# Sequential Change–Point Analysis based on Invariance Principles

Inaugural–Dissertation
zur
Erlangung des Doktorgrades
der Mathematisch–Naturwissenschaftlichen Fakultät
der Universität zu Köln
vorgelegt von

**Alexander Aue**
aus Marburg

Köln 2003

| | |
|---|---|
| Erster Berichterstatter: | Prof. Dr. Josef Steinebach |
| Zweiter Berichterstatter: | Prof. Dr. Lajos Horváth |

Tag der mündlichen Prüfung: 13.02.2004

## Zusammenfassung

Die Changepoint–Analyse befasst sich mit dem Aufdecken von Strukturbrüchen stochastischer Prozesse auf der Basis einer (längeren) Serie von Beobachtungen. In dieser Dissertation werden (nichtparametrische) sequentielle Testprozeduren hergeleitet, die neue Anregungen aus der Wirtschaftsstatistik aufgreifen. Das wesentliche Beweismittel bilden Invarianzprinzipien, die es erlauben, die statistische Analyse auf das Untersuchen von Eigenschaften des Grenzprozesses zu reduzieren. Basierend auf bestehenden Resultaten für lineare Modelle wird ein Lokationsmodell eingeführt, um auf einen möglichen Wechsel im Erwartungswert von zu Grunde liegenden Zufallsvariablen zu testen. Dabei wird das asymptotische Verhalten der Teststatistik untersucht und die Grenzverteilung der zugehörigen Stoppzeit bestimmt. In einem zweiten Teil werden sogenannte RCA(1) Zeitreihen betrachtet. Es zeigt sich, dass diese Prozesse ein starkes Invarianzprinzip mit einer gewissen Rate erfüllen und deshalb die vorangehenden Ergebnisse weiterhin Gültigkeit besitzen. Zudem werden a–posteriori Tests konzipiert, um die Stabilität eines Modellparameters zu untersuchen. Abschließend wird das Verhalten von Suprema stochastischer Prozesse mit linearem Drift diskutiert. Die erzielten Resultate können dazu verwendet werden, sequentielle Tests in mehrdimensionalen Modellen zu konstruieren.

## Abstract

Change–point analysis is concerned with detecting structural breaks of stochastic processes based on a (longer) series of observations. In this dissertation, we derive (nonparametric) sequential test procedures that take into account new motivation coming from econometrics. The main basis for the proofs are invariance principles which allow to reduce the statistical analysis to investigating the properties of the limit process. Taking into account results for linear models, a location model is introduced to test for possible changes in the mean of underlying random variables. Therein, we examine the asymptotic behaviour of the test procedure under both hypotheses and obtain the limit distribution of the corresponding stopping time. In a second part, so–called RCA(1) time series are studied. It turns out that these processes satisfy a strong invariance principle with a certain rate. This allows for retaining the previous results. Moreover, a–posteriori tests are provided to examine the stability of a model parameter. Finally, we discuss the behaviour of suprema of stochastic processes with linear drift. The results obtained can be utilized to construct sequential tests in multivariate settings.

# Preface

This dissertation is concerned with weak and strong invariance principles and their applications to the asymptotic theory (in probability and) in statistics, and therein especially to questions arising in the context of change–point analysis. The models of interest were introduced in the area of quality control to study the behaviour of (outputs of) certain production lines, but have been extended to numerous other fields since then. Thus, several time series models are now included in our approach such as (1) (G)ARCH processes, which have attracted a lot of attention recently because Robert Engle won the Nobel Prize in Economic Sciences in 2003, and (2) RCA time series, which have been successfully applied to investigate random perturbations of dynamical systems.

The main topic addressed here is the sequential testing of changes in the mean of underlying random variables (for example the time series of the previous paragraph). Sequential tests were introduced by Abraham Wald in the 1940s in order to construct more efficient inspection procedures. Motivated by developments in econometrics some modifications have been proposed and studied lately. Throughout the present work, the statistical analysis will be carried out via invariance principles, i.e., we are solely interested in asymptotic results.

The dissertation is organized as follows.

In Chapter 1, we shall introduce the notions of strong and weak invariance principles and briefly review the seminal results. Invariance proves useful for an elegant derivation of precise asymptotics in the following chapters.

In Chapter 2, a sequential test procedure is studied which detects possible changes in the mean of observations satisfying a weak invariance principle. Two introductory sections deal with the basic definitions and properties of change–point analysis and sequential statistics. In Section 2.3 and Section 2.4, we present a location model and examples covered by the given framework, while we continue with thoroughly examining the asymptotic properties of the test procedure in the following. Thus, we analyze the limiting behaviour under both hypotheses in Section 2.5. The main part of the chapter, Section 2.6, is devoted to deriving the limit distribution of the corresponding stopping rule. Provided the change is relatively small, it turns out, that this limit distribution is normal under a suitable standardization.

In Chapter 3, the results of the second chapter shall be extended to a certain non–linear time series arising from the well–known autoregressive time series by allowing for randomly disturbed coefficients. An important tool herein will be a strong invariance principle formulated for the partial sums of random variables whose dependence structure is described in terms of their conditional expectation. By recursive techniques, we shall retain the results of Chapter 2 for so–called random coefficient autoregressive time series of order one as well. In addition, an a–posteriori setting for a change in the mean and a testing procedure for a change in the deterministic part of the coefficients is provided in Section 3.4.

The final chapter will be concerned with the suprema (obtained in Euclidean norm) of vector–valued stochastic processes with drift defined on a certain closed (time) interval on the real line. It turns out that asymptotically these suprema are not far away from the rightmost point. Unlike in the univariate case, there are now two ways of defining a non–zero drift: either all components have non–zero drift or at least one. The approximation rate obtained is the same in both cases. However, different methods of proof are required. Moreover, the law of the iterated logarithm and some weak convergence results shall be discussed. In contrast to Chapters 2 and 3, which are based on certain concrete models, the results presented in Chapter 4 are of rather theoretical use. Nevertheless, they can be utilized to construct suitable (multivariate) test statistics for sequential procedures.

**Acknowledgements.** I am greatly indebted to my supervisor Prof. Dr. Josef Steinebach. His incredible thoroughness and astonishing insight in many fields of probability and statistics have always been a great source of inspiration. Moreover, he introduced me to Prof. Dr. Lajos Horváth from the University of Utah, who kindly agreed to be a member of my thesis committee. During two visits in Salt Lake City I highly benefited from countless discussions and stimulations, and from his extraordinary hospitality.

Finally, I would like to thank my family for all kinds of support. This is for you!

# Contents

# Chapter 1

# Invariance Principles

## 1.1 Introduction

There are numerous applications of partial sums of random variables in probability and statistics. Suppose, we have a sequence of independent, identically distributed, real–valued random variables $\{X_n\}_{n\in\mathbb{N}}$ on some probability space $(\Omega, \mathcal{A}, P)$. Define

$$S_0 = 0, \qquad S_n = X_1 + \ldots + X_n \quad (n \in \mathbb{N}).$$

Then, $X_n$ can for instance be seen as gain or loss of a gambler at time $n$. Obviously, $S_n$ describes the overall gain or loss. Intuitively, we would call the game fair if the expectation of the $X_n$ exists and is zero. Else, there would be a preference for either the gambler or his counterpart. More detailed considerations lead to the theory of martingales.

Another interpretation comes from Physics. Let the distribution of $X_n$ for all $n \in \mathbb{N}$ be given by $P(X_n = \pm 1) = \frac{1}{2}$. Then, $X_n$ steers the movement of a particle which at time $n$ chooses between the possibilities of going one step to the right or one step to the left with the same probability. $S_n$ can therefore be regarded as the position of the particle at time $n$ which by assumption $S_0 = 0$ starts in the origin. This situation corresponds to the so–called symmetric simple random walk on the integers $\mathbb{Z}$ and its generalizations are of major interest in the theory of discrete Markov chains.

A third example describes a familiar statistical problem. Suppose for the moment that $X_1, \ldots, X_n$ have finite expectation and unit variance. On account of several observations a statistician becomes suspicious and conjectures a violation of the iid assumption, for instance a break in the mean of the random variables at a (fixed and known) time–point $1 < k^* < n$ which is usually called a change–point. A suitable test statistic can be given in terms of the partial sums $S_n$ by the difference

$$T_n = \frac{1}{\sqrt{n}} \left( S_{k^*} - \frac{k^*}{n} S_n \right).$$

Great values of $T_n$ would indicate a change in the expectation at $k^*$. The given situation will be discussed in more generality in the second chapter below.

All three classical examples – and there are many more – show that it is natural to consider the behaviour of the sequence of partial sums $\{S_n\}_{n \in \mathbb{N}_0}$ in the long run. If $E|X_1| < \infty$, then the weak, respectively, strong law of large numbers tell us that

$$\frac{1}{n} S_n \longrightarrow EX_1 \qquad (n \to \infty)$$

in probability, respectively, almost surely. If we normalize the partial sums $S_n$ by $\frac{1}{n}$, the limit random variable is deterministic. Moreover, the partial sums $S_n$ will on average take on values of order $nEX_1$. On the other hand, if in addition $0 < \mathrm{Var} X_1 = \sigma^2 < \infty$, the central limit theorem gives

$$\lim_{n \to \infty} P\left( S_n - nEX_1 \leq \sqrt{n} \sigma x \right) = \Phi(x) \qquad (x \in \mathbb{R}),$$

where $\Phi$ denotes the distribution function of a standard normal random variable. Roughly speaking, the sum of many independent random variables will be approximately normally distributed if it is suitably normalized and the typical deviation from the average value $nEX_1$ is of order $\sqrt{n}$.

These limit theorems – completed by the law of the iterated logarithm which a.s. indicates the maximum order of the fluctuation of partial sums – have played a fundamental role in the development of probability and statistics as well in theory as in applications. A further instrument to examine the asymptotic behaviour of random variables was introduced by the notion of (weak) invariance in the papers of Erdős and Kac (1946), and Doob (1949). Starting with a theorem of Strassen (1964) we will shortly describe the development of the theory of strong invariance principles in the next section. One main advantage of the latter is the additional information on the rate of convergence provided in the theorems.

## 1.2 Strong approximations

The term strong invariance principle was first used by Strassen (1964) who used a technique due to Skorohod (the so–called Skorohod embedding scheme) to prove a functional law of the iterated logarithm. His results were strengthened by the work of the Hungarian construction school grouped around Csörgő and Révész or Komlós, Major and Tusnády. We are going to sketch their results now.

**The one–dimensional case.** Let again $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on a probability space $(\Omega, \mathcal{A}, P)$ and assign to it the sequence $\{S_n\}_{n\in\mathbb{N}_0}$ of their partial sums as in the previous section. Moreover, let $\{Y_n\}_{n\in\mathbb{N}}$ be a sequence of independent standard normal random variables on a suitable (but possibly different) probability space $(\Omega_0, \mathcal{A}_0, P_0)$ and let

$$T_0 = 0, \qquad T_n = Y_1 + \ldots + Y_n \quad (n \in \mathbb{N})$$

denote the corresponding partial sums. In the current section, we shall present best approximations of the partial sums of $X_n$ (redefined and normalized on the probability space $(\Omega_0, \mathcal{A}_0, P_0)$) by those of $Y_n$ in an a.s. sense. These are called strong approximations or strong invariance principles.

It is possible to consider a rich enough probability space on which both $\{X_n\}_{n\in\mathbb{N}}$ and $\{Y_n\}_{n\in\mathbb{N}}$ are defined. For the sake of convenience and without losing accuracy we hence set $(\Omega, \mathcal{A}, P) = (\Omega_0, \mathcal{A}_0, P_0)$ for the whole sequel.

If $X_1$ is standardized, Strassen could prove the following strong invariance.

**Theorem 1.2.1 (Strassen)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on $(\Omega, \mathcal{A}, P)$ with $EX_1 = 0$ and $\mathrm{Var}X_1 = 1$. Then, there exists a sequence of independent standard normal random variables $\{Y_n\}_{n\in\mathbb{N}}$ such that*

$$S_n - T_n = o\left(\sqrt{n \log\log n}\right) \qquad \text{a.s.}$$

*as $n \to \infty$.*

**Proof:** See Strassen (1964). □

Strassen's result led to the (optimal) extensions of the Hungarian construction. Their quantile technique was first applied by Csörgő and Révész (1975a,1975b). A substantial refinement of these results was given by Komlós, Major and Tusnády (1975,1976), and Major (1976a,1976b) in a series of now famous articles.

At first, we consider the case of existing moments of order greater than 2.

**Theorem 1.2.2 (Komlós, Major and Tusnády)**
*Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of independent, identically distributed random variables on* $(\Omega, \mathcal{A}, P)$ *with* $EX_1 = 0$ *and* $\mathrm{Var}X_1 = 1$. *If*

$$E|X_1|^\nu < \infty$$

*for some* $\nu > 2$, *there exists a sequence of independent standard normal random variables* $\{Y_n\}_{n\in\mathbb{N}}$ *such that*

$$S_n - T_n = o\left(n^{\frac{1}{\nu}}\right) \qquad a.s.$$

*as* $n \to \infty$.

**Proof:** See Komlós, Major and Tusnády (1975,1976), and Major (1976a). □

The result of Theorem 1.2.2 proved to be best possible in the following sense: if there are independent, identically distributed random variables $\{X_n\}_{n\in\mathbb{N}}$ satisfying $S_n - T_n = o(n^{\frac{1}{\nu}})$ a.s. as $n \to \infty$, then necessarily $E|X_1|^\nu < \infty$.

If we assume instead of a certain moment of order $\nu > 2$ the existence of the moment generating function of $X_1$ in a neighborhood of zero, the following theorem holds true. It is obtained from an exact probability inequality by an application of the Borel–Cantelli lemma.

**Theorem 1.2.3 (Komlós, Major, Tusnády)**
*Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of independent, identically distributed random variables on* $(\Omega, \mathcal{A}, P)$ *with* $EX_1 = 0$ *and* $\mathrm{Var}X_1 = 1$. *If*

$$E\exp(tX_1) < \infty$$

*for* $|t| \le t_0$, $t_0 > 0$, *there exists a sequence of independent standard normal random variables* $\{Y_n\}_{n\in\mathbb{N}}$ *such that*

$$S_n - T_n = \mathcal{O}(\log n) \qquad a.s.$$

*as* $n \to \infty$.

**Proof:** See Komlós, Major and Tusnády (1975,1976). □

The result of Theorem 1.2.3 is optimal under the given assumptions. Unless the $X_n$ ($n \in \mathbb{N}$) are not themselves standard normally distributed it is impossible to obtain a construction of $\{T_n\}_{n\in\mathbb{N}_0}$ such that the rate could be improved to $o(\log n)$ a.s. as $n \to \infty$. Moreover, if the statement of Theorem 1.2.3 is fulfilled for a sequence $\{X_n\}_{n\in\mathbb{N}}$, then $X_1$ necessarily has an existing moment generating function in a neighborhood of zero.

Along the lines of Csörgő and Horváth (1993), in most applications it is more convenient to approximate $\{S_n\}_{n\in\mathbb{N}_0}$ by a standard Wiener process $\{W(t) : t \geq 0\}$. Define therefore the partial sum process (in continuous time) $\{S(t) : t \geq 0\}$ by

$$S(0) = 0, \qquad S(t) = S_{[t]} \quad (t \in \mathbb{R}_+),$$

where $[\cdot]$ denotes the integer part. Under the given generalizations we can reformulate Theorems 1.2.2 – 1.2.3 as follows, since the increments $\sup_{k\leq t\leq k+1} |W(t) - W(k)|$ can be estimated appropriately (cf. for instance Csörgő and Réveśz (1981), Theorem 1.2.1 and Lemma 1.2.1).

**Theorem 1.2.4**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on $(\Omega, \mathcal{A}, P)$ with $EX_1 = 0$ and $\mathrm{Var}X_1 = 1$. If*

$$E|X_1|^\nu < \infty$$

*for some $\nu > 2$, there exists a standard Wiener process $\{W(t) : t \geq 0\}$ such that*

$$S(T) - W(T) = o\left(T^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as $T \to \infty$.*

**Proof:** See Komlós, Major and Tusnády (1975,1976), and Major (1976a). □

**Theorem 1.2.5**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on $(\Omega, \mathcal{A}, P)$ with $EX_1 = 0$ and $\mathrm{Var}X_1 = 1$ for all $n \in \mathbb{N}$. If*

$$E\exp(tX_1) < \infty$$

*for $|t| \leq t_0$, $t_0 > 0$, there exists a standard Wiener process $\{W(t) : t \geq 0\}$ such that*

$$S(T) - W(T) = \mathcal{O}(\log T) \qquad \text{a.s.}$$

*as $T \to \infty$.*

**Proof:** See Komlós, Major and Tusnády (1975,1976). □

As before, Theorems 1.2.4 and 1.2.5 provide the best possible rates under the given assumptions.

**The multi–dimensional case.** Soon after these seminal results were proved, first steps were taken to generalize them to a multivariate setting. To obtain similar statements for higher dimensions was not only forced by theoretical considerations, but also by a number of applications. First extensions of the Komlós, Major and Tusnády theorems are due to Berkes and Philipp (1979), Philipp (1979) or Berger (1982). While their results were essentially weaker, major progress goes back to Einmahl (1987,1989). His results will be cited now.

Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed $p$–dimensional, real–valued random vectors such that

$$E\boldsymbol{X}_1 = 0, \qquad \mathrm{Cov}\boldsymbol{X}_1 = \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is a positive definite $p \times p$ matrix. Let the sequence of partial sums $\{\boldsymbol{S}_n\}_{n\in\mathbb{N}_0}$ be defined by

$$\boldsymbol{S}_0 = 0, \qquad \boldsymbol{S}_n = \boldsymbol{X}_1 + \ldots + \boldsymbol{X}_n \quad (n \in \mathbb{N}).$$

We are interested in strong approximations of $\{\boldsymbol{S}_n\}_{n\in\mathbb{N}_0}$ by the partial sums of a sequence $\{\boldsymbol{Y}_n\}_{n\in\mathbb{N}}$ of independent, identically normally distributed random vectors having the same covariance structure $\boldsymbol{\Sigma}$ as the original sequence. So, let similarly $\{\boldsymbol{T}_n\}_{n\in\mathbb{N}_0}$ be defined by

$$\boldsymbol{T}_0 = 0, \qquad \boldsymbol{T}_n = \boldsymbol{Y}_1 + \ldots + \boldsymbol{Y}_n \quad (n \in \mathbb{N}).$$

Finally, let $\|\cdot\|$ denote the Euclidean norm in $\mathbb{R}^p$.

**Theorem 1.2.6 (Einmahl)**
*Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed $p$–dimensional random vectors on $(\Omega, \mathcal{A}, P)$ with $E\boldsymbol{X}_1 = 0$ and Cov$\boldsymbol{X}_1 = \boldsymbol{\Sigma}$. If*

$$E \|\boldsymbol{X}_1\|^{\nu} < \infty$$

*for some $\nu > 2$, there exists a sequence of independent, identically normal random vectors with Cov $\boldsymbol{Y}_1 = \boldsymbol{\Sigma}$ such that*

$$\|\boldsymbol{S}_n - \boldsymbol{T}_n\| = o\left(n^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as $n \to \infty$.*

**Proof:** See Einmahl (1987) if $2 < \nu < 4$ and Einmahl (1989) if $\nu > 3$. □

To establish the counterpart of Theorem 1.2.3 in the multivariate setting by an adaptation of the quantile transformation a further assumption is necessary. Let

$$M(\boldsymbol{t}) = E\exp(\langle \boldsymbol{t}, \boldsymbol{X}_1 \rangle) \qquad (t \in \mathbb{R}^p)$$

denote the moment generating function of $\boldsymbol{X}_1$, where $\langle \cdot, \cdot \rangle$ is the scalar product induced by the Euclidean norm.

**Theorem 1.2.7 (Einmahl)**
*Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed $p$–dimensional random vectors on $(\Omega, \mathcal{A}, P)$ with $E\boldsymbol{X}_1 = 0$ and Cov$\boldsymbol{X}_1 = \boldsymbol{\Sigma}$. If*

a) $M(\boldsymbol{t}) < \infty$ *for* $\| \boldsymbol{t} \| \leq t_0$, $t_0 > 0$ *and*

b) $\sup_{\|\boldsymbol{t}\|\leq t_0} \sup_{\|\boldsymbol{s}\|\leq\delta} \dfrac{|M(\boldsymbol{t}+i\boldsymbol{s})|}{M(\boldsymbol{t})} < 1$ *for all* $\delta > 0$,

*there exists a sequence of independent, identically normal random vectors with* Cov$\,\boldsymbol{Y}_1 = \boldsymbol{\Sigma}$ *such that*

$$\| \boldsymbol{S}_n - \boldsymbol{T}_n \| = \mathcal{O}(\log n) \qquad \text{a.s.}$$

*as* $n \to \infty$.

**Proof:** See Einmahl (1989). □

Without assumption b) the best available a.s. approximation rate would be of order $\mathcal{O}(\log n)^2$ only.

By the arguments already mentioned above, Theorems 1.2.6 and 1.2.7 can be reformulated in terms of a $p$–dimensional Wiener process, too. We say that $\{\boldsymbol{W}(t) : t \geq 0\}$ is a Wiener process in $\mathbb{R}^p$ with covariance matrix $\boldsymbol{\Sigma}$ if

a) $E\boldsymbol{W}(t) = \boldsymbol{0}$ for all $t \geq 0$,
b) the finite dimensional distributions of $\{\boldsymbol{W}(t) : t \geq 0\}$ are Gaussian and
c) $EW_i(s)W_j(t) = \sigma_{ij} \min\{s, t\}$ for all $s, t \geq 0$,

where $W_i(t)$ denotes the $i$–th component of $\boldsymbol{W}(t)$ and $\sigma_{ij}$ the $(i, j)$–th entry of $\boldsymbol{\Sigma}$. Let the $p$–dimensional partial sum process $\{\boldsymbol{S}(t) : t \geq 0\}$ be defined by

$$\boldsymbol{S}(0) = 0, \qquad \boldsymbol{S}(t) = \boldsymbol{S}_{[t]} \quad (t \in \mathbb{R}_+).$$

Then, we retain the previous theorems as the following strong approximations.

**Theorem 1.2.8**
*Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed $p$–dimensional random vectors on $(\Omega, \mathcal{A}, P)$ with $E\boldsymbol{X}_1 = 0$ and Cov$\boldsymbol{X}_1 = \boldsymbol{\Sigma}$. If*

$$E \| \boldsymbol{X}_1 \|^{\nu} < \infty$$

*for some $\nu > 2$, there exists a $p$–dimensional Wiener process $\{\boldsymbol{W}(t) : t \geq 0\}$ with covariance matrix $\boldsymbol{\Sigma}$ such that*

$$\| \boldsymbol{S}(T) - \boldsymbol{W}(T) \| = o\left(T^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as* $T \to \infty$.

**Proof:** See Einmahl (1987,1989). □

**Theorem 1.2.9**

*Let* $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ *be a sequence of independent, identically distributed* $p$*–dimensional random vectors on* $(\Omega, \mathcal{A}, P)$ *with* $E\boldsymbol{X}_1 = 0$ *and* $Cov\boldsymbol{X}_1 = \boldsymbol{\Sigma}$. *If*

*a)* $M(\boldsymbol{t}) < \infty$ *for* $\|\boldsymbol{t}\| \leq t_0$, $t_0 > 0$ *and*

*b)* $\sup_{\|\boldsymbol{t}\|\leq t_0} \sup_{\|\boldsymbol{s}\|\leq \delta} \frac{|M(\boldsymbol{t}+i\boldsymbol{s})|}{M(\boldsymbol{t})} < 1$ *for all* $\delta > 0$,

*there exists a* $p$*–dimensional Wiener process* $\{\boldsymbol{W}(t) : t \geq 0\}$ *with covariance matrix* $\boldsymbol{\Sigma}$ *such that*

$$\| \boldsymbol{S}(T) - \boldsymbol{W}(T) \| = \mathcal{O}(\log T) \qquad a.s.$$

*as* $T \to \infty$.

**Proof:** See Einmahl (1989). □

**Closing remarks.** Strassen's result (cf. Theorem 1.2.1) gives the optimal rate for just two existing moments. In case of existing higher moments, Theorems 1.2.2 and 1.2.4 (or their counterparts 1.2.6 and 1.2.8) have not been established under the most general point of view, since the given presentation suffices for our purposes. But Csörgő and Horváth (1993) point out, that it is possible to replace the condition $E|X_1|^\nu < \infty$ $(\nu > 2)$ by the general moment condition $EH(|X_1|) < \infty$, where $H$ is a strictly positive mapping on $\{x \geq 0\}$ satisfying

a) $\frac{H(x)}{x^{2+\eta}}$ is increasing for some $\eta > 0$ and

b) $\frac{\log H(x)}{x} = \frac{\log K(x)}{x}(1 + o(1))$

with a strictly positive mapping $K$ on $\{x \geq 0\}$ such that $x^{-1}\log K(x)$ is decreasing. The resulting a.s. rate changes from $o(n^{\frac{1}{\nu}})$ to $o(H^{-1}(n))$. There is also an analogue for random vectors.

Strong invariance principles are available not only for independent, identically distributed random variables and vectors, but also under a variety of dependence concepts. For a detailed discussion of specific examples we refer to Sections 2.4 or 4.3. Moreover, we shall derive a strong invariance principle for RCA (random coefficient autoregressive) time series in Chapter 3.

In Chapter 4 we shall discuss the behaviour of stochastic processes with drift and values in $\mathbb{R}^p$ which satisfy a strong invariance principle. It turns out that their suprema can be described appropriately by the suprema of the approximating $p$–dimensional Wiener process (with drift).

## 1.3 Weak approximations

Beside strong invariance principles also weak ones play an important role in the asymptotic theory. Weak approximations assure the convergence in distribution of certain functionals of partial sums. We will present a theorem due to Donsker (1951) and a stronger version traced back to Breiman (1968) which in turn implies Donsker's result.

**Donsker's theorem and an extension.** Construct from the partial sums of a sequence of independent, identically distributed random variables $\{X_n\}_{n\in\mathbb{N}}$ the sequence of stochastic processes $\{S_n(t) : 0 \le t \le 1\}_{n\in\mathbb{N}}$ on $\mathcal{C}[0,1] = \{f : [0,1] \to \mathbb{R} : f \text{ continuous}\}$ by

$$S_n(t) = \tfrac{1}{\sqrt{n}}\left(S_{[nt]} + (nt - [nt])X_{[nt]+1}\right) \qquad (t \in [0,1], n \in \mathbb{N}). \tag{1.1}$$

Then, for $n \in \mathbb{N}$ obviously

$$S_n\left(\tfrac{k}{n}\right) = \tfrac{1}{\sqrt{n}} S_k \qquad (k = 0, 1, \ldots, n)$$

and $S_n(t)$ is for $\frac{k}{n} < t < \frac{k+1}{n}$ obtained by linear interpolation. Donsker could prove the following convergence in distribution result for continuous functionals $h : \mathcal{C}[0,1] \to \mathbb{R}$, where continuity is regarded with respect to the supremum–norm.

**Theorem 1.3.1 (Donsker)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on $(\Omega, \mathcal{A}, P)$ with $EX_1 = 0$ and $\mathrm{Var}X_1 = 1$. For $n \in \mathbb{N}$ let the process $\{S_n(t) : 0 \le t \le 1\}$ be defined by (1.1). Then, there exists a standard Wiener process $\{W(t) : 0 \le t \le 1\}$ such that*

$$h(S_n(t)) \xrightarrow{\mathcal{D}} h(W(t))$$

*as $n \to \infty$ for every continuous functional $h : \mathcal{C}[0,1] \to \mathbb{R}$.*

**Proof:** See Donsker (1951). □

On choosing $h(f) = \sup_{0\le t\le 1} f(t)$, respectively, $h(f) = \sup_{0\le t\le 1} |f(t)|$, Theorem 1.3.1 implies immediately

$$\sup_{0\le t\le 1} S_n(t) \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1} W(t) \qquad (n \to \infty),$$

respectively,

$$\sup_{0\le t\le 1}|S_n(t)| \xrightarrow{\mathcal{D}} \sup_{0\le t\le 1}|W(t)| \qquad (n\to\infty)$$

by recognizing that both mappings are continuous with respect to the supremum–norm.

Although both proofs make use of the Skorohod embedding, the weak convergence result of Theorem 1.3.1 cannot be derived from Theorem 1.2.1. To obtain Donsker's theorem via Strassen's strong invariance an a.s. approximation rate of order $o(\sqrt{n})$ would be necessary. But Breiman (1967) as well as Major (1976b) could show that such a rate is impossible under the assumption of just two existing moments.

A different weak invariance principle is provided in the following theorem which conversely implies Donsker's theorem and which is due to Breiman (1968).

**Theorem 1.3.2**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on $(\Omega,\mathcal{A},P)$ with $EX_1=0$ and $VarX_1=1$. For $n\in\mathbb{N}$ let the process $\{S_n(t):0\le t\le 1\}$ be defined by (1.1). Then, there exists a probability space with a sequence of standard Wiener processes $\{W_n(t):0\le t\le 1\}$ and a sequence of stochastic processes $\{\tilde{S}_n(t):0\le t\le 1\}$ defined on it such that*

*a)* $\{\tilde{S}_n(t):0\le t\le 1\} \stackrel{\mathcal{D}}{=} \{S_n(t):0\le t\le 1\} \qquad (n\in\mathbb{N}),$

*b)* $\sup_{0\le t\le 1}|\tilde{S}_n(t)-W_n(t)| \xrightarrow{P} 0 \qquad (n\to\infty).$

**Proof:** See for example Theorem 2.1.2 in Csörgő and Révész (1981). □

**Closing remarks.** In fact, there is a kind of equivalence between the two concepts of strong and weak approximations which can be expressed via the Prohorov distance of probability measures. The precise connection can be found in Strassen (1965). For a detailed survey confer also Csörgő and Horváth (1993).

In Chapter 2, we shall use a weak invariance principle to determine the asymptotic behaviour of a delay time occurring in a sequential test procedure. Since we are only interested in the limit distribution, but not in a.s. results, we abstain from employing a strong approximation there.

# Chapter 2

# Delay Time in Sequential Detection of Change

The current chapter brings together the concepts of change–point analysis, sequential statistics and invariance principles. The sections are organized as follows.

Sections 2.1 and 2.2 provide preliminaries, the first one laying the foundations of change–point analysis in an a–posteriori setting. The second section describes the basics of sequential testing as far as necessary for our purposes.

In Section 2.3, we define a location model to test for a possible change in the mean of underlying random variables with a sequential procedure. We are interested in the asymptotic behaviour and hence introduce a weak invariance principle satisfying a certain approximation rate.

The following section is devoted to possible applications of the location model in concrete situations. We consider the case of independent, identically distributed random variables and linear processes. Furthermore, we shall discuss GARCH processes, which are successfully used to model the volatility of financial markets, and random variables fulfilling even more general dependence conditions (see Section 2.4).

In Section 2.5, the asymptotic behaviour of the test statistic under both the null hypothesis and the alternative shall be given.

The main section of this chapter contains the investigation of the stopping rule. It turns out that under some additional but not too restrictive assumptions, a limiting distribution can be given which only depends on mean and variance of the underlying sequence but not on its specific distribution. The problems of Section 2.6 have been proposed by Lajos Horváth from the University of Utah and have been submitted to the Statistics and Probability Letters as a joint paper (see Aue and Horváth (2003a)).

## 2.1 Change–point analysis

Change–point analysis provides tools to investigate whether a chronologically ordered set of observations remains stable over time (is "in order") or if the observations follow one pattern up to an unknown time–point and a different pattern afterwards (up to – maybe – another unknown time–point, and so on). This very general description can be specialized to numerous submodels. Some of them shall be discussed in this section. Firstly, we start with a summary of the historical developments. Then we move on to a statistical formulation of change–point problems and put special emphasis on the change in the mean scenario for abrupt and gradual changes.

**Overview.** Some of the first change–point problems were considered by Page (1954, 1955). He was interested in economic questions arising in quality control. Therein, the immediate detection of deviations from an acceptable level is of great importance for companies to minimize their expenses. To this matter, Page (1954) suggested an inspection scheme (suitably formalizing the heuristics from above) and also a test for a parameter change at an unknown time–point (1955).

More insight was gained some years later, when the results of the Hungarian construction school were applied to change–point analysis (cf. Chapter 1). Details on this topic and on many more may be found in Csörgő and Horváth (1997). A lot of important limit theorems were obtained which themselves led to asymptotic tests and to asymptotic estimators of the unknown parameters.

Today, change–point analysis plays a key role in a number of fields as for example in physics, medicine, biology or meteorology, and is applied even in areas as ethnology and archaeology. Classical sets of observations are for instance the Nile data giving the average annual water level of the Nile river in Aswan between 1871 and 1970, or the Clementinum data collecting the average temperatures in Prague from 1775–1989. Statistical investigations were carried out by Cobb (1978) in case of the Nile data, the Clementinum data was for example examined by Horváth, Kokoszka and Steinebach (1999) or Horváth, Hušková, Kokoszka and Steinebach (2003). One of the most remarkable examples is the so–called "Land's End data set" of 52 stone configurations in Cornwall (England). Kendall and Kendall (1980) employed a Poisson model to answer the question if there are too many straight line configurations.

All these applications can be distinguished by the different kind of data acquisition. If all observations are available at the beginning of the test procedure, we speak of *a–posteriori* change–point problems. If instead, a new test is applied after each arriving observation, we are in the area of *sequential* change–point analysis. While most of the papers of the past were devoted to a–posteriori problems, there have recently been some results for sequential procedures, too. Their description is postponed to the next section, while we will now state a–posteriori approaches only. Moreover, we focus on the discussion of nonparametric models. For various parametric settings confer for example Chen and Gupta (2000). Some of their asymptotic arguments however should only be used with great care.

