$, $M_t \subseteq M_{t'}$ can be inferred and hence $\inf M_t \geq \inf M_{t'} \geq \inf M > \infty$ holds true. Therefore, $\lim_{t \downarrow 0} \inf M_t \geq \inf M$ exists. Next, given any $x \in M$, there is a $t_x \in (0,1]$ such that $x \in M_{t_x}$. Thus, for $0 < t \leq t_x$, we can infer that $\inf M_t \leq \inf M_{t_x} \leq x$ holds true. Hence, $\lim_{t \downarrow 0} \inf M_t \leq x$ is satisfied for all $x \in M$ and therefore, $\lim_{t \downarrow 0} \inf M_t \leq \inf M$ is satisfied, too.

Part ii): According to Definition 2.1.8, F^- is the quantile function associated to F. Furthermore, since

$$(F^-)^+(x) = \sup\{u \in [0,1] \,|\, F^-(u) \leq x\} \overset{(7.38)}{=} \sup\{u \in [0,1] \,|\, F(x) \geq u\} = F(x), \quad x \in (a,b),$$

holds true, $(F^-)^+|_{(a,b)} = F|_{(a,b)}$ is satisfied. Here, $a, b, F^-(0)$ and $F^-(1)$ are defined as in Remark 2.1.9. It remains to prove $(F^-)^+(x) = F(x)$ for $x \in (-\infty, a] \cup [b, \infty)$. For $x < a$, we have $F(x) = 0 = (F^-)^+(x)$, since $F^-(0) = a$. Moreover, for $x \geq b$, it holds true that $F(x) = 1$ and, since $F^-(1) = b$, $(F^-)^+(x) = 1$ holds true, too. Finally, if $a < b$, since F and $(F^-)^+$ are continuous from the right,

$$F(a) = \lim_{\substack{x \downarrow a \\ x \in (a,b)}} F(x) = \lim_{\substack{x \downarrow a \\ x \in (a,b)}} (F^-)^+(x) = (F^-)^+(a)$$

is satisfied and if $a = b$, than $F(a) = 1 = (F^-)^+(a)$ is true by definition. This completes the proof. □

Lemma 7.1.5.

Let F be a distribution function with $supp(F) \subseteq [0,1]$ and $f_1, \ldots, f_p : [0,1] \longrightarrow \mathbb{R}$, $p \in \mathbb{N}$, be functions that are of bounded variation, continuous from the left and right continuous in zero.
Then, there is a distribution function G with $supp(G) \subseteq [0,1]$ and functions $g_1, \ldots, g_p : [0,1] \longrightarrow \mathbb{R}$ that are Lipschitz continuous, such that

$$f_i = g_i \circ G^-, \quad i = 1, \ldots, p.$$

Proof of Lemma 7.1.5 The proof will be divided into the following steps:

Step 1: Definition of a quantile function Q:
Without loss of generality, we suppose $f_i(0) = 0$ for $i = 1, \ldots, p$. This is no limitation since for $f_i(0) \neq 0$ for some $i \in \{1, \ldots, p\}$, we can consider $\hat{f}_i(\cdot) = f_i(\cdot) - f_i(0)$. We then have $\hat{f}_i(0) = 0$ and $\hat{f}_i$ is continuous from the left, right continuous in zero and of bounded variation, since f_i is. Moreover, if $\hat{f}_i = \widetilde{g}_i \circ G^-$ holds true for a distribution function G and some Lipschitz continuous $\widetilde{g}_i$, we then have $f_i = g_i \circ G^-$ with $g_i(\cdot) := \widetilde{g}_i(\cdot) + f_i(0)$ and g_i Lipschitz continuous holds true, too.
Moving on, since $f_1, \ldots, f_p$ are continuous from the left, right continuous in zero and of bounded variation, there are monotonically increasing functions $f_i^{(1)}, f_i^{(2)} : [0,1] \longrightarrow \mathbb{R}$, $i = 1, \ldots, p$, that are continuous from the left and right continuous in zero with $f_i = f_i^{(1)} - f_i^{(2)}$, $i = 1, \ldots, p$. This follows from the Jordan decomposition of bounded variation functions (see for example Aufgabe 1.10 of Kapitel VII in Elstrodt (2009)). Without loss of generality, we further suppose that

$$f_i^{(1)}(0) = 0 = f_i^{(2)}(0), \quad i = 1, \ldots, p.$$

This is not a limitation, since we already know that $0 = f_i(0) = f_i^{(1)}(0) - f_i^{(2)}(0)$ and

therefore, $f_i^{(1)}(0) = f_i^{(2)}(0)$ holds true. Now, if $f_i^{(1)}(0) = f_i^{(2)}(0) \neq 0$, we simply look at

$$\hat{f}_i^{(1)} := f_i^{(1)} - f_i^{(1)}(0) \text{ and } \hat{f}_i^{(2)} = f_i^{(2)} - f_i^{(2)}(0),$$

which satisfy $\hat{f}_i^{(1)}(0) = 0 = \hat{f}_i^{(2)}(0)$, are monotonically increasing, right continuous in zero and $f_i = \hat{f}_i^{(1)} - \hat{f}_i^{(2)}$ still holds true. Because $f_i^{(1)}, f_i^{(2)}$ are monotonically increasing, their left sided limits exist on $(0,1]$ and since f_i is continuous from the left hand side, for $x \in (0,1]$,

$$f_i(x) = f_i(x-) = \lim_{x_n \uparrow x} f_i(x_n) = \lim_{x_n \uparrow x} f_i^{(1)}(x_n) - \lim_{x_n \uparrow x} f_i^{(2)}(x_n) = f_i^{(1)}(x-) - f_i^{(2)}(x-)$$

holds true. Here, as usual, $h(x-)$ denotes the left sided limit $\lim_{x_n \uparrow x} h(x_n)$ of a function h. Therefore, without loss of generality, we can assume that the functions $f_i^{(1)}, f_i^{(2)}$ are continuous from the left.
Furthermore, since for $i = 1, \ldots, p$, the functions $f_i^{(1)}, f_i^{(2)} : [0,1] \to \mathbb{R}$ are monotonically increasing and are continuous from the left, there are finite Borel measures $\mu_i^{(1)}, \mu_i^{(2)}$ on $([0,1], \mathcal{B}_{[0,1]})$ with

$$f_i^{(1)}(x) = \mu_i^{(1)}([0,x)) \text{ and } f_i^{(2)}(x) = \mu_i^{(2)}([0,x)), \quad x \in [0,1],\ i = 1, \ldots, p,$$

(see for example Bauer (1992), p. 36). Hence, we define

$$\mu_i := \mu_i^{(1)} + \mu_i^{(2)},\ i = 1, \ldots, p, \quad \text{as well as} \quad \mu := \sum_{i=1}^{p} \mu_i.$$

Note that

$$\mu_i^{(j)}([0,t)) = f_i^{(j)}(t) \geq 0, \quad i = 1, \ldots, p,\ j = 1, 2,\ t \in [0,1],$$

since $f_i^{(j)}(0) = 0$ and $f_i^{(j)}$ is monotonically increasing and thus, $\mu([0,t)) \geq 0$ holds true for $t \in [0,1]$.
Now, if $\mu([0,1)) = 0$, it follows that $\mu_i^{(j)}([0,1)) = 0$ and thus $f_i^{(j)}(1) = 0$ for all $i = 1, \ldots, p$ and $j = 1, 2$. Since, furthermore, $f_i^{(j)}(0) = 0$ and $f_i^{(j)}$ is monotonically increasing for all $i = 1, \ldots, p$ and $j = 1, 2$, it can be deduced that the initial functions $f_1, \ldots, f_p$ are all constant functions and thus Lipschitz continuous. Thus, $f_i = g_i \circ G^-$, $i = 1, \ldots, p$, holds true for $g_i = f_i$, $i = 1, \ldots, p$, and $G(x) = x \cdot \mathbb{1}_{[0,1)}(x) + \mathbb{1}_{[1,\infty)}(x)$ and therefore Lemma 7.1.5 is proved in this case.
If, in contrast, $\mu([0,1)) > 0$ holds, we proceed as follows. Define

$$Q : [0,1] \to [0,1],\ t \mapsto \frac{\mu([0,t))}{\mu([0,1))}$$

and note that

i) Q is monotonically increasing,

ii) Q is continuous from the left. To see this, consider $t \in (0,1]$ and any sequence

$(t_n)_{n\in\mathbb{N}} \subseteq [0,t)$ with $t_n \uparrow t$. Then $[0,t) = \bigcup_{n=1}^{\infty} [0,t_n)$ and thus $\mu([0,t)) = \lim_{n\to\infty} \mu([0,t_n))$ is satisfied, since μ as a measure is continuous form below. We thus can infer that

$$\lim_{n\to\infty} Q(t_n) = \frac{1}{\mu([0,1))} \lim_{n\to\infty} \mu([0,t_n)) = \frac{1}{\mu([0,1))} \cdot \mu([0,t)) = Q(t)$$

is satisfied.

iii) $Q(0) = 0$ and $Q(1) = 1$,

iv) Q is right continuous in zero, since for any sequence $(t_n)_{n\in\mathbb{N}} \subseteq (0,1]$ with $t_n \downarrow 0$, we have

$$\begin{aligned}
\lim_{n\to\infty} Q(t_n) &= \lim_{n\to\infty} \frac{\mu([0,t_n))}{\mu([0,1))} = \frac{1}{\mu([0,1))} \lim_{n\to\infty} \sum_{i=1}^{p} \mu_i([0,t_n)) \\
&= \frac{1}{\mu([0,1))} \lim_{n\to\infty} \sum_{i=1}^{p} \left(\mu_i^{(1)}([0,t_n)) + \mu_i^{(2)}([0,t_n))\right) \\
&= \frac{1}{\mu([0,1))} \sum_{i=1}^{p} \lim_{n\to\infty} \left(f_i^{(1)}(t_n) + f_i^{(2)}(t_n)\right) = \frac{1}{\mu([0,1))} \sum_{i=1}^{p} \Big(\underbrace{f_i^{(1)}(0)}_{=0} + \underbrace{f_i^{(2)}(0)}_{=0}\Big) \\
&= 0 = Q(0).
\end{aligned}$$

Parts i) - iv) prove that Q is a quantile function that is right continuous in zero.

Step 2: Definition of a distribution function G corresponding to Q.
We define

$$G : \mathbb{R} \longrightarrow [0,1],\ x \longmapsto Q^+(x).$$

Then, according to Lemma 7.1.4, G is a distribution function and $G^- = (Q^+)^- = Q$ holds true. Thus, Q is the quantile function corresponding to G. We henceforth will use the notation G^- for Q.
Furthermore, because G is a distribution function and G^- is the respective quantile function

$$G^-(G(x)) \leq x, \quad x \in [0,1], \tag{7.58}$$

holds true (see Lemma 7.1.1). Together with the fact that G and G^- are monotonically increasing, we can infer that

$$t \leq G(G^-(t)) \Rightarrow G^-(t) \leq G^-(G(G^-(t))) \overset{(7.58)}{\leq} G^-(t)$$

as well as

$$G^-(G(x)) \overset{(7.58)}{\leq} x \Rightarrow G(x) \leq G(G^-(G(x))) \leq G(x)$$

are satisfied and therefore

$$G(G^-(G(x))) = G(x) \ \forall\, x \in [0,1] \quad \text{and} \quad G^-(G(G^-(t))) = G^-(t) \ \forall\, t \in [0,1] \tag{7.59}$$

hold true.

Step 3: Construction of the functions $g_1, \ldots, g_p$:
For $i = 1, \ldots, p$, we define the functions

$$g_{i,0} : Im(G^-) \longrightarrow \mathbb{R},\ x \longmapsto f_i(G(x)). \tag{7.60}$$

Then, for $x < y \in Im(G^-)$ there are $s, t \in [0,1]$ such that

$$x = G^-(s) \ \text{ and } \ y = G^-(t)$$

and it follows that for $i = 1, \ldots, p$,

$$\begin{aligned}
&|g_{i,0}(x) - g_{i,0}(y)| \\
&\quad = |f_i(G(x)) - f_i(G(y))| = |(f_i^{(1)}(G(x)) - f_i^{(2)}(G(x))) - (f_i^{(1)}(G(y)) - f_i^{(2)}(G(y)))| \\
&\quad \leq |f_i^{(1)}(G(x)) - f_i^{(1)}(G(y))| + |f_i^{(2)}(G(x)) - f_i^{(2)}(G(y))| \\
&\quad = |\mu_i^{(1)}([0, G(x))) - \mu_i^{(1)}([0, G(y)))| + |\mu_i^{(2)}([0, G(x))) - \mu_i^{(2)}([0, G(y)))| \\
&\quad \leq \sum_{j=1}^{p} |\mu_j^{(1)}([0, G(x))) - \mu_j^{(1)}([0, G(y)))| + |\mu_j^{(2)}([0, G(x))) - \mu_j^{(2)}([0, G(y)))| \\
&\quad = \sum_{j=1}^{p} \mu_j^{(1)}([G(x), G(y))) + \mu_j^{(2)}([G(x), G(y))) = \sum_{j=1}^{p} \mu_j([G(x), G(y))) \\
&\quad = \mu([G(x), G(y))) = \mu([0,1)) \cdot \frac{\mu([G(x), G(y)))}{\mu([0,1))} \\
&\quad = \mu([0,1)) \cdot \left(\frac{\mu([0, G(y)))}{\mu([0,1))} - \frac{\mu([0, G(x)))}{\mu([0,1))} \right) = \mu([0,1)) \cdot \big(G^-(G(y)) - G^-(G(x)) \big) \\
&\quad = \mu([0,1)) \cdot \big(G^-(G(G^-(t))) - G^-(G(G^-(s))) \big) \overset{(7.59)}{=} \mu([0,1)) \cdot \big(G^-(t) - G^-(s) \big) \\
&\quad = \mu([0,1)) \cdot (y - x) = \mu([0,1)) \cdot |y - x|
\end{aligned}$$

holds true. Thus, $g_{1,0}, \ldots, g_{p,0} : Im(G^-) \to \mathbb{R}$ are Lipschitz continuous with Lipschitz-constant $\mu([0,1))$.
According to Kirszbraun's theorem (see for example Federer (1969), Theorem 2.10.43 on page 201 and the note after that) by defining

$$g_i : [0,1] \longrightarrow \mathbb{R},\ x \longmapsto \inf\{g_{i,0}(y) + \mu([0,1)) \cdot |x - y| \,|\, y \in Im(G^-)\}, \quad i = 1 \ldots, p,$$

it is possible to extend the real-valued and Lipschitz continuous functions $g_{1,0}, \ldots, g_{p,0}$ from their domain $Im(G^-) \subseteq [0,1]$ to Lipschitz continuous functions $g_1, \ldots, g_p$ on the domain $[0,1]$ in such a way that $g_i|_{Im(G^-)} = g_{i,0}$ holds true for $i = 1, \ldots, p$ and such that each pair of these functions has the same Lipschitz-constant.

Step 4: Proof of $f_i = g_i \circ G^-$:
By definition (7.60), for $t \in [0,1]$ and $i = 1, \ldots, p$,

$$g_i(G^-(t)) = g_{i,0}(G^-(t)) = f_i(G(G^-(t)))$$

holds true,.It thus remains to show that

$$f_i(t) = f_i(G(G^-(t))), \quad t \in [0,1]$$

is valid. If a $t \in [0,1]$ with $G(G^-(t)) = t$ is given, then nothing is left to prove. Otherwise, according to equation (7.37), $G(G^-(t)) \geq t$ has to be true and due to equation (7.59) for $s := G(G^-(t))$, we have

$$G^-(s) = G^-(G(G^-(t))) = G^-(t),$$

and hence

$$\frac{(\mu([0,s))}{\mu[0,1))} = G^-(s) = G^-(t) = \frac{\mu([0,t))}{(\mu[0,1))}$$

is valid. Therefore, $\mu([t,s)) = 0$. Because of the definitions of μ and $\mu_1, \ldots, \mu_p$, we then have

$$0 = \mu([t,s)) = \sum_{i=1}^{p} \mu_i([t,s)) = \sum_{i=1}^{p} \left(\mu_i^{(1)}([t,s)) + \mu_i^{(2)}([t,s))\right)$$

and due to the non-negativity and monotony of the measures

$$\mu_1^{(1)}, \mu_1^{(2)}, \mu_2^{(1)}, \mu_2^{(2)}, \ldots, \mu_p^{(1)}, \mu_p^{(2)}$$

it follows that

$$\mu_i^{(j)}([t,s)) = 0, \quad i = 1, \ldots, p, \ j = 1, 2.$$

Because of

$$\begin{aligned}
f_i(s) - f_i(t) &= (f_i^{(1)}(s) - f_i^{(2)}(s)) - (f_i^{(1)}(t) - f_i^{(2)}(t)) \\
&= \mu_i^{(1)}([0,s)) - \mu_i^{(2)}([0,s)) - \mu_i^{(1)}([0,t)) + \mu_i^{(2)}([0,t)) \\
&= \mu_i^{(1)}([0,s)) - \mu_i^{(1)}([0,t)) - \left(\mu_i^{(2)}([0,s)) - \mu_i^{(2)}([0,t))\right) \\
&= \mu_i^{(1)}([t,s)) - \mu_i^{(2)}([t,s)) \\
&= 0
\end{aligned}$$

it follows that $f_i(s) = f_i(t)$ and thus

$$f_i(G(G^-(t))) = f_i(t)$$

holds true, which completes the proof. □

Corollary 7.1.6.

Let F be a distribution function with $supp(F) \subseteq [c,d]$, $c < d$, and let $f_1, \dots, f_p : [c,d] \longrightarrow \mathbb{R}$, $p \in \mathbb{N}$, be functions of bounded variation that are continuous from the left and right continuous in c. Furthermore, let $a < b$ be real numbers.
Then, there is a measure generating function $G : \mathbb{R} \longrightarrow [c,d]$ with $supp(G) \subseteq [a,b]$ and functions $g_1, \dots, g_p : [a,b] \longrightarrow \mathbb{R}$ that are Lipschitz continuous, such that

$$f_i = g_i \circ G^-, \quad i = 1, \dots, p.$$

Proof of Corollary 7.1.6: We consider the mappings

$$\varphi : [0,1] \longrightarrow [c,d],\ x \longmapsto (d-c)x + c$$

and

$$\psi : [a,b] \longrightarrow [0,1],\ x \longmapsto \frac{x-a}{b-a}$$

and note that φ and ψ are continuous, strictly monotonically increasing and therefore are invertible with the inverse functions φ^{-1} and ψ^{-1}. Hence, Lemma 7.1.5 applies to the functions $f_1 \circ \varphi, \dots, f_p \circ \varphi : [0,1] \longrightarrow \mathbb{R}$ that are of bounded variation, continuous from the left and right continuous in zero and yields the existence of a distribution function $\widetilde{G}$ with support in $[0,1]$ and functions $\widetilde{g}_1, \dots, \widetilde{g}_p : [0,1] \longrightarrow \mathbb{R}$ that are Lipschitz continuous such that

$$f_i \circ \varphi = \widetilde{g}_i \circ \widetilde{G}^-, \quad i = 1, \dots, p,$$

holds true. We thus can infer that

$$f_i = \widetilde{g}_i \circ \widetilde{G}^- \circ \varphi^{-1} = \widetilde{g}_i \circ \psi \circ \psi^{-1} \circ \widetilde{G}^- \circ \varphi^{-1} = (g_i \circ \psi) \circ (\varphi \circ \widetilde{G} \circ \psi)^-, \quad i = 1, \dots, p,$$

holds true, since

$$\begin{aligned}
(\varphi \circ \widetilde{G} \circ \psi)^-(y) &= \inf\{x \in [a,b] \,|\, \varphi(\widetilde{G}(\psi(x))) \geq y\} \\
&= \inf\{x \in [a,b] \,|\, \widetilde{G}(\psi(x)) \geq \varphi^{-1}(y)) \qquad (7.61) \\
&= \psi^{-1}\left(\inf\{x \in \mathbb{R} \,|\, \widetilde{G}(x) \geq \varphi^{-1}(y))\right) \qquad (7.62) \\
&= \psi^{-1}(G^-(\varphi^{-1}(y))), \quad y \in [c,d],
\end{aligned}$$

is satisfied. Note that equation (7.61) is true, as φ is invertible and equation (7.62) holds true, since the mapping ψ is strictly increasing and continuous. Hence, we have

$$f_i = g_i \circ G^-, \quad i = 1, \dots, p,$$

where $g_1 := \widetilde{g}_1 \circ \psi, \dots, g_p := \widetilde{g}_p \circ \psi : [a, b] \longrightarrow \mathbb{R}$ are Lipschitz continuous and where

$$G(x) := \varphi(\widetilde{G}(\psi(a)))\mathbb{1}_{(-\infty,a)}(x) + \varphi(\widetilde{G}(\psi(x)))\mathbb{1}_{[a,b]}(x) + \varphi(\widetilde{G}(\psi(b)))\mathbb{1}_{(b,\infty)}(x) : \mathbb{R} \longrightarrow [c, d]$$

is monotonically increasing and continuous from the right and therefore a measure generating function with $supp(G) \subseteq [a, b]$. □

Lemma 7.1.7.

In the situation of Corollary 7.1.6 the following two assertion are equivalent:

i) The functions $f_1, \dots, f_p$ are linearly independent in $\mathcal{L}^2([c, d], F)$.

ii) The functions $g_1, \dots, g_p$ are linearly independent in $\mathcal{L}^2([a, b], F \circ G)$.

Proof of Lemma 7.1.7 Note that in this proof, we use objects that are defined in the proof of Corollary 7.1.6. The proof consists in verifying the following chain of equivalences.

