Charles University in Prague
Faculty of Mathematics and Physics

DOCTORAL THESIS

Ondřej Chochola

Robust Monitoring Procedures for Dependent Data

Department of Probability and Mathematical Statistics
Supervisor: Prof. RNDr. Marie Hušková, DrSc.
Study Branch: Probability and Mathematical Statistics

Prague, 2013

Acknowledgements

First of all, I would like to thank my supervisor Prof. RNDr. Marie Hušková, DrSc., for her patient guidance, encouragement and valuable advice during my research. I am extremely grateful for having such a supportive and inspiring supervisor. Her comments and suggestions contributed enormously to the production of this thesis.

I would also like to thank to all my teachers from Charles University in Prague for introducing me into the world of higher mathematics and mathematical statistics, and providing me important theoretical knowledge.

I owe my deepest gratitude to my family for their support and patience during my whole studies. My deepest and warmest thanks belong to Jitka for love and encouragement.

This work was supported by the grants GAUK 162310, GAČR P201/12/1277 and SVV 261315/2010.

Statement of Honesty

I hereby declare that I have written this doctoral thesis separately, independently and entirely with using quoted resources. I agree that the University Library shall make it available to borrowers under rules of the Library.

Prague, June 10, 2013

Ondřej Chochola

Annotations

Title: Robust Monitoring Procedures for Dependent Data

Author: Ondřej Chochola

Department: Department of Probability and Mathematical Statistics

Supervisor: Prof. RNDr. Marie Hušková, DrSc.

Supervisor's e-mail address: `huskova@karlin.mff.cuni.cz`

Abstract: In the thesis we focus on sequential monitoring procedures. We extend some known results towards more robust methods. The robustness of the procedures with respect to outliers and heavy-tailed observations is introduced via use of M-estimation. Another extension is towards dependent and multivariate data. For several models, the appropriate test statistics are proposed and their asymptotic properties are studied both under the null hypothesis of no change as well as under the alternatives in order to derive proper critical values and show consistency of the tests. Finite sample properties are checked in a simulation study and by an application on real data as well.

Keywords: Robust monitoring, Change-point detection, M-estimates, Weak dependence, Capital asset pricing model

Název práce: Robustní monitorovací procedury pro závislá data

Autor: Ondřej Chochola

Katedra: Katedra pravděpodobnosti a matematické statistiky

Vedoucí disertační práce: Prof. RNDr. Marie Hušková, DrSc.

e-mail vedoucího: `huskova@karlin.mff.cuni.cz`

Abstrakt: V práci se zabýváme sekvenční analýzou změn. Některé známé výsledky rozšíříme na robutní metody. Robustnost vzhledem k odlehlým pozorováním a pozorováním s těžkými chvosty je dosažena využitím M-odhadů. Další rozšíření se týká mnohorozměrných a závislých dat. Pro několik modelů jsou navrženy vhodné testové statistiky a jejich asymptotické chování za nulové hypotézy žádné změny stejně jako za alternativ je studováno. Díky tomu můžeme odvodit správné kritické hodnoty a ukázat konzistenci testů. Simulační studie potvrdila použitelnost navržených procedur i pro konečné vzorky dat. Taktéž je ukázána možná aplikace.

Klíčová slova: robustní analýza změn, M-odhady, slabá závislost, model oceňování kapitálových aktiv

Notation

m	...	length of training period; basis for asymptotics
T	...	ratio of monitoring period to training one
d	...	number of dimensions for multivariate data
a.e.	...	almost everywhere
a.s.	...	almost surely
$\stackrel{\mathcal{D}}{=}$	...	equality in distribution
$\xrightarrow{\mathcal{D}}$	...	convergence in distribution
$\xrightarrow{\mathcal{D}[0,T]}$	...	weak convergence in Skorokhod space $D[0,T]$
$\xrightarrow{\mathcal{D}^d[0,T]}$	...	weak convergence in Skorokhod space $D^d[0,T]$
$\xrightarrow{P}$	...	convergence in probability
$:=$	...	defining symbol
O_P, o_p	...	stochastic Landau symbols, by default with $m \to \infty$
O, o	...	deterministic Landau symbols, by default with $m \to \infty$
$I\{\cdot\}$	...	identificator of a set
$(\cdot)^+$	...	positive part function ($\max(0,\cdot)$)
$\lfloor \cdot \rfloor$	...	integer part function
C	...	generic positive constant
$\alpha(k)$	...	α-mixing coefficient

Abbreviations

ARL	...	average run length
CAPM	...	capital asset pricing model
CLT	...	central limit theorem
CUSUM, MOSUM	...	cumulative sums, moving sums
DPC	...	delay-power curves (plot) - see p.82
DRL	...	density of run length (plot) - see p.82
FLT	...	flat top kernel (estimator)
FLT adapt	...	flat top kernel (estimator) with adaptive bandwidth choice
i.i.d.	...	independent and identically distributed
LAD	...	least absolute deviation
LHS	...	the left hand side
LRV	...	long-run variance
LS	...	least squares
RHS	...	the right hand side
SPC	...	size-power curves (plot) - see p.73
(V)AR	...	(vector) autoregression
WIP	...	weak invariance principle
WLOG	...,	without loss of generality

Contents

Annotations **2**

Notation **3**

Abbreaviations **4**

1 Introduction **7**

1.1 Change-point Analysis . . . 7
1.2 Retrospective Change-point Analysis . . . 8
1.3 Sequential Monitoring . . . 8
1.4 M-estimates . . . 11
1.5 Weak Dependence . . . 14
1.6 State of Art . . . 17
1.7 Aim and Structure of the Thesis . . . 18

2 Monitoring in Location Model **20**

2.1 Model, Assumptions and Test Statistic . . . 20
2.2 Main Results . . . 23
2.3 Auxiliary Results . . . 24
2.4 Proofs of Main Results . . . 30
2.5 Estimation of Long-run Variance . . . 33

3 Monitoring in Multivariate Location Model **38**

3.1 Model, Assumptions and Test Statistic . . . 38
3.2 Main Results . . . 40
3.3 Proofs . . . 42
3.4 Multiple Comparison . . . 45

4 Monitoring in Capital Asset Pricing Model **47**

4.1 Model, Assumptions and Test Statistic . . . 48
4.2 Main Results . . . 50
4.3 Auxiliary Results . . . 52
4.4 Proofs . . . 61

5 Retrospective Analysis **66**

5.1 Location Model . . . 66

5.2 Multivariate Location Model 68
5.3 CAPM 69

6 Computational Aspects **71**
6.1 Location Model 71
6.1.1 Long-run Variance Estimators 72
6.1.2 Boundary Function 77
6.1.3 Monitoring Procedures with Adaptive LRV Estimator 80
6.2 Multivariate Location Model 90
6.3 CAPM 94

7 Critical Values **99**
7.1 Online Monitoring 99
7.2 Retrospective Analysis 101

8 Conclusion **102**

A Some Useful Results **103**

Bibliography **105**

Chapter 1

Introduction

The *change-point analysis* is introduced first in both of its two general setups. A robustness of the procedures is introduced via use of M-estimators and thus we recall the theory in Section 1.4. As data in the change-point analysis are time ordered the assumption of independence is not usually feasible there. The dependence structure considered in the thesis is presented in Section 1.5. In next section the latest development of change-point analysis is presented. This chapter is concluded by an outline of the thesis.

1.1 Change-point Analysis

Modern world is changing at tremendous speed. High tech instruments from last year are average this year and will be outdated the next one. The change is present everywhere. Therefore the scientific study of change is of key importance.

A study of changes in statistical models is called *change-point analysis*. Otherwise said it is a study of stability of models. It provides tools to decide whether a given time ordered data remain stable over time (follow the same model all time) or whether some change occurs. It means that data follow a certain model up to an unknown time-point and a different model afterwards.

Two main areas of change-point analysis based on the data acquisition can be distinguished. When all data are at hand at the beginning of the analysis we speak about *Retrospective* (or *Offline*) analysis. We formulate this subject in the next section. However in modern applications it is common to receive the data online and monitor the stability online as well. In that case we speak about *Sequential* (or *Online*) change-point analysis or *monitoring*. This is set up in Section 1.3.

A history of change-point analysis goes back to 1950' and relates to names Wald and Page (see Wald [1947] and Page [1954] for example) and the area of statistical quality control. Since then the area of applications has grown a lot and nowadays it is exploited in a number of fields, for example in medicine, biology, climatology and economy. With a growing data availability in science and business it will become even more important.

Number of theoretical studies of the topic is also abundant, we summarize the recent development is Section 1.6.

1.2 Retrospective Change-point Analysis

The problem of stability of historical models can be formulated as follows. Observations $Y_1, \ldots, Y_n$ follow a statistical model, which may have changed during the observational period. There could have been one or more changes. We have to decide how many changes (if any) have occurred, then to locate them and finally to specify the way of the change.

The task is treated as a hypothesis testing problem, where we test the null hypothesis of no change against an alternative hypothesis, concerning a particular kind of change. Most popular alternatives are so called AMOC (at-most-one-change) ones, where as the name indicates we are looking for one abrupt change in the data. However the so-called gradual changes as well as multiple changes have been studied. As an example we introduce the AMOC alternative in a general case, where one looks for any change in distribution.

We denote the distribution functions of data $Y_1, \ldots, Y_n$ as $F_1, \ldots, F_n$ and we test the null hypothesis

$$H_0 : F_1 = \ldots = F_n \tag{1.1}$$

against

$$\begin{aligned} H_1 &: \text{there exists unknown time point } k^* < n \text{ such that} \\ &F_1 = \ldots = F_{k^*} \neq F_{k^*+1} = \ldots F_n. \end{aligned} \tag{1.2}$$

We have to define an appropriate test statistic and ensure that it behaves well under both null the hypothesis (i.e. the required level is kept) and the alternative hypothesis (test has appropriate power).

The problem is usually not treated in such a general setting but it is assumed that the distribution is dependent on some parameter, such as location or scale parameter or regression coefficient. Then the problem is reduced to the change in this parameter itself.

1.3 Sequential Monitoring

We speak about *(sequential) monitoring* when observations are coming sequentially and we want to monitor the stability online, i.e. after each new observation arrives, we make a decision, whether a change has occurred or not. We therefore observe a potentially infinite sequence $Y_1, Y_2, Y_3, \ldots$. Initially the generating process is "under control". However at some unknown time-point k^* the process changes and becomes "out of control". Our aim is to stop the observation (and take an appropriate action) as soon as possible after the change, but we also want to avoid false alarms caused by random fluctuations only.

Similarly to the retrospective change-point analysis we treat the problem as a hypothesis testing, where the null hypothesis of no change is testes against the alternative that a change occurs at some unknown time k^*. Denote the distribution functions of the observations Y_i again as $F_i,\ i = 1, 2, \ldots$. In the most general case the hypotheses of interest can represent any change in the distribution. Hence the null hypothesis of no change is

$$H_0 : F_i = F_0,\, 1 \leq i < \infty$$

and the alternative is

$$\begin{aligned} H_1 &: \text{there exists unknown } k^* \geq 1 \text{ such that} \\ &F_i = F_0,\, 1 \leq i \leq k^*, \quad F_i = F_*,\, k^* < i < \infty, \quad F_0 \neq F_*. \end{aligned}$$

The testing procedure is described by the so-called *stopping time*, which is the time, when the null hypothesis is rejected and the observations stopped. This time τ is defined as

$$\tau = \inf\{k \geq 1 : \Gamma(k) > c\},$$

with the understanding that $\inf\{\emptyset\} = \infty$. The test statistics $\Gamma(k) = \Gamma(Y_1, \ldots, Y_k)$ $k = 1, 2, \ldots$ (sometimes called detectors) are evaluated at time k and can be based on all available observations $Y_1, \ldots, Y_k$. The value c is a critical value according to which the decision is conducted.

The main task is to find a suitable detector $\Gamma(k)$ and also a method how to determine the critical value c such that the procedure fulfils two obvious requirements. The first one is that under the null hypothesis of no change we do not want to stop observing (i.e. cause the false alarm) whereas under the alternative we want to stop as soon as possible after the time k^*. Apparently there is a trade-off between these two requirements. There exists two main principles, how these requests can be balanced and the critical values determined.

The first one is based on the average run length (ARL). The idea is that we specify a lower bound for the average number of observations, which should be taken under the null hypothesis until the first alarm. Based on this request we can determine the critical values. Among the procedures that use ARL are also the famous Shewhart and CUSUM ones. Since the topic of this thesis is based on the second method, we do not give further details about ARL methods, those can be found in Siegmund [1985] for example.

The second one is based on requirements on probabilities of type I and type II errors. The idea is that we constrain the probability of false alarm (type I error) with a pre-defined small value $\alpha \in (0, 1)$ and then we minimize the probability of type II error (or try to make it small at least). We can formulate the requirements as follows:

$$P(\tau < \infty | H_0) \leq \alpha, \quad P(\tau < \infty | H_1) = 1.$$

With such requirements, the procedure detects the change with probability 1 if it occurs, but if there is no change, it rejects the null hypothesis with probability α at most.

Sequential Monitoring with Training Data Available
Now we introduce some more specific setting under which we work throughout the thesis. It is related to the one suggested by Chu et al. [1996].

Let us assume that the distribution function is characterized by some (generally multivariate) parameter $\boldsymbol{\theta}$ i.e. $F_i(\cdot) = F(\cdot, \boldsymbol{\theta}_i)$ and the change is allowed only in this parameter. In sequential change-point analysis, it is often assumed that there is available a training data set (historical data) with no change. Denote the size of this data set by m. A stability of the parameter of interest in this data set is called *non-contamination assumption*

$$\boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_m. \tag{1.3}$$

We can estimate the null hypothesis value of parameter from the historical data. Throughout the thesis we will work with the non-contamination assumption. The length of historical period m is also taken as a basis for the asymptotics i.e. the asymptotic relations are taken for $m \to \infty$ (this is usually assumed implicitly and thus not written).

The null hypothesis can be rewritten as

$$H_0 : \boldsymbol{\theta}_i = \boldsymbol{\theta}_0, \quad 1 \leq i < m + N + 1, \tag{1.4}$$

which is tested against the alternative that a change in the parameter occurs

$$\begin{aligned} &H_1 : \text{there exists } k^* = k^*_m < N \text{ such that} \\ &\qquad \boldsymbol{\theta}_0 = \boldsymbol{\theta}_1 = \ldots = \boldsymbol{\theta}_{m+k^*} \neq \boldsymbol{\theta}_{m+k^*+1} = \ldots = \boldsymbol{\theta}_*. \end{aligned}$$

The values of $\boldsymbol{\theta}_0, \boldsymbol{\theta}_*$ and k^* are unknown. Above we also introduced the maximal length of monitoring period N.

We can consider either a possibly infinite monitoring period or the so-called *closed-end monitoring* which is very useful in practical applications, since the maximal number of possible observations is (or easily can be) specified a priori and thus allow us to tailor the critical values appropriately to get more powerful tests. Further we assume that the maximal monitoring period is a fix multiple of the training period i.e. that $N = mT$ for some fixed $T > 0$.

The stopping time now depends both on the length of the historical data and the monitoring period, which is emphasized in the notation as well. It is defined as

$$\tau_{m,T} = \inf\{1 \leq k \leq mT : \Gamma(m,k) > c_{m,T}(\alpha)\}, \tag{1.5}$$

where again the detector $\Gamma(m,k)$ and the critical values $c_{m,T}(\alpha)$ have to be chosen such that the crucial requirements on the monitoring procedure

$$\lim_{m\to\infty} P(\tau_{m,T} < \infty | H_0) \quad \leq \alpha, \tag{1.6}$$

$$\lim_{m\to\infty} P(\tau_{m,T} < \infty | H_1) \quad = 1 \tag{1.7}$$

are fulfilled. Equation (1.6) ensures the level to be at most α asymptotically while the condition (1.7) corresponds to the consistency of the test, i.e. the probability of the type II error tends to 0 or, in other words, the power tends to 1, as $m \to \infty$.

The detector has often a form of a standardized statistic

$$\Gamma(m,k) = \frac{|\widehat{Q}(m,k)|}{g(k/m)}, \tag{1.8}$$

where $\widehat{Q}(m,k) = \widehat{Q}(m,k,Y_1,\ldots,Y_{m+k})$ uses all the observations available at time $m+k$ and $g(\cdot)$ is the so-called *boundary function* used for standardization.

Based on (1.5), the decision rule is as follows:

- The null hypothesis is rejected and the observation is stopped, whenever the ratio of $|\widehat{Q}(m,k)|$ and $g(k/m)$ exceeds the critical value $c_{m,T}(\alpha)$.
- Otherwise we take a new observation up to mT.

1.4 M-estimates

In this section we briefly recall robust estimation. Basic introduction can be found e.g. in Huber [1981] and Jurečková and Sen [1996].

Let us consider a random sample $Y_1,\ldots,Y_n$, where the observations have some unknown distribution function $F_\theta(\cdot) = F(\cdot - \theta)$, where $\theta \in \Theta$ is a location parameter. Then the classical least squares estimate of the true value of parameter is defined as

$$\widetilde{\theta}_n = \arg\min_{t\in\Theta} \sum_{i=1}^{n} (Y_i - t)^2.$$

This estimate is optimal if the underlying distribution is normal. However in case the distribution has heavier tails or some outliers are present, this estimate can be seriously distorted.

To overcome such difficulties Huber [1964] proposed a generalization which leads to the so-called *M-estimates*. These use a convex loss function ρ, i.e.

$$\widehat{\theta}_n = \arg\min_{t\in\Theta} \sum_{i=1}^{n} \rho(Y_i - t). \tag{1.9}$$

A choice of the loss function ρ influences sensitivity of the estimate with respect to outliers and heavy-tailed distributions i.e. the robustness of the given M-estimate.

If we assume that the loss function has a derivative (at least one sided) then the M-estimate can be defined via a score function ψ such that $\rho' = \psi$ almost everywhere. Since ρ is assumed convex then ψ is monotone. In case that ψ is also continuous then (1.9) is equivalent to finding solution of

$$\sum_{i=1}^{n} \psi(Y_i - t) = 0 \tag{1.10}$$

with respect to $t \in \Theta$. However, in general (e.g. when ψ is a step function) equation (1.10) may not have any solution. In that case we need to interpret a solution as a point where LHS changes a sign. Another problem can be an identification of a particular solution if there are more of them. Therefore we define

$$\widehat{\theta}_n^- = \sup\Big\{t \in \Theta : \sum_{i=1}^n \psi(Y_i - t) > 0\Big\}, \quad \widehat{\theta}_n^+ = \inf\Big\{t \in \Theta : \sum_{i=1}^n \psi(Y_i - t) < 0\Big\} \tag{1.11}$$

(note that $\psi(\cdot - t)$ is nonincreasing in t). Any $\widehat{\theta}_n^* \in [\widehat{\theta}_n^-, \widehat{\theta}_n^+]$ can serve as a M-estimate and minimizes (1.9). Moreover the asymptotic behavior is the same regardless of $\widehat{\theta}_n^*$ chosen (see below). Thus for uniqueness we usually choose the midpoint

$$\widehat{\theta}_n = \frac{1}{2}\big(\widehat{\theta}_n^- + \widehat{\theta}_n^+\big). \tag{1.12}$$

Now we present some basic facts about the asymptotic behavior of the estimates. As was said, the case when ψ function is not continuous can be problematic. This can however be overcome, if the distribution F_θ compensate for it. Thus we define function

$$\lambda(t) = -\int \psi(x - t) dF_\theta(x), \quad t \in \Theta,$$

which incorporates both previous aspects. This function is widely used in M-estimation theory.

Next two results about consistency and asymptotical normality under the general definition (1.12) are slightly adapted versions of Corollary 3.2 and 3.5 of Huber and Ronchetti [2009].

Lemma 1.1. *Let $Y_1, \ldots, Y_n$ be i.i.d. random variables, ψ nondecreasing function and θ_0 the unique solution of $\lambda(t) = 0$.*
Then $\widehat{\theta}_n \to \theta_0$ in probability and almost surely as $n \to \infty$.

Lemma 1.2. *Let $Y_1, \ldots, Y_n$ be i.i.d. random variables, ψ nondecreasing function and θ_0 the unique solution of $\lambda(t) = 0$. Further suppose that $\lambda(t)$ is differentiable at θ_0 such that $\lambda'(\theta_0) > 0$ and $\sigma^2(t) = \int_{\mathbb{R}} \psi^2(x-t) dF_\theta(x)$ is finite and continuous in neighborhood of θ_0.*
Then $\sqrt{n}(\widehat{\theta}_n - \theta_0)$ is asymptotically normal with mean 0 and variance $\sigma^2(\theta_0)/(\lambda'(\theta_0))^2$.

Later in the thesis we will need the root-n consistency of the estimate, which is a result in between of those of Lemma 1.1 and 1.2 and therefore we present the following proposition together with a proof.

Proposition 1.3. *Let $Y_1, \ldots, Y_n$ be i.i.d. random variables, ψ nondecreasing function and θ_0 the unique solution of $\lambda(t) = 0$. Further suppose that $\lambda(t)$ is differentiable at θ_0 such that $\lambda'(\theta_0) > 0$ and $\sigma^2(t) = \int_{\mathbb{R}} \psi^2(x - t) dF_\theta(x)$ is finite and continuous in neighborhood of θ_0.*
Then for any $\widehat{\theta}_n^* \in [\widehat{\theta}_n^-, \widehat{\theta}_n^+]$ is $\widehat{\theta}_n^* - \theta_0 = O_P(n^{-1/2})$ as $n \to \infty$.

Proof. For simplicity we assume that $\theta_0 = 0$.
Note that for $\widehat{\theta}_n^-, \widehat{\theta}_n^+$ from (1.11) holds

$$\{\widehat{\theta}_n^- < t\} \subseteq \Big\{\sum \psi(Y_i - t) \leq 0\Big\} \subseteq \{\widehat{\theta}_n^- \leq t\}, \tag{1.13}$$

$$\{\widehat{\theta}_n^+ < t\} \subseteq \Big\{\sum \psi(Y_i - t) < 0\Big\} \subseteq \{\widehat{\theta}_n^+ \leq t\}. \tag{1.14}$$

We focus on $\widehat{\theta}_n^-$ first. For $K > 0$

$$P(\widehat{\theta}_n^- \leq K n^{-1/2}) \geq P\Big(\sum_{i=1}^n \psi(Y_i - K n^{-1/2}) \leq 0\Big) =$$

$$= P\Big(n^{-1/2} \sum_{i=1}^n [\psi(Y_i - K n^{-1/2}) - \mathrm{E}\,\psi(Y_i - K n^{-1/2})] \leq -\sqrt{n}\,\mathrm{E}\,\psi(Y_i - K n^{-1/2})\Big) \tag{1.15}$$

Now we concentrate on the event within the last $P(\cdot)$. RHS of the inequality can be expressed as

$$n^{1/2}\lambda(K n^{-1/2}) = n^{1/2}[\lambda'(0) K n^{-1/2} + O(n^{-1})] = \lambda'(0)K + o(1) \text{ as } n \to \infty.$$

On the other hand we have by simple application of Chebyshev's inequality that LHS is $O_P(1)$ $(n \to \infty)$ and thus (1.15) can be made at least $1 - \varepsilon$ for all $\varepsilon > 0$ by choosing K properly large. Similarly we have

$$P(\widehat{\theta}_n^- < -K n^{-1/2}) \leq P\Big(\sum_{i=1}^n \psi(Y_i + K n^{-1/2}) \leq 0\Big) =$$

$$= P\Big(n^{-1/2} \sum_{i=1}^n [\psi(Y_i + K n^{-1/2}) - \mathrm{E}\,\psi(Y_i + K n^{-1/2})] \leq -\sqrt{n}\,\mathrm{E}\,\psi(Y_i + K n^{-1/2})\Big) \tag{1.16}$$

and the RHS in the last $P(\cdot)$ is equal to $-\lambda'(0)K + o(1)$ $(n \to \infty)$. Thus (1.16) can be bounded from above by ε. Together we have $\widehat{\theta}_n^- = O_P(n^{-1/2})$.
Using (1.14) instead of (1.13) we get the same also for $\widehat{\theta}_n^+$, which concludes the proof. □

The asymptotic normality of M-estimates is a useful result, enabling, for example, testing the significance of estimate. However even more insight into the asymptotic behavior can be gained via the representation similar to Bahadur [1966] developed for sample quantiles. This representation will be very useful for our later work, thus we present here results of He and Shao [1996]. For simplicity of notation we formulate the assumptions with $\theta_0 = 0$ as is the usual case for error distributions. Assume

(i) $Y_1, Y_2, \ldots$ are i.i.d.,

(ii) the function ψ is nondecreasing (i.e. ρ is convex); $\lambda(0) = 0$; the derivative $\lambda'(t)$ exists at $t = 0$ such that $\lambda'(0) > 0$ and is Lipschitz in the neighborhood of zero, i.e., for a positive constant D_3 and $|t| \leq D_3$, there exists $D_4 > 0$, such that

$$|\lambda(t) - \lambda(0) - t\lambda'(0)| \leq D_4 t^2,$$

(iii) it holds that $\int |\psi(e)|^{2+\xi} dF(e) < \infty$ for some $\xi > 0$ and further for some constants $a \geq 1$ and $D_1, D_2 > 0$ it holds that

$$\int (\psi(x - t_2) - \psi(x - t_1))^2 dF(x) \leq D_1 |t_2 - t_1|^a, \quad |t_j| \leq D_2, j = 1, 2,$$

(iv) $\int \psi^2(e) dF(e) \in (0, \infty)$.

The Bahadur almost sure representation of $\widehat{\theta}_n$, as $n \to \infty$, is

$$\widehat{\theta}_n = \theta_0 + (n\lambda'(0))^{-1} \sum_{i=1}^{n} \psi(Y_i - \theta_0) + O\left(\frac{\ln \ln n}{n}\right) \quad \text{a.s.} \tag{1.17}$$

From (1.17) we easily conclude that

$$\sqrt{n}(\widehat{\theta}_n - \theta_0) = \frac{1}{\sqrt{n}\lambda'(0)} \sum_{i=1}^{n} \psi(Y_i - \theta_0) + O_P(n^{-\eta}), \quad \text{for some } \eta > 0, \tag{1.18}$$

which is a form sufficient for us. It is also easy to see how the asymptotic normality follows from here. For further details about the asymptotic representation, see also Jurečková and Sen [1996].

Finally, typical choices of score functions ψ are given:

(a) $\psi(x) = x$, $x \in \mathbb{R}$ leads to least squares (LS,L_2) estimation

(b) $\psi(x) = \operatorname{sign} x$, $x \in \mathbb{R}$ leads to least absolute deviation (LAD, L_1) estimation

(c) So-called Huber ψ function

$$\psi(x) = xI\{|x| \leq K\} + K \operatorname{sign}(x) I\{|x| > K\},\ x \in \mathbb{R},\ \text{for some } K > 0 \tag{1.19}$$

is a compromise between the previous two and serves as a representative for robust estimation.

1.5 Weak Dependence

The classical assumption of independence of observations is often too strong to be realistic in many applications. Especially if data are collected sequentially over time which is indeed our case. It is then natural to expect that the current observation depends to some degree on the previous observations. However this dependence can be assumed to diminish as the (time) distance between the observations increases. This is a basic idea of the so-called *weak dependence*.

The concept of weak dependence is well known for many years and over the past decades has been formalized in many ways. Perhaps the most popular are various mixing conditions (see Doukhan [1994], Bradley [2005]), but in recent years several other approaches have also been introduced (see Doukhan and Louhichi [1999], Wu [2007], and Hörmann and Kokoszka [2010] among others).

Among the mixing conditions the two probably most used in statistical context are the strong and uniformly strong mixing which we now recall. Let us assume that $\{Y_i\}_i$ is a sequence of random elements on a probability space $(\Omega, \mathcal{F}, P)$. For sub-σ-fields $\mathcal{A}, \mathcal{B} \subseteq \mathcal{F}$, we define

$$\alpha(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}} |P(A \cap B) - P(A)P(B)|, \tag{1.20}$$

$$\varphi(\mathcal{A}, \mathcal{B}) := \sup_{A \in \mathcal{A}, B \in \mathcal{B}, P(A) > 0} |P(B|A) - P(B)|. \tag{1.21}$$

Intuitively it is clear that α, φ measure the dependence of the events in $\mathcal{B}$ on those in $\mathcal{A}$. Therefore, considering a filtration $\mathcal{F}_j^k := \sigma(Y_i, j \leq i \leq k)$, we can describe the fading dependence between the observations in a following way:
A sequence $\{Y_i\}_i$ of random elements is said to be strong mixing (α-mixing) if

$$\alpha(n) := \sup_{k \in \mathbb{N}} \alpha(\mathcal{F}_1^k, \mathcal{F}_{k+n}^\infty) \to 0, \quad n \to \infty,$$

and analogously for uniformly strong mixing (φ-mixing).

It is clear that for stationary sequence $\{Y_i\}_i$ we can omit the $\sup_{k \in \mathbb{N}}$ in the definition. The coefficients $\alpha(n)$ and $\varphi(n)$ measure how much dependence exists between events which are at least n observations or time periods apart. The rate of decay of $\alpha(n)$, $\varphi(n)$ (often called mixing rate) is usually characterized by summability of powers of the mixing coefficients i.e. that for some $\delta > 0$ is

$$\sum_{n=1}^{\infty} \alpha(n)^\delta < \infty.$$

The concept of the uniformly strong mixing was introduced by Rosenblatt [1956] and it holds that the uniformly strong mixing implies the strong mixing (see Lin and Lu [2010] for the proof), which was introduced later on by Ibragimov [1959].

This concept of the weak dependence has proven itself very useful in statistics as it effectively deals with the most common dependence structures. Indeed, it was shown in Anderson [1958] that *m-dependent* processes as well as finite order ARMA processes with innovations satisfying Doeblin's condition (Billingsley [1968], p. 168) are φ-mixing. Finite order processes, which do not satisfy Doeblin's condition were showed to be α-mixing instead (see Ibragimov and Linik [1971]). Moreover, Rosenblatt [1971] provides general conditions for stationary Markov Processes to be α-mixing as well.

In Withers [1981] and Davidson [1994] (Section 14.3) one can find conditions for linear processes to be α-mixing, which, in general, are smoothness of density of innovations and the coefficients of the process going to zero sufficiently fast. The smoothness is really important, as a famous example of Andrews [1984] shows that even AR(1) sequence with Bernoulli innovations needs not to be α-mixing. This is however not a serious drawback since in applications we usually have continuous data.

As a more serious drawback of mixing approach is sometimes pointed out the fact, that it is quite hard to verify that the sequence is mixing (either strong or uniformly strong). This is one of the reasons why other approaches to weak dependence arose and we introduce the one of Hörmann and Kokoszka [2010] at the end of this section. Nevertheless we have chosen to work under the mixing condition in this thesis and we concentrate on the α-mixing as it is weaker than φ-mixing.

All the classical results, such as law of large numbers (LLN) and central limit theorem (CLT) are available under the α-mixing framework. We present here only the weak invariance principle (WIP) which will be needed later several times. Weak invariance principle, also known as a functional central limit theorem, states a weak convergence of the partial sum process to the standard Wiener process. Thus we put

$$S_n := \sum_{i=1}^{n} Y_i \quad \text{and} \quad \sigma_n^2 := \operatorname{var} S_n$$

and define random elements on Skorokhod space $D[0,1]$ as follows

$$W_n(t) := \frac{S_{\lfloor nt \rfloor}}{\sigma_n}, \quad 0 \le t \le 1,$$

where $\lfloor \cdot \rfloor$ denotes the integer part function.

Theorem 1.1 (WIP for α-mixing)**.** *Let $\{Y_n\}_n$ be a sequence of zero mean α-mixing random variables with*

$$\sup_{n \in \mathbb{N}} \mathrm{E}\,|Y_n|^{2+\delta} < \infty \quad \text{and} \quad \sum_{n=1}^{\infty} \alpha(n)^{\delta/(2+\delta)} < \infty$$

for some $\delta > 0$. Suppose further that

$$\mathrm{E}\, S_n^2/n \to \sigma^2, \quad n \to \infty$$

for some $\sigma > 0$. Then

$$W_n(\cdot) \xrightarrow{\mathcal{D}[0,1]} W(\cdot), \quad n \to \infty,$$

where $\{W(t),\, 0 \le t \le 1\}$ is a standard Wiener process.

Proof. See Lin and Lu [2010] (Corollary 3.2.1).
Since the central limit theorem is just a special case of the WIP it can be obtained as a simple corollary of the previous theorem. Moreover as described in Davidson [1994] Section 13.4, a stationary strong mixing sequence is mixing in ergodic theory sense (this explains the name 'strong mixing') and this further implies ergodicity. Thus for stationary α-mixing sequences we can use the ergodic theorem. Another, more technical results on α-mixing are formulated in Appendix in Lemmas I and II.

We finish the part on mixing with an obvious but very useful observation about functions of mixing sequences. Since the (uniformly) strong mixing is property of σ-algebras generated by random elements, any Borel measurable function of these elements is also (uniformly) strong mixing with the same rate. More generally we have the following theorem.

Theorem 1.2. *Let $X_t = f(Y_t, \dots, Y_{t-p})$ be a measurable function for p finite. If Y_t is α-(φ-)mixing then X_t is also and the rate is the same.*

Proof. See Theorem 14.1 of Davidson [1994]. □

Another weak dependence framework called $L_p - m$-approximability was developed in Hörmann and Kokoszka [2010]. It is based on approximation of the original sequence by an m-dependent one. Verification of this approximability is usually simpler than verifying the mixing condition, but there is no analogue of theorem above. Moreover this property is restricted to sequences admitting representation $Y_n = f(\varepsilon_n, \varepsilon_{n-1}, \dots)$. And thus, they conclude, $L_p - m$-approximability is not directly comparable with classical mixing coefficients as introduced above. However it provides an interesting alternative.

1.6 State of Art

First we briefly summarize important contributions to recent development of sequential monitoring. The problem was treated in a number of papers. The pivotal one of Chu et al. [1996] provided a framework for others who followed.

Chu et al. considered regression setup which is briefly outlined in Section 1.3 and developed test procedures based on CUSUM (cumulative sums) type test statistic calculated from recursive residuals and a fluctuation test based on difference between estimates of the regression parameters from historical and monitoring periods.

The later test statistic was generalized in Leisch et al. [2000] to the so-called generalized fluctuation test. CUSUM type test statistics generally lose power when a change occurs long after the start of monitoring. Therefore Zeileis et al. [2005] suggested MOSUM (moving sums) type test statistic based on only last h ordinary residuals, which performs better in this respect. This was confirmed also by a simulation study they performed, comparing the three above mentioned test statistics. Both articles above assume a martingale difference structure for the errors.

A paper by Horváth et al. [2004] influenced majority of the ones mentioned below. They suggested two new kinds of test statistics to detect a change in a regression parameter working under the independence assumptions on error terms. The test statistic is based either on ordinary residuals or on the recursive ones. Both residuals are calculated from respective least squares estimates of the regression parameter. The disadvantage of these testing procedures is that they cannot detect changes in the regression coefficients which are orthogonal to the first column of the matrix $\lim_{m\to\infty} \sum_{i=1}^{m} \boldsymbol{X}_i \boldsymbol{X}_i^T$, where $\boldsymbol{X}_i$ is a vector of regressors for time i. This problem is overcome by the procedures developed by Hušková and Koubková [2005], where the test statistic is a quadratic form of weighted residuals.