**A–posteriori change–point analysis.** A typical formulation of the test problem can be given as follows. Let $X_1, \ldots, X_n$ be independent, real–valued random variables on a probability space $(\Omega, \mathcal{A}, P)$. Based on a given (historical) set of observations we wish to decide between the hypotheses

$$\begin{aligned} \mathcal{H}_\mathbf{0} &: X_1 \stackrel{\mathcal{D}}{=} \ldots \stackrel{\mathcal{D}}{=} X_n, \\ \mathcal{H}_\mathbf{A} &: \text{not } \mathcal{H}_\mathbf{0}. \end{aligned}$$

The null hypothesis $\mathcal{H}_\mathbf{0}$ describes the fact that all observations follow the same probability law, while the alternative $\mathcal{H}_\mathbf{A}$ just states that this is not the case. To apply reasonable test statistics a more concrete alternative is desirable. Here, we will put our emphasis on the most common submodel, the so–called AMOC ("at most one change–point") model for a change in the expectation of the underlying random variables. In this case $\mathcal{H}_\mathbf{0}$ and $\mathcal{H}_\mathbf{A}$ can be restated as

$$\begin{aligned} \mathcal{H}_\mathbf{0} &: EX_1 = \ldots = EX_n, \\ \mathcal{H}_\mathbf{A} &: \text{There exists a } k^* < n \text{ such that} \\ &\quad EX_1 = \ldots = EX_{k^*} \neq EX_{k^*+1} = \ldots = EX_n, \end{aligned}$$

where $k^*$ is unknown and $X_1, \ldots, X_n$ are assumed to have finite expectation. To apply (strong or weak) invariance principles and carry out the statistical analysis based on asymptotic properties, it is more convenient to use the representation

$$X_j = \begin{cases} \mu + \varepsilon_j & : \quad 1 \leq j \leq k^*, \\ \mu + \Delta + \varepsilon_j & : \quad k^* < j \leq n, \end{cases} \tag{2.1}$$

where $\mu$ and $\Delta = \Delta_n$ are unknown real parameters, the latter possibly depending on $n$. Moreover, $\varepsilon_1, \ldots \varepsilon_n$ are independent, identically distributed random variables satisfying

$$E\varepsilon_1 = 0, \qquad 0 < \sigma^2 = \operatorname{Var} \varepsilon_1 < \infty, \qquad E|\varepsilon_1|^\nu < \infty$$

for some $\nu > 2$. The third condition puts us in a position to apply the Komlós, Major and Tusnády theorems of Chapter 1, since enough moments are assured to construct an approximating Wiener process with an appropriate rate. As a consequence, the hypotheses simplify once more to

$$\mathcal{H}_{\mathbf{0}} \;:\; k^* = n,$$
$$\mathcal{H}_{\mathbf{A}} \;:\; k^* < n, \quad \Delta = \Delta_n \neq 0.$$

As already mentioned in Section 1.1, a test statistic can be given in terms of the partial sums of $X_1, \ldots, X_n$ by

$$T_{n,1} = \frac{1}{\sqrt{n}\sigma} \max_{k=1,\ldots,n} \left| S_k - \tfrac{k}{n} S_n \right|, \tag{2.2}$$

where $S_k = X_1 + \ldots + X_k$ for $k = 1, \ldots, n$. Therein, taking the maximum reflects the fact that $k^*$ is unknown in contrast to the example of Section 1.1 (confer also Example 2.2.1 in Section 2.2). Under $\mathcal{H}_{\mathbf{0}}$,

$$E\left(S_k - \tfrac{k}{n} S_n\right) = 0 \qquad (k = 1, \ldots, n),$$

since the deterministic parts of the difference cancel out and the remaining random parts have expectation zero by assumption on $\varepsilon_1, \ldots, \varepsilon_n$. Under $\mathcal{H}_{\mathbf{A}}$ however, the expectation is non–zero. So, heuristically we expect small values of $T_{n,1}$ under the null hypothesis, but on the other hand great absolute values under the alternative.
The finite sample performance of $T_{n,1}$ is illustrated by the following example, where we assume a specific distribution for $X_1, \ldots, X_n$.

**Example 2.1.1 (Finite samples)**
Let $n = 100$, $\mu = 0$, $\Delta = 1$, $k^* = 50$ and let $\varepsilon_1, \ldots, \varepsilon_n$ be standard normal random variables. Then, the first 50 observations follow the $\mathcal{N}(0,1)$ law, while the final 50 are $\mathcal{N}(1,1)$ distributed. One realization of $X_1, \ldots, X_{100}$ can be found in Figure 2.1. Then, Figure 2.2 shows the behaviour of $T_{100,1}$ applied to the data, or more exactly, the sample path of $k \mapsto S_k - \frac{k}{100} S_{100}$ based on the outcome. The critical value $c = 1.358$ has been obtained to the prescribed level $\alpha = 0.05$.

Example 2.1.1 and therein especially Figure 2.2 give us a good hint on how to define an estimator for the change–point $k^*$. We set

$$\hat{k}^* = \operatorname*{arg\,max}_{k=1,\ldots,n} \left| S_k - \tfrac{k}{n} S_n \right|. \tag{2.3}$$

The asymptotic properties of the test statistic $T_{n,1}$ and the corresponding change–point estimator $\hat{k}^*$ are given in the following theorems.

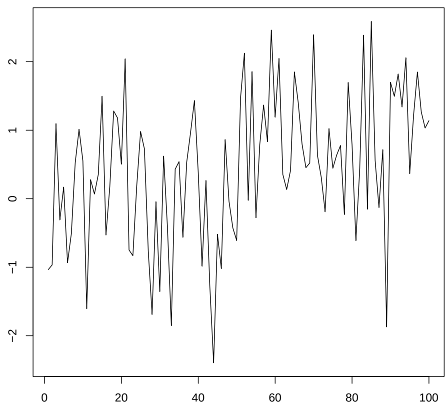


Figure 2.1: A realization of $X_1, \ldots, X_{100}$

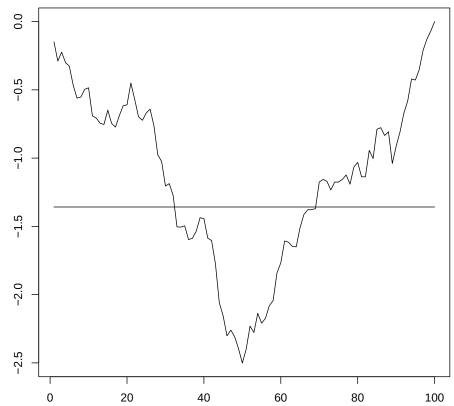


Figure 2.2: Behaviour of $S_k - \frac{k}{n} S_n$

**Theorem 2.1.1**
*Let the random variables $X_1, \ldots, X_n$ follow the model in (2.1) and let $T_{n,1}$ be defined as in (2.2). If $E\varepsilon_1^2 < \infty$, then, under $\mathcal{H}_\mathbf{0}$,*

$$T_{n,1} \xrightarrow{\mathcal{D}} \sup_{0 \le t \le 1} |B(t)|$$

*as $n \to \infty$, where $\{B(t) : 0 \le t \le 1\}$ denotes a standard Brownian bridge.*

**Proof:** It follows from Theorem 1.3.1. □

**Theorem 2.1.2**
*Let the random variables $X_1, \ldots, X_n$ follow the model in (2.1) and let $\hat{k}^*$ be defined as in (2.3). If*

$$\Delta_n \to 0, \qquad \frac{\sqrt{n}|\Delta_n|}{\sqrt{\log \log n}} \to \infty$$

*as $n \to \infty$, then, under $\mathcal{H}_\mathbf{A}$,*

$$\frac{\Delta_n^2}{\sigma^2} \left( \hat{k}^* - k^* \right) \xrightarrow{\mathcal{D}} \arg\max_{t \in \mathbb{R}} \left\{ W(t) - \frac{|t|}{2} \right\}$$

*as $n \to \infty$, where $\{W(t) : t \in \mathbb{R}\}$ denotes the two–sided standard Wiener process.*

**Proof:** See Csörgő and Horváth (1997), Theorem 2.8.2. □

More general stochastic processes covering the partial sum processes obtained from the random variables in (2.1) have been studied by Horváth and Steinebach (2000), who examined the testing procedure, and by Kühn and Steinebach (2002) in case of estimating the change–point.

**Gradual changes.** As mentioned in the overview, there is a great variety of further (sub)models. While the mean change was an abrupt one during our motivating example, jumping from $\mu$ to $\mu+\Delta$ at $k^*$, also finer changes – called gradual – have been considered. Hušková (1999), and Hušková and Steinebach (2000) (cf. also Hušková and Steinebach (2002)) investigated the random variables

$$Y_i = \mu + \Delta \left( \frac{i-k^*}{n} \right)_+^{\alpha} + \varepsilon_i \qquad (i=1,\dots,n), \tag{2.4}$$

where $\mu$, $\Delta = \Delta_n$ and $k^*$ are unknown parameters, while $\alpha \in (0,1]$ is supposed to be known. The errors $\varepsilon_1,\dots,\varepsilon_n$ are assumed to satisfy the conditions from above. Moreover, $x_+ = \max\{x,0\}$ for $x \in \mathbb{R}$. Hušková and Steinebach (2000) studied the hypotheses

$$\mathcal{H}_{\mathbf{0}} \;:\; k^* = n,$$
$$\mathcal{H}_{\mathbf{A}} \;:\; k^* < n, \quad \Delta = \Delta_n \neq 0.$$

It turned out that the test statistic

$$T_{n,2} = \frac{1}{\sigma} \max_{k=1,\dots,n-1} \frac{\left|\sum_{i=1}^{n} (x_{ik}(\alpha) - \bar{x}_k(\alpha)) Y_i\right|}{\left(\sum_{i=1}^{n} (x_{ik}(\alpha) - \bar{x}_k(\alpha))^2\right)^{\frac{1}{2}}}, \tag{2.5}$$

based on the partial sums of weighted residuals satisfies an extreme value asymptotic under $\mathcal{H}_{\mathbf{0}}$. Here we have used the notation

$$x_{ik}(\alpha) = \left( \frac{i-k}{n} \right)^{\alpha}, \qquad \bar{x}_k(\alpha) = \frac{1}{n} \sum_{j=1}^{n} x_{jk} \qquad (i,k=1,\dots,n).$$

**Theorem 2.1.3 (Hušková and Steinebach)**
*Let the random variables $Y_1,\dots,Y_n$ follow the model in (2.4) and let $T_{n,2}$ be defined as in (2.5). Then, under $\mathcal{H}_{\mathbf{0}}$,*

$$\lim_{n\to\infty} P\left(a_n T_{n,2} \le x + b_n(\alpha)\right) = \exp\left(-2e^{-x}\right)$$

*for all $x \in \mathbb{R}$, where*

$$a_n = \sqrt{2 \log\log n}$$

*and in case*

$$\alpha > \tfrac{1}{2} \;:\; b_n(\alpha) = 2\log\log n + \log\left( \frac{1}{4\pi} \sqrt{\frac{2\alpha+1}{2\alpha-1}} \right),$$
$$\alpha = \tfrac{1}{2} \;:\; b_n(\alpha) = 2\log\log n + \tfrac{1}{2} \log\log\log\log n - \log(4\pi),$$
$$0 < \alpha < \tfrac{1}{2} \;:\; b_n(\alpha) = 2\log\log n + \frac{1-2\alpha}{2(2\alpha+1)} \log\log\log n + \log\left( \frac{C_\alpha^{\frac{1}{2\alpha+1}} H_{2\alpha+1}}{\sqrt{\pi}\, 2^{\frac{2\alpha}{2\alpha+1}}} \right)$$

*with*

$$C_\alpha = -(2\alpha+1)\int_0^\infty y^\alpha \left((y+1)^\alpha - y^\alpha - \alpha y^{\alpha-1}\right) dy$$

*and $H_{2\alpha+1}$ being defined in Remark 12.2.10 of Leadbetter, Lindgren and Rootzén (1983).*

**Proof:** See Hušková and Steinebach (2000). □

It is worthwhile mentioning, that the results of Theorem 2.1.3 hold true also for $\alpha > 1$. Moreover, Hušková (1999) introduced the least squares type change–point estimator

$$\hat{k}^*(\alpha) = \underset{k=1,\dots,n-1}{\arg\max} \frac{|\sum_{i=1}^n (x_{ik}(\alpha) - \bar{x}_k(\alpha) Y_i)|}{(\sum_{i=1}^n (x_{ik}(\alpha) - \bar{x}_k(\alpha))^2)^{\frac{1}{2}}} \tag{2.6}$$

whose limiting behaviour is given by the next theorem. Again, the cases $0 < \alpha < \frac{1}{2}$, $\alpha = \frac{1}{2}$ and $\frac{1}{2} < \alpha \leq 1$ have to be considered separately.

**Theorem 2.1.4 (Hušková)**
*Let the random variables $Y_1, \dots, Y_n$ follow the model in (2.4) and let $\hat{k}^*(\alpha)$ be defined as in (2.6). If*

$$\Delta_n \to 0, \qquad \frac{\sqrt{n}|\Delta_n|}{\sqrt{\log\log n}} \to \infty \qquad (n \to \infty),$$

*and $k^* = [\theta n]$ with some $\theta \in (0,1)$, then under $\mathcal{H}_{\mathbf{A}}$, the following statements hold true.*

*a) Let $\alpha \in (0, \frac{1}{2})$. Then,*

$$\left(\frac{\Delta_n^2}{n^{2\alpha}}\right)^{\frac{1}{2\alpha+1}} \left(\hat{k}^*(\alpha) - k^*\right) \xrightarrow{\mathcal{D}} \sigma V_\alpha \qquad (n \to \infty),$$

*as $n \to \infty$, where*

$$V_\alpha = \underset{t\in\mathbb{R}}{\arg\max}\left\{W_\alpha(t) - \frac{1}{2}\int_{-\infty}^\infty \left((x+t)_+^\alpha - x_+^\alpha\right)^2 dx\right\}$$

*with a Gaussian process $\{W_\alpha(t) : t \in \mathbb{R}\}$ having zero mean and covariance function*

$$Cov\,(W_\alpha(s), W_\alpha(t)) = \int_{-\infty}^\infty \left((x+s)_+^\alpha - x_+^\alpha\right)\left((x+t)_+^\alpha - x_+^\alpha\right) dx \qquad (s,t \in \mathbb{R}).$$

*b) Let $\alpha = \frac{1}{2}$. Then,*

$$\frac{\Delta_n\sqrt{\log(n-k^*)}}{2\sqrt{n}}\left(\hat{k}^*(\tfrac{1}{2}) - k^*\right) \xrightarrow{\mathcal{D}} \sigma V_{\frac{1}{2}}$$

*as $n \to \infty$, where $V_{\frac{1}{2}}$ is a standard normal random variable.*

*c) Let* $\frac{1}{2} < \alpha \leq 1$*. Then,*

$$\frac{\Delta_n}{\sqrt{n}} \left( \hat{k}^*(\alpha) - k^* \right) \xrightarrow{\mathcal{D}} \sigma V_\alpha$$

*as* $n \to \infty$*, where* $V_\alpha$ *is normally distributed with mean zero and variance*

$$(1-\theta)^{2\alpha-1} \left( \frac{2\alpha+1}{4} \frac{(\alpha-1+2\theta)^2}{\alpha^2+\theta(1+2\alpha)} - \frac{(\alpha-1)^2+\theta(2\alpha-1)}{2\alpha-1} \right).$$

**Proof:** See Hušková (1999) if $\alpha \in (0,1)$ and Hušková (1998) if $\alpha = 1$. □

Alternative models assuming a stochastic process analogue of the random variables from above are due to Steinebach (1999). The estimation procedure for gradual changes was examined by Aue and Steinebach (2002), confer also Aue (2000) for details.

**Closing remarks.** Throughout the discussion of test statistics and change–point estimators the variance parameter $\sigma^2$ of the error variables $\varepsilon_1, \ldots, \varepsilon_n$ was assumed to be known. But the results of Theorems 2.1.2 – 2.1.4 hold true even if $\sigma^2$ is for instance replaced by consistent estimators $\hat{\sigma}_n^2$. (See the corresponding articles of Horváth and Steinebach (2000), Hušková (1999), and Hušková and Steinebach (2000).) Moreover, we have restricted the presentation to two–sided tests. One–sided alternatives can be introduced in a similar way.

Of course there are many more testing and estimation procedures for other parameters than the expectation. We skip the discussion of changes in the variance, covariance, etc. and refer to Chen and Gupta (2000) or to Csörgő and Horváth (1997). The results in one dimension have been transferred to multivariate settings, too. Moreover, there are approaches to investigate multiple change–point problems (for example epidemic alternatives, where the parameter in question changes after a first time–point and returns to its original value after a second time–point). Once again, detailed surveys may be found in Csörgő and Horváth (1997).

## 2.2 Sequential statistics

Aim of this section is the formulation of sequential tests and their distinction from the tests of the previous section. Moreover, we will discuss two extensions to linear models of Wald's (1947) testing approach by Chu, Stinchcombe and White (1996), and by Horváth, Hušková, Kokoszka and Steinebach (2003). Therein, we shall give another example revealing the connection between (log)likelihood ratios and the CUSUM test procedures introduced in the previous section. Finally, we close the section with some remarks on a different sequential approach carried out by Gut and Steinebach (2002,2003) for renewal processes.

**Sequential tests.** In Section 2.1, we have always considered a historical data set of given size. The typical approach was to carry out the statistical investigation via invariance principles, i.e., by an embedding into a Gaussian framework.

A different strategy introduced by Wald (cf. for example Wald's (1947) monograph) is the following. Instead of treating the sample size as a constant, in a (so–called) sequential test the number of observations required is itself a random variable. At each stage of the experiment one of the three decisions

- accept the null hypothesis $\mathcal{H}_{\mathbf{0}}$,
- reject the null hypothesis $\mathcal{H}_{\mathbf{0}}$ or
- continue taking observations

has to be chosen due to a given rule. If one of the first two decisions is made, the testing procedure is terminated. Else another trial will be performed. Since the total number of observations depends on the specific outcome, it is not predetermined but a random variable. This leads to the formulation of the sequential probability ratio test (shortly SPRT) for simple hypotheses. We will give a short presentation now. For more detailed information confer for example Siegmund (1985), Chapter 2.

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent random variables on some probability space $(\Omega, \mathcal{A}, P)$. We are interested in testing the simple hypotheses

$$\begin{aligned} \mathcal{H}_{\mathbf{0}} &: \quad f_n = f_0 \quad \text{ for } n \in \mathbb{N}, \\ \mathcal{H}_{\mathbf{A}} &: \quad f_n = f_1 \quad \text{ for } n \in \mathbb{N}, \end{aligned}$$

where $f_n$ denotes the probability density of $X_n$ $(n \in \mathbb{N})$. Define the likelihood ratio

$$L_n = L_n(X_1, \ldots, X_n) = \frac{\prod_{i=1}^n f_0(X_i)}{\prod_{i=1}^n f_1(X_i)} \qquad (n \in \mathbb{N}),$$

and stop the observations at the first $n \in \mathbb{N}$ such that $L_n \notin (A, B)$ for predetermined constants $0 < A < 1 < B < \infty$, i.e., we consider the stopping time

$$\tau = \min\{n \geq 1 : L_n \notin (A, B)\},$$

where $\min \emptyset = \infty$. Then, reject $\mathcal{H}_{\mathbf{0}}$ if $L_\tau \geq B$ and accept $\mathcal{H}_{\mathbf{0}}$ if $L_\tau \leq A$. The following example gives us a good idea about the connections between the likelihood ratios and the CUSUM procedures of the previous section.

**Example 2.2.1 (Normal distribution)**
Let $\{X_n\}_{n\in\mathbb{N}}$ be independent normally distributed with common and unit variance. Easy calculations show that for testing $\mathcal{H}_\mathbf{0} : \mu = \mu_0$ against $\mathcal{H}_\mathbf{A} : \mu = \mu_1$ with $\mu_0 < \mu_1$, the likelihood ratio is given by

$$L_n = \exp\left((\mu_1 - \mu_0)S_n - \tfrac{n}{2}(\mu_1^2 - \mu_0^2)\right),$$

where $S_n = X_1 + \ldots + X_n$. Now, the stopping rule can be equivalently stated as

$$\tau = \min\{n \geq 1 : S_n - \tfrac{n}{2}(\mu_1 + \mu_0) \notin (a, b)\},$$

where

$$a = \log \frac{A}{\mu_1 - \mu_0}, \qquad b = \log \frac{B}{\mu_1 - \mu_0}.$$

Even simpler is the symmetric case $\mu_1 = -\mu_0$. Then, we stop our testing procedure at the first $n \geq 1$ such that $|S_n| \geq b$.

One important question arises immediately: Does the sequential test terminate in finite time? From a practical point of view, infinitely many observations are intolerable if collecting data is assumed to be costly both under the null hypothesis $\mathcal{H}_\mathbf{0}$ and the alternative $\mathcal{H}_\mathbf{A}$. A positive answer to the problem is provided in the following theorem.

**Theorem 2.2.1**
*If $P\{L_n = 1\} \neq 1$ for all $n \in \mathbb{N}$, then*

*a)* $P\{\tau < \infty\} = 1$,
*b)* $E\tau < \infty$.

**Proof:** See Wald (1947). □

The constants $A$ and $B$ in the definition of the stopping rule $\tau$ can be related to the errors of first and second type, i.e., to

$$\alpha = P_{\mathcal{H}_\mathbf{0}}\{L_\tau \geq B\} \qquad \text{and} \qquad \beta = P_{\mathcal{H}_\mathbf{A}}\{L_\tau \leq A\}$$

by the simple approximations

$$\alpha \approx \frac{1 - A}{B - A}, \qquad \beta \approx \frac{A(B-1)}{B - A}.$$

So, tests of a certain level and power can be easily constructed from the latter.

Deeper investigations have shown that the SPRT is optimal for testing simple hypotheses for independent, identically distributed random variables. More exactly, the

SPRT minimizes the expected sample size under both hypotheses among all tests with no larger error probabilities. Unfortunately, this property is not retained for extensions of the SPRT to composite hypotheses (see Siegmund (1985), Sections 2.3 and 2.5).

**A modification.** We shall consider possible modifications of the SPRT which are applicable in the change–point analysis environment now. In their paper, Chu, Stinchcombe and White (1996) argue convincingly that in a great number of applications nowadays data arrive steadily and are cheaply obtainable. Hence, the basis of Wald's considerations can be weakened in the following direction. During the "in control" scenario the statistician is satisfied to keep on observing, since this process is (nearly or completely) cost free. So, under $\mathcal{H}_{\mathbf{0}}$, we do not necessarily have $P_{\mathcal{H}_{\mathbf{0}}}\{\tau < \infty\} = 1$. Confer however the discussion in Section 2.3. Under this point of view, Chu, Stinchcombe and White (1996) introduced the following linear regression model. Let the sequence of random variables $\{Y_n\}_{n\in\mathbb{N}}$ be defined by

$$Y_n = \boldsymbol{X'}_n \boldsymbol{\beta}_n + \varepsilon_n, \qquad (n \in \mathbb{N}),$$

where $\boldsymbol{X}_n$ is a random vector with values in $\mathbb{R}^p$ and $\boldsymbol{\beta}_n \in \mathbb{R}^p$ is deterministic.
One essential assumption is the so–called non–contamination

$$\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 \qquad (n = 1, \dots, m).$$

That means, the regresssion parameters stay "in control" for some historical data set of size $m$. Of interest are the hypotheses

$$\begin{aligned} \mathcal{H}_{\mathbf{0}} \quad &: \quad \boldsymbol{\beta}_n = \boldsymbol{\beta}_0 \qquad (n = m+1, m+2, \dots), \\ \mathcal{H}_{\mathbf{A}} \quad &: \quad \text{There exists a } k^* \geq 1 \text{ such that} \\ &\qquad \boldsymbol{\beta}_{m+1} = \dots = \boldsymbol{\beta}_{m+k^*-1} = \boldsymbol{\beta}_0 \neq \boldsymbol{\beta}_1 = \boldsymbol{\beta}_{m+k^*} = \boldsymbol{\beta}_{m+k^*+1} = \dots, \end{aligned}$$

where $\boldsymbol{\beta}_0, \boldsymbol{\beta}_1$ and $k^*$ are unknown parameters. The setting was picked up by Horváth, Hušková, Kokoszka and Steinebach (2003). Under assumptions on $\{\varepsilon_n\}_{n\in\mathbb{N}}$ which admit the usage of the Komlós, Major and Tusnády strong approximations, they provided two test procedures based on CUSUMs of residuals. Therefore, introduce the least squares estimator

$$\hat{\boldsymbol{\beta}}_n = \left( \sum_{i=1}^{n} \boldsymbol{X}_i \boldsymbol{X}'_i \right)^{-1} \sum_{j=1}^{n} Y_j \boldsymbol{X}_j$$

of the regression parameter $\beta$ at time $n$. Then, Horváth, Hušková, Kokoszka and Steinebach (2003) examined detectors based on the residuals

a) $\hat{\varepsilon}_i = Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}_m \qquad (i \in \mathbb{N}),$

b) $\tilde{\varepsilon}_i = Y_i - \boldsymbol{X}_i'\hat{\boldsymbol{\beta}}_{i-1} \qquad (i \in \mathbb{N} \setminus \{1\}).$

The first residual relies only on the historical data set, for which the regression parameters are assumed to be "in control". In contrast, the second residual is obtained recursively, making use of all previously taken observations. The residuals lead to two CUSUM procedures:

a) $\hat{Q}(m,k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i,$

b) $\tilde{Q}(m,k) = \sum_{i=m+1}^{m+k} \tilde{\varepsilon}_i.$

Before we discuss the asymptotics of $\hat{Q}(m,k)$, $\tilde{Q}(m,k)$ and establish corresponding stopping rules, we impose some conditions on the random variables $\{\varepsilon_n\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$.

**Assumption 2.2.1**

*a) Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be independent and identically distributed random variables with*

$$E\varepsilon_1 = 0, \qquad 0 < \sigma^2 = Var\,\varepsilon_1 < \infty, \qquad E|\varepsilon_1|^\nu < \infty$$

*for some $\nu > 2$.*

*b) Let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ and $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be independent.*

*c) There is a positive definite matrix $\boldsymbol{C} \in \mathbb{R}^{p\times p}$ such that*

$$\left\| \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{X}_i\boldsymbol{X}_i' - \boldsymbol{C} \right\|_\infty = \mathcal{O}\left(n^{-\beta}\right) \qquad a.s.$$

*as $n \to \infty$ for some constant $\beta > 0$, where here and in the sequel $\|\cdot\|_\infty$ denotes the maximum norm of vectors and matrices.*

*d) Finally, we assume that $X_{1,n} = 1$ for $n \in \mathbb{N}$.*

Now, the first stopping rule can be defined as follows:

$$\hat{\tau}_m = \min\{k \geq 1 : |\hat{Q}(m,k)| \geq \hat{g}(m,k)\},$$

where $\min \emptyset = \infty$ and

$$\hat{g}(m,k) = c\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^\gamma$$

with $\gamma \in [0, \min\{\beta, \frac{1}{2}\})$ and $c = c(\alpha)$ being prescribed by

$$\lim_{m\to\infty} P\{\hat{\tau}_m < \infty\} = \alpha.$$

For a more detailed discussion on choosing the boundary function $\hat{g}(m,k)$ confer the next section. We are now in a position to give the limiting behaviour under both hypotheses.

**Theorem 2.2.2**

*Let the conditions of Assumption 2.2.1 be satisfied and let*

$$\hat{\sigma}_m^2 = \frac{1}{m-p} \sum_{i=1}^{m} \left( Y_i - \boldsymbol{X}_i' \hat{\boldsymbol{\beta}}_m \right)^2$$

*be the least squares estimator of $\sigma^2$ based on the historical data. Then, the following statements hold true.*

*a) Under $\mathcal{H}_{\mathbf{0}}$,*

$$\lim_{m\to\infty} P\left\{ \frac{1}{\hat{\sigma}_m} \sup_{1\le k<\infty} \frac{|\hat{Q}(m,k)|}{\hat{g}(m,k)} \le 1 \right\} = P\left\{ \sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma} \le c \right\},$$

*where $\{W(t) : 0 \le t \le 1\}$ denotes a standard Wiener process.*

*b) If $\|\boldsymbol{C}_1'(\boldsymbol{\beta}_1 - \boldsymbol{\beta}_0)\|_\infty > 0$, then under $\mathcal{H}_{\mathbf{A}}$,*

$$\frac{1}{\hat{\sigma}_m} \sup_{1\le k<\infty} \frac{|\hat{Q}(m,k)|}{\hat{g}(m,k)} \xrightarrow{P} \infty$$

*as $m \to \infty$.*

**Proof:** See Horváth, Hušková, Kokoszka and Steinebach (2003). □

The situation turns out to be more complicated for $\tilde{Q}(m,k)$. Similarly, we define the stopping rule

$$\tilde{\tau}_m = \min\{k \ge 1 : |\tilde{Q}(m,k)| \ge \tilde{g}(m,k)\},$$

where

$$\tilde{g}(m,k) = \sqrt{m} h \left( \frac{k}{m} \right).$$

Therein, the function $h$ is assumed to satisfy the following conditions:

a) $\lim_{h\to 0} \dfrac{t^\gamma}{h(t)} = 0$ with some $\gamma \in [0, \min\{\beta, \frac{1}{2}\})$,

b) $\limsup_{t\to\infty} \frac{\sqrt{t\log\log t}}{h(t)} < \infty$ and

c) $h(t)$ is positive and continuous on $(0,\infty)$.

The precise asymptotics of $\tilde{\tau}_m$ are stated in the next theorem.

**Theorem 2.2.3**

*Let the assumptions of Theorem 2.2.2 be satisfied. Then, the following statements hold true.*

*a) Under $\mathcal{H}_{\mathbf{0}}$,*

$$\lim_{m\to\infty} P\left\{\frac{1}{\hat{\sigma}_m}\sup_{1\le k<\infty}\frac{|\tilde{Q}(m,k)|}{\tilde{g}(m,k)} \le 1\right\} = P\left\{\sup_{0<t<\infty}\frac{|W(t)|}{h(t)} \le 1\right\},$$

*where $\{W(t) : t \ge 0\}$ denotes a standard Wiener process.*

*b) If there is a $\tilde{k} \in \mathbb{N}$ such that*

$$\frac{m+k^*}{\sqrt{m}h\left(\frac{\tilde{k}}{m}\right)}\left\|\sum_{i=m+k^*}^{m+\tilde{k}}\frac{1}{i}\boldsymbol{X}_i'(\boldsymbol{\beta}_0-\boldsymbol{\beta}_1)\right\|_\infty \to \infty,$$

*then under $\mathcal{H}_{\mathbf{A}}$,*

$$\frac{1}{\hat{\sigma}_m}\sup_{1\le k<\infty}\frac{|\tilde{Q}(m,k)|}{\tilde{g}(m,k)} \to \infty$$

*as $m\to\infty$.*

**Proof:** See Horváth, Hušková, Kokoszka and Steinebach (2003). □

The approach of this paragraph will be transferred to a location model in the following section. We shall introduce a test procedure which is an analogue of $\hat{Q}(m,k)$ in a different setting.

**Closing remarks.** A further shortcoming of the SPRT is the following. If the probability densities $f_0$ and $f_1$ are "close to each other" in the sense that

$$E\left(\frac{f_0(X_1)}{f_1(X_1)}\right) \approx 1,$$

then occasionally very large sample sizes occur. One way to get around this difficulty is the truncation of the stopping rule: The null hypothesis $\mathcal{H}_{\mathbf{0}}$ is accepted if after a maximum number of observations the test procedure did not terminate. This approach was carried out by Gut and Steinebach (2002,2003) to detect changes in the parameters of a renewal counting process via invariance principles.

To the best of our knowledge, there has not been any attempt to establish an estimation procedure in a sequential setting to derive the limit distribution for the delay time of a stopping rule under the point of view of change–point analysis. We shall come back to this topic in Section 2.6 below.

## 2.3 The location model

As we have seen, a typical approach in change–point analysis is to carry out the statistical investigation based on a given data set of fixed size. Tests of this kind are often called to be of one–shot type and an extensive amount of literature is available for them (cf. for example Csörgő and Horváth (1997), Chapter 2).

On the other hand though, sequential procedures seem to be more realistic in many applications, for example in economics. Moreover, one–shot tests cannot be applied repeatedly each time new data arrives, as is shown in Example 2.3.1, which goes back to Robbins (1970) and uses the law of the iterated logarithm for Wiener processes.

So, we will introduce a sequential method, which falls back on the extension of Wald's SPRT obtained by Chu, Stinchcombe and White (1996) and which was refined by Horváth, Hušková, Kokoszka and Steinebach (2003). Instead of investigating a linear model, we will focus our attention on a location model, in which we are searching for a possible change in the mean based on observations of the random variables $\{X_i\}_{i\in\mathbb{N}}$ following the equations

$$X_i = \begin{cases} \mu + \varepsilon_i & : \quad i = 1, \ldots, m + k^* - 1, \\ \mu + \Delta_m + \varepsilon_i & : \quad i = m + k^*, m + k^* + 1, \ldots, \end{cases} \tag{2.7}$$

where $\mu$ and $\Delta_m$ are real numbers, the latter (possibly) depending on $m$. The centered random variables $\{\varepsilon_i\}_{i\in\mathbb{N}}$ are assumed to satisfy the following conditions.