$f_1, \dots, f_p$ are linearly independent in $\mathcal{L}^2([c, d], F)$.

$$\iff f_1 \circ \varphi, \dots, f_p \circ \varphi \text{ are linearly independent in } \mathcal{L}^2([0, 1], F \circ \varphi). \tag{7.63}$$

$$\iff \widetilde{g}_1 \circ \widetilde{G}^-, \dots, \widetilde{g}_p \circ \widetilde{G}^- \text{ are linearly independent in } \mathcal{L}^2([0, 1], F \circ \varphi). \tag{7.64}$$

$$\iff \widetilde{g}_1, \dots, \widetilde{g}_p \text{ are linearly independent in } \mathcal{L}^2([0, 1], F \circ \varphi \circ \widetilde{G}). \tag{7.65}$$

$$\iff \widetilde{g}_1 \circ \psi, \dots, \widetilde{g}_p \circ \psi \text{ are linearly independent in } \mathcal{L}^2([a, b], F \circ \varphi \circ \widetilde{G} \circ \psi). \tag{7.66}$$

$$\iff \widetilde{g}_1 \circ \psi, \dots, \widetilde{g}_p \circ \psi \text{ are linearly independent in } \mathcal{L}^2([a, b], F \circ G). \tag{7.67}$$

$$\iff g_1, \dots, g_p \text{ are linearly independent in } \mathcal{L}^2([a, b], F \circ G). \tag{7.68}$$

Here, the equivalences (7.63) and (7.66) are satisfied, since φ and ψ are bijective. equivalence (7.64) holds true, since $f_i \circ \varphi = \widetilde{g}_i \circ \widetilde{G}^-$, $i = 1, \dots, p$, and equivalence (7.67) is satisfied, since $G|_{[a,b]} = \varphi \circ \widetilde{G} \circ \psi$. Furthermore, equivalence (7.68) holds true, since $g_i := \widetilde{g}_i \circ \psi$, $i = 1, \dots, p$. It thus remains to prove equivalence (7.65). Indeed, equivalence (7.65) holds true, since

$$\widetilde{g}_1 \circ \widetilde{G}^-, \ldots, \widetilde{g}_p \circ \widetilde{G}^- \text{ are linearly independent in } \mathcal{L}^2([0,1], F \circ \varphi).$$

$$\iff \int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i (\widetilde{g}_i \circ \widetilde{G}^-) \right)^2 dF \circ \varphi > 0 \quad \forall\, \alpha \in \mathbb{R}^p \setminus \{0\}.$$

$$\iff \int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i \widetilde{g}_i \right)^2 dF \circ \varphi \circ \widetilde{G} > 0 \quad \forall\, \alpha \in \mathbb{R}^p \setminus \{0\}. \tag{7.69}$$

$$\iff \widetilde{g}_1, \ldots, \widetilde{g}_p \text{ are linearly independent in } \mathcal{L}^2([0,1], F \circ \varphi \circ \widetilde{G}).$$

Note that because of

$$\int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i \widetilde{g}_i \right)^2 dF \circ \varphi \circ \widetilde{G} = \int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i (\widetilde{g}_i \circ (F \circ \varphi \circ \widetilde{G})^-) \right)^2 d\lambda\!\!\!\lambda_{[0,1]} \tag{7.70}$$

$$= \int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i (\widetilde{g}_i \circ \widetilde{G}^- \circ (F \circ \varphi)^-) \right)^2 d\lambda\!\!\!\lambda_{[0,1]} \tag{7.71}$$

$$= \int\limits_{[0,1]} \left(\sum_{i=1}^{p} \alpha_i (\widetilde{g}_i \circ \widetilde{G}^-) \right)^2 dF \circ \varphi, \tag{7.72}$$

equivalence (7.69) holds true. Here, the equalities (7.70) and (7.72) hold true, according to Lemma 7.1.3, since all integrals exist, $F \circ \varphi \circ \widetilde{G}$ and $F \circ \varphi$ are distribution functions with support in $[0,1]$ and $\left(\sum_{i=1}^{p} \alpha_i \widetilde{g}_i \right)^2$ as well as $\left(\sum_{i=1}^{p} \alpha_i (\widetilde{g}_i \circ \widetilde{G}^-) \right)^2$ are measurable functions. Furthermore, equality (7.71) is satisfied, since

$$\begin{aligned}
(F \circ \varphi \circ \widetilde{G})^-(t) &= \inf\{x \in \mathbb{R} \,|\, F(\varphi(\widetilde{G}(x))) \geq t\} = \inf\{x \in \mathbb{R} \,|\, \varphi(\widetilde{G}(x)) \geq F^-(t)\} \\
&= \inf\{x \in \mathbb{R} \,|\, \widetilde{G}(x) \geq \varphi^-(F^-(t))\} = \widetilde{G}^-(\varphi^-(F^-(t))) \\
&= (\widetilde{G}^- \circ (F \circ \varphi)^-)(t)
\end{aligned}$$

(which follows using equation (7.38) from Lemma 7.1.1) holds true for all $t \in (0,1)$. □

Proof of Theorem 2.2.9

The proof will be divided into the following two steps:

Step 1 :
Let a linear regression model with fixed design be given. Then, we define the corresponding linear regression model with uniform fixed design by

$$\widetilde{Y}_n = \widetilde{X}_n\beta + \varepsilon_n, \quad n \in \mathbb{N},$$

where

i) $\left(\left(\frac{i}{n}\right)_{i=1}^{n}\right)_{n\in\mathbb{N}}$ is the new triangular array of equidistant design points,

ii) $\widetilde{X}_n := \left(\widetilde{f}_r\left(\frac{i}{n}\right)\right)_{i=1,r=1}^{n\quad p} := \left((f_r \circ F^-)\left(\frac{i}{n}\right)\right)_{i=1,r=1}^{n\quad p}$ is the new design matrix, wherein the transformed regression functions are

$$\widetilde{f}_1 := f_1 \circ F^-, \ldots, \widetilde{f}_p := f_p \circ F^- : [0,1] \longrightarrow \mathbb{R}.$$

Now, the proof consists in verifying the following assertions:

i) The transformed regression functions $\widetilde{f}_1, \ldots, \widetilde{f}_p : [0,1] \longrightarrow \mathbb{R}$ are

   α) of bounded variation,

   β) linear independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$.

ii) The sequence $(\overline{F}_n)_{n\in\mathbb{N}}$ of empirical distribution functions in the new design of equidistant design points converges uniformly to the distribution function of the uniform distribution on the unit interval, i.e. the distribution function corresponding to the probability measure $\lambda\!\!\!\lambda_{[0,1]}$.

iii) The residual partial sum process in the linear regression model with uniform fixed design admits a weak limit that is identical to the residual partial sum limit processes in the original model. That is, in a linear regression model with uniform fixed design, the weak convergence

$$\frac{1}{\sqrt{n}} T_n(\hat{\varepsilon}_n)(z) \xrightarrow[n\to\infty]{\mathcal{D}} B_{f,F}(z) \quad \text{in } C[0,1]$$

holds true.

As to i) α): Since the regression functions $f_1, \ldots, f_p$ are of bounded variation and the quantile function F^- is monotonically increasing, the variation of $f_i \circ F^-$ is bounded from above by the variation of f_i, $i = 1, \ldots, p$. Therefore, because $f_1, \ldots, f_p$ are of bounded variation, the transformed regression functions $\widetilde{f}_1, \ldots, \widetilde{f}_p$ are of bounded variation, too. To this regard note that

$$\begin{aligned}\|\widetilde{f}_i\|_V &= \|f \circ F^-\|_V \\ &= \sup_{\{0\le x_1\le\cdots\le x_n\le 1,\, n\in\mathbb{N}\}} \sum_{i=1}^{n-1} |f(F^-(x_{i+1})) - f(F^-(x_i))| \\ &\le \sup_{\{a\le t_1\le\cdots\le t_n\le b,\, n\in\mathbb{N}\}} \sum_{i=1}^{n-1} |f(t_{i+1}) - f(t_i)| \\ &= \|f_i\|_V, \quad i = 1,\ldots,p.\end{aligned}$$

As to i) β**)** We are given an arbitrary design distribution function F on the experimental region $[a,b]$, the original regression functions $f_1,\ldots,f_p$ that are linearly independent in $\mathcal{L}^2([a,b],F)$ and $\alpha\in\mathbb{R}^p$. In this situation due to the change of variables formula introduced in Lemma 7.1.3, we know that

$$\int\limits_{[a,b]} \left(\sum_{i=1}^p \alpha_i f_i(t)\right)^2 dF(t) = \int\limits_{[0,1]} \left(\sum_{i=1}^p \alpha_i \left(f_i \circ F^-\right)(x)\right)^2 d\lambda\!\!\!\lambda_{[0,1]}(x)$$

and therefore, the the following chain of equivalences holds:

$f_1,\ldots,f_p$ are linearly independent in $\mathcal{L}^2([a,b],F)$.

$$\iff \int\limits_{[a,b]} \left(\sum_{i=1}^p \alpha_i f_i\right)^2 dF > 0 \quad \forall \alpha\in\mathbb{R}^p\setminus\{0\}.$$

$$\iff \int\limits_{[0,1]} \left(\sum_{i=1}^p \alpha_i \left(f_i\circ F^-\right)\right)^2 d\lambda\!\!\!\lambda_{[0,1]} > 0 \quad \forall \alpha\in\mathbb{R}^p\setminus\{0\}.$$

$$\iff \widetilde{f}_1,\ldots,\widetilde{f}_p \text{ are linearly independent in } \mathcal{L}^2([0,1],\lambda\!\!\!\lambda_{[0,1]}).$$

As to ii) In the given situation, for $n\in\mathbb{N}$, the empirical distribution function is given by

$$\overline{F}_n(x) = \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{(-\infty,x]}\left(\frac{i}{n}\right), \quad x\in\mathbb{R},$$

and for the new limit distribution function we define

$$\overline{F}(u) = u\cdot\mathbb{1}_{[0,1]}(u) + \mathbb{1}_{(1,\infty)}(u), \quad u\in\mathbb{R}.$$

Note, that the measure which uniquely corresponds to the distribution function $\overline{F}$ is the Lebesgue-measure on the unit interval $\lambda\!\!\!\lambda_{[0,1]}$. A straight forward calculation shows that

$$\sup_{x\in\mathbb{R}} |\overline{F}_n(x) - \overline{F}(x)| \leq \frac{1}{n} \xrightarrow[n\to\infty]{} 0$$

and thus

$$\overline{F}_n(\cdot) \xrightarrow[n\to\infty]{} \overline{F}(\cdot) \text{ uniformly.}$$

As to iii) We defined the linear regression model with uniform fixed design corresponding to the given linear regression model with fixed design to be

$$\widetilde{Y}_n = \left((f_r \circ F^-) \left(\frac{i}{n} \right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N}.$$

We can rewrite this as the regression model

$$\widetilde{Y}_n = \left(f_r \left(F^- \left(\frac{i}{n} \right) \right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{7.73}$$

where $f_1, \ldots, f_p : [a,b] \to \mathbb{R}$ are the regression functions of the original linear regression model with fixed design and where

$$\left(F^- \left(\frac{i}{n} \right) \right)_{i=1}^{n} \subseteq [a,b]$$

is the triangular array of design points. Note the following two things:

- The triangular array $\left(F^- \left(\frac{i}{n}\right)\right)_{i=1}^n$ of design points in model (7.73) does not have to be the same as the triangular array $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of the original linear regression model with fixed design (see model (2.18)).
- The sequence of empirical distribution functions $(\widetilde{F}_n)_{n\in\mathbb{N}}$ corresponding to the triangular array $\left(F^- \left(\frac{i}{n}\right)\right)_{i=1}^n$ converges uniformly to the distribution function F. This can be verified considering the following straight forward calculation.

$$\begin{aligned}
\sup_{x\in\mathbb{R}} |\widetilde{F}_n(x) - F(x)| &= \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]} \left(F^{-1} \left(\frac{i}{n} \right) \right) - F(x) \right| \\
&= \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,nF(x)]}(i) - F(x) \right| = \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \left(\lfloor nF(x) \rfloor - nF(x) \right) \right| \\
&\leq \sup_{x\in\mathbb{R}} \left| \frac{1}{n} \right| = \frac{1}{n} \xrightarrow[n\to\infty]{} 0.
\end{aligned} \tag{7.74}$$

Thus, model (7.73) is a linear regression model with fixed design. Therefore, Theorem 2.2.4 is applicable, stating that the residual partial sum limit process of the regression residuals of model (7.73) converges weakly to $B_{f,F}$ in $C[0,1]$. Hence, the residual partial sum limit process is the same as in the original model (i.e. model (2.18)). This completes Step 1 of the proof.

Step 2 : Let a linear regression model

$$\widetilde{Y}_n = \left(\tilde{f}_r \left(\frac{i}{n} \right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{7.75}$$

with uniform fixed design be given. Note that this means in particular that the regression functions $\widetilde{f}_1, \ldots, \widetilde{f}_p : [0,1] \to \mathbb{R}$ are of bounded variation and linearly independent in $\mathcal{L}^2([0,1], \lambda\!\!\!\lambda_{[0,1]})$.
Without loss of generality we can assume that the functions $\widetilde{f}_1, \ldots, \widetilde{f}_p$ are continuous from the left and right continuous in zero. This is indeed possible, since it is assumed that $\widetilde{f}_1, \ldots, \widetilde{f}_p$ are of bounded variation. Therefore, $\widetilde{f}_1, \ldots, \widetilde{f}_p$ can have at most countably many discontinuity points, and moreover, limits from the left and limits from the right do always exist. As a reference for this, see Korollar 4.6 or Aufgabe 1.10 in Kapitel VII of Elstrodt (2009). Hence, we can change each of the functions $\widetilde{f}_1, \ldots, \widetilde{f}_p$ on a measure zero set such that the values at the discontinuity points in $(0,1]$ always agree with the left hand limit and such that the value at zero agrees with the right hand limit. This has no influence on the distribution of the residual partial sum limit processes.
Therefore, according to Lemma 2.2.7, there is a distribution function F with $supp(F) \subseteq [a,b]$ and functions $f_1, \ldots, f_p : [a,b] \longrightarrow \mathbb{R}$ that are Lipschitz continuous and linearly independent in $\mathcal{L}^2([a,b], F)$ such that

$$\widetilde{f}_i \; = \; f_i \circ F^-, \quad i = 1, \ldots, p,$$

holds true. We can thus define a corresponding linear regression model with fixed design

$$\widetilde{Y}_n = \left(f_r \left(F^- \left(\frac{i}{n} \right) \right) \right)_{i=1,r=1}^{n \quad p} \beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{7.76}$$

that is equal to the given linear regression model with uniform fixed design, see model (7.75). Note that in model (7.76)

$$((t_{ni})_{i=1}^n)_{n\in\mathbb{N}} := \left(\left(F^- \left(\frac{i}{n} \right) \right)_{i=1}^n \right)_{n\in\mathbb{N}}$$

is the triangular array of design points. Furthermore, similarly to (7.74), one can verify that the corresponding sequence of empirical distribution functions

$$F_n(x) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(t_{ni}), \quad x \in \mathbb{R}, \, n \in \mathbb{N},$$

satisfies the condition

$$F_n \xrightarrow[n\to\infty]{} F \quad \text{uniformly.}$$

Therefore Theorem 2.2.4 is applicable, stating that the residual partial sum limit process of the regression residuals of model (7.76) converges weakly to $B_{f,F}$ in $C[0,1]$. Furthermore, since we know that the regression models (7.75) and (7.74) are equal, their residual partial limit processes coincide. This completes the proof of Theorem 2.2.9. □

Proof of Lemma 2.2.10

The proof consists in verifying the following chain of equations:

$$B_{f\circ F^-,\lambda\!\!\!\lambda_{[0,1]}}(z) = B(z) - \left(\int_{[0,z]} f(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t)\right)^T \left(\int_{[0,1]} f(F^-(t)) f^T(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t)\right)^{-1} \times$$

$$\times \int_{[0,1]} f(F^-(t))\, dB(t)$$

$$= B(z) - \left(\int_{[a,b]} f(t)\, dF_z(t)\right)^T \left(\int_{[a,b]} f(t) f^T(t)\, dF(t)\right)^{-1} \int_{[0,1]} f(F^-(t))\, dB(t) \tag{7.77}$$

$$= B(z) - \left(\int_{[a,b]} f(t)\, dF_z(t)\right)^T \left(\int_{[a,b]} f(t) f^T(t)\, dF(t)\right)^{-1} \int_{[a,b]} f(t)\, dB(F(t)) \tag{7.78}$$

$$= B_{f,F}(z), \quad z \in [0,1].$$

Here, (7.77) is true, since, according to Lemma 7.1.3

$$\int_{[0,1]} f(F^-(t)) f^T(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t) = \int_{[a,b]} f(t) f^T(t)\, dF(t)$$

is satisfied and furthermore,

$$\int_{[a,b]} f(t)\, dF_z(t) = \lim_{x\uparrow F^-(z)} \int_{[a,x]} f(t)\, dF(t) + f(F^-(z)) \cdot \left(z - \lim_{x\uparrow F^-(z)} F(x)\right)$$

$$= \lim_{x\uparrow F^-(z)} \int_{[0,F(x)]} f(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t) + f(F^-(z)) \left(z - \lim_{x\uparrow F^-(z)} F(x)\right) \tag{7.79}$$

$$= \int_{[0,z]} f(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t) \tag{7.80}$$

holds true. Here, we again used Lemma 7.1.3 in equation (7.79), which is applicable since the regression functions $f = (f_1, \ldots, f_p)$ are measurable. To see that (7.80) holds true, we first consider the case that F is continuous at $F^-(z)$. We thus have $z \in Im(F)$ and $\lim_{x\uparrow F^-(z)} F(x) = F(F^-(z)) = z$ holds true. Hence, we can conclude

$$\lim_{x\uparrow F^-(z)} \int_{[0,F(x)]} f(F^-(t))\, d\lambda\!\!\!\lambda_{[0,1]}(t) + f(F^-(z)) \left(z - \lim_{x\uparrow F^-(z)} F(x)\right)$$

$$= \int\limits_{[0,z]} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t) + 0$$

$$= \int\limits_{[0,z]} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t).$$

In the case that F is discontinuous at $F^-(z)$, we can infer that F^- is constant in the interval $\left(\lim\limits_{x\uparrow F^-(z)} F(x), F(F^-(z))\right]$ and thus

$$\lim_{x\uparrow F^-(z)} \int\limits_{[0,F(x)]} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t) + f(F^-(z)) \left(z - \lim_{x\uparrow F^-(z)} F(x)\right)$$

$$= \int\limits_{\left[0,\ \lim\limits_{x\uparrow F^-(z)} F(x)\right)} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t) + \int\limits_{\left[\lim\limits_{x\uparrow F^-(z)} F(x), z\right]} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t)$$

$$= \int\limits_{[0,z]} f(F^-(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t)$$

holds true.

It remains to show that equation (7.78) is satisfied, which amounts to checking

$$\int\limits_{[0,1]} f(F^-(t))\, dB(t) = \int\limits_{[a,b]} f(t)\, dB(F(t)).$$

Transforming the measure integral into a Riemann-Stieltjes integral using the construction in Remark 2.2.3 (see equation (2.9)) and then applying integration by parts to the Riemann-Stieltjes integral on the left hand side, we obtain

$$\int\limits_{[0,1]} f(F^-(t))\, dB(t) = \overset{(R*)}{\int\limits_{[0,1]}} f(F^-(t))\, dB(t) = \overset{(R)}{\int\limits_{[-\infty,1]}} (\overline{f \circ F^-})(t)\, d\widetilde{B}(t) = \overset{(R)}{\int\limits_{[0,1]}} (f \circ F^-)(t)\, dB(t)$$

$$= f(F^-(1))B(1) - f(F^-(0))B(0) - \overset{(R)}{\int\limits_{[0,1]}} B(t)\, d(f \circ F^-)(t)$$

$$= f(F^-(1))B(1) - \overset{(R)}{\int\limits_{[0,1]}} B(t)\, d(f \circ F^-)(t). \tag{7.81}$$

Performing the same operations on the right hand side, we get

$$\begin{aligned}
\int_{[a,b]} f(t)\,dB(F(t)) &= \overset{(R*)}{\int_{[a,b]}} f(t)\,dB(F(t)) = \overset{(R)}{\int_{[-\infty,b]}} \overline{f}(t)\,d(\widetilde{B\circ F})(t) \\
&= f(a)B(F(a)) + \overset{(R)}{\int_{[a,b]}} f(t)\,d(B\circ F)(t) \\
&= f(a)B(F(a)) + f(b)B(F(b)) - f(a)B(F(a)) - \overset{(R)}{\int_{[a,b]}} B(F(t))\,df(t) \\
&= f(b)B(1) - \overset{(R)}{\int_{[a,b]}} B(F(t))\,df(t). \qquad (7.82)
\end{aligned}$$

Considering (7.81) and (7.82), two differences seem to be noticeable:

The first difference seems to be between $f(F^-(1))$ and $f(b)$. In practice, this difference does not exist, since, as already noted after Theorem 2.2.4, the experimental region $[a,b]$ can always be chosen in such a way that the last design point is b and therefore, $F^-(1) = b$ follows. The second difference seems to be between the integrals

$$\overset{(R)}{\int_{[0,1]}} B(t)\,d(f\circ F^-)(t) \quad \text{and} \quad \overset{(R)}{\int_{[a,b]}} B(F(t))\,df(t). \qquad (7.83)$$

We are going to prove their equality by first noticing that f is continuous and of bounded variation and therefore according to the Jordan decomposition there are functions $f^+, f^- : [a,b] \longrightarrow \mathbb{R}$ that are continuous and monotonically increasing such that $f = f^+ - f^-$. Therefore, we have

$$\overset{(R)}{\int_{[a,b]}} B(F(t))\,df(t) = \overset{(R)}{\int_{[a,b]}} B(F(t))\,df^+(t) - \overset{(R)}{\int_{[a,b]}} B(F(t))\,df^-(t)$$

and, since f^+ and f^- are continuous and monotonically increasing, one can define unique measures μ_{f^+} and μ_{f^-} on $\left([a,b], \mathcal{B}_{[a,b]}\right)$ (see for instance Bauer (1992) on page 36) such that

$$\overset{(R)}{\int_{[a,b]}} B(F(t))\,df^+(t) = \int_{(a,b]} B(F(t))\,d\mu_{f^+}(t) \quad \text{and} \quad \overset{(R)}{\int_{[a,b]}} B(F(t))\,df^-(t) = \int_{(a,b]} B(F(t))\,d\mu_{f^-}(t).$$

Altogether, we therefore have

$$\overset{(R)}{\int\limits_{[a,b]}} B(F(t))\, df(t) = \int\limits_{(a,b]} B(F(t))\, d\mu_f(t), \tag{7.84}$$

where μ_f is a signed measure with the Jordan decomposition $\mu_f := \mu_{f^+} - \mu_{f^-}$. For some theory on the Jordan decomposition of measures see Kapitel VII in Elstrodt (2009) and especially Theorem 1.12. To the integral on the right hand side of (7.84) the change of variables formula for measure integrals applies and we therefore can infer that

$$\int\limits_{(a,b]} B(F(t))\, d\mu_f(t) = \int\limits_{(0,1]} B(t)\, d\mu_f^F(t)$$

holds true. Here, μ_f^F denotes the pushforward measure of μ_f with respect to the measurable mapping F.
Similar to the above, when considering the left integral in (7.83) it becomes apparent that

$$\overset{(R)}{\int\limits_{[0,1]}} B(t)\, d(f \circ F^-)(t) = \int\limits_{(0,1]} B(t)\, d\mu_{(f\circ F^-)}(t)$$

holds true. Here, $\mu_{(f\circ F^-)}$ denotes the unique signed measure on $\big([0,1], \mathcal{B}_{[0,1]}\big)$ corresponding to $(f \circ F^-)^+ - (f \circ F^-)^-$. Note that the Riemann-Stieltjes integral on the left hand side exists, since the paths of $B(\cdot)$ are continuous and $f \circ F^-$ is of bounded variation and that it coincides with the measure integral on the right hand side according to Theorem 5.1.2 in Strook (1994). Furthermore, since for $0 \leq c < d \leq 1$,

$$\begin{aligned}
\mu_{(f\circ F^-)}([c,d)) &= (f \circ F^-)(d) - (f \circ F^-)(c) \\
&= \mu_f([F^-(c), F^-(d))) \\
&= \mu_f\left(F^{-1}([c,d))\right) \qquad (7.85) \\
&= \mu_f^F([c,d))
\end{aligned}$$

holds true, we know that $\mu_{(f\circ F^-)} = \mu_f^F$. Here, in (7.85) we used equivalence (7.38) of Lemma 7.1.1.

Combining the above facts, we thus have

$$\begin{aligned}
\overset{(R)}{\int\limits_{[0,1]}} B(t)\, df(F^-(t)) &= \int\limits_{(0,1]} B(t)\, d\mu_{(f\circ F^-)}(t) \\
&= \int\limits_{(0,1]} B(t)\, d\mu_f^F(t)
\end{aligned}$$

$$= \int\limits_{(a,b]} B(F(t))\, d\mu_f(t)$$

$$\overset{(7.84)}{=} \int\limits_{[a,b]}^{(R)} B(F(t))\, df(t).$$

This completes the proof. □

Proof of Theorem 2.2.11

Without loss of generality, we can assume that the functions $f_1, \ldots, f_p$ are continuous from the left and right continuous in a. This is indeed possible, since it is assumed that $f_1, \ldots, f_p$ are of bounded variation. Therefore, $f_1, \ldots, f_p$ can have at most countably many discontinuity points, and moreover, limits from the left and limits from the right do always exist. As a reference for this, see Korollar 4.6 or Aufgabe 1.10 in Kapitel VII of Elstrodt (2009). Hence, we can change each of the functions $f_1, \ldots, f_p$ on a measure zero set such that the values at the discontinuity points in $(a, b]$ always agree with the left hand limit and such that the value at a agrees with the right hand limit. This has no influence on the distribution of the residual partial sum limit processes.
Therefore, according to Lemma 2.2.7, in the situation of model (2.20), there is a measure generating function $G : \mathbb{R} \longrightarrow [a, b]$ with $supp(G) \subseteq [a, b]$ and functions $g_1, \ldots, g_p : [a, b] \longrightarrow \mathbb{R}$ that are Lipschitz continuous and linearly independent in $\mathcal{L}^2([a, b], F \circ G)$ such that

$$f_i \;=\; g_i \circ G^-, \quad i = 1, \ldots, p,$$

holds true. Thus, the regression model (2.20) is equal to the regression model

$$Y_n = \widetilde{X}_n \beta + \varepsilon_n, \quad n \in \mathbb{N}, \tag{7.86}$$

where the design matrix reads $\widetilde{X}_n = \left(g_r\left(\widetilde{t}_{ni}\right)\right)_{i=1,r=1}^{n\quad p}$, $\widetilde{t}_{ni} := G^-(t_{ni})$, $i = 1, \ldots, n$, and $g_1, \ldots, g_p$ are known and real-valued regression functions that are Lipschitz continuous and linearly independent in $\mathcal{L}^2([a, b], F \circ G)$. Note that $\widetilde{X}_n$ and X_n are different notations for the same design matrix. Furthermore,

$$\widetilde{F}_n(\cdot) := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,\cdot]}(\widetilde{t}_{ni}) = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,\cdot]}(G^-(t_{ni}))$$

$$= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,G(\cdot)]}(t_{ni}) \;\xrightarrow[n\to\infty]{}\; F(G(\cdot)) \quad \text{uniformly,}$$

where $F \circ G$ is the distribution function of the design in the transformed model. Now,

according to Theorem 2.2.2, in model (7.86) the weak convergence

$$\frac{1}{\sigma\sqrt{n}} T_n(\hat{\varepsilon}_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B_{g,F\circ G}(\cdot) \quad \text{in } C[0,1]$$

holds true. Here, $B_{g,F\circ G}(\cdot)$ is a Gaussian process that is given by

$$B_{g,F\circ G}(z) = B(z) + \left(\int_{[a,b]} g(t)\, d(F\circ G)_z(t) \right)^T \left(\int_{[a,b]} g(t) g^T(t)\, d(F\circ G)(t) \right)^{-1} \times$$
$$\times \left(\int_{[a,b]}^{(R)} B(F(G(t)))\, dg(t) - B(1) g((F\circ G)^-(1)) \right), \quad z \in [0,1].$$

Finally, we can conclude that

$$B_{g,F\circ G}(\cdot) = B_{g\circ (F\circ G)^-, \lambda\!\!\!\lambda_{[0,1]}}(\cdot) \tag{7.87}$$
$$= B_{g\circ G^-\circ F^-, \lambda\!\!\!\lambda_{[0,1]}}(\cdot) \tag{7.88}$$
$$= B_{f\circ F^-, \lambda\!\!\!\lambda_{[0,1]}}(\cdot) \tag{7.89}$$
$$= B_{f,F}(\cdot) \tag{7.90}$$

holds true. In this chain of equations, equation (7.87) is satisfied, since $F \circ G$ is a distribution function and thus Lemma 2.2.10 is applicable. Equation (7.88) holds true, since

$$\begin{aligned}(F\circ G)^-(t) &= \inf\{x \in \mathbb{R} : F(G(x)) \geq t\} \\ &= \inf\{x \in \mathbb{R} : G(x) \geq F^-(t)\} \\ &= G^-(F^-(t))\end{aligned}$$

holds true for all $t \in (0,1)$. Here, we used equation (7.38) from Lemma 7.1.1. Finally, equation (7.89) is satisfied, since $f_i = g_i \circ G^-$, $i = 1, \ldots, p$, and equation (7.90) holds true, according to Lemma 2.2.10. This completes the proof. □

7.2 Proofs for Chapter 3

The Glivenko-Cantelli theorem is a classical result in the theory of probability. We present a proof of it here for the sake of completeness.