Although the assumptions of independent identically distributed errors is very convenient from a mathematical point of view, it is typically violated in practice as was discussed earlier. Towards this issue, Aue et al. [2006] showed that the results of

Horváth et al. [2004] holds also for some dependent sequences including heteroskedastic augmented GARCH processes. The augmented GARCH model is a unification of numerous extensions of the popular and widely used (G)ARCH process and thus can be applied in various situations.

All the above mentioned methods use a least squares estimation. This can cause problems when outliers are present or data are heavy-tailed. Thus there was an effort to design more robust procedures. Koubková [2006] considered procedures based on least absolute deviation (LAD) estimation and also on general M-estimation, all for i.i.d. errors.

Possibility of change was studied not only for the location or regression parameter but also for other ones. For example Chochola [2008] considered change in scale. General change in distribution (i.e. infinite dimensional parameter) was considered in Hušková and Chochola [2010]. The test procedure used empirical distribution function and it was studied for both independent as well as dependent observations.

Another two papers relating to a sequential monitoring in a special regression model are recalled in Chapter 4. They deal with multivariate (one even functional) data.

There are even more papers and books devoted to the retrospective change-point analysis. For basic references see, e.g. Andrews [1993] or Csörgő and Horváth [1997]. As the retrospective analysis is not the main topic of the thesis we give not so much details about a past development and mention only papers closely related to robust procedures considered in Chapter 5. An M-estimation in the context of change-point analysis was studied in e.g. Antoch and Hušková [1989] or Hušková and Antoch [2001]. Hušková and Marušiaková [2012] extended some of known results on M-procedures for detection of changes in a location model to the situation with dependent observations, particularly when the error terms fulfill α-mixing condition. Prášková and Chochola [2013] studied a change in regression parameter when both the regressors and errors are weakly dependent in the sense of $L_p - m$-approximability.

1.7 Aim and Structure of the Thesis

Our aim is to extend some known results in change-point analysis towards more robust methods. We focus mainly on the sequential monitoring. The methods assume a stable historical period and thus it is desirable to have also robust retrospective procedures. Therefore we explore this area a bit as well. The robustness of the procedures with respect to outliers and heavy tailed observations is introduced via use of M-estimation instead of classical least squares estimation.

Another extension is towards dependent and multivariate data. It is assumed that the observations are weakly dependent, more specifically they fulfil strong mixing condition.

Our goal is to propose appropriate test statistics and study their asymptotic properties both under the null hypothesis of no change as well as under the alternatives, in order to derive proper critical values and show consistency of the tests. Finite sample properties of the tests need to be also examined. This is done in a simulation study and by application on some real data as well.

The thesis is structured as follows. In the next three chapters we develop the robust monitoring procedures in different models. Starting with a simple location model, which is generalized to a multivariate one and finishing with a special case of multivariate regression model.

Chapter 5 deals with the retrospective change-point analysis. Some results of an extensive simulation study are presented in Chapter 6, where one can also find an illustration of applications of the proposed methods. Derivation and tables of critical values for all the proposed procedures are presented in Chapter 7. Short conclusion follows afterwards.

Chapter 2

Monitoring in Location Model

In this chapter we propose robust sequential monitoring procedures for a detection of change in mean in a simple location model. The test statistic together with the assumptions is presented in Section 2.1 and its theoretical properties are studied in Section 2.2, their proofs are postponed to Sections 2.3 and 2.4. Finally, in Section 2.5, an estimator of the long-run variance is developed and is shown to have required properties.

2.1 Model, Assumptions and Test Statistic

We consider a location model

$$Y_i = \mu_i + e_i, i = 1, 2, \ldots, \tag{2.1}$$

where $\{Y_i\}_i$ are the observed data, $\{\mu_i\}_i$ the location parameters and $\{e_i\}_i$ the random errors fulfilling the assumptions specified below. Our goal is to monitor a change in mean in a robust way.

The setting was introduced in Section 1.3. The stability of the training data (non-contamination assumption) can be written as

$$\mu_1 = \ldots = \mu_m = \mu_0. \tag{2.2}$$

We test the null hypothesis of no change

$$H_0 : \mu_i = \mu_0, \quad 1 \le i \le m + mT, \tag{2.3}$$

against the alternative that at some time-point $m + k^*$ the model changes

$$\begin{aligned} H_1 : &\text{ there exists } k_m^* = k^* < mT \text{ such that} \\ &\mu_i = \mu_0,\ 1 \le i \le m + k^*, \quad \mu_i = \mu_*,\ m + k^* < i < \infty, \quad \mu_0 \ne \mu_*, \end{aligned} \tag{2.4}$$

where the values of μ_0, μ_* and k^* are unknown.

As was discussed earlier, the test procedures are typically based on the least squares estimation. Here we would like to develop a more robust procedure and therefore we will consider the detector based on CUSUM of M-residuals.

The M-residual $\psi(\widehat{e_i})$ for the score function ψ is defined as

$$\psi(\widehat{e_i}) = \psi(Y_i - \widehat{\mu}_m(\psi)), \tag{2.5}$$

where $\widehat{\mu}_m(\psi)$ is an M-estimate of μ_0 based on the training data as defined by (1.12) i.e. the solution of the minimization problem

$$\arg\min_{t\in\mathbb{R}} \sum_{i=1}^{m} \rho(Y_i - t),$$

where ρ is a convex loss function such that $\rho' = \psi$ a.e.

Then we can define

$$\widehat{S}_\psi(k,m) = \frac{1}{m^{1/2}} \sum_{i=m+1}^{m+k} \psi(\widehat{e_i}), \quad k = 1,\dots,mT \tag{2.6}$$

and a standardized test statistics

$$\widehat{Q}_\psi(m,k) = \frac{\left|\widehat{S}_\psi(k,m)\right|}{\widehat{\sigma}_m(\psi)}, \quad k = 1,\dots,mT, \tag{2.7}$$

where $\widehat{\sigma}_m(\psi)$ is a proper standardization. Since we have dependent data, $\widehat{\sigma}_m^2(\psi)$ is an estimate of the so-called *long-run variance* (LRV) of the test statistic, i.e. of the quantity

$$\sigma^2(\psi) = \lim_{m\to\infty} \operatorname{var}\Big(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i)\Big) = \mathrm{E}\,\psi^2(e_1) + 2\sum_{i=1}^{\infty} \mathrm{E}\,\psi(e_1)\psi(e_{1+i}). \tag{2.8}$$

We consider a class of boundary functions introduced in Horváth et al. [2004], namely

$$q_\gamma(t) = (1+t)\left(\frac{t}{1+t}\right)^\gamma, \ t\in(0,\infty), \quad \gamma\in[0,1/2). \tag{2.9}$$

The constant γ is called a *tuning parameter* as it allows us to tailor the procedure according to our expectation of possible time of change. More details are presented in Section 6.1.2.

As was already written in the general setup at the end of Section 1.3, the monitoring scheme is described through the stopping time $\tau_{m,T}$ defined as

$$\tau_{m,T} = \inf\{1 \le k \le mT : \widehat{Q}_\psi(m,k) > c_{m,T}(\alpha)\, q_\gamma(k/m)\}, \tag{2.10}$$

and the null hypothesis is rejected and the observation is stopped, whenever the ratio of $\widehat{Q}_\psi(m,k)$ and $q_\gamma(k/m)$ exceeds the critical value.

Assumptions
Now we discuss the assumptions under which we will be working:

(A.1) $\{e_i\}_i$ is a strictly stationary α-mixing sequence with coefficients $\{\alpha(i)\}_i$ and with a distribution function F such that for some $\Delta > 0$:

$$\sum_{k=0}^{\infty} \alpha(k)^{\Delta/(2+\Delta)} < \infty. \tag{2.11}$$

The score function ψ, the function $\lambda(t) = -\int \psi(e-t)dF(e)$, $t \in \mathbb{R}$ and the distribution function F satisfy:

(A.2) ψ is non-decreasing; the derivative $\lambda'(\cdot)$ of the function $\lambda(\cdot)$ exists and is Lipschitz in a neighborhood of zero, $\lambda(0) = 0$ and $\lambda'(0) > 0$.

(A.3) $\int |\psi(x)|^{2+\Delta} dF(x) < \infty$ and

$$\int |\psi(x + t_2) - \psi(x + t_1)|^{2+\Delta} dF(x) \leq D_1 |t_2 - t_1|^a, \quad |t_j| \leq D_2,\ j = 1, 2$$

for $1 \leq a \leq 2 + \Delta$, constant $\Delta > 0$ from assumption (A.1) and some positive constants D_1, D_2 depending on Δ.

(A.4) It holds
$$0 < \sigma^2(\psi) < \infty.$$

From Section 1.5 follows that the assumption (A.1) is satisfied for a quite large spectrum of situations. Due to the α-mixing property of $\{e_i\}_i$, $\{\psi(e_i)\}_i$ is also α-mixing with the same coefficients. Assumptions (A.2) – (A.3) are standard assumptions imposed on the score function ψ and the error distribution F. Theorem 1.1 asserts that under (A.1) is $\sigma^2(\psi) < \infty$ i.e. it is enough to assume just positivity in (A.4).

These assumptions are similar to the one considered in Hušková and Marušiaková [2012] in the retrospective setting, however it will be shown that we do not need their stronger assumption on the mixing rate, which can be relaxed to (2.11). This is however a standard condition for CLT to hold for strong mixing sequence and thus the assumptions are not overly restrictive.

Now, we look more specifically on the assumptions for typical choices of ψ's introduced in Section 1.4:

(a) For $\psi(x) = x$, $x \in \mathbb{R}$, the procedures reduce to classical L_2 ones. Assumptions (A.2) - (A.3) reduce to the $(2 + \Delta)$ moment restrictions, $a = 2 + \Delta$.

(b) For $\psi(x) = \operatorname{sign} x$, $x \in \mathbb{R}$, the procedures reduce to L_1 procedures. Assumptions (A.2) - (A.3) are satisfied if the error distribution F is symmetric with a continuous density f in a neighborhood of 0 with $f(0) > 0$. In this case $a = 1$ for any $\Delta > 0$.

(c) For Huber ψ function (see (1.19)) assumptions (A.2) - (A.3) are satisfied if the distribution function F is symmetric and if there exists a continuous density f in a neighborhood of $\pm K$ with $f(\pm K) > 0$. In this case $a = 2 + \Delta$.

2.2 Main Results

Here we formulate assertions on limit behavior of the test statistic under both null and alternative hypotheses.

Theorem 2.1. *Assume that $Y_1, Y_2, \ldots$ follow the model (2.1), assumptions (A.1) – (A.4) are satisfied. Further let $\widehat{\sigma}_m^2(\psi)$ be a consistent estimator of $\sigma^2(\psi)$, i.e.*

$$\widehat{\sigma}_m^2(\psi) - \sigma^2(\psi) = o_P(1). \tag{2.12}$$

Then under the null hypothesis (2.3) holds

$$\lim_{m\to\infty} P\left(\max_{1\le k\le mT} \frac{|\widehat{S}_\psi(k,m)|}{\widehat{\sigma}_m(\psi) q_\gamma(k/m)} \le c\right) = P\left(\sup_{0\le t\le T/(T+1)} \frac{|W(t)|}{t^\gamma} \le c\right) \tag{2.13}$$

for all $c > 0$, where $\{W(t),\, t \in [0,1]\}$ is Wiener process and the test detector is defined in (2.6)–(2.9) *with $T > 0$ fixed.*

The limit distribution can be used for approximation of the critical value $c_{m,T}(\alpha)$ introduced in (1.5). We can choose it as a value $c_T(\alpha, \gamma)$ such that

$$P\left(\sup_{0\le t\le T/(T+1)} \frac{|W(t)|}{t^\gamma} > c_T(\alpha,\gamma)\right) = \alpha, \tag{2.14}$$

where the dependence on the tuning parameter γ from the boundary function $q_\gamma(\cdot)$ is also denoted. Details, together with a table of these critical values can be found in Chapter 7.

Another way of finding the critical values is applying a suitable version of resampling methods. Particularly, the block bootstrap studied in Kirch [2006] can be adjusted to our situation.

Theorem 2.1 also accounts for the range of constant γ, since for $\gamma \ge 1/2$ the random process $W(t)/t^\gamma$ converges to infinity as $t \to 0+$ almost surely. As Aue et al. [2008] indicates $\gamma = 1/2$ can be also used in the boundary function, but it leads to a different asymptotic distribution of the detector and therefore we will consider only $\gamma \in [0, 1/2)$.

Now we consider asymptotic behavior of the test statistic under local alternatives.

Theorem 2.2. *Assume that $Y_1, Y_2, \ldots$ follow the model (2.1) under the alternative* (2.4) *with $k^* = k_m^* = \lfloor m\tau \rfloor$ for some $\tau \in [0, T)$ and $\mu_* = \mu_0 + \theta\delta_m$. Let assumptions (A.1) – (A.4) be satisfied and test detector is defined in* (2.6)–(2.9) *with $T > 0$ fixed. Further let $\widehat{\sigma}_m^2(\psi)$ be a consistent estimator of $\sigma^2(\psi)$, i.e. $\widehat{\sigma}_m^2(\psi) - \sigma^2(\psi) = o_P(1)$.*
(i) If $\delta_m = m^{-1/2}$, then

$$\max_{1\le k\le mT} \frac{\widehat{Q}_\psi(m,k)}{q_\gamma(k/m)} \xrightarrow{\mathcal{D}} \sup_{0\le t\le T/(T+1)} \frac{\left|W(t) + \theta\lambda'(0)\sigma^{-1}(\psi)p(t,\tau)\right|}{t^\gamma}, \tag{2.15}$$

where $\{W(t),\, 0 \le t \le 1\}$ is a Wiener process and

$$p(t,s) = (t - s(1-t))^+,\ 0 \le t \le T/(T+1),\ 0 \le s < T, \tag{2.16}$$

with $(x)^+ = \max(0, x)$.
(ii) If $\delta_m \to 0$ *and* $|\delta_m| m^{1/2} \to \infty$*, then*

$$\max_{1\le k\le mT} \frac{\widehat{Q}_\psi(m,k)}{q_\gamma(k/m)} \xrightarrow{P} \infty.$$

Theorem 2.2 (i) deals with the so-called *contiguous alternatives*. The asymptotic distribution is a maximum of absolute value of shifted Wiener process, where the shift depends on the change-point, the amount of change and also on the choice of the loss function via $\lambda'(0)$.
Theorem 2.2 (*ii*) ensures that the requirement (1.7) is fulfilled i.e. that the true change will be detected with probability tending to 1 as $m \to \infty$.

2.3 Auxiliary Results

Key tool in proving previous theorems is an asymptotic representation of M-estimate similar to (1.18). To show that under the assumptions (A.1)–(A.4) we proceed similarly as in Hušková and Marušiaková [2012], just assuring that our slightly weaker assumptions are sufficient and giving more detailed reasoning. Following Lemmas 2.1 – 2.4 are slight modifications of those of Hušková and Marušiaková [2012], Lemma 2.5 is new as it deals with the monitoring period.

In the sequel, $C > 0$ is a generic constant, which may vary from case to case. We also use A_m and similar to denote different variables in different lemmas.
Lemma 2.1. *Let assumptions (A.1) – (A.3) be satisfied. Then*

$$\sup_{|t|\le D} \frac{1}{\sqrt{m}(\sqrt{m}a_m)} \left| \sum_{i=1}^{m} \big(\psi(e_i - ta_m) - \psi(e_i) - \mathrm{E}\,\psi(e_i - ta_m)\big) \right| = O_P(m^{-\eta}) \qquad (2.17)$$

for some $\eta > 0$*, any* $D > 0$ *and any* $m^{-1/2} \le a_m \le D_2$*, where* D_2 *is from assumption (A.2).*

Proof. For proving the uniform result we make use of the monotonicity of ψ function. The interval $[-D, D]$ is split into segments $[t_{j-1}, t_j]$, defined by

$$-D = t_0 < t_1 < \cdots < t_N = D, \quad t_j - t_{j-1} = m^{-\kappa}, j < N \text{ and } D - t_{N-1} \le m^{-\kappa},$$

where $\kappa > 0$ is appropriately chosen.

Denote $y_i(t) := \psi(e_i - ta_m) - \psi(e_i),\ t \in \mathbb{R}$. Since $\psi(\cdot)$ is nondecreasing then $y_i(\cdot)$ is nonincreasing and thus we have for $t \in [t_{j-1}, t_j]$

$$y_i(t_j) - \mathrm{E}\,y_i(t_{j-1}) \le y_i(t) - \mathrm{E}\,y_i(t) \le y_i(t_{j-1}) - \mathrm{E}\,y_i(t_j).$$

From here easily follows

$$\begin{aligned}
&\big(y_i(t_j) - \mathrm{E}\,y_i(t_j)\big) - \big(\mathrm{E}\,y_i(t_{j-1}) - \mathrm{E}\,y_i(t_j)\big) \le \\
&\qquad \le y_i(t) - \mathrm{E}\,y_i(t) \le \\
&\qquad \le \big(y_i(t_{j-1}) - \mathrm{E}\,y_i(t_{j-1})\big) + \big(\mathrm{E}\,y_i(t_{j-1}) - \mathrm{E}\,y_i(t_j)\big).
\end{aligned}$$

Since $\big(\mathrm{E}\, y_i(t_{j-1}) - \mathrm{E}\, y_i(t_j)\big) \geq 0$ we get from previous that

$$\sup_{|t|\leq D} \left| \sum_{i=1}^{m} (y_i(t) - \mathrm{E}\, y_i(t)) \right| \leq \max_{1\leq j\leq N} \left| \sum_{i=1}^{m} (y_i(t_j) - \mathrm{E}\, y_i(t_j)) \right| + \max_{1\leq j\leq N} \sum_{i=1}^{m} \mathrm{E}\big(y_i(t_{j-1}) - y_i(t_j)\big). \tag{2.18}$$

Clearly, for the second term on RHS holds

$$\sum_{i=1}^{m} \mathrm{E}\big(y_i(t_{j-1}) - y_i(t_j)\big) = m\big(\lambda(t_j a_m) - \lambda(t_{j-1} a_m)\big) = m\lambda'(0) a_m (t_j - t_{j-1})(1 + o(1)) \leq$$
$$\leq C a_m m^{1-\kappa}(1 + o(1)) \tag{2.19}$$

uniformly in $j = 1, \ldots, N$. Thus the second term on RHS in (2.18) divided by $\sqrt{m}(\sqrt{m}a_m)$ is surely $O_P(m^{-\eta})$ for some $\eta > 0$.

For the first term on RHS in (2.18) we have

$$P\left(m^{\eta} \max_{1\leq j\leq N} \left| \sum_{i=1}^{m} (y_i(t_j) - \mathrm{E}\, y_i(t_j)) \right| \geq A\sqrt{m}(\sqrt{m}a_m) \right) \leq$$
$$\leq \sum_{j=1}^{N} P\left(\left| \sum_{i=1}^{m} (y_i(t_j) - \mathrm{E}\, y_i(t_j)) \right| \geq A m^{1/2-\eta}(\sqrt{m}a_m) \right) \leq$$
$$\leq \sum_{j=1}^{N} \mathrm{E}\left(\left| \sum_{i=1}^{m} (y_i(t_j) - \mathrm{E}\, y_i(t_j)) \right| \right)^2 A^{-2} m^{-1+2\eta} (\sqrt{m}a_m)^{-2} \tag{2.20}$$

By Lemma II

$$\mathrm{E}\left(\left| \sum_{i=1}^{m} (y_i(t_j) - \mathrm{E}\, y_i(t_j)) \right| \right)^2 \leq Cm \left\| y_i(t_j) - \mathrm{E}\, y_i(t_j) \right\|_{2+\Delta}^2, \tag{2.21}$$

where the L_p norm is defined as $\|X\|_p = (\mathrm{E}\,|X|^p)^{1/p}$.

Moreover by the assumptions on ψ and λ we get

$$\|y_i(t_j) - \mathrm{E}\, y_i(t_j)\|_{2+\Delta} = \|\psi(e_i - t_j a_m) - \psi(e_i) - \mathrm{E}\,\psi(e_i - t_j a_m)\|_{2+\Delta} \leq$$
$$\leq \|\psi(e_i - t_j a_m) - \psi(e_i)\|_{2+\Delta} + |\mathrm{E}\,\psi(e_i - t_j a_m)| \leq$$
$$\leq C(t_j a_m)^{a/(2+\Delta)} + C\lambda'(0) t_j a_m \leq C a_m^{a/(2+\Delta)} \tag{2.22}$$

uniformly in $j \in 1, \ldots, N$.
Thus from (2.21) and (2.22), RHS of (2.20) can be estimated from above by

$$\frac{C}{A^{-2}} m^{1+\kappa}\, a_m^{2a/(2+\Delta)}\, m^{-1+2\eta} (\sqrt{m}a_m)^{-2} = \frac{C}{A^{-2}} m^{\kappa+2\eta-1} a_m^{-2(1-a/(2+\Delta))}. \tag{2.23}$$

RHS of (2.23) is maximal for $a_m = m^{-1/2}$ and $a = 1$. This is however still only $O(m^{\kappa+2\eta-1/(2+\Delta)})$ and thus (2.23) can be made small enough for properly chosen κ and η, which concludes the proof.

□

Lemma 2.1 has few useful consequences, which are formulated in the following corollary.

Corollary 2.2. *Let assumptions (A.1) – (A.3) be satisfied. Then, for some* $\eta > 0$

$$\text{(i)} \quad \sup_{|t|\le D} \frac{1}{\sqrt{m}} \left| \sum_{i=1}^{m} \psi(e_i - tm^{-1/2}) - \psi(e_i) - \mathrm{E}\,\psi(e_i - tm^{-1/2}) \right| = O_P(m^{-\eta}) \tag{2.24}$$

$$\text{(ii)} \quad \sup_{|t|\le D} \frac{1}{\sqrt{m}} \left| \sum_{i=1}^{m} \psi(e_i - tm^{-1/2}) - \mathrm{E}\,\psi(e_i - tm^{-1/2}) \right| = O_P(1) \tag{2.25}$$

$$\text{(iii)} \quad \frac{1}{\sqrt{m}} \sum_{i=1}^{m} |\psi(e_i - a_m) - \mathrm{E}\,\psi(e_i - a_m)| = O_P(1), \ \textit{where } |a_m| \le D_2. \tag{2.26}$$

Proof. The proof is very easy.
(i) is just Lemma 2.1 with $a_m = m^{-1/2}$.
(ii) follows from (i) and $\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i) = O_P(1)$.
(iii) can be easily seen from the proof of Lemma 2.1. □

Previous Corollary provides the key part for the proof of Proposition 1.3 and thus this proposition gives us the root consistency of our estimate i.e.

$$\sqrt{m}(\widehat{\mu}_m(\psi) - \mu_0) = O_P(1). \tag{2.27}$$

Now we prove that even for the general definition of M-estimates (1.12), the sum $\sum_{i=1}^{m} \psi(Y_i - \widehat{\mu}_m(\psi))$ is small in probability. For the ease of writing the score function ψ is dropped from the notation of the M-estimate and without loss of a generality (WLOG) it is assumed that $\mu_0 = 0$.

Lemma 2.3. *Let assumptions (A.1) – (A.3) be satisfied. Then*

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i - \widehat{\mu}_m) = O_P(m^{-\eta})$$

for some $\eta > 0$.

Proof. Let $a_m > 0$ with maximal order of $O(m^{-1/2})$, the precise one will be specified later. Due to monotonicity of ψ we have

$$\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i - \widehat{\mu}_m - a_m) \le \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i - \widehat{\mu}_m) \le \frac{1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i - \widehat{\mu}_m + a_m),$$

where the lower bound is ≤ 0 and the upper one is ≥ 0 by the definition of $\widehat{\mu}_m$. Thus

$$\begin{aligned}\left|\frac{1}{\sqrt{m}}\sum_{i=1}^m \psi(e_i-\widehat{\mu}_m)\right| &\leq \frac{1}{\sqrt{m}}\sum_{i=1}^m \psi(e_i-\widehat{\mu}_m+a_m)-\frac{1}{\sqrt{m}}\sum_{i=1}^m \psi(e_i-\widehat{\mu}_m-a_m)= \\ &=\left[\frac{1}{\sqrt{m}}\sum_{i=1}^m\big(\psi(e_i-\widehat{\mu}_m+a_m)-\psi(e_i)+\sqrt{m}\lambda(\widehat{\mu}_m-a_m)\big)\right] \\ &-\left[\frac{1}{\sqrt{m}}\sum_{i=1}^m\big(\psi(e_i-\widehat{\mu}_m-a_m)-\psi(e_i)+\sqrt{m}\lambda(\widehat{\mu}_m+a_m)\big)\right] \\ &+\left[\sqrt{m}\lambda(\widehat{\mu}_m+a_m)-\sqrt{m}\lambda(\widehat{\mu}_m-a_m)\right] \\ &=: A_m^- - A_m^+ + A_m.\end{aligned}$$

To show that $A_m^- = O_P(m^{-\eta})$ we use

$$P(m^\eta|A_m^-|>K)=P(m^\eta|A_m^-|>K,\sqrt{m}|\widehat{\mu}_m|\leq\widetilde{K})+P(m^\eta|A_m^-|>K,\sqrt{m}|\widehat{\mu}_m|>\widetilde{K}). \tag{2.28}$$

Since $\sqrt{m}|\widehat{\mu}_m| = O_P(1)$, we can for every $\varepsilon > 0$ find $\widetilde{K}$ such that the second probability on RHS of (2.28) is smaller than ε. For the first part we can use Corollary 2.2 (i) to find K such that for m large enough the probability is again smaller than ε. Similarly we obtain $A_m^+ = O_P(m^{-\eta})$.

Previous reasoning is a standard way of treating functions of M-estimates and will not be presented in such a detail further. We see that the key ingredient is the uniformity of the result, the rest is straightforward.

Towards A_m we have by assumption (A.2)

$$0\leq\lambda(\widehat{\mu}_m+a_m)-\lambda(\widehat{\mu}_m-a_m)=\int_{\widehat{\mu}_m-a_m}^{\widehat{\mu}_m+a_m}\lambda'(x)dx=\int_{\widehat{\mu}_m-a_m}^{\widehat{\mu}_m+a_m}(\lambda'(x)-\lambda'(0))dx+2\lambda'(0)a_m\leq$$

$$\leq\int_{\widehat{\mu}_m-a_m}^{\widehat{\mu}_m+a_m}C\,xdx+2\lambda'(0)a_m=2C\widehat{\mu}_m a_m+2\lambda'(0)a_m\leq Ca_m(1+o_P(1)).$$

Choosing $a_m = m^{-1/2-\eta}$ we get that $A_m = O_P(m^{-\eta})$, which concludes the proof. □

Lemma 2.4 (Asymptotic representation). *Let assumptions (A.1) – (A.3) be satisfied. Then*

$$\sqrt{m}\widehat{\mu}_m=\frac{1}{\lambda'(0)\sqrt{m}}\sum_{i=1}^m\psi(e_i)+O_P(m^{-\eta}) \tag{2.29}$$

for some $\eta > 0$.

Proof. Define

$$A_m:=\frac{1}{\sqrt{m}}\sum_{i=1}^m\left(\psi(e_i-\widehat{\mu}_m)-\psi(e_i)+\lambda(\widehat{\mu}_m)\right). \tag{2.30}$$

Recall that we use notation $\lambda(t) = -\operatorname{E}\psi(e_i - t)$, which is especially useful when the argument is random.

First we show that

$$A_m = O_P(m^{-\eta}). \tag{2.31}$$

Using the fact that $\widehat{\mu}_m = O_P(m^{-1/2})$ similarly as in Lemma 2.3 it is enough to concentrate on the set where $|\widehat{\mu}_m| < Km^{-1/2}$. And thus we can conclude by Corollary 2.1 (i) that (2.31) holds.

On the other hand using Lemma 2.3 we have

$$\begin{aligned} A_m &= \frac{1}{\sqrt{m}} \left(-\sum_{i=1}^{m} \psi(e_i) + m\lambda(\widehat{\mu}_m) \right) + O_P(m^{-\eta}) = \\ &= \frac{-1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i) + \sqrt{m} \left(\lambda'(0)\widehat{\mu}_m + O_P(m^{-1}) \right) + O_P(m^{-\eta}) = \\ &= \frac{-1}{\sqrt{m}} \sum_{i=1}^{m} \psi(e_i) + \sqrt{m}\lambda'(0)\widehat{\mu}_m + O_P(m^{-\eta}) \end{aligned} \tag{2.32}$$

for some $\eta > 0$. Combining (2.31) and (2.32), the asymptotic representation easily follows. □

Now we can formulate an analogue of Lemma 2.1 focusing however on the monitoring period.

Lemma 2.5 (Monitoring period). *Let assumption (A.1) – (A.3) be satisfied. Then for $\gamma < 1/2$ holds*

$$\sup_{|t| \le D} \max_{1 \le k \le mT} \frac{1}{\sqrt{m} \left(\frac{k}{m}\right)^{\gamma}} \left| \sum_{i=m+1}^{m+k} \psi(e_i - tm^{-1/2}) - \psi(e_i) - \mathrm{E}\, \psi(e_i - tm^{-1/2}) \right| = O_P(m^{-\zeta}),$$

for some $\zeta > 0$.

Proof. Analogous to Lemma 2.1. We define again auxiliary variables $z_i(t)$, but now only with $m^{-1/2}$ instead of general a_m and shifted index, i.e.

$$z_i(t) = \psi(e_{m+i} - tm^{-1/2}) - \psi(e_{m+i}), \quad i = 1, \dots, mT,\ t \in \mathbb{R}$$

and also the grid

$$-D = t_0 < t_1 < \cdots < t_N = D, \quad t_j - t_{j-1} = m^{-\kappa}, j < N \text{ and } D - t_{N-1} \le m^{-\kappa},$$

for some $\kappa > 0$. The analogue of (2.18) is

$$\sup_{|t| \le D} \left| \sum_{i=1}^{k} (z_i(t) - \mathrm{E}\, z_i(t)) \right| \le \max_{1 \le j \le N} \left| \sum_{i=1}^{k} (z_i(t_j) - \mathrm{E}\, z_i(t_j)) \right| + \max_{1 \le j \le N} \sum_{i=1}^{k} |\mathrm{E}\,(z_i(t_j) - z_i(t_{j-1}))| \tag{2.33}$$

and similarly to (2.19) we have uniformly in $1 \le j \le N$ that

$$|\mathrm{E}\,(z_i(t_j) - z_i(t_{j-1}))| \le Cm^{-\zeta} m^{-1/2}.$$

Thus if we split the LHS of the statement of this Lemma according to (2.33), we get for the second part that

$$\max_{1\le j\le N}\max_{1\le k\le mT}\frac{1}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\sum_{i=1}^{k}\left|\mathrm{E}\left(z_i(t_j)-z_i(t_{j-1})\right)\right|\le C\max_{1\le k\le mT}\frac{km^{-\kappa-1/2}}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}=O(m^{-\kappa}).$$

And for the first part

$$\begin{aligned}
&P\left(\max_{1\le j\le N}\max_{1\le k\le mT}\frac{1}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\left|\sum_{i=1}^{k}\left(z_i(t_j)-\mathrm{E}\,z_i(t_j)\right)\right|\ge A\right)\le\\
&\le\sum_{j=1}^{N}P\left(\max_{1\le k\le mT}k^{-\gamma}\left|\sum_{i=1}^{k}\left(z_i(t_j)-\mathrm{E}\,z_i(t_j)\right)\right|\ge Am^{1/2-\gamma}\right)\le\\
&\le C\sum_{j=1}^{N}\mathrm{E}\left(\max_{1\le k\le mT}k^{-\gamma}\left|\sum_{i=1}^{k}\left(z_i(t_j)-\mathrm{E}\,z_i(t_j)\right)\right|\right)^{2}m^{-(1-2\gamma)}.
\end{aligned}\tag{2.34}$$

From Lemma II we have

$$\mathrm{E}\Big(\max_{1\le k\le mT}k^{-\gamma}\Big|\sum_{i=1}^{k}\left(z_i(t_j)-\mathrm{E}\,z_i(t_j)\right)\Big|\Big)^{2}\le D(\log 2Tm)^2\sum_{k=1}^{mT}k^{-2\gamma}\left(\|z_i(t_j)-\mathrm{E}\,z_i(t_j)\|_{2+\Delta}\right)^2.$$

Since

$$N=O(m^{\kappa}),\quad \sum_{k=1}^{mT}k^{-2\gamma}=O(m^{1-2\gamma})\text{ for }\gamma<1/2$$

and from (2.22) is

$$\left(\|z_i(t_j)-\mathrm{E}\,z_i(t_j)\|_{2+\Delta}\right)^2=O(m^{-a/(2+\Delta)}),$$

we have that RHS of (2.34) is of order $O\left((\log m)^2m^{\kappa-a/(2+\Delta)}\right)$. Thus there surely exists $\zeta>0$ such that

$$\max_{1\le j\le N}\max_{1\le k\le mT}\frac{1}{\sqrt{m}\left(\frac{k}{m}\right)^{\gamma}}\left|\sum_{i=1}^{k}\left(z_i(t_j)-\mathrm{E}\,z_i(t_j)\right)\right|=O_P(m^{-\zeta}),$$

which concludes the proof. □

The next lemma allows us to approximate the M-residuals with their theoretical counterparts.

Lemma 2.6. *Let the assumptions of Theorem 2.1 be satisfied. Then*

$$\max_{1\le k\le mT}\frac{1}{m^{1/2}q_{\gamma}(k/m)}\left|\sum_{i=m+1}^{m+k}\psi(\widehat{e}_i)-\left(\sum_{i=m+1}^{m+k}\psi(e_i)-\frac{k}{m}\sum_{j=1}^{m}\psi(e_j)\right)\right|=o_P(1).\tag{2.35}$$

Proof. One can find constants $0 < c_1 < c_2 < \infty$ such that

$$c_1 < \max_{1\le k\le mT} \frac{m^{1/2}\left(\frac{k}{m}\right)^\gamma}{m^{1/2}q_\gamma(k/m)} < c_2.$$

Thus it is sufficient to prove the lemma with $m^{1/2}\left(\frac{k}{m}\right)^\gamma$ instead of $m^{1/2}q_\gamma(k/m)$. Recall that $\widehat{e}_i = e_i - \widehat{\mu}_m$. We make use of the fact that $\sqrt{m}\widehat{\mu}_m = O_P(1)$ similarly as in Lemmas 2.3 and 2.4 and thus it is enough to concentrate on the event $\{|\widehat{\mu}_m| \le Km^{-1/2}\}$. This allows us to use Lemma 2.5. For

$$A_m := \max_{1\le k\le mT} \frac{1}{m^{1/2}\left(\frac{k}{m}\right)^\gamma} \sum_{i=m+1}^{m+k} \psi(\widehat{e}_i)$$

we thus get

$$A_m = \max_{1\le k\le mT} \frac{1}{m^{1/2}\left(\frac{k}{m}\right)^\gamma} \left(\sum_{i=m+1}^{m+k} \psi(e_i) + k\,\mathrm{E}\,\psi(e_i - z)\Big|_{z=\widehat{\mu}_m} \right) + O_P(m^{-\zeta}) \tag{2.36}$$

for some $\zeta > 0$. Further

$$\mathrm{E}\,\psi(e_i - z)\Big|_{z=\widehat{\mu}_m} = -\lambda(\widehat{\mu}_m) = -\lambda'(0)\widehat{\mu}_m + O_P(m^{-1}) = -m^{-1}\sum_{j=1}^{m} \psi(e_i) + O_p(m^{-1/2-\eta}),$$

where we used the properties of λ and the asymptotic representation of $\widehat{\mu}_m$ derived in Lemma 2.4. Since

$$\max_{1\le k\le mT} \frac{k}{m^{1/2}\left(\frac{k}{m}\right)^\gamma} O_p(m^{-1/2-\eta}) = O_P(m^{-\eta}),$$

we get

$$A_m = \max_{1\le k\le mT} \frac{1}{m^{1/2}\left(\frac{k}{m}\right)^\gamma} \left(\sum_{i=m+1}^{m+k} \psi(e_i) - \frac{k}{m}\sum_{j=1}^{m} \psi(e_j) \right) + O_P(m^{-\eta}) + O_P(m^{-\zeta}),$$

which concludes the proof.