**Assumption 2.3.1 (Weak invariance)**
*The following assumptions will be used.*

*a) Let*

$$\left| \sum_{i=1}^{m} \varepsilon_i \right| = \mathcal{O}_P(\sqrt{m}) \qquad (m \to \infty) \tag{2.8}$$

*hold for* $\varepsilon_1, \ldots, \varepsilon_m$*. This condition is somewhat weaker than the asymptotic normality of the partial sums of the errors* $\{\varepsilon_i\}_{i\in\mathbb{N}}$*, but sufficient for our purposes.*

b) *There is a sequence of Wiener processes* $\{W_m(t) : t \geq 0\}_{m\in\mathbb{N}}$ *and a positive constant* $\sigma > 0$ *such that*

$$\sup_{\frac{1}{m}\leq t<\infty} \frac{1}{(mt)^{\frac{1}{\nu}}} \left| \sum_{i=m+1}^{m+mt} \varepsilon_i - \sigma W_m(mt) \right| = \mathcal{O}_P(1) \qquad (m \to \infty) \tag{2.9}$$

*with some* $\nu > 2$*. That is, the sequence* $\{\varepsilon_i\}_{i\geq m+1}$ *satisfies a weak invariance principle with a given rate.*

Moreover, we assume that there is no change in the mean in the "training period" of size $m$. Therefore, we have from (2.7) that

$$EX_1 = \ldots = EX_m = \mu. \tag{2.10}$$

In this setting, we wish to test between the null hypothesis

$$\mathcal{H}_{\mathbf{0}} : \ \Delta_m = 0$$

and the alternative of a change in the mean, i.e.,

$$\mathcal{H}_{\mathbf{A}} : \ \Delta_m \neq 0.$$

Our test statistic will be based on the CUSUM detector

$$Q(m,k) = \sum_{i=m+1}^{m+k} X_i - \frac{k}{m}\sum_{i=1}^{m} X_i. \tag{2.11}$$

Similarly to the tests in Section 2.2, the procedure stops and we state a mean change in $m+k$ if $Q(m,k)$ hits or crosses the value of a boundary function $g(m,k)$ for the first time. That is, we are waiting for the time–point

$$\tau_m = \min\{\, k \geq 1 : \ |Q(m,k)| \geq g(m,k)\,\} \tag{2.12}$$

with the understanding $\min\emptyset = \infty$. This leads to the natural interpretation of infinitely often repeated observations corresponding to our assumption that data is cheaply available. Mathematically spoken, $\tau_m = \infty$ if $|Q(m,k)| < g(m,k)$ for all $k \geq 1$.

So far, no conditions were imposed on the boundary function in (2.12). The following example teaches us that $g$ cannot be arbitrary, but has to be carefully chosen indeed.

**Example 2.3.1 (Robbins)**
Let $\{Y_i\}_{i\in\mathbb{N}}$ be a sequence of independent random variables with a common variance $\operatorname{Var} Y_i = \sigma^2 > 0$ $(i \in \mathbb{N})$. Suppose, one is interested in testing the hypotheses

$$\mathcal{H}_{\mathbf{0}}: \; EY_i = 0 \qquad (i \in \mathbb{N})$$

and

$$\mathcal{H}_{\mathbf{A}}: \; \text{Not } \mathcal{H}_{\mathbf{0}}.$$

Then (cf. Ploberger, Krämer and Kontrus (1989)), a suitable one–shot test statistic is given by

$$T_{n,3} = \frac{1}{\sqrt{n}\sigma} \max_{k=1,\ldots,n} |S_k|,$$

where $S_k = Y_1 + \ldots + Y_k$ for $k = 1, \ldots, n$.

If the partial sums $S_k$ satisfy a (weak) invariance principle, the critical value $c$ can be calculated from the hitting time of the approximating Wiener process. This leads to the stopping rule

$$\tau_n' = \min\{\, n \geq 1 : \; |S_n| \geq c_n \,\},$$

where $c_n = \sqrt{n}\sigma c$. Now, the law of the iterated logarithm for Wiener processes gives that

$$P\{\, S_n \in [-c_n, c_n] \text{ for all } n \geq 1 \,\} = 0$$

even under $\mathcal{H}_{\mathbf{0}}$, so that we get

$$\lim_{n\to\infty} P_{\mathcal{H}_{\mathbf{0}}}\{\tau_n' < \infty\} = 1,$$

and we eventually cause false alarm. For additional information confer the paper of Robbins (1970).

Example 2.3.1 shows that there is a need for a more sophisticated approach to choose a boundary function $g$ in (2.12). Making use of the invariance in (2.9), it turns out that

$$g(m,k) = c\sqrt{m}\left(1 + \frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}, \tag{2.13}$$

where $\gamma \in [0, \frac{1}{2})$, is a good choice for deriving precise asymptotics. Moreover, it is in accordance with the results of Chu, Stinchcombe and White (1996) and has also been used by Horváth, Hušková, Kokoszka and Steinebach (2003) to examine a linear regression model (cf. Section 2.2). Therein, the constant parameter $c = c(\alpha)$ is prescribed by limiting the error of first kind to a given level $\alpha \in (0,1)$, i.e.,

$$\lim_{m\to\infty} P_{\mathcal{H}_\mathbf{0}}\{\tau_m < \infty\} = \alpha.$$

In case $\Delta_m \neq 0$ is a constant, Horváth,Hušková, Kokoszka and Steinebach (2003) proved that the asymptotic power is one, that is

$$\lim_{m\to\infty} P_{\mathcal{H}_\mathbf{A}}\{\tau_m < \infty\} = 1.$$

The parameter $\gamma$ determines the sensitivity of the test and has to be fixed by the statistician in advance. Horváth, Hušková, Kokoszka and Steinebach (2003) showed in a simulation study, that in their linear regression model detectors with $\gamma$ close to $\frac{1}{2}$ have the shortest delay time if $k^*$ is small compared to $m$, that is a change occurs shortly after the beginning of monitoring. We will give a formal proof of this observation in Section 2.6 based on our model. (Else smaller choices of $\gamma$ seemed to be more advisable.)

In Section 2.4 we will describe some settings, that can be treated by our model. Section 2.5 contains the behaviour of our test statistic under both the null hypothesis and the alternative. Section 2.6 is devoted to determining the limit distribution of the stopping rule $\tau_m$.

## 2.4 Examples

This section contains several applications, which are included in the framework presented in the preceding section. We will discuss independent, identically distributed random variables, linear processes or moving average time series of infinite order and GARCH processes. Two final examples point out that the approximations in (2.8) and (2.9) can be obtained under rather general conditions such as strong mixing and as well for martingale differences.

**Example 2.4.1 (Independent, identically distributed random variables)**
Let the sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ in the definition of the $\{X_i\}_{i\in\mathbb{N}}$ in (2.7) be independent and identically distributed with the further assumptions

a) $E\varepsilon_1 = 0$,

b) $E\varepsilon_1^2 = \sigma^2 > 0$,

c) $E|\varepsilon_1|^\nu < \infty$ with some $\nu > 2$.

In this setting, the seminal approximations of Komlós, Major and Tusnády (1975,1976), and Major (1976a) immediately imply (2.8) and (2.9).

Now, independence and identical distribution of the underlying random variables are very restrictive assumptions. In a great variety of applications therefore models allowing also dependence structures are of importance. One crucial example in time series analysis is the moving average process of infinite order (shortly MA($\infty$)) or linear process. It will be explained next.

**Example 2.4.2 (Linear processes)**
Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ fulfill the equations

$$\varepsilon_i = \sum_{j=1}^{\infty} c_j \delta_{j-i} \qquad (i \in \mathbb{N}),$$

where $\{\delta_i\}_{i\in\mathbb{N}_0}$ are independent and identically distributed random variables with

a) $E\delta_0 = 0$,
b) $E\delta_0^2 = \sigma^2 > 0$,
c) $E|\delta_0|^\nu < \infty$ with some $\nu > 2$.

The sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ is called a linear process.
Horváth (1997) showed that (2.8) and (2.9) are satisfied if $\delta_0$ has a smooth density function $f$ such that

$$\sup_{-\infty<s<\infty} \frac{1}{|s|} \int_{-\infty}^{\infty} |f(t+s) - f(t)| \, dt < \infty,$$

if the condition

$$c_j = \mathcal{O}\left(j^{-\beta}\right) \qquad (j \to \infty) \tag{2.14}$$

with $\beta > \frac{3}{2}$ holds for the coefficient sequence $\{c_j\}_{j\in\mathbb{N}}$ and if moreover

$$g(z) = \sum_{j=1}^{\infty} c_j z^j \neq 0 \qquad (|z| \leq 1), \tag{2.15}$$

where $z \in \mathbb{C}$. Conditions (2.14) and (2.15) yield for example

$$\begin{aligned} E\varepsilon_i &= 0 && (i \in \mathbb{N}), \\ E\varepsilon_i^2 &= \sigma^2 \sum_{j=1}^{\infty} c_j^2 < \infty && (i \in \mathbb{N}). \end{aligned}$$

That means, expectation and variance of the sequence $\{\varepsilon_i\}_{i\in\mathbb{N}}$ do not depend on $i$.

One reason, why linear processes are of major interest for statisticians, comes from time series analysis: a sequence of random variables $\{X_n\}_{n\in\mathbb{Z}}$ is called ARMA$(p,q)$ process[1], if it is the second–order stationary solution of the set of equations

$$X_n - \sum_{j=1}^{p} a_j X_{n-j} = e_n + \sum_{j=1}^{q} b_j e_{n-j} \qquad (n \in \mathbb{Z}),$$

where $\{e_n\}_{n\in\mathbb{Z}}$ are random variables satisfying $Ee_n = 0$ and $0 < Ee_n^2 = \sigma^2 < \infty$ $(n \in \mathbb{Z})$. A sequence of this type is often called *white noise*.

Now, the following theorem, which classifies the representation of ARMA$(p,q)$ time series as linear processes, states the link to Example 2.4.2. But firstly, set

$$\begin{aligned} a(z) &= 1 - a_1 z - \ldots - a_p z^p, \\ b(z) &= 1 + b_1 z + \ldots + b_q z^q, \end{aligned}$$

where $z \in \mathbb{C}$.

**Theorem 2.4.1**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an ARMA$(p,q)$ process such that the polynomials $a(\cdot)$ and $b(\cdot)$ have no common zeroes. Then, the following statements are equivalent:*

*a) There exists a sequence $\{c_j\}_{j\in\mathbb{N}_0}$ such that*

$$X_n = \sum_{j=0}^{\infty} c_j e_{n-j} \qquad (n \in \mathbb{Z})$$

*with $\sum_{j=0}^{\infty} |c_j| < \infty$.*

*b) $a(z) \neq 0$ for all $z \in \mathbb{C}$ with $|z| \leq 1$.*

*Moreover, the coefficients in a) are determined by the complex power series*

$$c(z) = \sum_{j=0}^{\infty} c_j z^j = \frac{b(z)}{a(z)} \qquad (|z| \leq 1).$$

The property in a) is called causality. A proof of Theorem 2.4.1 can, for instance, be found in Brockwell and Davis (1991), Theorem 3.1.2.

The ARMA concept can be further generalized. In finance, many time series exhibit non–stationary behaviour and/or heteroskedasticity. Engle (1982), who won the Nobel prize in 2003, and Bollerslev (1986) introduced the so–called ARCH and GARCH[2] models,

[1] ARMA is the acronym for *autoregressive moving average*.
[2] GARCH is the acronym for *generalized autoregressive conditionally heteroskedastic*.

which allow for time dependent conditional variances of the random variables in consideration, while the variance itself remains constant. These models were successfully applied to various economic data, for instance to the inflation rate in the United Kingdom (cf. Engle (1982)) or to stock prices in financial markets. Today, volatility models are indispensable for financial analysts. The following example shows, that GARCH (and therefore ARCH) processes are included in our framework under certain additional assumptions.

**Example 2.4.3 (GARCH processes)**
Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ be a sequence of random variables which satisfy the following two sets of equations:

$$\begin{aligned} \varepsilon_i &= \sigma_i \delta_i && (i \in \mathbb{Z}), \\ \sigma_i^2 &= \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{i-j}^2 + \sum_{j=1}^{q} \beta_j \sigma_{i-j}^2 && (i \in \mathbb{Z}), \end{aligned}$$

where $\{\delta_i\}_{i\in\mathbb{Z}}$ is a further sequence of random variables with finite second moments, and where

$$\omega > 0, \qquad \alpha_1, \ldots, \alpha_p \geq 0, \qquad \beta_1, \ldots, \beta_q \geq 0.$$

If $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is a second–order stationary and ergodic solution of these equations, it is called a GARCH$(p, q)$ process. In the special case $\beta_1 = \ldots = \beta_q = 0$, the solution is called ARCH$(p)$ process. If furthermore $\alpha_1 = \ldots = \alpha_p = 0$, then $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is simply a white noise sequence.

Nelson (1990) derived necessary and sufficient conditions for the existence of a unique second–order stationary solution if $p = q = 1$. The general case has been treated by Berkes, Horváth and Kokoszka (2003a). Moreover, it has been shown that $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is strong mixing under additional assumptions (cf. Bougerol and Picard (1992a,1992b)). Eventually, Carrasco and Chen (2002) showed that also a large class of GARCH type processes is strong mixing under suitable regularity conditions.

The next example illustrates, that Examples 2.4.2 and 2.4.3 are special cases of second–order stationary and mixing sequences and can therefore be included in the following approach.

**Example 2.4.4 (Mixing sequences)**
Let $\{\varepsilon_i\}_{i\in\mathbb{N}}$ be a second–order stationary, mixing sequence of random variables. Then, (2.8) and (2.9) can be verified under regularity conditions given in Philipp (1986) and Shao (1993).

Finally, we introduce an approach built upon martingales and filtrations. A filtration is a collection of $\sigma$–fields $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$ which satisfy $\mathcal{F}_i \subset \mathcal{F}_j$ for all $i \leq j$.

**Example 2.4.5 (Martingale differences)**
A sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is called a martingale difference sequence with respect to some filtration $\{\mathcal{F}_i\}_{i\in\mathbb{Z}}$ if

$$E(\varepsilon_i|\mathcal{F}_{i-1}) = 0 \qquad \text{a.s.}$$

for all $i \in \mathbb{Z}$. Then, the approximations in Philipp and Stout (1986) and Eberlein (1986) can be used to establish (2.8) and (2.9).

Even so–called RCA (random coefficient autoregressive) time series exhibit a dependence structure that can be covered by the strong invariance principles provided in Eberlein (1986). Details will be discussed in Chapter 3.

## 2.5 Asymptotics of the test procedure

In this section, we state the limit theorems for the test procedure both under the null hypothesis $\mathcal{H}_\mathbf{0}$ and the alternative $\mathcal{H}_\mathbf{A}$. The latter can be further distinguished. We say, that there is a

- *fixed change* if $\Delta_m = \Delta \neq 0$ is constant and therefore independent of $m$,
- *local change* if $\Delta_m \to 0$ as $m \to \infty$.

Not surprisingly, to obtain asymptotics under $\mathcal{H}_\mathbf{A}$ we need different assumptions according to the distinction above.

### 2.5.1 Asymptotics under the null hypothesis

At first, we give the limit theorem for $Q(m,k)$ under $\mathcal{H}_\mathbf{0}$. We shall use the invariance stated in Assumption 2.3.1 to obtain a supremum of a Wiener process as limiting random variable. Details are addressed in the following theorem.

**Theorem 2.5.1 (Asymptotic under $\mathcal{H}_\mathbf{0}$)**
*Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables according to (2.7) such that (2.8)–(2.10) hold. Then, under $\mathcal{H}_\mathbf{0}$,*

$$\lim_{m\to\infty} P\left\{\frac{1}{\hat{\sigma}_m}\sup_{k\geq 1}\frac{|Q(m,k)|}{g(m,k)} \leq 1\right\} = P\left\{\sup_{0\leq t\leq 1}\frac{|W(t)|}{t^\gamma} \leq c\right\},$$

| $\gamma \backslash \alpha$ | 0.010 | 0.025 | 0.050 | 0.100 | 0.250 |
|---|---|---|---|---|---|
| 0.00 | 2.7912 | 2.4948 | 2.2365 | 1.9497 | 1.5213 |
| 0.15 | 2.8516 | 2.5475 | 2.2996 | 2.0273 | 1.6126 |
| 0.25 | 2.9445 | 2.6396 | 2.3860 | 2.1060 | 1.7039 |
| 0.35 | 3.0475 | 2.7394 | 2.5050 | 2.2433 | 1.8467 |
| 0.45 | 3.3015 | 3.0144 | 2.7992 | 2.5437 | 2.1729 |
| 0.49 | 3.5705 | 3.2944 | 3.0722 | 2.8259 | 2.4487 |

Table 2.1: Selected critical values $c_\alpha(\gamma)$.

*where* $\{W(t) : 0 \leq t \leq 1\}$ *denotes a Wiener process and* $\hat{\sigma}_m^2$ *is a consistent variance estimator.*

**Proof:** It is an adaptation of the proof of Theorem 2.1 in Horváth, Hušková, Kokoszka and Steinebach (2003). □

We close this subsection with two final remarks. Firstly, the limiting distribution of $X_\gamma = \{t^{-\gamma}|W(t)| : 0 \leq t \leq 1\}$ is only known for $\gamma = 0$. Else, one has to rely on simulation results to obtain critical values. For a few commonly used $\alpha$, we will list some selected critical values $c_\alpha(\gamma)$ given in Horváth, Hušková, Kokoszka and Steinebach (2003). They were gained by 50,000 repetitions of $X\gamma$, where the Wiener process $\{W(t) : 0 \leq t \leq 1\}$ was approximated on a grid of $10,000$ equi–spaced points in the interval $[0, 1]$. Secondly, as the proof of Theorem 2.5.1 shows, a suitable estimator $\hat{\sigma}_m^2$ must satisfy the condition

$$\left| \frac{1}{\hat{\sigma}_m^2} - \frac{1}{\sigma^2} \right| = o_P(1) \qquad (m \to \infty),$$

which is true for instance if the estimator is consistent.

### 2.5.2 Asymptotics under the alternative

As already mentioned above, we will differ between fixed and local changes, resulting in two limit statements in Theorems 2.5.2 and 2.5.3. It turns out that in both cases the test procedure converges in probability to infinity if the null hypothesis is violated, but different assumptions are required.

We start with the fixed change.

**Theorem 2.5.2 (Asymptotic under $\mathcal{H}_{\mathbf{A}}$, fixed change)**
*Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables according to (2.7) such that (2.8)–(2.10) hold. Let $\Delta_m = \Delta$ be constant and $k^* = o(m)$ as $m \to \infty$. Then, under $\mathcal{H}_{\mathbf{A}}$,*

$$\frac{1}{\hat{\sigma}_m} \sup_{k\geq 1} \frac{|Q(m,k)|}{g(m,k)} \xrightarrow{P} \infty$$

*as $m \to \infty$, where $\hat{\sigma}_m^2$ is a consistent variance estimator.*

**Proof:** Set $\tilde{k} = m + k^*$. Then

$$\begin{aligned} Q(m,\tilde{k}) &= \sum_{i=m+1}^{m+\tilde{k}} X_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} X_i \\ &= \sum_{i=m+1}^{m+\tilde{k}} \varepsilon_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} \varepsilon_i + \Delta_m\left(\tilde{k} - k^* + 1\right), \end{aligned}$$

where $\Delta_m \neq 0$ under $\mathcal{H}_{\mathbf{A}}$. By Theorem 2.5.1, we have

$$\frac{1}{g(m,\tilde{k})}\left(\sum_{i=m+1}^{m+\tilde{k}} \varepsilon_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} \varepsilon_i\right) = \mathcal{O}_P(1) \qquad (m \to \infty).$$

It remains to analyze the deterministic part. Since $k^* = o(m)$, we get

$$2^{1-\gamma} \leftarrow \left(1 + \frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma} = \left(2 + \frac{k^*}{m}\right)\left(\frac{1+\frac{k^*}{m}}{2+\frac{k^*}{m}}\right)^{\gamma} \leq 3 \cdot 2^{-\gamma}.$$

Combining this result with the assumption that $\Delta_m = \Delta$ is constant, yields

$$\begin{aligned} &\left|\frac{|\Delta_m|(m+1)}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} - \frac{\sqrt{m}|\Delta|}{2^{1-\gamma}}\right| \\ &\leq \frac{|\Delta_m|}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} + \sqrt{m}|\Delta|\left|\frac{2^{1-\gamma} - \left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}}{2^{1-\gamma}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}}\right| \end{aligned}$$

as $m \to \infty$. Hence,

$$\frac{|\Delta_m|(m+1)}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} = \frac{\sqrt{m}|\Delta|}{2^{1-\gamma}}\left(1+o(1)\right) \qquad (m \to \infty).$$

This completes the proof. □

Analogously, we can prove the next theorem for a local change in the mean of the underlying random variables. Since the arguments in the proof only slightly differ from those of the previous theorem, we just sketch the main parts.

**Theorem 2.5.3 (Asymptotic under $\mathcal{H}_{\mathbf{A}}$, local change)**
*Let $\{X_i\}_{i\in\mathbb{N}}$ be a sequence of random variables according to (2.7) such that (2.8)–(2.10) hold. Let $\Delta_m \to 0$, but $\sqrt{m}|\Delta_m| \to \infty$ and let $k^* = [\beta m]$, $\beta > 0$ fixed, as $m \to \infty$. Then, under $\mathcal{H}_{\mathbf{A}}$,*

$$\frac{1}{\hat{\sigma}_m} \sup_{k\geq 1} \frac{|Q(m,k)|}{g(m,k)} \xrightarrow{P} \infty$$

*as $m \to \infty$, where $\hat{\sigma}_m^2$ is a consistent variance estimator.*

**Proof:** Set $\tilde{k} = m + k^*$. By assumption on $k^*$ we get $k^* m^{-1} = [\beta m]m^{-1} \to \beta > 0$ as $m \to \infty$. Therefore,

$$\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma} = \left(2+\frac{k^*}{m}\right)\left(\frac{1+\frac{k^*}{m}}{2+\frac{k^*}{m}}\right)^{\gamma} \longrightarrow (2+\beta)^{1-\gamma}(1+\beta)^{\gamma}$$

as $m \to \infty$, resulting in

$$\begin{aligned}
&\left|\frac{|\Delta_m|(m+1)}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} - \frac{\sqrt{m}|\Delta_m|}{(2+\beta)^{1-\gamma}(1+\beta)^{\gamma}}\right| \\
&\leq \quad \frac{|\Delta_m|}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} + \sqrt{m}|\Delta_m| \left|\frac{(2+\beta)^{1-\gamma}(1+\beta)^{\gamma} - \left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}}{\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}(2+\beta)^{1-\gamma}(1+\beta)^{\gamma}}\right|
\end{aligned}$$

as $m \to \infty$. So,

$$\frac{|\Delta_m|(m+1)}{\sqrt{m}\left(1+\frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^{\gamma}} = \frac{\sqrt{m}|\Delta_m|}{(2+\beta)^{1-\gamma}(1+\beta)^{\gamma}}(1+o(1)) \qquad (m \to \infty).$$

Since $\sqrt{m}|\Delta_m| \to \infty$ by assumption, the assertion of Theorem 2.5.3 follows if we treat the random part as in the proof of Theorem 2.5.2. □

## 2.6 Estimation of delay time

We now move on to the main section of this chapter. It deals with determining the limit distribution of the stopping rule $\tau_m$, which can be interpreted as the delay time of the sequential test procedure. Therefore, additional assumptions on $\Delta_m$, which are commonly used in change–point analysis, and a suitable normalization of $\tau_m$ are needed.

We start with a set of conditions which are imposed on $\Delta_m$ and $k^*$. We assume that the change is relatively small and does not occur too late in the sample. Moreover, we assume without loss of generality that $\Delta_m > 0$ throughout this section.

**Assumption 2.6.1**
*For $m \to \infty$, let the following three conditions hold:*

$$\Delta_m \longrightarrow 0, \tag{2.16}$$

$$\sqrt{m}\Delta_m \longrightarrow \infty, \tag{2.17}$$

$$k^* = \mathcal{O}\left(m^\theta\right) \qquad \text{with some } 0 \le \theta < \left(\frac{\frac{1}{2}-\gamma}{1-\gamma}\right)^2, \tag{2.18}$$

*where $\gamma \in [0, \frac{1}{2})$ according to (2.13).*

Assumption 2.6.1 puts us in a position to formulate the main theorem.

**Theorem 2.6.1 (Limit distribution of the stopping time)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables according to (2.7) such that (2.8)–(2.9) and (2.16)–(2.18) are satisfied. Then, under $\mathcal{H}_\mathbf{A}$,*

$$\lim_{m\to\infty} P\left\{\frac{\tau_m - a_m}{b_m} \le x\right\} = \Phi(x),$$

*where $\Phi$ denotes the standard normal distribution function. Moreover, for $m \in \mathbb{N}$,*

$$\begin{aligned} a_m &= \left(\frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m}\right)^{\frac{1}{1-\gamma}}, \\ b_m &= \frac{\sqrt{a_m}\sigma}{(1-\gamma)\Delta_m}. \end{aligned}$$

The proof of Theorem 2.6.1 is divided into several auxiliary lemmas. Before we start, we give one short remark which clarifies a statement from Section 2.3.

**Remark 2.6.1**
*Let the assumptions of Theorem 2.6.1 be satisfied. Then $\tau_m$ is small if $\gamma$ is close to $\frac{1}{2}$.*

**Proof:** Theorem 2.6.1 implies that

$$\tau_m \approx a_m = \left(\frac{c}{\Delta_m}\right)^{\frac{1}{1-\gamma}} m^{\frac{\frac{1}{2}-\gamma}{1-\gamma}}.$$

Both factors of the product on the right hand side get smaller if $\gamma$ is chosen to be closer to $\frac{1}{2}$. □

The main idea in the proof of Theorem 2.6.1 is the following observation:

$$P\{\tau_m > N\} = P\left\{\max_{1\le k\le N} \frac{|Q(m,k)|}{g(m,k)} \le 1\right\}, \tag{2.19}$$

which is easily obtained from the definition of the stopping rule $\tau_m$ in (2.12). Equation (2.19) suggests that we have to find a normalizing sequence $N = N(m, x)$ depending on $m$ and $x \in \mathbb{R}$ such that the probabilities $P\{\tau_m > N(m, x)\}$ converge to the corresponding value $\Phi(x)$ for all $x$ as $m \to \infty$ by using the symmetry $\Phi(x) = 1 - \Phi(-x)$ of the standard normal distribution function and its continuity.

Instead of giving a constructive proof, we will now define $N$ and then show that this $N$ is the correct choice. Set

$$N = N(m, x) = \left( \frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m} - \sigma x \left( \frac{c^{\frac{1}{2}-\gamma} m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}} \right)^{\frac{1}{1-\gamma}} \right)^{\frac{1}{1-\gamma}}, \tag{2.20}$$

where $x \in \mathbb{R}$.

Next, we will state a technical lemma which contains some easy, but very helpful results. They explain the connections between the involved model parameters $m$, $k^*$ and $\Delta_m$, and $N$ as defined in (2.20) above.

**Lemma 2.6.1**
*Let $\gamma \in [0, \frac{1}{2})$ and let (2.16)–(2.18) be satisfied. Then,*

*a)* (i) $\dfrac{N}{m} \longrightarrow 0$, (ii) $\sqrt{N}\Delta_m \longrightarrow \infty$,

(iii) $\dfrac{k^*}{N} \longrightarrow 0$, (iv) $\dfrac{k^*}{m} \longrightarrow 0$

*for all $x \in \mathbb{R}$ as $m \to \infty$, where $N$ is defined in (2.20).*

*b) Furthermore,*

$$\lim_{m\to\infty} \frac{1}{\sigma} \left( \frac{N}{m} \right)^{\gamma - \frac{1}{2}} \left( c - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^\gamma} \right) = x,$$

*where $N$ is defined in (2.20).*

**Proof:** Firstly, we show that the first part of the difference in (2.20) is dominating the second one, i.e.,

$$N \approx \left( \frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m} \right)^{\frac{1}{1-\gamma}} = a_m \tag{2.21}$$

for large $m$. It holds

$$\begin{aligned}
& \frac{m^{\frac{1}{2}-\gamma}}{\Delta_m}\left(\frac{m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{-\frac{1}{1-\gamma}} \\
= & \left(\frac{m^{(\frac{1}{2}-\gamma)(1-\gamma)}}{\Delta_m^{1-\gamma}}\frac{\Delta_m^{\frac{3}{2}-2\gamma}}{m^{(\frac{1}{2}-\gamma)^2}}\right)^{\frac{1}{1-\gamma}} \\
= & \left(m^{(\frac{1}{2}-\gamma)(1-\gamma)-(\frac{1}{2}-\gamma)^2}\Delta_m^{\frac{1}{2}-\gamma}\right)^{\frac{1}{1-\gamma}} \\
= & \left(\sqrt{m}\Delta_m\right)^{\frac{\frac{1}{2}-\gamma}{1-\gamma}} \longrightarrow \infty
\end{aligned}$$

as $m \to \infty$, by condition (2.17) and $\gamma \in [0, \frac{1}{2})$.

a) (i) For large enough $m$ we can write

$$\begin{aligned}
\frac{N}{m} & \leq (1+o(1))\left(\frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m}m^{\gamma-1}\right)^{\frac{1}{1-\gamma}} \\
& \approx \left(\frac{c}{\sqrt{m}\Delta_m}\right)^{\frac{1}{1-\gamma}} \longrightarrow 0
\end{aligned}$$

as $m \to \infty$, using (2.17) again. Thus, (i) is proved.

(ii) By (2.21),

$$\begin{aligned}
\sqrt{N}\Delta_m & \approx \Delta_m\left(\frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m}\right)^{\frac{1}{2(1-\gamma)}} \\
& = c^{\frac{1}{2}(1-\gamma)}\left(\sqrt{m}\Delta_m\right)^{\frac{\frac{1}{2}-\gamma}{1-\gamma}} \longrightarrow \infty
\end{aligned}$$

as $m \to \infty$ by (2.17), finishing the proof of (ii).

(iii) Applying assumption (2.18) and the approximation of (2.21), we obtain

$$\frac{k^*}{N} = \mathcal{O}\left(\Delta_m^{\frac{1}{1-\gamma}} m^{\left(\frac{\frac{1}{2}-\gamma}{1-\gamma}\right)^2 - \frac{\frac{1}{2}-\gamma}{1-\gamma}}\right) = o(1) \qquad (m \to \infty),$$

since

$$\left(\frac{\frac{1}{2}-\gamma}{1-\gamma}\right)^2 - \frac{\frac{1}{2}-\gamma}{1-\gamma} < 0,$$

completing the proof of (iii).

(iv) From (i) and (iii) we have
$$\frac{k^*}{m} = \frac{k^*}{N}\frac{N}{m} \longrightarrow 0$$
as $m \to \infty$.

b) The relation is a consequence of the definition of $N$ in (2.20). It holds

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c - \frac{\Delta_m}{m^{\frac{1}{2}-\gamma}}N^{1-\gamma}\right) \\
&= \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c - \frac{\Delta_m}{m^{\frac{1}{2}-\gamma}}\left(\frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m} - \sigma x\left(\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}}\right)\right) \\
&= \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\sigma x\frac{\Delta_m}{m^{\frac{1}{2}-\gamma}}\left(\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}} \\
&= \sigma x N^{\gamma-\frac{1}{2}}\Delta_m\left(\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}}.
\end{aligned}$$

By (2.21),

$$\begin{aligned}
&N^{\gamma-\frac{1}{2}}\Delta_m\left(\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}} \\
&\approx \left(\Delta_m^{1-\gamma}\frac{c^{\gamma-\frac{1}{2}}m^{(\gamma-\frac{1}{2})(\frac{1}{2}-\gamma)}}{\Delta_m^{\gamma-\frac{1}{2}}}\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}} \\
&= 1,
\end{aligned}$$

giving the desired result. □

With the technical support of Lemma 2.6.1, we are able to proceed with the proof of Theorem 2.6.1 by showing that the observations before $k^*$ do not contribute to the asymptotic.

**Lemma 2.6.2**
*Let $\gamma \in [0, \frac{1}{2})$. If (2.8)–(2.10) and (2.16)–(2.18) hold, then*

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\max_{1\le k<k^*}\frac{|Q(m,k)|}{g_1(m,k)} - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^{\gamma}}\right) \xrightarrow{P} -\infty$$

*as $m \to \infty$, where $g_1 = c^{-1}g$.*

**Proof:** From the definition of $Q(m,k)$ in (2.11) we obtain the following representation:

$$Q(m,k) = \sum_{i=m+1}^{m+k} \varepsilon_i - \frac{k}{m}\sum_{i=1}^{m} \varepsilon_i + \Delta_m(k-k^*+1)I_{\{k\geq k^*\}}. \tag{2.22}$$

We will work on the right hand side of (2.22) term by term.

a) The first term of equation (2.22) can be approximated by a Wiener process according to assumption (2.9). This yields

$$\begin{aligned}
&\max_{1\leq k<k^*} \frac{1}{g_1(m,k)}\left|\sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_m(k)\right| \\
&= \mathcal{O}_P(1) \max_{1\leq k<k^*} \frac{k^{\frac{1}{\nu}}}{\sqrt{m}(\frac{k}{m})^{\gamma}} \\
&= \mathcal{O}_P\left(\left(\frac{k^*}{m}\right)^{\frac{1}{2}-\gamma}\right) = o_P(1)
\end{aligned}$$

as $m\to\infty$ by Lemma 2.6.1a)(iv). So, it suffices to consider the behaviour of $W_m(k)$. We note that

$$\max_{1\leq k<k^*} \frac{|W_m(k)|}{g_1(m,k)} = \mathcal{O}_P(1) \sup_{0<t\leq k^*} \frac{|W_m(t)|}{\sqrt{m}(\frac{t}{m})^{\gamma}}$$

by using Lemma 2.6.1a)(iv).