Theorem 7.2.1. Glivenko-Cantelli

Let F be a distribution function and $X_1, X_2, \dots$ be a sequence of random variables with

$$X_1, X_2, \dots \overset{\text{iid}}{\sim} F.$$

Then the sequence of empirical distribution functions $(F_n)_{n\in\mathbb{N}}$, with

$$F_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i), \quad x \in \mathbb{R},\ n \in \mathbb{N},$$

converges uniformly to F almost surely, that is

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to\infty]{\text{a.s.}} 0\,.$$

Proof: The proof of this theorem is going to be separated in three parts.
Part 1: Because the sequence $(X_i)_{i\in\mathbb{N}}$ is an iid sequence of random variables, we can conclude that for $x \in \mathbb{R}$,

$$\left(\mathbb{1}_{(-\infty,x]}(X_i)\right)_{i\in\mathbb{N}}$$

is a iid sequence of random variables, too. Furthermore, we have

$$\mathbb{E}\left(\mathbb{1}_{(-\infty,x]}(X_1)\right) = \mathbb{P}(X_1 \le x) = F(x)$$

and therefore, the strong law of large numbers yields

$$F_n(x) = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i) \xrightarrow[n\to\infty]{\text{a.s.}} \mathbb{E}\left(\mathbb{1}_{(-\infty,x]}(X_1)\right) = F(x), \quad x \in \mathbb{R}.$$

Part 2: For

$$U_1, U_2, \dots \overset{\text{iid}}{\sim} U([0,1])$$

and

$$\overline{F}_n(x) := \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(U_i), \quad x \in (-\infty,\infty),\ n \in \mathbb{N},$$

the following statement holds true: For all $m \in \mathbb{N}$ and for all $t \in [0,1)$ there is a $k \in \{1,\dots,m\}$ such that

$$\overline{F}_n(t) - t \le \overline{F}_n\left(\frac{k}{m}\right) - \frac{k}{m} + \frac{1}{m} \quad \text{or} \quad t - \overline{F}_n(t) \le \frac{k-1}{m} - \overline{F}_n\left(\frac{k-1}{m}\right) + \frac{1}{m}$$

are satisfied. As a consequence,

$$\sup_{t\in[0,1]} |\overline{F}_n(t) - t| \;\leq\; \max_{0\leq k\leq m} \left| \overline{F}_n\left(\frac{k}{m}\right) - \frac{k}{m} \right| + \frac{1}{m}, \quad m \in \mathbb{N},$$

is satisfied and because of Part 1, for $n \longrightarrow \infty$ and subsequently $m \longrightarrow \infty$,

$$\sup_{t\in[0,1]} |\overline{F}_n(t) - t| \;\xrightarrow[n\to\infty]{\text{a.s.}}\; 0\,.$$

holds true.

Part 3: Since $X_1, X_2, \ldots \overset{\text{iid}}{\sim} F$ and due to the quantile transformation (see for example Chapter 21 in Vaart (1998)), we know that there is a sequence of random variables $U_1, U_2, \ldots \overset{\text{iid}}{\sim} U([0,1])$ with

$$X_i \sim F^-(U_i), \;\; i \in \mathbb{N}.$$

Therefore, for $x \in \mathbb{R}$, we have

$$\begin{aligned} F_n(x) &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(X_i) \\ &\overset{\mathcal{D}}{=} \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(F^-(U_i)) \\ &= \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,F(x)]}(U_i) \\ &= \overline{F}_n(F(x)) \end{aligned}$$

and thus

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \overset{\mathcal{D}}{=} \sup_{x\in\mathbb{R}} |\overline{F}_n(F(x)) - F(x)|.$$

Due to Part 2 we can conclude that

$$\sup_{x\in\mathbb{R}} |\overline{F}_n(F(x)) - F(x)| \leq \sup_{t\in[0,1]} |\overline{F}_n(t) - t| \xrightarrow[n\to\infty]{\text{a.s.}} 0$$

holds true and thus,

$$\sup_{x\in\mathbb{R}} |F_n(x) - F(x)| \xrightarrow[n\to\infty]{\text{a.s.}} 0.$$

This completes the proof. □

Note that the fact that the statement of the Glivenko-Cantelli theorem holds true "only" almost surely does not imply any restrictions to statements or analyses about distributional properties like weak convergence, since zero sets are irrelevant in statistical applications and weak limits of stochastic processes are only almost surely unique.

7.3 Proofs for Chapter 4

Proof of Theorem 4.1.3

Obviously, $H(R)$ is a vector space of functions $f : T \to \mathbb{R}$ and $S(R) \subseteq H(R)$, because for $f \in S(R)$, we can choose the constant sequence $(f_n(\cdot))_{n\in\mathbb{N}} := (f)_{n\in\mathbb{N}}$ which is Cauchy in $S(R)$ and converges pointwise to f. We now define a form denoted by $\langle\cdot,\cdot\rangle_{H(R)}$ on $H(R) \times H(R)$ via

$$\langle\cdot,\cdot\rangle_{H(R)} : H(R)\times H(R) \longrightarrow \mathbb{R},\ (f,g) \longmapsto \langle f,g\rangle_{H(R)} = \lim_{n\to\infty}\langle f_n, g_n\rangle_{S(R)}$$

for Cauchy sequences $(f_n(\cdot))_{n\in\mathbb{N}}$ and $(g_n(\cdot))_{n\in\mathbb{N}}$ in $S(R)$ with

$$f(t) = \lim_{n\to\infty} f_n(t) \text{ pointwise for all } t \in T$$

and

$$g(t) = \lim_{n\to\infty} g_n(t) \text{ pointwise for all } t \in T.$$

In order to proof that $\langle\cdot,\cdot\rangle_{H(R)}$ is well defined, we have to show that $\big(\langle f_n, g_n\rangle_{S(R)}\big)_{n\in\mathbb{N}}$ converges and that $\langle f,g\rangle_{H(R)}$ is independent of the Cauchy sequences converging pointwise to f and g respectively.
Proving the first assertion, we begin by noting that $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are Cauchy sequences and therefore bounded, thus there is a $C > 0$ such that $\|f_n\|_{S(R)} < C$ and $\|g_n\|_{S(R)} < C$ for all $n \in \mathbb{N}$. Hence,

$$\begin{aligned}
|\langle f_m, g_m\rangle_{S(R)} - \langle f_n, g_n\rangle_{S(R)}| &= |\langle f_m - f_n, g_m\rangle_{S(R)} + \langle f_n, g_m - g_n\rangle_{S(R)}| \\
&\leq \|g_m\|_{S(R)} \cdot \|f_m - f_n\|_{S(R)} + \|f_n\|_{S(R)} \cdot \|g_m - g_n\|_{S(R)} \\
&\leq C \cdot \big(\|f_m - f_n\|_{S(R)} + \|g_m - g_n\|_{S(R)}\big) \\
&\xrightarrow[m,n\to\infty]{} 0,
\end{aligned}$$

which means that $\big(\langle f_n, g_n\rangle_{S(R)}\big)$ is a Cauchy sequence in $\mathbb{R}$ and therefore convergent, since $\mathbb{R}$ is complete. Note that in the inequality above the Cauchy-Schwarz inequality was used, which has been proven for the inner product $\langle\cdot,\cdot\rangle_{S(R)}$ in Step 3 of Section 4.1.
To prove the second assertion, let $(\widetilde{f}_n))_{n\in\mathbb{N}}$ and $(\widetilde{g}_n)_{n\in\mathbb{N}}$ be another — possibly different — pair of Cauchy sequences in $S(R)$ with

$$f(t) = \lim_{n\to\infty} \widetilde{f}_n(t) \text{ pointwise for all } t \in T$$

and

$$g(t) = \lim_{n\to\infty} \widetilde{g}_n(t) \text{ pointwise for all } t \in T.$$

We then have to show that

$$\lim_{n\to\infty} \langle f_n(\cdot), g_n(\cdot)\rangle_{S(R)} = \lim_{n\to\infty} \langle \widetilde{f}_n(\cdot), \widetilde{g}_n(\cdot)\rangle_{S(R)}.$$

Because $(f_n)_{n\in\mathbb{N}}$ and $(\widetilde{f}_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $S(R)$, $(f_n - \widetilde{f}_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $S(R)$, which, for $t \in T$, converges to zero pointwise. The same is true for $(g_n - \widetilde{g}_n)_{n\in\mathbb{N}}$. Since $(f_n - \widetilde{f}_n)_{n\in\mathbb{N}}$ is a Cauchy sequence, it is bounded and therefore there is a $C > 0$ such that $\|f_n - \widetilde{f}_n\|_{S(R)} \leq C$ for all $n \in \mathbb{N}$. Hence, for an arbitrary $\varepsilon > 0$ there is a $N \in \mathbb{N}$ such that

$$\|(f_n - \widetilde{f}_n) - (f_m - \widetilde{f}_m)\|_{S(R)} \leq \frac{\varepsilon}{2C} \quad \forall\, n, m \geq N.$$

Furthermore, since $f_N - \widetilde{f}_N \in S(R)$, there is a representation

$$(f_N - \widetilde{f}_N)(\cdot) = \sum_{i=1}^{l} a_i\, R(s_i, \cdot) \quad \text{with } l \in \mathbb{N},\ a_1, \ldots, a_l \in \mathbb{R} \text{ and } s_1, \ldots, s_l \in T.$$

Because we know that $(f_n - \widetilde{f}_n)(t) \xrightarrow[n\to\infty]{} 0$ is satisfied for all $t \in T$ there is a $N' \in \mathbb{N}$ such that $\left| \sum_{i=1}^{l} a_i (f_n - \widetilde{f}_n)(s_i) \right| \leq \frac{\varepsilon}{2}$ for all $n \geq N'$. For $n \geq \max\{N, N'\}$, we now can conclude that

$$\begin{aligned}
\|f_n - \widetilde{f}_n\|^2_{S(R)} &= \langle f_n - \widetilde{f}_n, f_n - \widetilde{f}_n\rangle_{S(R)} \\
&= \langle (f_n - \widetilde{f}_n) - (f_N - \widetilde{f}_N), f_n - \widetilde{f}_n\rangle_{S(R)} + \langle (f_N - \widetilde{f}_N), f_n - \widetilde{f}_n\rangle_{S(R)} \\
&\leq \|(f_n - \widetilde{f}_n) - (f_N - \widetilde{f}_N)\|^2_{S(R)} \cdot \|f_n - \widetilde{f}_n\|_{S(R)} + \left\langle \sum_{i=1}^{l} a_i\, R(s_i, \cdot), (f_n - \widetilde{f}_n)(\cdot) \right\rangle_{S(R)} \\
&\leq \frac{\varepsilon}{2C} \cdot C + \sum_{i=1}^{l} a_i\, (f_n - \widetilde{f}_n)(s_i) \\
&\leq \varepsilon
\end{aligned}$$

holds true, due to the Cauchy-Schwarz inequality and the reproducing kernel property. Thus, $\|f_n - \widetilde{f}_n\|_{S(R)} \xrightarrow[n\to\infty]{} 0$ and the same applies for $\|g_n - \widetilde{g}_n\|_{S(R)}$. Therefore,

$$\begin{aligned}
|\langle f_n, g_n\rangle_{S(R)} - \langle \widetilde{f}_n, \widetilde{g}_n\rangle_{S(R)}| &= |\langle f_n, g_n - \widetilde{g}_n\rangle_{S(R)} + \langle f_n - \widetilde{f}_n, \widetilde{g}_n\rangle_{S(R)}| \\
&\leq \|f_n\|_{S(R)} \cdot \|f_n - \widetilde{g}_n\|_{S(R)} + \|\widetilde{g}_n\|_{S(R)} \cdot \|f_n - \widetilde{f}_n\|_{S(R)} \\
&\xrightarrow[n\to\infty]{} 0
\end{aligned}$$

follows, since again $\|f_n\|_{S(R)}$ and $\|\widetilde{g}_n\|_{S(R)}$ are bounded, as $(f_n)_{n\in\mathbb{N}}$ and $(g_n)_{n\in\mathbb{N}}$ are Cauchy sequences in $S(R)$. Thus, the value of $\langle f, g\rangle_{H(R)}$ does not depend on the choice of the pointwise convergent Cauchy sequences. Since we now know that $\langle\cdot,\cdot\rangle_{H(R)}$ is well-defined

the following properties are easy to be verify:

i) $\langle \cdot, \cdot \rangle_{H(R)}$ is biliniar.

ii) $\langle \cdot, \cdot \rangle_{H(R)}\big|_{S(R)} = \langle \cdot, \cdot \rangle_{S(R)}$.

iii) The reproducing kernel property on $H(R)$ with respect to $\langle \cdot, \cdot \rangle_{H(R)}$ holds true. This holds true, since for $f \in H(R)$

$$\begin{aligned} \langle f(\cdot), R(t, \cdot) \rangle_{H(R)} &= \lim_{n \to \infty} \langle f_n(\cdot), R(t, \cdot) \rangle_{S(R)} \\ &= \lim_{n \to \infty} f_n(t) = f(t) \qquad \forall\, t \in T\,. \end{aligned}$$

iv) For $f \in H(R)$, we have

$$\langle f, f \rangle_{H(R)} = \lim_{n \to \infty} \langle f_n, f_n \rangle_{S(R)} \geq 0$$

and therefore $\langle \cdot, \cdot \rangle_{H(R)}$ is non-negative definite. Thus, the Cauchy-Schwarz inequality holds true for $\langle \cdot, \cdot \rangle_{H(R)}$ and $||f||_{H(R)} := \sqrt{\langle f, f \rangle_{H(R)}}$ for all $f \in H(R)$. Hence, due to

$$|f(t)| = |\langle f(\cdot), R(t, \cdot) \rangle_{H(R)}| \leq \|f\|_{H(R)} \cdot \|R(t, \cdot)\|_{H(R)} \quad \forall\, t \in T,$$

we see that

$$\|f\|_{H(R)} = 0 \iff f \equiv 0\,.$$

Taking the above considerations together, we know that $\langle \cdot, \cdot \rangle_{H(R)}$ is an inner product on $H(R)$. It remains to show that $H(R)$ is complete to finish the proof of Theorem 4.1.3. We start with a function $f \in H(R)$ and a Cauchy sequence $(f_n)_{n \in \mathbb{N}} \subseteq S(R)$ with $f = \lim_{n \to \infty} f_n$ pointwise on T. For an arbitrary $\varepsilon > 0$, we choose $N \in \mathbb{N}$ big enough such that for all $n, m \geq N$, $\|f_n - f_m\|_{S(R)} \leq \varepsilon$. Thus,

$$\|f - f_N\|_{H(R)} = \lim_{n \to \infty} \|f_n - f_N\|_{S(R)} \leq \varepsilon,$$

since $(f_n - f_N)_{n \in \mathbb{N}}$ is a Cauchy sequence in $S(R)$ that converges pointwise to $f - f_N \in H(R)$ on T. As a consequence,

$$\|f - f_n\|_{H(R)} \xrightarrow[n \to \infty]{} 0$$

holds. We have thus shown that $S(R)$ is dense in $H(R)$. To prove completeness of $H(R)$, let $(f_n)_{n \in \mathbb{N}}$ be a Cauchy sequence in $H(R)$. We have to show that there is a $f \in H(R)$ with $\|f - f_n\|_{H(R)} \xrightarrow[n \to \infty]{} 0$. In order to do so, and because $S(R)$ is dense in $H(R)$ for each $n \in \mathbb{N}$, we choose $g_n \in S(R)$ with $\|f_n - g_n\|_{H(R)} \leq \frac{1}{n}$. Thus,

$$\begin{aligned} \|g_n - g_m\|_{S(R)} &= \|g_n - g_m\|_{H(R)} \\ &\leq \|g_n - f_n\|_{H(R)} + \|f_n - f_m\|_{H(R)} + \|f_m - g_m\|_{H(R)} \\ &\leq \frac{1}{n} + \|f_n - f_m\|_{H(R)} + \frac{1}{m} \xrightarrow[n,m \to \infty]{} 0 \end{aligned}$$

can be inferred, which shows that $(g_n)_{n\in\mathbb{N}}$ is a Cauchy sequence in $S(R)$ and due to Step 4 of Section 4.1, we therefore know that $(g_n)_{n\in\mathbb{N}}$ is pointwise convergent on T. Finally, let f denote the pointwise limit of $(g_n)_{n\in\mathbb{N}}$. Due to the definition of $H(R)$, we have $f \in H(R)$ and due to the definition of $\|\cdot\|_{H(R)}$ and $\langle\cdot,\cdot\rangle_{H(R)}$, we have

$$\|f - g_n\|_{H(R)} = \sqrt{\langle f, g_n\rangle_{H(R)}} = \lim_{n\to\infty} \sqrt{\langle g_n, g_n\rangle_{S(R)}} = 0.$$

We therefore conclude

$$\|f - f_n\|_{H(R)} \leq \|f - g_n\|_{H(R)} + \|g_n - f_n\|_{H(R)} \leq \|f - g_n\|_{H(R)} + \frac{1}{n} \xrightarrow[n\to\infty]{} 0\,,$$

which means that for a Cauchy sequence $(f_n)_{n\in\mathbb{N}} \subseteq H(R)$ we have found a limit $f \in H(R)$ with $f_n \to f$ in $H(R)$ for $n \to \infty$. Altogether, $H(R)$ satisfies all properties demanded by Definition 4.1.2 and therefore is the RKHS with respect to the reproducing kernel R. □

7.4 Proofs for Chapter 5

Proof of Theorem 5.0.2

Without loss of generality we can assume that

$$\int_{[0,1]} \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) = 1,$$

since, if

$$\widetilde{Y} = \widetilde{f}^T(t)\beta + \widetilde{\varepsilon}(t)$$

is a heteroscedastic generic linear model like the model (5.1) with

$$\int_{[0,1]} \widetilde{\sigma}^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) = c \neq 1,$$

we define

$$Y(t) := \frac{\widetilde{Y}(t)}{\sqrt{c}}, \quad f(t) := \frac{\widetilde{f}(t)}{\sqrt{c}} \quad \text{and} \quad \varepsilon(t) := \frac{\widetilde{\varepsilon}(t)}{\sqrt{c}}.$$

Then

$$\begin{aligned}
\int_{[0,1]} \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) &= \int_{[0,1]} \mathbb{V}ar(\varepsilon(t))\, d\lambda\!\!\!^{-}_{[0,1]}(t) \\
&= \int_{[0,1]} \mathbb{V}ar\left(\frac{\widetilde{\varepsilon}(t)}{\sqrt{c}}\right) d\lambda\!\!\!^{-}_{[0,1]}(t) \\
&= \frac{1}{c} \int_{[0,1]} \mathbb{V}ar\left(\widetilde{\sigma}^2(t)\right) d\lambda\!\!\!^{-}_{[0,1]}(t) \\
&= 1
\end{aligned}$$

holds true and we get the heteroscedastic generic linear regression model

$$Y = f^T(t)\beta + \varepsilon(t)$$

where $\int_0^1 \sigma^2(t)\, d\lambda\!\!\!^{-}_{[0,1]}(t) = 1$ is valid. This model is equivalent to the non-normalized model $\widetilde{Y} = \widetilde{f}^T(t)\beta + \widetilde{\varepsilon}(t)$ we started with, since the regression coefficients $\beta = (\beta_1, \ldots, \beta_p)^T$ did not change and β is the quantity the practitioner usually wants to study via estimation or hypothesis testing.

The following Lemma is a well-known invariance principle of Prokhorov (see Theorem 3.1 in Prokhorov (1956) page 190) which can also be found in Billingsley (1968) (see Problem 1 on page 77). We will use it in the context of this proof.

Lemma 7.4.1. **[Prokhorov (1956), Billingsley (1968)]**

Let a triangular array $((\xi_{ni})_{i=1,\ldots,n})_{n\in\mathbb{N}}$ of line by line independent random variables with $\mathbb{E}(\xi_{ni}) = 0$ and $\mathbb{V}ar(\xi_{ni}) = \sigma_{ni}^2 \in (0,\infty)$, $i = 1,\ldots,n$, $n \in \mathbb{N}$, be given. We use the notations

$$s_{ni}^2 := \sum_{j=1}^{i} \sigma_{nj}^2, \quad s_n^2 := s_{nn}^2 \quad \text{and} \quad z_{ni} := \frac{s_{ni}^2}{s_n^2}$$

for simplification. Then, the random process which is linear on each interval $[z_{ni-1}, z_{ni}]$, $i = 1,\ldots,n$ and for $i = 0,1,\ldots,n$, has the value $\frac{1}{s_n}\sum_{j=1}^{i}\xi_{nj}$ at z_{ni} converges for $n \to \infty$ in distribution to a standard Brownian motion $B(\cdot)$ if Lindeberg's condition

$$\frac{1}{s_n^2}\sum_{j=1}^{n}\int \mathbb{1}_{[\varepsilon s_n,\infty)}(|\xi_{nj}|)\cdot\xi_{nj}^2\, d\mathbb{P} \xrightarrow[n\to\infty]{} 0$$

is fulfilled for every $\varepsilon > 0$.

In the present situation of the heteroscedastic linear regression model (5.1), we define $\xi_{ni} := \varepsilon\left(\frac{i}{n}\right)$ so that $\mathbb{E}\left(\varepsilon\left(\frac{i}{n}\right)\right) = 0$ and $\mathbb{V}ar\left(\varepsilon\left(\frac{i}{n}\right)\right) = \sigma^2\left(\frac{i}{n}\right) =: \sigma_{ni}^2$ hold true and that $\varepsilon\left(\frac{1}{n}\right), \varepsilon\left(\frac{2}{n}\right), \ldots, \varepsilon(1)$ are stochastically independent for $n \in \mathbb{N}$. Therefore, according to Lemma 7.4.1, we can infer that for the random process $V_n \equiv (V_n(t))_{t\in[0,1]}$, which is linear on each interval $[z_{ni-1}, z_{ni}]$, $i = 1,\ldots,n$ and takes the value $\frac{1}{s_n}\sum_{j=1}^{i}\varepsilon\left(\frac{i}{n}\right)$ at z_{ni}, the convergence

$$V_n(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B(\cdot) \quad \text{in } C[0,1], \tag{7.91}$$

holds true. Note that Lindeberg's condition is satisfied in this situation due to Assumption vi) of model (5.1), since

$$\begin{aligned}
&\frac{1}{s_n^2}\sum_{j=1}^{n}\int \mathbb{1}_{[\varepsilon s_n,\infty)}\left(\left|\varepsilon\left(\frac{j}{n}\right)\right|\right)\cdot\varepsilon^2\left(\frac{j}{n}\right) d\mathbb{P} \\
&\qquad \le \frac{1}{s_n^2}\sum_{j=1}^{n}\int \mathbb{1}_{[\varepsilon s_n,\infty)}\left(\left|\varepsilon\left(\frac{j}{n}\right)\right|\right)\cdot\frac{\left|\varepsilon\left(\frac{j}{n}\right)\right|^{2+\delta}}{(\varepsilon s_n)^{\delta}}\, d\mathbb{P} \\
&\qquad \le \frac{1}{\varepsilon^{\delta} s_n^{2+\delta}}\sum_{j=1}^{n}\mathbb{E}\left(\left|\varepsilon\left(\frac{j}{n}\right)\right|^{2+\delta}\right) \\
&\qquad \le \frac{n}{s_n^2}\cdot\frac{C}{\varepsilon^{\delta} s_n^{\delta}} \xrightarrow[n\to\infty]{} 0
\end{aligned} \tag{7.92}$$

holds true, since we have

$$\mathbb{E}\left(\left|\varepsilon\left(\frac{j}{n}\right)\right|^{2+\delta}\right) \leq C$$

and according to Assumption vi) of (5.1),

$$\frac{n}{s_n^2} = \left(\frac{1}{n}\sum_{j=1}^{n}\sigma_{nj}^2\right)^{-1} = \left(\int_{[0,1]}\sigma^2(t)\,d\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{(-\infty,\cdot]}\left(\frac{j}{n}\right)\right)(t)\right)^{-1}$$

$$\xrightarrow[n\to\infty]{} \left(\int_{[0,1]}\sigma^2(t)\,d\lambda\!\!\lambda_{[0,1]}(t)\right) = 1 \tag{7.93}$$

is satisfied and the fact that

$$s_n^{\delta} = \left(\sum_{i=1}^{n}\sigma_{ni}^2\right)^{\frac{\delta}{2}} = \left(n\cdot\int_{[0,1]}\sigma^2(t)\,d\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{(-\infty,\cdot]}\left(\frac{j}{n}\right)\right)(t)\right)^{\frac{\delta}{2}} \xrightarrow[n\to\infty]{} \infty$$

holds true, as $\delta > 0$ and $\int_{[0,1]}\sigma^2(t)\,d\left(\frac{1}{n}\sum_{j=1}^{n}\mathbb{1}_{(-\infty,\cdot]}\left(\frac{j}{n}\right)(t) \xrightarrow[n\to\infty]{} 1\right.$.