□

2.4 Proofs of Main Results

Proof of Theorem 2.1. Lemma 2.6 implies that the limit behavior of

$$\max_{1\le k\le mT} \frac{1}{m^{1/2}q_\gamma(k/m)} \left| \sum_{i=m+1}^{m+k} \psi(\widehat{e}_i) \right|$$

is the same as that of

$$\max_{1\le k\le mT} \frac{1}{m^{1/2}q_\gamma(k/m)} \left| \sum_{i=m+1}^{m+k} \psi(e_i) - \frac{k}{m}\sum_{j=1}^{m} \psi(e_j) \right|.$$

Now we use the WIP for mixing sequence (see Theorem 1.1), i.e. the fact that the process

$$\left\{\frac{1}{\sigma(\psi)\sqrt{m}}\sum_{i=1}^{\lfloor mt\rfloor}\psi(e_i),\ t\in[0,T+1]\right\}$$

converges to a Wiener process $\{W(t),\ t\in[0,T+1]\}$ in Skorokhod topology on $D[0,T+1]$ (denoted by $\overset{\mathcal{D}[0,T+1]}{\longrightarrow}$). Thus defining a process $\{Z_m(t),\ t\in[0,T]\}$ as

$$Z_m(t)=\frac{1}{\sigma(\psi)\sqrt{m}}\Big(\sum_{i=m+1}^{m+\lfloor mt\rfloor}\psi(e_i)-t\sum_{j=1}^{m}\psi(e_j)\Big) \tag{2.37}$$

we get

$$\{Z_m(t),\ t\in[0,T]\}\overset{\mathcal{D}[0,T]}{\longrightarrow}\{W(1+t)-W(1)-tW(1),\ t\in[0,T]\}.$$

Using the properties of Wiener process we can write

$$\{W(1+t)-W(1)-tW(1),\ t\in[0,T]\}\overset{D}{=}\{W_1(t)-tW_2(1),\ t\in[0,T]\},$$

where $\{W_1(t),t\in[0,T]\}$ and $\{W_2(t),t\in[0,T]\}$ are two independent Wiener processes. Computing the covariance functions one can easily verify that

$$\{W_1(t)-tW_2(1),\ t\in[0,T]\}\overset{D}{=}\left\{(1+t)W\left(\frac{t}{1+t}\right),\ t\in[0,T]\right\},$$

where $\{W(t),\ t\in[0,1]\}$ is again a Wiener process. Hence

$$\sup_{0\le t\le T}\frac{|W_1(t)-tW_2(1)|}{(1+t)\left(\frac{t}{1+t}\right)^{\gamma}}\overset{D}{=}\sup_{0\le t\le T}\frac{\left|W\left(\frac{t}{1+t}\right)\right|}{\left(\frac{t}{1+t}\right)^{\gamma}}\overset{D}{=}\sup_{0\le t\le T/(T+1)}\frac{|W(t)|}{t^{\gamma}}, \tag{2.38}$$

since $\frac{|W(t)|}{t^{\gamma}}$ is continuous at $t=0$ a.s. for all $\gamma\in[0,1/2)$. Thus finally, since we assume $\widehat{\sigma}^2_m(\psi)-\sigma^2(\psi)=o_P(1)$, we get

$$\max_{1\le k\le mT}\frac{|\widehat{S}_\psi(k,m)|}{\widehat{\sigma}_m(\psi)q_\gamma(k/m)}\overset{D}{\to}\sup_{0\le t\le T/(T+1)}\frac{|W(t)|}{t^{\gamma}},$$

which concludes the proof. □

Proof of Theorem 2.2. Assume WLOG again that $\mu_0=0$. Then

$$\widehat{e}_i=\begin{cases}e_i-\widehat{\mu}_m & i\le m+k^*_m,\\ e_i+\theta\delta_m-\widehat{\mu}_m & i>m+k^*_m.\end{cases}$$

(i)
Since $\delta_m=m^{-1/2}$, i.e. is of the same order as $\widehat{\mu}_m$, we can proceed as in Lemma 2.6. The main part within the brackets of (2.36) is now

$$\Big(\sum_{i=m+1}^{m+k}\psi(e_i)-\sum_{i=m+1}^{m+k}\lambda\big(\widehat{\mu}_m-\theta\delta_m I\{i>m+k^*_m\}\big)\Big).$$

Focusing on the second term for $k \geq k_m^*$ we have

$$\sum_{i=m+1}^{m+k} \lambda\big(\widehat{\mu}_m - \theta\delta_m I\{i > m + k_m^*\}\big) = k_m^* \lambda(\widehat{\mu}_m) + (k - k_m^*)\lambda(\widehat{\mu}_m - \theta\delta_m) =$$

$$= k\lambda'(0)\widehat{\mu}_m - \theta\delta_m\lambda'(0)(k - k_m^*) + C_m,$$

where the remainder term C_m can be treated as in Lemma 2.6. Thus the analogue of (2.35) is

$$\max_{1\leq k\leq mT} \frac{\left|\sum_{i=m+1}^{m+k} \psi(\widehat{e}_i) - \left(\sum_{i=m+1}^{m+k} \psi(e_i) - \frac{k}{m}\sum_{j=1}^{m} \psi(e_j) + \theta\delta_m\lambda'(0)(k - k_m^*)^+\right)\right|}{m^{1/2} q_\gamma(k/m)} = o_P(1). \tag{2.39}$$

Following the lines of proof of Theorem 2.1 we define the analogue of process $\{Z_m(t)\}$ as

$$Z_m^*(t) = Z_m(t) + \frac{1}{\sigma(\psi)\sqrt{m}}\theta\delta_m\lambda'(0)(t-\tau)^+ \, m = Z_m(t) + \frac{1}{\sigma(\psi)}\theta\lambda'(0)(t-\tau)^+,$$

$t \in [0, T]$, with $Z_m(t)$ from (2.37). The term after the first $\stackrel{D}{=}$ in (2.38) thus becomes

$$\sup_{0\leq t\leq T} \frac{\left|(1+t)W\left(\frac{t}{1+t}\right) + \theta\lambda'(0)(t-\tau)^+ \, \sigma^{-1}(\psi)\right|}{(1+t)\left(\frac{t}{1+t}\right)^\gamma} \stackrel{D}{=}$$

$$\stackrel{D}{=} \sup_{0\leq x\leq T/(T+1)} \frac{\left|W(x) + \theta\lambda'(0)(x - \tau(1-x))^+ \, \sigma^{-1}(\psi)\right|}{x^\gamma},$$

where making the transformation $t/(t+1) = x$ as is used in Theorem 2.1 leads to

$$\frac{(t-\tau)^+}{(1+t)(t/(t+1))^\gamma} = \frac{(x/(1-x) - \tau)^+}{x^\gamma/(1-x)} = \frac{(x - \tau(1-x))^+}{x^\gamma}, \tag{2.40}$$

giving us the function $p(\cdot, \tau)$ from (2.16) and thus concluding the first part of the proof.

(ii)
Choose k^0 far enough after the change, i.e. with $k^0 - k_m^* = \lfloor m\varepsilon \rfloor$ for some $\varepsilon > 0$. Divide the detector on a part before and after the change. The part before the change is $O_P(1)$ according to Theorem 2.1.
After the change is $\widehat{e}_i = e_i + \theta\delta_m - \widehat{\mu}_m$. Since $\widehat{\mu}_m = O_P(m^{-1/2})$ and $m^{1/2}\delta_m \to \infty$, δ_m dominates. Therefore we can consider only

$$\sum_{i=m+k_m^*+1}^{m+k^0} \psi(e_i + \theta\delta_m)$$

for the part after the change.

Since $q_\gamma(t) = (1+t)\,(t/(1+t))^\gamma \le (1+T)$ for $t \in (0,T]$, we have

$$\frac{1}{q_\gamma(k^0/m)\sqrt{m}}\left|\sum_{i=m+k_m^*+1}^{m+k^0}\psi(e_i+\theta\delta_m)\right| \ge \frac{1}{(T+1)\sqrt{m}}\left|\sum_{i=m+k_m^*+1}^{m+k^0}\psi(e_i+\theta\delta_m)\right| \ge$$

$$\ge \frac{k^0-k_m^*}{\sqrt{m}}\left|\mathrm{E}\,\psi(e_1+\theta\delta_m)\right| - \frac{\sqrt{\lfloor m\varepsilon\rfloor}}{\sqrt{m}}\frac{1}{\sqrt{k^0-k_m^*}}\left|\sum_{i=m+k_m^*+1}^{m+k^0}\psi(e_i+\theta\delta_m) - \mathrm{E}\,\psi(e_i+\theta\delta_m)\right|. \tag{2.41}$$

The last term in (2.41) is $O_P(1)$ analogically to Corollary 2.2 (iii). Further assume WLOG that $\theta > 0,\ \delta_m > 0$. Then

$$0 \le \mathrm{E}\,\psi(e_1+\theta\delta_m) = -\lambda(-\theta\delta_m) = \lambda'(0)\theta\delta_m + o(\delta_m)$$

and thus

$$\frac{\lfloor m\varepsilon\rfloor}{\sqrt{m}}\,\mathrm{E}\,\psi(e_1+\theta\delta_m) \ge C\sqrt{m}\delta_m \to \infty,$$

which finishes the proof of part *(ii)*. □

2.5 Estimation of Long-run Variance

Now we will focus on finding a suitable consistent estimator of the long-run variance $\sigma^2(\psi)$ defined in (2.8).
We consider a Bartlett type estimator

$$\widehat{\sigma}_m^2(\psi) = \widehat{R}_m(0,\psi) + 2\sum_{k=1}^{\Lambda_m} w(k/\Lambda_m)\widehat{R}_m(k,\psi), \tag{2.42}$$

where

$$\widehat{R}_m(k,\psi) = \frac{1}{m}\sum_{i=1}^{m-k}\psi(\widehat{e}_i)\psi(\widehat{e}_{i+k}), \tag{2.43}$$

$$w(t) = (1-|t|)I\{|t|\le 1\}, \quad t\in\mathbb{R}. \tag{2.44}$$

To prove the consistency we need an additional assumption:

(A.5) For some $q > 4$

$$\mathrm{E}\,|\psi(e_i)|^q < \infty, \quad \sum_{j=1}^{\infty}\alpha(j)^{1-4/q} < \infty.$$

Theorem 2.3. *Let $Y_1,\dots,Y_m$ follow model* (2.1) *under the non-contamination assumption* (2.2). *Let assumptions (A.1) – (A.5) be satisfied and let, as m $\to\infty$*

$$\Lambda_m \to \infty, \qquad \Lambda_m m^{-\nu} \to 0,\ \textit{where}\ \nu = \min\left(\frac{1}{3}, \frac{a}{2(2+\Delta)}\right).$$

Then for $\widehat{\sigma}^2_m(\psi)$ *defined in* (2.42) *is*

$$\widehat{\sigma}^2_m(\psi) - \sigma^2(\psi) = o_P(1).$$

Remark 2.1. Instead of the Bartlett kernel (2.44) we can use also other types of kernels w, e.g., a flat top kernel

$$w(t) = \begin{cases} 1 & |t| \leq 1/2 \\ 2(1-|t|) & 1/2 < |t| < 1 \\ 0 & |t| \geq 1, \end{cases}$$

and Theorem 2.3 remains true. More information on suitability of the respective estimators can be found is Section 6.1.1.

Proof of Theorem 2.3. The proof is done in two steps. In the first step we show that we can replace $\psi(\widehat{e}_i)$ by $\psi(e_i)$ in the definition of $\widehat{\sigma}^2_m(\psi)$ i.e. that

$$\widehat{\sigma}^2_m(\psi) - \widetilde{\sigma}^2_m(\psi) = o_P(1), \tag{2.45}$$

where

$$\begin{aligned} \widetilde{\sigma}^2_m(\psi) &= \widetilde{R}(0,\psi) + 2\sum_{k=1}^{\Lambda_m} w(k/m)\widetilde{R}_m(k,\psi), \\ \widetilde{R}_m(k,\psi) &= \frac{1}{m}\sum_{i=1}^{m-k} \psi(e_i)\psi(e_{i+k}). \end{aligned}$$

In the second step we show that

$$\widetilde{\sigma}^2_m(\psi) - \sigma^2(\psi) = o_P(1). \tag{2.46}$$

We start the first step by decomposition

$$\widehat{R}_m(k,\psi) - \widetilde{R}_m(k,\psi) = A_k(\psi) + B_k(\psi) + C_k(\psi),$$

where

$$\begin{aligned} A_k(\psi) &= \frac{1}{m}\sum_{i=1}^{m-k} (\psi(\widehat{e}_i) - \psi(e_i))(\psi(\widehat{e}_{i+k}) - \psi(e_{i+k})), \\ B_k(\psi) &= \frac{1}{m}\sum_{i=1}^{m-k} (\psi(\widehat{e}_i) - \psi(e_i))\psi(e_{i+k}), \\ C_k(\psi) &= \frac{1}{m}\sum_{i=1}^{m-k} (\psi(\widehat{e}_{i+k}) - \psi(e_{i+k}))\psi(e_i). \end{aligned}$$

Towards estimating $A_k(\psi)$ we first consider

$$\begin{aligned} &\mathrm{E}\sup_{|t|\leq K} \frac{1}{m}\left|\sum_{i=1}^{m-k} \Big(\psi(e_i - tm^{-1/2}) - \psi(e_i)\Big)\Big(\psi(e_{i+k} - tm^{-1/2}) - \psi(e_{i+k})\Big)\right| \leq \\ &\leq \mathrm{E}\sup_{|t|\leq K} \frac{1}{m}\sum_{i=1}^{m-k} \Big|\psi(e_i - tm^{-1/2}) - \psi(e_i)\Big|\,\Big|\psi(e_{i+k} - tm^{-1/2}) - \psi(e_{i+k})\Big|. \end{aligned} \tag{2.47}$$

Since $\psi(.)$ is nondecreasing we can estimate RHS of (2.47) by

$$\begin{aligned}
&\mathrm{E}\Big\{\frac{1}{m}\sum_{i=1}^{m-k}\Big(\Big|\psi(e_i-Km^{-1/2})-\psi(e_i)\Big|+\Big|\psi(e_i+Km^{-1/2})-\psi(e_i)\Big|\Big)\cdot\\
&\quad\cdot\Big(\Big|\psi(e_{i+k}-Km^{-1/2})-\psi(e_{i+k})\Big|+\Big|\psi(e_{i+k}+Km^{-1/2})-\psi(e_{i+k})\Big|\Big)\Big\}\\
&\quad =: \mathrm{E}\Big\{\frac{1}{m}\sum_{i=1}^{m-k}\left(|Z_i^-|+|Z_i^+|\right)\left(|Z_{i+k}^-|+|Z_{i+k}^+|\right)\Big\}.
\end{aligned}$$

Now considering for example the term $\mathrm{E}\left(|Z_i^-|\,|Z_{i+k}^-|\right)$ we get by assumption (A.3)

$$\begin{aligned}
\mathrm{E}\left(|Z_i^-|\,|Z_{i+k}^-|\right) \le \|Z_i^-\|_{2+\Delta}\,\|Z_{i+k}^-\|_{2+\Delta} &\le C\left(m^{-a/2(2+\Delta)}K^{a/(2+\Delta)}\right)^2 \le\\
&\le Cm^{-a/(2+\Delta)} \qquad (2.48)
\end{aligned}$$

uniformly in $k=1,\dots,\Lambda_m,\ i=1,\dots,m-k$.

Using the fact that $\sqrt{m}(\widehat{\mu}_m(\psi)-\mu_0)=O_P(1)$ similarly as in Lemma 2.3 we get uniformly in $k=1,\dots,\Lambda_m$ $A_k(\psi)=O_P(m^{-a/(2+\Delta)})$. Thus we can easily conclude that

$$A_0(\psi)+2\sum_{k=1}^{\Lambda_m}w(k/m)A_k(\psi)=O_P(m^{-a/(2+\Delta)}\Lambda_m)=o_P(1).$$

Quite analogously, we get (using now both parts of assumption (A.3)) that

$$\begin{aligned}
B_0(\psi)+2\sum_{k=1}^{\Lambda_m}w(k/m)B_k(\psi) &= O_P(m^{-a/2(2+\Delta)}\Lambda_m)=o_P(1),\\
C_0(\psi)+2\sum_{k=1}^{\Lambda_m}w(k/m)C_k(\psi) &= O_P(m^{-a/2(2+\Delta)}\Lambda_m)=o_P(1).
\end{aligned}$$

Note that the power of m is half compared to $A_k(\psi)$, this is due to fact that only one factor is a difference in the definition of $B_k(\psi)$ and $C_k(\psi)$. Then assertion (2.45) follows directly.

It remains to show (2.46). Since

$$\begin{aligned}
\mathrm{E}\,\widetilde{\sigma}_m^2(\psi) &= \mathrm{E}\Big(\frac{1}{m}\sum_{i=1}^{m}\psi^2(e_i)+2\frac{1}{m}\sum_{k=1}^{\Lambda_m}\Big(1-\frac{k}{\Lambda_m}\Big)\sum_{i=1}^{m-k}\psi(e_i)\psi(e_{i+k})\Big)\\
&= \mathrm{E}\,\psi^2(e_1)+2\sum_{k=1}^{\Lambda_m}\mathrm{E}\,\psi(e_1)\psi(e_{1+k})\left(1-\frac{k}{\Lambda_m}\right)\frac{m-k}{m}\\
&= \sigma^2(\psi)+o(1)
\end{aligned}$$

it suffices to show

$$\widetilde{\sigma}_m^2(\psi)-\mathrm{E}\,\widetilde{\sigma}_m^2(\psi)=o_P(1). \qquad (2.49)$$

Towards this,

$$\widetilde{\sigma}_m^2(\psi) - \mathrm{E}\,\widetilde{\sigma}_m^2(\psi) = \frac{1}{m}\sum_{i=1}^{m}\big(\psi^2(e_i) - E\psi^2(e_i)\big) + \\ + 2\frac{1}{m}\sum_{k=1}^{\Lambda_m}\Big(1 - \frac{k}{\Lambda_m}\Big)\sum_{i=1}^{m-k}\Big(\psi(e_i)\psi(e_{i+k}) - \mathrm{E}\,\psi(e_i)\psi(e_{i+k})\Big). \quad (2.50)$$

By Lemma II with $g_n(e_i) = \psi^2(e_i) - \mathrm{E}\,\psi^2(e_i)$ and $2 + \xi = q/2 > 2$,

$$\mathrm{E}\left(\frac{1}{m}\sum_{i=1}^{m}\big(\psi^2(e_i) - \mathrm{E}\,\psi^2(e_i)\big)\right)^2 \le Dm^{-1}\sum_{j=1}^{\infty}\alpha(j)^{(q-4)/q} \to 0$$

and thus the first term on RHS of (2.50) is $o_P(1)$. For the second part

$$\mathrm{E}\left|\frac{1}{m}\sum_{k=1}^{\Lambda_m}\left(1 - \frac{k}{\Lambda_m}\right)\sum_{i=1}^{m-k}\big(\psi(e_i)\psi(e_{i+k}) - \mathrm{E}\,\psi(e_i)\psi(e_{i+k})\big)\right| \le \\ \le \frac{1}{m}\sum_{k=1}^{\Lambda_m}\left(1 - \frac{k}{\Lambda_m}\right)\left(\mathrm{E}\left(\sum_{i=1}^{m-k}\big(\psi(e_i)\psi(e_{i+k}) - \mathrm{E}\,\psi(e_i)\psi(e_{i+k})\big)\right)^2\right)^{1/2}. \quad (2.51)$$

Due to the stationarity

$$\mathrm{E}\Big(\sum_{i=1}^{m-k}\Big(\psi(e_i)\psi(e_{i+k}) - \mathrm{E}\,\psi(e_i)\psi(e_{i+k})\Big)\Big)^2 =: I_1(k) + 2I_2(k) + 2I_3(k),$$

where

$$\begin{aligned} I_1(k) &= (m-k)\,\mathrm{E}\big(\psi(e_1)\psi(e_{1+k}) - \mathrm{E}\,\psi(e_1)\psi(e_{1+k})\big)^2, \\ I_2(k) &= \sum_{i_1=1}^{m-k}\sum_{\{i_2;0<i_2-i_1\le\Lambda_m\}}\mathrm{E}\Big\{\Big(\psi(e_{i_1})\psi(e_{i_1+k}) - \mathrm{E}\,\psi(e_{i_1})\psi(e_{i_1+k})\Big)\cdot \\ &\quad \Big(\psi(e_{i_2})\psi(e_{i_2+k}) - \mathrm{E}\,\psi(e_{i_2})\psi(e_{i_2+k})\Big)\Big\}, \\ I_3(k) &= \sum_{i_1=1}^{m-k}\sum_{\{i_2;i_2-i_1>\Lambda_m\}}\mathrm{E}\Big\{\Big(\psi(e_{i_1})\psi(e_{i_1+k}) - \mathrm{E}\,\psi(e_{i_1})\psi(e_{i_1+k})\Big)\cdot \\ &\quad \Big(\psi(e_{i_2})\psi(e_{i_2+k}) - \mathrm{E}\,\psi(e_{i_2})\psi(e_{i_2+k})\Big)\Big\}. \end{aligned}$$

Since we assume that $q > 4$ moment of $\psi(e_1)$ is finite, we easily get

$$I_1(k) \le Cm, \quad I_2(k) \le Cm\Lambda_m,$$

uniformly in k. For $I_3(k)$ we apply Lemma I with

$$Z_1 = \psi(e_{i_1})\psi(e_{i_1+k}) - \mathrm{E}\,\psi(e_{i_1})\psi(e_{i_1+k}) \text{ and } Z_2 = \psi(e_{i_2})\psi(e_{i_2+k}) - \mathrm{E}\,\psi(e_{i_2})\psi(e_{i_2+k})$$

and thus get uniformly in k

$$I_3(k) \leq Cm \sum_{j=1}^{\infty} (\alpha(j))^{(q-4)/q}.$$

Therefore RHS of (2.51) is of the same order as

$$\frac{1}{m} \sum_{k=1}^{\Lambda_m} \left(1 - \frac{k}{\Lambda_m}\right) \sqrt{m\Lambda_m} = O\left(\sqrt{\Lambda_m^3/m}\right) \to 0,$$

which finishes the proof of (2.49) and thus (2.46), completing the proof.

□

Chapter 3

Monitoring in Multivariate Location Model

In this chapter we consider a multivariate generalization of the monitoring procedure from the previous chapter. We use again the M-residuals which are now based on multivariate M-estimates. We also show how the multiple comparison can be used to detect which component of the data is responsible for the alarm.

3.1 Model, Assumptions and Test Statistic

We consider sequentially arriving d-dimensional observations following model

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{e}_i,\ i = 1, 2, \dots, \tag{3.1}$$

where $\boldsymbol{Y}_i = (Y_{i1}, \dots, Y_{id})^T$ are observed data, $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{id})^T$ the location parameters and $\boldsymbol{e}_i = (e_{i1}, \dots, e_{id})^T$ strictly stationary random errors forming α-mixing sequence with properties specified bellow. (Vector (matrix) quantities are denoted with bold symbols.)

Both the non-contamination assumption and the hypotheses are the same as in the univariate case, referring now to the multivariate parameters. Thus it is assumed to have stable historical period with

$$\boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_m = \boldsymbol{\mu}_0 \tag{3.2}$$

and we are testing stability of the multivariate location parameter in the monitoring period

$$H_0 : \boldsymbol{\mu}_i = \boldsymbol{\mu}_0, \quad 1 \le i \le m + mT,$$

against the alternative

$$H_1 : \exists k_m^* = k^* < mT \text{ such that } \boldsymbol{\mu}_i = \boldsymbol{\mu}_0,\ 1 \le i \le m + k^*,$$
$$\boldsymbol{\mu}_i = \boldsymbol{\mu}_*,\ m + k^* < i \le m + mT, \quad \boldsymbol{\mu}_0 \ne \boldsymbol{\mu}_*,$$

where $\boldsymbol{\mu}_0, \boldsymbol{\mu}_*$ and k^* are unknown. We consider the closed-end monitoring procedure with $T > 0$ fixed again.

Monitoring scheme is described through the stopping time $\tau_{m,T}$ defined as

$$\tau_{m,T} = \inf\{1 \le k \le mT : |\widehat{Q}_{\boldsymbol{\psi}}(m,k)| > c_{m,T}(\alpha)\, g(k/m)\}, \tag{3.3}$$

where the detector $\widehat{Q}_{\boldsymbol{\psi}}(m,k)$ is a quadratic form of cumulative sums of M-residuals, $g(\cdot)$ is a boundary function described later and $c_{m,T}(\alpha)$ is a critical value, which we need to find. Thus we can follow the decision rule described at the end of Section 1.3, i.e. we reject the null hypothesis as soon as the critical value is crossed.

For the multivariate M-estimate a component-wise version of the definition is used, i.e. the estimate $\widehat{\boldsymbol{\mu}}_m(\boldsymbol{\psi})$ of the location parameter $\boldsymbol{\mu}_0$ is defined as

$$\widehat{\boldsymbol{\mu}}_m(\boldsymbol{\psi}) = \arg\min_{\boldsymbol{t}\in\mathbb{R}^d} \sum_{j=1}^{d} \sum_{i=1}^{m} \rho_j(Y_{ij} - t_j),$$

using possibly different convex loss functions ρ_j and respective a.e. derivatives ψ_j for each component. Similarly as indicated in Section 1.4, finding the minimum usually reduces to solving the following set of equations

$$\sum_{i=1}^{m} \psi_j(Y_{ij} - \mu_j) = 0,\ j = 1, \dots, d,$$

w.r.t. $\mu_j,\ j = 1, \dots, d$.

The M-residual is then defined as

$$\boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i) = (\psi_1(\widehat{e}_{i1}), \dots, \psi_d(\widehat{e}_{id}))^T \text{ with } \widehat{e}_{ij} = Y_{ij} - \widehat{\mu}_{jm}(\psi_j),\ j = 1, \dots, d. \tag{3.4}$$

Finally this leads us to the test statistic

$$\widehat{Q}_{\boldsymbol{\psi}}(m,k) = \left(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)\right)^T \widehat{\boldsymbol{\Sigma}}_m^{-1} \left(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)\right), \tag{3.5}$$

where $\widehat{\boldsymbol{\Sigma}}_m$ is an estimator of the asymptotic (long-run) variance matrix

$$\boldsymbol{\Sigma} = \lim_{m\to\infty} \operatorname{var}\left(\frac{1}{\sqrt{m}} \sum_{i=1}^{m} \boldsymbol{\psi}(\boldsymbol{e}_i)\right) = \sum_{i\in\mathbb{Z}} \mathrm{E}\, \boldsymbol{\psi}(\boldsymbol{e}_0)\boldsymbol{\psi}(\boldsymbol{e}_i)^T \tag{3.6}$$

based on the training data. Details will be discussed later. Dependence on the score function $\boldsymbol{\psi}$ is not indicated but this should not bring any ambiguity.

We again make use of class of functions from (2.9), i.e.

$$q_\gamma(t) = (1+t)\big(t/(1+t)\big)^\gamma,\ \gamma \in [0, 1/2)$$

and define the boundary function $g(\cdot)$ as a square of it.

Assumptions
The assumptions are analogous to those considered in Chapter 2, therefore they are denoted by *.

(A*.1) $\{\boldsymbol{e}_i\}_i$ is a strictly stationary α-mixing sequence with coefficients $\{\alpha(i)\}_i$ such that

$$\sum_{k=0}^{\infty} \alpha(k)^{\Delta/(2+\Delta)} < \infty \text{ for some } \Delta > 0. \tag{3.7}$$

Next two assumptions are considered coordinate-wise, for $j = 1, \ldots, d$. The score function ψ_j, the distribution function F_j and the function $\lambda_j(t) := -\int \psi_j(e_j - t) dF_j(e_j)$ are assumed to satisfy:

(A*.2) ψ_j is non-decreasing; the derivative $\lambda_j'(\cdot)$ of the function $\lambda_j(\cdot)$ exists and is Lipschitz in a neighborhood of zero, $\lambda_j(0) = 0$ and $\lambda_j'(0) > 0$.

(A*.3) $\int |\psi_j(x)|^{2+\Delta} dF_j(x) < \infty$ and

$$\int |\psi_j(x + t_2) - \psi_j(x + t_1)|^{2+\Delta} dF_j(x) \leq D_1 |t_2 - t_1|^a, \quad |t_l| \leq D_2,\, l = 1, 2 \tag{3.8}$$

for $1 \leq a \leq 2 + \Delta$ and some positive constants D_1, D_2 depending on Δ.

(A*.4) The long run variance matrix $\boldsymbol{\Sigma}$ defined in (3.6) is positive definite and finite.

A nice property of α-mixing is that the assumption (A*.1) gives us that $\{\psi_j(e_{ij})\}_i$ is also α-mixing with the same rate for every $j = 1, \ldots, d$.

3.2 Main Results

Now we formulate assertions on limit behavior of the test statistic both under the null and alternative hypotheses. Proofs are deferred to the next section.

Theorem 3.1. *Assume that $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ follow the model* (3.1), *assumptions (A*.1) – (A*.4) are satisfied and test statistic is defined in* (3.5). *Further let $\widehat{\boldsymbol{\Sigma}}_m$ be a consistent estimate of $\boldsymbol{\Sigma}$.*
Then under the null hypothesis H_0 it holds

$$\max_{1 \leq k \leq mT} \frac{\widehat{Q}_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \xrightarrow{\mathcal{D}} \sup_{0 \leq t \leq T/(T+1)} \frac{\sum_{j=1}^d W_j^2(t)}{t^{2\gamma}},$$

where $\{W_j(t),\, t \in [0,1]\}, j = 1, \ldots, d,$ are independent Wiener processes.

The situation under local alternatives is described in the next theorem.

Theorem 3.2. *Assume that $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ follow the model* (3.1), *assumptions (A*.1) – (A*.4) are satisfied and test statistic is defined in* (3.5). *Further let $\widehat{\boldsymbol{\Sigma}}_m$ be a consistent estimate of $\boldsymbol{\Sigma}$.*
Finally assume that for the alternative hypothesis H_1 holds $\boldsymbol{\mu}_* = \boldsymbol{\mu}_0 + \delta_m \boldsymbol{\theta},\ \boldsymbol{\theta} \neq \boldsymbol{0}$ and

$k_m^* = \lfloor m\tau \rfloor$, $0 < \tau < T$.
(i) If $\delta_m = m^{-1/2}$, then

$$\max_{1\le k\le mT} \frac{\widehat{Q}_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \xrightarrow{\mathcal{D}} \sup_{0\le t\le T/(T+1)} \frac{\sum_{j=1}^d \big(W_j(t)+h_j(t,\tau)\big)^2}{t^{2\gamma}},$$

where $\boldsymbol{h}(t,s) = (t - s(1-t))^+ \boldsymbol{\Sigma}^{-1/2}\big(\lambda_1'(0)\theta_1,\ldots,\lambda_d'(0)\theta_d\big)^T$, $0 \le t \le T/(T+1)$, $0 \le s < T$ and $\boldsymbol{\Sigma}^{1/2}$ is square root matrix of $\boldsymbol{\Sigma}$.

(ii) If $\delta_m \to 0$ and $|\delta_m| m^{1/2} \to \infty$, then

$$\max_{1\le k\le mT} \frac{\widehat{Q}_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \xrightarrow{P} \infty.$$

Remark 3.1. (i) Theorem 3.1 provides a way to approximate the critical value so that the test procedure fulfills (1.6) under H_0. Critical value $c_{m,T}(\alpha)$ is approximated by c such that

$$P\Bigg(\sup_{0\le t\le T/(T+1)} \frac{\sum_{j=1}^d W_j^2(t)}{t^{2\gamma}} > c\Bigg) = \alpha.$$

For more details, together with a table of these critical values, please refer to Chapter 7.
(ii) Theorem 3.2 (i) deals with the contiguous alternatives. As expected the asymptotic distribution a maximum of weighted sum of squares of shifted Wiener processes, where the shifts depend on the change-point, the amount of change and also on the choice of the loss functions $\rho_1,\ldots,\rho_d$ (through $\lambda_1'(0),\ldots,\lambda_d'(0)$). A time dependent part of the shift is the same as in the univariate case, i.e. $p(t,\tau)$ from (2.16).
(iii) Theorem 3.2 (ii) implies the consistency of the test, i.e., the validity of (1.7) (asymptotic power of test procedure is 1).

Now we focus our attention on estimation of the long run variance matrix $\boldsymbol{\Sigma}$ defined in (3.6). We use the Bartlett type estimator

$$\widehat{\boldsymbol{\Sigma}}_m = \sum_{k=-\Lambda_m}^{\Lambda_m} w(k/\Lambda_m)\widehat{\boldsymbol{R}}_m(k), \tag{3.9}$$

where $\widehat{\boldsymbol{R}}_m(k)$ is the k-th lag sample covariance

$$\widehat{\boldsymbol{R}}_m(k) = \begin{cases} \frac{1}{m}\sum_{i=1}^{m-k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)\boldsymbol{\psi}(\widehat{\boldsymbol{e}}_{i+k})^T, & k \ge 0, \\ \widehat{\boldsymbol{R}}_m(-k)^T, & k < 0, \end{cases} \tag{3.10}$$

and $w(t) = (1-|t|)I\{|t|\le 1\}$, $\quad t \in \mathbb{R}$ is again the Bartlett kernel.

Theorem 3.3. *Assume that $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots$ follow the model* (3.1) *with non-contamination assumption* (3.2), *assumptions (A*.1) – (A*.4) are satisfied and $\widehat{\boldsymbol{\Sigma}}_m$ is defined in* (3.9). *Let further for some $q > 4$*

$$E|\psi_j(e_{ij})|^q < \infty,\ j = 1,\ldots,d, \quad \sum_{k=1}^\infty \alpha(k)^{1-4/q} < \infty$$

and as $m \to \infty$

$$\Lambda_m \to \infty, \qquad \Lambda_m m^{-\nu} \to 0 \textit{ where } \nu = \min\left(\frac{1}{3}, \frac{a}{2(2+\Delta)}\right). \tag{3.11}$$

Then $\widehat{\mathbf{\Sigma}}_m - \mathbf{\Sigma} = o_P(1), \quad m \to \infty.$

As can be seen from the proof, the theorem remains valid also for the flat top kernel instead of the Bartlett one.