Applying the scale transformation of Wiener processes, we arrive at

$$\sup_{0<t\leq k^*} \frac{|W_m(t)|}{\sqrt{m}(\frac{t}{m})^{\gamma}} \stackrel{\mathcal{D}}{=} \sup_{0<t\leq \frac{k^*}{m}} \frac{|W(t)|}{t^{\gamma}},$$

where $\{W(t): t\geq 0\}$ denotes another Wiener process. Now, its a.s. continuity also implies the a.s. continuity of $\{t^{-\gamma}W(t): t\geq 0\}$ (recall that $\gamma<\frac{1}{2}$). Combining this property with Lemma 2.6.1a)(iv), we get

$$\lim_{m\to\infty} \sup_{0<t\leq \frac{k^*}{m}} \frac{|W(t)|}{t^{\gamma}} = 0 \qquad \text{a.s.}$$

Thus,

$$\max_{1\leq k<k^*} \frac{1}{g_1(m,k)}\left|\sum_{i=m+1}^{m+k} \varepsilon_i\right| = o_P(1)$$

as $m\to\infty$, i.e., the first term is negligible.

b) The second term of the right hand side in (2.22) contains only the errors $\varepsilon_1, \dots, \varepsilon_m$. So, we can apply assumption (2.8) and get

$$\max_{1 \le k < k^*} \frac{1}{g_1(m,k)} \left| \frac{k}{m} \sum_{i=1}^{m} \varepsilon_i \right| = \mathcal{O}_P \left( \left( \frac{k^*}{m} \right)^{1-\gamma} \right) = o_P(1)$$

as $m \to \infty$ using Lemma 2.6.1a)(iv) once again.

c) Finally, we study the deterministic third part of equation (2.22). On combining Lemma 2.6.1a)(iv) and part b) of this proof, we arrive at

$$\lim_{m \to \infty} \frac{\Delta_m N}{\sqrt{m} (\frac{N}{m})^\gamma} = c > 0.$$

Putting together a)–c), Lemma 2.6.2 is proved, since

$$\left( \frac{N}{m} \right)^{\gamma - \frac{1}{2}} \longrightarrow \infty \qquad (m \to \infty)$$

by Lemma 2.6.1a)(i). □

The next aim is to show that the asymptotic behaviour of the maximum taken over those $k$ ranging from $k^*$ to $N$ can be appropriately approximated with the corresponding maximum of the Wiener process $\{W_m(t) : t \ge 0\}$ plus an additional drift term traced back to the change in the mean under $\mathcal{H}_{\mathbf{A}}$. Again, we will strongly rely on the invariance assumed in (2.9). The following lemma contains the exact formulation.

**Lemma 2.6.3**
*Let $\gamma \in [0, \frac{1}{2})$. If (2.8), (2.9) and (2.16)–(2.18) hold, then*

$$\left( \frac{N}{m} \right)^{\gamma - \frac{1}{2}} \max_{k^* \le k \le N} \frac{1}{g_1(m,k)} |Q(m,k) - (\sigma W_m(k) + \Delta_m k)| = o_P(1)$$

*as $m \to \infty$, where $g_1 = c^{-1} g$.*

**Proof:** The proof is given in three steps.

a) First of all, we have

$$\left( \frac{N}{m} \right)^{\gamma - \frac{1}{2}} \max_{k^* \le k \le N} \frac{1}{g_1(m,k)} \left| \frac{k}{m} \sum_{i=1}^{m} \varepsilon_i \right| = \mathcal{O}_P \left( \sqrt{\frac{N}{m}} \right) = o_P(1)$$

as $m \to \infty$, following from (2.8) and Lemma 2.6.1a)(i).

b) Secondly, the deterministic part is negligible as well, since

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^*\le k\le N} \frac{\Delta_m(k-1)}{g_1(m,k)} = \mathcal{O}(1)\left(\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{\Delta_m k^*}{\sqrt{m}}\left(\frac{m}{k^*}\right)^{\gamma}\right) \tag{2.23}$$

as $m \to \infty$. The order can be further estimated by using the dominating term of $N$ as in (2.21) and assumption (2.18) on the change-point. Thus, we can continue with (2.23) by writing

$$\mathcal{O}(1)\left(\Delta_m\left(\frac{m^{\frac{1}{2}-\gamma}}{\Delta_m}\right)^{\frac{\gamma-\frac{1}{2}}{1-\gamma}} k^{*1-\gamma}\right) = \mathcal{O}(1)\left(\Delta_m^{1+\frac{\frac{1}{2}-\gamma}{1-\gamma}} m^{-\frac{(\frac{1}{2}-\gamma)^2}{1-\gamma}+\theta(1-\gamma)}\right) = o(1)$$

as $m \to \infty$, since

$$1+\frac{\frac{1}{2}-\gamma}{1-\gamma} > 0 \qquad \text{and} \qquad -\frac{(\frac{1}{2}-\gamma)^2}{1-\gamma}+\theta(1-\gamma) < 0,$$

the last inequality following from (2.18).

c) Finally, an application of (2.9) yields

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^*\le k\le N} \frac{1}{g_1(m,k)}\left|\sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_m(k)\right| \\
&= \mathcal{O}_P(1)\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^*\le k\le N} \frac{k^{\frac{1}{\nu}}}{g_1(m,k)} \\
&= \mathcal{O}_P(1)\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \frac{1}{\sqrt{m}(1+\frac{N}{m})}\left(\frac{m+N}{m}\right)^{\gamma} \max_{k^*\le k\le N} \frac{k^{\frac{1}{\nu}}}{(\frac{k}{m})^{\gamma}} \\
&= \mathcal{O}_P(1) N^{\gamma-\frac{1}{2}} \max_{k^*\le k\le N} k^{\frac{1}{2}-\gamma} k^{\frac{1}{\nu}-\frac{1}{2}} \\
&= o_P\left(k^{*\frac{1}{\nu}-\frac{1}{2}}\right) \\
&= o_P(1)
\end{aligned}$$

as $m \to \infty$. Therein, Lemma 2.6.1a)(iv) has been used to obtain the third equality sign. Moreover, we mention that $\frac{1}{\nu}-\frac{1}{2}<0$.

On combining a)–c), the proof is complete. □

The previous lemma can be further refined by simplifying the boundary function $g$ (or $g_1$) through an asymptotic equivalent expression.

**Lemma 2.6.4**
*Let $\gamma \in [0, \frac{1}{2})$. If (2.8),(2.9) and (2.16)–(2.18) hold, then*

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^* \le k \le N} \left| \frac{\sigma W_m(k)}{g_1(m,k)} - \frac{\sigma W_m(k)}{\sqrt{m}(\frac{k}{m})^\gamma} \right| = o_P(1)$$

*as $m \to \infty$, where $g_1 = c^{-1} g$.*

**Proof:**

a) Set

$$a_{N,m}(\gamma) = \max\left\{ \left| \left(1 + \frac{N}{m}\right)^{\gamma-1} - 1 \right|, \left| \left(1 + \frac{k^*}{m}\right)^{\gamma-1} - 1 \right| \right\}.$$

Then $a_{N,m}(\gamma) \to 0$ $(m \to \infty)$ by Lemma 2.6.1a)(i) and (iv). Therefore, we get

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^* \le k \le N} \frac{|W_m(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} \left| \frac{(\frac{k}{m})^\gamma}{(1+\frac{k}{m})(\frac{k}{m+k})^\gamma} - 1 \right| \\
&= \mathcal{O}(1) \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^* \le k \le N} \frac{|W_m(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} \left| \left(1 + \frac{k}{m}\right)^{\gamma-1} - 1 \right| \\
&= \mathcal{O}(1) \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} a_{N,m}(\gamma) \max_{k^* \le k \le N} \frac{|W_m(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} \\
&= o(1) \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^* \le k \le N} \frac{|W_m(k)|}{\sqrt{m}(\frac{k}{m})^\gamma},
\end{aligned}$$

as $m \to \infty$ by the above named property of $a_{N,m}(\gamma)$.

b) Let $\{W(t) : t \ge 0\}$ be an arbitrary Wiener process. Then obviously

$$\max_{k^* \le k \le N} \frac{|W(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} \le \sup_{0<t\le N} \frac{|W(t)|}{\sqrt{m}(\frac{t}{m})^\gamma}.$$

Moreover, by the scale transformation of Wiener processes we get

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \sup_{0<t\le N} \frac{|W(t)|}{\sqrt{m}(\frac{t}{m})^\gamma} \stackrel{\mathcal{D}}{=} \sup_{0<t\le 1} \frac{|W(t)|}{t^\gamma}.$$

c) Now recall, that the distribution of $\{W_m(t) : t \geq 0\}$ does not depend on $m$. So, we can apply b) to achieve

$$\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \max_{k^* \leq k \leq N} \frac{|W_m(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} \left| \frac{(\frac{k}{m})^\gamma}{(1+\frac{k}{m})(\frac{k}{m+k})^\gamma} - 1 \right| = o_P(1)$$

as $m \to \infty$.

On combining a)–c) the proof of Lemma 2.6.4 is complete. □

The remaining and final auxiliary lemma contains the previously claimed convergence (in distribution) result to the normal law. To simplify notation, we introduce

$$a_m(\gamma) = \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}} \tag{2.24}$$

$$b_m(\gamma) = a_m(\gamma)\left(c - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^\gamma}\right) \tag{2.25}$$

Recall, that both expressions depend on $x \in \mathbb{R}$ via $N = N(m, x)$ and on the constant $c = c(\alpha)$ determined by the prescribed number $\alpha$, which controls the error of first type (asymptotically).

**Lemma 2.6.5**
*Let $\gamma \in [0, \frac{1}{2})$. If (2.8), (2.9) and (2.16)–(2.10) hold, then*

$$P\left\{\frac{a_m(\gamma)}{\sigma}\left(\max_{k^* \leq k \leq N} \frac{\sigma W(k) + \Delta_m k}{\sqrt{m}(\frac{k}{m})^\gamma} - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^\gamma}\right) \leq \frac{b_m(\gamma)}{\sigma}\right\} \longrightarrow \Phi(x)$$

*as $m \to \infty$, where $a_m(\gamma)$ and $b_m(\gamma)$ are defined in (2.24) and (2.25), respectively.*

**Proof:** Again, the proof is given in three steps.

a) On rephrasing step b) in the proof of Lemma 2.6.4, we see

$$\frac{|W(k)|}{\sqrt{m}(\frac{k}{m})^\gamma} = \mathcal{O}_P\left(\left(\frac{N}{m}\right)^{\frac{1}{2}-\gamma}\right) = o_P\left(\frac{\Delta_m N^{1-\gamma}}{m^{\frac{1}{2}-\gamma}}\right)$$

uniformly in $k^* \leq k \leq N$ as $m \to \infty$, since by Lemma 2.6.1a)(ii)

$$\left(\frac{N}{m}\right)^{\frac{1}{2}-\gamma}\left(\frac{\Delta_m N^{1-\gamma}}{m^{\frac{1}{2}-\gamma}}\right)^{-1} = \frac{1}{\sqrt{N}\Delta_m} = o(1)$$

as $m \to \infty$.

b) Let $\delta \in (0,1)$ be a fixed number. Then,

$$\max_{k^* \le k \le (1-\delta)N} \frac{\Delta_m k}{\sqrt{m}(\frac{k}{m})^\gamma} \le (1-\delta)^{1-\gamma} \Delta_m m^{\gamma - \frac{1}{2}} N^{1-\gamma}.$$

Now, step a) implies

$$\lim_{m\to\infty} P\left\{ \max_{k^* \le k \le N} \frac{|\sigma W(k) + \Delta_m k|}{\sqrt{m}(\frac{k}{m})^\gamma} = \max_{(1-\delta)N \le k \le N} \frac{|\sigma W(k) + \Delta_m k|}{\sqrt{m}(\frac{k}{m})^\gamma} \right\} = 1,$$

and furthermore

$$\lim_{m\to\infty} P\left\{ \max_{k^* \le k \le N} \frac{|\sigma W(k) + \Delta_m k|}{\sqrt{m}(\frac{k}{m})^\gamma} = \max_{(1-\delta)N \le k \le N} \frac{\sigma W(k) + \Delta_m k}{\sqrt{m}(\frac{k}{m})^\gamma} \right\} = 1,$$

since the positive deterministic part is the dominating term, i.e., the maximum will be reached near the right endpoint $N$ with arbitrary high probability.

c) Next, we use the scale transformation of Wiener processes to obtain

$$\left(\frac{N}{m}\right)^{\gamma - \frac{1}{2}} \max_{(1-\delta)N \le k \le N} \frac{|W(k) - W(N)|}{\sqrt{m}(\frac{k}{m})^\gamma} \stackrel{\mathcal{D}}{=} \sup_{1-\delta \le t \le 1} \frac{|W(t) - W(1)|}{t^\gamma}.$$

The almost sure continuity of $\{W(t) : t \ge 0\}$ gives

$$\lim_{\delta\to 0} \sup_{1-\delta \le t \le 1} \frac{|W(t) - W(1)|}{t^\gamma} = 0 \qquad \text{a.s.},$$

since $\sup_{1-\delta \le t \le 1} t^{-\gamma} = ((1-\delta)t)^{-\gamma} \to 1$ as $\delta \to 0$. Moreover,

$$\max_{(1-\delta)N \le k \le N} \frac{|W(N)|}{\sqrt{m}(\frac{k}{m})^\gamma} = \mathcal{O}_P\left(N^{\frac{1}{2}-\gamma} m^{\gamma - \frac{1}{2}}\right) = o_P\left(\Delta_m N^{1-\gamma} m^{\gamma - \frac{1}{2}}\right),$$

since by Lemma 2.6.1a)(ii)

$$\frac{N^{\frac{1}{2}-\gamma} m^{\gamma-\frac{1}{2}}}{\Delta_m N^{1-\gamma} m^{\gamma - \frac{1}{2}}} = \frac{1}{\sqrt{N}\Delta_m} = o(1)$$

as $m \to \infty$.

Putting together a)–c), we arrive at

$$\lim_{m\to\infty} P\left\{\max_{k^*\le k\le N} \frac{\sigma W(k)+\Delta_m k}{\sqrt{m}(\frac{k}{m})^\gamma} \le c\right\} = \lim_{m\to\infty} P\left\{\frac{a_m(\gamma)}{\sigma}\frac{\sigma W(N)}{\sqrt{m}(\frac{N}{m})^\gamma} \le \frac{b_m(\gamma)}{\sigma}\right\}.$$

Hence, the proof of Lemma 2.6.5 is complete after an application of Lemma 2.6.1b) by recognizing that $N^{-\frac{1}{2}}W(N)$ is a standard normal random variable. □

Putting together Lemmas 2.6.2 to 2.6.5, we are now in a position to give the proof of Theorem 2.6.1.

**Proof of Theorem 2.6.1:**

a) It only remains to reformulate the results of the auxiliary lemmas. As already mentioned in (2.19), from the definition of $\tau_m$ in (2.12) we have

$$\begin{aligned} &P\{\tau_m \ge N(m,x)\} \\ &= P\left\{\max_{1\le k\le N}\frac{|Q(m,k)|}{g_1(m,k)} \le c\right\} \\ &= P\left\{\left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(\max_{1\le k\le N}\frac{|Q(m,k)|}{g_1(m,k)} - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^\gamma}\right)\right. \\ &\qquad\qquad \left.\le \left(\frac{N}{m}\right)^{\gamma-\frac{1}{2}}\left(c - \frac{\Delta_m}{\sqrt{m}(\frac{N}{m})^\gamma}\right)\right\}. \end{aligned}$$

In view of Lemmas 2.6.3 – 2.6.5, we obtain

$$\lim_{m\to\infty} P\left\{a_m(\gamma)\left(\max_{k^*\le k\le N}\frac{|Q(m,k)|}{g_1(m,k)} - \frac{\Delta_m N}{\sqrt{m}(\frac{N}{m})^\gamma}\right) \le b_m(\gamma)\right\} = \Phi(x),$$

where $a_m(\gamma)$ and $b_m(\gamma)$ are defined in (2.24) and (2.25), respectively.

b) Next, we will replace $N(m,x)$ by an expression containing the parameters $a_m$ and $b_m$ defined in Theorem 2.6.1. Using part a), we get

$$\begin{aligned} \Phi(x) &= 1-\Phi(-x) \\ &= 1-\lim_{m\to\infty} P\{\tau_m \ge N(m,-x)\} \\ &= 1-\lim_{m\to\infty} P\left\{\tau_m^{1-\gamma} \ge N(m,-x)^{1-\gamma}\right\} \\ &= 1-\lim_{m\to\infty} P\left\{\tau_m^{1-\gamma} \ge \frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m} + \sigma x\left(\frac{c^{\frac{1}{2}-\gamma}m^{(\frac{1}{2}-\gamma)^2}}{\Delta_m^{\frac{3}{2}-2\gamma}}\right)^{\frac{1}{1-\gamma}}\right\} \end{aligned}$$

$$
\begin{aligned}
&= 1 - \lim_{m\to\infty} P\left\{\tau_m^{1-\gamma} - a_m^{1-\gamma} \geq \sigma x a_m^{\frac{1}{2}-\gamma}\Delta_m^{-1}\right\} \\
&= \lim_{m\to\infty} P\left\{\tau_m^{1-\gamma} - a_m^{1-\gamma} \leq \sigma x a_m^{\frac{1}{2}-\gamma}\Delta_m^{-1}\right\},
\end{aligned}
$$

where we have used symmetry properties and the continuity of the standard normal distribution function.

c) Part a) and equation (2.21) yield

$$
\frac{\tau_m}{a_m} \xrightarrow{P} 1 \tag{2.26}
$$

as $m \to \infty$. Now (2.26) combined with the mean-value theorem imply

$$
\begin{aligned}
\frac{\tau_m - a_m}{b_m} &= \frac{(\tau_m^{1-\gamma})^{\frac{1}{1-\gamma}} - (a_m^{1-\gamma})^{\frac{1}{1-\gamma}}}{b_m} \\
&= \frac{1}{1-\gamma}\left(a_m^{1-\gamma}(1+o_P(1))\right)^{\frac{1}{1-\gamma}-1}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m} \\
&= \frac{1}{1-\gamma}a_m^{\gamma}(1+o_P(1))\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m}.
\end{aligned}
$$

Therefore, Slutsky's lemma implies that the random variables

$$
\frac{\tau_m - a_m}{b_m} \qquad \text{and} \qquad \frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m}
$$

have the same limit distribution. Plugging this relation into part b), we obtain

$$
\begin{aligned}
\Phi(x) &= \lim_{m\to\infty} P\left\{\tau_m^{1-\gamma} - a_m^{1-\gamma} \leq \sigma x a_m^{\frac{1}{2}-\gamma}\Delta_m^{-1}\right\} \\
&= \lim_{m\to\infty} P\left\{\frac{a_m^{\gamma}}{1-\gamma}\frac{\tau_m^{1-\gamma} - a_m^{1-\gamma}}{b_m} \leq \sigma x \frac{a_m^{\gamma}}{1-\gamma}\frac{a_m^{\frac{1}{2}-\gamma}\Delta_m^{-1}}{b_m}\right\} \\
&= \lim_{m\to\infty} P\left\{\frac{\tau_m - a_m}{b_m} \leq \sigma x \frac{a_m^{\gamma}}{1-\gamma}\frac{a_m^{\frac{1}{2}-\gamma}\Delta_m^{-1}}{\sigma a_m^{\frac{1}{2}}(1-\gamma)^{-1}\Delta_m^{-1}}\right\} \\
&= \lim_{m\to\infty} P\left\{\frac{\tau_m - a_m}{b_m} \leq x\right\}.
\end{aligned}
$$

This finishes the proof of Theorem 2.6.1. □

# Chapter 3

# Monitoring Changes in RCA(1) Time Series

In this chapter, we shall consider random coefficient autoregressive (RCA) time series, which are defined as generalizations of autoregressive time series. Their basic properties are discussed in Section 3.1.

As pointed out in the previous chapter (cf. Section 2.4), (weak or strong) invariance can be verified for a variety of dependence concepts. In Section 3.2, we will show that also in the case of RCA time series of order one a strong approximation holds true, although the random variables form (for instance) no martingale difference array. Furthermore, it is still an open question if they are strong mixing (cf. Lee (2003)). Fortunately, it turns out that a theorem of Eberlein (1986) is applicable in the given situation. Moreover, we derive a second strong approximation for a sequence applied to test for a change in one of the parameters.

In Section 3.3, we shall use the strong invariance principles to obtain the corresponding theorems from Section 2.5 and 2.6 for RCA time series of order one as well. Furthermore, in Section 3.4 we add some a–posteriori procedures for a change in the mean scenario and for a change in one of the parameters determining an RCA time series.

Some of the results of the current chapter have been submitted to the Statistics and Probability Letters (cf. Aue (2003)).

## 3.1 Introduction

Originally, random coefficient autoregressive (shortly RCA) time series have been used in the context of random perturbations of dynamical systems, but are now used in a variety of applications, as for example in finance or biology (cf. Tong (1990)). RCA time series are generalizations of autoregressive time series, since they allow for randomly disturbed coefficients as well. There are variations allowing a dependence on $p$ of the previous random variables, called RCA(p) time series, and moreover even multidimensional models (cf. Nicholls and Quinn (1982)). We shall focus our attention on RCA time series of order one here.

**Assumption 3.1.1**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series, that is a solution of the equations*

$$X_n = (\varphi + b_n)X_{n-1} + e_n \qquad (n \in \mathbb{Z}), \tag{3.1}$$

*where*

*(i)* $\begin{pmatrix} b_n \\ e_n \end{pmatrix} \overset{iid}{\sim} \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \omega^2 & 0 \\ 0 & \sigma^2 \end{pmatrix} \right),$

*(ii)* $\varphi^2 + \omega^2 < 1.$

*The sequences $\{b_n\}_{n\in\mathbb{Z}}$ and $\{e_n\}_{n\in\mathbb{Z}}$, respectively, are called (white) noise. Furthermore, define*

$$\mathcal{F}_n = \sigma\left(b_k, e_k : k \le n\right) \qquad (n \in \mathbb{Z}) \tag{3.2}$$

*as the filtration generated by the noise sequences.*

While part (i) of Assumption 3.1.1 describes the properties of and the connections between the noise sequences $\{b_n\}_{n\in\mathbb{Z}}$ and $\{e_n\}_{n\in\mathbb{Z}}$, the second condition appears less familiar. But in fact, it is a necessary and sufficient condition for the existence of an $\{\mathcal{F}_n\}$–measurable solution of the defining equations (3.1) and for its stationarity. A time series $\{Y_n\}_{n\in\mathbb{Z}}$ is weakly stationary if for all $n \in \mathbb{Z}$, the mean $EY_n$ is (the same) constant, $EY_n^2$ is finite and if there is a function $\gamma = \gamma_Y$ such that

$$\operatorname{Cov}(Y_n, Y_{n+h}) \;=\; \gamma_Y(h) \qquad (n, h \in \mathbb{Z}),$$

i.e., the covariance depends only on the difference of the time–points under consideration. A more restricted approach is the notion of strict stationarity defined via the finite dimensional distributions by the equality

$$(X_{n_1}, \ldots, X_{n_k}) \overset{\mathcal{D}}{=} (X_{n_1+h}, \ldots, X_{n_k+h}) \qquad (n_1, \ldots, n_k, h \in \mathbb{Z},\ k \in \mathbb{N}).$$

In case of an RCA(1) time series, (weak) stationarity can be characterized as follows.

**Theorem 3.1.1 (Nicholls and Quinn)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series. Then, the following statements are equivalent:*

*a) There is a unique $\{\mathcal{F}_n\}$–measurable, weakly stationary solution of (3.1).*
*b) $\varphi^2 + \omega^2 < 1$, i.e., condition (ii) holds true.*

**Proof:** See Corollary 2.3.2 in Nicholls and Quinn (1982). □

It is an obvious fact, that from (i) we immediately get the strict stationarity of $\{X_n\}_{n\in\mathbb{Z}}$. Now, the following properties are easy consequences of the previous remarks.

**Lemma 3.1.1**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i) and (ii), and let $\gamma_X$ denote its covariance function. Then,*

$$EX_n = 0, \qquad EX_n^2 = \frac{\sigma^2}{1-\varphi^2-\omega^2} \qquad \text{and} \qquad \gamma_X(n) = \frac{\varphi^n\sigma^2}{1-\varphi^2-\omega^2}$$

*for all $n \in \mathbb{Z}$.*

**Proof:** By (ii), $E(\varphi + b_1)^2 = \varphi^2 + \omega^2 < 1$. So, using Feigin and Tweedie (1985), $EX_1^2 < \infty$ and thus mean and variance exist. Since $\{X_n\}_{n\in\mathbb{Z}}$ is strictly stationary, it suffices to consider $X_1$.

a) Condition (ii) implies $|\varphi| < 1$. Hence,

$$EX_1 \;=\; E(\varphi + b_1)X_0 + Ee_1 \;=\; \varphi EX_0.$$

Substituting $EX_0 = EX_1$ gives the first claim.

b) Next, we calculate

$$EX_1^2 \;=\; E\left((\varphi + b_1)X_0 + e_1\right)^2 \;=\; \varphi^2 EX_0^2 + \omega^2 EX_0^2 + \sigma^2.$$

Substituting $EX_0^2 = EX_1^2$ and condition (ii) finish the proof.

c) Finally, we consider the covariance structure. $\gamma_X(0)$ has already been calculated in part b). Let the statement be true for $n-1 \in \mathbb{N}$. Then,

$$\gamma_X(n) = EX_{n+1}X_1 \;=\; \varphi EX_nX_1 \;=\; \varphi\gamma_X(n-1) \;=\; \frac{\varphi\varphi^{n-1}\sigma^2}{1-\varphi^2-\omega^2},$$

where we have used the iid properties of the sequences $\{b_n\}_{n\in\mathbb{Z}}$ and $\{e_n\}_{n\in\mathbb{Z}}$. □

## 3.2 A strong invariance principle

As we have seen, strong approximations have played a fundamental role in probability and statistics ever since the seminal papers of Strassen (1964) and Komlós, Major and Tusnády (1975,1976). Their results for partial sums of independent and identically distributed random variables have been extended to various dependence concepts. We are going to verify a strong invariance principle for the partial sums of an RCA(1) time series $\{X_n\}_{n\in\mathbb{Z}}$ here. Let

$$S_n(m) = X_{m+1} + \ldots + X_{m+n} \qquad (m \in \mathbb{N}_0,\, n \in \mathbb{N}), \tag{3.3}$$

where we will abbreviate $S_n(0) = S_n$. Throughout the section, we assume the underlying probability space to be rich enough such that both $\{X_n\}_{n\in\mathbb{Z}}$ and the approximating Wiener process can be defined on it. Our result will follow from a theorem of Eberlein (1986), which we state in a simpler version here, since we are only interested in real–valued RCA(1) time series (and not in vector–valued ones). Let $\|\cdot\|_1$ denote the $L_1$–norm.

**Theorem 3.2.1 (Eberlein)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of real–valued random variables such that*

- *a)* $EX_n = 0$ *for all* $n \in \mathbb{N}$,
- *b)* $\|E(S_n(m)|\mathcal{F}_m)\|_1 = \mathcal{O}(n^{\frac{1}{2}-\theta})$ *uniformly in* $m \in \mathbb{N}_0$ *for some* $\theta \in (0, \frac{1}{2})$,
- *c) there exists a constant* $\sigma_W$ *such that uniformly in* $m \in \mathbb{N}_0$,

$$ES_n^2(m) - \sigma_W \;=\; \mathcal{O}\left(n^{-\rho}\right)$$

  *as* $n \to \infty$ *for some* $\rho > 0$,
- *d) there exists a* $\gamma > 0$ *such that uniformly in* $m \in \mathbb{N}_0$,

$$\|E\left(S_n^2(m)|\mathcal{F}_m\right) - ES_n^2(m)\|_1 \;=\; \mathcal{O}\left(n^{1-\gamma}\right) \qquad \text{a.s.}$$

  *as* $n \to \infty$,
- *e) there exist a constant* $M < \infty$ *and* $\kappa > 2$, *such that* $E|X_n|^\kappa < M$ *for all* $n \in \mathbb{N}$.

*Then, there exists a Wiener process* $\{W(t) : t \geq 0\}$, *such that*

$$\sum_{n=1}^{[t]} X_n - \sigma_W W(t) \;=\; \mathcal{O}\left(t^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as* $t \to \infty$ *for some* $\nu > 2$.

**Proof:** See Theorem 1 in Eberlein (1986). □

By verifying the assumptions made in Theorem 3.2.1, we get the following strong approximation for RCA(1) time series.

**Theorem 3.2.2 (Strong invariance for RCA(1) time series)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let*

$$E|e_1|^\kappa < \infty \qquad \text{and} \qquad E|\varphi + b_1|^\kappa < 1 \tag{3.4}$$

*for some $\kappa > 2$. Then, there exists a Wiener process $\{W(t) : t \geq 0\}$ such that*

$$S_{[t]} - \sigma_S W(t) = \mathcal{O}\left(t^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as $t \to \infty$ for some $\nu > 2$, where*

$$\sigma_S^2 = \frac{\sigma^2}{1-\varphi^2-\omega^2}\frac{1+\varphi}{1-\varphi}.$$

We start with the proof of Theorem 3.2.2 and firstly obtain the order of the conditional expectation of $S_n(m)$. Here and in the sequel we shall use the following property of conditional expectations. If $\mathcal{C}_1 \subset \mathcal{C}_2$ are $\sigma$–fields, then $E(X|\mathcal{C}_1) = E(E(X|\mathcal{C}_2)|\mathcal{C}_1)$.

**Lemma 3.2.1**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i) and (ii). Then, uniformly in $m \in \mathbb{N}_0$,*

$$\|E\left(S_n(m)|\mathcal{F}_m\right)\|_1 = \mathcal{O}\left(1\right)$$

*as $n \to \infty$.*

**Proof:** Firstly,

$$\begin{aligned} E\left(S_n(m)|\mathcal{F}_m\right) &= \sum_{i=1}^{n} E\left(X_{m+i}|\mathcal{F}_m\right) \\ &= X_m \sum_{i=1}^{n} \varphi^i \\ &= \frac{\varphi X_m(1-\varphi^n)}{1-\varphi} \longrightarrow \frac{\varphi X_m}{1-\varphi} \qquad \text{a.s.} \end{aligned}$$

as $n \to \infty$, since

$$E\left(X_{m+i}|\mathcal{F}_m\right) = E(E(X_{m+i}|\mathcal{F}_{m+i-1})|\mathcal{F}_m) = \varphi E(X_{m+i-1}|\mathcal{F}_m) = \varphi^i X_m \qquad \text{a.s.}$$

for $i = 1, \dots, n$ by iteration. Therefore, we get

$$\begin{aligned}
\|E(S_n(m)|\mathcal{F}_m)\|_1 &= E(|E(S_n(m)|\mathcal{F}_m)|) \\
&= E\left|X_m \frac{\varphi(1-\varphi^n)}{1-\varphi}\right| \\
&\le |\varphi| \sum_{k=0}^{\infty} |\varphi|^k E|X_m| \\
&= \frac{|\varphi| E|X_0|}{1-|\varphi|} \\
&= \mathcal{O}\left(\frac{|\varphi|}{1-|\varphi|}\right) = \mathcal{O}(1)
\end{aligned}$$

as $n \to \infty$ uniformly in $m$. □

Secondly, we determine the asymptotic variance of $S_n(m)$.

**Lemma 3.2.2**

*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i) and (ii). Then, uniformly in $m \in \mathbb{N}_0$,*

$$\frac{1}{n} E S_n^2(m) = \frac{\sigma^2}{1-\varphi^2-\omega^2} \frac{1+\varphi}{1-\varphi} + \mathcal{O}\left(\frac{1}{n}\right)$$

*as $n \to \infty$.*

**Proof:** It holds,

$$\frac{1}{n} E\left(S_n(m) S_n(m)\right) = \frac{1}{n} E\left(\sum_{k=m+1}^{m+n} X_k^2\right) + 2 \sum_{k>l}^{m+n} E(X_k X_l)$$

and

$$\frac{1}{n} E\left(\sum_{k=m+1}^{m+n} X_k^2\right) = \frac{1}{n} n \gamma_X(0) = \frac{\sigma^2}{1-\varphi^2-\omega^2},$$

$$\frac{2}{n} \sum_{k>l}^{m+n} E(X_k X_l) = \frac{2}{n} \sum_{k>l}^{m+n} \gamma_X(k-l) = \frac{2\sigma^2}{1-\varphi^2-\omega^2} \frac{1}{n} \sum_{i=1}^{n-1} i \varphi^{n-i}.$$

Now

$$\begin{aligned}
\frac{1}{n} \sum_{i=1}^{n-1} i \varphi^{n-i} &= \frac{1}{n} \sum_{i=1}^{n-1} (n-i) \varphi^i \\
&= \sum_{i=1}^{n-1} \varphi^i - \frac{1}{n} \sum_{i=1}^{n-1} i \varphi^i
\end{aligned}$$

$$= \frac{\varphi(1-\varphi^n)}{1-\varphi} - \frac{\varphi}{n}\left(\frac{\varphi(1-\varphi^n)}{1-\varphi}\right)'$$
$$= \frac{\varphi}{1-\varphi} + \mathcal{O}\left(\frac{1}{n}\right)$$

as $n \to \infty$, since

$$\left(\frac{\varphi(1-\varphi^n)}{1-\varphi}\right)' \longrightarrow -\frac{1}{(1-\varphi)^2} \qquad \text{and} \qquad \frac{\varphi(1-\varphi^n)}{1-\varphi} \longrightarrow \frac{\varphi}{1-\varphi}$$

exponentially fast as $n \to \infty$, finishing the proof. □

Finally, we calculate the order of the difference between the variance and the conditional variance of $S_n(m)$.