The central idea for the remainder of the proof is to find a transformation

$$\Phi_n : C[0,1] \to C[0,1], \quad n \in \mathbb{N},$$

with the following two properties:

- $(\Phi_n(V_n))(\cdot) = \frac{1}{\sqrt{n}}T_n\left(\varepsilon_n\right)(\cdot)$ on $[0,1]$. (7.94)
- $$(\Phi_n(V_n))(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (\Phi(V_n))(\cdot) = B\left(\int_{[0,\cdot]}\sigma^2(t)\,d\lambda\!\!\lambda_{[0,1]}(t)\right) \quad \text{in } C[0,1] \tag{7.95}$$

  for an appropriate $\Phi : C[0,1] \to C[0,1]$.

Proof of (7.94):

Since for all $n \in \mathbb{N}$ the random process $\frac{1}{\sqrt{n}}T_n\left(\varepsilon_n\right)(\cdot)$ takes the value $\frac{1}{\sqrt{n}}\sum_{j=1}^{i}\varepsilon\left(\frac{j}{n}\right)$ at the point $\frac{i}{n}$, $i = 0,\ldots,n$, and is piecewise linear in between the points $\frac{1}{n}, \frac{2}{n}, \ldots, 1$, for (7.94) to be true it suffices to define Φ_n in such a way that $(\Phi_n(V_n))(\cdot)$ is piecewise linear in between the points $\frac{1}{n}, \frac{2}{n}, \ldots, 1$ and that

$$(\Phi_n(V_n))\left(\frac{i}{n}\right) = \frac{1}{\sqrt{n}}\,T_n\left(\varepsilon_n\right)\left(\frac{i}{n}\right), \quad i = 0,1,\ldots,n.$$

To achieve this we define

i) the mapping $p_n : [0,1] \to \{0,1,\ldots,n\}$ with $p_n(t) := \max\{i \in 1,\ldots,n \,|\, \frac{i}{n} \leq t\}$.

ii) the mapping $\phi_n : [0,1] \to [0,1]$ with

$$\phi_n(t) := z_{n,p_n(t)} + (z_{n,p_n(t)+1} - z_{n,p_n(t)}) \cdot (nt - p_n(t)).$$

iii) the operator $\Phi_n : C[0,1] \to C[0,1]$ with

$$(\Phi_n(x))(t) := \frac{s_n}{\sqrt{n}} \cdot x(\phi_n(t)) = \frac{s_n}{\sqrt{n}} \cdot x\left(z_{n,p_n(t)} + (z_{n,p_n(t)+1} - z_{n,p_n(t)}) \cdot (nt - p_n(t))\right).$$

Then, for $i \in 0,1,\ldots,n$, we have

$$\begin{aligned}
(\Phi_n(V_n))\left(\frac{i}{n}\right) &= \frac{s_n}{\sqrt{n}} \cdot V_n\left(\phi_n\left(\frac{i}{n}\right)\right) \\
&= \frac{s_n}{\sqrt{n}} \cdot V_n\left(z_{n,p_n(i/n)} + (z_{n,p_n(i/n)+1} - z_{n,p_n(i/n)}) \cdot (n \cdot i/n - p_n(i/n))\right) \\
&= \frac{s_n}{\sqrt{n}} \cdot V_n\left(z_{ni} + (z_{n,i+1} - z_{n,i}) \cdot (i - i)\right) \\
&= \frac{s_n}{\sqrt{n}} \cdot V_n\left(z_{ni}\right) \\
&= \frac{1}{\sqrt{n}} \sum_{j=1}^{i} \varepsilon\left(\frac{i}{n}\right) \\
&= \frac{1}{\sqrt{n}} T_n\left(\varepsilon_n\right)\left(\frac{i}{n}\right).
\end{aligned}$$

Furthermore, for $i = 1,\ldots,n$ and $t \in \left[\frac{i-1}{n}, \frac{i}{n}\right]$

$$(\Phi_n(V_n))(t) = \frac{s_n}{\sqrt{n}} \cdot V_n\left(z_{n,i-1} + (z_{n,i} - z_{n,i-1}) \cdot (n \cdot t - (i-1))\right)$$

is linear since V_n is linear in $[z_{ni-1}, z_{ni}]$.

Proof of (7.95):
For $t \in [0,1]$, we have

$$\frac{p_n(t)}{n} = \max\{i = 0, 1/n, \ldots, 1 \,|\, i \leq t\} \xrightarrow[n\to\infty]{} t$$

and thus,

$$\begin{aligned}
\phi_n(t) &= z_{n,p_n(t)} + (z_{n,p_n(t)+1} - z_{n,p_n(t)}) \cdot (nt - p_n(t)) \\
&= \frac{s^2_{np_n(t)}}{s^2_n} + \frac{\sigma^2\left(\frac{p_n(t)+1}{n}\right)}{s^2_n} \cdot (nt - p_n(t))
\end{aligned}$$

$$
\begin{aligned}
&= \frac{\frac{1}{n}\sum_{i=1}^{p_n(t)} \sigma^2\left(\frac{i}{n}\right)}{\frac{1}{n}\sum_{i=1}^{n} \sigma^2\left(\frac{i}{n}\right)} + \frac{\sigma^2\left(\frac{p_n(t)+1}{n}\right)}{\sum_{i=1}^{n} \sigma^2\left(\frac{i}{n}\right)} \cdot n \cdot \left(t - \frac{p_n(t)}{n}\right) \\
&= \frac{\int_0^{\frac{p_n(t)}{n}} \sigma^2(x)\, dF_n(x)}{\int_0^1 \sigma^2(x)\, dF_n(x)} + \frac{\sigma^2\left(\frac{p_n(t)+1}{n}\right)}{\frac{1}{n}\sum_{i=1}^{n} \sigma^2\left(\frac{i}{n}\right)} \cdot \left(t - \frac{p_n(t)}{n}\right) \\
&\xrightarrow[n\to\infty]{} \int_{[0,t]} \sigma^2(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) =: \phi(t)\,, \qquad (7.96)
\end{aligned}
$$

holds true. Here, F_n denotes the empirical distribution function corresponding to the design points $\frac{1}{n}, \frac{2}{n}, \ldots, 1$. The above convergence holds true, since due to

$$\int_{[0,1]} \sigma^2(x)\, dF_n(x) \xrightarrow[n\to\infty]{} \int_{[0,1]} \sigma^2(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x) = 1$$

and $p_n(t)/n \xrightarrow[n\to\infty]{} t$ the first summand converges as asserted and the second summand converges to zero because $\sigma^2\left(\frac{p_n(t)+1}{n}\right)$ is bounded for $t \in [0,1]$ and $n \to \infty$ since $\sigma : [0,1] \to (0,\infty)$ is of bounded variation, hence bounded and

$$t - \frac{p_n(t)}{n} \xrightarrow[n\to\infty]{} 0.$$

Therefore, for $x(\cdot) \in C[0,1]$ and $t \in [0,1]$,

$$
\begin{aligned}
\lim_{n\to\infty} (\Phi_n(x))(t) &= \lim_{n\to\infty} \frac{s_n}{\sqrt{n}} \cdot x(\phi_n(t)) \\
&= x\left(\int_{[0,t]} \sigma^2(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x)\right) \\
&= x\left(\phi(t)\right) \\
&=: (\Phi(x))(t)
\end{aligned}
$$

holds true, since (7.93) implies that $\frac{s_n}{\sqrt{n}} \xrightarrow[n\to\infty]{} 1$.

In the sequel of this proof we use the following Lemma. We point out that we could not find the lemma in the literature, which is surprising because it can be of interests on its own in the context of approximation theory.

Lemma 7.4.2.

Let $I \subset \mathbb{R}$ be a compact interval and $f : I \longrightarrow \mathbb{R}$ continuous. For $n \in \mathbb{N}$ let $T_n = \{t_{n1}, \ldots, t_{ni_n}\}$, $i_n \in \mathbb{N}$ with $\min(I) \leq t_{n1} \leq \cdots \leq t_{ni_n} \leq \max(I)$ be a partition of I and $f_n : I \longrightarrow \mathbb{R}$ be the piecewise linear interpolation of f with respect to the partition T_n. That is, f_n is the continuous and piecewise linear function with the fraction points $(t_{n1}, f(t_{n1})), \ldots, (t_{ni_n}, f(t_{ni_n}))$, the starting point $(\min(I), f(\min(I)))$ and the endpoint $(\max(I), f(\max(I)))$.

Then the sequence $(f_n)_{n\in\mathbb{N}}$ is equicontinuous.

Proof of Lemma 7.4.2 :

Since f is a continuous function on a compact interval $\max(f(I))$ and $\min(f(I))$ exist and therefore, without loss of generality, we can assume

$$\max(f(I)) - \min(f(I)) = 1$$

since, if otherwise, either f is constant or $\max(f(I)) - \min(f(I)) = k \neq 1$. In the first case f_n would be constant for all $n \in \mathbb{N}$ and therefore $(f_n)_{n\in\mathbb{N}}$ would be equicontinuous. In the latter case, we consider $\widetilde{f} : I \to \mathbb{R}$ with $\widetilde{f}(x) = \frac{1}{k} \cdot f(x)$ so that $\max(\widetilde{f}(x)) - \min(\widetilde{f}(x)) = 1$. Here, for every piecewise linear interpolation f_n of f and $\widetilde{f}_n$ on $\widetilde{f}$, relative to the same partition $t_{n1}, \ldots, t_{ni_n}$,

$$\widetilde{f}_n = \frac{1}{k} f_n$$

holds true. Thus, it suffices to show that $(\widetilde{f}_n)_{n\in\mathbb{N}}$ is equicontinuous in order to prove that $(f_n)_{n\in\mathbb{N}} = (k \cdot f)_{n\in\mathbb{N}}$ is equicontinuous, too.

Because f is continuous on the compact interval I and f is uniformly continuous, for every $\varepsilon > 0$ there exists a $\delta > 0$ such that for all $|x - y| < \delta$,

$$|f(x) - f(y)| < \frac{\varepsilon}{3}$$

holds true. Now, let $\varepsilon > 0$ be given. In order to prove that $(f_n)_{n\in\mathbb{N}}$ is equicontinuous, we have to prove that there is a $\widetilde{\delta} > 0$ such that for all $n \in \mathbb{N}$, all $\min(I) \leq t_{n1} \leq, \ldots, \leq t_{ni_n} \leq \max(I)$ arbitrary partitions of I and all $|x - y| < \widetilde{\delta}$,

$$|f_n(x) - f_n(y)| \leq \varepsilon$$

holds true. We therefore define

$$\widetilde{\delta} = \min(\delta, \frac{\varepsilon\,\delta}{3}),$$

where δ is the positive number for which according to the uniformly continuity of f,

$$|x - y| < \delta \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{3}$$

holds. Furthermore, let $t_{n0} = \min(I)$, $t_{ni_n+1} = \max(I)$, f_n be the piecewise linear inter-

polation of f corresponding to the given partition and $x \leq y \in I$ with $|x-y| \leq \widetilde{\delta}$. Then, in the cases we will list down below the following assertions about

$$|f_n(x) - f_n(y)|$$

hold true:

Case 1: $t_{nk} \leq x \leq y \leq t_{nk+1}$ for a $k \in \{0, \ldots, i_n\}$.

Subcase a: $t_{nk+1} - t_{nk} < \delta$ holds true and therefore

$$|f_n(x) - f_n(y)| \leq |f_n(t_{nk}) - f_n(t_{nk+1})| = |f(t_{nk}) - f(t_{nk+1})| < \frac{\varepsilon}{3} < \varepsilon,$$

since f_n is linear on $[t_{nk}, t_{nk+1}]$, $[x, y] \subseteq [t_{nk}, t_{nk+1}]$ and $f_n(t_{nj}) = f(t_{nj})$ for $j = 0, \ldots, i_n + 1$.

Subcase b: $t_{nk+1} - t_{nk} \geq \delta$ holds true and therefore

$$\Delta_{1b} := \left| \frac{f(t_{nk+1}) - f(t_{nk})}{t_{nk+1} - t_{nk}} \right| \leq \frac{1}{\delta},$$

since $\max(f(I)) - \min(f(I)) = 1$ holds true and as a consequence we can infer, that

$$|f_n(x) - f_n(y)| = \Delta_{1b} \cdot (y - x) \leq \frac{y-x}{\delta} \leq \frac{\widetilde{\delta}}{\delta} = \min(1, \frac{\varepsilon}{3}) \leq \frac{\varepsilon}{3}$$

holds true, since f_n is linear on $[t_{nk}, t_{nk+1}]$ and $[x, y] \subseteq [t_{nk}, t_{nk+1}]$.

Case 2: For a $k \in \{0, \ldots, i_n - 1\}$ there is a $l \geq 1$ such that $t_{nk} \leq x \leq t_{nk+1} \leq \cdots \leq t_{nk+l} \leq y \leq t_{nk+l+1}$ holds true. In this case we have

$$\begin{aligned} |f_n(x) - f_n(y)| &= |f_n(x) - f_n(t_{nk+1}) + f_n(t_{nk+1}) - f_n(t_{nk+l}) + f_n(t_{nk+l}) - f_n(y)| \\ &\leq |f_n(x) - f_n(t_{nk+1})| + |f_n(t_{nk+1}) - f_n(t_{nk+l})| + |f_n(t_{nk+l}) - f_n(y)|. \end{aligned} \tag{7.97}$$

- Concerning the summand $|f_n(t_{nk+1}) - f_n(t_{nk+l})|$:

$$|f_n(t_{nk+1}) - f_n(t_{nk+l})| = |f(t_{nk+1}) - f(t_{nk+l})| < \frac{\varepsilon}{3},$$

since $f_n(t_{nj}) = f(t_{nj})$ for $j = 0, \ldots, i_n + 1$ and $|t_{nk+1} - t_{nk+l}| \leq |x - y| \leq \delta$.

- Concerning the summand $|f_n(t_{nk+l}) - f_n(y)|$:

Subcase a: If $t_{nk+l+1} - t_{nk+l} < \delta$ holds true, we have

$$|f_n(t_{nk+l}) - f_n(y)| \leq |f_n(t_{nk+l}) - f_n(k+l+1)| = |f(t_{nk+l}) - f(k+l+1)| < \frac{\varepsilon}{3},$$

since f_n is linear on $[t_{nk+l}, t_{nk+l+1}]$, $[t_{nk+l}, y] \subseteq [t_{nk+l}, t_{nk+l+1}]$, $f_n(t_{nj}) = f(t_{nj})$ for $j = 0, \ldots, i_n + 1$ and $t_{nk+l+1} - t_{nk+l} < \delta$.

Subcase b: If $t_{nk+l+1} - t_{nk+l} \geq \delta$ holds true, we have

$$\Delta_{2b} := \left| \frac{f(t_{nk+l+1}) - f(t_{nk+l})}{t_{nk+l+1} - t_{nk+l}} \right| \leq \frac{1}{\delta},$$

since $\max(f(I)) - \min(f(I)) = 1$ and therefore

$$|f_n(t_{nk+l}) - f_n(y)| = \Delta_{2b} \cdot (y - t_{nk+l}) \leq \frac{y - t_{nk+l}}{\delta} \leq \frac{\widetilde{\delta}}{\delta} = \min(1, \frac{\varepsilon}{3}) \leq \frac{\varepsilon}{3},$$

since f_n is linear on $[t_{nk+l}, t_{nk+l+1}]$ and $[t_{nk+l}, y] \subseteq [t_{nk+l}, t_{nk+1}]$.

- Concerning the summand $|f_n(x) - f_n(t_{nk+1})|$:

Subcase a: In the case $t_{nk+1} - t_{nk} < \delta$ we can show similarly to Subcase a) from above that $|f_n(x) - f_n(t_{nk+1})| < \frac{\varepsilon}{3}$ holds true.

Subcase b: In the case $t_{nk+1} - t_{nk} \geq \delta$ we can show analogously to Subcase b) from above that $|f_n(x) - f_n(t_{nk+1})| \leq \frac{\varepsilon}{3}$ holds true.

Taking the above together, we know that in Case 2 for equation (7.97) we have

$$|f_n(x) - f_n(y)| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon.$$

Taking our findings in Case 1 and Case 2 together, we have shown that

$$|f_n(x) - f_n(y)| \leq \varepsilon$$

holds true independent of $n \in \mathbb{N}$ and the partition $t_{n1} \leq, \ldots, \leq t_{ni_n}$. Consequently, we have shown that $(f_n)_{n\in\mathbb{N}}$ is equicontinuous. □

Moving on, considering $(x_n)_{n\in\mathbb{N}} \subseteq C[0,1]$ and $x \in C[0,1]$ such that $||x_n - x||_\infty \xrightarrow[n\to\infty]{} 0$, we have

$$\begin{aligned}
\|(\Phi_n(x_n))(\cdot) - (\Phi(x))(\cdot)\|_\infty &\leq \| (\Phi_n(x_n))(\cdot) - (\Phi_n(x))(\cdot) \|_\infty + \| (\Phi_n(x))(\cdot) - (\Phi(x))(\cdot) \|_\infty \\
&= \left\| \frac{s_n}{\sqrt{n}} x_n(\phi_n(\cdot)) - \frac{s_n}{\sqrt{n}} x(\phi_n(\cdot)) \right\|_\infty + \left\| \frac{s_n}{\sqrt{n}} x(\phi_n(\cdot)) - x(\phi(\cdot)) \right\|_\infty \\
&\leq \frac{s_n}{\sqrt{n}} \|x_n(\cdot) - x(\cdot)\|_\infty + \left\| \frac{s_n}{\sqrt{n}} x(\phi_n(\cdot)) - x(\phi_n(\cdot)) \right\|_\infty + \|x(\phi_n(\cdot)) - x(\phi(\cdot))\|_\infty \\
&\leq \frac{s_n}{\sqrt{n}} \|x_n(\cdot) - x(\cdot)\|_\infty + \left| \frac{s_n}{\sqrt{n}} - 1 \right| \cdot \|x(\cdot)\|_\infty + \|x(\phi_n(\cdot)) - x(\phi(\cdot))\|_\infty.
\end{aligned}$$

For $n \to \infty$ all three summands converge to zero. The first summand converges to zero due to (7.93) and the uniform convergence of x_n to x. The second summand converges to zero because of (7.93) and the fact that $x(\cdot)$ is a continuous function on $[0,1]$, hence bounded. Concerning the third summand, since x is continuous on $[0,1]$ it is uniformly

continuous. Hence for convergence to zero it would suffice to show that ϕ_n converges to ϕ uniformly. To see this, note that $(\phi_n)_{n\in\mathbb{N}}$ is the piecewise linear interpolation of the continuous function $\phi(\cdot) = \int_0^{\cdot} \sigma^2(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x)$ with respect to the partition points

$$0,\ \frac{s_{n1}^2}{s_n^2},\ \frac{s_{n2}^2}{s_n^2},\ \ldots,\ \frac{s_n^2}{s_n^2} = 1$$

on $[0,1]$. Due to Lemma 7.4.2 we therefore know that $(\phi_n)_{n\in\mathbb{N}}$ is equicontinuous on the compact set $[0,1]$ and due to (7.96) we already know that ϕ_n converges to ϕ pointwise, hence

$$\phi_n \xrightarrow[n\to\infty]{} \phi \quad \text{uniformly.}$$

Taking everything together according to (7.91) we have

$$V_n(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B(\cdot) \quad \text{in } C[0,1]$$

and due to Rubin's theorem (see Anderson (1963) page 140 and pages 146 ff., Topsoe (1967) and Theorem 5.5 in Billingsley (1968)) and the uniform convergence of ϕ_n to ϕ, we can conclude that

$$(V_n(\phi_n))(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} (B(\phi))(\cdot) \quad \text{in } C[0,1].$$

We thus have proven assertion (7.95) since

$$\begin{aligned}(\Phi_n(V_n))(\cdot) &= \frac{s_n}{\sqrt{n}} \cdot V_n(\phi_n)(\cdot) \\ &\xrightarrow[n\to\infty]{\mathcal{D}} B(\phi)(\cdot) \\ &= B\left(\int_{[0,\cdot]} \sigma^2(t)\, d\lambda\!\!\!\lambda_{[0,1]}(t)\right).\end{aligned}$$

Note that in the equation above

$$\frac{s_n}{\sqrt{n}} = \sqrt{\frac{1}{n}\sum_{i=1}^n \sigma_{ni}^2} = \sqrt{\int_{[0,1]} \sigma^2(x)\, dF_n(x)} \xrightarrow[n\to\infty]{} \sqrt{\int_{[0,1]} \sigma^2(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x)} = 1$$

is satisfied.

The now proven assertions (7.94) and (7.95) prove Theorem 5.0.2. □

Proof of Theorem 5.0.3

According to Lemma 4.2.2 and the linearity of the partial sum operator T_n we can infer

$$\begin{aligned}\frac{1}{\sqrt{n}}T_n(\hat{\varepsilon}_n)(\cdot) &= \frac{1}{\sqrt{n}}T_n(pr_{W_n{}^{\perp}}(Y_n))(\cdot)\\ &= \frac{1}{\sqrt{n}}T_n(pr_{W_n{}^{\perp}}(X_n\beta+\varepsilon_n))(\cdot)\\ &= \frac{1}{\sqrt{n}}T_n(pr_{W_n{}^{\perp}}(\varepsilon_n))(\cdot)\\ &= \frac{1}{\sqrt{n}}T_n(\varepsilon_n)(\cdot)-\frac{1}{\sqrt{n}}T_n(pr_{W_n}(\varepsilon_n))(\cdot).\end{aligned}$$

As we already know from Theorem 5.0.2

$$\frac{1}{\sqrt{n}}T_n(\varepsilon_n)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} B\left(\int_{[0,\cdot]}\sigma^2(t)\,d\lambda\!\!\!^{-}_{[0,1]}(t)\right) \quad \text{in } C[0,1]$$

holds true and thus it remains to prove that

$$\begin{aligned}\frac{1}{\sqrt{n}}T_n(pr_{W_n}(\varepsilon_n))(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} &\left(\int_{[0,\cdot]}f(t)\,d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^T\left(\int_{[0,1]}f(t)f^T(t)\,d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^{-1}\times\\ &\times\int_{[0,1]}f(t)\,dB\left(\int_{[0,1]}\sigma^2(u)\,d\lambda\!\!\!^{-}_{[0,1]}(u)\right) \quad \text{in } C[0,1].\end{aligned} \tag{7.98}$$

According to Lemma 3.1 in Bischoff (2002) for $a\in\mathbb{R}^n$ we have

$$T_n(pr_{W_n}(a)) = pr_{W_n\mathcal{H}_{\mathcal{B}}}(T_n(a)).$$

Here, $pr_{W_n\mathcal{H}_B}$ denotes the orthogonal projection on $W_n\mathcal{H}_B = span\{h_{f_{1,n}},\ldots,h_{f_{p,n}}\}$, where

$$h_{f_{i,n}} : [0,1]\longrightarrow\mathbb{R},\ x\longmapsto\frac{1}{n}T_n(f_{i,n})(x), \quad i=1,\ldots,p.$$

Therefore, for the left hand side of (7.98)

$$\frac{1}{\sqrt{n}}T_n(pr_{W_n}(\varepsilon_n))(\cdot) = \frac{1}{\sqrt{n}}pr_{W_n\mathcal{H}_{\mathcal{B}}}(T_n(\varepsilon_n)(\cdot))(\cdot) = pr_{W_n\mathcal{H}_{\mathcal{B}}}\left(\frac{1}{\sqrt{n}}T_n(\varepsilon_n)(\cdot)\right)(\cdot)$$

holds true.