3.3 Proofs

Proof of Theorem 3.1. We make use of the facts that have been already shown in the univariate case. Similarly as in Lemma 2.6 we have for $j = 1, \dots, d$

$$\max_{1 \le k \le mT} \frac{1}{m^{1/2} q_\gamma(k/m)} \left| \sum_{i=m+1}^{m+k} \psi_j(\widehat{e}_{ij}) - \left(\sum_{i=m+1}^{m+k} \psi_j(e_{ij}) - \frac{k}{m} \sum_{l=1}^{m} \psi_j(e_{lj}) \right) \right| = o_P(1).$$

Denoting

$$\widehat{\boldsymbol{S}}(k,m) := \frac{1}{m^{1/2}} \sum_{i=m+1}^{m+k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)$$

and

$$\widetilde{\boldsymbol{S}}(k,m) := \frac{1}{m^{1/2}} \left(\sum_{i=m+1}^{m+k} \boldsymbol{\psi}(\boldsymbol{e}_i) - \frac{k}{m} \sum_{l=1}^{m} \boldsymbol{\psi}(\boldsymbol{e}_l) \right), \tag{3.12}$$

we have that the limit behavior of

$$\max_{1 \le k \le mT} \widehat{Q}_{\boldsymbol{\psi}}(m,k)/q_\gamma^2(k/m)$$

is the same as that of

$$\max_{1 \le k \le mT} \widetilde{Q}_{\boldsymbol{\psi}}(m,k)/q_\gamma^2(k/m),$$

where

$$\widetilde{Q}_{\boldsymbol{\psi}}(m,k) = \widetilde{\boldsymbol{S}}(k,m)^T \mathbf{\Sigma}^{-1} \widetilde{\boldsymbol{S}}(k,m),$$

since

$$\max_{1 \le k \le mT} \frac{\left| \widehat{Q}_{\boldsymbol{\psi}}(m,k) - \widetilde{Q}_{\boldsymbol{\psi}}(m,k) \right|}{q_\gamma^2(k/m)} \le \max_{1 \le k \le mT} \frac{\left| \widehat{\boldsymbol{S}}(k,m)^T (\widehat{\mathbf{\Sigma}}_m^{-1} - \mathbf{\Sigma}^{-1}) \widehat{\boldsymbol{S}}(k,m) \right|}{q_\gamma^2(k/m)} +$$

$$+ \max_{1 \le k \le mT} \frac{\left| \widehat{\boldsymbol{S}}(k,m)^T \mathbf{\Sigma}^{-1} (\widehat{\boldsymbol{S}}(k,m) - \widetilde{\boldsymbol{S}}(k,m)) \right| + \left| (\widehat{\boldsymbol{S}}(k,m) - \widetilde{\boldsymbol{S}}(k,m))^T \mathbf{\Sigma}^{-1} \widetilde{\boldsymbol{S}}(k,m)) \right|}{q_\gamma^2(k/m)} =$$

$$= o_p(1).$$

Now we turn to the limit behavior of the partial sum process

$$\boldsymbol{Z}_m(t) = \frac{1}{\sqrt{m}} \sum_{i=1}^{\lfloor mt \rfloor} \boldsymbol{\psi}(\boldsymbol{e}_i), \quad 0 \le t \le T+1.$$

We show that

$$\boldsymbol{Z}_m(\cdot) \overset{\mathcal{D}^d[0,T+1]}{\longrightarrow} \boldsymbol{W}_{\boldsymbol{\Sigma}}(\cdot), \tag{3.13}$$

where $\{\boldsymbol{W}_{\boldsymbol{\Sigma}}(t),\, 0 \le t \le T+1\}$ is a centered Gaussian process with covariance function

$$\mathrm{E}[\boldsymbol{W}_{\boldsymbol{\Sigma}}(t)\boldsymbol{W}_{\boldsymbol{\Sigma}}^T(s)] = \min(t,s)\boldsymbol{\Sigma}$$

and $\overset{\mathcal{D}^d[0,T+1]}{\longrightarrow}$ denotes weak convergence in the Skorokhod space $D^d[0,T+1]$ (see Appendix for further details).

Toward this we use Lemma III to transform the multivariate problem to the univariate one. The Lemma states that in order to show (3.13) it is enough to show that, for any set of constants $\boldsymbol{c} = (c_1, \dots, c_d)^T$, we have

$$\boldsymbol{c}^T\boldsymbol{Z}_m(\cdot) \overset{\mathcal{D}[0,T+1]}{\longrightarrow} \boldsymbol{c}^T\boldsymbol{W}_{\boldsymbol{\Sigma}}(\cdot).$$

For LHS we have

$$\boldsymbol{c}^T\boldsymbol{Z}_m(t) = \sum_{j=1}^{d} c_j \frac{1}{m^{1/2}} \sum_{i=1}^{\lfloor mt \rfloor} \psi_j(e_{ij}) = \frac{1}{m^{1/2}} \sum_{i=1}^{\lfloor mt \rfloor}\sum_{j=1}^{d} c_j\psi_j(e_{ij}).$$

Since $\{\boldsymbol{e}_i\}_i$ is strong mixing then $\{\sum_{j=1}^d c_j\psi_j(e_{ij})\}_i$ is also strong mixing (with the same rate) and we can thus use the WIP for strong mixing sequences - Theorem 1.1. By definition of $\boldsymbol{\Sigma}$

$$\lim_{m\to\infty} \mathrm{var}(\boldsymbol{c}^T\boldsymbol{Z}_m(1)) = \boldsymbol{c}^T\boldsymbol{\Sigma}\boldsymbol{c},$$

thus we can conclude that (3.13) holds true.

In the next step, we study the process $\{\boldsymbol{H}_m(t), 0 \le t \le T\}$ defined as

$$\boldsymbol{H}_m(t) = \boldsymbol{Z}_m(t+1) - \boldsymbol{Z}_m(1) - t\boldsymbol{Z}_m(1) = \boldsymbol{Z}_m(t+1) - (t+1)\boldsymbol{Z}_m(1), \tag{3.14}$$

which is closely related to (3.12).

Since $\boldsymbol{\Sigma}$ is symmetric positive definite matrix, there exists the so-called *square root matrix* $\boldsymbol{C}$ such that $\boldsymbol{\Sigma} = \boldsymbol{C}\boldsymbol{C}^T$. This matrix is therefore regular and will be denoted also $\boldsymbol{\Sigma}^{1/2}$. Via the continuous mapping theorem,

$$\boldsymbol{C}^{-1}\boldsymbol{H}_m(\cdot) \overset{\mathcal{D}^d[0,T]}{\longrightarrow} \widetilde{\boldsymbol{W}}(\cdot),$$

where $\{\widetilde{\boldsymbol{W}}(t), 0 \le t \le T\}$ is a centered Gaussian process with covariance function $E[\widetilde{\boldsymbol{W}}(t)\widetilde{\boldsymbol{W}}^T(s)] = (t+1)s \cdot \boldsymbol{I}_d$, for $0 \le s \le t \le T$, with $\boldsymbol{I}_d$ denoting the d-dimensional unity matrix. Thus, via another application of the continuous mapping theorem,

$$\boldsymbol{C}^{-1}\boldsymbol{H}_m(\cdot)/(\,\cdot+1) \overset{\mathcal{D}^d[0,T]}{\longrightarrow} \widetilde{\boldsymbol{W}}(\cdot)/(\,\cdot+1) = \boldsymbol{W}^*(\cdot),$$

for which by computing covariance function is easy to checked that

$$\{\boldsymbol{W}^*(t), 0 \le t \le T\} \overset{\mathcal{D}}{=} \{\boldsymbol{W}\big(t/(t+1)\big), 0 \le t \le T\}, \tag{3.15}$$

with $\{\boldsymbol{W}(t), t \geq 0\}$ denoting a standard d-dimensional Brownian motion (i.e. having independent components).

To complete the proof we recall that $q_\gamma(t) = (1+t)(t/(t+1))^\gamma$ and thus in view of the law of iterated logarithm for a Brownian motion (see Csörgő and Horváth [1993])

$$\boldsymbol{W}\Big(\frac{t}{t+1}\Big)\Big/\Big(\frac{t}{t+1}\Big)^\delta \;\rightarrow\; \mathbf{0} \quad \text{a.s.} \quad \text{as} \quad t \rightarrow 0+,$$

for every $0 \leq \delta < 1/2$. Since $0 \leq \gamma < 1/2$, by yet another application of continuous mapping theorem

$$\boldsymbol{C}^{-1}\boldsymbol{H}_m(\cdot)/q_\gamma(\cdot) \;\overset{\mathcal{D}^d[0,T]}{\longrightarrow}\; \boldsymbol{W}\left(\frac{\cdot}{\cdot+1}\right)\Big/\left(\frac{\cdot}{\cdot+1}\right)^\gamma.$$

Finally realizing that

$$\sup_{0\leq t\leq T} |\lfloor mt\rfloor/m - t| \rightarrow\; 0 \quad \text{and}$$

$$\sup_{1/m\leq t\leq T} \Big|\frac{q_\gamma(t)}{q_\gamma(\lfloor mt\rfloor/m)} - 1\Big| \rightarrow\; 0,$$

we get that

$$\boldsymbol{C}^{-1}\boldsymbol{H}_m(\lfloor m\cdot\rfloor/m)/q_\gamma(\lfloor m\cdot\rfloor/m) \;\overset{\mathcal{D}^d[0,T]}{\longrightarrow}\; \boldsymbol{W}\left(\frac{\cdot}{\cdot+1}\right)\Big/\left(\frac{\cdot}{\cdot+1}\right)^\gamma, \tag{3.16}$$

where the argument of the process runs from 0 to T. The asymptotic distribution of supremum of quadratic form is now clear. □

Proof of Theorem 3.2. (*i*) We again use the facts already known from the univariate case and keep the notation from the previous proof. Thus similarly as in (2.39) we have for $j = 1, \ldots, d$

$$\max_{1\leq k\leq mT} \frac{\Big|\sum_{i=m+1}^{m+k}\psi_j(\widehat{e}_{ij}) - \Big(\sum_{i=m+1}^{m+k}\psi_j(e_{ij}) - \frac{k}{m}\sum_{l=1}^{m}\psi_j(e_{lj}) + \delta_m\theta_j\lambda_j'(0)(k-k_m^*)^+\Big)\Big|}{m^{1/2}q_\gamma(k/m)} = o_P(1)$$

Defining

$$\boldsymbol{H}_m^*(t) = \boldsymbol{H}_m(t) + m^{1/2}\delta_m(t-\tau)^+(\theta_1\lambda_1'(0), \ldots, \theta_d\lambda_d'(0))^T$$

with $\boldsymbol{H}_m(t)$ from (3.14), we get similarly as in (3.16)

$$\frac{\boldsymbol{C}^{-1}\boldsymbol{H}_m^*(\lfloor m\cdot\rfloor/m)}{q_\gamma(\lfloor m\cdot\rfloor/m)} \;\overset{\mathcal{D}^d[0,T]}{\longrightarrow}\; \frac{\boldsymbol{W}\big(\cdot/(\cdot+1)\big)}{\big(\cdot/(\cdot+1)\big)^\gamma} + \boldsymbol{C}^{-1}(\theta_1\lambda_1'(0), \ldots, \theta_d\lambda_d'(0))^T\frac{(\cdot-\tau)^+}{q_\gamma(\cdot)}. \tag{3.17}$$

The term $\frac{(t-\tau)^+}{q_\gamma(t)}$ was already treated in (2.40) and is thus give us the $p(t,\tau)$ part of the function $\boldsymbol{h}(t,\tau)$. Therefore the proof of part (*i*) is finished.

(*ii*) There exists at least one coordinate l such that $\theta_l \neq 0$ and according to Theorem 2.2 (*ii*) we know that

$$\max_{1\leq k\leq mT} \frac{\big|\sum_{i=m+1}^{m+k}\psi_l(\widehat{e}_{il})\big|}{\sqrt{m}\sqrt{\widehat{\sigma}_{ll}}q_\gamma(k/m)} \overset{P}{\rightarrow} \infty,$$

where $\widehat{\boldsymbol{\Sigma}}_m = (\widehat{\sigma}_{kl})_{kl}$. Thus the same divergence in probability is true for our detector. □

Proof of Theorem 3.3. The idea of the proof is the same as in the univariate case. It consists of 2 steps.

Firstly one shows that $\widehat{\boldsymbol{\Sigma}}_m - \widetilde{\boldsymbol{\Sigma}}_m = o_P(1)$, where

$$\widetilde{\boldsymbol{\Sigma}}_m = \sum_{k=-\Lambda_m}^{\Lambda_m} w(k/\Lambda_m)\widetilde{\boldsymbol{R}}_m(k)$$

with

$$\widetilde{\boldsymbol{R}}_m(k) = \frac{1}{m}\sum_{i=1}^{m-k} \boldsymbol{\psi}(\boldsymbol{e}_i)\boldsymbol{\psi}(\boldsymbol{e}_{i+k})^T, k \geq 0 \quad \text{and} \quad \widetilde{\boldsymbol{R}}_m(k) = \widetilde{\boldsymbol{R}}_m(-k)^T, k < 0.$$

Then it is shown that $\widetilde{\boldsymbol{\Sigma}}_m - \boldsymbol{\Sigma} = o_P(1)$.

We consider the previously mentioned matrices $\boldsymbol{\Sigma}$, $\widehat{\boldsymbol{\Sigma}}_m$, $\widetilde{\boldsymbol{\Sigma}}_m$ element-wise. For the (j,l) element of the matrix $\boldsymbol{\Sigma}$ we have

$$\sigma_{jl} = \sum_{i\in\mathbb{Z}} \mathrm{E}\,\psi_j(e_{ij})\psi_l(e_{il}) = \mathrm{E}\,\psi_j(e_{1j})\psi_l(e_{1l}) + 2\sum_{i=2}^{\infty} \mathrm{E}\,\psi_j(e_{1j})\psi_l(e_{il})$$

and the key parts of (j,l) elements of matrices $\widehat{\boldsymbol{\Sigma}}_m$, $\widetilde{\boldsymbol{\Sigma}}_m$ are

$$\widehat{r}_{jl}(k) = \frac{1}{m}\sum_{i=1}^{m-k} \psi_j(\widehat{e}_{ij})\psi_l(\widehat{e}_{i+k,\,l})$$

and $\widetilde{r}_{jl}(k)$ which is defined accordingly.

Realizing that $\{\psi_j(e_{ij})\psi_l(e_{il})\}_i,\ 1 \leq j,l \leq d$ is strong mixing sequence we can follow the proof of Theorem 2.3 line by line to show both above mentioned assertions. $\square$

3.4 Multiple Comparison

In case when the procedure signals a change it is of interest to find out which of the components has changed. Towards this we can use a variant of the well known Scheffé method used in multiple comparison of ANOVA. We first recall the underlying theorem, see Anděl [1985] (p.147, Lemma 1) for example.

Theorem 3.4 (Scheffé)**.** *Let $\boldsymbol{A}$ be a $d\times d$ positive definite matrix. Then for every $c > 0$*

$$[\boldsymbol{x}^T\boldsymbol{A}^{-1}\boldsymbol{x} \leq c] \Leftrightarrow [(\boldsymbol{h}^T\boldsymbol{x})^2 \leq c\,\boldsymbol{h}^T\boldsymbol{A}\boldsymbol{h} \text{ for all } \boldsymbol{h} \in \mathbb{R}^d]. \tag{3.18}$$

Now recall that our test detector has a form

$$\sup_{1\leq k\leq mT} \boldsymbol{U}(m,k)^T\widehat{\boldsymbol{\Sigma}}_m^{-1}\boldsymbol{U}(m,k),$$

where we denoted

$$\boldsymbol{U}(m,k) := \frac{1}{\sqrt{m}q_\gamma(k/m)} \sum_{i=m+1}^{m+k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i).$$

Moreover the procedure signals a change as soon as for some $k = 1, \ldots, mT$

$$\boldsymbol{U}(m,k)^T \widehat{\boldsymbol{\Sigma}}_m^{-1} \boldsymbol{U}(m,k) > c_T^{(d)}(\alpha, \gamma),$$

where $c_T^{(d)}(\alpha,\gamma)$ is a the critical value for level α and d-dimensional data introduced in Section 7.1. Under the null hypothesis this happens with probability α asymptotically and by Theorem 3.4 it is equivalent to the fact that there exists $\boldsymbol{h} \in \mathbb{R}^d$ such that

$$(\boldsymbol{h}^T \boldsymbol{U}(m,k))^2 > c_T^{(d)}(\alpha,\gamma) \boldsymbol{h}^T \widehat{\boldsymbol{\Sigma}}_m \boldsymbol{h}. \tag{3.19}$$

When we choose $\boldsymbol{h}$ as a canonical vector $\boldsymbol{h}_j$ having one in j-th coordinate and zeros otherwise, equation (3.19) transforms to

$$\frac{1}{\sqrt{m}q_\gamma(k/m)\sqrt{\widehat{\sigma}_{jj}}} \left| \sum_{i=m+1}^{m+k} \psi_j(\widehat{e}_{ij}) \right| > \sqrt{c_T^{(d)}(\alpha,\gamma)}, \tag{3.20}$$

where the elements of $\widehat{\boldsymbol{\Sigma}}_m$ are denoted again as $\widehat{\sigma}_{jk}$. Note that LHS is exactly the same as the test detector from the univariate case (since $\widehat{\sigma}_m^2(\psi)$ corresponds to $\widehat{\sigma}_{jj}$), only the critical value differs. Moreover (3.20) happens for any coordinate $j = 1, \ldots, d$ with probability smaller than α asymptotically and thus we can use it to determine which component has caused the alarm.

It is also of interest to compare the obtained critical values for multiple comparison with those coming from simple Bonferroni correction. This is done in Table 3.1 for $d = 5$ dimensions. Since the Bonferroni method does not cover so many alternatives, it gives even sharper results.

	$\backslash\gamma$	0	0.15	0.25	0.45	0.49
Bonferroni	$c_\infty(0.01,\gamma)$	2.79	2.85	2.94	3.30	3.57
Scheffé	$\sqrt{c_\infty^{(5)}(0.05,\gamma)}$	3.53	3.59	3.65	4.02	4.36

Table 3.1: Critical values for multiple comparison with Bonferroni and Scheffé methods, $d = 5,\ \alpha = 5\%$.

Chapter 4

Monitoring in Capital Asset Pricing Model

In this chapter we consider robust sequential monitoring in a situation of Capital Asset Pricing Model (CAPM), which is one of the famous models used in econometrics. In fact, this procedure is an elaborate extension of the one considered in the previous chapter to a multivariate simple regression model. Naturally analogous robust sequential monitoring procedures can be designed for general linear regression models, the proofs become however more technical. Thus we restrict ourselves to this special case, which is due to its widespread application of interest on its own. As the main interest of CAPM lies in the so-called *portfolio betas*, we test for a change in the slope parameter only.

This robust monitoring was already considered in Chochola et al. [2013], however under different assumptions on the dependence structure of observations. Since we want to have unified approach to modeling dependence structures in the thesis, we consider again the α-mixing as opposed to $L_p{-}m$-approximability concept used in Chochola et al. [2013]. The main obtained results are naturally the same, however the technique of the proofs differs.

CAPM, introduced by Sharpe [1964] and subsequently modified by many authors (see, e.g. Lintner [1965] and Merton [1973]), is an important and widely used model for evaluating the risk of a portfolio of assets with respect to the market risk. Despite of some shortcomings pointed out by theoreticians and practitioners as well, the wide-spread use of the CAPM is also well-documented (cf., e.g., the report of Martin and Simin [2003]). A main advantage of the model is its simplicity in describing the sensitivity of an asset's risk against the market risk, which is essentially expressed through one parameter, the portfolio beta.

On the other hand, it is also well-known that the corresponding pricing of a portfolio asset heavily relies on the constancy of the betas over time. Confer, for example, the discussion in Ghysels [1998] and recently Caporale [2012]. Therefore, it may be of great interest to find out whether portfolio betas change significantly over time or not. This was a main motivation in Aue et al. [2012] for constructing a sequential monitoring procedure for testing the stability of portfolio betas, taking high-frequency nature of data also into account. Along the lines of Chu et al. [1996], the corresponding stopping rules of Aue et al. [2012] are based on comparing the (ordinary) least squares estimate (OLS) of the beta from a historical data set (training period) to that from sequentially incoming new observations. A structural break (change) in the model is then confirmed when the beta significantly changes, that is, when the newly estimated beta exceeds a critical distance from the historical one.

However, it is well-known that OLS estimators are sensitive with respect to outliers and deviations from normality assumptions. Concerning the possible application of the CAPM this has led to an extensive discussion and numerous suggestions for "robustifying" the use of beta estimates in the prediction of portfolio risks (confer, e.g., Genton and Ronchetti [2008] and Martin and Simin [2003] together with the works mentioned therein).

Indeed, this robustification lead us to the idea of applying procedure similar to the ones considered in previous chapters for testing the stability of CAPM portfolio betas. Thus we use the M-estimates (and M-residuals) in order to reduce the sensitivity against outliers and non-normality assumptions. Moreover we suggest a multivariate approach allowing for dependencies within the portfolio and also for possible dependencies over time.

4.1 Model, Assumptions and Test Statistic

In the sequel our statistical framework will be as follows. We consider the model

$$\boldsymbol{r}_i = \boldsymbol{\alpha}_i + \boldsymbol{\beta}_i r_{iM} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, \tag{4.1}$$

where $\boldsymbol{r}_i = (r_{i,1}, \dots, r_{i,d})^T$ is a d-dimensional vector of daily log-returns at time i, r_{iM} is the log-return of the market portfolio at time i, and $\boldsymbol{\varepsilon}_i = (\varepsilon_{i,1}, \dots, \varepsilon_{i,d})^T$ is a d-dimensional error term. The $\boldsymbol{\alpha}_i$'s and $\boldsymbol{\beta}_i$'s are d-dimensional unknown parameters, and the $\boldsymbol{\beta}_i$'s are the parameters of interest, the "portfolio betas".

We assume the non-contamination condition. i.e. that a training sample of size m with no instabilities is available

$$\boldsymbol{\alpha}_1 = \dots = \boldsymbol{\alpha}_m =: \boldsymbol{\alpha}^0, \quad \boldsymbol{\beta}_1 = \dots = \boldsymbol{\beta}_m =: \boldsymbol{\beta}^0, \tag{4.2}$$

where $\boldsymbol{\alpha}^0$ and $\boldsymbol{\beta}^0$ are unknown parameters. The problem of the instability of the portfolio betas is formulated as a testing problem, that is, we want to test the null hypothesis

$$H_0: \ \boldsymbol{\beta}_1 = \dots = \boldsymbol{\beta}_m = \boldsymbol{\beta}_{m+1} = \dots$$

of no change versus the alternative

$$H_1: \ \boldsymbol{\beta}_1 = \dots, = \boldsymbol{\beta}_{m+k^*} \neq \boldsymbol{\beta}_{m+k^*+1} = \dots$$

of a structural break at an unknown change-point $k^* = k^*_m$ in the monitoring period.

For later convenience we reformulate our model as follows:

$$r_{i,j} = \alpha_j^0 + \beta_j^0 \widetilde{r}_{iM} + (\alpha_j^1 + \beta_j^1 \widetilde{r}_{iM})\delta_m I\{i > m + k^*\} + \varepsilon_{i,j}, \quad j = 1, \dots, d, \quad i = 1, 2, \dots, \tag{4.3}$$

where $k^* = k_m^*$ is the change-point, $\alpha_j^0, \beta_j^0, \alpha_j^1, \beta_j^1, \delta_m$ are unknown parameters, and

$$\widetilde{r}_{iM} = r_{iM} - \overline{r}_{mM}, \quad \text{with} \quad \overline{r}_{mM} = \frac{1}{m}\sum_{i=1}^{m} r_{iM}. \tag{4.4}$$

The M-estimates are similarly as in Section 3.1 generated by convex loss functions $\rho_1, \dots, \rho_d$ with a.e. derivatives $\rho_j' = \psi_j$ (score functions) having further properties to be specified later. We only have to take the regressors $\widetilde{r}_{iM}$ into account. Thus the estimators $\widehat{\alpha}_{jm} = \widehat{\alpha}_{jm}(\psi_j), \widehat{\beta}_{jm} = \widehat{\beta}_{jm}(\psi_j)$ of α_j^0, β_j^0 based on the training sample are defined as minimizers of

$$\sum_{i=1}^{m} \rho_j(r_{i,j} - a_j - b_j \widetilde{r}_{iM}) \tag{4.5}$$

w.r.t. a_j, b_j for $j = 1, \dots, d$.

The M-residuals are then defined as

$$\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i) = (\psi_1(\widehat{\varepsilon}_{i,1}), \dots, \psi_d(\widehat{\varepsilon}_{i,d}))^T \tag{4.6}$$

with

$$\begin{aligned} \widehat{\boldsymbol{\varepsilon}}_i &= (\widehat{\varepsilon}_{i,1}, \dots, \widehat{\varepsilon}_{i,d})^T, \\ \widehat{\varepsilon}_{i,j} &= r_{i,j} - \widehat{\alpha}_{jm} - \widetilde{r}_{iM}\widehat{\beta}_{jm}. \end{aligned} \tag{4.7}$$

A test statistic based on the first $m + k$ observations is a quadratic form of CUSUM of weighted M-residuals

$$\widehat{Q}_{\boldsymbol{\psi}}^C(k, m) = \Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\Big)^T \widehat{\boldsymbol{\Sigma}}_m^{-1} \Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM} \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\Big), \tag{4.8}$$

where the matrix $\widehat{\boldsymbol{\Sigma}}_m$ is an estimator of the long-run variance matrix

$$\boldsymbol{\Sigma} = \lim_{m\to\infty} \operatorname{var}\Big\{\frac{1}{\sqrt{m}} \sum_{i=1}^{m} (r_{iM} - \mathrm{E}\, r_{iM}) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\Big\} \tag{4.9}$$

based on the first m observations. Details will be discussed later.

We reject the null hypothesis as soon as the detector exceeds a critical level for the first time, i.e., when

$$\widehat{Q}_{\boldsymbol{\psi}}^C(m, k)/q_\gamma^2(k/m) > c$$

for an appropriately chosen critical value c, where $q_\gamma(\cdot)$ is a boundary (weight) function defined in (2.9). In this case we stop the procedure and confirm a structural break, otherwise we continue monitoring at most up to time mT, i.e. we design again the closed-end procedure.

The associated stopping rule is thus given by (3.3) with the test statistic (4.8). We only need to choose the critical value c such that (1.6) and (1.7) hold true, i.e. the test has asymptotic level α and is asymptotically consistent. Towards this we use the asymptotic behavior of the test statistic which is derived in the next section.

Assumptions
The assumptions on ψ_j, distribution function F_j of $\varepsilon_{i,j}$ and the derived function λ_j, $j = 1, \dots, d$ are the same as in Chapter 3, i.e. (A*.1)– (A*.4), where in (A*.4) we now consider the long run variance matrix $\mathbf{\Sigma}$ defined in (4.9).
Moreover we introduce one assumption on $\{r_{iM}\}_i$

(B) $\{r_{iM}\}_i$ is a strictly stationary α-mixing sequence independent of $\{\boldsymbol{\varepsilon}_i\}_i$ with the same mixing rate, i.e. $\sum_{k=0}^{\infty} \alpha_M(k)^{\Delta/(2+\Delta)} < \infty$, where Δ is from assumption (A*.1) and $\{\alpha_M(i)\}_i$ are mixing coefficients of $\{r_{iM}\}_i$. Moreover $\mathrm{E}\,|r_{iM}|^{2+\Delta} < \infty$.

Remark 4.1. a) In formula (3.8) of (A*.3) the upper bound is $|t_2 - t_1|^a$, for $1 \le a \le 2+\Delta$ and $|t_l| \le D_2$, $l = 1, 2$. Thus $|t_2 - t_1|$ is small and in proofs the least favorable situation is for $a = 1$. Therefore it suffices to consider $a = 1$ and it is not a problem to use the symbol a in different context in this chapter.
b) Due to the independence and equal mixing rate of $\{\boldsymbol{\varepsilon}_i\}_i$ and $\{r_{iM}\}_i$ we have that $\{r_{iM}\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\}_i$ is also α-mixing with the same rate.

4.2 Main Results

We present and discuss our results on the limit behavior of the test procedures both under the null hypothesis H_0 as well as under the alternative H_1. The proofs of theorems are postponed to Section 4.4.

Theorem 4.1. *Let the observations follow model* (4.3) *and assumptions (A*.1)–(A*.4), (B) be satisfied. Further let $\widehat{\mathbf{\Sigma}}_m$ be a consistent estimate of $\mathbf{\Sigma}$*

$$\widehat{\mathbf{\Sigma}}_m - \mathbf{\Sigma} = o_P(1). \tag{4.10}$$

Then, under the null hypothesis H_0

$$\max_{1 \le k \le \lfloor mT \rfloor} \left(\frac{\widehat{Q}^C_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \right) \xrightarrow{\mathcal{D}} \sup_{0 \le t \le T/(T+1)} \left(\frac{\sum_{j=1}^d W_j^2(t)}{t^{2\gamma}} \right),$$

where $\{W_j(t),\ t \in [0,1]\}$, $j = 1, \dots, d$, are independent (standard) Wiener processes.

Now we consider local alternatives, i.e. the model (4.3) with $\delta_m \to 0$ and $k^* < \lfloor mT \rfloor$.
Theorem 4.2. *Let the observations follow model* (4.3) *and assumptions (A*.1)–(A*.4), (B) be satisfied. Further let $\widehat{\mathbf{\Sigma}}_m$ be a consistent estimate of $\mathbf{\Sigma}$. Then*
(i) *under (4.3) with $\delta_m = m^{-1/2}$ and $k^* = \lfloor ms \rfloor$, $0 \le s < T$,*

$$\max_{1 \le k \le \lfloor mT \rfloor} \left(\frac{\widehat{Q}^C_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \right) \xrightarrow{\mathcal{D}} \sup_{0 \le t \le T/(T+1)} \left(\frac{\sum_{j=1}^d (W_j(t) + h_j(t,s))^2}{t^{2\gamma}} \right),$$

where $\{W_j(t),\ t \in [0,1]\}$, $j = 1, \ldots, d$, *are independent Wiener processes,*

$$\boldsymbol{h}(t,s) = (t - s(1-t))^+ \operatorname{var}\{r_{0M}\}\boldsymbol{\Sigma}^{-1/2}(\lambda_1'(0)\beta_1^1, \ldots, \lambda_d'(0)\beta_d^1)^T,$$

$$\boldsymbol{h}(t,s) = (h_1(t,s), \ldots, h_d(t,s))^T, \quad 0 \le t \le T/(T+1), 0 \le s < T.$$

(ii) *under (4.3) with* $\delta_m \to 0$, $|\delta_m| m^{1/2} \to \infty$, $\liminf_{m\to\infty}(\lfloor mT \rfloor - k^*)/m > 0$, *and* $\beta_j^1 \neq 0$ *for at least one* j*,*

$$\max_{1 \le k \le \lfloor mT \rfloor} \left(\frac{\widehat{Q}^C_{\boldsymbol{\psi}}(m,k)}{q_\gamma^2(k/m)} \right) \xrightarrow{P} \infty.$$

Remark 4.2. (i) By Theorem 4.1 the asymptotic behavior of the test statistic under the null hypothesis is the same as the corresponding one in Chapter 3. Thus we can use the same critical values (see Chapter 7) to ensure that (1.6) is fulfilled.

(ii) Notice also that the limit distribution in Theorem 4.2 (i) is only sensitive w.r.t. a change in the β_j's, but not w.r.t. a change in the α_j's. Moreover, on checking the proof one can conclude that, in case of a contiguous change in the α_j's only, the limit distribution is the same as under H_0 .

(iii) A time dependent part of the shift is again the function $p(t,s)$ from (2.16). Note that there is a misprint in Chochola et al. [2013] in this regard.

Estimation of the variance matrix

In this section we deal with an estimator of the long-run variance matrix $\boldsymbol{\Sigma}$ given in (4.9). Notice that

$$\boldsymbol{\Sigma} = \sum_{k=-\infty}^{\infty} \boldsymbol{\Gamma}_k,$$

where $\boldsymbol{\Gamma}_k = \mathrm{E}[(r_{0M} - \mathrm{E}\, r_{0M})(r_{kM} - \mathrm{E}\, r_{kM})\boldsymbol{\psi}(\boldsymbol{\varepsilon}_0)\boldsymbol{\psi}(\varepsilon_k)^T]$ for $k \ge 0$ and $\boldsymbol{\Gamma}_{-k} = \boldsymbol{\Gamma}_k^T$ for $k < 0$.

We consider again a kernel type estimator of $\boldsymbol{\Sigma}$ based on the first m observations defined as

$$\widehat{\boldsymbol{\Sigma}}_m = \sum_{|k| < \Lambda_m} w(k/\Lambda_m)\widehat{\boldsymbol{\Gamma}}_k \tag{4.11}$$

where $\widehat{\boldsymbol{\Gamma}}_k$ is the k-th lag sample covariance corresponding to $\boldsymbol{\Gamma}_k$, i.e.,

$$\widehat{\boldsymbol{\Gamma}}_k = \begin{cases} \frac{1}{m}\sum_{i=1}^{m-k} \widetilde{r}_{iM}\widetilde{r}_{i+k,M}\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})^T, & k \ge 0, \\ \widehat{\boldsymbol{\Gamma}}_{-k}^T, & k < 0, \end{cases} \tag{4.12}$$

with $\widetilde{r}_{iM}$ given in (4.4) and $\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i)$ in (4.6).

We work with the Bartlett kernel again, i.e.,

$$w(t) = (1 - |t|)\, I\{|t| < 1\}, \quad t \in \mathbb{R}, \tag{4.13}$$

same results can be however also obtained for the flat-top kernel.

Theorem 4.3. *Let Assumptions (A*.1)–(A*.4), and (B) be satisfied and moreover for some $q > 4$ holds*

$$\mathrm{E}\,|r_{iM}\psi_j(e_{ij})|^q < \infty,\ j = 1,\dots,d, \quad \sum_{k=1}^{\infty}\alpha(k)^{1-4/q} < \infty, \quad \sum_{k=1}^{\infty}\alpha_M(k)^{1-4/q} < \infty,$$

Let

$$\Lambda_m \to \infty, \qquad \Lambda_m m^{-1/(2(2+\Delta))} \to 0.$$

Then for $\widehat{\boldsymbol{\Sigma}}_m$ given in (4.11) with the kernel (4.13)

$$\widehat{\boldsymbol{\Sigma}}_m = \boldsymbol{\Sigma} + o_P(1).$$

4.3 Auxiliary Results

Let us recall that $C > 0$ is a generic constant, which may vary from case to case.