**Lemma 3.2.3**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i) and (ii). Then, uniformly in $m \in \mathbb{N}_0$,*

$$\left\|E\left(S_n^2(m)|\mathcal{F}_m\right) - ES_n^2(m)\right\|_1 = \mathcal{O}(1),$$

*as $n \to \infty$.*

**Proof:** The proof is given in three steps.

a) Firstly,

$$E\left(S_n^2(m)|\mathcal{F}_m\right) = \sum_{k=m+1}^{m+n} E\left(X_k^2|\mathcal{F}_m\right) + 2\sum_{k>l}^{m+n} E\left(X_kX_l|\mathcal{F}_m\right) \qquad \text{a.s.}$$

Consider

$$X_{m+i}^2 = \left(\varphi^2 + 2b_{m+i}\varphi + b_{m+i}^2\right)X_{m+i-1}^2 + 2(\varphi + b_{m+i})X_{m+i-1}e_{m+i} + e_{m+i}^2.$$

Then, we recursively get

$$\begin{aligned} E\left(X_{m+i}^2|\mathcal{F}_m\right) &= E\left(E\left(X_{m+i}^2|\mathcal{F}_{m+i-1}\right)|\mathcal{F}_m\right) \\ &= E\left(\left(\varphi^2+\omega^2\right)X_{m+i-1}^2 + \sigma^2|\mathcal{F}_m\right) \\ &= \left(\varphi^2+\omega^2\right)E\left(X_{m+i-1}^2|\mathcal{F}_m\right) + \sigma^2 \\ &= \left(\varphi^2+\omega^2\right)^i X_m^2 + \sigma^2\sum_{j=1}^{i}(\varphi^2+\omega^2)^{j-1} \qquad \text{a.s.} \end{aligned}$$

Furthermore,

$$\begin{aligned}
\sum_{k=m+1}^{m+n} E\left(X_k^2|\mathcal{F}_m\right) &= \sum_{i=1}^{n} E\left(X_{m+i}^2|\mathcal{F}_m\right) \\
&= \sum_{i=1}^{n}\left(\varphi^2+\omega^2\right)^i X_m^2 + \sigma^2\sum_{i=1}^{n}\sum_{j=1}^{i}\left(\varphi^2+\omega^2\right)^{j-1} \\
&= \sum_{i=1}^{n}\left(\varphi^2+\omega^2\right)^i X_m^2 + \sigma^2\sum_{i=1}^{n}\frac{1-(\varphi^2+\omega^2)^i}{1-\varphi^2-\omega^2} \\
&= \frac{(\varphi^2+\omega^2)X_m^2(1-(\varphi^2+\omega^2)^n)}{1-\varphi^2-\omega^2} \\
&\qquad + \frac{n\sigma^2}{1-\varphi^2-\omega^2} - \frac{\sigma^2(\varphi^2+\omega^2)(1-(\varphi^2+\omega^2)^n)}{1-\varphi^2-\omega^2}
\end{aligned}$$

a.s. and for $i > j$ we have a.s.

$$\begin{aligned}
E(X_{m+i}X_{m+j}|\mathcal{F}_m) &= E(E(X_{m+i}X_{m+j}|\mathcal{F}_{m+j})\mathcal{F}_m) \\
&= \varphi^{i-j}E\left(X_{m+j}^2|\mathcal{F}_m\right) \\
&= \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j X_m^2 + \sigma^2\varphi^{i-j}\sum_{k=1}^{j}\left(\varphi^2+\omega^2\right)^{k-1} \\
&= \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j X_m^2 + \sigma^2\varphi^{i-j}\frac{1-(\varphi^2+\omega^2)^j}{1-\varphi^2-\omega^2}.
\end{aligned}$$

Thus a.s.,

$$\begin{aligned}
\sum_{k>l}^{m+n} E(X_kX_l|\mathcal{F}_m) &= \sum_{i>j}^{n} E(X_{m+i}X_{m+j}|\mathcal{F}_m) \\
&= \sum_{i>j}^{n}\varphi^{i-j}\left(\varphi^2+\omega^2\right)^j X_m^2 + \sigma^2\sum_{i>j}^{n}\varphi^{i-j}\frac{1-(\varphi^2+\omega^2)^j}{1-\varphi^2-\omega^2}.
\end{aligned}$$

b) It holds,

$$\begin{aligned}
ES_n^2(m) &= \sum_{k=m+1}^{m+n} EX_k^2 + 2\sum_{k>l}^{m+n} E(X_kX_l) \\
&= \sum_{i=1}^{n}\gamma_X(0) + 2\sum_{i>j}^{n}\gamma_X(i-j) \\
&= \frac{n\sigma^2}{1-\varphi^2-\omega^2} + 2\sigma^2\sum_{i>j}^{n}\frac{\varphi^{i-j}}{1-\varphi^2-\omega^2}.
\end{aligned}$$

c) From a) and b),

$$\begin{aligned}
&E\left(S_n^2(m)|\mathcal{F}_m\right) - ES_n^2(m) \\
&= \sum_{i=1}^{n} E\left(X_{m+i}^2|\mathcal{F}_m\right) - \sum_{i=1}^{n} EX_{m+i}^2 \\
&\qquad +2\sum_{i>j}^{n} E(X_{m+i}X_{m+j}|\mathcal{F}_m) - 2\sum_{i>j}^{n} E(X_{m+i}X_{m+j}) \\
&= D_1 + D_2 \qquad \text{a.s.}
\end{aligned}$$

Now,

$$\begin{aligned}
D_1 &= \left(X_m^2 - \sigma^2\right) \frac{(\varphi^2+\omega^2)(1-(\varphi^2+\omega^2)^n)}{1-\varphi^2-\omega^2} \\
&\longrightarrow \left(X_m^2 - \sigma^2\right) \frac{\varphi^2+\omega^2}{1-\varphi^2-\omega^2} \qquad \text{a.s.}
\end{aligned}$$

as $n \to \infty$ and hence $ED_1 = \mathcal{O}(1)$ uniformly in $m \in \mathbb{N}_0$. Moreover, $ED_2 = \mathcal{O}(1)$ if $\varphi = 0$. If $\varphi \neq 0$, we get

$$D_2 = \frac{2(X_m^2 - \sigma^2)}{1-\varphi^2-\omega^2} \sum_{i>j}^{n} \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j \qquad \text{a.s.}$$

(and hence $ED_2 = \mathcal{O}(1)$ uniformly in $m \in \mathbb{N}_0$) as $n \to \infty$ as follows. Set

$$I = \sum_{i>j}^{n} \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j = \sum_{j=1}^{n}\sum_{i=j+1}^{n} \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j$$

and consider the inner sum

$$\begin{aligned}
\sum_{i=j+1}^{n} \varphi^{i-j}\left(\varphi^2+\omega^2\right)^j &= \varphi^{-j}\left(\varphi^2+\omega^2\right)^j \sum_{i=j+1}^{n} \varphi^i \\
&= \varphi^{-j}\left(\varphi^2+\omega^2\right)^j \frac{\varphi^{j+1}-\varphi^{n+1}}{1-\varphi}.
\end{aligned}$$

This yields

$$\begin{aligned}
I &= \sum_{j=1}^{n} \left(\frac{\varphi^2+\omega^2}{\varphi}\right)^j \frac{\varphi^{j+1}-\varphi^{n+1}}{1-\varphi} \\
&= \frac{\varphi}{1-\varphi}\sum_{j=1}^{n}\left(\varphi^2+\omega^2\right)^j - \frac{1}{1-\varphi}\sum_{j=1}^{n}\left(\varphi^2+\omega^2\right)^j \varphi^{n-j+1} \\
&= I_1 + I_2
\end{aligned}$$

with

$$\lim_{n\to\infty} I_1 \;=\; \frac{\varphi(\varphi^2+\omega^2)}{(1-\varphi)(1-\varphi^2-\omega^2)}$$

and

$$I_2 \;=\; \mathcal{O}\left(\frac{1}{1-\varphi}\sum_{j=1}^{n}|\varphi|^{n-j+1}\right) \;=\; \mathcal{O}\left(\frac{1}{1-\varphi}\sum_{j=1}^{n}|\varphi|^{j}\right) \;=\; \mathcal{O}(1)$$

since $\varphi^2+\omega^2<1$ and the sum converges to $(1-|\varphi|)^{-1}$. □

Thus, we have proved Theorem 3.2.2.

**Proof of Theorem 3.2.2:** Applying Feigin and Tweedie (1985), from (3.4), we immediately get that $E|X_1|^\kappa < \infty$. Hence, the assertion follows from Theorem 3.2.1 in combination with Lemmas 3.2.1 – 3.2.3. □

In Section 3.4, we shall construct a test for a change of the parameter $\varphi$, too. In order to do so, we need a strong invariance principle for the partial sums of the random variables $\{Z_n\}_{n\in\mathbb{Z}}$ defined via the equations

$$Z_n \;=\; X_{n-1}(X_n-\varphi X_{n-1}) \qquad (n\in\mathbb{Z}), \tag{3.5}$$

where $\{X_n\}_{n\in\mathbb{Z}}$ denotes an RCA(1) time series. Set

$$R_n(m) \;=\; Z_{m+1}+\ldots+Z_{m+n} \qquad (m\in\mathbb{N}_0,\, n\in\mathbb{N})$$

and $R_n = R_n(0)$. Under suitable moment conditions we arrive at the following strong approximation.

**Theorem 3.2.3 (Strong invariance)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let $\{Z_n\}_{n\in\mathbb{Z}}$ be defined in (3.5). If*

$$E|e_1|^\kappa<\infty \qquad \text{and} \qquad E|\varphi+b_1|^\kappa<1$$

*with some $\kappa>4$, then, there exists a Wiener process $\{W(t):t\ge 0\}$ such that*

$$R_{[t]}-\sigma_R W(t) \;=\; \mathcal{O}\left(t^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

*as $t\to\infty$ for some $\nu>2$, where*

$$\sigma_R^2 \;=\; \omega^2 EX_1^4+\sigma^2 EX_1^2.$$

Now, we prove the corresponding lemmas for the sequence $\{Z_n\}_{n\in\mathbb{Z}}$ by replacing $S_n(m)$ with $R_n(m)$ in the statements of the previous lemmas. We start again with the consideration of the conditional expectation of $R_n(m)$ with respect to the filtration given in (3.2).

**Lemma 3.2.4**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let $\{Z_n\}_{n\in\mathbb{Z}}$ be defined in (3.5). Then, for all $n \in \mathbb{N}$, $m \in \mathbb{N}_0$,*

$$\|E(R_n(m)|\mathcal{F}_m)\|_1 \;=\; 0.$$

**Proof:** Since

$$E(X_{i-1}(X_i - \varphi X_{i-1})|\mathcal{F}_{i-1}) \;=\; X_{i-1}E(X_i|\mathcal{F}_{i-1}) - \varphi X_{i-1}^2 \;=\; 0 \qquad \text{a.s.}$$

for all $i \in \mathbb{Z}$, $\{Z_n\}_{n\in\mathbb{Z}}$ is a sequence of martingale differences and the assertion follows. □

Next, we estimate the variance of $R_n(m)$.

**Lemma 3.2.5**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let $\{Z_n\}_{n\in\mathbb{Z}}$ be defined in (3.5). If*

$$Ee_1^4 < \infty \qquad \textit{and} \qquad E(\varphi + b_1)^4 < 1,$$

*then, for all $n \in \mathbb{N}$, $m \in \mathbb{N}_0$,*

$$\frac{1}{n}ER_n^2(m) \;=\; \omega^2 EX_1^4 + \sigma^2 EX_1^2.$$

**Proof:** It follows from our assumptions that $EX_1^4 < \infty$ (cf. Feigin and Tweedie (1985)). Now,

$$\begin{aligned}\frac{1}{n}ER_n^2(m) \;&=\; \frac{1}{n}\sum_{i=m+1}^{m+n} E(X_{i-1}^2 b_i + X_{i-1}e_i)^2 \\ &\qquad + \frac{2}{n}\sum_{i>j}^{m+n} E(X_{i-1}^2 b_i + X_{i-1}e_i)(X_{j-1}^2 b_j + X_{j-1}e_j) \\ &=\; J_1 + J_2.\end{aligned}$$

By strong stationarity, the first term gives

$$\begin{aligned}J_1 \;&=\; \frac{1}{n}\sum_{i=m+1}^{m+n} EX_{i-1}^4 b_i^2 + \frac{1}{n}\sum_{i=m+1}^{m+n} EX_{i-1}^2 e_i^2 \\ &=\; \omega^2 EX_1^4 + \sigma^2 EX_1^2 \;<\; \infty,\end{aligned}$$

while the second is zero, since for $i > j$:

$$\begin{aligned} & E(X_{i-1}^2 b_i + X_{i-1}e_i)(X_{j-1}^2 b_j + X_{j-1}e_j) \\ &= EX_{i-1}^2 X_{j-1}^2 b_j Eb_i + EX_{i-1}^2 X_{j-1}e_j Eb_i \\ &\qquad + EX_{i-1}X_{j-1}^2 b_j Ee_i + EX_{i-1}X_{j-1}e_j Ee_i \\ &= 0 \end{aligned}$$

by (i). This completes the proof. □

We need the following proposition which gives a recursion for the fourth conditional moments of the $X_n$.

**Proposition 3.2.1**

*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i) and (ii). If*

$$Ee_1^4 < \infty \qquad \text{and} \qquad E(\varphi + b_1)^4 < 1,$$

*then, the recursion*

$$\begin{aligned} E\left(X_{m+i-1}^4 | \mathcal{F}_m\right) &= c_0^{i-1} X_m^4 + \sum_{j=1}^{i-1} c_0^{j-1} c_1 E\left(X_{m+i-j-1}^2 | \mathcal{F}_m\right) \\ &\qquad + \sum_{j=1}^{i-1} c_0^{j-1} c_2 E(X_{m+i-j-1} | \mathcal{F}_m) + \frac{1 - c_0^{i-1}}{1 - c_0} c_3 \qquad \text{a.s.} \end{aligned}$$

*holds true for all $i = 1, \ldots, n$, where*

$$c_0 = E(\varphi + b_1)^4, \quad c_1 = 6\sigma^2(\varphi^2 + \omega^2), \quad c_2 = \varphi E e_1^3, \quad c_3 = E e_1^4.$$

**Proof:** We have

$$\begin{aligned} E\left(X_{m+i-1}^4 | \mathcal{F}_m\right) &= c_0 E\left(X_{m+i-2}^4 | \mathcal{F}_m\right) + N_{m+i-2} \\ &= c_0^{i-1} X_m^4 + \sum_{j=1}^{i-1} c_0^{j-1} N_{m+i-j-1} \qquad \text{a.s.,} \end{aligned}$$

where

$$N_{m+i-j-1} = c_1 E\left(X_{m+i-j-1}^2 | \mathcal{F}_m\right) + c_2 E(X_{m+i-j-1} | \mathcal{F}_m) + c_3 \qquad \text{a.s.}$$

Now,

$$\begin{aligned} \sum_{j=1}^{i-1} c_0^{j-1} N_{m+i-j-1} &= \sum_{j=1}^{i-1} c_0^{j-1} c_1 E\left(X_{m+i-j-1}^2 | \mathcal{F}_m\right) \\ &\qquad + \sum_{j=1}^{i-1} c_0^{j-1} c_2 E(X_{m+i-j-1} | \mathcal{F}_m) + \sum_{j=1}^{i-1} c_0^{j-1} c_3 \qquad \text{a.s.,} \end{aligned}$$

finishing the proof. □

Finally, we obtain a third lemma.

**Lemma 3.2.6**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let $\{Z_n\}_{n\in\mathbb{Z}}$ be defined in (3.5). If*

$$Ee_1^4 < \infty \qquad \text{and} \qquad E(\varphi + b_1)^4 < 1,$$

*then, uniformly in $m \in \mathbb{N}_0$,*

$$\|E(R_n^2(m)|\mathcal{F}_m) - ER_n^2(m)\|_1 \;=\; \mathcal{O}(1)$$

*as $n \to \infty$.*

**Proof:**

a) Similarly as in the proof of Lemma 3.2.5,

$$\begin{aligned} ER_n^2(m) &= E\left(\sum_{j=m+1}^{m+n}\left(X_{j-1}^2 b_j + X_{j-1}e_j\right)\right)^2 \\ &= n\omega^2 EX_0^4 + n\sigma^2 EX_0^2 \end{aligned}$$

and therefore the variance is independent of $m$.

b) By (i) and the definition of $\{\mathcal{F}_m\}_{m\in\mathbb{Z}}$ in (3.2),

$$\begin{aligned} &E\left(\left(\sum_{i=m+1}^{m+n}\left(X_{i-1}^2 b_i + X_{i-1}e_i\right)\right)^2 |\mathcal{F}_m\right) \\ &= \sum_{i=m+1}^{m+n} E\left(X_{i-1}^4 b_i^2|\mathcal{F}_m\right) + \sum_{i=m+1}^{m+n} E\left(X_{i-1}^2 e_i^2|\mathcal{F}_m\right) \\ &= L_1 + L_2 \qquad \text{a.s.} \end{aligned}$$

Using the recursions from the proof of Lemma 3.2.3, for $i = 1, \ldots, n$ we get

$$\begin{aligned} &E\left(X_{m+i-1}^2 e_{m+i}^2|\mathcal{F}_m\right) \\ &= E\left(E\left(X_{m+i-1}^2 e_{m+i}^2|\mathcal{F}_{m+i-1}|\right)\mathcal{F}_m\right) \\ &= \sigma^2 E\left(X_{m+i-1}^2|\mathcal{F}_m\right) \\ &= \sigma^2\left(X_m^2 - EX_0^2\right)\left(\varphi^2+\omega^2\right)^{i-1} + \sigma^2 EX_0^2 \qquad \text{a.s.} \end{aligned}$$

Thus,

$$\begin{aligned} L_2 &= \sum_{i=1}^{n} E\left(X_{m+i-1}^2 e_{m+i}^2 | \mathcal{F}_m\right) \\ &= \sigma^2 \left(X_m^2 - EX_0^2\right) \sum_{i=1}^{n} \left(\varphi^2 + \omega^2\right)^{i-1} + n\sigma^2 EX_0^2 \qquad \text{a.s.} \end{aligned}$$

and hence $EL_2 = \mathcal{O}(1)$ as $n \to \infty$. An application of Proposition 3.2.1 yields

$$EL_1 = n\omega^2 EX_0^4 + \mathcal{O}(1) \qquad \text{a.s.} \tag{3.6}$$

as $n \to \infty$ in a similar way: Firstly,

$$\sum_{i=m+1}^{m+n} E\left(X_{i-1}^4 b_i^2 | \mathcal{F}_m\right) = \omega^2 \sum_{i=1}^{n} E\left(X_{m+i-1}^4 | \mathcal{F}_m\right) \qquad \text{a.s.}$$

Now, we are able to use the recursion from Proposition 3.2.1.

(i) The first term yields

$$\sum_{i=1}^{n} c_0^{i-1} X_m^4 = X_m^4 \frac{1 - c_0^n}{1 - c_0} \qquad \text{a.s.}$$

as $n \to \infty$, since $c_0 = E(\varphi + b_1)^4 < 1$ by assumption.

(ii) For the second term, it holds true

$$\begin{aligned} &\sum_{i=1}^{n} \sum_{j=1}^{i-1} c_0^{j-1} c_1 E\left(X_{m+i-j-1}^2 | \mathcal{F}_m\right) \\ &= c_1 \sum_{i=1}^{n} \sum_{j=1}^{i-1} c_0^{j-1} X_m^2 \left(\varphi^2 + \omega^2\right)^{i-j+1} + \frac{c_1 \sigma^2}{1 - \varphi^2 - \omega^2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} c_0^{j-1} \\ &\qquad - \frac{c_1 \sigma^2}{1 - \varphi^2 - \omega^2} \sum_{i=1}^{n} \sum_{j=1}^{i-1} c_0^{j-1} \left(\varphi^2 + \omega^2\right)^{i-j+1} \\ &= L_a + L_b + L_c \qquad \text{a.s.} \end{aligned}$$

Therein,

$$L_a \le c_1 X_m^2 \sum_{i=1}^{n} (i-1) \left(\varphi^2 + \omega^2\right)^{i-2} \qquad \text{a.s.}$$

and $EL_a = \mathcal{O}(1)$ as $n \to \infty$, by assumption on $c_0$. Analogously,

$$L_c = \mathcal{O}(1)$$

as $n \to \infty$. Finally,

$$\begin{aligned} L_b &= \frac{nc_1\sigma^2}{(1-\varphi^2-\omega^2)(1-c_0)} - \frac{c_1\sigma^2(1-c_0^n)}{(1-\varphi^2-\omega^2)(1-c_0)} \\ &= \frac{nc_1EX_1^2}{1-c_0} + \mathcal{O}(1) \end{aligned}$$

as $n \to \infty$.

(iii) Next, we consider the third term of the recursion. It holds,

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{j=1}^{i-1} c_0^{j-1}c_2E\left(X_{m+i-j-1}|\mathcal{F}_m\right) \\ &= c_2\sum_{i=1}^{n}\sum_{j=1}^{i-1} c_0^{j-1}\varphi^{i-j-1}X_m \qquad \text{a.s.} \end{aligned}$$

as $n \to \infty$, by the same arguments used in the proof of Lemma 3.2.3, step c). Recall again, that $c_0 < 1$.

(iv) The last term yields

$$c_3\sum_{i=1}^{n}\frac{1-c_0^{i-1}}{1-c_0} = \frac{nc_3}{1-c_0} + \mathcal{O}(1)$$

as $n \to \infty$.

On recognizing

$$EX_1^4 = \frac{1}{1-c_0}\left(c_1EX_1^2 + c_3\right),$$

relation (3.6) is proved.

On combining a) and b) the proof of Lemma 3.2.6 is complete. □

Therefore, Theorem 3.2.3 is readily proved.

**Proof of Theorem 3.2.3:** It is obtained by an application of Lemmas 3.2.4 – 3.2.6 and Theorem 3.2.1. □

## 3.3 A sequential test procedure

In this section, we shall apply the strong invariance principle proved in Theorem 3.2.2 to provide sequential test procdeures. Hence, we deal again with the situation of Chapter 2, i.e., in more detail, a change in the mean scenario for the random variables $\{Y_n\}_{n\in\mathbb{N}}$ defined by

$$Y_n = \begin{cases} \mu + X_n & : \quad n = 1, \ldots, m + k^* - 1, \\ \mu + \Delta_m + X_n & : \quad n = m + k^*, m + k^* + 1, \ldots, \end{cases}$$

where $\mu$, $k^*$ and $\Delta_m$ are unknown parameters and $\{X_n\}_{n\in\mathbb{Z}}$ is an RCA(1) time series. In particular, we assume that there is no change in a historical data set of size $m$. We are interested in testing the hypotheses

$$\mathcal{H}_{\mathbf{0}} \ : \ \Delta_m = 0,$$
$$\mathcal{H}_{\mathbf{A}} \ : \ \Delta_m \neq 0.$$

Rephrasing the arguments of Section 2.3, we stop and reject $\mathcal{H}_{\mathbf{0}}$, if

$$\tau_m \ = \ \inf \{k \geq 1 \ : \ |Q(m,k)| \geq g(m,k)\}$$

with the CUSUM detector

$$Q(m,k) \ = \ \sum_{j=m+1}^{m+k} Y_j - \frac{k}{m} \sum_{j=1}^{m} Y_j$$

and the boundary function

$$g(m,k) \ = \ cm^{1/2} \left(1 + \frac{k}{m}\right) \left(\frac{k}{m+k}\right)^{\gamma},$$

where $\gamma \in [0, \frac{1}{2})$ and $c = c(\alpha)$. Then, we can derive the following limit theorems for the test statistic under the null and alternative hypothesis.

**Theorem 3.3.1 (Asymptotic under the null hypothesis)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let*

$$E|e_1|^{\kappa} < \infty \qquad \text{and} \qquad E|\varphi + b_1|^{\kappa} < 1$$

*for some $\kappa > 2$. Then, under $\mathcal{H}_{\mathbf{0}}$,*

$$\lim_{m\to\infty} P\left\{\frac{1}{\sigma_S} \sup_{k\geq 1} \frac{|Q(m,k)|}{g(m,k)} \leq 1\right\} = P\left\{\sup_{0\leq t\leq 1} \frac{|\hat{W}(t)|}{t^{\gamma}} \leq c\right\},$$

*where $\{\hat{W}(t) : 0 \leq t \leq 1\}$ denotes a Wiener process and $\sigma_S^2$ is given in Theorem 3.2.2.*

**Proof:**

a) From Theorem 3.2.2, for all $k \geq 1$ we get

$$\sum_{i=m+1}^{m+k} X_i = \sigma_S(W(m+k) - W(m)) + \mathcal{O}\left((m+k)^{\frac{1}{\nu}} + m^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

and

$$\sum_{i=1}^{m} X_i = \sigma_S W(m) + \mathcal{O}\left(m^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

as $m \to \infty$. Set $\tilde{W}_m(k) = W(m+k) - W(m)$. Then,

$$\begin{aligned}
&\sup_{k\geq 1} \frac{1}{g(m,k)} \left| \sum_{i=m+1}^{m+k} X_i - \frac{k}{m}\sum_{i=1}^{m} X_i - \sigma_S\left(\tilde{W}_m(k) - \frac{k}{m}W(m)\right)\right| \\
&= \mathcal{O}\left(\sup_{k\geq 1} \frac{(m+k)^{\frac{1}{\nu}} + m^{\frac{1}{\nu}} + \frac{k}{m}m^{\frac{1}{\nu}}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}}\right) \qquad \text{a.s.}
\end{aligned}$$

as $m \to \infty$. Now,

$$\begin{aligned}
&\sup_{1\leq k\leq m} \frac{(m+k)^{\frac{1}{\nu}} + m^{\frac{1}{\nu}} + \frac{k}{m}m^{\frac{1}{\nu}}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} \\
&\leq \sup_{1\leq k\leq m} \frac{(m+k)^{\frac{1}{\nu}} + 2m^{\frac{1}{\nu}}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} \\
&\leq \left(2^{\frac{1}{\nu}} + 2\right) m^{\frac{1}{\nu}-\frac{1}{2}} \sup_{1\leq k\leq m}\left(\frac{k}{m+k}\right)^{1-\gamma} \\
&\leq \left(2^{\frac{1}{\nu}} + 2\right) m^{\frac{1}{\nu}-\frac{1}{2}} 2^{\gamma-1} \\
&= o(1)
\end{aligned}$$

as $m \to \infty$, since $\nu > 2$. Similarly,

$$\sup_{m<k<\infty} \frac{(m+k)^{\frac{1}{\nu}} + m^{\frac{1}{\nu}} + \frac{k}{m}m^{\frac{1}{\nu}}}{\sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m+k}\right)^{\gamma}} = \mathcal{O}\left(m^{\frac{1}{\nu}-\frac{1}{2}}\right) = o(1)$$

as $m \to \infty$. Thus, we have proved

$$\sup_{k\geq 1} \frac{1}{g(m,k)} \left| \sum_{i=m+1}^{m+k} X_i - \frac{k}{m}\sum_{i=1}^{m} X_i - \sigma_S\left(\tilde{W}_m(k) - \frac{k}{m}W(m)\right)\right| = o(1) \qquad \text{a.s.}$$

as $m \to \infty$.

b) The distribution of $\{\tilde{W}_m(k)\}$ is independent of $m$. Hence,

$$\sup_{k\geq 1}\frac{|\tilde{W}_m(k)-\frac{k}{m}W(m)|}{g(m,k)} \stackrel{\mathcal{D}}{=} \sup_{k\geq 1}\frac{|\tilde{W}(k)-\frac{k}{m}W(m)|}{g(m,k)},$$

where $\{\tilde{W}(t) : t \geq 0\}$ is a Wiener process independent of $\{W(t) : t \geq 0\}$. Now, we are exactly at the starting point of the proof of Theorem 2.1 in Horváth, Hušková, Kokoszka and Steinebach (2003). Along their lines, the proof is complete. □

**Theorem 3.3.2 (Asymptotic under the alternative)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let*

$$E|e_1|^\kappa < \infty \qquad \text{and} \qquad E|\varphi + b_1|^\kappa < 1$$

*for some $\kappa > 2$. If $\Delta_m = \Delta$ and $k^* = o(m)$, then, under $\mathcal{H}_{\mathbf{A}}$,*

$$\frac{1}{\sigma_S}\sup_{k\geq 1}\frac{|Q(m,k)|}{g(m,k)} \xrightarrow{P} \infty$$

*as $m \to \infty$, where $\sigma_S^2$ is given in Theorem 3.2.2.*

**Proof:** Set $\tilde{k} = m + k^*$. Then,

$$\begin{aligned} Q(m,\tilde{k}) &= \sum_{i=m+1}^{m+\tilde{k}} Y_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} Y_i \\ &= \sum_{i=m+1}^{m+\tilde{k}} X_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} X_i + \Delta_m(\tilde{k} - k^* + 1) \end{aligned}$$

with $\Delta_m = \Delta \neq 0$ under $\mathcal{H}_{\mathbf{A}}$. By Theorem 3.3.1, we have

$$\frac{1}{g(m,\tilde{k})}\left(\sum_{i=m+1}^{m+\tilde{k}} X_i - \frac{\tilde{k}}{m}\sum_{i=1}^{m} X_i\right) = \mathcal{O}_P(1)$$

as $m \to \infty$. Since $k^* = o(m)$ by assumption,

$$1 \geq \left(1 + \frac{\tilde{k}}{m}\right)\left(\frac{\tilde{k}}{m+\tilde{k}}\right)^\gamma \to 2^{1-\gamma}$$

as $m \to \infty$. Therefore,

$$\frac{|\Delta_m|(\tilde{k} - k^* + 1)}{g(m,\tilde{k})} \sim \frac{\sqrt{m}|\Delta|}{2^{1-\gamma}} \to \infty$$

as $m \to \infty$, where $a_n \sim b_n$ means $\frac{a_n}{b_n} \to 1$ as $n \to \infty$. This completes the proof. $\square$

Our next goal is to replace the variance parameter $\sigma_S^2$ of the approximating Wiener process $\{W(t) : 0 \le t \le 1\}$ from Theorem 3.2.2 by a suitable estimator $\hat{\sigma}_{S,m}^2$ $(m \ge 1)$ (which is obtained by plugging in consistent estimators of the parameters of the RCA(1) time series).

**Lemma 3.3.1**

*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and for every $m \ge 1$ let $\hat{\varphi}_m$, $\hat{\sigma}_m^2$ and $\hat{\omega}_m^2$ be (weakly) consistent estimators of the parameters $\varphi$, $\sigma^2$ and $\omega^2$, respectively. Then,*

$$\hat{\sigma}_{S,m}^2 = \frac{\hat{\sigma}_m^2}{1-\hat{\varphi}_m^2-\hat{\omega}_m^2}\frac{1+\hat{\varphi}_m}{1-\hat{\varphi}_m} \xrightarrow{P} \sigma_S^2 \tag{3.7}$$

*as $m \to \infty$, i.e., $\hat{\sigma}_{S,m}^2$ is a (weakly) consistent estimator of $\sigma_S^2$.*

**Proof:** $(\hat{\varphi}_m, \hat{\sigma}_m^2, \hat{\omega}_m^2)'$ is a (weakly) consistent estimator for the vector valued parameter $(\varphi, \sigma^2, \omega^2)'$. The estimator $\hat{\sigma}_{S,m}^2$ is obtained from $(\hat{\varphi}_m, \hat{\sigma}_m^2, \hat{\omega}_m^2)'$ through a continuous transformation and hence (weakly) consistent for $\sigma_S^2$. $\square$

Now, we get the following corollaries of Theorems 3.3.1 and 3.3.2.