We will show below that if $(u_n)_{n\in\mathbb{N}} \subseteq C[0,1]$ satisfies $u_n \xrightarrow[n\to\infty]{} u \in C[0,1]$ uniformly,

$$\sup_{z\in[0,1]} |pr_{W_n\mathcal{H}_B}(u_n)(z) - pr_{W_{\mathcal{H}_B}}(u)(z)| \xrightarrow[n\to\infty]{} 0. \tag{7.99}$$

Thus, similarly to the end of the proof of Theorem 5.0.2, we deduce due to Rubin's theorem and Theorem 5.0.2 the convergence

$$\begin{aligned}\frac{1}{\sqrt{n}}T_n(pr_{W_n}(\varepsilon_n))(\cdot) &= pr_{W_{n,\mathcal{H}_\mathcal{B}}}\left(\frac{1}{\sqrt{n}}T_n(\varepsilon_n)(\cdot)\right)(\cdot) \xrightarrow[n\to\infty]{\mathcal{D}} pr_{W_{\mathcal{H}_\mathcal{B}}}\left(B\left(\int_{[0,\cdot]}\sigma^2(t)\,d\lambda\!\!\!^{-}_{[0,1]}(t)\right)\right)(\cdot)\\ &= \left(\int_{[0,\cdot]} f(t)d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^T\left(\int_{[0,1]} f(t)f^T(t)d\lambda\!\!\!^{-}_{[0,1]}(t)\right)^{-1}\int_{[0,1]} f(t)dB\left(\int_{[0,t]}\sigma^2(u)\,d\lambda\!\!\!^{-}_{[0,1]}(u)\right),\end{aligned}$$

i.e., (7.98), completing the proof. Note that the last equality follows from what we have already seen in Section 4.2, namely in Lemma 4.2.2, the extension of $pr_{W_{\mathcal{H}_B}}\colon \mathcal{H}_B \longrightarrow W_{\mathcal{H}_B}$ to $pr_{W_{\mathcal{H}_B}}\colon C[0,1] \longrightarrow W_{\mathcal{H}_B}$ after that and Theorem 4.2.3. Therefore, it is sufficient to prove (7.99) in order to complete the proof of Theorem 5.0.3.

The prove of (7.99) is divided into the following six parts.

Part 1:
Let $i \in \{1,\dots,p\}$ and define

$$h'_{f_{i,n}} : [0,1] \longrightarrow \mathbb{R},\ x \longmapsto \sum_{j=1}^{n} f_i\left(\frac{j}{n}\right)\mathbb{1}_{\left[\frac{j-1}{n},\frac{j}{n}\right)}(x) + f_i(1)\mathbb{1}_{\{1\}}(x).$$

Then,

$$\begin{aligned}\int_{[0,x]} h'_{f_{i,n}}(u)\,d\lambda\!\!\!^{-}_{[0,1]}(u) &= \int_{[0,x]}\sum_{j=1}^{n} f_i\left(\frac{j}{n}\right)\mathbb{1}_{\left[\frac{j-1}{n},\frac{j}{n}\right)}(u)\,d\lambda\!\!\!^{-}_{[0,1]}(u)\\ &= \frac{1}{n}\left[\sum_{j=1}^{\lfloor nx\rfloor} f_i\left(\frac{j}{n}\right) + (nx - \lfloor nx\rfloor)\ f_i\left(\frac{\lfloor nx\rfloor + 1}{n}\right)\right]\\ &= \frac{1}{n}T_n(f_{i,n})(x)\\ &= h_{f_{i,n}}(x), \quad x \in [0,1),\end{aligned}$$

holds true and due to

$$\int_{[0,1]} h'_{f_{i,n}}(u)\,d\lambda\!\!\!^{-}_{[0,1]}(u) = \frac{1}{n}\sum_{j=1}^{n} f_i\left(\frac{j}{n}\right) = \frac{1}{n}T_n(f_{i,n})(1) = h_{f_{i,n}}(1),$$

we thus have

$$h_{f_{i,n}}(x) = \int\limits_{[0,x]} h'_{f_{i,n}}(u)\, d\lambda_{[0,1]}(u), \quad x \in [0,1]. \tag{7.100}$$

Furthermore, we have

$$h'_{f_i}(x) = \frac{d}{dx} \int\limits_{[0,x]} f_i(u)\, d\lambda_{[0,1]}(u) = f_i(x), \quad x \in [0,1],$$

up to a zero set with respect to the Lebesgue-measure.

We are going to prove that

$$\|h'_{f_{i,n}} - h'_{f_i}\|_{\mathcal{L}^1([0,1],\lambda_{[0,1]})} \xrightarrow[n\to\infty]{} 0, \quad i = 1, \dots, p. \tag{7.101}$$

Since $f_i : [0,1] \to \mathbb{R}$ is a function of bounded variation, it is bounded and has at most countably many discontinuities. In particular, the set of discontinuities is a zero set with respect to the Lebesgue-measure. By a well-known characterization of Riemann-integrable functions (see for example the end of Kapitel 3 in Deiser (2015)) this implies that f_i is Riemann-integrable. Therefore, we have

$$\lim_{\Delta(P)\to 0} (U(f_i, P) - L(f_i, P)) = 0,$$

where

- $P : a_0 = 0 < a_1 < \cdots < a_n = 1$ denotes a partition of $[0,1]$ and $\Delta(P) := \max\limits_{1\le j\le n} |a_j - a_{j-1}|$ denotes the maximal width of the partition P,
- $U(f_i, P) := \sum\limits_{j=1}^{n} (a_j - a_{j-1}) \sup\limits_{a_{j-1}\le x\le a_j} f_i(x)$ denotes the upper integral of f_i with respect to the partition P,
- $L(f_i, P) := \sum\limits_{j=1}^{n} (a_j - a_{j-1}) \inf\limits_{a_{j-1}\le x\le a_j} f_i(x)$ denotes the lower integral of f_i with respect to the partition P.

Thus, let $\varepsilon > 0$ be given. Then there is a $\delta > 0$ such that

$$U(f_i, P) - L(f_i, P) < \varepsilon$$

holds true for $\Delta(P) \le \delta$. As a consequence, for $n \ge \frac{1}{\delta}$, $n \in \mathbb{N}$ and $P = (a_j)_{j=0}^n = \left(\frac{j}{n}\right)_{j=0}^n$, we have $\Delta(P) = \frac{1}{n} \le \delta$ and hence,

$$\varepsilon > U(f_i, P) - L(f_i, P) = \sum_{j=1}^{n} (a_j - a_{j-1}) \left(\sup_{a_{j-1}\le x\le a_j} f_i(x) - \inf_{a_{j-1}\le x\le a_j} f_i(x) \right)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left(\sup_{\frac{j-1}{n} \leq x \leq \frac{j}{n}} f_i(x) - \inf_{\frac{j-1}{n} \leq x \leq \frac{j}{n}} f_i(x) \right)$$

holds true. Note that $(f_{i,n})_j = f_i\left(\frac{j}{n}\right)$ and therefore

$$\inf_{\frac{j-1}{n} \leq x \leq \frac{j}{n}} f_i(x) \leq f_i\left(\frac{j}{n}\right) \leq \sup_{\frac{j-1}{n} \leq x \leq \frac{j}{n}} f_i(x)$$

holds true. We can thus infer for $x \in [0,1)$,

$$|h'_{f_{i,n}}(x) - f_i(x)| = \left| \sum_{j=1}^{n} \mathbb{1}_{\left[\frac{j-1}{n}, \frac{j}{n}\right)}(x) \left(f_i\left(\frac{j}{n}\right) - f_i(x) \right) \right| \leq \sum_{j=1}^{n} \mathbb{1}_{\left[\frac{j-1}{n}, \frac{j}{n}\right)}(x) \left| f_i(x) - f_i\left(\frac{j}{n}\right) \right|$$

$$\leq \sum_{j=1}^{n} \mathbb{1}_{\left[\frac{j-1}{n}, \frac{j}{n}\right)}(x) \left(\sup_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) - \inf_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) \right),$$

which leads us to conclude

$$\|h'_{f_{i,n}} - h'_{f_i}\|_{\mathcal{L}^1([0,1],\lambda_{[0,1]})} = \int_{[0,1]} |h'_{f_{i,n}}(x) - f_i(x)| \, d\lambda_{[0,1]}(x)$$

$$\leq \int_{[0,1]} \sum_{j=1}^{n} \mathbb{1}_{\left[\frac{j-1}{n}, \frac{j}{n}\right)}(x) \left(\sup_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) - \inf_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) \right) d\lambda_{[0,1]}(x)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \left(\sup_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) - \inf_{\frac{j-1}{n} \leq t \leq \frac{j}{n}} f_i(t) \right) < \varepsilon$$

for all $n \geq \frac{1}{\delta}$. This completes the proof of (7.101).

Part 2:

Let $(\mathcal{H}, \langle \cdot, \cdot \rangle)$ be an inner product space, $v_1, \ldots, v_p \in \mathcal{H}$ be linearly independent and $(v_i^{(n)})_{n \in \mathbb{N}} \subset \mathcal{H}$ with $v_i^{(n)} \xrightarrow[n \to \infty]{} v_i$, $i = 1, \ldots, p$, be given. We additionally denote by $\widetilde{v}_1^{(n)}, \ldots, \widetilde{v}_p^{(n)}$ the output of the Gram-Schmidt orthonormalization of $v_1^{(n)}, \ldots, v_p^{(n)}$ and by $\widetilde{v}_1, \ldots, \widetilde{v}_p$, we denote the output of the Gram-Schmidt orthonormalization of $v_1, \ldots, v_p$. Then the following assertions hold true:

i) $v_1^{(n)}, \ldots, v_p^{(n)}$ are linearly independent for $n \geq n_0$ for a suitable $n_0 \in \mathbb{N}$.

ii) $\widetilde{v}_i = \sum_{j=1}^{i} \alpha_{ij} v_j$ for some $\alpha_{i1}, \ldots, \alpha_{ii} \in \mathbb{R}$ for all $i = 1, \ldots, p$.

iii) $\widetilde{v}_i^{(n)} = \sum_{j=1}^{i} \alpha_{ij}^{(n)} v_j^{(n)}$ for some $\alpha_{i1}^{(n)}, \ldots, \alpha_{ii}^{(n)} \in \mathbb{R}$ for all $i = 1, \ldots, p$ and $n \geq n_0$.

iv) $\alpha_{ij}^{(n)} \xrightarrow[n \to \infty]{} \alpha_{ij}$, $j = 1, \ldots, i$, and $\widetilde{v}_i^{(n)} \xrightarrow[n \to \infty]{} \widetilde{v}_i$ for all $i = 1, \ldots, p$.

Proof: Consider the Gram matrix $G := (\langle v_i, v_j\rangle)_{i,j=1}^p \in \mathbb{R}^{p\times p}$ as well as the Gram matrices

$$G^{(n)} := (\langle v_i^{(n)}, v_j^{(n)}\rangle)_{i,j=1}^p \in \mathbb{R}^{p\times p}, \quad n \in \mathbb{N}.$$

Then, for all $c \in \mathbb{R}^p$,

$$\begin{aligned}\langle G^{(n)}c, c\rangle &= \sum_{i=1}^p (G^{(n)}c)_i\, c_i = \sum_{i=1}^p \sum_{j=1}^p (G^{(n)})_{i,j}\, c_j\, c_i \\ &= \left\langle \sum_{i=1}^p c_i v_i^{(n)}, \sum_{j=1}^p c_j v_j^{(n)} \right\rangle = \left\| \sum_{i=1}^p c_i v_i^{(n)} \right\|^2 \\ &\geq 0\end{aligned}$$

holds true, showing that $G^{(n)}$ is symmetric and positive-semidefinite, and that $v_1^{(n)}, \ldots, v_p^{(n)}$ are linearly independent if and only if $G^{(n)}$ is positive definite, i.e. invertible. By the same arguments, it follows that G is invertible. Since $G^{(n)} \xrightarrow[n\to\infty]{} G$, we see that $G^{(n)}$ is invertible for $n \geq n_0$ for some $n_0 \in \mathbb{N}$. This proves assertion i).

We now prove assertions ii) and iii) by induction over $i \in \{1, \ldots, p\}$ for $n \geq n_0$.
For $i = 1$, we have

$$\widetilde{v}_1 = \frac{1}{\|v_1\|}\, v_1 \quad \text{and} \quad \widetilde{v}_1^{(n)} = \frac{1}{\|v_1^{(n)}\|}\, v_1^{(n)}$$

so that the claims hold true in this case. If ii) and iii) hold true for all $i \leq i_0 \leq p-1$, then

$$\widetilde{v}_{i_0+1} = \frac{1}{\|w_{i_0+1}\|}\, w_{i_0+1} \quad \text{with} \quad w_{i_0+1} = v_{i_0+1} - \sum_{j=1}^{i_0} \langle v_{i_0+1}, \widetilde{v}_j\rangle \widetilde{v}_j$$

and

$$\widetilde{v}_{i_0+1}^{(n)} = \frac{1}{\|w_{i_0+1}^{(n)}\|}\, w_{i_0+1}^{(n)} \quad \text{with} \quad w_{i_0+1}^{(n)} = v_{i_0+1}^{(n)} - \sum_{j=1}^{i_0} \langle v_{i_0+1}^{(n)}, \widetilde{v}_j^{(n)}\rangle \widetilde{v}_j^{(n)}$$

hold true. Hence, ii) and iii) hold true for $i = i_0 + 1$ with

$$\alpha_{i_0+1,j} = \begin{cases} \frac{1}{\|w_{i_0+1}\|}, & j = i_0+1 \\ -\frac{\langle v_{i_0+1}, \widetilde{v}_j\rangle}{\|w_{i_0+1}\|}, & j \leq i_0 \end{cases}$$

and

$$\alpha_{i_0+1,i}^{(n)} = \begin{cases} \frac{1}{\|w_{i_0+1}^{(n)}\|}, & j = i_0+1 \\ -\frac{\langle v_{i_0+1}^{(n)}, \widetilde{v}_j^{(n)}\rangle}{\|w_{i_0+1}^{(n)}\|}, & j \leq i_0. \end{cases}$$

Note that as $v_i^{(n)} \xrightarrow[n\to\infty]{} v_i$, $i = 1, \ldots, p$, according to the above construction $\alpha_{ij}^{(n)} \xrightarrow[n\to\infty]{} \alpha_{ij}$, $j = 1, \ldots, i$, and $\widetilde{v}_i^{(n)} \xrightarrow[n\to\infty]{} \widetilde{v}_i$ for all $i = 1, \ldots, p$, which proves assertion iv).

Part 3:
For $i = 1, \dots, p$,

$$\begin{aligned}
\|h_{f_{i,n}} - h_{f_i}\|_{\mathcal{H}_B} &= \|h'_{f_{i,n}} - h'_{f_i}\|_{\mathcal{L}^2([0,1],\lambda\!\!\lambda_{[0,1]})} = \|h'_{f_{i,n}} - f_i\|_{\mathcal{L}^2([0,1],\lambda\!\!\lambda_{[0,1]})} \\
&= \sqrt{\int\limits_{[0,1]} |h'_{f_{i,n}}(x) - f_i(x)| \cdot |h'_{f_{i,n}}(x) - f_i(x)| \, d\lambda\!\!\lambda_{[0,1]}(x)} \\
&\leq \sqrt{\|h'_{f_{i,n}} - f_i\|_\infty \cdot \int\limits_{[0,1]} |h'_{f_{i,n}} - f_i|(x) \, d\lambda\!\!\lambda_{[0,1]}(x)} \\
&= \|h'_{f_{i,n}} - f_i\|_\infty^{\frac{1}{2}} \cdot \|h'_{f_{i,n}} - h'_{f_i}\|_{\mathcal{L}^1([0,1],\lambda\!\!\lambda_{[0,1]})}^{\frac{1}{2}} \\
&\xrightarrow[n\to\infty]{} 0
\end{aligned}$$

holds true, since $\|h'_{f_{i,n}}\|_\infty \leq \|f_i\|_\infty$, f_i is bounded and assertion (7.101) holds true.
In view of Part 2, since for $i = 1, \dots, p$, $\widetilde{h}_{f_{i,n}}$ is a linear combination of $h_{f_{1,n}}, \dots, h_{f_{i,n}}$ with convergent and thus bounded coefficients, this implies $\widetilde{h}_{f_{i,n}} \xrightarrow[n\to\infty]{} \widetilde{h}_{f_i}$ in $\mathcal{H}_B$ for $i \in \{1, \dots, p\}$.

Part 4:
Let $BV_D(\mathcal{H}_B) = \{h \in \mathcal{H}_B \,|\, h' \text{ is of bounded variation}\}$ and

$$\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]} : BV_D(\mathcal{H}_B) \times C[0,1] \longrightarrow \mathbb{R}, \ (h, u) \longmapsto \int\limits_{[0,1]}^{(R)} h' du$$

be as defined in Section 4.2. Thus, for the extension $pr_{W_n\mathcal{H}_B}$ of the orthogonal projection onto $W_{n\mathcal{H}_\mathcal{B}}$ with respect to the bilinear mapping $\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}$ we have

$$pr_{W_n\mathcal{H}_B} : C[0,1] \longrightarrow C[0,1], \ u \longmapsto \sum_{i=1}^{p} \langle \widetilde{h}_{f_{i,n}}, u \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]} \widetilde{h}_{f_{i,n}},$$

where, for $n \in \mathbb{N}$, $\widetilde{h}_{f_{1,n}}, \dots, \widetilde{h}_{f_{p,n}}$ is the output of the Gram-Schmidt orthonormalization of $h_{f_{1,n}}, \dots, h_{f_{p,n}}$ and therefore an orthonormal basis of $W_{n\mathcal{H}_\mathcal{B}}$.
Then, there is a $K > 0$ such that

$$\|pr_{W_n\mathcal{H}_B}\|_{C[0,1]\to C[0,1]} \leq K \tag{7.102}$$

holds true for all $n \geq n_0$ with n_0 as in Part 2. Here $\|\cdot\|_{C[0,1]\to C[0,1]}$ denotes the operator norm on the space of linear operators $C[0,1] \to C[0,1]$, thus for $g \in C[0,1]$ and a linear $\phi : C[0,1] \to C[0,1]$, we have $\|\phi\|_{C[0,1]\to C[0,1]} := \sup\limits_{\|g\|_\infty = 1} \|\phi(g)\|_\infty$.
Note, that it is not immediate that $\|pr_{W_n\mathcal{H}_B}\|_{C[0,1]\to C[0,1]} \leq K$ holds true for all $n \in \mathbb{N}$ even though $pr_{W_n\mathcal{H}_B}$ is the extension of an orthogonal projection, since $pr_{W_n\mathcal{H}_B}$ is a projection

with respect to the bilinear mapping $\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}$ and therefore bounded regarding to the norm induced by $\langle \cdot, \cdot \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}$, but not necessary with respect to the norm on $C[0,1]$ (which does not come from an inner product).

Proof of (7.102): For $i \in \{1, \dots, p\}$, with the help of (7.100), we get

$$\begin{aligned} |h_{f_{i,n}}(x)| &= \left| \int\limits_{[0,x]} h'_{f_{i,n}}(u)\, d\lambda\!\!\!^{-}_{[0,1]}(u) \right| \leq \int\limits_{[0,x]} |h'_{f_{i,n}}(u)|\, d\lambda\!\!\!^{-}_{[0,1]}(u) \\ &= \int\limits_{[0,x]} \left| \sum_{j=1}^{n} f_i\left(\frac{j}{n}\right) \mathbb{1}_{\left[\frac{j-1}{n}, \frac{j}{n}\right)}(u) \right| d\lambda\!\!\!^{-}_{[0,1]}(u) \;\leq\; \|f_i\|_\infty. \end{aligned}$$

In combination with Part 2 and Part 3, this shows that

$$\|\widetilde{h}_{f_{i,n}}\|_\infty \leq K_1$$

for all $n \geq n_0$ with n_0 as in Part 2 for $K_1 > 0$ suitable, since $\widetilde{h}_{f_{i,n}}$ is a linear combination of $h_{f_{1,n}}, \dots, h_{f_{i,n}}$ with convergent and thus bounded coefficients. Furthermore, for $u \in C[0,1]$ due to partial integration (see for example Satz 6.3 in Walter (2002))

$$\begin{aligned} |\langle h_{f_{i,n}}, u \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}| &= \left| \int\limits_{[0,1]}^{(R)} h'_{f_{i,n}}(t)\, du(t) \right| \\ &= \left| h'_{f_{i,n}}(1)u(1) - h'_{f_{i,n}}(0)u(0) - \int\limits_{[0,1]}^{(R)} u(t)\, dh'_{f_{i,n}}(t) \right| \\ &= \left| f_i(1)u(1) - f_i\left(\frac{1}{n}\right) u(0) - \int\limits_{[0,1]} u(t)\, d \left[\sum_{j=1}^{n} \left(f_i\left(\frac{j}{n}\right) - f_i\left(\frac{j-1}{n}\right) \right) \delta_{\frac{j-1}{n}} \right](t) \right| \\ &\leq |f_i(1)| \cdot |u(1)| + \left| f_i\left(\frac{1}{n}\right) \right| \cdot |u(0)| + \|u\|_\infty \left[\sum_{j=2}^{n} \left| f_i\left(\frac{j}{n}\right) - f_i\left(\frac{j-1}{n}\right) \right| + \left| f_i\left(\frac{1}{n}\right) \right| \right] \\ &\leq [\|f_i\|_\infty + \|f_i\|_V] \cdot \|u\|_\infty \\ &\leq K_2 \cdot \|u\|_\infty \end{aligned}$$

holds true for some $K_2 > 0$, since f_i is of bounded variation. Note that K_2 is independent of n and note furthermore, that we interpreted $f_i\left(\frac{j-1}{n}\right)$ as zero for $j = 1$. Again, by Part 2 and Part 3, this implies the existence of $K_3 > 0$ satisfying

$$|\langle \widetilde{h}_{f_{i,n}}, u \rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}| \leq K_3 \cdot \|u\|_\infty$$

for all $n \geq n_0$, $i \in \{1, \dots, p\}$ and $u \in C[0,1]$.

Taking the above facts together, we can infer:

$$\|pr_{W_{n\mathcal{H}_B}}(u)\|_\infty \leq \sum_{i=1}^{p} |\langle \widetilde{h}_{f_{i,n}}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}| \cdot \|\widetilde{h}_{f_{i,n}}\|_\infty$$

$$\leq p \cdot K_3 \cdot \|u\|_\infty \cdot K_1$$

holds true for all $n \geq n_0$ with n_0 as in Part 2 and all $u \in C[0,1]$. This proves (7.102).