At first we gather some properties of the sequence $\{r_{iM}\}$. From now on let us denote

$$r_{iM}^0 := r_{iM} - \mathrm{E}\,r_{iM}. \tag{4.14}$$

Lemma 4.1. *Let Assumption (B) be satisfied. Then,*

(i) *there is a constant $D > 0$ such that, for every $\ell \in \mathbb{Z}$ and $m \in \mathbb{N}$,*

$$\mathrm{E}\Big|\sum_{i=\ell+1}^{\ell+m} r_{iM}^0\Big|^2 \le Dm,$$

and, for $b_1 \ge b_2 \ge \ldots \ge b_m > 0$,

$$\mathrm{E}\max_{1\le k\le m}\Big|b_k\sum_{i=\ell+1}^{\ell+m} r_{iM}^0\Big|^2 \le C\sum_{k=1}^{m} b_k^2; \tag{4.15}$$

(ii)

$$\sum_{i=1}^{m} r_{iM}^0 = O_P\big(m^{1/2}\big),$$

$$\max_{1\le i\le m}|r_{iM}^0| = O_P\big(m^{1/(2+\Delta)}\big),$$

$$\sum_{i=1}^{m}|r_{iM}^0|^a = O_P\big(m^{\max(1,a/(2+\Delta))}\big)$$

for $a > 0$;

(iii) *for some $D > 0$ and $0 \le \gamma < 1/2$,*

$$\mathrm{E}\left(\max_{1\le k\le\lfloor mT\rfloor}\frac{|\sum_{i=m+1}^{m+k} r_{iM}^0|}{\sqrt{m}\,(k/m)^{\gamma}}\right)^2 \le D;$$

(iv)

$$\sup_{0\le t\le T}\left|\frac{\sum_{i=m+1}^{m+\lfloor mt\rfloor}(r_{iM}^0)^2}{\sum_{i=1}^{m}(r_{iM}^0)^2}-t\right| \xrightarrow{P} 0.$$

Proof. (i) It follows directly from Lemma II.
(ii) By Chebyshev's inequality and assertion (i) above

$$P\Big(\Big|\sum_{i=1}^{m} r_{iM}^0\Big| \ge \lambda\Big) \le \frac{D}{\lambda^2}m.$$

Next, note that

$$\max_{1\le i\le m}|r_{iM}^0| \le D\Big(\frac{1}{m}\sum_{i=1}^{m}|r_{iM}^0|^{2+\delta}\Big)^{1/(2+\delta)}m^{1/(2+\delta)}$$

for any $\delta \ge 0$. Since the sequence $\{r_{iM}\}$ is stationary and ergodic, also $\{g(r_{iM})\}$ is stationary and ergodic, where g is a measurable function, and, if $E|g(r_{iM})| < \infty$, the ergodic theorem implies

$$\frac{1}{m}\sum_{i=1}^{m} g(r_{iM}) \to Eg(r_{iM}) \quad \text{a.s.} \tag{4.16}$$

Hence,

$$\frac{1}{m}\sum_{i=1}^{m}|r_{iM}^0|^{2+\Delta} \to E|r_{0M}^0|^{2+\Delta} \quad \text{a.s.,}$$

and therefore

$$\max_{1\le i\le m}|r_{iM}^0| = O_P\big(m^{1/(2+\Delta)}\big),$$

which easily implies

$$\begin{aligned}\sum_{i=1}^{m}|r_{iM}^0|^a &= O_P\Big(\sum_{i=1}^{m}|r_{iM}^0|^{\min(a,2+\Delta)}\max_{1\le i\le m}|r_{iM}^0|^{\max(0,a-(2+\Delta))}\Big) =\\ &= O_P\big(m^{\max(1,a/(2+\Delta))}\big).\end{aligned}$$

(iii) It follows immediately from (4.15) since $\sum_{k=1}^{\lfloor mT\rfloor} k^{-2\gamma} = O(m^{1-2\gamma})$.
(iv) Note that, by (4.16),

$$\frac{1}{m}\sum_{i=1}^{m}(r_{iM}^0)^2 \;\to\; \text{var}(r_{0M}) \quad \text{a.s.,}$$

hence, due to the strict stationarity, also

$$\sup_{0\le t\le T}\left|\frac{1}{m}\Big\{\sum_{i=m+1}^{m+\lfloor mt\rfloor}(r_{iM}^0)^2 - \lfloor mt\rfloor\,\text{var}(r_{0M})\Big\}\right| \xrightarrow{P} 0. \tag{4.17}$$

On combining the above two assertions, the proof of (iv) can be completed. □

Remark 4.3. The following two lemmas are crucial assertions for the proof of the limit behavior of the estimators $\widehat{\alpha}_{jm},\ \widehat{\beta}_{jm},\ j=1,\dots,d$. It is however convenient to introduce auxiliary estimators $\widehat{\alpha}^*_{jm}$ and $\widehat{\beta}^*_{jm}$ as minimizers of

$$\sum_{i=1}^{m} \rho_j\big(\varepsilon_{i,j} - (a^*_j + b^*_j \widetilde{r}_{iM})/\sqrt{m}\big) \tag{4.18}$$

w.r.t. a^*_j and b^*_j for $j=1,\dots,d$, since their theoretical counterparts are equal to zero. Clearly,

$$\widehat{\alpha}^*_{jm} = \sqrt{m}(\widehat{\alpha}_{jm} - \alpha^0_j), \quad \widehat{\beta}^*_{jm} = \sqrt{m}(\widehat{\beta}_{jm} - \beta^0_j). \tag{4.19}$$

Usually, the estimators $\widehat{\alpha}^*_{jm}$ and $\widehat{\beta}^*_{jm}$ can be obtained as solutions of the equations

$$\sum_{i=1}^{m} \psi_j(\varepsilon_{i,j} - (a^*_j + b^*_j \widetilde{r}_{iM})/\sqrt{m}) = 0, \tag{4.20}$$

$$\sum_{i=1}^{m} \psi_j(\varepsilon_{i,j} - (a^*_j + b^*_j \widetilde{r}_{iM})/\sqrt{m})\widetilde{r}_{iM} = 0, \tag{4.21}$$

w.r.t. a^*_j, b^*_j for $j=1,\dots,d$.

We would like to use the theory of α-mixing, but the problem is that $\widetilde{r}_{iM}$ is not α-mixing due to the centering by $\overline{r}_{mM}$. Therefore we introduce one more reparametrization and define $\widehat{\alpha}'_{jm}$ and $\widehat{\beta}'_{jm}$ as minimizers of

$$\sum_{i=1}^{m} \rho_j(\varepsilon_{i,j} - (a'_j + b'_j r^0_{iM})/\sqrt{m}) \tag{4.22}$$

w.r.t. a'_j and b'_j for $j=1,\dots,d$. Since $\overline{r}_{mM} - \mathrm{E}\, r_{iM}$ does not depend on i and $\widetilde{r}_{iM} = r^0_{iM} - (\overline{r}_{mM} - \mathrm{E}\, r_{iM})$ one can easily see that

$$\widehat{\beta}'_{jm} = \widehat{\beta}^*_{jm} \text{ and } \widehat{\alpha}'_{jm} = \widehat{\alpha}^*_{jm} - \widehat{\beta}^*_{jm}(\overline{r}_{mM} - \mathrm{E}\, r_{iM}).$$

By Lemma 4.1 is $\overline{r}_{mM} - \mathrm{E}\, r_{iM} = O_P(m^{-1/2})$ and thus these estimators are sufficiently close to each other. Therefore the results derived for primed estimates holds also for the starred ones.

In Lemmas 4.2, 4.3 and 4.4 we omit the index j, i.e., we write $\varepsilon_i, \psi, \dots$ instead of $\varepsilon_{ij}, \psi_j, \dots$. In the following E^* denotes the conditional expectation given $r_{iM},\ i = 1,\dots,m$.

Lemma 4.2. *Let the assumptions of Theorem 4.1 be satisfied. Then, for arbitrary* $D > 0$

$$\sup_{|a|+|b|\le D} \Big| Z_m(a,b) - \frac{\lambda'(0)}{2}\big(a^2 + b^2 \frac{1}{m}\sum_{i=1}^{m}(r^0_{iM})^2\big)\Big| = O_P(m^{-\eta}),$$

for some $\eta > 0$*, where*

$$Z_m(a,b) = \sum_{i=1}^{m} \Big(\rho(\varepsilon_i - (a + br^0_{iM})/\sqrt{m}) - \rho(\varepsilon_i) + ((a + br^0_{iM})/\sqrt{m})\, \psi(\varepsilon_i)\Big).$$

Proof. Uniformity of the result will be treated similarly as in Lemma 2.1, i.e. discretizing the situation to a maximum over a finite number of points. Then, for fixed $a,\ b$, we just need to derive a proper approximation for the conditional expectation and variance of $Z_m(a,b)$.

First note that due to Lemma 4.1 (ii) we have

$$P\Big(\sup_{|a|+|b|\le D}\max_{i=1,\dots,m}|a+br_{iM}^0|/\sqrt{m}\le D_2\Big)\to 1,$$

where D_2 is from Assumption (A*.3), and thus we can restrict ourselves to this set only.

Now we introduce some convenient short-hand notations

$$d_i := a+br_{iM}^0 \quad \text{and} \quad f(\varepsilon_i,x,d_i) := \operatorname{sign} d_i\big(-\psi(\varepsilon_i - x\operatorname{sign} d_i)+\psi(\varepsilon_i)\big),\ x>0. \tag{4.23}$$

Note that, for any $\delta\in\mathbb{R}$, $x>0$ and $i\in\mathbb{Z}$, we have $f(\varepsilon_i,x,\delta)\ge 0$ and

$$\rho(\varepsilon_i-\delta)-\rho(\varepsilon_i)+\delta\psi(\varepsilon_i)=\int_0^{|\delta|}\operatorname{sign}\delta\big(-\psi(\varepsilon_i-x\operatorname{sign}\delta)+\psi(\varepsilon_i)\big)dx=\int_0^{|\delta|}f(\varepsilon_i,x,\delta)dx\ge 0.$$

Thus

$$Z_m(a,b)=\sum_{i=1}^m\int_0^{|d_i|/\sqrt{m}}f(\varepsilon_i,x,d_i)dx =: \sum_{i=1}^m Y_i(a,b).$$

We can start with the conditional expectation. Direct calculations in combination with Lemma 4.1 result in

$$\begin{aligned}
\mathrm{E}^*\,Z_m(a,b) &= \mathrm{E}^*\sum_{i=1}^m\operatorname{sign} d_i\int_0^{|d_i|/\sqrt{m}}f(\varepsilon_i,x,d_i)dx=\sum_{i=1}^m\int_0^{|d_i|/\sqrt{m}}\lambda(x\operatorname{sign} d_i)dx=\\
&=\sum_{i=1}^m\lambda'(0)d_i^2\frac{1}{2m}+O_P\Big(\sum_{i=1}^m|d_i|^3\frac{1}{m^{3/2}}\Big)=\\
&=\frac{1}{2}\lambda'(0)\big(a^2+b^2\frac{1}{m}\sum_{i=1}^m(r_{iM}^0)^2\big)+O_P\big(m^{-1/2}|a|^3+|b|^3m^{-3/2+\max(1,\,3/(2+\Delta))}\big),
\end{aligned} \tag{4.24}$$

uniformly in $|a|+|b|\le D$, where we used the fact that for δ small enough holds

$$\Big|\int_0^\delta\lambda(x)\,dx-\int_0^\delta\lambda'(0)x\,dx\Big|=O(|\delta|^3).$$

Now we study the uniformity of the result. Define a grid of points (a_j,b_k), $j,k=0,\dots,N$, such that $a_0=b_0=-D$, $a_N=b_N=D$, $(0,0)$ is part of the grid and $a_j-a_{j-1}=b_k-b_{k-1}=m^{-\xi}$ for some $\xi>0$, $j,k=1,\dots,N-1$ and distance is not more than $m^{-\xi}$ for the first and last pair.

Now assume we have $a_{j-1}\le a\le a_j,\ b_{k-1}\le b\le b_k$ for j,k fixed. Since $f(\varepsilon_i,x,d_i)\ge 0$, the estimation of $Y_i(a,b)$ (and thus also of $Z_m(a,b)$) from below and above is equivalent to estimation of $|d_i|=\big|a+br_{iM}^0\big|$.

We need to distinguish four cases:

$$
\begin{aligned}
\text{i) } a+br_{iM}^0 \ge 0,\ r_{iM}^0 \ge 0: &\quad a_{j-1}+b_{k-1}r_{iM}^0 \le \left|a+br_{iM}^0\right| \le a_j+b_kr_{iM}^0\\
\text{ii) } a+br_{iM}^0 \ge 0,\ r_{iM}^0 < 0: &\quad a_{j-1}+b_kr_{iM}^0 \le \left|a+br_{iM}^0\right| \le a_j+b_{k-1}r_{iM}^0\\
\text{iii) } a+br_{iM}^0 < 0,\ r_{iM}^0 \ge 0: &\quad \left|a_j+b_kr_{iM}^0\right| \le \left|a+br_{iM}^0\right| \le \left|a_{j-1}+b_{k-1}r_{iM}^0\right|\\
\text{iv) } a+br_{iM}^0 < 0,\ r_{iM}^0 < 0: &\quad \left|a_j+b_{k-1}r_{iM}^0\right| \le \left|a+br_{iM}^0\right| \le \left|a_{j-1}+b_kr_{iM}^0\right|
\end{aligned}
$$

For the ease of further writing denote

$$
d_i^{UU} := a_j+b_kr_{iM}^0, \quad d_i^{LL} := a_{j-1}+b_{k-1}r_{iM}^0, \quad d_i^{UL} := a_j+b_{k-1}r_{iM}^0, \quad d_i^{LU} := a_{j-1}+b_kr_{iM}^0,
$$

where U means the upper boundary and L the lower one in turns for a, b. Also note that since $(0,0)$ is part of the grid, the sign $a+br_{iM}^0$ can change only on the grid itself.

Lets focus on i)

$$
\begin{aligned}
Y_i(a,b) - \mathrm{E}^* Y_i(a,b) &\le \int_0^{|d_i^{UU}|/\sqrt{m}} f(\varepsilon_i, x, d_i^{UU})dx - \mathrm{E}^* \int_0^{|d_i^{LL}|/\sqrt{m}} f(\varepsilon_i, x, d_i^{UU})dx =\\
&= Y_i(a_j,b_k) - \mathrm{E}^* Y_i(a_j,b_k) + \mathrm{E}^* \int_{|d_i^{LL}|/\sqrt{m}}^{|d_i^{UU}|/\sqrt{m}} f(\varepsilon_i, x, d_i^{UU})dx
\end{aligned}
\tag{4.25}
$$

and similarly

$$
Y_i(a,b) - \mathrm{E}^* Y_i(a,b) \ge Y_i(a_{j-1},b_{k-1}) - \mathrm{E}^* Y_i(a_{j-1},b_{k-1}) - \mathrm{E}^* \int_{|d_i^{LL}|/\sqrt{m}}^{|d_i^{UU}|/\sqrt{m}} f(\varepsilon_i, x, d_i^{UU})dx.
\tag{4.26}
$$

We estimate now the last term present both in (4.25) and (4.26), which will be useful for us later

$$
\begin{aligned}
\mathrm{E}^* \int_{|d_i^{LL}|/\sqrt{m}}^{|d_i^{UU}|/\sqrt{m}} f(\varepsilon_i, x, d_i^{UU})dx &= \int_{|d_i^{LL}|/\sqrt{m}}^{|d_i^{UU}|/\sqrt{m}} \lambda(x)dx =\\
&= \frac{\lambda'(0)}{2m}\left[(d_i^{UU})^2 - (d_i^{LL})^2 + O_P(m^{-1/2})\right] \le\\
&\le Cm^{-1-\xi}(1 + r_{iM}^0 + (r_{iM}^0)^2),
\end{aligned}
\tag{4.27}
$$

where the constant C can be chosen the same for all $i = 1, \ldots, m$ and all cases i) – iv).

By (4.25),(4.26) and (4.27)

$$
\begin{aligned}
\sup_{|a|+|b|\le D} |Z_m(a,b) - E^* Z_m(a,b)| \le \max_{0\le j,k\le N} &|Z_m(a_j,b_k) - E^* Z_m(a_j,b_k)| +\\
&+ Cm^{-1-\xi}\Big(m + \sum_{i=1}^m \left|r_{iM}^0\right| + \sum_{i=1}^m (r_{iM}^0)^2\Big).
\end{aligned}
\tag{4.28}
$$

By Lemma 4.1 we have $\sum_{i=1}^m \left|r_{iM}^0\right| = O_P(m)$ and $\sum_{i=1}^m (r_{iM}^0)^2 = O_P(m)$ thus the last term on RHS in (4.27) is of order $O_P(m^{-\xi})$. Thus the supremum can be reduced to maximum over the finite grid, which we will treat now.

Assume that (a, b) is a fixed point of the grid, i.e. $a = a_j,\ b = b_k$ for some $0 \leq j, k \leq N$. Towards estimation of $|Z_m(a, b) - E^* Z_m(a, b)|$ it is enough to calculate the variance of Z_m since

$$\mathrm{E}(Z_m - \mathrm{E}^* Z_m)^2 \leq \mathrm{E}(Z_m - \mathrm{E}\, Z_m)^2.$$

Moreover by Lemma 4.1 we have $\max_{1\leq i\leq m} \left|r_{iM}^0\right| = O_P(m^{1/(2+\Delta)})$ and thus we can make

$$P\Big(\max_{1\leq i\leq m} \left|r_{iM}^0\right| \geq K m^{1/(2+\Delta)}\Big) \tag{4.29}$$

arbitrary small by choosing K large enough. Thus it is sufficient to focus on trimmed version $\widehat{Z}_m(a, b)$ of $Z_m(a, b)$

$$\widehat{Z}_m(a, b) := \sum_{i=1}^m Y_i(a, b) I\big\{\left|r_{iM}^0\right| \leq K m^{1/(2+\Delta)}\big\}.$$

We will further use notation $I_i := I\big\{\left|r_{iM}^0\right| \leq K m^{1/(2+\Delta)}\big\}$ for the trimming term. Since $Y_i(a, b)I_i$ is α-mixing we can use Lemma II and thus

$$\mathrm{E}(\widehat{Z}_m(a, b) - \mathrm{E}\,\widehat{Z}_m(a, b))^2 \leq Cm \sum_{j=1}^{\infty} (\alpha(j))^{\frac{\Delta}{2+\Delta}} \Big(\mathrm{E}\left|Y_i(a, b)I_i\right|^{2+\Delta}\Big)^{\frac{2}{2+\Delta}}. \tag{4.30}$$

So we need to study $\mathrm{E}\left|Y_i(a, b)I_i\right|^{2+\Delta}$ and due to monotonicity of ψ we have

$$\begin{aligned}
&\mathrm{E}\left|\int_0^{|d_i|/\sqrt{m}} \operatorname{sign} d_i\big(-\psi(\varepsilon_i - x \operatorname{sign} d_i) + \psi(\varepsilon_i)\big)dx\ I_i\right|^{2+\Delta} \leq \\
&\leq C\,\mathrm{E}\left|\int_0^{|d_i|/\sqrt{m}} \big(-\psi(\varepsilon_i - |d_i|/\sqrt{m}) + \psi(\varepsilon_i)\big)dx\ I_i\right|^{2+\Delta} + \\
&+ C\,\mathrm{E}\left|\int_0^{|d_i|/\sqrt{m}} \big(\psi(\varepsilon_i + |d_i|/\sqrt{m}) - \psi(\varepsilon_i)\big)dx\ I_i\right|^{2+\Delta} =: A_{m1} + A_{m2},
\end{aligned}$$

where both integrands on RHS are non-negative constants. Thus

$$\begin{aligned}
A_{m1} &\leq C\,\mathrm{E}\Big[(|d_i|/\sqrt{m})^{2+\Delta}\left|-\psi(\varepsilon_i - |d_i|/\sqrt{m}) + \psi(\varepsilon_i)\right|^{2+\Delta} I_i\Big] \leq \\
&\leq C\,\mathrm{E}\left[\left(\frac{\left|a + b r_{iM}^0\right|}{\sqrt{m}}\right)^{2+\Delta} \frac{\left|a + b r_{iM}^0\right|}{\sqrt{m}} I_i\right] \leq \\
&\leq C m^{-\frac{3+\Delta}{2}} \Big(|a|^{3+\Delta} + |b|^{3+\Delta}\,\mathrm{E}\big(\left|r_{iM}^0\right|^{3+\Delta} I_i\big)\Big)
\end{aligned}$$

and the similar upper bound holds also for A_{m2}. Since $\sum_{j=1}^{\infty} (\alpha(j))^{\frac{\Delta}{2+\Delta}} < \infty$ by the assumptions, we have from (4.30)

$$\mathrm{E}(\widehat{Z}_m(a, b) - \mathrm{E}\,\widehat{Z}_m(a, b))^2 \leq Cm\Big[m^{-\frac{3+\Delta}{2+\Delta}} |a|^{2\frac{3+\Delta}{2+\Delta}} + |b|^{2\frac{3+\Delta}{2+\Delta}} m^{-\frac{3+\Delta}{2+\Delta}} \Big(\mathrm{E}\big(\left|r_{iM}^0\right|^{3+\Delta} I_i\big)\Big)^{\frac{2}{2+\Delta}}\Big]. \tag{4.31}$$

Since $\mathrm{E}\left|r_{iM}^0\right|^{2+\Delta} < \infty$ we can estimate

$$\Big(\mathrm{E}\big(\left|r_{iM}^0\right|^{3+\Delta} I\big[\left|r_{iM}^0\right| \le m^{\frac{1}{2+\Delta}}\big]\big)\Big)^{\frac{2}{2+\Delta}} \le Cm^{\frac{1}{2+\Delta}\frac{2}{2+\Delta}} = O(m^{2/(2+\Delta)^2})$$

and thus the order of the term with $|b|$ in (4.31) is $m^{-\frac{1}{2+\Delta}+\frac{2}{(2+\Delta)^2}} = m^{-\frac{\Delta}{(2+\Delta)^2}}$ and of the one with $|a|$ is $m^{-\frac{1}{2+\Delta}}$. Thus we can choose $\xi > 0$ (the grid parameter) such that

$$m^{2\xi}\left|\widehat{Z}_m(a,b) - \mathrm{E}\,\widehat{Z}_m(a,b)\right| = O_P(m^{-\eta})$$

for some $\eta > 0$, where the term $m^{2\xi}$ comes from treating

$$\max_{1\le j,k\le N}\left|\widehat{Z}_m(a_j,b_k) - \mathrm{E}\,\widehat{Z}_m(a_j,b_k)\right|$$

similarly as in Lemma 2.1. Due to definition of the trimming, the same is true for the $Z_m(a,b)$ and we can thus conclude that

$$\max_{1\le j,k\le N}\left|Z_m(a_j,b_k) - \mathrm{E}\,Z_m(a_j,b_k)\right| = O_P(m^{-\eta})$$

which finishes the proof of Lemma 4.2. □

Lemma 4.3. *Let the assumptions of Theorem 4.1 be satisfied. Then, for arbitrary* $D > 0$

$$\sup_{|a|+|b|\le D}\left|\boldsymbol{M}_m(a,b) + \lambda'(0)(a, b\,\frac{1}{m}\sum_{i=1}^m (r_{iM}^0)^2)^T\right| = O_P\big(m^{-\eta}\big)$$

for some $\eta > 0$, *where*

$$\boldsymbol{M}_m(a,b) = \frac{1}{\sqrt{m}}\sum_{i=1}^m (1, r_{iM}^0)^T\big(\psi(\varepsilon_i - (a + br_{iM}^0)/\sqrt{m}) - \psi(\varepsilon_i)\big).$$

Proof. The idea of the proof is similar to that of the previous lemma, therefore we will not give all the details here. We concentrate on the key part only. Again one has to get suitable approximations for the conditional expectation of $\boldsymbol{M}_m(a,b)$ and the (2×2)-variance matrix

$$\mathrm{E}\Big(\boldsymbol{M}_n(a,b) - \mathrm{E}^*\,\boldsymbol{M}_n(a,b)\Big)\Big(\boldsymbol{M}_n(a,b) - \mathrm{E}^*\,\boldsymbol{M}_n(a,b)\Big)^T. \tag{4.32}$$

We start with the conditional expectation. Keeping the notation $d_i = a + br_{iM}^0$ we get similarly as in (4.24)

$$\begin{aligned}\mathrm{E}^*\,\boldsymbol{M}_m^T(a,b) = &= \frac{1}{\sqrt{m}}\sum_{i=1}^m (1, r_{iM}^0)(-\lambda(d_i/\sqrt{m})) = \\ &= -\frac{1}{\sqrt{m}}\lambda'(0)\sum_{i=1}^m (1, r_{iM}^0)\big(d_i/\sqrt{m} + O_P\big(|d_i/\sqrt{m}|^2\big)\big) = \\ &= -\lambda'(0)\Big(a,\, b\frac{1}{m}\sum_{i=1}^m (r_{iM}^0)^2\Big) + O_P\big((a^2+b^2)m^{-1/2} + b^2 m^{-3/2+\max(1,3/(2+\Delta))}\big),\end{aligned}$$

uniformly in $|a| + |b| \le D$.

For estimation the variance matrix (4.32) we introduce again the trimming with the same notation $I_i = I\{ |r^0_{iM}| \le Km^{1/(2+\Delta)}\}$. Similarly as earlier we can substitute the inner conditional expectation in (4.32) by the unconditional one. As all the elements of the matrix are similar we focus on one specific only. By Lemma II

$$\begin{aligned}
&\operatorname{var}\Big\{\frac{1}{\sqrt{m}}\sum_{i=1}^{m} r^0_{iM}\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big)I_i\Big\} \le \\
&\le C\frac{m}{m}\sum_{j=1}^{\infty}(\alpha(j))^{\frac{\Delta}{2+\Delta}}\Big(\mathrm{E}\left|r^0_{iM}\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big)I_i\right|^{2+\Delta}\Big)^{\frac{2}{2+\Delta}}, \qquad (4.33)
\end{aligned}$$

where we used the fact that for any $p \ge 1$

$$\mathrm{E}\,|X - \mathrm{E}\,X|^p \le 2^{p-1}\,\mathrm{E}(|X|^p + |\mathrm{E}\,X|^p) \le 2^p\,\mathrm{E}\,|X|^p\,.$$

Hence again it suffices to estimate

$$\begin{aligned}
\mathrm{E}\left|r^0_{iM}\big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big)I_i\right|^{2+\Delta} &\le Cm^{-1/2}\,\mathrm{E}\Big(\left|r^0_{iM}\right|^{2+\Delta}\left|a + br^0_{iM}\right| I_i\Big) \le \\
&\le Cm^{-1/2}\big(|a| + |b|\,\mathrm{E}\big[\left|r^0_{iM}\right|^{3+\Delta} I_i\big]\big) \le Cm^{-1/2}(|a| + |b|\,m^{1/(2+\Delta)}) \\
&\le C(|a| + |b|)m^{-\Delta/(2(2+\Delta))} \qquad (4.34)
\end{aligned}$$

uniformly in $|a| + |b| \le D$.

Coming back to (4.33) we have that the RHS is if order $m^{-\Delta/(2+\Delta)^2}$, which concludes the proof. □

Remark 4.4. Proceeding in a standard way, Lemmas 4.2 and 4.3 ensure that $\widehat{\alpha}'_{jm} = O_P(1)$ and $\widehat{\beta}'_{jm} = O_P(1)$ and, moreover, we get the asymptotic representations

$$\widehat{\alpha}'_{jm} = \frac{1}{\sqrt{m}\lambda'_j(0)}\sum_{i=1}^{m}\psi_j(\varepsilon_{i,j}) + O_P(m^{-\eta}), \qquad (4.35)$$

$$\widehat{\beta}'_{jm} = \frac{\sqrt{m}}{\lambda'_j(0)}\frac{1}{\sum_{i=1}^{m}(r^0_{iM})^2}\sum_{i=1}^{m}\psi_j(\varepsilon_{i,j})r^0_{iM} + O_P(m^{-\eta}), \qquad (4.36)$$

for some $\eta > 0$. Note that due to Remark 4.3 the same asymptotic representation holds as well for

$$\widehat{\alpha}^*_{jm} = \sqrt{m}(\widehat{\alpha}_{jm} - \alpha^0_j), \text{ and } \widehat{\beta}^*_{jm} = \sqrt{m}(\widehat{\beta}_{jm} - \beta^0_j).$$

Lemma 4.4. *Let the assumptions of Theorem 4.1 be satisfied. Then, for any $T > 0$, as $m \to \infty$,*

$$\max_{1\le k\le \lfloor mT\rfloor}\left|N_{k,m}(\widehat{\alpha}'_m, \widehat{\beta}'_m) + \lambda'(0)\frac{1}{m}\Big(\widehat{\alpha}'_m\sum_{i=m+1}^{m+k} r^0_{iM} + \widehat{\beta}'_m\sum_{i=m+1}^{m+k}(r^0_{iM})^2\Big)\right|\Big/(k/m)^{\gamma} = O_P(m^{-\eta}),$$

for some $\eta > 0$, where

$$N_{k,m}(a,b) = \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} r_{iM}^0 \big(\psi(\varepsilon_i - a/\sqrt{m} - br_{iM}^0/\sqrt{m}) - \psi(\varepsilon_i)\big)$$

and $\widehat{\alpha}'_m, \widehat{\beta}'_m$ represent the proper coordinate of estimators defined in (4.22).

Proof. The proof is based on the uniform result for $N_{k,m}(a,b)$ over $|a| + |b| \leq D$ for any $D > 0$, i.e.

$$\sup_{|a|+|b|\leq D} \max_{1\leq k\leq \lfloor mT \rfloor} \left| N_{k,m}(a,b) + \lambda'(0)\frac{1}{m}\Big(a \sum_{i=m+1}^{m+k} r_{iM}^0 + b \sum_{i=m+1}^{m+k} (r_{iM}^0)^2\Big) \right| \Big/ (k/m)^\gamma = O_P(m^{-\eta}). \tag{4.37}$$

Plugging in the estimators is then accomplished similarly as in proof of Lemma 2.3, since by Remark 4.4 we know that $\widehat{\alpha}'_m = O_P(1)$ and $\widehat{\beta}'_m = O_P(1)$, i.e. $P\big(|\widehat{\alpha}'_m| + |\beta'_m| > D\big)$ can be made arbitrary small by choosing D large.

Towards (4.37) we proceed similarly to the proof of Lemma 4.3. We need now however conditional expectation given $r_{iM},\ i = m+1, \ldots, m(T+1)$, which we denote again E^*. Direct calculations give

$$\begin{aligned} \mathrm{E}^*\, N_{k,m}(a,b) &= -\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} r_{iM}^0 \lambda\big((a + br_{iM}^0)/\sqrt{m}\big) = \\ &= -\lambda'(0)\frac{1}{m}\Big(a \sum_{i=m+1}^{m+k} r_{iM}^0 + b \sum_{i=m+1}^{m+k} (r_{iM}^0)^2\Big) + O_P\big(m^{-\eta}\big), \end{aligned}$$

uniformly for $|a| + |b| \leq D$, with some $\eta > 0$.

Next we want to find an upper bound for

$$\mathrm{E}\Big(\max_{1\leq k\leq \lfloor mT \rfloor} (m/k)^\gamma |N_{k,m}(a,b) - \mathrm{E}^*\, N_{k,m}(a,b)|\Big)^2.$$

We proceed as in previous proof, only have to take the maximum of the cumulative sums into account. Using Lemma II we have

$$\begin{aligned} \mathrm{var}\Big\{\max_{1\leq k\leq \lfloor mT \rfloor} (m/k)^\gamma N_{k,m}(a,b)\Big\} &\leq C \log^2(2\lfloor mT \rfloor) m^{2\gamma-1} \sum_{k=1}^{\lfloor mT \rfloor} k^{-2\gamma}\ \cdot \\ &\cdot \Big(\mathrm{E}\left| r_{iM}^0 \big(\psi(\varepsilon_i - d_i/\sqrt{m}) - \psi(\varepsilon_i)\big) I_i \right|^{2+\Delta}\Big)^{\frac{2}{2+\Delta}}. \end{aligned} \tag{4.38}$$

The expected value part of RHS was already studied in (4.34) and thus we can conclude that RHS of (4.38) is $O(\log^2(m)\, m^{-\Delta/(2(2+\Delta))})$. This similarly as in Lemma 4.2 finishes the proof.

$\square$

4.4 Proofs

Proof of Theorem 4.1. The proof will proceed in two steps.
1. First we show that the limit behavior of the weighted partial sums

$$\widehat{\boldsymbol{H}}(m,k) = (\widehat{H}_1(m,k),\ldots,\widehat{H}_d(m,k))^T = \frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k}\widetilde{r}_{iM}\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i), \quad k=1,\ldots,\lfloor mT\rfloor$$

is the same as that of

$$\widetilde{\boldsymbol{H}}(m,k) = \frac{1}{\sqrt{m}}\Big(\sum_{i=m+1}^{m+k} r_{iM}^0\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) - \frac{\sum_{i=m+1}^{m+k}(r_{iM}^0)^2}{\sum_{i=1}^{m}(r_{iM}^0)^2}\sum_{i=1}^{m} r_{iM}^0\boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\Big). \tag{4.39}$$

Using Lemma 4.4 we have for every $j=1,\ldots,d$

$$\begin{aligned}\max_{1\le k\le\lfloor mT\rfloor}\Big|&\sum_{i=m+1}^{m+k} r_{iM}^0\psi_j\big(\varepsilon_{i,j} - (\widehat{\alpha}'_{jm} + \widehat{\beta}'_{jm}r_{iM}^0)/\sqrt{m}\big) - \sum_{i=m+1}^{m+k} r_{iM}^0\psi_j(\varepsilon_{i,j}) + \\ &+ \lambda'(0)\frac{1}{\sqrt{m}}\Big(\widehat{\alpha}'_{jm}\sum_{i=m+1}^{m+k} r_{iM}^0 + \widehat{\beta}'_{jm}\sum_{i=m+1}^{m+k}(r_{iM}^0)^2\Big)\Big|\Big/\big(\sqrt{m}(k/m)^\gamma\big) = O_P(m^{-\eta}) \end{aligned}\tag{4.40}$$

for some $\eta > 0$. Since $\frac{1}{\sqrt{m}}\widehat{\alpha}'_{jm} = O_P(m^{-1/2})$ and by Lemma 4.1

$$\max_{1\le k\le\lfloor mT\rfloor}\frac{|\sum_{i=m+1}^{m+k} r_{iM}^0|}{\sqrt{m}\,(k/m)^\gamma} = O_P(1),$$

we can simplify (4.40) accordingly. Using the asymptotical representation (4.36) of $\widehat{\beta}'_{jm}$ we arrive at

$$\begin{aligned}\max_{1\le k\le\lfloor mT\rfloor}\Big|\sum_{i=m+1}^{m+k} r_{iM}^0\psi_j(\widehat{\varepsilon}_{i,j}) - \Big(\sum_{i=m+1}^{m+k} r_{iM}^0\psi_j(\varepsilon_{i,j}) - \frac{\sum_{i=m+1}^{m+k}(r_{iM}^0)^2}{\sum_{i=1}^{m}(r_{iM}^0)^2}\sum_{i=1}^{m} r_{iM}^0\psi_j(\varepsilon_{i,j})\Big)\Big|\Big/ \\ \sqrt{m}\left(1+\frac{k}{m}\right)\left(\frac{k}{m}\right)^\gamma = O_P(m^{-\eta}).\end{aligned}$$

By Remark 4.3

$$\widehat{\boldsymbol{H}}(m,k) = \frac{1}{\sqrt{m}}\sum_{i=m+1}^{m+k} r_{iM}^0\boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i) + O_P(m^{-\eta})$$

and thus the asymptotic equivalence of $\widehat{\boldsymbol{H}}(m,k)$ and $\widetilde{\boldsymbol{H}}(m,k)$ is proven. This together with Assumption (4.10) further implies that the limit behavior of

$$\max_{1\le k\le\lfloor mT\rfloor}\widehat{Q}^C_{\boldsymbol{\psi}}(m,k)/q_\gamma^2(k/m),$$

is the same as that of

$$\max_{1\le k\le\lfloor mT\rfloor}\widetilde{Q}^C_{\boldsymbol{\psi}}(m,k)/q_\gamma^2(k/m),$$

where
$$\widetilde{Q}^C_{\boldsymbol{\psi}}(m,k) = \widetilde{\boldsymbol{H}}(m,k)^T \boldsymbol{\Sigma}^{-1} \widetilde{\boldsymbol{H}}(m,k). \tag{4.41}$$

2. Now we study the limit behavior of
$$\sum_{i=m+1}^{m+k} r^0_{iM} \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i), \quad k = 1, \ldots, \lfloor mT \rfloor,$$
and that of the related maximum of weighted quadratic forms
$$\max_{1 \le k \le \lfloor mT \rfloor} \widetilde{Q}^C_{\boldsymbol{\psi}}(m,k)/q^2_\gamma(k/m).$$
The proof follows lines of the one of Theorem 3.1. Define
$$\boldsymbol{Z}_m(t) = \frac{1}{\sqrt{m}} \sum_{i=1}^{\lfloor mt \rfloor} r^0_{iM} \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i), \quad 0 \le t \le T+1.$$
Since $\{r^0_{iM} \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i)\}$ is α-mixing we can show similarly to (3.13) that
$$\boldsymbol{Z}_m(\cdot) \overset{\mathcal{D}^d[0,T+1]}{\longrightarrow} \boldsymbol{W}_{\boldsymbol{\Sigma}}(\cdot),$$
where $\{\boldsymbol{W}_{\boldsymbol{\Sigma}}(t),\, t \in [0, T+1]\}$ is a centered Gaussian process with covariance function
$$E[\boldsymbol{W}_{\boldsymbol{\Sigma}}(t) \boldsymbol{W}^T_{\boldsymbol{\Sigma}}(s)] = \min(t,s) \boldsymbol{\Sigma}$$
and $\boldsymbol{\Sigma}$ is defined in (4.9). Recall that $\overset{\mathcal{D}^d[0,T+1]}{\longrightarrow}$ denotes weak convergence in the Skorokhod space $D^d[0, T+1]$. Next we study the process
$$\boldsymbol{H}(m, \lfloor mt \rfloor) = \boldsymbol{Z}_m(t+1) - \boldsymbol{Z}_m(1) - t\boldsymbol{Z}_m(1) = \boldsymbol{Z}_m(t+1) - (t+1)\boldsymbol{Z}_m(1), \quad 0 \le t \le T.$$
Following the lines from (3.14) to (3.15) we can conclude that
$$\boldsymbol{\Sigma}^{-1/2} \boldsymbol{H}(m, \lfloor m \cdot \rfloor)/(\cdot + 1) \overset{\mathcal{D}^d[0,T]}{\longrightarrow} \boldsymbol{W}\Big(\frac{\cdot}{\cdot + 1}\Big), \tag{4.42}$$
with $\{\boldsymbol{W}(t),\, t \ge 0\}$ denoting a standard Brownian motion.