**Corollary 3.3.1**

*Let the assumptions of Theorem 3.3.1 be satisfied. Then, under $\mathcal{H}_\mathbf{0}$,*

$$\lim_{m\to\infty} P\left\{\frac{1}{\hat{\sigma}_{S,m}}\sup_{k\ge 1}\frac{|Q(m,k)|}{g(m,k)} \le 1\right\} = P\left\{\sup_{0\le t\le 1}\frac{|\hat{W}(t)|}{t^\gamma} \le c\right\},$$

*where $\hat{\sigma}_{S,m}^2$ is defined in (3.7).*

**Proof:** The assertion follows from Theorem 3.3.1 and Lemma 3.3.1. $\square$

**Corollary 3.3.2**

*Let the assumptions of Theorem 3.3.2 be satisfied. Then, under $\mathcal{H}_\mathbf{A}$,*

$$\frac{1}{\hat{\sigma}_{S,m}}\sup_{k\ge 1}\frac{|Q(m,k)|}{g(m,k)} \xrightarrow{P} \infty$$

*as $m \to \infty$, where $\hat{\sigma}_{S,m}^2$ is defined in (3.7).*

**Proof:** The assertion follows from Theorem 3.3.2 and Lemma 3.3.1. $\square$

One way to obtain consistent estimators of $\varphi, \omega^2$ and $\sigma^2$ is the application of a two step procedure which leads to (conditional) least squares estimators (LSE) $\hat{\varphi}_{m,L}, \hat{\omega}^2_{m,L}$ and $\hat{\sigma}^2_{m,L}$. Along the lines of Section 3.2 in Nicholls and Quinn (1982), firstly

$$\hat{\varphi}_{m,L} = \left(\sum_{i=1}^{m} X_{i-1}^2\right)^{-1} \sum_{i=1}^{m} X_{i-1}X_i \tag{3.8}$$

can be calculated by minimizing the sum

$$\sum_{i=1}^{m}(X_i - \varphi X_{i-1})^2.$$

with respect to $\varphi$. Then, in a similar way, we get

$$\begin{aligned}
\hat{\omega}^2_{m,L} &= \left(\sum_{i=1}^{m}(X_{i-1}^2 - \bar{X}_m^2)^2\right)^{-1} \sum_{i=1}^{m}(X_{i-1}^2 - \bar{X}_m^2)(X_i - \hat{\varphi}_{m,L}X_{i-1})^2, \\
\hat{\sigma}^2_{m,L} &= \frac{1}{m}\sum_{i=1}^{m}(X_i - \hat{\varphi}_{m,L}X_{i-1})^2 - \hat{\omega}^2_{m,L}\bar{X}_m^2
\end{aligned}$$

as minimizers (with respect to $\omega^2$ and $\sigma^2$, respectively) of

$$\sum_{i=1}^{m}\left((X_i - \hat{\varphi}_{m,L}X_{i-1})^2 - \omega^2 X_{i-1}^2 - \sigma^2\right)^2.$$

Recall, that under $\mathcal{H}_\mathbf{0}$ we a.s. have $E((X_i - \varphi X_{i-1})^2|\mathcal{F}_{i-1}) = \omega^2 X_{i-1}^2 + \sigma^2$ for $i = 1, \dots, m$.

How long does it take to detect the change–point and how big is the difference between the stopping time and the true change–point given in the underlying model? We will answer this question by giving the limit distribution of the stopping time $\tau_m$. Thereby, we will follow closely the lines of Section 2.6. So, we establish the following conditions on $k^*$ and $\Delta_m$, which is moreover assumed to be positive without loss of generality from now on. Let

$$\Delta_m \to 0, \qquad \sqrt{m}\Delta_m \to 0, \tag{3.9}$$

$$k^* = \mathcal{O}\left(m^\theta\right) \qquad \text{with some} \qquad 0 \le \theta < \left(\frac{\frac{1}{2} - \gamma}{1 - \gamma}\right)^2 \tag{3.10}$$

as $m \to \infty$. Recall that $\gamma \in [0, \frac{1}{2})$.

**Theorem 3.3.3 (Limit distribution of the stopping time)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and*

$$E|e_1|^\kappa < \infty \qquad \text{and} \qquad E|\varphi + b_1|^\kappa < 1$$

*for some $\kappa > 2$. Let the conditions (3.9) and (3.10) be satisfied. Then, under $\mathcal{H}_{\mathbf{A}}$,*

$$\lim_{m\to\infty} P\left\{\frac{\tau_m - a_m}{b_m} \le x\right\} = \Phi(x),$$

*where $\Phi(x)$ denotes the standard normal distribution function. Moreover,*

$$\begin{aligned} a_m &= \left(\frac{cm^{\frac{1}{2}-\gamma}}{\Delta_m}\right)^{\frac{1}{1-\gamma}}, \\ b_m &= \frac{\sqrt{a_m}\sigma_S}{(1-\gamma)\Delta_m} = \frac{\sqrt{a_m}}{(1-\gamma)\Delta_m}\frac{\sigma^2}{1-\varphi^2-\omega^2}\frac{1+\varphi}{1-\varphi}, \end{aligned}$$

*where $\sigma_S^2$ is given in Theorem 3.2.2.*

**Proof:** Since

$$\sum_{i=1}^m X_i = \sigma_S W(m) + \mathcal{O}\left(m^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

as $m \to \infty$ by Theorem 3.2.2, we have that

$$\sum_{i=1}^m X_i = \mathcal{O}_P\left(\sqrt{m}\right)$$

as $m \to \infty$. Therefore, the assumptions of Theorem 2.6.1 are satisfied and the assertion follows. □

## 3.4 A–posteriori tests

Beside the sequential testing of Section 3.3, Theorem 3.2.2 can also be utilized to construct asymptotic tests for a given data set of fixed size. Recently, there has been an article by Lee, Ha, Na and Na (2003), who used a weak invariance principle to test for the stability of a certain general parameter vector determining the underlying time series. In particular, they derive a test which is able to detect changes in all of the parameters $\varphi$, $\omega^2$ and $\sigma^2$ of an RCA(1) time series at the same time. Here, we shall focus on the testing for a change in the mean and for the stability of the deterministic component $\varphi$ of the coefficients.

Let $Y_1, \ldots, Y_m$ be observations of the random variables

$$Y_n = \begin{cases} \mu + X_n & : \quad n = 1, \ldots, k^*, \\ \mu + \Delta_m + X_n & : \quad n = k^* + 1, \ldots, m, \end{cases}$$

where $\mu$, $k^*$ and $\Delta_m$ are unknown parameters and $\{X_n\}_{n \in \mathbb{Z}}$ is an RCA(1) time series. Instead of a sequential monitoring scheme, we shall apply a test procedure based on a fixed data set of $m$ observations. Again, we are interested in testing the change in the mean hypotheses

$$\begin{aligned} \mathcal{H}_{\mathbf{0}} &: \quad \Delta_m = 0, \\ \mathcal{H}_{\mathbf{A}} &: \quad \Delta_m \neq 0, \quad k^* < m. \end{aligned}$$

The test is based on the CUSUM

$$Q(m,k) = S_k - \frac{k}{m} S_m = \sum_{i=1}^{k} Y_i - \frac{k}{m} \sum_{i=1}^{m} Y_i. \tag{3.11}$$

But instead of (3.11), we will consider the functional versions

$$T_m(t) = S_{[mt]} - tS_m = \sum_{i=1}^{[mt]} Y_i - t \sum_{i=1}^{m} Y_i \qquad (t \in [0,1]),$$

where $[\cdot]$ denotes the integer part.

**Theorem 3.4.1 (Asymptotic under the null hypothesis)**
*Let $\{X_n\}_{n \in \mathbb{Z}}$ be an RCA(1) time series with (i), (ii) and let*

$$E|e_1|^\kappa < \infty \qquad \text{and} \qquad E|\varphi + b_1|^\kappa < 1$$

*for some $\kappa > 2$. Then, under $\mathcal{H}_{\mathbf{0}}$,*

$$\begin{aligned} \sup_{t \in [0,1]} \frac{|T_m(t)|}{\sqrt{m}\sigma_S} &\xrightarrow{\mathcal{D}} \sup_{t \in [0,1]} |B(t)|, \\ \sup_{t \in [0,1]} \frac{T_m(t)}{\sqrt{m}\sigma_S} &\xrightarrow{\mathcal{D}} \sup_{t \in [0,1]} B(t) \end{aligned}$$

*as $m \to \infty$, where $\{B(t) : 0 \leq t \leq 1\}$ denotes a Brownian bridge and $\sigma_S^2$ is given in Theorem 3.2.2.*

**Proof:** It follows from Theorem 3.2.2, that there exist a Wiener process $\{W(t) : t \geq 0\}$ and a $\nu > 2$ such that

$$\sum_{i=1}^{k} X_i \;=\; \sigma_S W(k) + \mathcal{O}\left(k^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

as $k \to \infty$. Hence,

$$\begin{aligned}
&\sup_{t\in[0,1]} \left| \frac{|T_m(t)|}{\sqrt{m}} - \frac{\sigma_S}{\sqrt{m}} \left|W([mt]) - tW(m)\right| \right| \\
&= \mathcal{O}\left( \sup_{t\in[0,1]} \frac{[mt]^{\frac{1}{\nu}} + tm^{\frac{1}{\nu}}}{\sqrt{m}} \right) \\
&= \mathcal{O}\left( m^{\frac{1}{\nu}-\frac{1}{2}} \right) \\
&= o(1) \qquad \text{a.s.}
\end{aligned}$$

as $m \to \infty$. Now,

$$\begin{aligned}
&\sup_{t\in[0,1]} \left| \frac{|W([mt]) - tW(m)|}{\sqrt{m}} - \frac{|W(mt) - tW(m)|}{\sqrt{m}} \right| \\
&\leq \sup_{t\in[0,1]} \frac{|W([mt]) - W(mt)|}{\sqrt{m}} \\
&= o(1) \qquad \text{a.s.}
\end{aligned}$$

as $m \to \infty$, since the order of the increments of the Wiener process can be estimated appropriately (cf. Csörgő and Réveśz (1981), Theorem 1.2.1 and Lemma 1.2.1). Finally, the scale transformation yields

$$\left\{ \frac{1}{\sqrt{m}} \left(W(mt) - tW(m)\right) \right\} \stackrel{\mathcal{D}}{=} \{W(t) - tW(1)\} \stackrel{\mathcal{D}}{=} \{B(t)\},$$

where $\{B(t) : 0 \leq t \leq 1\}$ denotes a Brownian bridge. The second claim is proved in a similar way and hence is omitted. □

As in the previous section, $\sigma_S^2$ can be replaced by the consistent estimator $\hat{\sigma}_{S,m}^2$ defined in (3.7) without losing the convergence results of Theorem 3.4.1.

**Corollary 3.4.1**

*Let the assumptions of Theorem 3.4.1 be satisfied. Then, under $\mathcal{H}_{\mathbf{0}}$,*

$$\begin{aligned}
\sup_{t\in[0,1]} \frac{|T_m(t)|}{\sqrt{m}\hat{\sigma}_{S,m}} &\stackrel{\mathcal{D}}{\longrightarrow} \sup_{t\in[0,1]} |B(t)|, \\
\sup_{t\in[0,1]} \frac{T_m(t)}{\sqrt{m}\hat{\sigma}_{S,m}} &\stackrel{\mathcal{D}}{\longrightarrow} \sup_{t\in[0,1]} B(t)
\end{aligned}$$

*as $m \to \infty$, where $\hat{\sigma}_{S,m}^2$ is defined in (3.7).*

**Proof:** The assertion follows from Theorem 3.4.1 and Lemma 3.3.1. □

The final paragraph is devoted to testing the constancy of $\varphi$ via the least squares estimator $\hat{\varphi}_m = \hat{\varphi}_{m,L}$ in (3.8). The estimator itself does not satisfy the conditions of Theorem 3.2.1, as is easily checked. But in Section 3.2, we have already seen that the sequence

$$Z_n = X_{n-1}(X_n - \varphi X_{n-1}) \qquad (n \in \mathbb{Z})$$

does. Furthermore,

$$\begin{aligned} \hat{\varphi}_m - \varphi &= \left(\sum_{i=1}^m X_{i-1}^2\right)^{-1} \sum_{i=1}^m X_i X_{i-1} - \left(\sum_{i=1}^m X_{i-1}^2\right)^{-1} \sum_{i=1}^m \varphi X_{i-1}^2 \\ &= \left(\sum_{i=1}^m X_{i-1}^2\right)^{-1} \sum_{i=1}^m X_{i-1}(X_i - \varphi X_{i-1}) \\ &= \left(\sum_{i=1}^m X_{i-1}^2\right)^{-1} \sum_{i=1}^m (X_{i-1}^2 b_i + X_{i-1} e_i). \end{aligned}$$

Hence, a test statistic can be obtained through imitating the CUSUM procedure for a change in the mean by comparing the estimators $\hat{\varphi}_k$ and $\hat{\varphi}_m$ or by using the functional version

$$U_m(t) = [mt]\left(\hat{\varphi}_{[mt]} - \hat{\varphi}_m\right) \qquad (t \in [0,1]),$$

where $[\cdot]$ denotes the integer part. We are able to proof the following theorem.

**Theorem 3.4.2 (Asymptotic under the null hypothesis)**
*Let $\{X_n\}_{n\in\mathbb{Z}}$ be an RCA(1) time series with (i),(ii) and let*

$$E|e_1|^\kappa < \infty \qquad \text{and} \qquad E|\varphi + b_1|^\kappa < 1$$

*for some $\kappa > 4$. Then, under $\mathcal{H}_\mathbf{0}$,*

$$\sup_{t\in[0,1]} \frac{|U_m(t)|}{\sqrt{m}\sigma_U} \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |B(t)|,$$

$$\sup_{t\in[0,1]} \frac{U_m(t)}{\sqrt{m}\sigma_U} \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} B(t),$$

*as $m \to \infty$, where $\{B(t) : 0 \le t \le 1\}$ denotes a Brownian bridge and*

$$\sigma_U = \frac{\sigma_R}{EX_1^2}.$$

**Proof:** We have

$$\begin{aligned}
&\sup_{t\in[0,1]}\left|\frac{U_m(t)}{\sqrt{m}\sigma_U}-\frac{1}{\sqrt{m}}(W([mt])-tW(m))\right| \\
\leq\ &\sup_{t\in[0,1]}\left|\frac{U_m(t)}{\sqrt{m}\sigma_U}-\frac{1}{\sqrt{m}\sigma_R}\left(R_{[mt]}-\frac{[mt]}{m}R_m\right)\right| \\
&\qquad+\sup_{t\in[0,1]}\left|\frac{1}{\sqrt{m}\sigma_R}\left(R_{[mt]}-\frac{[mt]}{m}R_m\right)-\frac{1}{\sqrt{m}}(W([mt])-tW(m))\right| \\
=\ &K_1+K_2.
\end{aligned}$$

Clearly, $K_2=o(1)$ a.s. as $m\to\infty$ by Theorem 3.2.3 (cf. also the proof of Theorem 3.4.1). Moreover,

$$\begin{aligned}
K_1 &= \sup_{t\in[0,1]}\left|\frac{U_m(t)}{\sqrt{m}\sigma_U}-\frac{1}{\sqrt{m}\sigma_R}\left(R_{[mt]}-\frac{[mt]}{m}R_m\right)\right| \\
&\leq \sup_{t\in[0,1]}\left|\frac{1}{\sqrt{m}\sigma_R}\left(R_{[mt]}-\frac{[mt]}{m}R_m\right)\left(\frac{EX_1^2}{\frac{1}{m}\sum_{i=1}^m X_{i-1}^2}-1\right)\right| \\
&= o_P(1)
\end{aligned}$$

as $m\to\infty$, since $\{R_n\}_{n\in\mathbb{N}}$ is a square integrable martingale with respect to the filtration $\{\mathcal{F}_n\}_{n\in\mathbb{N}}$ and

$$\frac{1}{m}\sum_{i=1}^m X_{i-1}^2\longrightarrow EX_1^2 \qquad \text{a.s.}$$

as $m\to\infty$ by the ergodic theorem (cf. Feigin and Tweedie (1985)). □

Finally, $\sigma_U$ can be replaced by consistent estimators, but the estimation procedure ist tedious because estimators of all moments up to fourth order are involved.

# Chapter 4

# Maximum Approximations

The present chapter is devoted to the maxima of partial sums of random vectors with non–zero mean or – in greater generality – to suprema of multivariate stochastic processes having a drift term. First results for maxima of partial sums of random variables with positive expectation are due to Teicher (1973), who proved a central limit theorem. Furthermore, he could show that the problem of determining the maximum or its limiting distribution is closely related to the normality of stopping rules appearing in sequential testing (see Siegmund (1968)). So, the results presented are not only interesting for their own sake, but also for possible applications in statistics – for instance in a multivariate modification of the model introduced in Chapter 2.

The chapter is organized as follows. In Section 4.1, we discuss the results preceding the paper of Berkes and Horváth (2003). Their generalization to suprema of real–valued stochastic processes with positive drift will be stated in Section 4.2, where we present the outset of our considerations, too.

Section 4.3 contains the corresponding results in the multi–dimensional case. Similarly, we define stochastic processes with linear drift and values in $\mathbb{R}^p$ satisfying a strong invariance principle. It turns out that the approximation rates can be established by an adaptation of the proof in the scalar case if all drift components are non–zero. To obtain the same rate in the general case, a different method is necessary. Moreover, we give an approximation of the Euclidean norm of our stochastic processes – itself regarded as a scalar stochastic process – in terms of a scalar Wiener process with drift (cf. Subsection 4.3.2). Finally, we derive the law of the iterated logarithm and some weak convergence theorems (cf. Subsection 4.3.3) for the considered vector–valued processes.

The results of Section 4.3 have been proposed by Lajos Horváth from the University of Utah and have already been published as a joint article (see Aue and Horváth (2003b)).

## 4.1 Introduction

Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables on some $(\Omega, \mathcal{A}, P)$ with $EX_1 = \mu > 0$ and $0 < \sigma^2 = \operatorname{Var} X_1 < \infty$. We are interested in the maxima of partial sums

$$S_0 = 0, \qquad S_n = X_1 + \ldots + X_n \qquad (n \in \mathbb{N}).$$

More exactly, we consider the random variables $\max_{j=1,\ldots,n} j^{-\alpha} S_j$, where $\alpha \in [0,1)$. We start with a review of existing results.

**Maxima and stopping rules.** Teicher (1973) was one of the first studying maxima of partial sums of independent random variables. He could prove a central limit theorem for these partial sums without using reflection or invariance principles, the latter appearing for the first time in an article by Erdős and Kac (1946). Teicher's proof showed that the central limit theorem is strongly related to first passage times studied in Siegmund (1968).

**Theorem 4.1.1 (Teicher)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with $EX_1 = \mu > 0$ and $0 < \sigma^2 = \operatorname{Var} X_1 < \infty$. Then,*

$$\lim_{n\to\infty} P\left\{\max_{j=1,\ldots,n} \frac{S_j}{j^\alpha} - \mu n^{1-\alpha} \le x\sigma n^{\frac{1}{2}-\alpha}\right\} = \Phi(x),$$

*where $\alpha \in [0,1)$ and $\Phi$ denotes the standard normal distribution function.*

**Proof:** See Teicher (1973). □

The connection to stopping rules or first passage times is the following. Set

$$\tau_c = \min\left\{j \ge 1 : \frac{S_j}{j^\alpha} > c\right\},$$

where $c > 0$, then Teicher's proof used the fact

$$P\left\{\max_{j=1,\ldots,n} \frac{S_j}{j^\alpha} > c\right\} = P\left\{\tau_c \le n\right\}. \tag{4.1}$$

So studying the limiting distribution of the maximum on the left hand side of equation (4.1) is simply the same as determining the asymptotics of $\tau_c$.

In more generality, Siegmund (1968) obtained the asymptotic normality for a larger class of (one–sided) stopping rules. Let $f_c$ be a concave function defined on $\mathbb{R}_+$ such that $f_c \to \infty$ as $c \to \infty$. Set

$$\tilde{\tau}_c = \min\left\{j \ge 1 : S_j > f_c(j)\right\}.$$

The special choice $f_c(j) = cj^\alpha$ with $\alpha \in [0,1)$ provides Teicher's result. We have already discussed stopping rules of this kind in Chapter 2, where we derived the asymptotic normality for a first passage time, given a special boundary function. Moreover, the stopping times $\tau_c$ or $\tilde{\tau}_c$ are of interest in renewal theory (see Gut (1974)) as well as in a variety of other sequential applications (cf. for example, Bhattacharya and Mallik (1973), Cabilio (1977) or Robbins and Siegmund (1974)).

Now, via the classical central limit theorem, also

$$\lim_{n\to\infty} P\left\{\frac{S_n}{n^\alpha} - \mu n^{1-\alpha} \le x\sigma n^{\frac{1}{2}-\alpha}\right\} = \Phi(x).$$

Comparing the latter with the statement of Theorem 4.1.1 leads to the conjecture

$$\max_{j=1,\ldots,n} \frac{S_j}{j^\alpha} - \frac{S_n}{n^\alpha} = o_P\left(n^{\frac{1}{2}-\alpha}\right) \tag{4.2}$$

as $n \to \infty$. In the next paragraph, we deal with the conjecture (4.2) for more general processes, but typically under more restrictive assumptions. It turns out, that the rate even holds true with probability 1, implying that the random variables $j^{-\alpha}S_j$ attain their maximum near the endpoint $n$ of the underlying index set $\{1,\ldots,n\}$.

**Partial sums with errors.** Teicher's approach has been further extended to partial sums which can be observed with errors only. Results are due to Chow and Hsiung (1976), and Chow, Hsiung and Yu (1980). Therefore, let in addition $\{Y_n\}_{n\in\mathbb{N}}$ be real–valued random variables such that $Y_n = o(1)$ a.s. as $n \to \infty$ and let $\{a_n\}_{n\in\mathbb{N}}$ be a sequence of positive constants satisfying $n^{-\alpha}a_n \to 1$ as $n \to \infty$, where $\alpha > 0$. Put

$$U_n = \tfrac{1}{n}S_n + Y_n \qquad (n \in \mathbb{N}). \tag{4.3}$$

Then, Chow, Hsiung and Yu (1980) obtained the limiting normal distribution for the random variables

$$\begin{array}{ll} \max\limits_{j=1,\ldots,n} a_j U_j, & \inf\limits_{j\ge n} a_j U_j, \\ \min\limits_{j=1,\ldots,n} a_j^{-1} U_j, & \sup\limits_{j\ge n} a_j^{-1} U_j \end{array}$$

and the related stopping times

$$\begin{aligned} \tau_{1,c} &= \min\{n \ge n_c : U_n \ge (ca_n)^{-1}\}, \\ \tau_{2,c} &= \min\{n \ge n_c : 0 < U_n \le ca_n\}, \end{aligned}$$

where $n_c$ is allowed to depend on $c > 0$. Moreover, they could prove the following theorem.

**Theorem 4.1.2 (Chow, Hsiung and Yu)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with $EX_1 = \mu > 0$. Let $\{Y_n\}_{n\in\mathbb{N}}$ be random variables such that $Y_n = o(1)$ a.s. as $n \to \infty$. Moreover, choose positive constants $\alpha$, $\{a_n\}_{n\in\mathbb{N}}$ and $\{f_n\}_{n\in\mathbb{N}}$ satisfying*

[1] $a_n = n^\alpha(1 + o(1)) \qquad (n \to \infty)$,

[2] $(a_n - n^\alpha)\, f_n = a + o(1) \qquad (n \to \infty)$
*for some $a \in \mathbb{R}$ and*

$$\lim_{\beta\to 1} \limsup_{n\to\infty} \max\left\{\frac{f_j}{f_k} : n\beta \le j,\, k \le \frac{n}{\beta}\right\} = 1.$$

*Then, the following statements hold true.*

*a) If*

*(i)* $E|X_1|^\nu < \infty$ *for some* $\nu \in [1, 4)$,

*(ii)* $n^\alpha f_n Y_n = c\mu + o(1)$ a.s. $\qquad (n \to \infty)$,

*(iii)* $f_n = \mathcal{O}\left(n^{1-\alpha-\frac{1}{\nu}}\right)$ a.s. $\qquad (n \to \infty)$,

*then as $n \to \infty$, we get a.s.*

$$\begin{aligned}
\max_{j=1,\dots,n} (a_j U_j - a_n U_n)\, f_n &= o(1), \\
\sup_{j\ge n} (a_n U_n - a_j U_j)\, f_n &= o(1), \\
\max_{n_0\le j\le n} (a_j U_n - a_n U_j)\, f_n &= o(1), \\
\sup_{j\ge n} (a_n U_j - a_j U_n)\, f_n &= o(1),
\end{aligned}$$

*where $n_0$ is a random variable such that $U_n > 0$ for $n \ge n_0$.*

*b) Let $X_1 = \mu$ a.s., let $\{\varepsilon_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with $E\varepsilon_1 = 0$ and $E|\varepsilon_1|^\nu < \infty$ for some $\nu \ge 2$. Set*

$$Y_n = Y_n' - \left(\frac{1}{n}\sum_{i=1}^n \varepsilon_i\right)^2$$

*with $n^\alpha f_n Y_n' = c\mu + o(1)$ a.s. as $n \to \infty$.*

*(i) If $f_n = \mathcal{O}\left(n^{\frac{3}{2}-\frac{1}{\nu}-\alpha}\right)$ as $n \to \infty$, then the results of a) hold in probability.*

(ii) *If* $f_n = \mathcal{O}\left(\frac{n^{\frac{3}{2}-\frac{1}{\nu}-\alpha}}{\sqrt{\log\log n}}\right)$ *as* $n \to \infty$, *then the results of a) hold a.s.*

c) *If* $f_n$ *is replaced by* $n^\alpha f_n a_j^{-1}$ *or by* $n^\alpha f_n a_n^{-1}$, *then results of a) and b) still hold true.*

**Proof:** See Chow, Hsiung and Yu (1980). □

The next result is due to Chow and Hsiung (1976) and is a specialization of the previous theorem. It gives the promised proof that the conjecture (4.2) holds true a.s.

**Theorem 4.1.3 (Chow and Hsiung)**
*Let* $\{X_n\}_{n\in\mathbb{N}}$ *be a sequence of independent, identically distributed random variables with* $EX_1 > 0$ *and* $\mathrm{Var}X_1 = \sigma^2 < \infty$. *Then,*

$$\max_{j=1,\ldots,n} \frac{S_j}{j^\alpha} - \frac{S_n}{n^\alpha} = o\left(n^{\frac{1}{2}-\alpha}\right) \qquad a.s.$$

*as* $n \to \infty$, *where* $\alpha \in [0, 1)$.

**Proof:** See Chow and Hsiung (1976). □

We close this section by an example which underlines the applicability of the above approach in sequential analysis.

**Example 4.1.1 (Sequential estimation)**
Let $\{Z_n\}_{n\in\mathbb{N}}$ be independent, identically distributed random variables with $EZ_1 = \tilde{\mu}$ and $0 < \tilde{\sigma}^2 = \mathrm{Var}\, Z_1 < \infty$. Define

$$\begin{aligned} \bar{Z}_n &= \frac{1}{n}(Z_1 + \ldots + Z_n) \qquad (n \in \mathbb{N}), \\ V_n &= \frac{1}{n}\sum_{i=1}^n \left(Z_i - \bar{Z}_n\right)^2 + b_n \qquad (n \in \mathbb{N}) \end{aligned}$$

with a sequence $\{b_n\}_{n\in\mathbb{N}}$ of constants satisfying $b_n \to 0$ as $n \to \infty$. Now, set

$$\begin{aligned} X_n &= (Z_n - \tilde{\mu})^2 \qquad (n \in \mathbb{N}), \\ R_n &= \frac{1}{n}\sum_{i=1}^n \left(Z_i - \bar{Z}_n\right)^2 + b_n \qquad (n \in \mathbb{N}). \end{aligned}$$

Then, $\{X_n\}_{n\in\mathbb{N}}$ is a sequence of independent, identically distributed random variables and by assumption on $\{b_n\}_{n\in\mathbb{N}}$, the random variables $\{R_n\}_{n\in\mathbb{N}}$ satisfy the condition $R_n = o(1)$ a.s. as $n \to \infty$, since the sample mean $\bar{Z}_n$ converges a.s. to the expectation $\mu$ by the strong law of large numbers. Furthermore, the sample variance can be rewritten as

$$V_n = \frac{1}{n}\sum_{i=1}^n X_i + R_n \qquad (n \in \mathbb{N}),$$

i.e., $V_n$ is of the form claimed in (4.3). Stopping times related to $V_n$ are often used in sequential analysis (see Robbins (1959) or the papers cited above).

The theorems of this section and mainly Theorem 4.1.3 can be derived from more general results, which were obtained by Berkes and Horváth (2003). Their setting will be stated in the following section.

## 4.2 Results in the scalar case

Like the classical limit theorems were extended to functional versions (see for example Strassen (1964) in case of the law of the iterated logarithm), Berkes and Horváth (2003) proved that also Theorem 4.1.3 can be generalized to suprema of any stochastic process which can be approximated by a Wiener process with a positive drift term under a certain rate. Their results are pictured in the following lines.

Let $\{\Gamma(t) : t \geq 1\}$ be a stochastic process such that there exists a Wiener process $\{W(t) : t \geq 1\}$ and positive constants $\delta$ and $\gamma$ with

$$\Gamma(t) - (\delta W(t) + \gamma t) = o\left(t^{\frac{1}{\nu}}\right) \qquad \text{a.s.} \tag{4.4}$$

as $t \to \infty$ with some $\nu > 2$.

**Theorem 4.2.1 (Berkes and Horváth)**
*Let $\{\Gamma(t) : t \geq 1\}$ be a stochastic process such that (4.4) is satisfied. Then,*

$$\sup_{1\leq t\leq T} \frac{\Gamma(t)}{t^\alpha} - \frac{\Gamma(T)}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \textit{a.s.},$$

$$\sup_{1\leq t\leq T} \frac{\Gamma(t)}{t^\alpha} - \frac{\delta W(T) + \gamma T}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \textit{a.s.}$$

*as $T \to \infty$, where $\alpha \in [0, 1)$.*

**Proof:** See Berkes and Horváth (2003). □

Under the additional and stronger moment condition on $X_1$ imposed by (4.4), the rate of convergence in Theorem 4.1.3 can be improved.

**Theorem 4.2.2 (Berkes and Horváth)**
*Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random variables with $EX_1 > 0$, $VarX_1 = \sigma^2 < \infty$ and $E|X_1|^\nu < \infty$ for some $\nu > 2$. Then,*

$$\max_{j=1,\ldots,n} \frac{S_j}{j^\alpha} - \frac{S_n}{n^\alpha} = o\left(n^{\frac{1}{\nu}-\alpha}\right) \qquad \textit{a.s.}$$

*as $n \to \infty$, where $\alpha \in [0, 1)$.*

**Proof:** See Berkes and Horváth (2003). □

Besides, Berkes and Horváth (2003) proved the law of the iterated logarithm for the supremum of $\{t^{-\alpha}\Gamma(t) : t \in [1,T]\}$ and obtained some weak convergence results in $\mathcal{D}[a,b]$ for suitable choices of $a$ and $b$. Confer Subsection 4.3.3 for further information.

## 4.3 Vector–valued processes

We are going to extend the results presented in the previous section to random vectors and vector–valued stochastic processes on $\mathbb{R}^p$. It turns out in Subsection 4.3.1 that the weighted suprema of the Euclidean norm of the processes will be attained near the end-point of the underlying index set.

In addition, Subsection 4.3.2 shows that even the Euclidean norm of a vector–valued Wiener process with linear drift can be approximated by a scalar Wiener process with a certain variance parameter.

Finally, Subsection 4.3.3 contains some corollaries which are easily obtained from the theorems proved before.

The main tool in the proofs to come is a strong invariance principle employed on the considered processes. Under the point of view of limit distributions (as for instance regarded in Chapter 2), it would have been sufficient to use weak approximations instead. But since we are not interested in exact asymptotics here, we derive the rates almost surely.

### 4.3.1 Approximations in Euclidean norm

As in the scalar case described in Section 4.2, we are interested in approximations of the weighted suprema of partial sums (in Euclidean norm). Again, a more general approach using an invariance principle is used to obtain our results.

Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of random vectors with values in $\mathbb{R}^p$ and

$$E\boldsymbol{X}_1 = \boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_p)' \quad \text{and} \quad \mathrm{Cov}\boldsymbol{X}_1 = \boldsymbol{\Sigma}, \tag{4.5}$$

where $\boldsymbol{\Sigma}$ denotes a positive definite matrix. We define the partial sums

$$\boldsymbol{S}_0 = 0, \qquad \boldsymbol{S}_n = \boldsymbol{X}_1 + \ldots + \boldsymbol{X}_n \qquad (n \in \mathbb{N}) \tag{4.6}$$

and the Euclidean norm of an element $\boldsymbol{z} = (z_1, \ldots, z_p)' \in \mathbb{R}^p$ as

$$\| \boldsymbol{z} \| = \sqrt{z_1^2 + \ldots + z_p^2}.$$

We will prove the following analogues of Theorems 4.2.1 and 4.2.2 in this section. The first theorem assumes all coordinates to have a non–zero drift, that is $\gamma_j \neq 0$ for all $j = 1, \dots, p$ in (4.5).

**Theorem 4.3.1**
*Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random vectors such that (4.5) holds and let $\alpha \in [0, 1)$. If*

- *a) $\gamma_j \neq 0$ for all $j = 1, \dots, p$ and*
- *b) $E \parallel \boldsymbol{X}_1 \parallel^\nu < \infty$ with some $\nu > 2$,*

*then*

$$\max_{1\le k\le m} \frac{\parallel \boldsymbol{S}_k \parallel}{k^\alpha} - \frac{\parallel \boldsymbol{S}_m \parallel}{m^\alpha} = o\left(m^{\frac{1}{\nu}-\alpha}\right) \qquad a.s.$$

*as $m \to \infty$, where $\{\boldsymbol{S}_k\}_{k\in\mathbb{N}_0}$ is defined in (4.6).*

Theorem 4.3.1 will be an immediate consequence of the methods used in Berkes and Horváth (2003). The second theorem only assumes that $\boldsymbol{\gamma} \neq \boldsymbol{0}$, i.e., $\parallel \boldsymbol{\gamma} \parallel$ is strictly positive, while no conditions are imposed on a single coordinate. Under this more general assumption the statement of Theorem 4.3.1 retains with the same upper bound for the rate of convergence. However, it will require a somewhat different proof.