Part 5:
The extension $pr_{W_{\mathcal{H}_\mathcal{B}}}$ of orthogonal projection onto $W_{\mathcal{H}_\mathcal{B}}$ with respect to the bilinear mapping $\langle\cdot,\cdot\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}$ has the form

$$pr_{W_{\mathcal{H}_\mathcal{B}}} : C[0,1] \longrightarrow C[0,1],\ u \longmapsto \sum_{i=1}^{p} \langle \widetilde{h}_{f_i}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}\, \widetilde{h}_{f_i},$$

where $\widetilde{h}_{f_1}, \ldots, \widetilde{h}_{f_p}$ is the output of the Gram-Schmidt orthonormalization of $h_{f_1}, \ldots, h_{f_p}$ and therefore an orthonormal basis of $W_{\mathcal{H}_\mathcal{B}}$.
Then, there is a dense subspace $V \subset C[0,1]$, such that

$$\|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \xrightarrow[n\to\infty]{} 0, \quad u \in V.$$

Proof: Note that $\{u \in C^1([0,1]) \,|\, u(0) = 0\} \subseteq \mathcal{H}_B$ and that

$$span\{\mathbb{1}_{[0,1]}\} + \{u \in C^1([0,1]) \,|\, u(0) = 0\} = C^1([0,1])$$

is dense in $C([0,1])$. It is thus sufficient to show

$$\|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_\mathcal{B}}}(u)\|_\infty \xrightarrow[n\to\infty]{} 0$$

for $u = \mathbb{1}_{[0,1]}$ and for $u \in \mathcal{H}_B$. First, for $u = \mathbb{1}_{[0,1]}$ and $f \in BV_D(\mathcal{H}_B)$ we have

$$\langle f, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]} = \int_{[0,1]}^{(R)} f'\, du = 0,$$

which implies $pr_{W_{n\mathcal{H}_B}}(\mathbb{1}_{[0,1]}) = 0 \xrightarrow[n\to\infty]{} 0 = pr_{W_{\mathcal{H}_B}}(u)$. Furthermore, for $u \in \mathcal{H}_B$ and $f \in BV_D(\mathcal{H}_B)$ we have

$$\langle f, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]} = \int_{[0,1]}^{(R)} f'(x)\, du(x) = \int_{[0,1]} f'(x)u'(x)\, d\lambda_{[0,1]}(x) = \langle f, u\rangle_{\mathcal{H}_B}.$$

Therefore, together with Part 3, we can infer that for $u \in \mathcal{H}_B$

$$
\begin{aligned}
&\|pr_{W_n\mathcal{H}_B}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_{\mathcal{H}_B} \\
&\qquad = \left\| \sum_{i=1}^{p} \langle \widetilde{h}_{f_{i,n}}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}\, \widetilde{h}_{f_{i,n}} - \sum_{i=1}^{p} \langle \widetilde{h}_{f_i}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}\, \widetilde{h}_{f_i} \right\|_{\mathcal{H}_B} \\
&\qquad \leq \sum_{i=1}^{p} \| \langle \widetilde{h}_{f_{i,n}}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}\, \widetilde{h}_{f_{i,n}} - \langle \widetilde{h}_{f_i}, u\rangle_{BV_D(\mathcal{H}_B)\times C[0,1]}\, \widetilde{h}_{f_i} \|_{\mathcal{H}_B} \\
&\qquad = \sum_{i=1}^{p} \| \langle \widetilde{h}_{f_{i,n}}, u\rangle_{\mathcal{H}_B}\, \widetilde{h}_{f_{i,n}} - \langle \widetilde{h}_{f_i}, u\rangle_{\mathcal{H}_B}\, \widetilde{h}_{f_i} \|_{\mathcal{H}_B} \\
&\qquad \leq \sum_{i=1}^{p} \| \langle \widetilde{h}_{f_{i,n}} - \widetilde{h}_{f_i}, u\rangle_{\mathcal{H}_B}\, \widetilde{h}_{f_{i,n}} \|_{\mathcal{H}_B} + \| \langle \widetilde{h}_{f_i}, u\rangle_{\mathcal{H}_B}\, (\widetilde{h}_{f_{i,n}} - \widetilde{h}_{f_i}) \|_{\mathcal{H}_B} \\
&\qquad \xrightarrow[n\to\infty]{} \sum_{i=1}^{p} \| \langle \widetilde{h}_{f_i} - \widetilde{h}_{f_i}, u\rangle_{\mathcal{H}_B}\, \widetilde{h}_{f_i} \|_{\mathcal{H}_B} + \| \langle \widetilde{h}_{f_i}, u\rangle_{\mathcal{H}_B}\, (\widetilde{h}_{f_i} - \widetilde{h}_{f_i}) \|_{\mathcal{H}_B} = 0
\end{aligned}
$$

holds true.

Now, for $g \in \mathcal{H}_B$, $\|g\|_\infty \leq \|g\|_{\mathcal{H}_B}$ holds true, which can be seen in the following way: According to the Cauchy-Schwarz inequality for all $t \in (0,1]$

$$
|\langle k, l\rangle_{\mathcal{L}^2([0,t],\lambda\!\!\!\lambda_{[0,t]})}| \leq \|k\|_{\mathcal{L}^2([0,t],\lambda\!\!\!\lambda_{[0,t]})} \cdot \|l\|_{\mathcal{L}^2([0,t],\lambda\!\!\!\lambda_{[0,t]})}
$$

holds true. If we set $k = g'$ and $l \equiv 1$, we thus get

$$
\begin{aligned}
\left| \int_{[0,t]} g'(x)\, d\lambda\!\!\!\lambda_{[0,t]}(x) \right| &\leq \sqrt{\int_{[0,t]} g'(x)^2\, d\lambda\!\!\!\lambda_{[0,t]}(x)} \cdot \sqrt{\int_{[0,t]} 1\, d\lambda\!\!\!\lambda_{[0,t]}(x)} \\
&\leq \sqrt{\int_{[0,1]} g'(x) g'(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x)}
\end{aligned}
$$

and therefore

$$
|g(t)| = |g(t) - g(0)| \leq \sqrt{\int_{[0,1]} g'(x) g'(x)\, d\lambda\!\!\!\lambda_{[0,1]}(x)} = \|g\|_{\mathcal{H}_B}, \quad t \in (0,1],
$$

holds true, since $g(0) = 0$. As a consequence, $\sup_{t\in[0,1]} |g(t)| \leq \|g\|_{\mathcal{H}_B}$ follows.

We thus conclude that,

$$
\|pr_{W_n\mathcal{H}_B}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \xrightarrow[n\to\infty]{} 0
$$

holds true for $u \in \mathcal{H}_B$, which, as seen above, completes the proof of Part 4.

Part 6:
According to Part 4 there is a $K > 0$ such that

$$\|pr_{W_{n\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} \leq K$$

and according to Part 5 there is a dense subspace $V \subset C[0,1]$ such that

$$\|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \xrightarrow[n\to\infty]{} 0, \quad u \in V,$$

holds true. Therefore, since $\|pr_{W_{\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} < \infty$, which can be shown similar to $\|pr_{W_{n\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} < \infty$, by enlarging K, we can assume that

$$\|pr_{W_{\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} \leq \widetilde{K}$$

holds true as well and then denote $\widetilde{K}$ by K. We can then infer that

$$\|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \xrightarrow[n\to\infty]{} 0, \quad u \in C[0,1],$$

holds true, since given such an $u \in C[0,1]$ and $\varepsilon > 0$, one can choose $v \in V$ such that $\|u - v\|_\infty \leq \frac{\varepsilon}{3K}$ and $n_0 \in \mathbb{N}$ such that $\|pr_{W_{n\mathcal{H}_B}}(v) - pr_{W_{\mathcal{H}_B}}(v)\|_\infty \leq \frac{\varepsilon}{3}$ for all $n \geq n_0$ and thus

$$\begin{aligned}
&\|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \\
&\leq \|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{n\mathcal{H}_B}}(v)\|_\infty + \|pr_{W_{n\mathcal{H}_B}}(v) - pr_{W_{\mathcal{H}_B}}(v)\|_\infty + \|pr_{W_{\mathcal{H}_B}}(v) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \\
&\leq \|pr_{W_{n\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} \cdot \|u - v\|_\infty + \frac{\varepsilon}{3} + \|pr_{W_{\mathcal{H}_B}}\|_{C[0,1]\to C[0,1]} \cdot \|v - u\|_\infty \\
&\leq K \cdot \frac{\varepsilon}{3K} + \frac{\varepsilon}{3} + K \cdot \frac{\varepsilon}{3K} \\
&= \varepsilon, \quad n \geq n_0.
\end{aligned}$$

Finally, if $(u_n)_{n\in\mathbb{N}} \subset C[0,1]$ with $u_n \xrightarrow[n\to\infty]{} u \in C[0,1]$ then, as desired,

$$\begin{aligned}
\|pr_{W_{n\mathcal{H}_B}}(u_n) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty &\leq \|pr_{W_{n\mathcal{H}_B}}(u_n) - pr_{W_{n\mathcal{H}_B}}(u)\|_\infty + \|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \\
&\leq K \cdot \|u_n - u\|_\infty + \|pr_{W_{n\mathcal{H}_B}}(u) - pr_{W_{\mathcal{H}_B}}(u)\|_\infty \\
&\xrightarrow[n\to\infty]{} 0
\end{aligned}$$

holds true, which proves (7.99). □

7.5 Proofs for Chapter 6

The following lemma is the result of a strict forward computation, for which we provide an explicit proof for the convenience of the reader.

Lemma 7.5.1.

Let the presumptions and notations of Section 6.1.1 and Section 6.1.2 be given. In particular, $\mathcal{F} := \{F(\cdot,\theta) : \mathbb{R} \to [0,1] \,|\, \theta \in \Theta\}$ is the hypothesised model family of absolutely continuous distribution functions $F(\cdot,\theta)$, parametrised by a parameter θ from an open set $\Theta \subseteq \mathbb{R}^p$, $\theta_0 \in \Theta$, $f(t,\theta) = \frac{\partial F(F^-(t,\theta_0),\theta)}{\partial \theta}$ and $f'(t,\theta_0) = \frac{\partial f(t,\theta_0)}{\partial t}$. Then

$$f'(t,\theta_0) = \left(\frac{\partial}{\partial\theta}\log\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)$$

holds true.

Proof: It holds true that

$$\begin{aligned}
f'(t,\theta_0) &= \frac{\partial}{\partial t}\frac{\partial}{\partial\theta}\left[F(F^-(t,\theta_0),\theta)\right]\Big|_{\theta=\theta_0} = \frac{\partial}{\partial t}\left[\left(\frac{\partial}{\partial\theta}F\right)(F^-(t,\theta_0),\theta)\right]\Big|_{\theta=\theta_0}\cdot\frac{\partial}{\partial\theta}\theta \\
&= \left(\frac{\partial}{\partial t}\frac{\partial}{\partial\theta}F\right)(F^-(t,\theta_0),\theta)\Big|_{\theta=\theta_0}\cdot\left(\frac{\partial}{\partial t}F^-\right)(t,\theta_0) \\
&= \frac{\left(\frac{\partial}{\partial t}\frac{\partial}{\partial\theta}F\right)(F^-(t,\theta_0),\theta_0)}{\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)} \qquad (7.103) \\
&= \frac{\left(\frac{\partial}{\partial\theta}\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)}{\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)} \qquad (7.104) \\
&= \frac{\left(\frac{\partial}{\partial\theta}\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)\cdot\frac{\partial}{\partial\theta}\theta}{\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)} = \frac{\frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta)\right]\big|_{\theta=\theta_0}}{\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)} \\
&= \frac{\partial}{\partial\theta}\left[\log\left(\left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta)\right)\right]\Big|_{\theta=\theta_0} \\
&= \left(\frac{\partial}{\partial\theta}\log\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0).
\end{aligned}$$

Note that we could change the order of derivatives in equation (7.104) since, according to our assumptions, $F(F^-(t,\theta_0),\theta)$ is continuously differentiable in t and θ. Furthermore, equation (7.103) holds true, since $F(F^-(t,\theta_0),\theta_0) = t$ is fulfilled, from which

$$\frac{\partial}{\partial t}\left[F(F^-(t,\theta_0),\theta)\right]\Big|_{\theta=\theta_0} = \left(\frac{\partial}{\partial t}F\right)(F^-(t,\theta_0),\theta_0)\cdot\left(\frac{\partial}{\partial t}F^-\right)(t,\theta_0) = \frac{\partial}{\partial t}t$$

can be inferred by derivation on both sides, which in turn is equivalent to

$$\left(\frac{\partial}{\partial t}F^{-}\right)(t,\theta_0) = \frac{1}{\left(\frac{\partial}{\partial t}F\right)\left(F^{-}(t,\theta_0),\theta_0\right)}.$$

□

Proof of the two assertions in Remark 6.3.1

It holds true that

$$\left(\frac{1}{n}\underbrace{\begin{pmatrix}(f_1\cdot\mathbb{1}_{(c,b]})(X_{n1}) & \dots & (f_1\cdot\mathbb{1}_{(c,b]})(X_{nn})\\ \vdots & \ddots & \vdots\\ (f_p\cdot\mathbb{1}_{(c,b]})(X_{n1}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(X_{nn})\end{pmatrix}\begin{pmatrix}(f_1\cdot\mathbb{1}_{(c,b]})(X_{n1}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(X_{n1})\\ \vdots & \ddots & \vdots\\ (f_1\cdot\mathbb{1}_{(c,b]})(X_{nn}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(X_{nn})\end{pmatrix}}_{=:A_n}\right)_{k,l}$$

$$= \left(\frac{1}{n}\sum_{i=1}^{n}(f_k\cdot f_l\cdot\mathbb{1}_{(c,b]})(X_{ni})\right)_{k,l} \xrightarrow[n\to\infty]{a.s.} \left(\int_{[-\infty,\infty]}\left(f_k\cdot f_l\cdot\mathbb{1}_{(c,b]}\right)(x)\,dF(x)\right),\ 1\le k,l\le p,$$

and thus,

$$\frac{1}{n}A_n^TA_n \xrightarrow[n\to\infty]{a.s.} \int_{[-\infty,\infty]}\begin{pmatrix}f_1\cdot\mathbb{1}_{(c,b]}\\ \vdots\\ f_p\cdot\mathbb{1}_{(c,b]}\end{pmatrix}(x)\left(f_1\cdot\mathbb{1}_{(c,b]},\dots,f_p\cdot\mathbb{1}_{(c,b]}\right)(x)\,dF(x). \tag{7.105}$$

Since the matrix on the right hand side of (7.105) is invertible (as $f_1\cdot\mathbb{1}_{(c,b]},\dots,f_p\cdot\mathbb{1}_{(c,b]}$ are linearly independent in $\mathcal{L}^2([a,b],F)$) and since $\det:\mathbb{R}^{p\times p}\longrightarrow\mathbb{R}$ is continuous, for almost all realisations $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ there exists an $n_0\in\mathbb{N}$ such that

$$\underbrace{\begin{pmatrix}(f_1\cdot\mathbb{1}_{(c,b]})(t_{n1}) & \dots & (f_1\cdot\mathbb{1}_{(c,b]})(t_{nn})\\ \vdots & \ddots & \vdots\\ (f_p\cdot\mathbb{1}_{(c,b]})(t_{n1}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(t_{nn})\end{pmatrix}\begin{pmatrix}(f_1\cdot\mathbb{1}_{(c,b]})(t_{n1}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(t_{n1})\\ \vdots & \ddots & \vdots\\ (f_1\cdot\mathbb{1}_{(c,b]})(t_{nn}) & \dots & (f_p\cdot\mathbb{1}_{(c,b]})(t_{nn})\end{pmatrix}}_{=:B_n}$$

is invertible for all $n\ge n_0$ and thus $B_n^TB_n$ has rank p for all $n\ge n_0$. Furthermore, since

$$rank(B_n^TB_n) = rank(B_n) = rank\left(\begin{pmatrix}f_1(t_{nj_n(c)+1}) & \dots & f_p(t_{nj_n(c)+1})\\ \vdots & \ddots & \vdots\\ f_1(t_{nn}) & \dots & f_p(t_{nn})\end{pmatrix}\right)$$

$$= rank\left(X_n^{(j_n(c)+1)}\right) = rank\left((X_n^{(j_n(c)+1)})^TX_n^{(j_n(c)+1)}\right),$$

we can conclude that for almost all realisations $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ there exists

an $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$, $(X_n^{(j_n(c)+1)})^T X_n^{(j_n(c)+1)} \in \mathbb{R}^{p\times p}$ has rank p. Note that then, of course, $X_n^{(1)}, \ldots, X_n^{(j_n(c))}$ have full rank, too.
Therefore, we have show that almost all realisations $((t_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ of $((X_{ni})_{i=1}^n)_{n\in\mathbb{N}}$ satisfy Assumption (6.24). □

Lemma 7.5.2.

The following statements about the backwards recursive residuals, defined in Definition 6.3.2, are true:

i) $\mathbb{E}(r_{ni}) = 0$ and $\mathbb{V}ar(r_{ni}) = \sigma^2, \quad i = 1, \ldots, j_n(c)$.

ii) The i-th backwards recursive residual r_{ni}, $i = 1, \ldots, j_n(c)$, is a linear combination of the last $n - i + 1$ regression errors $\varepsilon_{ni}, \ldots, \varepsilon_{nn}$. To be more specific,

$$r_{ni} = \frac{1}{c_{ni}}(1, b_{ni})\, \varepsilon_n^{(i)}$$

with

$$b_{ni} := -f^T(t_{ni}) \left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1} (X_n^{(i+1)})^T \in \mathbb{R}^{1\times(n-i)}$$

and

$$c_{ni} := \sqrt{1 + f^T(t_{ni}) \left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1} f(t_{ni})}.$$

iii) The backwards recursive residual vector r_n can be computed at once, using a linear transformation of the regression error vector ε_1. To be more specific,

$$r_n = M_n\, \varepsilon_n^{(1)}$$

holds true with

$$M_n := \begin{pmatrix} \frac{1}{c_{n1}} & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \frac{1}{c_{nj_n(c)}} \end{pmatrix} \begin{pmatrix} 1 & b_{n1} & & & \\ 0 & 1 & b_{n2} & & \\ \vdots & \ddots & \ddots & \ddots & \\ 0 & \ldots & 0 & 1 & b_{nj_n(c)} \end{pmatrix} \in \mathbb{R}^{j_n(c)\times n}.$$

iv) The rows of M_n are orthogonal to the columns of $X_n^{(1)}$, i.e.

$$M_n X_n^{(1)} = 0 \in \mathbb{R}^{j_n(c)\times p}.$$

v) The rows of M_n are an orthonormal system and thus $Cov(r_n) = \sigma^2 I_{j_n(c)}$ holds true. In particular, for $\varepsilon_n^{(1)} \sim N_n(0, \sigma^2 I_n)$, the random variables $r_{n1}, \ldots, r_{nj_n(c)}$ are stochastically independent.

vi) The relation $M_n^T M_n = pr_{X_n^{(1)\perp}}$ holds true. Therefore, we have $\hat{\varepsilon}_n = M_n^T\, M_n\, \varepsilon_n^{(1)}$. As a consequence, $r_n = M_n\, \hat{\varepsilon}_n$ holds true.

Lemma 7.5.3.

Let the prerequisites of Section 6.3 be given. Let $\mathcal{X}_{\min}(A)$ and $\mathcal{X}_{\max}(A)$ denote the minimal and maximal eigenvalue of a symmetric matrix A. Then the following statements hold true:

i) Let $H(t) := \int\limits_{[t,1]} (ff^T) \circ F^- \, d\lambda\!\!\!^{-}_{[0,1]}$, $t \in [0, F(c)]$. Then there are $\mathcal{X}_{\min}, \mathcal{X}_{\max} \in (0, \infty)$ such that

$$\mathcal{X}_{\min} \leq \mathcal{X}_{\min}(H(t)) \leq \mathcal{X}_{\max}(H(t)) \leq \mathcal{X}_{\max}$$

holds true for all $t \in [0, F(c)]$.

ii) For all $m \in \mathbb{N}_0$

$$\sup_{t \in [0,1]} \left\| \frac{1}{n} (X_n^{(\lceil tn \rceil + m)})^T X_n^{(\lceil tn \rceil + m)} - \int\limits_{[t,1]} (ff^T) \circ F^- \, d\lambda\!\!\!^{-}_{[0,1]} \right\| \xrightarrow[n \to \infty]{} 0$$

and

$$\sup_{t \in [0, F(c)]} \left\| \left(\frac{1}{n} (X_n^{(\lceil tn \rceil + m)})^T X_n^{(\lceil tn \rceil + m)} \right)^{-1} - \left(\int\limits_{[t,1]} (ff^T) \circ F^- \, d\lambda\!\!\!^{-}_{[0,1]} \right)^{-1} \right\| \xrightarrow[n \to \infty]{} 0$$

hold true for all matrix norms (since matrix norms are equivalent in finite dimensional vector spaces).

iii) There are $\widetilde{\mathcal{X}}_{\min}, \widetilde{\mathcal{X}}_{\max} \in (0, \infty)$ such that for any fixed $m \in \mathbb{N}_0$ there is a $n_0 \in \mathbb{N}$ such that

$$\begin{aligned} \widetilde{\mathcal{X}}_{\min} &\leq \mathcal{X}_{\min} \left(\frac{1}{n} (X_n^{(\lceil tn \rceil + m)})^T X_n^{(\lceil tn \rceil + m)} \right) \\ &\leq \mathcal{X}_{\max} \left(\frac{1}{n} (X_n^{(\lceil tn \rceil + m)})^T X_n^{(\lceil tn \rceil + m)} \right) \leq \widetilde{\mathcal{X}}_{\max} \end{aligned}$$

holds true for all $t \in [0, F(c)]$ and all $n \geq n_0$.

iv) There are $r > 0$ and $n_0 \in \mathbb{N}$ such that $1 \leq c^2_{n\lceil tn \rceil} \leq 1 + \frac{r}{n}$ holds true for all $t \in [0, F(c)]$ and $n \geq n_0$. Therefore,

$$\sup_{i \in \{1, \ldots, j_n(c)\}} c_{ni} \xrightarrow[n \to \infty]{} 1$$

is satisfied.

Proofs: The proofs of Lemma 7.5.2 and Lemma 7.5.3 are similar to those in Brown, Durbin, and Evans (1975) and Evers (2022) for "normal" (i.e. non backwards) recursive residuals and are therefore omitted here. The only difference is that in the case of backwards recursive residuals, we have to apply the proofs given therein to the time inverted data. □

Proof of Theorem 6.3.3

For $x \in (-\infty, c]$ the equations

$$(T^*(R_n^1))(x) = R_n^1(x) - \int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \left(\int\limits_{(y,\infty)} f(z)\, dR_n^1(z) \right) dF(y) \tag{7.106}$$

$$\begin{aligned} &= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,x]}(t_{ni})\, \hat{\varepsilon}_{ni} - \int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, dF(u) \right)^{-1} \times \\ &\quad \times \frac{1}{\sigma\sqrt{n}} \, (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\, dF(y) \\ &= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \hat{\varepsilon}_{ni} \; - \; \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, dF(u) \right)^{-1} \times \\ &\quad \times \; (X_n^{(j_n(y)+1)})^T \, \hat{\varepsilon}_n^{(j_n(y)+1)}\, dF(y) \end{aligned} \tag{7.107}$$

$$\begin{aligned} &= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \hat{\varepsilon}_{ni} \; - \; \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(u) \right)^{-1} \times \\ &\quad \times \; (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(y) \; + \; o_{\mathbb{P}}(1) \end{aligned} \tag{7.108}$$

$$\begin{aligned} &= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \hat{\varepsilon}_{ni} - \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} f^T(t_{ni}) \left((X_n^{(i+1)})^T X_n^{(i+1)} \right)^{-1} (X_n^{(i+1)})^T \hat{\varepsilon}_n^{(i+1)} \; + \; o_{\mathbb{P}}(1) \\ &= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} \left(\hat{\varepsilon}_{ni} - f^T(t_{ni}) \left((X_n^{(i+1)})^T \, X_n^{(i+1)} \right)^{-1} (X_n^{(i+1)})^T \, \hat{\varepsilon}_n^{(i+1)} \right) \; + \; o_{\mathbb{P}}(1) \end{aligned} \tag{7.109}$$

$$= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} r_{ni} \; + \; o_{\mathbb{P}}(1)$$

$$= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(c)} \mathbb{1}_{(-\infty,x]}(t_{ni}) \cdot r_{ni} \; + \; o_{\mathbb{P}}(1)$$

$$= \widetilde{R}_n^1(x) \; + \; o_{\mathbb{P}}(1).$$

are satisfied for $n \to \infty$. Here, as defined in Remark 6.3.1, $j_n : \mathbb{R} \longrightarrow \{0, 1, \ldots, n\},\ y \longmapsto \max\{0, \sup\{i \in \{1, \ldots, n\} \,|\, t_{ni} \leq y\}\}$ and $\delta_{\{t_{ni}\}}$ denotes the one point measure in t_{ni}.

In order to thoroughly prove the above equations, we take a closer look at the enumerated equalities (7.106) - (7.109). Note that in the context of this proof, we denote by $\|M\| = \sup_{\|x\|_2=1} \|Mx\|_2$ the operator norm on the space $\mathbb{R}^{m\times n}$ of $m \times n$ matrices.

Regarding equation (7.106)**:** Since

$$\begin{aligned}
\int_{(y,\infty)} f(z)\, dR_n^1(z) &= \int_{(y,\infty)} f(z)\, d\left(\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} \mathbb{1}_{(-\infty,\cdot]}(t_{ni}) \cdot \hat{\varepsilon}_{ni}\right)(z) \\
&= \frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{n} \hat{\varepsilon}_{ni} \int_{(y,\infty)} f(z)\, d\delta_{\{t_{ni}\}}(z) \\
&= \frac{1}{\sigma\sqrt{n}} \sum_{i=j_n(y)+1}^{n} \hat{\varepsilon}_{ni}\, f(t_{ni}) \\
&= \frac{1}{\sigma\sqrt{n}} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\,.
\end{aligned}$$

is valid, the equation in question holds true.