By Lemma 4.1
$$\sup_{0 \le t \le T} \left| \frac{\sum_{i=m+1}^{m+\lfloor mt \rfloor} (r^0_{iM})^2}{\sum_{i=1}^m (r^0_{iM})^2} - t \right| \overset{P}{\longrightarrow} 0,$$
which implies that (4.42) holds also with $\widetilde{\boldsymbol{H}}(m, \lfloor m \cdot \rfloor)$ in place of $\boldsymbol{H}(m, \lfloor m \cdot \rfloor)$. The rest of the proof is exactly the same as the one of Theorem 3.1. □

Proof of Theorem 4.2. Recall that the model under the considered alternative has the form
$$r_{i,j} = \alpha^0_j + \beta^0_j \widetilde{r}_{iM} + (\alpha^1_j + \beta^1_j \widetilde{r}_{iM}) \delta_m I\{i > m + k^*\} + \varepsilon_{i,j}, \quad j = 1, \ldots, d, \quad i = 1, 2, \ldots,$$

where $\delta_m \to 0$. Notice that in this situation we have, for $k^* < k \le \lfloor mT \rfloor$,

$$\sum_{i=m+1}^{m+k} \widetilde{r}_{iM}\psi_j(\widehat{\varepsilon}_{i,j}) =$$
$$\sum_{i=m+1}^{m+k} \widetilde{r}_{iM}\psi_j\big(\varepsilon_{i,j} - (\widehat{\alpha}'_{jm} + \widehat{\beta}'_{jm} r^0_{iM})/\sqrt{m} + (\alpha^1_j + \beta^1_j \widetilde{r}_{iM})\delta_m I\{i > m + k^*\}\big) \quad (4.43)$$

with
$$\widehat{\alpha}'_{jm} = O_P(1), \quad \widehat{\beta}'_{jm} = O_P(1),$$
based on the training sample only. Moreover
$$\alpha^1_j + \beta^1_j \widetilde{r}_{iM} = \alpha^1_j - \beta^1_j(\overline{r}_{mM} - \mathrm{E}\, r_{iM}) + \beta^1_j r^0_{iM} =: \alpha^{11}_j + \beta^1_j r^0_{iM},$$
where α^{11}_j can be conditionally on $(r_{1M}, \dots, r_{mM})$ treated as a constant.

Similar to the proof of Theorem 4.1 we need to study
$$\widehat{L}_j(\widehat{\alpha}'_{jm}, \widehat{\beta}'_{jm}, m, k) := \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM}\psi_j(\widehat{\varepsilon}_{i,j}),$$

where the notation is based on the formula below. Towards this we define similarly to Lemma 4.4

$$L_j(a, b, m, k) := \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} r^0_{iM}\psi_j\big(\varepsilon_{i,j} - (a + b r^0_{iM})/\sqrt{m} + (\alpha^{11}_j + \beta^1_j r^0_{iM})\delta_m I\{i > m + k^*\}\big).$$

Along the lines of the proof of Lemma 4.4 we get that

$$\max_{1 \le k \le \lfloor mT \rfloor} \left(\frac{|\{L_j(a, b, m, k) - \mathrm{E}^*(L_j(a, b, m, k)\}|}{(k/m)^\gamma} \right) = O_P(m^{-\eta}),$$

uniformly in $|a| + |b| \le D$ for some $\eta > 0$ and arbitrary $D > 0$, where E^* denotes the conditional expectation given $r_{iM},\ i = m + 1, \dots, m(T + 1)$.

The conditional expectation of $L_j(a, b, m, k)$ has to be calculated carefully. Using again the notation $d_i = a + b r^0_{iM}$

$$\begin{aligned}
\mathrm{E}^*\, L_j(a, b, m, k) &= \frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} r^0_{iM}\, \mathrm{E}^*\, \psi_j(\varepsilon_{i,j} - d_i/\sqrt{m} + (\alpha^{11}_j + \beta^1_j r^0_{iM})\delta_m I\{i > m + k^*\}) = \\
&= -\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} r^0_{iM}\Big(\lambda'_j(0)(d_i/\sqrt{m} - (\alpha^{11}_j + \beta^1_j r^0_{iM})\delta_m I\{i > m + k^*\}) + \\
&\quad + O_P\Big(\big((a + b r^0_{iM})/\sqrt{m} + (\alpha^{11}_j + \beta^1_j r^0_{iM})\delta_m I\{i > m + k^*\}\big)^2\Big),
\end{aligned}$$

uniformly in $|a| + |b| \le D$ and in $1 \le k \le \lfloor mT \rfloor$.

(i) Then, in case of $\delta_m = m^{-1/2}$, an application of Lemma 4.1 results in

$$\begin{aligned}\mathrm{E}^* L_j(a,b,m,k) = &- b\lambda_j'(0)\frac{1}{m}\sum_{i=m+1}^{m+k}(r_{iM}^0)^2 + \beta_j^1\lambda_j'(0)\frac{1}{m}\sum_{i=m+k^*+1}^{m+k}(r_{iM}^0)^2\, I\{k>k^*\}\\ &- a\lambda_j'(0)\frac{1}{m}\sum_{i=m+1}^{m+k} r_{iM}^0 + \alpha_j^{11}\lambda_j'(0)\frac{1}{m}\sum_{i=m+k^*+1}^{m+k} r_{iM}^0\, I\{k>k^*\}\\ &+ O_P\big(\{a^2+b^2+(\alpha_j^{11})^2+(\beta_j^1)^2\}m^{-\xi}\big),\end{aligned}$$

uniformly in $|a|+|b|\le D$ and in $1\le k\le \lfloor mT\rfloor$, for some $\xi>0$.

Now, since $\widehat{\alpha}_{jm}' = O_P(1)$ and $\widehat{\beta}_{jm}' = O_P(1)$, we can plug-in the estimates $\widehat{\alpha}_{jm}'$ and $\widehat{\beta}_{jm}'$ for a and b in a standard way and similarly as in proof of Theorem 4.1 we get that the limit behavior of

$$\Big\{\widetilde{L}_j(\widehat{\alpha}_{jm}', \widehat{\beta}_{jm}', m, \lfloor mt\rfloor)\Big/(q_\gamma(\lfloor mt\rfloor/m),\ t\in[1/m,T]\Big\}$$

is the same as that of

$$\begin{aligned}&\Big\{\Big(\frac{1}{\sqrt{m}}\sum_{i=m+1}^{\lfloor mt\rfloor} r_{iM}^0\psi_j(\varepsilon_{i,j}) - \frac{\sum_{i=m+1}^{\lfloor mt\rfloor}(r_{iM}^0)^2}{\sum_{i=1}^{m}(r_{iM}^0)^2}\sum_{i=1}^{m} r_{iM}^0\psi_j(\varepsilon_{i,j}) + \lambda_j'(0)\beta_j^1\frac{1}{m}\sum_{i=m+k^*+1}^{m+\lfloor mt\rfloor}(r_{iM}^0)^2\Big)\Big/\\ &\qquad q_\gamma(\lfloor mt\rfloor/m),\ t\in[1/m,T]\Big\} =: \big\{H_j^*(m,\lfloor mt\rfloor)/q_\gamma(\lfloor mt\rfloor/m),\ t\in[1/m,T]\big\}\end{aligned}$$

for $0\le\gamma<1/2$. Note that

$$H_j^*(m,\lfloor mt\rfloor) = \widetilde{H}_j(m,\lfloor mt\rfloor) + \lambda_j'(0)\beta_j^1\frac{1}{m}\sum_{i=m+k^*+1}^{m+\lfloor mt\rfloor}(r_{iM}^0)^2,$$

where $\widetilde{H}_j$ is j-th coordinate of $\widetilde{\boldsymbol{H}}$ defined in (4.39). Thus we know the asymptotics of $\widetilde{H}_j$ from Theorem 4.1. Regarding the last term of H_j^* we use (4.17) of Lemma 4.1, i.e.

$$\sup_{0\le t\le T}\left|\frac{1}{m}\sum_{i=m+k^*+1}^{m+\lfloor mt\rfloor}(r_{iM}^0)^2 - (t-s)^+\operatorname{var}\{r_{0M}\}\right| \xrightarrow{P} 0,$$

which leads to the shift

$$\frac{(t-s)^+}{q_\gamma(t)}\operatorname{var}\{r_{0M}\}\boldsymbol{\Sigma}^{-1/2}(\lambda_1'(0)\beta_1^1,\ldots,\lambda_d'(0)\beta_d^1)^T$$

in an analogue of (3.17) in the proof of Theorem 3.2. Thus we can follow the lines there to see the form of the function $\boldsymbol{h}(t,s)$. Then the rest of the proof of (i) is clear.

(ii) In case of $\sqrt{m}\,|\delta_m|\to\infty$ and $\liminf_{m\to\infty}(\lfloor mT\rfloor - k^*)/m>0$, the term with β_j^1 in $\mathrm{E}^* L_j(a,b,m,mT)$ dominates. More precisely,

$$|\,\mathrm{E}^* L_j(a,b,m,mT)| = \sqrt{m}\,|\delta_m|\,\lambda_j'(0)|\,\beta_j^1|\,\frac{1}{m}\sum_{i=m+k^*+1}^{m+\lfloor mT\rfloor}(r_{iM}^0)^2\,\big(1+o_P(1)\big) \xrightarrow{P} \infty,$$

uniformly in $|a| + |b| \leq D$, if $\beta_j^1 \neq 0$. Therefore also the test statistic converges to ∞ in probability, which proves (ii) and completes the proof of Theorem 4.2. $\square$

Proof of Theorem 4.3. The proof follows lines of the one of Theorem 3.3. Only in addition we have to deal with the regressors $\widetilde{r}_{iM}$. Thus we need one extra step in estimation.

Define, for $k \geq 0$,

$$\widetilde{\mathbf{\Gamma}}_k = \frac{1}{m} \sum_{i=1}^{m-k} r_{iM}^0 r_{i+k,M}^0 \boldsymbol{\psi}(\boldsymbol{\varepsilon}_i) \boldsymbol{\psi}(\boldsymbol{\varepsilon}_{i+k})^T,$$

$$\overline{\mathbf{\Gamma}}_k = \frac{1}{m} \sum_{i=1}^{m-k} r_{iM}^0 r_{i+k,M}^0 \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_i) \boldsymbol{\psi}(\widehat{\boldsymbol{\varepsilon}}_{i+k})^T,$$

and, for $k < 0$, put $\widetilde{\mathbf{\Gamma}}_k = \widetilde{\mathbf{\Gamma}}_{-k}^T$ and $\overline{\mathbf{\Gamma}}_k = \overline{\mathbf{\Gamma}}_{-k}^T$, respectively.

Further let

$$\widetilde{\mathbf{\Sigma}}_m = \sum_{|k|<\Lambda_m} w(k/\Lambda_m) \widetilde{\mathbf{\Gamma}}_k, \quad \overline{\mathbf{\Sigma}}_m = \sum_{|k|<\Lambda_m} w(k/\Lambda_m) \overline{\mathbf{\Gamma}}_k.$$

Thus we get following decomposition

$$\widehat{\mathbf{\Sigma}}_m = (\widehat{\mathbf{\Sigma}}_m - \overline{\mathbf{\Sigma}}_m) + (\overline{\mathbf{\Sigma}}_m - \widetilde{\mathbf{\Sigma}}_m) + (\widetilde{\mathbf{\Sigma}}_m - \mathbf{\Sigma}) + \mathbf{\Sigma}.$$

According to Remark 4.3, the first term can be shown to be $o_p(1)$ similarly as in Theorem 4.1. The second and the third term can be treated as in Theorem 3.3 (using the techniques of this chapter) and thus shown to be $o_p(1)$ as well. $\square$

Chapter 5

Retrospective Analysis

In this chapter we explore the robust retrospective change-point analysis which is a crucial prerequisite to online monitoring since it enables to check the stability of the historical (training) data. This fact leads us to denote the sample size m as compared to standard n.

The topic of robust retrospective change-point analysis in a univariate location model with dependent data based on M-estimation was studied in Hušková and Marušiaková [2012] under slightly more restrictive assumptions than we have presented in Section 2.1. However their conclusions hold under Assumptions (A.1) - (A.5) as well. Their results are summarized in next section and are generalized to a multivariate location model in Section 5.2 and to CAPM in Section 5.3. Due to the similarity to the procedures of previous chapters, the presentation will be shorter.

5.1 Location Model

It is assumed that the one dimensional observations $Y_1, \ldots, Y_m$ follow the model:

$$Y_i = \mu_0 + \delta_m I\{i > k_m^*\} + e_i, \quad i = 1 \ldots, m, \tag{5.1}$$

where k_m^* $(\leq m)$, μ_0 and $\delta_m \neq 0$ are unknown parameters and $e_1, \ldots, e_m$ are random errors. Function $I\{A\}$ denotes again the indicator of the set A.

The hypothesis testing problem introduced in Section 1.2 thus transforms to

$$\widetilde{H}_0 : k_m^* = m \quad \text{versus} \quad \widetilde{H}_1 : k_m^* < m. \tag{5.2}$$

The test statistic is again based on the cumulative sums of ψ-residuals

$$\widehat{S}_{m,\psi}(k) = \frac{1}{m^{1/2}} \sum_{i=1}^{k} \psi(\widehat{e}_i), \quad k = 1, \ldots, m, \tag{5.3}$$

where $\psi(\widehat{e}_i)$ is defined by (2.5).

Hušková and Marušiaková [2012] chose to use max-type test and thus defined

$$T_{m,\psi} = \max_{1\le k\le m} \frac{|\widehat{S}_{m,\psi}(k)|}{\widehat{\sigma}_m(\psi)} \tag{5.4}$$

and the trimmed version

$$T_{m,\psi}(\eta) = \max_{\eta m\le k\le m(1-\eta)} \sqrt{\frac{m^2}{k(m-k)}} \frac{|\widehat{S}_{m,\psi}(k)|}{\widehat{\sigma}_m(\psi)}, \tag{5.5}$$

where $\eta \in (0, 1/2)$ and $\widehat{\sigma}_m(\psi)$ is a proper standardization.

The asymptotic behavior of the test statistics is summarized in following two theorems.

Theorem 5.1. *Let* $Y_1, \dots, Y_m$ *follow model* (5.1). *Let assumptions (A.1) – (A.4) from Section 2.1 be satisfied and let* $\widehat{\sigma}^2_m(\psi)$ *be a consistent estimator of* $\sigma^2(\psi)$ *defined by* (2.8). *Then under* $\widetilde{H}_0$, *as* $m \to \infty$,

$$T_{m,\psi} \xrightarrow{\mathcal{D}} \sup_{0<t<1} |B(t)| \tag{5.6}$$

and

$$T_{m,\psi}(\eta) \xrightarrow{\mathcal{D}} \sup_{\eta<t<1-\eta} \frac{|B(t)|}{\sqrt{t(1-t)}}, \tag{5.7}$$

where $\{B(t),\, t \in (0, 1)\}$ *is a Brownian bridge and* $0 < \eta < 1/2$.

Theorem 5.2. *Let* $Y_1, \dots, Y_m$ *follow model* (5.1) *with* $k^*_m = \lfloor m\tau \rfloor$ *for some* $\tau \in (0, 1)$. *Let assumptions (A.1) – (A.4) from Section 2.1 be satisfied and let* $\widehat{\sigma}^2_m(\psi)$ *be a consistent estimator of* $\sigma^2(\psi)$ *defined by* (2.8). *If* $\delta_m = \theta m^{-1/2}$ *then* $T_{m,\psi}$ *and* $T_{m,\psi}(\eta)$ *have the same limit distribution as*

$$\sup_{0<t<1} \left|B(t) - \theta\lambda'(0)\widetilde{p}(t, \tau)/\sigma(\psi)\right|$$

and

$$\sup_{\eta<t<1-\eta} \left\{ \left|B(t) - \theta\lambda'(0)\widetilde{p}(t, \tau)/\sigma(\psi)\right| / \sqrt{t(1-t)} \right\},$$

respectively, where $\{B(t),\, t \in (0, 1)\}$ *is a Brownian bridge,* $0 < \eta < 1/2$ *and*

$$\widetilde{p}(t, \tau) = \min(t, \tau)(1 - \max(t, \tau)),\ t \in (0, 1),\ \tau \in (0, 1). \tag{5.8}$$

For proofs please see Hušková and Marušiaková [2012] or the following remark for the main ideas.

Remark 5.1. (i) Proof of Theorem 5.1 is similar (even simpler) than the one of Theorem 2.1, one need to use just Corollary 2.2 instead of Lemmas 2.5 and 2.6 and realize that $\{W(t) - tW(1),\, t \in [0, 1]\}$ is Brownian bridge for $\{W(t),\, t \in [0, 1]\}$ being Wiener process.

(ii) Under the alternative a situation is a bit more complicated compared to the online monitoring. The problem is that the estimate $\widehat{\mu}_m(\psi)$ is influenced by the amount of change. Therefore instead of the asymptotic representation (2.29), i.e.

$$\sqrt{m}(\widehat{\mu}_m(\psi) - \mu_0) = \frac{1}{\lambda'(0)\sqrt{m}} \sum_{i=1}^{m} \psi(e_i) + o_P(1),$$

we have for $\delta_m = m^{-1/2}$

$$\sqrt{m}(\widehat{\mu}_m(\psi) - \mu_0) = \frac{1}{\lambda'(0)\sqrt{m}} \sum_{i=1}^{m} \psi(e_i) + (1-\tau) + o_P(1). \tag{5.9}$$

Otherwise the proof goes similarly to the one of Theorem 2.2 and thus give us the form of the shift function $\widetilde{p}(t,\tau)$.

(iii) As an estimate of the long run variance $\sigma^2(\psi)$, we can take $\widehat{\sigma}^2_m(\psi)$ defined in (2.42). Under the additional assumption (A.5) it is consistent not only under the null hypothesis, but also under the local alternatives, which can be checked going through the proof of Theorem 2.3. Estimators of $\sigma^2(\psi)$ can be modified to take into account a possible change in order to improve the power, confer Hušková and Kirch [2010]. The idea is that we calculate the residuals differently before and after the supposed change, which is estimated as

$$\widehat{k}^*_m = \arg\max_{1\le k\le m} \left|\widehat{S}_{m,\psi}(k)\right|$$

and analogously for $T_{m,\psi}(\eta)$.

(iv) The boundary function of (5.5) is derived from the standard deviation of limit distribution as can be seen in (5.7) and thus the procedure should have better properties. However the trimming is needed at both ends. Another possibility is to use function $(t(1-t))^{\xi}$ for some $\xi \in (0, 1/2)$ cf. Antoch et al. [2002] for example. For simplicity we restrict ourselves to analogue of detector (5.4) in further generalizations.

5.2 Multivariate Location Model

Results from previous section are generalized here to the multivariate model, which allow to test stability of training data for online monitoring studied in Chapter 3.

Thus the d-dimensional observations are assumed to follow model

$$\boldsymbol{Y}_i = \boldsymbol{\mu}_i + \boldsymbol{\delta}_m I\{i > k^*_m\} + \boldsymbol{e}_i,\ i = 1, \dots, m, \tag{5.10}$$

where again $\boldsymbol{\delta}_m \neq \boldsymbol{0}$. Hypotheses of (5.2) remain the same, referring now to the model (5.10).

The test statistic is again a quadratic form of ψ-residuals

$$T^*_{m,\boldsymbol{\psi}} = \max_{1\le k\le m} \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)\right)^T \widehat{\boldsymbol{\Sigma}}^{-1}_m \left(\frac{1}{\sqrt{m}} \sum_{i=1}^{k} \boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)\right),$$

where M-residual $\boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)$ is defined in (3.4) and $\widehat{\boldsymbol{\Sigma}}_m$ is an estimator of the asymptotic variance matrix $\boldsymbol{\Sigma}$ defined in (3.6).

Under the assumptions (A*.1)–(A*.4) from Chapter 3 we have the following asymptotic behavior of the test statistic.

Theorem 5.3. *Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ follow model* (5.10)*. Let assumptions (A*.1) – (A*.4) from Section 3.1 be satisfied and let $\widehat{\boldsymbol{\Sigma}}_m$ be a consistent estimator of $\boldsymbol{\Sigma}$. Then under $\widetilde{H}_0$, as $m \to \infty$,*

$$T^*_{m,\boldsymbol{\psi}} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \sum_{j=1}^{d} B_j^2(t), \tag{5.11}$$

where $\{B_j(t),\, t \in (0,1)\},\ j = 1, \ldots, d$ are independent Brownian bridges.

Theorem 5.4. *Let $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_m$ follow model* (5.1) *with $\boldsymbol{\delta}_m = \boldsymbol{\theta} m^{-1/2}$, $\boldsymbol{\theta} \neq \mathbf{0}$ and $k_m^* = \lfloor m\tau \rfloor$ for some $\tau \in (0,1)$. (A*.1) – (A*.4) from Section 3.1 be satisfied and let $\widehat{\boldsymbol{\Sigma}}_m$ be a consistent estimator of $\boldsymbol{\Sigma}$. Then*

$$T^*_{m,\boldsymbol{\psi}} \xrightarrow{\mathcal{D}} \sup_{0<t<1} \sum_{j=1}^{d} (B_j(t) - s_j(t,\tau))^2, \tag{5.12}$$

where $\{B_j(t),\, t \in (0,1)\},\ j = 1, \ldots, d$ are independent Brownian bridges,

$$\boldsymbol{s}(t,\tau) = \widetilde{p}(t,\tau)\boldsymbol{\Sigma}^{-1/2}\big(\lambda_1'(0)\theta_1, \ldots, \lambda_d'(0)\theta_d\big)^T,$$

with $\widetilde{p}(t,\tau)$ defined in (5.8) *and $\boldsymbol{\Sigma}^{1/2}$ is square root matrix of $\boldsymbol{\Sigma}$.*

Proofs of both previous theorems are a straight-forward combinations of Theorems 5.1 and 3.1, respectively 5.2 and 3.2. As an estimate of $\boldsymbol{\Sigma}$ we can use $\widehat{\boldsymbol{\Sigma}}_m$ from (3.9), which under additional assumptions of Theorem 3.3 is proven to be consistent under the null hypothesis $\widetilde{H}_0$. This holds similarly under the local alternative of Theorem 5.4.

The limit distribution of Theorem 5.3 can be used to determine critical values such that the test has required level. For more details see Chapter 7.

5.3 CAPM

Similarly as in previous section we design here a retrospective procedure for testing stability of portfolio betas in CAPM by analogy with the monitoring procedure.

Instead of model (4.3) we consider

$$r_{i,j} = \alpha_j^0 + \beta_j^0 \widetilde{r}_{iM} + (\alpha_j^1 + \beta_j^1 \widetilde{r}_{iM})\delta_m I\{i > k_m^*\} + \varepsilon_{i,j}, \quad j = 1, \ldots, d, \quad i = 1, \ldots, m \tag{5.13}$$

with $\boldsymbol{\beta}^1 \neq \mathbf{0}$, where the notation comes from Chapter 4. For this model hypotheses (5.2) need to be tested. Towards this we use test statistic

$$T^C_{m,\boldsymbol{\psi}} = \max_{1 \le k \le m} \Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM}\boldsymbol{\psi}(\widehat{\varepsilon}_i)\Big)^T \widehat{\boldsymbol{\Sigma}}_m^{-1} \Big(\frac{1}{\sqrt{m}} \sum_{i=m+1}^{m+k} \widetilde{r}_{iM}\boldsymbol{\psi}(\widehat{\varepsilon}_i)\Big), \tag{5.14}$$

where M-residual $\boldsymbol{\psi}(\widehat{\boldsymbol{e}}_i)$ is defined in (4.6) and $\widehat{\boldsymbol{\Sigma}}_m$ is an estimator of the asymptotic variance matrix $\boldsymbol{\Sigma}$ defined in (4.9).

Limit distributions of this detector are analogous to those of $T^*_{m,\boldsymbol{\psi}}$ thus we present them only as a remark.

Remark 5.2. (i) With the assumptions (A*.1) – (A*.4), (B) as considered in Chapter 4, $\widehat{\boldsymbol{\Sigma}}_m$ being a consistent estimator of $\boldsymbol{\Sigma}$, the limit distribution of the detector $T^C_{m,\boldsymbol{\psi}}$ under the null hypothesis is described by (5.11).

(ii) With the assumptions (A*.1) – (A*.4), (B) as considered in Chapter 4, $\widehat{\boldsymbol{\Sigma}}_m$ being a consistent estimator of $\boldsymbol{\Sigma}$, the limit distribution of the detector $T^C_{m,\boldsymbol{\psi}}$ under the contiguous alternative with $k^*_m = \lfloor m\tau \rfloor$ for some $\tau \in (0,1)$ is described by (5.12), where the shift function $\boldsymbol{s}(t,\tau)$ has now form

$$\boldsymbol{s}(t,\tau) = \widetilde{p}(t,\tau)\,\mathrm{var}\{r_{0M}\}\boldsymbol{\Sigma}^{-1/2}(\lambda_1'(0)\beta_1^1,\ldots,\lambda_d'(0)\beta_d^1)^T.$$

Chapter 6

Computational Aspects

Until now the behavior of monitoring procedures has beeb studied asymptotically – as we have considered the length of training data m tending to infinity. However, for real data problems this is always limited and thus it is important to investigate some finite sample properties for the proposed procedures as well.

In this chapter we will focus on results of simulation studies conducted for each procedure introduced, as well as a real data example. All programming work was provided using the statistical software R 2.15.1.

6.1 Location Model

We start with the monitoring procedure for a change in location model described in Chapter 2.

Let us first list different options we have used for the simulation study in order to be able to describe the influence of particular parameters:

- ψ-functions corresponding to the L_2, L_1 and Huber estimators as described in Section 1.4 (in the Huber case with constant $K = 1.345$ which is the default value in the R-package);
- length of historical period $m = 80, 200, 400$;
- multiple T for monitoring period $T = 10$;
- tuning constant $\gamma = 0, 0.25, 0.45$;
- type of dependence: AR(1) or MA(1) with coefficient $\rho = -0.5, -0.25, 0, 0.25, 0.5, 0.75$;
- distribution of random errors (innovations): $N(0, 1), t_3, t_1$, Laplace, contaminated normal ones;
- $\mu_0 = 0$, $\delta = 1/2, 1, 2$;
- $k^* = 10, 80, 200, 400$.

One important parameter choice is missing in the previous list. It is the estimator $\widehat{\sigma}^2_m(\psi)$ of the long-run variance (LRV), square root of which is used for normalizing the test statistic (2.7). For dependent data it is crucial to have a good estimator, since it generally influences significantly the whole test procedure (see e.g. Kirch [2006]).

In Section 2.5 we proposed a class of kernel estimators, namely Bartlett kernel and the flat-top kernel variance estimators (shortly the Bartlett and FLT estimators). Nevertheless the question of a proper choice of the number of lags Λ_m (i.e. bandwidth of the kernel) remains to be addressed.

Since the choice of the LRV estimator is of great importance for our procedures, we focus on this in the following subsection.

6.1.1 Long-run Variance Estimators

The Bartlett estimator used to be a traditional estimator of the LRV. Results of Antoch et al. [1997] suggest using $\Lambda_m = m/10$. This setting was also adopted by Hušková and Marušiaková [2012], which dealt with retrospective change-point analysis using M-estimates, as described earlier.

Recently FLT estimator became more popular. This was also supported by Politis [2003] who has shown that the bandwidth can be chosen adaptively in such a way that it captures the theoretically optimal rates very well. This is not generally true if we use this selection procedure with the Bartlett estimator.

We present the selection procedure as summarized in Hušková and Kirch [2010], where it was used in large simulation study which compared the performance of respective estimators in L_2 setting.

Adaptive Selection of the bandwidth Λ_m:
Let ℓ be the smallest positive integer such that

$$\left|\widehat{R}_m(\ell + k, \psi)/\widehat{R}_m(0, \psi)\right| < c\sqrt{\log_{10} m/m} \text{ for } k = 1, \dots, K_m,$$

where $\widehat{R}_m(k, \psi)$ is defined in (2.43), $c > 0$ is a fixed constant, $\log_{10}$ is the logarithm to base 10 and K_m is a positive, nondecreasing integer valued function of m such that $K_m = o(\log m)$. Then choose $\Lambda_m = 2\ell$.

The choice of the parameters c and K_m is left to practitioners. However, Politis [2003] suggests to use $c = 2$ and $K_m = 5$ for the usual sample sizes approximately between 100 and 1000. Results of Hušková and Kirch [2010] confirm that these values indeed give good results. But different choices give very similar results showing that the procedure is rather robust with respect to the choice of c and K_m. For practical applications they recommend the choice $c = 1.4$ and $K_m = 3$, since it is a good compromise between $c = 2, K_m = 5$ which works best under positive correlation and $c = 1, K_m = 1$ which works best under negative correlation. Moreover, they showed that the adaptive FLT estimator has better performance than the Bartlett or FLT estimator with a variety of choices for L_m expressed as fractions of sample size m.

It is also necessary to mention that, unlike the Bartlett estimator, FLT one can be negative, especially if $\sigma^2(\psi)$ is small. On one side this can be an advantage since in this case the Bartlett estimator usually overestimates $\sigma^2(\psi)$ (while for large values of the LRV both estimators usually underestimate it). FLT estimator manages to capture small $\sigma^2(\psi)$ quite well and large value still better than the Bartlett one. Nevertheless it is not desirable for many applications to have the estimate too close to zero (or even negative). For example in testing problems this estimate forms a part of a denominator of the test statistic and thus the statistic becomes quite large leading to rejecting the null hypothesis wrongly. This is why the FLT estimator is usually modified as follows

$$\widetilde{\sigma}_m^2(\psi) = \max(\widehat{\sigma}_m^2(\psi), 1/\log^2 m)$$

in order to assure that it stays bounded away from zero. We also use this modification.

In spirit of Hušková and Kirch [2010] we extend the study of the Bartlett, FLT and adaptive FLT estimators to M-estimate setting, which deepens results published in Hušková and Marušiaková [2012]. Moreover, we do not focus only on the LRV estimators themselves, but we consider their influence on the testing procedure as well.

We consider five different parameters for each estimator, namely

- $\Lambda_m = 4, 8, 10, 20, 40$ for the Bartlett and FLT ones, and
- $c = 1, 1.2, 1.4, 1.7, 2$, with $K = 1, \ldots, 5$ for adaptive FLT one (denoted FLT adapt).

We believe that a proper chart conveys much more information than plentiful of numbers in a table, therefore we try to present the results mainly in form of figures. As the number of parameters of the simulation study is quite large, we present here the typical representative of each figure only. The rest of figures can be found attached in a supplementary file (reffered to as Attachment further) and we comment them in the text when it is useful.

The influence of different estimators (and parameters) on the test procedure under the null hypothesis is illustrated by the so-called *size-power curve* plots (SPC) introduced by Kirch [2006]. SPC plot shows the empirical size and power (i.e. the empirical α-errors resp. 1-(β-errors)) on the y-axis for the chosen nominal level on the x-axis. So, the graph for the null hypothesis should be close to the diagonal (which is given by the thick dash-and-dot line) and for the alternatives it should be as steep as possible. The nominal levels correspond to asymptotic distribution of the test statistic, i.e. the critical values for $\alpha = 1\%, 2\%, \ldots, 20\%$ are obtained in the same way as in Chapter 7.