**Theorem 4.3.2**
*Let $\{\boldsymbol{X}_n\}_{n\in\mathbb{N}}$ be a sequence of independent, identically distributed random vectors such that (4.5) holds and let $\alpha \in [0, 1)$. If*

- *a) $\parallel \boldsymbol{\gamma} \parallel > 0$ and*
- *b) $E \parallel \boldsymbol{X}_1 \parallel^\nu < \infty$ with some $\nu > 2$,*

*then*

$$\max_{1\le k\le m} \frac{\parallel \boldsymbol{S}_k \parallel}{k^\alpha} - \frac{\parallel \boldsymbol{S}_m \parallel}{m^\alpha} = o\left(m^{\frac{1}{\nu}-\alpha}\right) \qquad a.s.$$

*as $m \to \infty$, where $\{\boldsymbol{S}_k\}_{k\in\mathbb{N}_0}$ is defined in (4.6).*

The proofs of both theorems are postponed until the end of this subsection, since they will follow from a more general approach by an application of results due to Einmahl (1989).

Introduce a $p$–dimensional stochastic process $\{\boldsymbol{\Gamma}(t) : t \geq 1\}$ on $[1, \infty)$ with components $\Gamma_1(t), \dots, \Gamma_p(t)$. We impose a strong invariance principle.

**Assumption 4.3.1 (Strong invariance)**
*There is a p–dimensional Wiener process* $\{\boldsymbol{W}(t) : t \geq 1\}$ *with* $\boldsymbol{W}(t) = (W_1(t), \ldots, W_p(t))'$, *such that the following two conditions hold true:*

*a)* $Cov(\boldsymbol{W}(s), \boldsymbol{W}(t)) = \min\{s,t\}\boldsymbol{\Sigma}$,
*b)* $\| \boldsymbol{\Gamma}(t) - (\boldsymbol{W}(t) + t\boldsymbol{\gamma}) \| = o\left(t^{\frac{1}{\nu}}\right)$ *a.s. as* $t \to \infty$ *with some* $\nu > 2$.

The strong invariance principle stated in b) is satisfied for large classes of random vectors. We will not go into the details here, but will give references for further information. We can apply our theory for instance to

- independent random vectors (cf. Einmahl (1989) and in addition Theorems 4.3.1 and 4.3.2 of this section),
- random vectors satisfying a mixing condition (cf. Kuelbs and Philipp (1980)),
- vector–valued martingale differences (cf. Eberlein (1986)).

We will now restate Theorems 4.3.1 and 4.3.2 in terms of the process $\{\boldsymbol{\Gamma}(t) : t \geq 1\}$, i.e., we will give functional analogues. Again, we have to separate the cases

- $\gamma_j \neq 0$ for all $j = 1, \ldots, p$ and
- $\| \boldsymbol{\gamma} \| > 0$.

We start with the more restrictive first condition.

**Theorem 4.3.3**
*Let* $\{\boldsymbol{\Gamma}(t) : t \geq 1\}$ *be a p–dimensional stochastic process, such that Assumption 4.3.1 holds, and let* $\gamma_j \neq 0$ *for all* $j = 1, \ldots, p$. *Then,*

$$\sup_{1 \leq t \leq T} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^\alpha} - \frac{\| \boldsymbol{\Gamma}(T) \|}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad a.s.$$

$$\sup_{1 \leq t \leq T} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^\alpha} - \frac{\| \boldsymbol{W}(T) + T\boldsymbol{\gamma} \|}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad a.s.$$

*as* $T \to \infty$, *where* $\alpha \in [0, 1)$.

**Proof:** The proof of the theorem is organized as follows. At first, we will divide the process $\{\boldsymbol{\Gamma}(t) : t \geq 1\}$ into the parts with positive and negative drift. Moreover, if the drift term is negative, the supremum of $\Gamma_i(t)$ on $[1, T]$ is asymptotically smaller than the supremum of $-\Gamma_i(t)$ taken over the same interval. By invariance, we can approximate both $\{\Gamma_i(t) : t \geq 1\}$ and $\{-\Gamma_i(t) : t \geq 1\}$ by scalar Wiener processes with positive drift and are therefore able to apply Theorem 4.2.1. Some more technical calculations are needed to establish the upper bound for the rate of convergence. The details are addressed below.

a) Since all drift components $\gamma_j$ are non–zero by assumption, we assume without loss of generality that

$$\gamma_1, \ldots, \gamma_{p(1)} > 0, \qquad \gamma_{p(1)+1}, \ldots, \gamma_p < 0,$$

where $0 \leq p(1) \leq p$. (If $p(1) = 0$ or $= p$, respectively, all coordinates have a strictly negative or positive drift parameter.) Consider the Euclidean norm of $T^{-\alpha} \,\|\, \mathbf{\Gamma}(T) \,\|$. Then,

$$\begin{aligned}
\frac{\|\, \mathbf{\Gamma}(T) \,\|}{T^\alpha} &\leq \sup_{1 \leq t \leq T} \frac{\|\, \mathbf{\Gamma}(t) \,\|}{t^\alpha} \\
&\leq \left( \sum_{i=1}^{p(1)} \left( \sup_{1 \leq t \leq T} \frac{|\Gamma_i(t)|}{t^\alpha} \right)^2 + \sum_{i=p(1)+1}^{p} \left( \sup_{1 \leq t \leq T} \frac{|\Gamma_i(t)|}{t^\alpha} \right)^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

b) By the law of the iterated logarithm for Wiener processes we have

$$\limsup_{T \to \infty} \frac{1}{\sqrt{T \log\log T}} \sup_{1 \leq t \leq T} \|\, \boldsymbol{W}(t) \,\| = C \qquad \text{a.s.}, \tag{4.7}$$

where $C$ is a constant. So, the drift term is the dominating one. Hence, there is a random variable $T_1$ such that

$$\begin{aligned}
&\left( \sum_{i=1}^{p(1)} \left( \sup_{1 \leq t \leq T} \frac{|\Gamma_i(t)|}{t^\alpha} \right)^2 + \sum_{i=p(1)+1}^{p} \left( \sup_{1 \leq t \leq T} \frac{|\Gamma_i(t)|}{t^\alpha} \right)^2 \right)^{\frac{1}{2}} \\
&= \mathcal{O}(1) \left( \sum_{i=1}^{p(1)} \left( \sup_{1 \leq t \leq T} \frac{\Gamma_i(t)}{t^\alpha} \right)^2 + \sum_{i=p(1)+1}^{p} \left( \sup_{1 \leq t \leq T} \frac{-\Gamma_i(t)}{t^\alpha} \right)^2 \right)^{\frac{1}{2}}
\end{aligned}$$

for all $T \geq T_1$. Now, $\{-\Gamma_i(t) : t \geq 1\}$ is a scalar stochastic process which can be approximated with the scalar process $\{-W_i(t) - \gamma_i t : t \geq 1\}$ $(i > p(1))$. For it is well–known that $\{-W(t)\}$ is a Wiener process in case $\{W(t)\}$ is, an application of Theorem 4.2.1 yields

$$\begin{aligned}
&\left( \sum_{i=1}^{p(1)} \left( \sup_{1 \leq t \leq T} \frac{\Gamma_i(t)}{t^\alpha} \right)^2 + \sum_{i=p(1)+1}^{p} \left( \sup_{1 \leq t \leq T} \frac{-\Gamma_i(t)}{t^\alpha} \right)^2 \right)^{\frac{1}{2}} \\
&= \left( \sum_{i=1}^{p(1)} \left( \frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu} - \alpha}\right) \right)^2 + \sum_{i=p(1)+1}^{p} \left( -\frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu} - \alpha}\right) \right)^2 \right)^{\frac{1}{2}}
\end{aligned}$$

a.s. as $T \to \infty$.

c) Next, we use the mean value theorem for the mapping $f(x) = \sum_{i=1}^{p} x_i^2$ with $x \in \mathbb{R}^p$. Then,

$$\begin{aligned}
&\left| \left( \sum_{i=1}^{p(1)} \left( \frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 + \sum_{i=p(1)+1}^{p} \left( -\frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 \right)^{\frac{1}{2}} \right. \\
&\qquad\qquad \left. - \left( \sum_{i=1}^{p} \left( \frac{\Gamma_i(T)}{T^\alpha} \right)^2 \right)^{\frac{1}{2}} \right| \\
&= o\left( T^{\frac{1}{\nu}-\alpha} \sum_{i=1}^{p} \frac{\Gamma_i(T)}{T^\alpha} T^{\alpha-1} \right) \\
&= o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$, where the last equality follows from (4.7) and once more from the fact that the drift term is the dominating part.

On combining a)–c), we arrive at

$$\frac{\| \mathbf{\Gamma}(T) \|}{T^\alpha} \leq \sup_{1 \leq t \leq T} \frac{\| \mathbf{\Gamma}(t) \|}{t^\alpha} \leq \frac{\| \mathbf{\Gamma}(T) \|}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.}$$

as $T \to \infty$ and the first assertion of Theorem 4.3.3 follows. The second statement is an easy consequence of Assumption 4.3.1b), since

$$\begin{aligned}
&\left| \frac{\| \mathbf{\Gamma}(T) \|}{T^\alpha} - \frac{\| \boldsymbol{W}(T) + T\boldsymbol{\gamma} \|}{T^\alpha} \right| \\
&\leq \frac{1}{T^\alpha} \| \mathbf{\Gamma}(T) - (\boldsymbol{W}(T) + T\boldsymbol{\gamma}) \| = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$. This completes the proof. □

If just $\| \boldsymbol{\gamma} \| > 0$ is assumed, we arrive at the same upper bound for the rate of convergence already established in Theorem 4.3.2 above.

**Theorem 4.3.4**
*Let $\{\mathbf{\Gamma}(t) : t \geq 1\}$ be a $p$–dimensional stochastic process, such that Assumption 4.3.1 holds and let $\| \boldsymbol{\gamma} \| > 0$. Then,*

$$\sup_{1 \leq t \leq T} \frac{\| \mathbf{\Gamma}(t) \|}{t^\alpha} - \frac{\| \mathbf{\Gamma}(T) \|}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.,}$$

$$\sup_{1 \leq t \leq T} \frac{\| \mathbf{\Gamma}(t) \|}{t^\alpha} - \frac{\| \boldsymbol{W}(T) + T\boldsymbol{\gamma} \|}{T^\alpha} = o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.}$$

*as $T \to \infty$, where $\alpha \in [0, 1)$.*

**Proof:** Since some of the drift components are allowed to be zero by assumption, Theorem 4.3.4 is not as easy traced back to the scalar situation as in Theorem 4.3.3. In a first step it is shown that the location of the largest value of $t^{-\alpha} \parallel \mathbf{\Gamma}(t) \parallel$ on the interval $[1, T]$ is close to the right endpoint $T$ up to some asymptotic rate specified below. In a second step, we investigate the components without drift and in a third part those with drift.

We assume without loss of generality

$$\gamma_1, \ldots, \gamma_{p(1)} > 0, \qquad \gamma_{p(1)+1}, \ldots, \gamma_{p(1)+p(2)} < 0, \qquad \gamma_{p(1)+p(2)+1}, \ldots, \gamma_p = 0$$

and set $q = p(1) + p(2)$.

a) Our first aim is to examine the order of the difference $T - \eta(T)$, where $\eta(T)$ is defined as the location of the largest value of $t^{-\alpha} \parallel \mathbf{\Gamma}(t) \parallel$ on $[1, T]$, i.e.,

$$\eta(T) = \sup \left\{ t \in [1, T] \, : \, \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} = \sup_{1 \le s \le T} \frac{\parallel \mathbf{\Gamma}(s) \parallel}{s^\alpha} \right\}.$$

By the law of the iterated logarithm for Wiener processes

$$\limsup_{T \to \infty} \frac{1}{\sqrt{T \log \log T}} \sup_{1 \le t \le T} \parallel \boldsymbol{W}(t) \parallel = C \qquad \text{a.s.,} \tag{4.8}$$

where $C$ is a positive constant. Now, (4.8) and the strong invariance of Assumption 4.3.1b) yield

$$\begin{aligned}
& \sup_{1 \le t \le T - c\sqrt{T \log \log T}} \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} \\
= & \sup_{1 \le t \le T - c\sqrt{T \log \log T}} \frac{\parallel \boldsymbol{W}(t) + t\boldsymbol{\gamma} \parallel}{t^\alpha} + o\left(T^{\frac{1}{\nu} - \alpha}\right) \\
\le & \sup_{1 \le t \le T} \frac{\parallel \boldsymbol{W}(t) \parallel}{t^\alpha} + \sup_{1 \le t \le T - c\sqrt{T \log \log T}} \frac{\parallel t\boldsymbol{\gamma} \parallel}{t^\alpha} + o\left(T^{\frac{1}{\nu} - \alpha}\right) \\
= & \, C \sup_{1 \le t \le T} \frac{\sqrt{t \log \log t}}{t^\alpha} + \parallel \boldsymbol{\gamma} \parallel \left(T - c\sqrt{T \log \log T}\right)^{1-\alpha} + o\left(T^{\frac{1}{\nu} - \alpha}\right) \\
\le & \left(T - c_1\sqrt{T \log \log T}\right)^{1-\alpha} \qquad \text{a.s.}
\end{aligned} \tag{4.9}$$

for all $T \ge T_0$ if $c_1 = \frac{c}{2}$ and $c$ is a constant chosen large enough. On the other hand by similar arguments

$$\sup_{1 \le t \le T} \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} \ge \parallel \boldsymbol{\gamma} \parallel T^{1-\alpha} - 2C \frac{\sqrt{T \log \log T}}{T^\alpha} \qquad \text{a.s.} \tag{4.10}$$

for all $T \geq T_0$. Now, on choosing $c$ large enough, the upper bound for the supremum over the restricted range $[1, T - c\sqrt{T \log\log T}]$ in (4.9) is smaller than the lower bound for the supremum over the complete interval $[1, T]$ in (4.10). Hence,

$$T - \eta(T) = \mathcal{O}\left(\sqrt{T \log\log T}\right) \qquad \text{a.s.}$$

as $T \to \infty$.

b) Similarly as in the proof of Theorem 4.3.3 we obtain

$$\begin{aligned}
\frac{\|\mathbf{\Gamma}(T)\|}{T^\alpha} &\leq \sup_{1 \leq t \leq T} \frac{\|\mathbf{\Gamma}(t)\|}{t^\alpha} \\
&= \frac{\|\mathbf{\Gamma}(\eta(T))\|}{\eta^\alpha(T)} \\
&\leq \frac{1}{\eta^\alpha(T)} \left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(\eta(T)) \right)^{\frac{1}{2}} \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$, where the sum on the right hand side has been split up according to existing or non–existing drifts. We consider the second sum first. Observe that for any $q < i \leq p$ it holds by Csörgő and Révész (1981), Theorem 1.2.1

$$\begin{aligned}
&\left|\Gamma_i^2(\eta(T)) - \Gamma_i^2(T)\right| \\
&= |\Gamma_i(\eta(T)) - \Gamma_i(T)| \, |\Gamma_i(\eta(T)) + \Gamma_i(T)| \\
&= \left|W_i(\eta(T)) - W_i(T) + o\left(T^{\frac{1}{\nu}}\right)\right| \left|W_i(\eta(T)) + W_i(T) + o\left(T^{\frac{1}{\nu}}\right)\right| \\
&= \mathcal{O}\left((T \log\log T)^{\frac{1}{4}} \sqrt{\log T} \sqrt{T \log\log T}\right) \\
&= \mathcal{O}\left(T^{\frac{3}{4}} \sqrt{\log T} (\log\log T)^{\frac{3}{4}}\right) \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$, which establishes an upper bound for the components without drift. In view of this estimation it is already possible to replace $\eta(T)$ by $T$ for the zero drift components. In detail, using

$$\sqrt{x+h} = \sqrt{x} + \frac{h}{2\sqrt{\xi}}, \qquad \xi \in (x, x+h), \tag{4.11}$$

we get

$$\frac{1}{\eta^\alpha(T)} \left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(\eta(T)) \right)^{\frac{1}{2}}$$

$$
\begin{aligned}
&= \frac{1}{\eta^\alpha(T)} \left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(T) + \mathcal{O}\left(T^{\frac{3}{4}}\sqrt{\log T}(\log\log T)^{\frac{3}{4}}\right)\right)^{\frac{1}{2}} \\
&= \frac{1}{\eta^\alpha(T)} \left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(T)\right)^{\frac{1}{2}} + \mathcal{O}\left(T^{-\frac{1}{4}-\alpha}\sqrt{\log T}(\log\log T)^{\frac{3}{4}}\right)
\end{aligned}
$$

a.s. as $T \to \infty$, since

$$
\liminf_{t\to\infty} \frac{1}{T}\left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(T)\right) > 0 \qquad \text{a.s.}
$$

c) Using Theorem 4.2.1 for the components with drift, we get that

$$
\begin{aligned}
&\frac{1}{\eta^\alpha(T)} \left( \sum_{i=1}^{q} \Gamma_i^2(\eta(T)) + \sum_{i=q+1}^{p} \Gamma_i^2(T)\right)^{\frac{1}{2}} \\
&= \left( \sum_{j=1}^{q} \frac{\Gamma_i^2(\eta(T))}{\eta^{2\alpha}(T)} + \sum_{i=q+1}^{p} \frac{\Gamma_i^2(T)}{\eta^{2\alpha}(T)}\right)^{\frac{1}{2}} \\
&= \left( \sum_{i=1}^{q} \left(\frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 + \sum_{i=q+1}^{p} \frac{\Gamma_i(T)}{\eta^{2\alpha}(T)}\right)^{\frac{1}{2}} \qquad \text{a.s.} \qquad (4.12)
\end{aligned}
$$

as $T \to \infty$. We still have to replace $\eta(T)$ by $T$ in the second sum on the right hand side in (4.12). This will be done by an application of the law of the iterated logarithm and the mean value theorem applied to $f(x) = x^{-2\alpha}$. Since

$$
\eta(T) = T(1 + o(1)) \qquad \text{a.s.}
$$

as $T \to \infty$ by part a), we can proceed by recognizing $\xi \in (\eta(T), T)$ with

$$
\begin{aligned}
&\left|\frac{1}{\eta^{2\alpha}(T)} - \frac{1}{T^{2\alpha}}\right| \sum_{i=q+1}^{p} \Gamma_i^2(T) \\
&= \frac{2\alpha(T-\eta(T))}{|\xi^{2\alpha+1}|} \sum_{i=q+1}^{p} \Gamma_i^2(T) \\
&= \mathcal{O}\left(\frac{\sqrt{T\log\log T}}{T^{2\alpha+1}} T\log\log T\right) \\
&= \mathcal{O}\left(T^{\frac{1}{2}-2\alpha}(\log\log T)^{\frac{3}{4}}\right) \qquad \text{a.s.}
\end{aligned}
$$

as $T \to \infty$. Now by (4.11),

$$\left( \sum_{i=1}^{q} \left( \frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 + \sum_{i=q+1}^{p} \frac{\Gamma_i^2(T)}{T^{2\alpha}} + \mathcal{O}\left(T^{\frac{1}{2}-2\alpha}(\log\log T)^{\frac{3}{4}}\right)\right)^{\frac{1}{2}}$$

$$= \left( \sum_{i=1}^{q} \left( \frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 + \sum_{i=q+1}^{p} \frac{\Gamma_i^2(T)}{T^{2\alpha}} \right)^{\frac{1}{2}} + \mathcal{O}\left(T^{-\frac{1}{2}-\alpha}(\log\log T)^{\frac{3}{4}}\right)$$

a.s. as $T \to \infty$. As in the proof of Theorem 4.3.3, we finally arrive at

$$\left| \left( \sum_{i=1}^{q} \left( \frac{\Gamma_i(T)}{T^\alpha} + o\left(T^{\frac{1}{\nu}-\alpha}\right)\right)^2 + \sum_{i=q+1}^{p} \frac{\Gamma_i^2(T)}{T^{2\alpha}} \right)^{\frac{1}{2}} - \left( \sum_{i=1}^{p} \frac{\Gamma_i^2(T)}{T^{2\alpha}} \right)^{\frac{1}{2}} \right|$$

$$= o\left(T^{\frac{1}{\nu}-\alpha}\right) \qquad \text{a.s.}$$

as $T \to \infty$.

Putting together parts a)–c), the proof is complete. □

Berkes and Horváth (2003) approximated the Gaussian process with drift with a sum of specially constructed random variables such that the corresponding weighted sum becomes an increasing function of time. We used a different method in the previous proof. However, their method applies to the vector valued case if $\alpha = 0$.

**Theorem 4.3.5**
*Let $\{\mathbf{\Gamma}(t) : t \geq 1\}$ be a $p$–dimensional stochastic process, such that Assumption 4.3.1 holds. Then,*

$$\sup_{1\leq t\leq T} \| \mathbf{\Gamma}(t) \| - \| \mathbf{\Gamma}(T) \| = o\left(T^{\frac{1}{\nu}}\right) \qquad \textit{a.s.},$$

$$\sup_{1\leq t\leq T} \| \mathbf{\Gamma}(t) \| - \| \boldsymbol{W}(T) + \boldsymbol{\gamma} T \| = o\left(T^{\frac{1}{\nu}}\right) \qquad \textit{a.s.}$$

*as $T \to \infty$.*

The proof of Theorem 4.3.5 is based on the following lemma. Let $X$ be a uniformly distributed random variable on some interval $[a, b] \subset \mathbb{R}$. Then, mean and variance of $X$ are given by

$$EX = \tfrac{1}{2}(b + a), \qquad \text{Var}X = \tfrac{1}{12}(b - a)^2.$$

Moreover, $X$ is standardized if $b = -a = \sqrt{3}$.

In the next lemma, we will approximate the supremum of $\| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \|$ on $[1, T]$ by a sequence of $p$–dimensional random variables whose components are independent and uniformly distributed on $[-\sqrt{3}, \sqrt{3}]$.

**Lemma 4.3.1**
*Let* $\{\boldsymbol{W}(t) : t \geq 1\}$ *be a* $p$*–dimensional Wiener process on* $[1, \infty)$ *with*

$$Cov(\boldsymbol{W}(s), \boldsymbol{W}(t)) = \min\{s, t\} \boldsymbol{\Sigma} \qquad (s, t \geq 1).$$

*Then,*

$$\sup_{1 \leq t \leq T} \| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \| - \| \boldsymbol{W}(T) + \boldsymbol{\gamma} T \| = \mathcal{O}(\log T) \qquad a.s.$$

*as* $T \to \infty$.

**Proof:** As before, we assume that all non–zero drift terms are positive. Let

$$\gamma_1, \ldots, \gamma_{p(1)} > 0, \qquad \gamma_{p(1)+1}, \ldots, \gamma_p = 0,$$

where $\gamma_1 = \min\{\gamma_i : \gamma_i > 0\}$ and $p(1) \leq p$.

a) Let $\xi_1, \ldots, \xi_p$ be independent, identically distributed random variables, uniform on the interval $[-\sqrt{3}, \sqrt{3}]$ and set $\boldsymbol{\xi} = (\xi_1, \ldots, \xi_p)'$. Then,

$$E\boldsymbol{\xi} = 0, \qquad E\boldsymbol{\xi}\boldsymbol{\xi}' = I_{p \times p},$$

where $I_{p \times p}$ denotes the identity matrix. Since $\boldsymbol{\Sigma}$ is a positive definite matrix, $\boldsymbol{\Sigma}^{\frac{1}{2}}$ exists and we can introduce the random vector $\boldsymbol{\eta} = \boldsymbol{\Sigma}^{\frac{1}{2}} \boldsymbol{\xi}$ with

$$E\boldsymbol{\eta} = 0, \qquad E\boldsymbol{\eta}\boldsymbol{\eta}' = \boldsymbol{\Sigma}.$$

Moreover, by definition of $\boldsymbol{\eta}$, there exists a positive constant $a$ such that

$$|\eta_i| \leq a \qquad (1 \leq i \leq p),$$

where $\eta_1, \ldots, \eta_p$ denote the components of $\boldsymbol{\eta}$.

b) Let $\{\boldsymbol{\eta}_n\}_{n \in \mathbb{N}}$ be a sequence of independent random vectors distributed as $\boldsymbol{\eta}$ and set

$$c = \left( \frac{2a(1 + \sqrt{p})}{\gamma_1} \right)^2. \tag{4.13}$$

We are going to prove that the sum of the $\boldsymbol{\eta}_n$'s (with drift term $\sqrt{c}\boldsymbol{\gamma}$) is increasing. Therefore, we consider

$$\begin{aligned}
&\left\|\sum_{i=1}^{k+1}\left(\boldsymbol{\eta}_i+\sqrt{c}\boldsymbol{\gamma}\right)\right\|^2-\left\|\sum_{i=1}^{k}\left(\boldsymbol{\eta}_i+\sqrt{c}\boldsymbol{\gamma}\right)\right\|^2 \\
&= \sum_{j=1}^{p}\left(\sum_{i=1}^{k+1}\left(\eta_{i,j}+\sqrt{c}\gamma_j\right)\right)^2-\sum_{j=1}^{p}\left(\sum_{i=1}^{k}\left(\eta_{i,j}+\sqrt{c}\gamma_j\right)\right)^2.
\end{aligned}$$

For $j=1$, the difference of the inner sums can be reduced in the following way:

$$\begin{aligned}
&\left(\sum_{i=1}^{k+1}\left(\eta_{i,1}+\sqrt{c}\gamma_1\right)\right)^2-\left(\sum_{i=1}^{k}\left(\eta_{i,1}+\sqrt{c}\gamma_1\right)\right)^2 \\
&= \sum_{i=1}^{k+1}\sum_{l=1}^{k+1}\left(\eta_{i,1}+\sqrt{c}\gamma_1\right)\left(\eta_{l,1}+\sqrt{c}\gamma_1\right)-\sum_{i=1}^{k}\sum_{l=1}^{k}\left(\eta_{i,1}+\sqrt{c}\gamma_1\right)\left(\eta_{l,1}+\sqrt{c}\gamma_1\right) \\
&= \left(\eta_{k+1,1}+\sqrt{c}\gamma_1\right)\left(\left(\eta_{k+1,1}+\sqrt{c}\gamma_1\right)+2\sum_{i=1}^{k}\left(\eta_{i,1}+\sqrt{c}\gamma_1\right)\right).
\end{aligned}$$

Summing over all $j=1,\ldots,p$ and observing that $\gamma_{p(1)+1}=\ldots=\gamma_p=0$ yields

$$\begin{aligned}
&\left\|\sum_{i=1}^{k+1}\left(\boldsymbol{\eta}_i+\sqrt{c}\boldsymbol{\gamma}\right)\right\|^2-\left\|\sum_{i=1}^{k}\left(\boldsymbol{\eta}_i+\sqrt{c}\boldsymbol{\gamma}\right)\right\|^2 \\
&= \sum_{j=1}^{p(1)}\left(\eta_{k+1,j}+\sqrt{c}\gamma_j\right)\left(\left(\eta_{k+1,j}+\sqrt{c}\gamma_j\right)+2\sum_{i=1}^{k}\left(\eta_{i,j}+\sqrt{c}\gamma_j\right)\right) \\
&\quad+\sum_{j=p(1)+1}^{p}\eta_{k+1,j}\left(\eta_{k+1,j}+2\sum_{i=1}^{k}\eta_{i,j}\right) \\
&= I_1+I_2.
\end{aligned}$$

Now, by the definition of $c$ in (4.13) and part a) of the proof

$$\eta_{k+1,j}+\sqrt{c}\gamma_j\geq -a+\sqrt{c}\gamma_1=a(1+2\sqrt{p})>0.$$

Hence,

$$\begin{aligned}
I_1 &\geq \left(-a+\sqrt{c}\gamma_1\right)\sum_{j=1}^{p(1)}\left(\left(\eta_{k+1,j}+\sqrt{c}\gamma_j\right)+2\sum_{i=1}^{k}\left(\eta_{i,j}+\sqrt{c}\gamma_j\right)\right) \\
&\geq (2k+1)\left(-a+\sqrt{c}\gamma_1\right)^2.
\end{aligned}$$

Similar arguments yield

$$I_2 \;\geq\; -(2k+1)pa^2.$$

So, finally

$$\begin{aligned} I_1 + I_2 \;&\geq\; (2k+1)\left(a^2 - 2a\sqrt{c}\gamma_1 + c\gamma_1^2 - pa^2\right) \\ &=\; (2k+1)a^2(1 + 4\sqrt{p} + 3p) \;>\; 0 \end{aligned}$$

by plugging in the definition of $c$.

c) We can rewrite the supremum of $\|\boldsymbol{W}(t) + \boldsymbol{\gamma} t\|$ on $[1, T]$ by using the scale transformation of Wiener processes as

$$\begin{aligned} \sup_{1 \leq t \leq T} \|\boldsymbol{W}(t) + \boldsymbol{\gamma} t\| \;&=\; \sup_{\frac{1}{c} \leq t \leq \frac{T}{c}} \|\boldsymbol{W}(ct) + c\boldsymbol{\gamma} t\| \\ &=\; \sqrt{c} \sup_{\frac{1}{c} \leq t \leq \frac{T}{c}} \|\boldsymbol{W}_*(t) + \sqrt{c}\boldsymbol{\gamma} t\|, \end{aligned}$$

where $\boldsymbol{W}_*(t) = \frac{1}{\sqrt{c}}\boldsymbol{W}(ct)$ is a Wiener process. It is easy to check that

$$E\boldsymbol{W}_*(t) = 0, \qquad E\boldsymbol{W}_*'(s)\boldsymbol{W}_*(t) = \min\{s,t\}\boldsymbol{\Sigma}.$$

By Einmahl (1989), we can use the sequence $\{\boldsymbol{\eta}_n\}_{n\in\mathbb{N}}$ defined in b) to obtain the following strong approximation:

$$\left\| \boldsymbol{W}_*(t) - \sum_{i=1}^{t} \boldsymbol{\eta}_i \right\| = \mathcal{O}(\log t) \qquad \text{a.s.}$$

as $t \to \infty$, where the rate cannot be further improved since the $\boldsymbol{\eta}_n$ $(n \in \mathbb{N})$ are not normal random vectors themselves. The above considerations result in

$$\begin{aligned} &\sup_{1 \leq t \leq T} \|\boldsymbol{W}(t) + \boldsymbol{\gamma} t\| \\ &\quad= \sqrt{c} \sup_{\frac{1}{c} \leq t \leq \frac{T}{c}} \|\boldsymbol{W}_*(t) + \sqrt{c}\boldsymbol{\gamma} t\| \\ &\quad\leq \sqrt{c} \sup_{\frac{1}{c} \leq t \leq \frac{T}{c}} \left\| \sum_{i=1}^{t} \boldsymbol{\eta}_i + \sqrt{c}\boldsymbol{\gamma} t \right\| + \mathcal{O}(\log T) \\ &\quad= \sqrt{c} \left\| \sum_{i=1}^{\frac{T}{c}} \boldsymbol{\eta}_i + \frac{\boldsymbol{\gamma}}{\sqrt{c}} T \right\| + \mathcal{O}(\log T) \\ &\quad= \sqrt{c} \left\| \boldsymbol{W}_*\left(\frac{T}{c}\right) + \frac{\boldsymbol{\gamma}}{\sqrt{c}} T \right\| + \mathcal{O}(\log T) \\ &\quad= \|\boldsymbol{W}(T) + \boldsymbol{\gamma} T\| + \mathcal{O}(\log T) \qquad \text{a.s.} \end{aligned}$$

as $T \to \infty$. Since furthermore $\| \boldsymbol{W}(T)+\boldsymbol{\gamma}T \| \leq \sup_{1\leq t\leq T} \| \boldsymbol{W}(t)+\boldsymbol{\gamma}t \|$, the assertion of Lemma 4.3.1 is proved. □

Now, the proof of Theorem 4.3.5 is an immediate consequence of the previous lemma.

**Proof of Theorem 4.3.5:** Applying the invariance claimed in Assumption 4.3.1b), we arrive at

$$\begin{aligned}\sup_{1\leq t\leq T} \| \boldsymbol{\Gamma}(t) \| &= \sup_{1\leq t\leq T} \| \boldsymbol{W}(t) + \boldsymbol{\gamma}t \| + o\left(T^{\frac{1}{\nu}}\right) \\ &= \| \boldsymbol{W}(T) + \boldsymbol{\gamma}T \| + \mathcal{O}(\log T) + o\left(T^{\frac{1}{\nu}}\right) \\ &= \| \boldsymbol{W}(T) + \boldsymbol{\gamma}T \| + o\left(T^{\frac{1}{\nu}}\right) \qquad \text{a.s.}\end{aligned}$$

as $T \to \infty$. Moreover,

$$\| \boldsymbol{W}(T) + \boldsymbol{\gamma}T \| = \| \boldsymbol{\Gamma}(T) \| + o\left(T^{\frac{1}{\nu}}\right) \qquad \text{a.s.}$$

as $T \to \infty$, finishing the proof. □

The final paragraph of this subsection is devoted to Theorems 4.3.1 and 4.3.2. Since the conditions a) and b) of Assumption 4.3.1 have already been proved by Einmahl (1989), they follow readily from the associated Theorem 4.3.3 and 4.3.4, respectively.

### 4.3.2 More approximations

In this subsection we study the real–valued stochastic process $\{\| \boldsymbol{W}(t) + \boldsymbol{\gamma}t \|: t \geq 1\}$. It turns out that this process obtained by taking the Euclidean norm of a $p$–dimensional Wiener process with (positive) drift can a.s. be approximated by a scalar Wiener process $\{W(t) : t \geq 1\}$ plus an additional drift term up to some rate of convergence. Therein, the parameters are completely determined by the vector–valued Wiener process: The drift parameter is given by $\| \boldsymbol{\gamma} \|$ while the variance of the scalar process $\{W(t) : t \geq 1\}$ is obtained from the vector drift term $\boldsymbol{\gamma}$ in combination with the covariance structure of $\{\boldsymbol{W}(t) : t \geq 1\}$ through

$$\delta = \frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}}{\| \boldsymbol{\gamma} \|^2}. \tag{4.14}$$

The exact formulation is as follows.