Regarding equation (7.107)**:** It has to be verified that

$$\frac{1}{\sigma\sqrt{n}} \int_{(-\infty,x]} f^T(y) \left(\int_{(y,\infty)} f(u)\, f^T(u)\, dF(u)\right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\, dF(y)$$

$$= \frac{1}{\sigma\sqrt{n}} \int_{(-\infty,x]} f^T(y) \left(\int_{(y,\infty)} f(u) f^T(u) d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\{t_{ni}\}}\right)(u)\right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} d\left(\frac{1}{n}\sum_{i=1}^{n} \delta_{\{t_{ni}\}}\right)(y) + o_{\mathbb{P}}(1)$$

holds true for $x \in (-\infty, c]$ and $n \to \infty$. By using the notations

$$k_n(y) := f^T(y) \left(\int_{(y,\infty)} f(u)\, f^T(u)\, dF(u)\right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\,,$$

$$l_n(y) := f^T(y) \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(u) \right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} ,$$

$$\hat{m}_n(y) := (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} ,$$

$$\mu_n := \frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}}$$

and identifying the distribution function F with the unique probability measure μ corresponding to it, equation (7.107) reads as follows:

$$\frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) = \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} l_n(y)\, d\mu_n(y) \; + \; o_{\mathbb{P}}(1)$$

holds true for $x \in (-\infty, c]$ and $n \to \infty$. By definition of the $o_P(\mathbb{P})$ notation, proving this is equivalent to proving

$$\left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) \; - \int\limits_{(-\infty,x]} l_n(y)\, d\mu_n(y) \right) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0\,. \tag{7.110}$$

By using the triangle inequality, we get

$$\begin{aligned} &\left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) \; - \int\limits_{(-\infty,x]} l_n(y)\, d\mu_n(y) \right) \right| \\ &\qquad \le \left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) - \int\limits_{(-\infty,x]} k_n(y)\, d\mu_n(y) \right) \right| \\ &\qquad\quad + \left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu_n(y) - \int\limits_{(-\infty,x]} l_n(y)\, d\mu_n(y) \right) \right| . \end{aligned}$$

Thus, for (7.110) to be satisfied, it suffices to show that both

$$\left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) \; - \int\limits_{(-\infty,x]} k_n(y)\, d\mu_n(y) \right) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0 \tag{7.111}$$

and

$$\left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu_n(y) \; - \int\limits_{(-\infty,x]} l_n(y)\, d\mu_n(y) \right) \right| \xrightarrow[n\to\infty]{\mathbb{P}} 0 \tag{7.112}$$

are true. Therefore, in **Step I,** (7.111) will be proven and in **Step II,** we will prove

(7.112), which together completes the proof of equation (7.107).

Step I:

By identifying the probability measures μ and μ_n with their corresponding distribution functions F and F_n, we can rewrite (7.111) into

$$\left| \frac{1}{\sigma\sqrt{n}} \left(\int\limits_{(-\infty,x]} k_n(y)\, d\mu(y) \; - \int\limits_{(-\infty,x]} k_n(y)\, d\mu_n(y) \right) \right| = \left| \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} k_n(y)\, d(F - F_n)(y) \right| . \tag{7.113}$$

The integrand k_n can be divided in a deterministic and a random part in the following way:

$$\begin{aligned}
k_n(y) &= f^T(y) \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, dF(u) \right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} \\
&= \left(\sum_{i=1}^{p} f_i(y) \left[\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right]_{i1} , \ldots \right. \\
&\qquad \left. \ldots , \sum_{i=1}^{p} f_i(y) \left[\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right]_{ip} \right) (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} \\
&= \left(\sum_{q=1}^{p} \left[\sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{iq} \right] f_q(t_{nj_n(y)+1}) , \ldots \right. \\
&\qquad \left. \ldots , \sum_{q=1}^{p} \left[\sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{iq} \right] f_q(t_{nn}) \right) \hat{\varepsilon}_n^{(j_n(y)+1)} \\
&= \sum_{r=1}^{n-j_n(y)} \left(\sum_{q=1}^{p} \left[\sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{iq} \right] f_q(t_{nj_n(y)+r}) \right) \hat{\varepsilon}_n^{(j_n(y)+1)} \\
&= \sum_{r=1}^{n} \underbrace{\left(\mathbb{1}_{(y,\infty)}(t_{nr}) \cdot \sum_{q=1}^{p} \left[\sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{iq} \right] f_q(t_{nr}) \right)}_{=: F_r^{(n)}(y)} \cdot \hat{\varepsilon}_{nr} .
\end{aligned}$$

Thus, together with (7.113), we have

$$\begin{aligned}\left|\frac{1}{\sigma\sqrt{n}}\left(\int\limits_{(-\infty,x]} k_n(y)\,d\mu(y) - \int\limits_{(-\infty,x]} k_n(y)\,d\mu_n(y)\right)\right| &= \left|\frac{1}{\sigma\sqrt{n}}\int\limits_{(-\infty,x]}\sum_{r=1}^{n} F_r^{(n)}(y)\,\hat{\varepsilon}_{nr}\,d(F-F_n)(y)\right| \\ &= \left|\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\underbrace{\int\limits_{(-\infty,x]} F_r^{(n)}(y)\,d(F-F_n)(y)}_{=:d_{nr}(x)}\right| \\ &= \left|\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)\right|.\end{aligned}$$

Therefore, for (7.111) to be true, it suffices to show that

$$\left|\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)\right| \xrightarrow[n\to\infty]{\mathbb{P}} 0, \quad x\in(-\infty,c]. \tag{7.114}$$

To do so, first note that

$$\begin{aligned}\mathbb{E}\left(\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)\right) &= \frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n} d_{nr}(x)\mathbb{E}\left(\hat{\varepsilon}_{nr}\right) \\ &= \frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n} d_{nr}(x) f^T(t_{nr})\left[\beta - \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\mathbb{E}\left((X_n^{(1)})^T\beta + \varepsilon_n^{(1)}\right)\right] \\ &= 0\,.\end{aligned}$$

In the remainder of Step I, we are going to show that for all $x\in(-\infty,c]$,

$$\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)\right) \xrightarrow[n\to\infty]{} 0 \tag{7.115}$$

is satisfied. Therefore, because $\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)$ is centred and due to Chebyshev's inequality, (7.114) holds true, completing the proof. We thus assess

$$\begin{aligned}\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}}\sum_{r=1}^{n}\hat{\varepsilon}_{nr}\,d_{nr}(x)\right) &= \frac{1}{\sigma^2 n}\,\mathbb{V}ar\left(d_n^T(x)\hat{\varepsilon}_n^{(1)}\right) \\ &= \frac{1}{\sigma^2 n}\,d_n^T(x)\mathbb{V}ar\left(\hat{\varepsilon}_n^{(1)}\right)d_n(x) \\ &= \frac{1}{\sigma^2 n}\,d_n^T(x)\cdot\sigma^2\left(I_n - X_n^{(1)}\left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\right)d_n(x) \\ &= \frac{1}{n}\left\|d_n(x)\right\|_2^2 - \frac{1}{n}\,d_n^T(x)\left(X_n^{(1)}\left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\right)d_n(x)\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{n}\left\|d_n(x)\right\|_2^2 + \frac{1}{n}\left\|d_n(x)\right\|_2^2 \left\|X_n^{(1)}\left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\right\| \\
&= \frac{1}{n}\left\|d_n(x)\right\|_2^2 \left(1 + \left\|X_n^{(1)}\left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\right\|\right) \\
&\leq \frac{2}{n}\left\|d_n(x)\right\|_2^2 \\
&= \frac{2}{n}\sum_{l=1}^{n} d_{nl}^2(x), \qquad (7.116)
\end{aligned}
$$

wherein I_n denotes the unity matrix in $\mathbb{R}^{n\times n}$, $d_n(x) := (d_{n1}(x), \ldots, d_{nn}(x))^T$ and the last inequality is true due to the following lemma.

Lemma 7.5.4.

In the situation of Definition 6.3.2, page 94, it holds true that

$$\left\|X_n^{(1)}\left((X_n^{(1)})^T X_n^{(1)}\right)^{-1}(X_n^{(1)})^T\right\| \leq 1\,.$$

Proof: Let

$$X_n^{(1)} = U\Sigma V^T$$

be the singular value decomposition of $X_n^{(1)} \in \mathbb{R}^{n\times p}$ with $U \in \mathbb{R}^{n\times n}$ and $V \in \mathbb{R}^{p\times p}$ orthonormal matrices and

$$\Sigma = \begin{pmatrix} \sigma_1 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \sigma_p \\ 0 & \ldots & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & \ldots & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{n\times p},$$

where

$$\sigma_1 \geq \cdots \geq \sigma_p > 0$$

are the p strictly positive singular values (since $rank(X_1) = p$). Note that because Σ has linear independent columns, there is a Moore-Penrose inverse

$$\Sigma^\dagger = \begin{pmatrix} \frac{1}{\sigma_1} & 0 & \ldots & 0 & 0 & \ldots & 0 \\ 0 & \ddots & \ddots & \vdots & \vdots & \ldots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots & \ldots & \vdots \\ 0 & \ldots & 0 & \frac{1}{\sigma_p} & 0 & \ldots & 0 \end{pmatrix} \in \mathbb{R}^{p\times n},$$

since

$$\Sigma^\dagger \Sigma = I_p.$$

Furthermore, since $X_n^{(1)}$ has linear independent columns, there is the Moore-Penrose inverse

$$(X_n^{(1)})^\dagger = \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T$$

of $X_n^{(1)}$, which is a left inverse. Thus,

$$(X_n^{(1)})^\dagger = \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T = (V\Sigma^T U^T U \Sigma V^T)^{-1} V \Sigma^T U^T = V \Sigma^\dagger U^T .$$

As a consequence,

$$\begin{aligned} X_n^{(1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T = X_n^{(1)} (X_n^{(1)})^\dagger = U\Sigma V^T V \Sigma^\dagger U^T \\ = U \Sigma \Sigma^\dagger U^T = U \underbrace{\left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array}\right)}_{\in \mathbb{R}^{n\times n}} U^T. \end{aligned}$$

Hence, for all vectors $v \in \mathbb{R}^n$

$$\left\| X_n^{(1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T v \right\|_2 = \left\| U \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array}\right) U^T v \right\|_2 = \left\| \left(\begin{array}{c|c} I_p & 0 \\ \hline 0 & 0 \end{array}\right) v \right\|_2 \leq \| v \|_2$$

holds true, thus

$$\left\| X_n^{(1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T \right\| \leq 1,$$

which had to be proved. □

For (7.115) to be true, due to (7.116), it suffices to show that for all $x \in (-\infty, c]$

$$\frac{2}{n} \sum_{r=1}^{n} d_{nr}^2(x) \xrightarrow[n\to\infty]{} 0. \tag{7.117}$$

By the definition of $d_{nr}(x)$ and integration by parts, for $x \in (-\infty, c]$, we can infer

$$\begin{aligned} d_{nr}(x) &= \int_{(-\infty,x]} F_r^{(n)}(y)\, d(F - F_n)(y) \\ &= F_r^{(n)}(x)\,(F - F_n)(x) - F_r^{(n)}(-\infty)\,(F - F_n)(-\infty) - \int_{(-\infty,x]} (F - F_n)(y)\, dF_r^{(n)}(y). \end{aligned} \tag{7.118}$$

$F_r^{(n)}(x)$ is bounded and $F_r^{(n)}(-\infty)$ is finite, because $f = (f_1, \ldots, f_p)$ is bounded and

$\left(\int\limits_{(x,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1}$ is bounded, since, according to Assertion 1 of the proof of Lemma 7.5.2, part iii), the eigenvalues of $\int\limits_{(x,\infty)} f(u) f^T(u)\, dF(u)$ are bounded away from zero for all $x \in (-\infty, c]$. Furthermore, since

$$F_n \xrightarrow[n\to\infty]{} F \quad \text{uniformly,} \tag{7.119}$$

we have $(F - F_n)(x) \xrightarrow[n\to\infty]{} 0$ for all $x \in (-\infty, c]$ and

$$(F - F_n)(-\infty) = F(-\infty) - F_n(-\infty) = 0, \tag{7.120}$$

since both are distribution functions. Therefore,

$$F_r^{(n)}(x)\,(F - F_n)(x) \xrightarrow[n\to\infty]{} 0$$

as well as

$$F_r^{(n)}\,(-\infty)(F - F_n)(-\infty) \;=\; 0\,.$$

It remains to show that the last term in (7.118) is bounded for $x \in (-\infty, c]$. First, the inequality

$$\left| \int\limits_{(-\infty,x]} (F - F_n)(y)\, dF_l^{(n)}(y) \right| \le \|F - F_n\|_\infty \cdot \|F_l^{(n)}\|_V \tag{7.121}$$

holds true. We already know that $\|F - F_n\|_\infty \xrightarrow[n\to\infty]{} 0$ as F_n converges to F uniformly. Therefore, if we were able to show that

$$\sup_{n\in\mathbb{N}} \max_{r=1,\ldots,n} \|F_r^{(n)}\|_V < \infty \tag{7.122}$$

is true, we can conclude that

$$\max_{r=1,\ldots,n} d_{nr}(x) \xrightarrow[n\to\infty]{} 0\,. \tag{7.123}$$

With this in mind, we start our examination of $\|F_r^{(n)}\|_V$. Firstly, $\|f_j\|_V < \infty$ for $j = 1, \ldots, p$ by assumption. Since the product of any two functions of bounded variation is again of bounded variation, we have $\|f_i \cdot f_j\|_V < \infty$ for $1 \le i, j \le p$, hence

$$\|(f\, f^T)_{i,j}\|_V < \infty, \quad 1 \le i, j \le p.$$

Now, denote by g_{ij}^+ and g_{ij}^- the positive part and the negative part of $f_i \cdot f_j$ respectively.

Then, $\left(f\, f^T\right)_{i,j} = g_{ij}^+ - g_{ij}^-$ thus

$$\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)_{ij} = \int\limits_{(y,\infty)} g_{ij}^+\, dF(u) - \int\limits_{(y,\infty)} g_{ij}^-\, dF(u)\,.$$

Both integrals on the right hand side of the equation above are monotonically decreasing in y, are bounded below by zero and bounded above by $\|g_{ij}^+\|_\infty$ and $\|g_{ij}^-\|_\infty$ respectively. Thus,

$$\left\| \int\limits_{(y,\infty)} g_{ij}^+\, dF(u) \right\|_V \leq \|g_{ij}^+\|_\infty \quad \text{and} \quad \left\| \int\limits_{(y,\infty)} g_{ij}^-\, dF(u) \right\|_V \leq \|g_{ij}^-\|_\infty$$

hold true and hence, for $1 \leq i, j \leq p$,

$$\left\| \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)_{ij} \right\|_V \leq \|g_{ij}^+\|_\infty + \|g_{ij}^-\|_\infty \leq 2 \cdot \|(f\, f^T)_{ij}\|_\infty < \infty$$

holds true. Since $\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u)$ is invertible for all $y \in (-\infty, c]$, by using Cramer's rule for computing inverse matrices, we get

$$\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} = \frac{1}{\det\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)} \cdot adj \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right),$$

where for a square matrix M, $adj(M)$ denotes the adjugate matrix to M. As a consequence, we can compute the ij-th component of $\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \in \mathbb{R}^{p \times p}$ in the following way:

$$\begin{aligned}
&\left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{ij} \\
&\qquad = \frac{(-1)^{i+j}}{\det\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)} \cdot \det \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^p_{\substack{k,l=1 \\ k \neq i, j \neq l}} \right) \\
&\qquad = \frac{(-1)^{i+j} \cdot \sum\limits_{\nu \in S_{p-1}} \left[sgn(\nu) \cdot \prod\limits_{s=1}^{p-1} \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^p_{\substack{k,l=1 \\ k \neq i, j \neq l}} \right)_{s\nu(s)} \right]}{\sum\limits_{\nu \in S_p} \left[sgn(\nu) \cdot \prod\limits_{i=1}^{p} \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)_{i\nu(i)} \right]}\,.
\end{aligned}$$

Here, in the first equation the component in the i-th row and the j-th column of the adjugate matrix is computed and in the second equation the Leibniz-formula is used to compute the determinants. Furthermore, S_p denotes the symmetric group of permutations of the elements of the set $S = \{1, \dots, p\}$ and $sgn(\nu)$ denotes the signature of a permutation $\nu \in S_p$. Considering the equation above, it becomes obvious that

$$\left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{ij}$$

is a concatenation of Lipschitz continuous functions of sums and products of the functions in the set

$$\left\{ \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)_{ij} \;\middle|\; 1 \le i, j \le p \right\},$$

which are all of bounded variation. Therefore, since Lipschiz continuous functions, sums and products of functions of bounded variation are of bounded variation,

$$\left\| \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{ij} \right\|_V < \infty, \quad 1 \le i, j \le p$$

is satisfied. Following this argument,

$$\left\| \sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{(y,\infty)} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{ij} \right\|_V < \infty, \quad 1 \le j \le p,$$

and consequently, since f_j is of bounded variation for all $j = 1, \dots, p$,

$$\left\| \sum_{j=1}^{p} \left(\sum_{i=1}^{p} f_i(y) \left(\left(\int\limits_{y}^{\infty} f(u) f^T(u)\, dF(u) \right)^{-1} \right)_{ij} \right) f_j(t_{nl}) \right\|_V < \infty$$

uniformly in $l \in \{1, \dots, n\}$ and $n \in \mathbb{N}$.v This completes the proof of (7.122). As already mentioned, we now know that (7.123) is true and therefore,

$$\max_{l=1,\dots,n} d_{nl}^2(x) \xrightarrow[n\to\infty]{} 0$$

holds true, too. Hence, for $\varepsilon > 0$ there is a $n \in \mathbb{N}$ for which $\max\limits_{l=1,\dots,n} d_{nl}^2(x) \le \varepsilon$ for all $n \ge N$. Therefore, for all $n \ge N$

$$\frac{1}{n} \sum_{r=1}^{n} d_{nr}^2(x) \le \frac{1}{n} \sum_{r=1}^{n} \max_{l=1,\dots,n} d_{nl}^2(x) \le \frac{1}{n} \sum_{r=1}^{n} \varepsilon = \varepsilon$$

holds true and thus,

$$\frac{2}{n}\sum_{r=1}^{n} d_{nr}^2(x) \xrightarrow[n\to\infty]{} 0$$

holds true, too, which in view of (7.117), (7.116) and (7.115) completes the proof.

Remark 7.5.5.

The stochastic convergence asserted in (7.111) is true even uniformly on $(-\infty, c]$. To prove this, it suffices to show that the convergence in (7.115) holds true uniformly in $x \in (-\infty, c]$ and due to (7.116) it therefore suffices to show that

$$\sup_{x\in(-\infty,c]} \frac{2}{n}\sum_{r=1}^{n} d_{nr}^2(x) \xrightarrow[n\to\infty]{} 0\,. \tag{7.124}$$

By performing partial integration (see (7.118)) it becomes evident that for the validity of (7.124) it is sufficient to show that the three summands

$$F_r^{(n)}(x)\,(F-F_n)(x),\quad F_r^{(n)}(-\infty)\,(F-F_n)(-\infty),\quad \int\limits_{(-\infty,x]} (F-F_n)(y)\,dF_r^{(n)}(y)$$

converge to zero uniformly in $x \in (-\infty, c]$. Because of (7.120), the second summand is equal to zero. Furthermore, $F_r^{(n)}(x)$ is bounded uniformly in r and x and therefore, we have the uniform convergence of the first summand. Concerning the third summand, it's uniform convergence to zero can be seen from inequality (7.121), the uniform convergence of F_n to F and the bound in (7.122).

Step II:

Using the definitions for $k_n(y)$, $l_n(y)$ and μ_n, we define for $x \in (-\infty, c]$ and $n \in \mathbb{N}$

$$Z_n^{(x)} := \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} (k_n(y) - l_n(y))\; d\mu_n(y).$$

In the following, we are going to show that

$$\|Z_n^{(x)}\|_{\mathcal{L}^2(\mathbb{P})} \xrightarrow[n\to\infty]{} 0 \tag{7.125}$$

is satisfied. This means that $Z_n^{(x)} \xrightarrow[n\to\infty]{\mathcal{L}^2(\mathbb{P})} 0$ and thus, $Z_n^{(x)} \xrightarrow[n\to\infty]{\mathbb{P}} 0$ could be inferred, meaning that (7.112) is satisfied. In order to prove (7.125), we first use the Minkowski-inequality for integrals which implies that

$$\begin{aligned}\|Z_n^{(x)}\|_{\mathcal{L}^2(\mathbb{P})} &= \left\| \int\limits_{(-\infty,x]} \frac{1}{\sigma\sqrt{n}}\,(k_n(y) - l_n(y))\; d\mu_n(y) \right\|_{\mathcal{L}^2(\mathbb{P})} \\ &\le \int\limits_{(-\infty,x]} \frac{1}{\sigma\sqrt{n}} \|k_n(y) - l_n(y)\|_{\mathcal{L}^2(\mathbb{P})}\, d\mu_n(y)\,. \end{aligned} \tag{7.126}$$

Furthermore, it holds true, that for all $n \in \mathbb{N}$ and $y \in (-\infty, c]$,

$$
\begin{aligned}
&\frac{1}{\sigma\sqrt{n}} \left| k_n(y) - l_n(y) \right| \\
&= \frac{1}{\sigma\sqrt{n}} \left| f^T(y) \left[\left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu_n(u) \right)^{-1} \right] \hat{m}_n(y) \right| \\
&\leq \left\| f^T(y) \left[\left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu_n(u) \right)^{-1} \right] \right\|_2 \cdot \frac{1}{\sigma\sqrt{n}} \left\| \hat{m}_n(y) \right\|_2 \\
&\leq \sup_{y \in (-\infty,c]} \left\| f^T(y) \right\|_2 \cdot \sup_{y \in (-\infty,c]} \left\| \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \times \\
&\qquad \times \frac{1}{\sigma\sqrt{n}} \sqrt{\sum_{j=1}^{p} \left(\hat{m}_n(y) \right)_j^2} \\
&\leq \sup_{y \in (-\infty,c]} \left\| f^T(y) \right\|_2 \cdot \sup_{y \in (-\infty,c]} \left\| \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \times \\
&\qquad \times \frac{1}{\sigma\sqrt{n}} \sum_{j=1}^{p} \left| \left(\hat{m}_n(y) \right)_j \right|
\end{aligned}
$$

and therefore,

$$
\begin{aligned}
&\frac{1}{\sigma\sqrt{n}} \left\| k_n(y) - l_n(y) \right\|_{\mathcal{L}^2(\mathbb{P})} \\
&\leq \sup_{y \in (-\infty,c]} \left\| f^T(y) \right\|_2 \cdot \sup_{y \in (-\infty,c]} \left\| \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \times \\
&\qquad \times \frac{1}{\sigma\sqrt{n}} \left\| \sum_{j=1}^{p} \left| \left(\hat{m}_n(y) \right)_j \right| \right\|_{\mathcal{L}^2(\mathbb{P})} \\
&\leq \sup_{y \in (-\infty,c]} \left\| f^T(y) \right\|_2 \cdot \sup_{y \in (-\infty,c]} \left\| \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu(u) \right)^{-1} - \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \times \\
&\qquad \times \sum_{j=1}^{p} \left\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y) \right)_j \right\|_{\mathcal{L}^2(\mathbb{P})}
\end{aligned}
\tag{7.127}
$$

is satisfied. Taking together (7.126) and (7.127), we get

$$\begin{aligned}\|Z_n^{(x)}\|_{\mathcal{L}^2(\mathbb{P})} \leq & \sup_{y\in(-\infty,c]} \|f^T(y)\|_2 \ \times \\ & \times \sup_{y\in(-\infty,c]} \left\| \left(\int_{(y,\infty)} f(u) f^T(u)\, d\mu(u) \right)^{-1} - \left(\int_{(y,\infty)} f(u) f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \times \\ & \times \int_{-\infty}^{x} \sum_{j=1}^{p} \left\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y)\right)_j \right\|_{\mathcal{L}^2(\mathbb{P})} d\mu_n(y)\,.\end{aligned}$$

Therein

$$\sup_{y\in(-\infty,c]} \|f^T(y)\|_2 \ \leq \ \sqrt{\sum_{j=1}^{p} \left(\sup_{y\in(-\infty,c]} f_j(y) \right)^2} \ < \ \infty,$$

since $f_j(y)$ is of bounded variation and therefore bounded for $1 \leq j \leq m$ and

$$\begin{aligned}& \sup_{y\in(-\infty,c]} \left\| \left(\int_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u) \right)^{-1} - \left(\int_{(y,\infty)} f(u)\, f^T(u)\, d\mu_n(u) \right)^{-1} \right\| \\ & \qquad = \sup_{y\in(-\infty,c]} \left\| \left(\frac{1}{n} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \right)^{-1} - \left(\int_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u) \right)^{-1} \right\| \\ & \qquad \xrightarrow[n\to\infty]{} 0\end{aligned}$$

holds true, which up to a change of variables (see Lemma 7.1.3) has been stated in Lemma 7.5.3. Thus, for (7.125) to be true, it suffices to show that for all $n \in \mathbb{N}$ and $y \in (-\infty, c]$ there is a $C \in \mathbb{R}$ such that

$$\int_{-\infty}^{x} \sum_{i=1}^{p} \left\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y)\right)_i \right\|_{\mathcal{L}^2(\mathbb{P})} d\mu_n(y) \ \leq C \tag{7.128}$$

holds true. The remainder of Step II deals with the proof of (7.128). Note that

$$\mathbb{E}\left(\frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y)\right)_j \right) = 0$$

holds true for all $i = 1, \ldots, p$, $n \in \mathbb{N}$ and $y \in (-\infty, c]$, as can be seen in the following way:

Since

$$\begin{aligned}\hat{m}_n(y) &= (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)} \;=\; (X_n^{(j_n(y)+1)})^T (Y_{j_n(y)+1} - X_n^{(j_n(y)+1)} \hat{\beta}_n^{(1)}) \\ &= (X_n^{(j_n(y)+1)})^T \left(X_n^{(j_n(y)+1)} \beta + \varepsilon_n^{(j_n(y)+1)} \right) - \\ &\quad - (X_n^{(j_n(y)+1)})^T \left(X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T (X_n^1 \beta + \varepsilon_n^{(1)}) \right) \\ &= (X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} - (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^1 \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)}\end{aligned} \tag{7.129}$$

and because of $\mathbb{E}(\varepsilon_n^{(i)}) = 0 \in \mathbb{R}^{n-i+1}$, $i \in \{1, \ldots, n\}$, we conclude that

$$\mathbb{E}\left((X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} \right) = 0 \tag{7.130}$$

and

$$\mathbb{E}\left((X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)} \right) = 0 \tag{7.131}$$

and therefore,

$$\mathbb{E}\left(\hat{m}_n(y) \right) = 0 \in \mathbb{R}^p$$

for all $n \in \mathbb{N}$ and $y \in (-\infty, c]$. Due to equations (7.129), (7.130) and (7.131), for $i = 1, \ldots, p$, we obtain

$$\begin{aligned}&\left\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y) \right)_i \right\|_{\mathcal{L}^2(\mathbb{P})} \\ &= \left\| \frac{1}{\sigma\sqrt{n}} \left[(X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} - (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)} \right]_i \right\|_{\mathcal{L}^2(\mathbb{P})} \\ &\leq \left\| \frac{1}{\sigma\sqrt{n}} \left((X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} \right)_i \right\|_{\mathcal{L}^2(\mathbb{P})} + \\ &\quad + \left\| \frac{1}{\sigma\sqrt{n}} \left((X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)} \right)_i \right\|_{\mathcal{L}^2(\mathbb{P})} \\ &= \sqrt{\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}} \left((X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} \right)_i \right)} + \\ &\quad + \sqrt{\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}} \left((X_n^{j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)} \right)_i \right)} \\ &= \sqrt{\left(\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}} (X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)} \right) \right)_{ii}} + \\ &\quad + \sqrt{\left(\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)} \right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)} \right) \right)_{ii}}.\end{aligned}$$

In what fallows, we are going to show that the covariances in both of the above summands are bounded independent of $n \in \mathbb{N}$ and $y \in (-\infty, c]$ so that there is a $\widetilde{C} \in \mathbb{R}$ with

$$\Big\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y)\right)_i \Big\|_{\mathcal{L}^2(\mathbb{P})} \leq \widetilde{C}, \quad i = 1, \ldots, p,$$

independent of $n \in \mathbb{N}$ and $y \in (-\infty, c]$. Hence, in (7.128) we would get

$$\int\limits_{(-\infty,x]} \sum_{i=1}^{p} \Big\| \frac{1}{\sigma\sqrt{n}} \left(\hat{m}_n(y)\right)_i \Big\|_{\mathcal{L}^2(\mathbb{P})} d\mu_n(y) \leq \int\limits_{(-\infty,x]} \sum_{i=1}^{p} \widetilde{C}\, d\mu_n(y) \leq p\widetilde{C} =: C,$$

which would complete the proof of (7.112). To prove this boundedness of the covariances, we first note that

$$\begin{aligned} \mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}}(X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)}\right) &= \frac{1}{n}(X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \\ &\xrightarrow[n\to\infty]{} \int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u) \end{aligned}$$

holds true. The convergence, up to a change of variables (see Lemma 7.1.3), has been subject of Lemma 7.5.3. Thus, the covariance matrix of $\frac{1}{\sigma\sqrt{n}}(X_n^{(j_n(y)+1)})^T \varepsilon_n^{(j_n(y)+1)}$ is bounded independent of $n \in \mathbb{N}$ and $y \in (-\infty, c]$.

In the further on, we need the following lemma about the Loewner order of matrices. Here "$\prec$" denotes smaller with respect to the Loewner order and for $A, B \in R^{p\times p}$, $A \prec B$ holds by definition if and only if $B - A$ is positive definite. The lemma is a folklore result for which, for the reader's convenience, we provide a proof.

Lemma 7.5.6.

Let $A, B \in R^{p\times p}$, $p \in \mathbb{N}$ be positive definite matrices with $A \prec B$, $C \in R^{p\times p}$ invertible and I_p the $p \times p$ unit matrix. Then the following assertions hold true:

i) $C^T A C \prec C^T B C$.

ii) For $I_p \prec B$ it follows that B is regular and $B^{-1} \prec I_p$.

iii) $B^{-1} \prec A^{-1}$.

Proof:

i) For $v \in \mathbb{R}^p \setminus \{0\}$,

$$v^T C^T (B - A) C v = (Cv)^T (B - A) C v > 0$$

holds true since $A \prec B$ and $Cv \in \mathbb{R}^p \setminus \{0\}$ thus

$$C^T AC \prec C^T BC\,.$$

ii) Let I_p be the $p \times p$ unit matrix and $I_p \prec B$. Since the matrix B is symmetrical with real values it admits a decomposition

$$B = EDE^T,$$

where E is an orthogonal matrix and D is a diagonal matrix. As a consequence of this and part i),

$$I_p = E^T I_p E \prec E^T BE = E^T EDE^T E = D$$

holds true and thus all eigenvalues of B are bigger that 1. As a consequence, B^{-1} exists and

$$B^{-1} = B^{-\frac{1}{2}} I_p B^{-\frac{1}{2}} \prec B^{-\frac{1}{2}} BB^{-\frac{1}{2}} = I_p$$

where $B^{-\frac{1}{2}} = ED^{-\frac{1}{2}}E^T$.

iii) Since $A \prec B$ and A, B positive definite, we know that A and B are regular and $B - A \succ 0$. Thus, because of part i)

$$A^{-\frac{1}{2}}(B - A)A^{-\frac{1}{2}} \succ 0$$

and therefore

$$A^{-\frac{1}{2}} BA^{-\frac{1}{2}} \succ I_p$$

are satisfied. Now, because of part ii), we get

$$A^{\frac{1}{2}} B^{-1} A^{\frac{1}{2}} \prec I_p$$

and due to part i)

$$B^{-1} = A^{-\frac{1}{2}}(A^{\frac{1}{2}} B^{-1} A^{\frac{1}{2}})A^{-\frac{1}{2}} \prec A^{-\frac{1}{2}} I_p A^{-\frac{1}{2}} = A^{-1}.$$

is satiated. □

Concerning the variance of the second random vector

$$\frac{1}{\sigma\sqrt{n}}(X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)},$$

we know that for

$$\widetilde{X}_n := \begin{pmatrix} f^T(t_{n1}) \\ \vdots \\ f^T(t_{nj_n(y)}) \end{pmatrix} \in \mathbb{R}^{j_n(y) \times p},$$

we have

$$\widetilde{X}_n^T \widetilde{X}_n \succ 0$$

for $n \in \mathbb{N}$ large enough and $j_n(y) \geq p$ due to the linear independence of $f_1, \ldots, f_p$. Thus,

$$(X_n^{(1)})^T X_n^{(1)} = \left(\widetilde{X}_n^T \middle| (X_n^{(j_n(y)+1)})^T\right) \begin{pmatrix} \widetilde{X}_n \\ \hline X_n^{(j_n(y)+1)} \end{pmatrix} = \widetilde{X}_n^T \widetilde{X}_n + (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

$$\succ (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \succ 0$$

holds true, since

$$\widetilde{X}_n^T \widetilde{X}_n + (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} - (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} = \widetilde{X}_n^T \widetilde{X}_n$$

is positive definite. Therefore, according to Lemma 7.5.6 part iii), we have

$$(X_1^T X_1)^{-1} \prec \left((X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}\right)^{-1}$$

and again due to Lemma 7.5.6 part i),

$$(X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

$$\prec (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}\right)^{-1} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

$$= (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

holds true. Hence,

$$\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)}\right)$$

$$= \frac{1}{n} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

$$\prec \frac{1}{n} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}$$

$$\xrightarrow[n \to \infty]{} \int\limits_{(y,\infty)} f(u)\, f^T(u)\, d\mu(u)\,.$$

Note that the convergence in the last step of the above inequality chain has been proved in Lemma 7.5.3, up to a change of variables, see Lemma 7.1.3. We therefore can conclude that the covariance matrix of

$$\frac{1}{\sigma\sqrt{n}} (X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)} \left((X_n^{(1)})^T X_n^{(1)}\right)^{-1} (X_n^{(1)})^T \varepsilon_n^{(1)}$$

is bounded independent of $n \in \mathbb{N}$ and $y \in (-\infty, c]$.

Remark 7.5.7.

Since for $x \in (-\infty, c]$

$$\left|Z_n^{(x)}\right| = \left| \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} k_n(y) - l_n(y)\, d\mu_n(y) \right| \leq \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,x]} |k_n(y) - l_n(y)|\, d\mu_n(y)$$

$$\leq \frac{1}{\sigma\sqrt{n}} \int\limits_{(-\infty,c]} |k_n(y) - l_n(y)|\, d\mu_n(y) =: \widetilde{Z}_n$$

holds true and according to our findings in Step II, we can follow:

$$\begin{aligned} \|\widetilde{Z}_n\|_{\mathcal{L}^2(\mathbb{P})} &= \left\| \int\limits_{(-\infty,c]} \frac{1}{\sigma\sqrt{n}} |k_n(y) - l_n(y)|\, d\mu_n(y) \right\|_{\mathcal{L}^2(\mathbb{P})} \\ &\leq \int\limits_{(-\infty,c]} \frac{1}{\sigma\sqrt{n}} \|k_n(y) - l_n(y)\|_{\mathcal{L}^2(\mathbb{P})}\, d\mu_n(y) \\ &\xrightarrow[n\to\infty]{} 0. \end{aligned}$$

Thus,

$$\sup_{x\in]-\infty,x_{\max}]} \left|Z_n^{(x)}\right| \leq \widetilde{Z}_n \xrightarrow[n\to\infty]{\mathbb{P}} 0$$

and therefore, we have (7.112) uniformly in $x \in (-\infty, c]$.

Regarding equation (7.108)**:** For $1 \leq l, m \leq p$, the l, m-th component of the matrix

$$\sum_{i=j_n(y)+1}^{n} f(t_{ni}) f^T(t_{ni}) \in \mathbb{R}^{p\times p}$$

can be represented as

$$\left(\sum_{i=j_n(y)+1}^{n} f(t_{ni})\, f^T(t_{ni}) \right)_{l,m} = \sum_{i=j_n(y)+1}^{n} f_l(t_{ni}) f_m(t_{ni}) = \left((X_n^{(j_n(y)+1)})^T X_n^{(i_n(y)+1)} \right)_{l,m}.$$

It is therefore true that

$$\int\limits_{(-\infty,x]} f^T(y) \left(\int\limits_{(y,\infty)} f(u) f^T(u)\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(u) \right)^{-1} (X_n^{(i_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(y)$$

$$= \int\limits_{(-\infty,x]} f^T(y) \left(\frac{1}{n} \sum_{i=j_n(y)+1}^{n} f(t_{ni}) f^T(t_{ni}) \right)^{-1} (X_n^{(j_n(y)+1)})^T \hat{\varepsilon}_n^{(j_n(y)+1)}\, d\left(\frac{1}{n} \sum_{i=1}^{n} \delta_{\{t_{ni}\}} \right)(y)$$

$$= \int\limits_{(-\infty,x]} f^T(y)\left(\frac{1}{n}(X_n^{(j_n(y)+1)})^T X_n^{(j_n(y)+1)}\right)^{-1}(X_n^{(j_n(y)+1)})^T\hat{\varepsilon}_n^{(j_n(y)+1)}\, d\left(\frac{1}{n}\sum_{i=1}^{n}\delta_{\{t_{ni}\}}\right)(y)$$

$$= \sum_{i=1}^{j_n(x)} f^T(t_{ni})\left((X_n^{(j_n(t_{ni})+1)})^T X_n^{(j_n(t_{ni})+1)}\right)^{-1}(X_n^{(j_n(t_{ni})+1)})^T\hat{\varepsilon}_n^{(j_n(t_{ni})+1)}$$

$$= \sum_{i=1}^{j_n(x)} f^T(t_{ni})\left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1}(X_n^{(i+1)})^T\hat{\varepsilon}_n^{(i+1)}\,.$$

Regarding equation (7.109)**:** According to Lemma 7.5.2, part ii),

$$c_{ni}\, r_{ni} = \hat{\varepsilon}_{ni} - f^T(t_{ni})\left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1}(X_n^{(i+1)})^T\hat{\varepsilon}_n^{(i+1)}$$

holds true. Thus,

$$\begin{aligned}
&\left|\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)}\left(\hat{\varepsilon}_{ni} - f^T(t_{ni})\left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1}(X_n^{i+1})^T\hat{\varepsilon}_n^{(i+1)}\right) - \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)} r_{ni}\right| \\
&\qquad = \left|\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)}\left[\left(\hat{\varepsilon}_{ni} - f^T(t_{ni})\left((X_n^{(i+1)})^T X_n^{(i+1)}\right)^{-1}(X_n^{(i+1)})^T\hat{\varepsilon}_n^{(i+1)}\right) - r_{ni}\right]\right| \\
&\qquad = \left|\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)}\left[c_{ni}\, r_{ni} - r_{ni}\right]\right| \\
&\qquad = \left|\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)} r_{ni}\cdot(c_{ni}-1)\right| \\
&\qquad \xrightarrow[n\to\infty]{\mathbb{P}} 0
\end{aligned} \tag{7.132}$$

is satisfied. Here, the stochastic convergence can be justified as follows. According to Lemma 7.5.2,

$$\mathbb{E}\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)} r_{ni}\cdot(c_{ni}-1)\right) = \frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)}(c_{ni}-1)\mathbb{E}\,(r_{ni}) = 0,$$

$Cov(r_{ni}, r_{nj}) = 0$, $i \neq j$, $\mathbb{V}ar(r_{ni}) = \sigma^2$ and $\sup\limits_{i\in\{1,\ldots,i_n(c)\}} c_{ni} \xrightarrow[n\to\infty]{} 1$ are satisfied. As a consequence,

$$\mathbb{V}ar\left(\frac{1}{\sigma\sqrt{n}}\sum_{i=1}^{j_n(x)} r_{ni}\cdot(c_{ni}-1)\right) = \frac{1}{n}\sum_{i=1}^{j_n(x)}(c_{ni}-1)^2 \leq \frac{1}{n}\cdot n\cdot\sup_{i\in\{1,\ldots,i_n(c)\}}(c_{ni}-1)^2$$

$$\xrightarrow[n\to\infty]{} 0 \tag{7.133}$$

holds true. Therefore, (7.132) can be deduced, according to Chebyshev's inequality. This proofs equation (7.109).

So far we have proved the assertion $(T^*(R_n^1))(x) = \widetilde{R}_n^1(x)+o_P(1)$ pointwise for $x \in (-\infty, c]$. However, the assertion holds true uniformly in $x \in (-\infty, c]$, as Theorem 6.3.3 asserts. To see this, note first that it is obvious that the equations (7.106),(7.108) and (7.109) hold uniformly in x, note secondly that equation (7.107) is true, uniformly in $x \in (-\infty, c]$, according to Remarks 7.5.5 and 7.5.7 and that finally equation (7.109) holds true uniformly in x because, due to (7.133), the variance of $\frac{1}{\sigma\sqrt{n}} \sum_{i=1}^{j_n(x)} r_{ni} \cdot (c_{ni} - 1)$ converges to zero uniformly in $(-\infty, c]$. □

Index

Bibliography

Adler, R.J. (1990). *An Introduction to Continuity, Extrema, and Related Topics for General Gaussian Processes*. Ims Lecture Series. Institute of Mathematical Statistics.

Aki, Sigeo (1986). "Some test statistics based on the martingale term of the empirical distribution function". In: *Annals of the Institute of Statistical Mathematics* 38, pp. 1–21.

Anderson, T. W. (1963). "Asymptotic Theory for Principal Component Analysis". In: *The Annals of Mathematical Statistics* 34.1, pp. 122–148.

Aronszajn, Nachman (1950). "Theory of reproducing Kernels". In: *Transactions of the American Mathematical Society* 68, pp. 337–404.

Bai, Jushan (Aug. 2003). "Testing Parametric Conditional Distributions of Dynamic Models". In: *The Review of Economics and Statistics* 85.3, pp. 531–549.

Bauer, Heinz (1992). *Bauer: mass-und Integrationstheorie 2a Geb Lg*. Berlin, New York: Walter de Gruyter.

Beiglboeck, Mathias, Walter Schachermayer, and Bezirgen Veliyev (2012). "A short proof of the Doob–Meyer theorem". In: *Stochastic Processes and their applications* 122.4, pp. 1204–1209.

Billingsley, Patrick (1968). *Convergence of Probability Measures*. New York, Chichester, Brisbane, Toronto: John Wiley and Sons.

Bischoff, Wolfgang (1998). "A functional central limit theorem for regression models". In: *The Annals of Statistics* 26.4, pp. 1398–1410.

– (2002). "The structure of residual partial sums limit processes of linear regression models". In: *Theory of Stochastic Processes* 8.

Bischoff, Wolfgang and Frank Miller (2000). "Asymptotically optimal tests and optimal designs for testing the mean in regression models with applications to change-point problems". In: *Annals of the Institute of Statistical Mathematics* 52.4, pp. 658–679.

Bischoff, Wolfgang et al. (2005). "Analysis of a change-point regression problem in quality control by partial sums processes and Kolmogorov type tests". In: *Metrika* 62.1, pp. 85–98.

Brown, Robert L, James Durbin, and James M Evans (1975). "Techniques for testing the constancy of regression relationships over time". In: *Journal of the Royal Statistical Society: Series B (Methodological)* 37.2, pp. 149–163.

Dehardt, J. (1971). "Generalizations of the Glivenko-Cantelli Theorem". In: *The Annals of Mathematical Statistics* 42.6, pp. 2050–2055.

Deiser, Oliver (2015). *Analysis 2*. 2nd ed. Springer Spektrum.

Durbin, James (1973a). *Distribution theory for tests based on the sample distribution function*. SIAM.

– (1973b). "Weak convergence of the sample distribution function when parameters are estimated". In: *The Annals of Statistics*, pp. 279–290.

Elstrodt, Jürgen (2009). *Maß- und Integrationstheorie*. Vol. 6. Berlin, Heidelberg: Springer-Verlag.

Eubank, Randall L and Jeffrey D Hart (1992). "Testing goodness-of-fit in regression via order selection criteria". In: *The annals of Statistics*, pp. 1412–1425.

Eubank, Randall L and Clifford H Spiegelman (1990). "Testing the goodness of fit of a linear model via nonparametric regression techniques". In: *Journal of the American Statistical Association* 85.410, pp. 387–392.

Evers, Karsten (2022). *A Functional Central Limit Theorem for Recursive Residuals and Applications in Asymptotoc Statistics*. KU Eichstätt-Ingolstadt.

Farebrother, RW (1978). "An historical note on recursive residuals". In: *Journal of the Royal Statistical Society. Series B (Methodological)* 40.3, pp. 373–375.

Federer, Herbert (1969). *Geometric Measure Theory*. Berlin, Heidelberg, New York: Springer-Verlag.

Firth, D, J Glosup, and DV Hinkley (1991). "Model checking with nonparametric curves". In: *Biometrika* 78.2, pp. 245–252.

Gänsler, Peter and Winfried Stute (1977). *Wahrscheinlichkeitstheorie*. Berlin, Heidelberg, New York: Springer-Verlag.

Hawkins, Douglas M. (1987). *Diagnosis of Some Model Deficiencies Using Recursive Residuals*. University of Minnesota: Retrieved from the University of Minnesota Digital Conservancy.

He, Kaiming et al. (2016). "Deep Residual Learning for Image Recognition". In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 770–778.

Hocking, Ronald R. (2013). *Methods and applications of linear models: regression and the analysis of variance*. John Wiley & Sons.

Khmaladze, E. V. (1980). "The use of ω^2 tests for testing parametric hypotheses". In: *Theory of Probability & Its Applications* 24.2, pp. 283–301.

– (1982). "Martingale approach in the theory of goodness-of-fit tests". In: *Theory of Probability & Its Applications* 26.2, pp. 240–257.

Kim, Taehooie et al. (2020). "A stepwise interpretable machine learning framework using linear regression (LR) and long short-term memory (LSTM): City-wide demand-side prediction of yellow taxi and for-hire vehicle (FHV) service". In: *Transportation Research Part C: Emerging Technologies* 120, p. 102786.

Klenke, Achim (2013). *Wahrscheinlichkeitstheorie*. 3rd ed. Springer-Lehrbuch Masterclass. Springer Spektrum.

Koenker, Roger and Zhijie Xiao (2002). "Inference on the quantile regression process". In: *Econometrica* 70.4, pp. 1583–1612.

Krämer, Walter, Werner Ploberger, and Raimund Alt (1988). "Testing for structural change in dynamic models". In: *Econometrica: Journal of the Econometric Society*, pp. 1355–1369.

Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton (2012). "ImageNet classification with deep convolutional neural networks". In: *Communications of the ACM* 60, pp. 84–90.

Leshno, Moshe et al. (1993). "Multilayer feedforward networks with a nonpolynomial activation function can approximate any function". In: *Neural Networks* 6.6, pp. 861–867.

Ly, Alexander et al. (2017). "A tutorial on Fisher information". In: *Journal of Mathematical Psychology* 80, pp. 40–55.

MacNeill, Ian B. (1978a). "Limit processes for sequences of partial sums of regression residuals". In: *The Annals of Probability*, pp. 695–698.

– (1978b). "Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times". In: *The Annals of Statistics*, pp. 422–433.

MacNeill, Ian B., Y Mao, and L Xie (1994). "Modeling heteroscedastic age-period-cohort cancer data". In: *Canadian Journal of Statistics* 22.4, pp. 529–539.

Maulud, Dastan and Adnan M Abdulazeez (2020). "A review on linear regression comprehensive in machine learning". In: *Journal of Applied Science and Technology Trends* 1.4, pp. 140–147.

Mnih, Volodymyr et al. (2013). "Playing Atari with Deep Reinforcement Learning". In: *CoRR* abs/1312.5602. arXiv: 1312.5602. URL: http://arxiv.org/abs/1312.5602.

Müller, Hans-Georg (1992). "Goodness-of-fit diagnostics for regression models". In: *Scandinavian journal of statistics*, pp. 157–172.

Otto, Sven and Jörg Breitung (2020). "Backward CUSUM for testing and monitoring structural change". In: *arXiv preprint arXiv:2003.02682*.

Parzen, Emanuel (1959). *Statistical Inference on Time Series by Hilbert Space methods, I.* San Francisco, Cambridge, London, Amsterdam: Holden-Day.

Pesaran, M. Hashem and Allan Timmermann (2002). "Market timing and return prediction under model instability". In: *Journal of Empirical Finance* 9.5, pp. 495–510.

Prokhorov, Yu. V. (1956). "Convergence of random processes and limit theorems in probability theory". In: *Theory of Probability & Its Applications* 1.2, pp. 157–214.

R Core Team (2020). *R: A Language and Environment for Statistical Computing*. Vienna, Austria. URL: https://www.R-project.org/.

Schuld, Maria, Ilya Sinayskiy, and Francesco Petruccione (2016). "Prediction by linear regression on a quantum computer". In: *Physical Review A* 94.2, p. 022342.

Searle, Shayle R. (1971). *Linear Models*. New York: John Wiley and Sons.

Sen, Pranab Kumar (1982). "Invariance principles for recursive residuals". In: *The Annals of Statistics* 10.1, pp. 307–312.

Somayasa, Wayan (2011). *Model-Check Based on Residual Partial Sums Process of Heteroscedastic spatial Linar Regression Models*. Jurusan Matematika FMIPA Universitas Haluoloeo.

Strook, Daniel W. (1994). *A Concise Introduction to the Theory of Integration*. Boston, Basel, Berlin: Birkhäuser.

Stute, Winfried (1997). "Nonparametric model checks for regression". In: *The Annals of Statistics*, pp. 613–641.

Stute, Winfried, Silke Thies, and Li-Xing Zhu (1998). "Model checks for regression: an innovation process approach". In: *The Annals of Statistics* 26.5, pp. 1916–1934.

Topsoe, Flemming (1967). "Preservation of Weak Convergence Under Mapping". In: *The Annals of Mathematical Statistics* 38.6, pp. 1661–1665.

Vaart, A. W. V. d. (1998). *Asymptotic Statistics*. Cambridge University Press.

Walter, Wolfgang (2002). *Analysis 2, 5. Auflage*. 5. Springer-Lehrbuch. Springer.