Figure 6.1 presents SPC plots for a monitoring procedure using Huber ψ function (shortly Huber procedure) because it can be viewed as a compromise between L_1 and L_2 procedures, and the tuning parameter $\gamma = 0.25$ as it is generally the best choice - more on this topic later. Plots are arranged in a matrix form, where in columns there are the three different LRV estimators (the Bartlett, FLT and FLT adapt), whereas row panels represent various degrees of dependence of random errors - they form AR(1) sequence (of $N(0, 1)$ innovations) with coefficient ρ listed at LHS of each panel. Each chart is SPC under the null hypothesis (i.e. the empirical level against the nominal one),

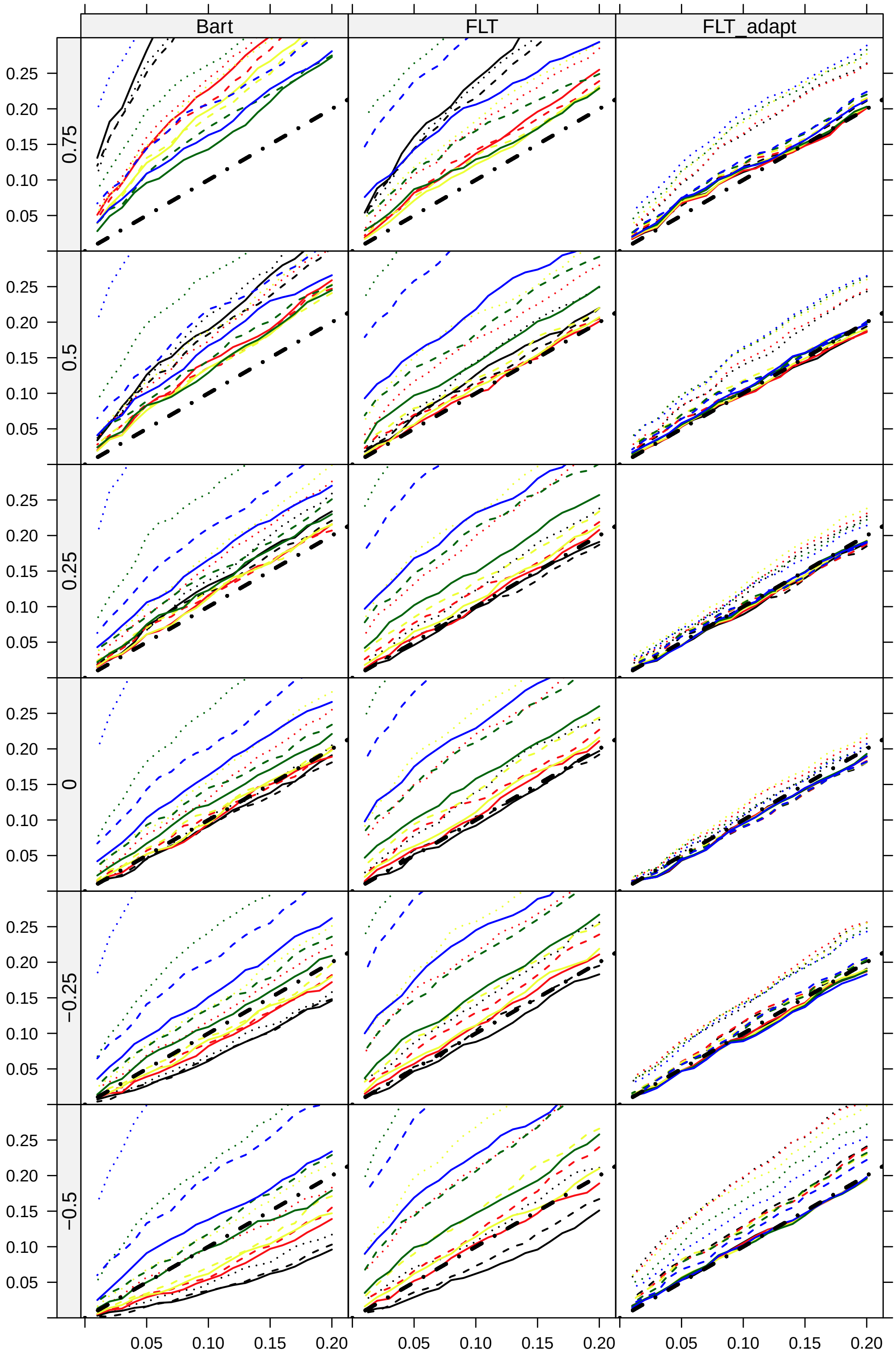

Figure 6.1: SPC for different LRV estimators, Huber procedure under H_0, $\gamma = 0.25$, errors being AR(1) of $N(0, 1)$ innovations with ρ indicated at LHS of each panel. Λ_m : 4 - black, 8 - red, 10- yellow, 20 - green, 40 - blue and accordingly for FLT adapt; m : 80 - dotted, 200 - dashed, 400 - solid.

where different line types (dotted, dashed and solid) correspond to the three chosen lengths of historical period ($m = 80, 200, 400$) - this coding is used in all figures if not specified otherwise. Black, red, yellow, green and blue colors correspond to the chosen parameter of the LRV estimators from the list above. The empirical sizes are based on 1000 repetitions and we use the same sequence of innovations for each setting in order to allow easier comparison.

We can see that the adaptive FLT estimator gives the best results among all the three estimators considered. From the moderate size historical period of $m = 200$, the empirical sizes follow the ideal diagonal line almost perfectly for all degrees of dependence. For short training data $m = 80$ and high degree of dependence (both positive and negative) the nominal level is not kept. Only here we can also see a difference caused by different parameters used. It is in line with the above mentioned choices of constants c and K regarding the sign of dependence.

The FLT estimator with a fixed number of lags gives the best results for $\Lambda_m = 4$ (except for the very strong dependence, where this is clearly insufficient). When the training period is long enough (i.e. $m = 400$), choices $\Lambda_m = 8,\ 10$ work even better - they are not so conservative for a negative dependence and perform better for the strong positive one.

Performance of the Bartlett estimator with recommended Λ_m is represented by solid blue, dashed green and dotted red lines respectively. We can see that the nominal level is not kept by far even under the independence. But it is hard to recommend any other rule regarding the number of lags, as the nominal level is not kept for stronger dependence by any combination of m and Λ_m. For small dependence the choices up to $\Lambda_m = 10$ seem reasonable.

Results for other types of procedures are similar (although the differences for L_1 procedures are smaller). Also the other types of dependence (MA(1) sequences) and tuning constant $\gamma = 0,\ 0.45$ do not play any significant role.

Now we focus on the estimators of the LRV themselves. Figure 6.2 shows their estimated density for the Huber procedure (using the standard R routine). The vertical line indicates the true value of LRV being estimated. For L_2 procedure this is easy to find since the LRV of AR(1) sequence is

$$\sigma^2(L_2) = \frac{s^2}{(1-\rho)^2},$$

where s^2 is the variance of innovations. For other procedures we adopt the approach of Hušková and Marušiaková [2012], where the theoretical values $\sigma^2(\psi)$ were approximated via simulations. Since we need the LRV for different values of ρ, we repeated the simulation. For each parameter ρ we generated 10 000 of $AR(1)$ error sequences $\{e_i\}_i$ and for each such sequence we calculated

$$\frac{1}{N}\sum_{i=1}^{N}\psi^2(e_i) + 2\sum_{k=1}^{L}\frac{1}{N}\sum_{i=1}^{N-k}\psi(e_i)\psi(e_{i+k}), \tag{6.1}$$

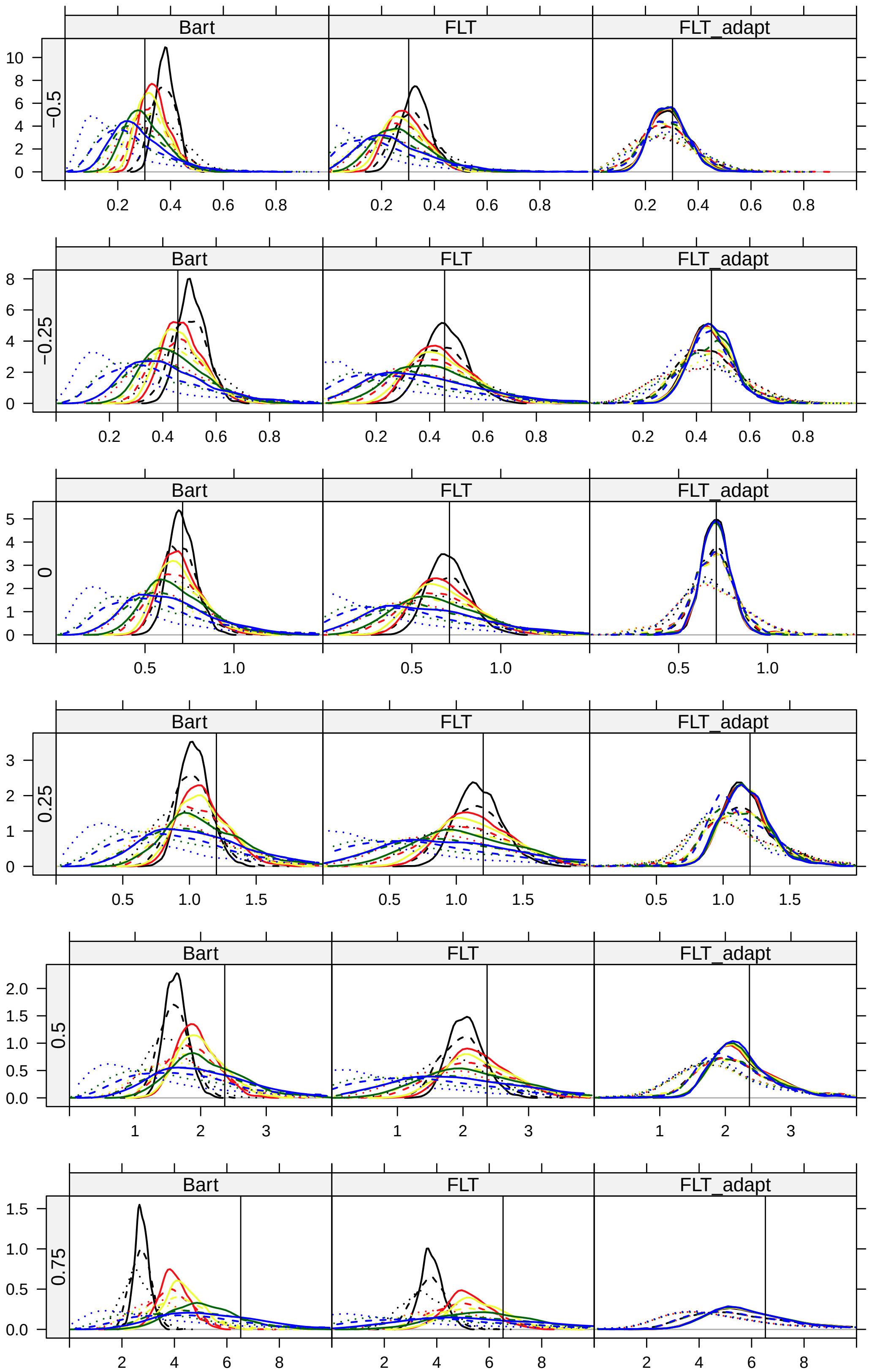

Figure 6.2: Kernel density estimates of the LRV for different estimators, Huber procedure, errors being AR(1) of $N(0,1)$ innovations with ρ indicated at LHS of each panel. Vertical black line represent true LRV. (Note different scale of each chart.)
Λ_m : 4 - black, 8 - red, 10- yellow, 20 - green, 40 - blue and accordingly for FLT adapt; m : 80 - dotted, 200 - dashed, 400 - solid.

where $N = 10\,000$ and number of lags $L = 100$. The mean of these quantities serves as an approximation of the theoretical LRV. On the L_2 procedure we verified that the approximation is really very precise.

Lets get back to Figure 6.2. In columns we have again the three variance estimators and the row panels represent various degrees of dependence in AR(1) sequence of standard normal innovations. The line types and colors coding is the same as in Figure 6.1. Please notice also different scale of each panel.

Again we can see that FLT adapt estimates the theoretical LRV in the best way. There is not much difference between the parameters considered and the estimates naturally gets better with increasing data available. (Recall that the LRV is estimated from the historical period only and thus the sample sizes are $80, 200, 400$.) For a strong positive dependence FLT adapt underestimates the variance, however it is still much better than the Bartlett estimator or FLT with fixed number of lags. FLT with $\Lambda_m = 4$ works well for negative and mild positive dependence, $\Lambda_m = 8, 10$ is fine for moderate one. The choice of a proper bandwidth for the Bartlett estimator is more difficult. Similar conclusions hold for L_1 and L_2 procedures as well, figures of which can be found attached.

It is also interesting to know what number of lags the adaptive estimator usually selects. By design it can be only a positive even number. Figure 6.3 shows histograms of number of lags Λ_m selected by the adaptive FLT estimator for Huber procedure. We used the recommended constants $c = 1.4$, $K = 3$, figures for the others are attached. Columns of the figure represent various sample sizes, rows again different degree of dependence expressed by the ρ coefficient of AR(1) sequence with standard normal innovations. We see that for independent observations the adaptive procedure selects generally the smallest number of lags possible. For the dependent data the number of lags selected naturally grows with increasing dependence (both positive and negative) and also with increasing sample size. However it usually does not exceed 8 even for $m = 400$ (except for the case of extreme dependence).

6.1.2 Boundary Function

A boundary function was in general terms introduced in (1.8) and for the location model procedures was specified in (2.9). The tuning parameter $\gamma \in [0, 1/2)$ was also introduced there and now we focus on its influence.

For better perspective we integrate the $1/\sqrt{m}$ from (2.6) into a new boundary function

$$g(m, k, \gamma) = \sqrt{m} q_\gamma(k/m) = \sqrt{m} \left(\frac{m+k}{m}\right) \left(\frac{k}{m+k}\right)^\gamma, \quad k = 1, 2, \ldots, \tag{6.2}$$

thus leaving the detector to be only the (absolute value of) cumulative sum of ψ-residuals (normalized by a square root of the LRV).

Figure 6.4 shows the boundary functions $g(m, k, \gamma)$ multiplied by the proper asymptotical critical values $c_{10}(0.05, \gamma)$ for $m = 200$, $\gamma = 0, 0.25, 0.45$ and $k = 1, \ldots, 1000$. Left

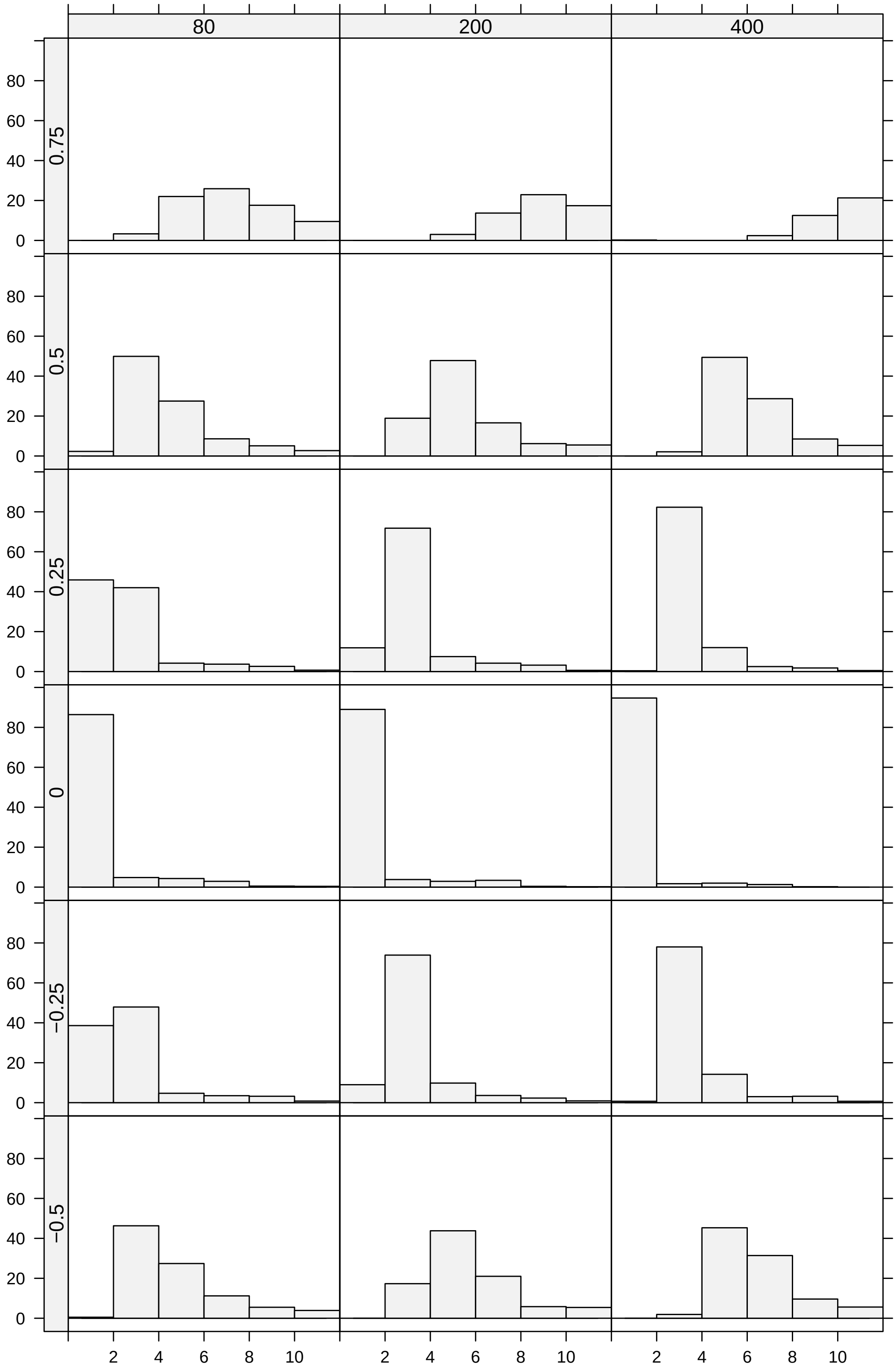

Figure 6.3: Histograms of number of lags Λ_m selected by the adaptive FLT estimator for $c = 1.4$, $K = 3$. Length of training period m in columns, dependence coefficient ρ in rows.

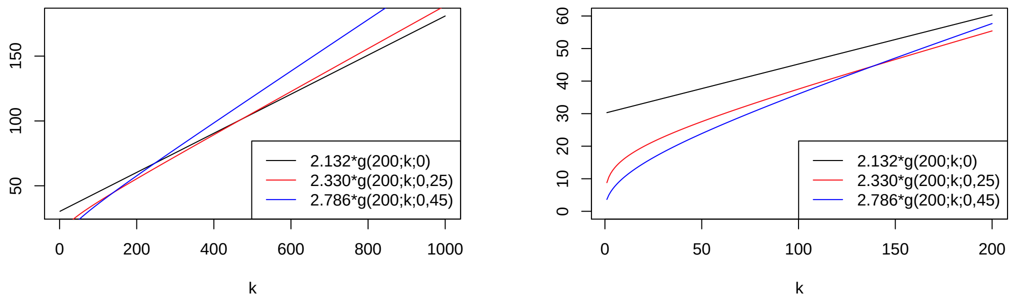

Figure 6.4: Boundary function $g(m, k, \gamma)$ multiplied by the corresponding critical values $c_{10}(0.05, \gamma)$ for the choices $\gamma = 0.00, 0.25, 0.45$, $m = 200$.

chart presents the overall picture whereas the right one is zoomed on the beginning of the monitoring period. We can see that for a large value of k (at least $2m$) the smallest boundary function is generated by $\gamma = 0$ and thus it is the most sensitive to detect the change far in the monitoring period. On the other hand, for small k (smaller than $m/2$ approximately) it is the best to use $\gamma = 0.45$. For the period between $m/2$ and $2m$ the most appropriate is to use $\gamma = 0.25$ and we can see that this choice is not much worse than the previous recommendations also outside this interval. Thus if we have some prior belief about the change-point, we can tune the procedure to suit it better. Otherwise the middle choice of $\gamma = 0.25$ works reasonably well for all situations.

Location of the change point k^* and the choice of parameter γ also determine how the procedure reacts when we have longer training period available. Clearly then we have more precise estimates of the mean and the LRV and the asymptotic approximations work better, but we focus here only on the influence of the boundary function. Each panel of Figure 6.5 shows for different $\gamma = 0, 0.25, 0.45$ the boundary functions $g(m, k, \gamma)$ with different lengths of the training period (critical value is asymptotical i.e. the same for all m and thus it was not included). We see that for $\gamma = 0.45$, the effect of prolongation of the training period is positive throughout the whole monitoring period. For $\gamma = 0.25$ there is a short period on the beginning of the monitoring, where the increase in m leads to a higher boundary function and thus to prolongation of the detection delay for a change-point located short after the start of the monitoring. However, the difference between the boundary functions (at the beginning) is not so significant. For a change-point located at time at least $m_0/2$, the prolongation of m_0 brings improvement also with regards to the boundary function. For $\gamma = 0$ the improvement is visible from $k^* = 3/2m_0$ and it thus confirms the above mentioned conclusion, that this value of the tuning parameter should be used only when expecting a change late enough.

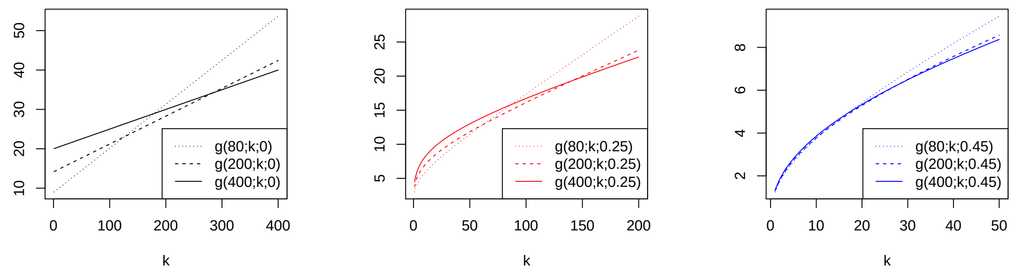

Figure 6.5: Boundary function $g(m, k, \gamma)$ for $m = 80, 200, 400$ and $\gamma = 0.00, 0.25, 0.45$ in respective charts.

6.1.3 Monitoring Procedures with Adaptive LRV Estimator

Now we get back to the monitoring procedures in the location model, where based on Section 6.1.1 we chose the adaptive FLT estimator with $c = 1.4$, $K = 3$ for an estimation of the LRV.

Null hypothesis
To summarize the behavior of the test procedure under the null hypothesis we present once again the matrix of SPC plots in Figure 6.6. It shows side by side all considered procedures (i.e. L_1, Huber and L_2) and various degrees of dependence (in AR(1) sequence of $N(0, 1)$ errors). We can see that the difference between the procedures is not significant and for a moderate length of the training period they perform very well also for the very strong dependence. Only for $m = 80$ and strongly dependent errors the required level is not kept. Regarding the tuning constant, the performance for $\gamma = 0.45$ is slightly worse than for the other two choices, however the difference is not significant.

Now we focus on the robustness of the procedures. Towards this we present Figure 6.7 of SPC for independent errors having various heavy tailed distributions. We consider Laplace, Student t_3 and t_1 (Cauchy) and also two contaminated standard normal ones. The first, denoted Mix1, is contaminated with 2% $N(0, 400)$ while the second Mix2 is contaminated with 5% $N(0, 100)$. The L_1 and Huber procedures keep the required level very well for all the distributions however the L_2 fails to do so if the contamination is strong or the data are really heavy tailed. Note that for Cauchy distribution, the SPC is even for $m = 400$ out of the range of the plot (about 50% of cases are rejected already for 1% nominal level). Thus we see that L_2 procedure performs unsatisfactory in a presence of outliers.

Alternative hypothesis
Now we concentrate on the alternative hypothesis where a change in mean of amount δ happens at point k^* in the monitoring period.

Since the procedures are designed to have power one (asymptotically), the SPC under the alternative hypothesis does not convey much information as the power is one even

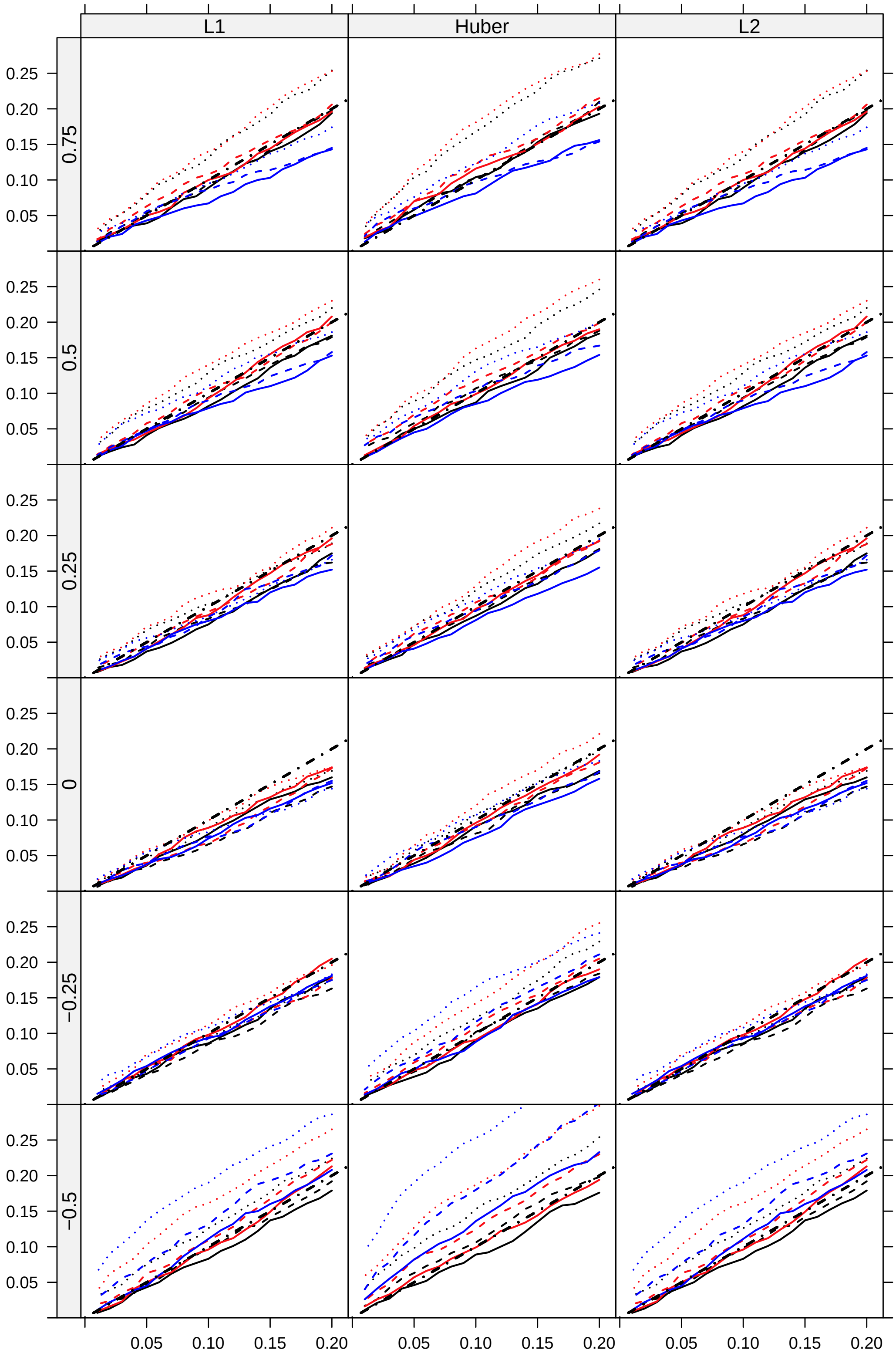

Figure 6.6: SPC for different procedures with adaptive FLT estimator under H_0, errors being AR(1) of $N(0,1)$ innovations with ρ indicated at LHS of each panel.
γ : 0 - black, 0.25 - red, 0.45 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

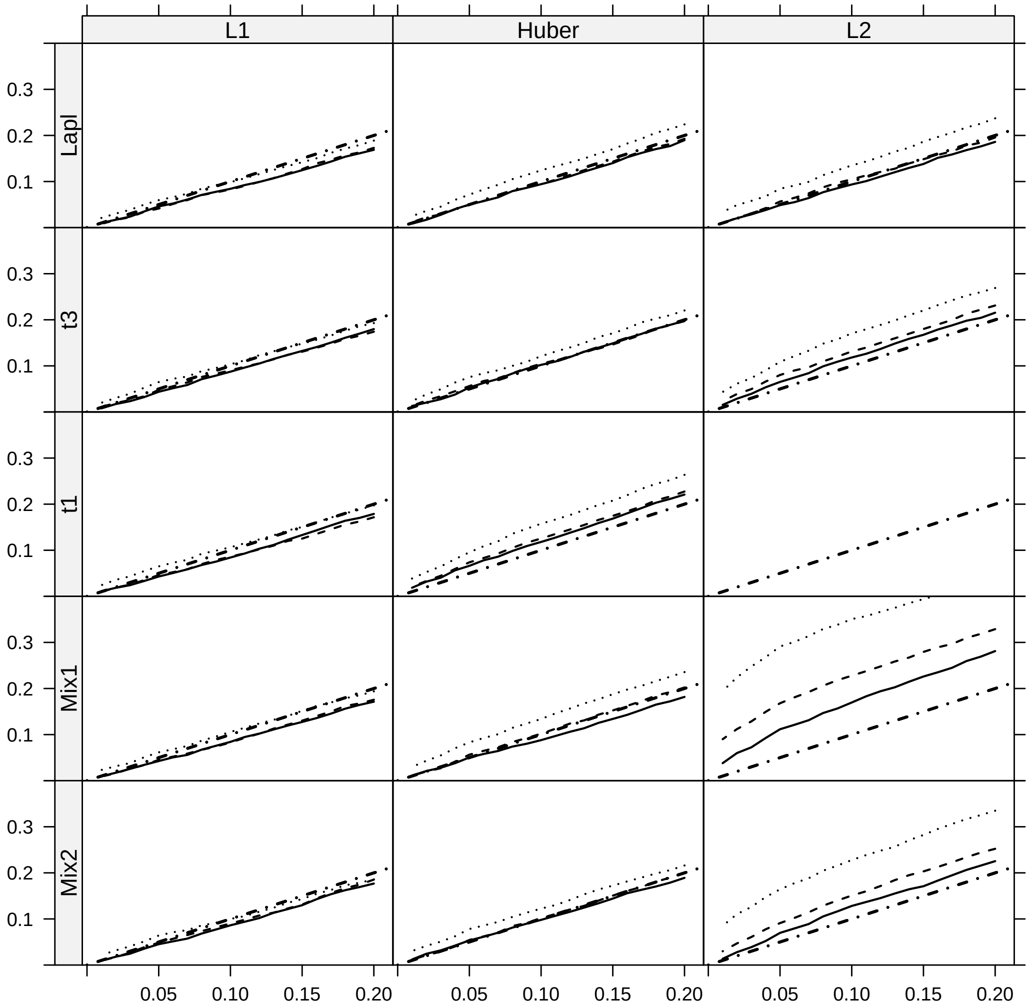

Figure 6.7: SPC for different procedures with adaptive FLT estimator under H_0, different distribution of errors in each panel, $\gamma = 0.25$.
m : 80 - dotted, 200 - dashed, 400 - solid.

for the smallest nominal sizes in the vast majority of cases considered. More informative is to measure power (on y-axis again) with respect to detection delay $k - k^*$ (on x-axis) for chosen nominal level (we chose as usual $\alpha = 5\%$). We can call this chart DPC (*Delay-Power Curve*) by analogy of SPC. Procedures gain power quite quickly with the number of out-of-control observations increasing, thus it is enough to consider a maximum delay of 200 even in the most adverse cases.

Second way to illustrate the performance under the alternative hypothesis is via the stopping time, sometimes also called the run length (RL). Usually the average run length (ARL) is shown as a single number summary. As has been said already, graphical representation delivers more information and thus we introduce the graph of *density of run length* (DRL for short). From the charts we can see for example that the distribution of the RL is not symmetric and thus it is sometimes argued that the median describes the distribution better that the classical mean.

Firstly we concentrate on the influence of the tuning parameter γ. Figure 6.8 shows DRL for all procedures and all change-points k^* considered. Each chart represents DRL for different m and γ. For a very early change ($k^* = 10$) the best performance is for $\gamma = 0.45$. For an early change (e.g. $k^* = 80$ for $m = 400$) $\gamma = 0.45$ still slightly outperforms 0.25, the situation is reversed for a moderate change-point. For a later change (e.g. $k^* = 400$ for $m = 200$) $\gamma = 0$ and 0.25 clearly outperforms 0.45 and for a very late change $\gamma = 0$ is the most appropriate. Thus the theoretical conclusions about choices of the tuning constant γ from Section 6.1.2 are confirmed and we see that $\gamma = 0.25$ performs almost as good as the best possible choice in all circumstances considered. We can also see that the difference between the procedures is not large and that the prolongation of the training period brings quite an improvement, especially from $m = 80$ to 200 (with the exceptions described in the second paragraph of Section 6.1.2). The same conclusions follow from Figure 6.9 where the DPC are displayed in the same layout as described above.

Figure 6.10 shows DRL in different arrangement, thus allowing for better comparison of the procedures and also sizes of the change. We used the middle value of $\gamma = 0.25$. Columns represent now the length of the training period m, rows again change-points k^*. Performance of L_2 and Huber procedure is quite similar even for a large change $\delta = 2$, whereas L_1 is inferior. In case of smaller change $\delta = 1$ the difference is not so large. A comparison of powers is done in Figure 6.11, where we also included a small change of 0.5. For that one, L_1 procedure is better than L_2 procedure. All the observations regarding the amount of change are in line with the theoretical expectations resulting from the boundedness of the residuals (for L_1 procedure and small change, this is a boundedness from below, as the ψ-residual is always at least 1 in absolute value).

Figure 6.12 illustrates the robustness aspect of the procedures. It shows DRL in the left column and DPC in the right one for various heavy-tailed/contaminated distributions. We see that the L_2 procedure clearly fails in these situations, whereas L_1 and Huber procedures still perform well even for Cauchy distribution.

In the same setting Figure 6.13 illustrates influence of the dependence. Row panels represent various degrees of dependence in AR(1) sequence of $N(0, 1)$ innovations. We see the procedures have the best performance for negatively correlated errors and with growing positive dependence the performance decreases, however having still 80% power at delay 200. These results are in line with those of Hušková and Marušiaková [2012].