**Theorem 4.3.6**

*Let $\{\boldsymbol{W}(t) : t \geq 1\}$ be a $p$–dimensional Wiener process with*

$$Cov(\boldsymbol{W}(s), \boldsymbol{W}(t)) = \min\{s,t\}\boldsymbol{\Sigma} \qquad (s,t \geq 1)$$

*and let* $\| \boldsymbol{\gamma} \| > 0$. *Then, there exists a scalar Wiener process* $\{W(t) : t \geq 1\}$ *such that*

$$\| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \| - \| \boldsymbol{\gamma} \| \, t - \delta W(t) = \mathcal{O}\left(\log\log t\right) \qquad \text{a.s.}$$

*as* $t \to \infty$, *where* $\delta$ *is defined in (4.14).*

**Proof:** The basic idea of the proof is to consider the difference $\| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \| - \| \boldsymbol{\gamma} t \|$, since it might be an intuitively promising approach to examine 'how far' the random part of $\| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \|$ differs from the deterministic part $\boldsymbol{\gamma} t$, which equals $\| \boldsymbol{\gamma} \| \, t$ and already gives the desired drift.

a) Using the mean value theorem for $f(x) = \sqrt{x}$, we get

$$\begin{aligned} &\| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \| - \| \boldsymbol{\gamma} t \| \\ =\ & \left( \sum_{j=1}^{p} \left(W_j(t) + \gamma_j t\right)^2 \right)^{\frac{1}{2}} - \left( \sum_{j=1}^{p} \gamma_j^2 t^2 \right)^{\frac{1}{2}} \\ =\ & \frac{1}{2\sqrt{\xi}} \left( \sum_{j=1}^{p} W_i^2(t) + 2t \sum_{j=1}^{p} \gamma_j W_j(t) \right) \\ =\ & \frac{1}{2\sqrt{\xi}} \left( \boldsymbol{W}(t)' \boldsymbol{W}(t) + 2t \boldsymbol{\gamma}' \boldsymbol{W}(t) \right), \end{aligned} \tag{4.15}$$

where $\xi$ satisfies the condition

$$\begin{aligned} \left| \xi - \| \boldsymbol{\gamma} \|^2 t^2 \right| &\leq \left| \| \boldsymbol{W}(t) + \boldsymbol{\gamma} t \|^2 - \| \boldsymbol{\gamma} t \|^2 \right| \\ &\leq \boldsymbol{W}(t)' \boldsymbol{W}(t) + 2t \left| \boldsymbol{\gamma}' \boldsymbol{W}(t) \right|. \end{aligned} \tag{4.16}$$

Applying the law of the iterated logarithm to the random parts of the right hand side of the inequality in (4.16), we see that

$$\limsup_{t \to \infty} \frac{1}{t \log\log t} \sum_{j=1}^{p} W_j^2(t) < \infty \qquad \text{a.s.}$$

as well as

$$\limsup_{t \to \infty} \frac{t |\boldsymbol{\gamma}' \boldsymbol{W}(t)|}{t^{\frac{3}{2}} \sqrt{\log\log t}} < \infty \qquad \text{a.s.,}$$

i.e., $2t|\boldsymbol{\gamma}' \boldsymbol{W}(t)|$ is dominating $\boldsymbol{W}(t)' \boldsymbol{W}(t)$.

Using the latter two statements and dividing both sides of the inequality (4.16) by $t^2$, we arrive at

$$\frac{\xi}{t^2} \longrightarrow \|\boldsymbol{\gamma}\|^2 \qquad \text{a.s.}$$

as $t \to \infty$.

b) We already know that $\{\boldsymbol{\gamma}'\boldsymbol{W}(t) : t \geq 1\}$ is a Gaussian process. In view of $\delta$ defined in (4.14) and equation (4.15) we need to replace $\sqrt{\xi}$ by $\|\boldsymbol{\gamma}\|\, t$. Therefore write

$$\left|\frac{1}{\sqrt{\xi}} - \frac{1}{\|\boldsymbol{\gamma}\|\, t}\right| \leq \frac{1}{\zeta^{\frac{3}{2}}} \left|\xi - \|\boldsymbol{\gamma}\|^2\, t^2\right|$$

using again the mean value theorem. Therein, $\zeta$ fulfills

$$\begin{aligned} \left|\zeta - \|\boldsymbol{\gamma}\|^2\, t^2\right| &\leq \left|\xi - \|\boldsymbol{\gamma}\|^2\, t^2\right| \\ &\leq \boldsymbol{W}(t)'\boldsymbol{W}(t) + 2t|\boldsymbol{\gamma}'\boldsymbol{W}(t)| \end{aligned}$$

by part a) of the proof. Also by the first part, we obtain

$$\begin{aligned} \left|\frac{1}{\sqrt{\xi}} - \frac{1}{\|\boldsymbol{\gamma}\|\, t}\right| &= \mathcal{O}\left(t^{-3} t^{\frac{3}{2}} \sqrt{\log\log t}\right) \\ &= \mathcal{O}\left(t^{-\frac{3}{2}} \sqrt{\log\log t}\right) \qquad \text{a.s.} \end{aligned}$$

as $t \to \infty$, since

$$\zeta = \|\boldsymbol{\gamma}\|^2\, t^2 + \mathcal{O}\left(t^{\frac{3}{2}} \sqrt{\log\log t}\right) \qquad \text{a.s.}$$

as $t \to \infty$ implies $\zeta^{\frac{3}{2}} = \mathcal{O}(t^3)$ a.s. as $t \to \infty$.

c) Putting together the previous results, we arrive at

$$\left|\frac{1}{\sqrt{\xi}} - \frac{1}{\|\boldsymbol{\gamma}\|\, t}\right| |t\boldsymbol{\gamma}'\boldsymbol{W}(t)| = \mathcal{O}(\log\log t) \qquad \text{a.s.}$$

as $t \to \infty$. Furthermore, $\{\|\boldsymbol{\gamma}\|^{-1}\, \boldsymbol{\gamma}'\boldsymbol{W}(t) : t \geq 1\}$ is a Gaussian process with mean $E\boldsymbol{\gamma}'\boldsymbol{W}(t) = 0$ for all $t \geq 1$. A computation of the covariance function yields

$$\begin{aligned} \frac{1}{\|\boldsymbol{\gamma}\|^2} E\boldsymbol{\gamma}'\boldsymbol{W}(s)\,(\boldsymbol{\gamma}'\boldsymbol{W}(t))' &= \frac{1}{\|\boldsymbol{\gamma}\|^2} \boldsymbol{\gamma}' E\boldsymbol{W}(s)\boldsymbol{W}(t)'\boldsymbol{\gamma} \\ &= \min\{s,t\} \frac{\boldsymbol{\gamma}'\boldsymbol{\Sigma}\boldsymbol{\gamma}}{\|\boldsymbol{\gamma}\|^2} \\ &= \min\{s,t\}\delta \end{aligned}$$

for all $s, t \geq 1$. So, there exists a scalar Wiener process $\{W(t) : t \geq 1\}$ as stated in Theorem 4.3.6 which satisfies the desired asymptotic rate. □

### 4.3.3 Some implications

In the final subsection, the law of the iterated logarithm for a weighted version of the norm of the supremum of $\{\mathbf{\Gamma}(t) : t \geq 1\}$ is given as well as some weak convergence results for the stochastic process $\{\sup t^{-\alpha} \parallel \mathbf{\Gamma}(t) \parallel : t \geq 1\}$.

A first corollary establishes the law of the iterated logarithm.

**Corollary 4.3.1**
*Let* $\{\mathbf{\Gamma}(t) : t \geq 1\}$ *be a* $p$*–dimensional stochastic process such that Assumption 4.3.1 holds and let* $\parallel \boldsymbol{\gamma} \parallel > 0$. *Then,*

$$\limsup_{T\to\infty} \frac{T^\alpha}{\sqrt{2T \log\log T}} \left( \sup_{1\leq t\leq T} \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} - \parallel \boldsymbol{\gamma} \parallel T^{1-\alpha} \right) = \delta \qquad a.s.,$$

*where* $\delta$ *is defined in (4.14).*

**Proof:** The proof is given by a joint application of the law of the iterated logarithm and Theorems 4.3.4 and 4.3.6. Firstly, by Theorem 4.3.4

$$\sup_{1\leq t\leq T} \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} = \frac{\parallel \boldsymbol{W}(T) + T\boldsymbol{\gamma} \parallel}{T^\alpha} + o\left(T^{\beta-\alpha}\right) \qquad \text{a.s.}$$

as $T \to \infty$, where $\beta > \max\{\frac{1}{4}, \frac{1}{\nu}\}$. By Theorem 4.3.6 we can replace $\parallel \boldsymbol{W}(T) + T\boldsymbol{\gamma} \parallel$ by a scalar Wiener process whose covariance structure is determined by $\delta$ defined in (4.14), i.e.,

$$\parallel \boldsymbol{W}(T) + T\boldsymbol{\gamma} \parallel = \parallel \boldsymbol{\gamma} \parallel T + \delta W(T) + \mathcal{O}(\log\log T) \qquad \text{a.s.}$$

as $T \to \infty$. Hence,

$$\begin{aligned} &\frac{T^\alpha}{\sqrt{2T \log\log T}} \left( \sup_{1\leq t\leq T} \frac{\parallel \mathbf{\Gamma}(t) \parallel}{t^\alpha} - \parallel \boldsymbol{\gamma} \parallel T^{1-\alpha} \right) \\ &= \frac{T^\alpha}{\sqrt{2T \log\log T}} \left( \frac{\parallel \boldsymbol{W}(T) + \boldsymbol{\gamma} T \parallel}{T^\alpha} - \parallel \boldsymbol{\gamma} \parallel T^{1-\alpha} + o\left(T^{\beta-\alpha}\right) \right) \\ &= \frac{T^\alpha}{\sqrt{2T \log\log T}} \left( \frac{\delta W(T)}{T^\alpha} + o\left(T^{\beta-\alpha}\right) + \mathcal{O}(\log\log T) \right) \\ &= \frac{T^\alpha}{\sqrt{2T \log\log T}} \frac{\delta W(T)}{T^\alpha} + o(1) \qquad \text{a.s.} \end{aligned}$$

as $T \to \infty$. Since

$$\limsup_{T\to\infty} \frac{W(T)}{\sqrt{2T \log\log T}} = 1 \qquad \text{a.s.}$$

by the law of the iterated logarithm, Corollary 4.3.1 is readily proved. □

**Corollary 4.3.2**

*Let $\{\boldsymbol{\Gamma}(t) : t \geq 1\}$ be a p–dimensional stochastic process such that Assumption 4.3.1 holds and let $\| \boldsymbol{\gamma} \| > 0$.*

*a) If $\alpha \in [0, \frac{1}{2})$, then as $T \to \infty$ we get*

$$T^{\alpha-\frac{1}{2}} \left( \sup_{1 \leq t \leq [Tu]+1} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^{\alpha}} - \| \boldsymbol{\gamma} \| \left([Tu]+1\right)^{1-\alpha} \right) \xrightarrow{\mathcal{D}[0,1]} \delta \frac{W(u)}{u^{\alpha}}.$$

*b) If $\alpha \in (\frac{1}{2}, 1)$, then as $T \to \infty$ we get*

$$T^{\alpha-\frac{1}{2}} \left( \sup_{1 \leq t \leq [Tu]+1} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^{\alpha}} - \| \boldsymbol{\gamma} \| \left([Tu]+1\right)^{1-\alpha} \right) \xrightarrow{\mathcal{D}[1,\infty]} \delta \frac{W(u)}{u^{\alpha}}.$$

*c) If $0 < c_1 < c_2 < \infty$, then as $T \to \infty$ we get*

$$T^{\alpha-\frac{1}{2}} \left( \sup_{1 \leq t \leq [Tu]+1} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^{\alpha}} - \| \boldsymbol{\gamma} \| \left([Tu]+1\right)^{1-\alpha} \right) \xrightarrow{\mathcal{D}[c_1,c_2]} \delta \frac{W(u)}{u^{\alpha}}.$$

*Therein, $\{W(t) : t \geq 1\}$ denotes a scalar Wiener process on a suitable interval. Moreover, $\delta$ is defined in (4.14).*

**Proof:** All three parts of the corollary are proved by taking advantage of the approximations of Theorems 4.3.4 and 4.3.6, the scale transformation and the a.s. uniform continuity of Wiener processes on suitable intervals.

a) (i) Fix a real number $C > 0$. Using Theorem 4.3.4 for any $0 \leq u \leq C$ we get

$$\sup_{1 \leq t \leq [Tu]+1} \frac{\| \boldsymbol{\Gamma}(t) \|}{t^{\alpha}} - \frac{\| \boldsymbol{W}([Tu]+1) + \boldsymbol{\gamma}([Tu]+1) \|}{([Tu]+1)^{\alpha}} = o\left(([Tu]+1)^{\beta-\alpha}\right)$$

a.s. as $T \to \infty$. Hence

$$\begin{aligned} & T^{\alpha-\frac{1}{2}} \sup_{0 \leq u \leq C} \left| \sup_{1 \leq t \leq [Tu]+1} \frac{\| \boldsymbol{\Gamma}(t) \|}{([Tu]+1)^{\alpha}} - \frac{\| \boldsymbol{W}([Tu]+1) + \boldsymbol{\gamma}([Tu]+1) \|}{([Tu]+1)^{\alpha}} \right| \\ &= o\left( T^{\alpha-\frac{1}{2}} \sup_{0 \leq u \leq C} ([Tu]+1)^{\beta-\alpha} \right) \\ &= o\left( T^{\alpha-\frac{1}{2}} (CT)^{\beta-\alpha} \right) \\ &= o(1) \qquad \text{a.s.} \end{aligned}$$

as $T \to \infty$ by assumption on $\beta$.

(ii) Since $\alpha \in [0, \frac{1}{2})$, Theorem 4.3.6 yields

$$\begin{aligned}
&T^{\alpha-\frac{1}{2}} \sup_{0\le u\le 1} \left| \frac{\|\boldsymbol{W}([Tu]+1)+\boldsymbol{\gamma}([Tu]+1)\|}{([Tu]+1)^\alpha} \right. \\
&\qquad\qquad \left. - \|\boldsymbol{\gamma}\|\,([Tu]+1)^{1-\alpha} - \delta \frac{W([Tu]+1)}{([Tu]+1)^\alpha} \right| \\
&= \mathcal{O}\left(T^{\alpha-\frac{1}{2}} \sup_{0\le u\le 1} \frac{\log|\log([Tu]+1)|}{([Tu]+1)^\alpha}\right) \\
&= \mathcal{O}\left(T^{\alpha-\frac{1}{2}} \log\log T\right) \\
&= o(1) \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$.

(iii) Considering the Gaussian process $\{T^{\alpha-\frac{1}{2}}([Tu]+1)^{-\alpha}W([Tu]+1) : 0 \le u \le C\}$, we see by an application of the scale transformation for Wiener processes that

$$\left\{T^{\alpha-\frac{1}{2}} \frac{W([Tu]+1)}{([Tu]+1)^\alpha} : 0 \le u \le C\right\} \stackrel{\mathcal{D}}{=} \left\{ \frac{W\left(\frac{[Tu]+1}{T}\right)}{\left(\frac{[Tu]+1}{T}\right)^\alpha} : 0 \le u \le C \right\}.$$

Clearly,

$$\frac{[Tu]+1}{T} \longrightarrow u \qquad (T \to \infty)$$

uniformly on $[0, C]$. Now, part a) of Corollary 4.3.2 is proved, since $t \mapsto t^{-\alpha}W(t)$ is a.s. uniformly continuous on $[0, 1]$.

b) Fix $C > 0$. By similar arguments as in part a) we see by Theorem 4.3.4 that

$$\begin{aligned}
&T^{\alpha-\frac{1}{2}} \sup_{C\le u<\infty} \left| \sup_{1\le t\le [Tu]+1} \frac{\|\boldsymbol{\Gamma}(t)\|}{t^\alpha} - \frac{\|\boldsymbol{W}([Tu]+1)+\boldsymbol{\gamma}([Tu]+1)\|}{([Tu]+1)^\alpha} \right| \\
&= \mathcal{O}\left(T^{\alpha-\frac{1}{2}} \sup_{C\le u<\infty} ([Tu]+1)^{\beta-\alpha}\right) \\
&= \mathcal{O}\left(T^{\alpha-\frac{1}{2}} (TC)^{\beta-\alpha}\right) \\
&= o(1) \qquad \text{a.s.}
\end{aligned}$$

as $T \to \infty$. By Theorem 4.3.6 we get

$$\begin{aligned}
&T^{\alpha-\frac{1}{2}} \sup_{C\le u<\infty} \left| \frac{\|\boldsymbol{W}([Tu]+1)+\boldsymbol{\gamma}([Tu]+1)\|}{([Tu]+1)^\alpha} \right. \\
&\qquad\qquad \left. - \|\boldsymbol{\gamma}\|\,([Tu]+1)^{1-\alpha} - \delta \frac{W([Tu]+1}{([Tu]+1)^\alpha} \right|
\end{aligned}$$

$$= \mathcal{O}\left(T^{\alpha-\frac{1}{2}}\frac{\log\log TC}{(TC)^\alpha}\right)$$
$$= o(1) \qquad \text{a.s.}$$

as $T \to \infty$. Since the law of the iterated logarithm implies

$$\lim_{C\to\infty}\sup_{C\le t<\infty}\frac{|W(t)|}{t^\alpha} = 0 \qquad \text{a.s.},$$

we get

$$\lim_{C\to\infty}\limsup_{T\to\infty} P\left\{T^{\alpha-\frac{1}{2}}\sup_{C\le u<\infty}\frac{|W([Tu]+1)|}{([Tu]+1)^\alpha} \ge \varepsilon\right\} = 0$$

for any $\varepsilon > 0$. Therefore, we can neglect the tail part of the interval $[1,\infty)$ if the constant $C$ is chosen to be large enough, for our considerations now imply

$$\lim_{C\to\infty}\limsup_{T\to\infty} P\left\{T^{\alpha-\frac{1}{2}}\sup_{C\le u<\infty}\left|\sup_{1\le t\le[Tu]+1}\frac{\|\mathbf{\Gamma}(t)\|}{t^\alpha} - \|\boldsymbol{\gamma}\|\,([Tu]+1)^{1-\alpha}\right| > \varepsilon\right\}$$
$$= 0$$

for any $\varepsilon > 0$.

On the other hand, for any compact interval $[c_1, c_2] \subset [1,\infty)$ with $0 < c_1 < c_2 < \infty$, we obtain similarly to part a) of the proof that

$$T^{\alpha-\frac{1}{2}}\left(\sup_{1\le t\le[Tu]+1}\frac{\|\mathbf{\Gamma}(t)\|}{t^\alpha} - \|\boldsymbol{\gamma}\|\,([Tu]+1)^{1-\alpha}\right) \xrightarrow{\mathcal{D}[c_1,c_2]} \delta\frac{W(u)}{u^\alpha}$$

completing the proof of b).

c) The last implication of Corollary 4.3.2 has already been proved in part b). □

# Bibliography


[1] Aue, A. (2000). Zur Schätzung von graduellen Veränderungen auf der Basis von Invarianzprinzipien. Diploma thesis, Philipps University Marburg.

[2] Aue, A. (2003). Strong approximation for RCA(1) time series with applications. *Stat. Probab. Letters* (submitted).

[3] Aue, A., and Horváth, L. (2003a). Delay time in sequential detection of change. *Stat. Probab. Letters* (submitted).

[4] Aue, A., and Horváth, L. (2003b). Approximations for the maximum of a vector–valued stochastic process with drift. *Period. Math. Hung.* **47**, 1–15.

[5] Aue, A., and Steinebach, J. (2002). A note on estimating the change–point of a gradually changing stochastic process. *Stat. Probab. Letters* **56**, 177–191.

[6] Bhattacharya, P.K., and Mallik, A. (1973). Asymptotic normality of the stopping times of some sequential procedures. *Ann. Statist.* **1**, 1203–1211.

[7] Basseville, M., and Nikiforov, I.V. (1993). *Detection of Abrupt Changes: Theory and Applications*. Prentice Hall, Upper Saddle River.

[8] Berger, E. (1982). Fast sichere Approximationen von Partialsummen unabhängiger und stationärer ergodischer Folgen von Zufallsvektoren. Dissertation, University of Göttingen.

[9] Berkes, I., and Horváth, L. (2003). Approximations for the maximum of stochastic processes with drift. *Kybernetika* (to appear).

[10] Berkes, I., Horváth, L., and Kokoszka, P. (2003a). GARCH processes: structure and estimation. *Bernoulli* (to appear).

[11] Berkes, I., Horváth, L., and Kokoszka, P. (2003b). Probabilistic and Statistical Properties of GARCH Processes. In: *Asymptotic Methods in Stochastics*, Fields Inst. Comm. AMS, Providence (to appear).

[12] Berkes, I., and Philipp, W. (1979). Approximation theorems for independent and weakly dependent random vectors. *Ann. Probab.* **7**, 29–54.

[13] Breiman, L. (1967). On the tail behaviour of independent random vectors. *Z. Wahrsch. Verw. Gebiete* **9**, 20–25.

[14] Breiman, L. (1968). *Probability.* Addison–Wesley, Reading, Mass.

[15] Bollarslev, T. (1986). Generalized autoregressive conditional heteroscedasticity. *J. Econ.* **31**, 307–327.

[16] Bougerol, P., and Picard, N. (1992a). Strict stationarity of generalized autoregressive processes. *Ann. Probab.* **20**, 1714–1730.

[17] Bougerol, P., and Picard, N. (1992b). Stationarity of GARCH processes and of some nonnegative time series. *J. Econ.* **52**, 115–127.

[18] Brockwell, P.D., and Davis, R.A. (1991). *Time Series: Theory and Methods*, 2nd Edition. Springer, New York.

[19] Cabilio, P. (1977). Sequential estimation in Bernoulli trials. *Ann. Statist.* **5**, 342–356.

[20] Carrasco, M., and Chen, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* **18**, 17–39.

[21] Chen, J., and Gupta, A.K. (2000). *Parametric Statistical Change Point Analysis.* Birkhäuser, Boston.

[22] Chow, Y.S., and Hsiung, A.C. (1976). Limiting behavior of $\max_{j\le n} S_j/j^\alpha$ and the first passage times in a random walk with positive drift. *Bull. Inst. Math. Acad. Sinica* **4**, 35–44.

[23] Chow, Y.S., Hsiung, A.C., and Yu, K.F. (1980). Limit theorems for a positively drifting process and its related first passage times. *Bull. Inst. Math. Acad. Sinica* **8**, 141–172.

[24] Chu, C.-S., Stinchcombe, J., and White, H. (1996). Monitoring structural change. *Econometrica* **64**, 1045–1065.

[25] Cobb, G.W. (1978). The problem of the Nile: Conditional solution to a change–point problem. *Biometrika* **65**, 243–251.

[26] Csörgő, M., and Horváth, L. (1993). *Weighted Approximations in Probability and Statistics*. Wiley, Chichester.

[27] Csörgő, M., and Horváth, L. (1997). *Limit Theorems in Change–Point Analysis.* Wiley, Chichester.

[28] Csörgő, M., and Réveśz, P. (1975a). A new method to prove Strassen type laws of invariance principle I. *Z. Wahrsch. Verw. Gebiete* **31**, 255–260.

[29] Csörgő, M., and Réveśz, P. (1975b). A new method to prove Strassen type laws of invariance principle II. *Z. Wahrsch. Verw. Gebiete* **31**, 261–269.

[30] Csörgő, M., and Réveśz, P. (1981). *Strong Approximations in Probability and Statistics.* Academic Press, New York.

[31] Donsker, M. (1951). An invariance principle for certain probability limit theorems. *Four Papers on Probability. Mem. Amer. Math. Soc. No.* **6**.

[32] Doob, J.L. (1949). Heuristic approach to the Kolmogorov–Smirnov theorem. *Ann. Math. Statist.* **20**, 393–403.

[33] Eberlein, E. (1986). On strong invariance principles under dependence. *Ann. Probab.* **14**, 260–270.

[34] Einmahl, U. (1987). Strong invariance principles for partial sums of independent random vectors. *Ann. Probab.* **15**, 1419–1440.

[35] Einmahl, U. (1989). Extensions of results of Komlós, Major, and Tusnády to the multivariate case. *J. Mult. Anal.* **28**, 20–68.

[36] Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* **50**, 987–1007.

[37] Erdős, P., and Kac, M. (1946). On certain limit theorems of the theory of probability. *Bull. Amer. Math. Soc.* **52**, 292–302.

[38] Feigin, P.D., and Tweedie, R.L. (1985). Random coefficient autoregressive processes: a Markov chain analysis of stationarity and finiteness of moments. *J. Time Ser. Anal.* **6**, 1–14.

[39] Gut, A. (1974). On the moments and limit distributions of some first passage times. *Ann. Probab.* **2**, 277–308.

[40] Gut, A., and Steinebach, J. (2002). Truncated sequential change–point detection based on renewal counting processes. *Scand. J. Statist.* **29**, 693–719.

[41] Gut, A., and Steinebach, J. (2004). EWMA charts for a sequential change detection in the drift of a stochastic process. *Sequential Analysis* (to appear).

[42] Horváth, L. (1997). Detection of changes in linear sequences. *Ann. Inst. Statist. Math.* **49**, 271–283.

[43] Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2003). Monitoring changes in linear models. *J. Stat. Plann. Infer* (to appear).

[44] Horváth, L., Kokoszka, P., and Steinebach, J. (1999). Testing for changes in multivariate dependent observations with an application to temperature changes. *J. Mult. Anal.* **68**, 96–119.

[45] Horváth, L., and Steinebach, J. (2000). Testing for changes in the mean or variance of a stochastic process under weak invariance. *J. Stat. Plann. Infer.* **91**, 365–376.

[46] Hušková, M. (1998). Estimators in location model with gradual changes. *Comment. Math. Univ. Carolinae* **39**, 147–157.

[47] Hušková, M. (1999). Gradual changes versus abrupt changes. *J. Stat. Plann. Infer.* **76**, 109–125.

[48] Hušková, M., and Steinebach, J. (2000). Limit theorems for a class of tests of gradual changes. *J. Stat. Plann. Infer.* **89**, 57–77.

[49] Hušková, M., and Steinebach, J. (2002). Asymptotic tests for gradual changes. *Statist. Decis.* **20**, 137–151.

[50] Kendall, D.G., and Kendall, W.S. (1980). Alignments in two–dimensional random sets of points. *Adv. Appl. Prob.* **12**, 380–424.

[51] Komlós, J., Major, P., and Tusnády, G. (1975). An approximation of partial sums of independent r.v.'s and the sample d.f. I. *Z. Wahrsch. Verw. Gebiete* **32**, 111–131.

[52] Komlós, J., Major, P., and Tusnády, G. (1976). An approximation of partial sums of independent r.v.'s and the sample d.f. II. *Z. Wahrsch. Verw. Gebiete* **34**, 33–58.

[53] Kuelbs, J., and Philipp, W. (1980). Almost sure invariance principles for partial sums of mixing $B$–valued random variables. *Ann. Probab.* **8**, 1003–1036.

[54] Kühn, C., and Steinebach, J. (2002). On the estimation of change parameters based on weak invariance principles. *Proc. "Fourth Colloquium on Limit Theorems in Probability and Statistics"* (Eds. I. Berkes, E. Csáki, M. Csörgő), Balatonlelle, Ungarn, 28. Juni – 2. Juli 1999, Budapest, 2002, pp. 237–260.

[55] Leadbetter, M.R., Lindgren, L., and Rootzén, H. (1983). Extremes and Related Properties of Random Sequences and Processes. Springer, New York.

[56] Lee, S. (2003). The sequential estimation in stochastic regression model with random coefficients. *Stat. Probab. Letters* **61**, 71–81.

[57] Lee, S., Ha, J., Na, O., and Na, S. (2003). The cusum test for parameter change in time series models. *Scand. J. Statist.* **30**, 781–796.

[58] Major, P. (1976a). The approximation of partial sums of independent r.v.'s. *Z. Wahrsch. Verw. Gebiete* **35**, 213–220.

[59] Major, P. (1976b). Approximation of partial sums of i.i.d.r.v.'s when the summands have only two moments. *Z. Wahrsch. Verw. Gebiete* **35**, 221–230.

[60] Nelson, D.B. (1990). Stationarity and persistence in the GARCH(1,1) model. *Econometric Theory* **6**, 318–334.

[61] Nicholls, D.F., and Quinn, B.G. (1982). *Random Coefficient Autoregressive Models: An Introduction.* Springer–Verlag, New York.

[62] Page, E.S. (1954). Continuous inspection schemes. *Biometrika* **41**, 100–105.

[63] Page, E.S. (1955). A test for a change in a parameter occuring at an unknown point. *Biometrika* **42**, 523–526.

[64] Philipp, W. (1979). Almost sure invariance principles for sums of $B$–valued random variables. In: *Probability in Banach spaces II*, Lecture Notes in Mathematics Vol. 709. Springer–Verlag, Berlin/Heidelberg/New York, pp. 171–193.

[65] Philipp, W. (1986). Invariance principles for independent and weakly dependent random variables. In: *Dependence in Probability and Statistics.* Progress in Probability and Statistics **11**, Birkhäuser, Boston, pp. 225–268.

[66] Philipp, W., and Stout, W. (1986). Invariance principles for martingales and sums of independent random variables. *Math. Z.* **192**, 253–264.

[67] Ploberger, W., Krämer, W., and Kontrus K. (1989). A new test for structural stability in the linear regression model. *J. Econ.* **40**, 307–318.

[68] Robbins, H. (1959). Sequential estimation of the mean of a normal population. *Probability and Statistics* (Harald Cramér Volume). Almquist and Wiksell, Uppsala.

[69] Robbins, H. (1970). Statistical methods related to the law of the iterated logarithm. *Ann. Math. Statist.* **41**, 1397–1409.

[70] Robbins, H., and Siegmund, D. (1974). Sequential estimation of $p$ in Bernoulli trials. *Studies in Probability and Statistics* (Ed. E.J. Williams), University of Melbourne.

[71] Shao, Q.M. (1993). Almost sure invariance principles for mixing sequences of random variables. *Stoch. Proc. Appl.* **48**, 319–334.

[72] Siegmund, D. (1968). On the asymptotic normality of one–sided stopping rules. *Ann. Math. Statist.* **39**, 1493–1497.

[73] Siegmund, D. (1985). *Sequential Analysis*. Springer, New York.

[74] Steinebach, J. (1999). Some remarks on the testing of smooth changes in the linear drift of a stochastic process. *Theory Probab. Math. Statist.* **61**, 164–175.

[75] Strassen, V. (1964). An invariance principle for the law of the iterated logarithm. *Z. Wahrsch. Verw. Gebiete* **3**, 211–226.

[76] Strassen, V. (1965). The existence of probability measures with given marginals. *Ann. Math. Statist.* **36**, 423–439.

[77] Teicher, H. (1973). A classical limit theorem without invariance or reflection. *Ann. Probab.* **1**, 702–704.

[78] Tong, H. (1990). *Non–linear Time Series: A Dynamical System Approach.* Clarendon, Oxford.

[79] Vervaat, W. (1972). Functional central limit theorems for processes with positive drift and their inverses. *Z. Wahrsch. Verw. Gebiete* **23**, 245–253.

[80] Wald, A. (1946). Some impromvements inletting limits for the expected number of observations required by a sequential probability ratio test. *Ann. Math. Statist.* **17**, 466–474.

[81] Wald, A. (1947). *Sequential Analysis*. Wiley, New York.

[82] Wald, A., and Wolfowitz, J. (1948). Optimum character of the sequential probability ratio test. *Ann. Math. Statist.* **19**, 326–339.

## Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen dieser Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Josef Steinebach betreut worden.

Köln, im Dezember 2003

____________________
Alexander Aue

Teilpublikationen:

[1] Approximations for the maximum of a vector–valued stochastic process with drift. *Period. Math. Hung.* **47** (2003), 1–15 (mit L. Horváth).

[2] Delay time in sequential detection of change. *Stat. Probab. Letters* (2003), eingereicht (mit L. Horváth).

[3] Strong approximation for RCA(1) time series with applications. *Stat. Probab. Letters* (2003), eingereicht.

# Lebenslauf

## Perönliche Daten

| | |
|---|---|
| Name | Alexander Aue |
| Geburtsdatum | 18.10.1974 |
| Geburtsort | Marburg/Lahn |
| Familienstand | ledig |

## Schulbildung

| | |
|---|---|
| 1981–1985 | Grundschule Jesberg |
| 1985–1994 | Gymnasium Oberurff |
| Juni 1994 | Abitur |

## Zivildienst

| | |
|---|---|
| 1994–1995 | Ableistung des Zivildienstes |

## Studium

| | |
|---|---|
| 1995–2000 | Mathematik (Diplom) mit Nebenfach Physik an der Philipps–Universität Marburg |
| Okt. 1997 | Vordiplom |
| Sep. 2000 | Diplom |

## Berufstätigkeit

| | |
|---|---|
| 2000–2002 | Wissenschaftlicher Mitarbeiter am Fachbereich Mathematik und Informatik der Philipps–Universität Marburg |
| seit 2002 | Wissenschaftlicher Mitarbeiter am Mathematischen Institut der Universität zu Köln |