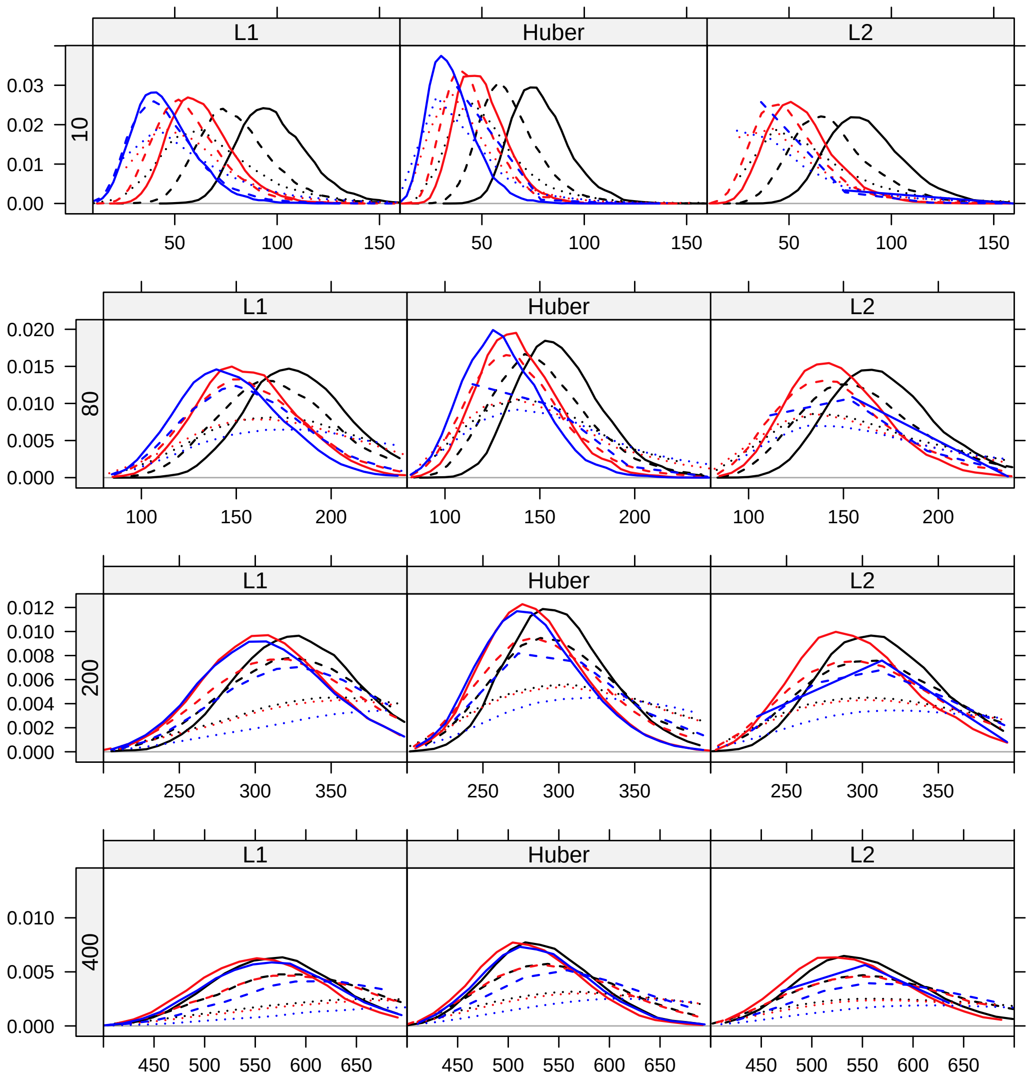

Figure 6.8: DRL for different procedures, change point k^* indicated at left of each panel, $\delta = 1$, independent Laplace errors.
γ : 0 - black, 0.25 - red, 0.45 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

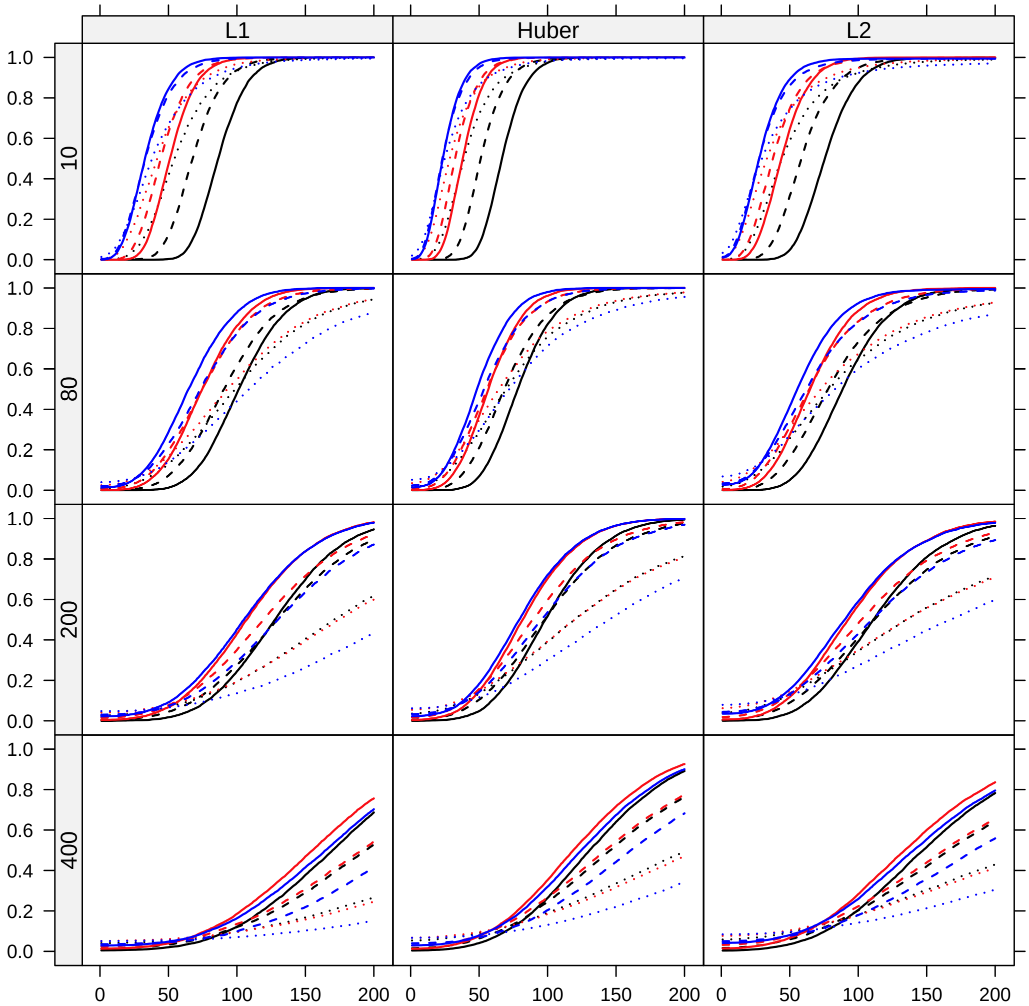

Figure 6.9: DPC for different procedures, change-point k^* indicated at left of each panel, $\delta = 1$, independent Laplace errors.
γ : 0 - black, 0.25 - red, 0.45 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

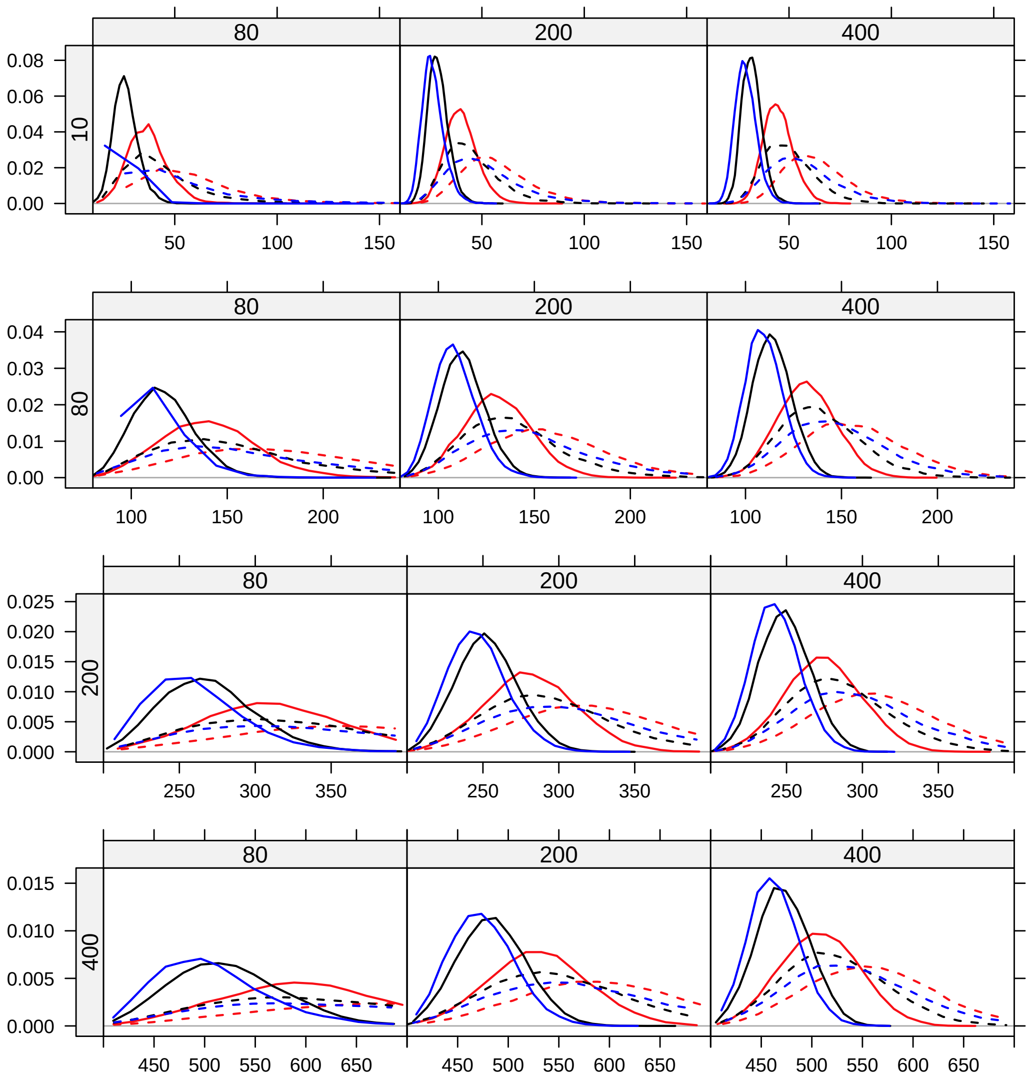

Figure 6.10: DRL for different lengths of training data m, change-point k^* indicated at left of each panel, $\gamma = 0.25$.
Procedure: Huber - black, L_1 - red, L_2 - blue; δ : 2 - solid, 1 - dashed.

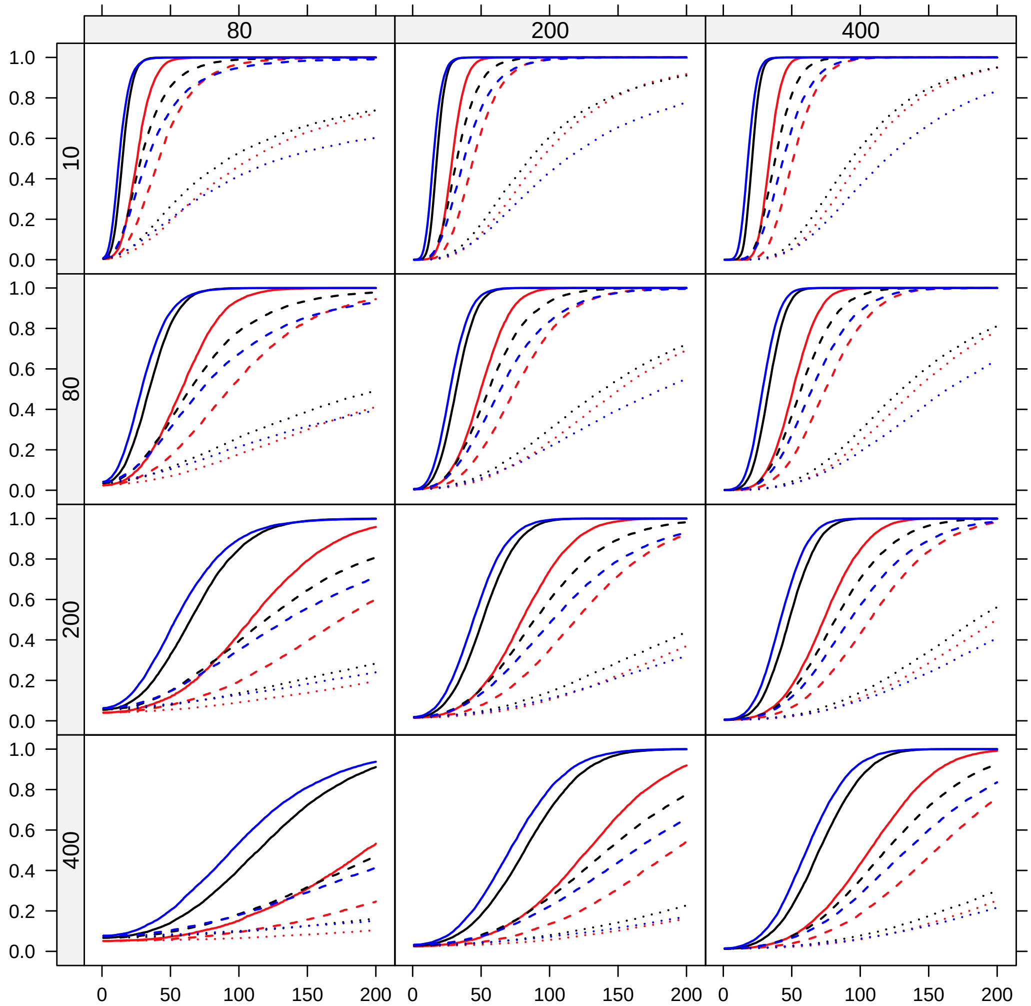

Figure 6.11: DPC for different lengths of training data m, change-point k^* indicated at left of each panel, $\gamma = 0.25$.
Procedure: Huber - black, L_1 - red, L_2 - blue; δ : 2 - solid, 1 - dashed, 0.5 - dotted.

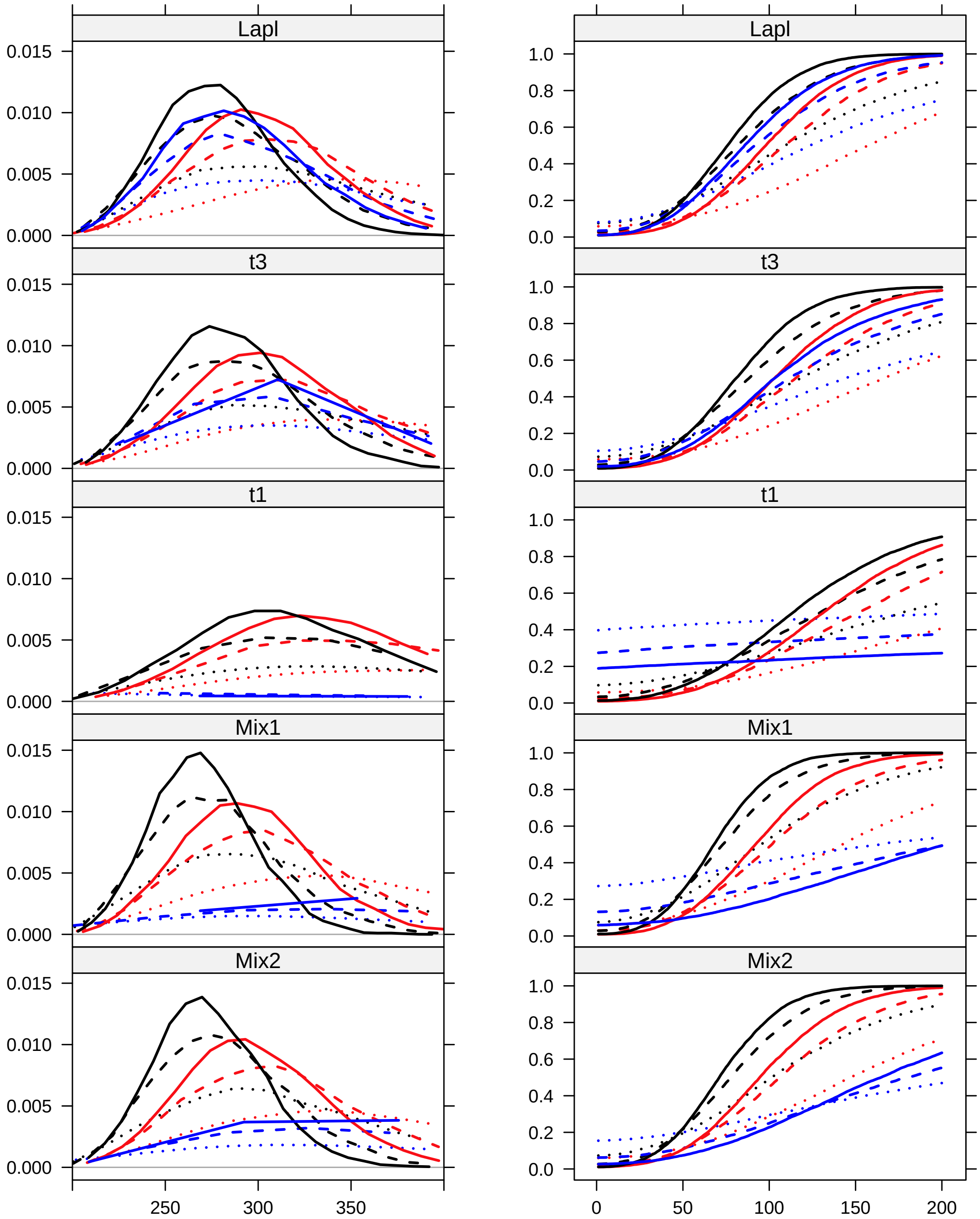

Figure 6.12: DRL (on the left) and DPC (on the right) of the procedures for different distributions, $k^* = 200$, $\gamma = 0.25$, $\delta = 1$.
Procedure: Huber - black, L_1 - red, L_2 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

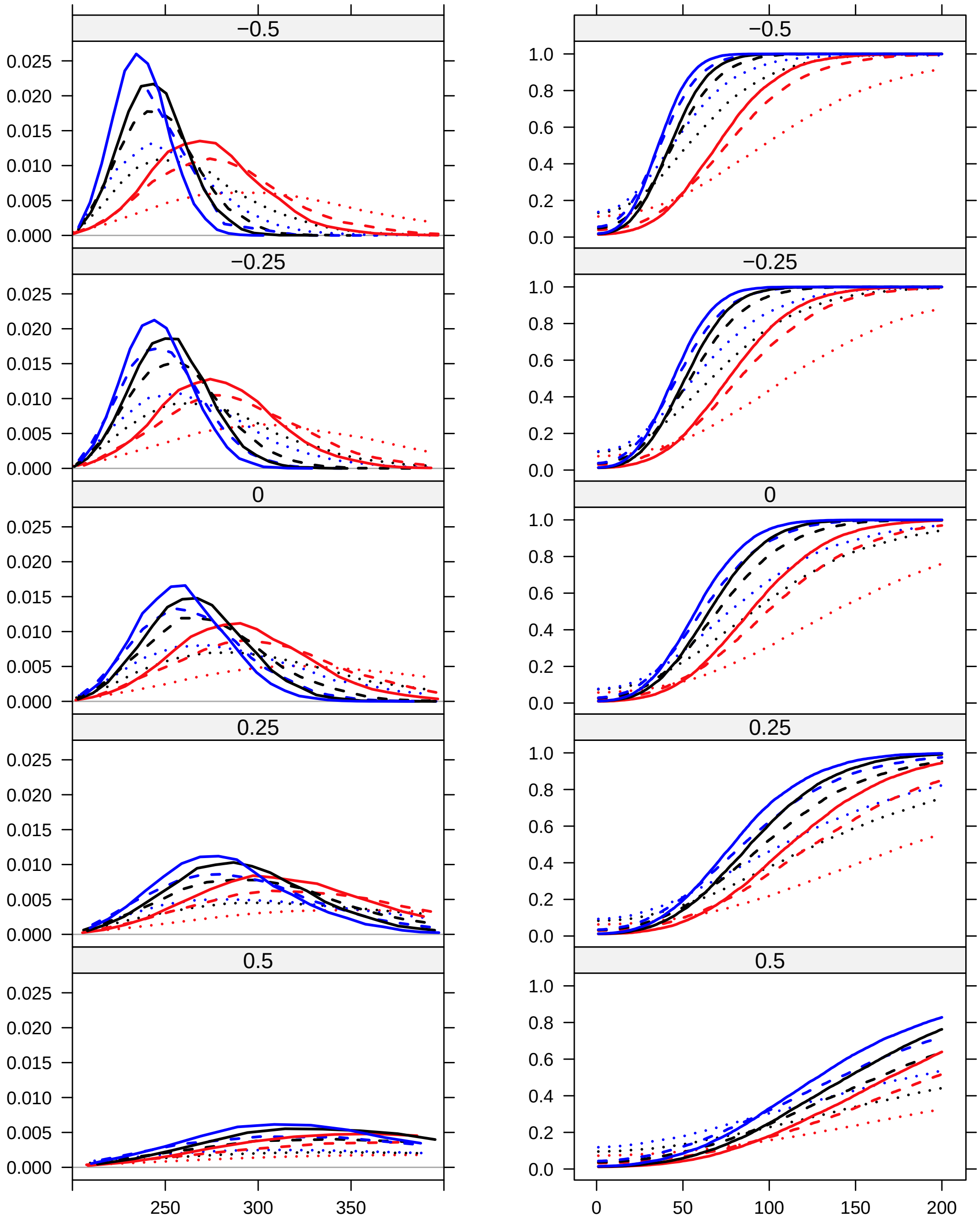

Figure 6.13: DRL (on the left) and DPC (on the right) of the procedures for different degrees of dependence (AR(1) sequences of N(0,1) innovations), $k^* = 200$, $\gamma = 0.25$, $\delta = 1$.
Procedure: Huber - black, L_1 - red, L_2 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

6.2 Multivariate Location Model

In this section we concentrate on the performance of the monitoring procedures introduced in Chapter 3. A lot of aspects are similar to the univariate case (we used an analogous setting) and thus the presentation will be omitted. For example the relationship between the tuning parameter γ and the change-point k^* is exactly the same and thus only the central value $\gamma = 0.25$ is considered. Moreover we restrict ourselves to $d = 2$ dimensional model with $\boldsymbol{\mu}_0 = (1,1)^T$. For estimation we use the same ψ_j function for both coordinates $j = 1, 2$ and thus refer to the monitoring procedures as L_1, Huber, L_2 procedures again.

Lets start again with a choice of an estimator of the long-run variance matrix (LRV) from (3.6). We are not aware of an analogue of the adaptive bandwidth choice procedure for FLT estimator (from the previous section) in a multivariate case. Thus we use older results of Andrews [1991] which are implemented in `R` package `sandwich` (see Zeileis [2004]). There the adaptive bandwidth choice exists for several kernels, from which we choose the Bartlett, Truncated and Quadratic Spectral (QS – this is recommended) ones. We compare the performance of the monitoring procedure using this LRV estimators with the ones using fixed lags Bartlett and FLT kernels, where $\Lambda_m = 4, 10, 20$ is chosen.

Figure 6.14 is an analogue of Figure 6.1 showing a matrix of SPC plots for the Huber procedure, where in columns there are different LRV estimators (the Bartlett, FLT with fixed number of lags and the above mentioned adaptive ones), whereas row panels represent various degrees of dependence of random errors - they form a vector autoregression VAR(1)

$$\boldsymbol{e}_i = \boldsymbol{A}(\rho)\boldsymbol{e}_{i-1} + \boldsymbol{\zeta}_i, \tag{6.3}$$

with coefficient matrix $\boldsymbol{A}(\rho) = \begin{pmatrix} \rho & 0 \\ 0 & \rho \end{pmatrix}$, coefficient ρ listed at LHS of each panel and i.i.d. innovations with two-dimensional normal distribution $\boldsymbol{\zeta}_i \sim N_2(\mathbf{0}, \boldsymbol{A}(0.25))$. The empirical sizes are based on 1000 repetitions and we use the same sequence of innovations for each setting in order to allow easier comparison.

We can see that the adaptive choice of bandwidth delivers the best results in general. The Truncated kernel does not function for the negative dependence, otherwise the performance of the three "adaptive" kernels is comparable. The FLT kernel with $\Lambda_m = 4$ works well except for the very strong dependence. For the Bartlett kernel the optimal number of lags seems to be 10, however the performance is worse than for FLT kernel. We can also see a big difference between the fixed number and the adaptive choice of lags for the Bartlett kernel. Based on these conclusions we decided to use the Quadratic Spectral kernel with an adaptive bandwidth choice in following figures. For comparison we ran the same simulations with the FLT kernel with $\Lambda_m = 4$, figures can be found attached. Results are based on 2000 repetitions.

Figure 6.15 is an analogue of Figure 6.7, where the SPC plots are presented for the L_1, Huber, L_2 procedures with the adaptive QS kernel for various distributions of random errors. We consider either independent components with i.i.d errors (heavy-tailed t_1

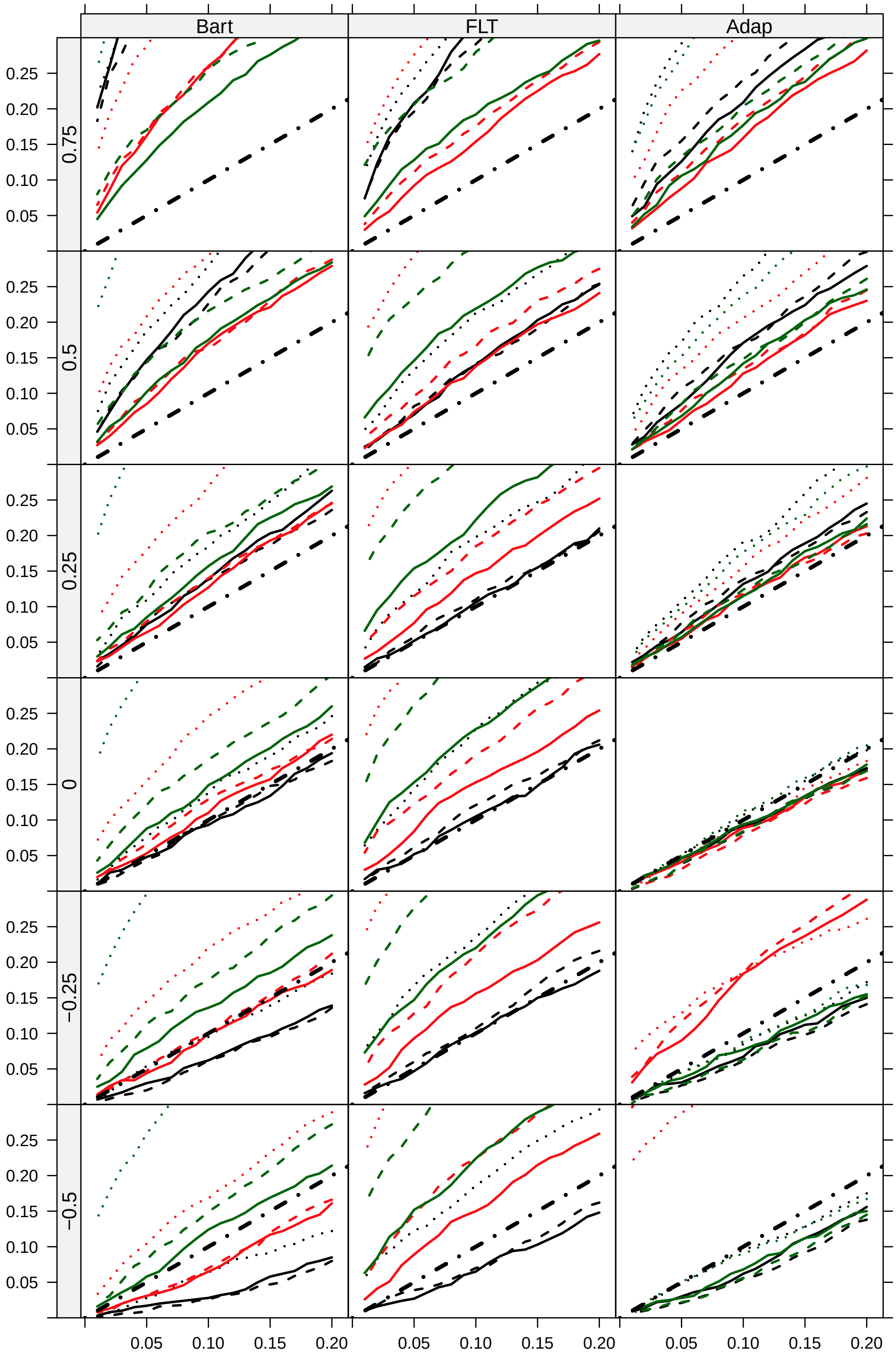

Figure 6.14: SPC for different LRV estimators, Huber procedure under H_0, $\gamma = 0.25$, errors being VAR(1) with $\boldsymbol{A}(\rho)$, ρ indicated at LHS of each panel, of $N_2(\mathbf{0}, \boldsymbol{A}(0.25))$ innovations.
Λ_m/adaptive kernel : 4/Bartlett - black, 10/Truncated - red, 20/QS - green;
m : 80 - dotted, 200 - dashed, 400 - solid.

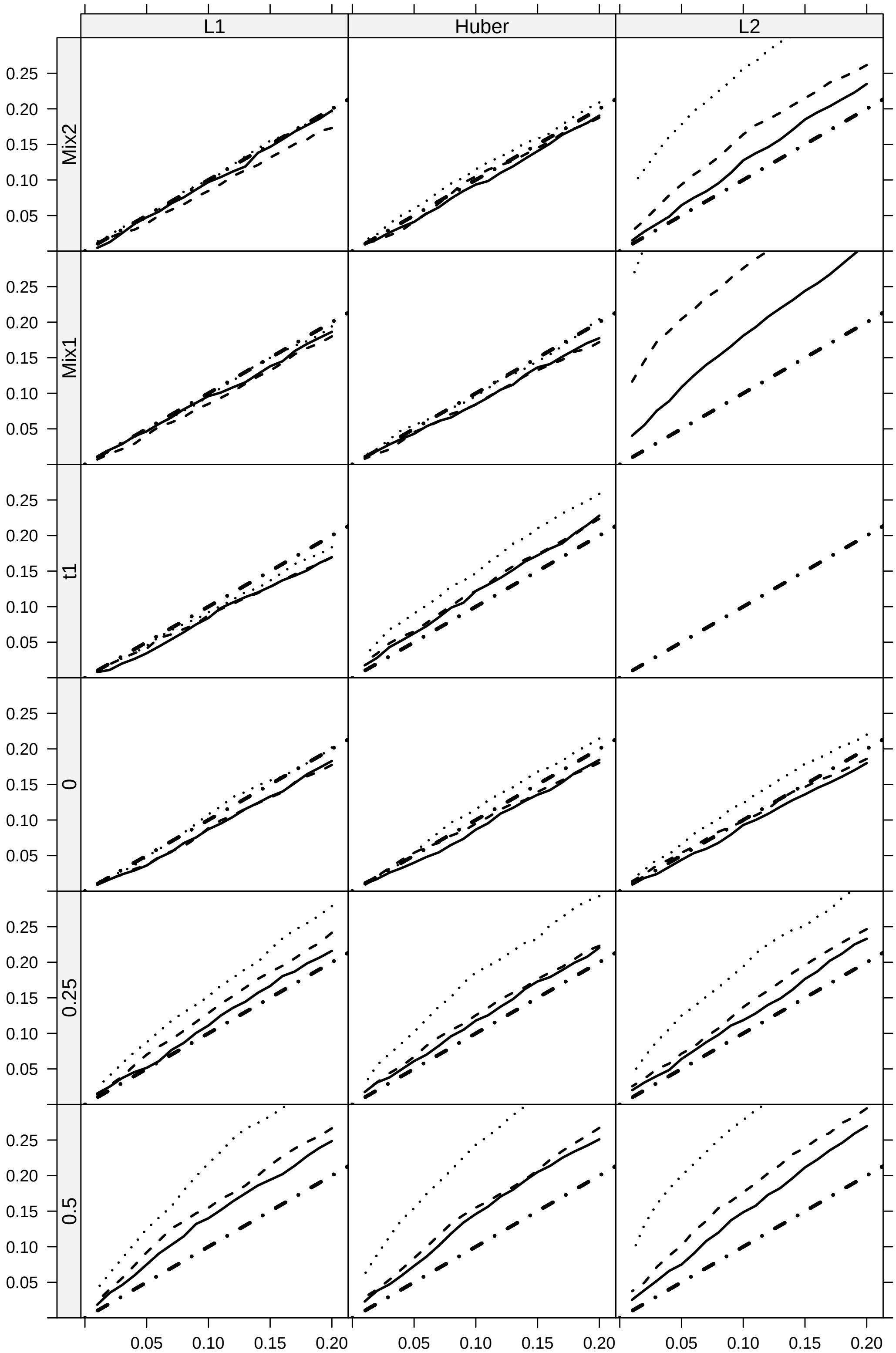

Figure 6.15: SPC for different procedures with adaptive QS estimator under H_0, different distributions of errors in each panel, $\gamma = 0.25$.
m : 80 - dotted, 200 - dashed, 400 - solid.

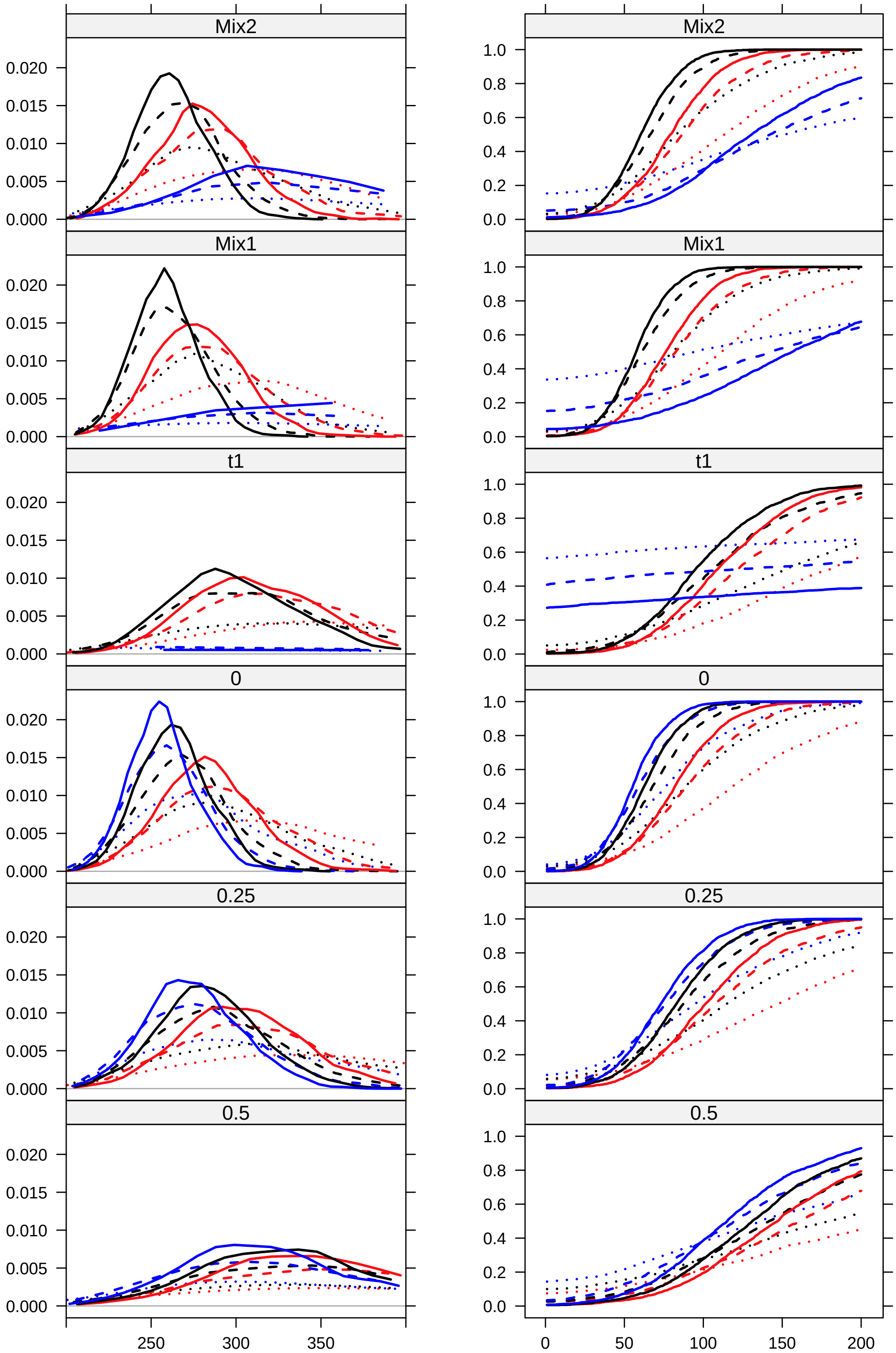

Figure 6.16: DRL (on the left) and DPC (on the right) of the procedures for different distributions, $k^* = 200$, $\gamma = 0.25$, $\delta_m \boldsymbol{\theta} = (1, 1)^T$.
Procedure: Huber - black, L_1 - red, L_2 - green; m : 80 - dotted, 200 - dashed, 400 - solid.

and two contaminated normal ones – Mix1, Mix2 introduced in the previous section) or dependent ones (the vector autoregression (6.3) with $\rho = 0, 0.25, 0.5$).

We can see again that the L_2 procedure performs unsatisfactory in a presence of outliers, whereas the Huber and L_1 procedures are not influenced by these. With a growing dependence the procedure requires more training data to get close to the nominal level. For the FLT kernel $m = 80$ is not enough in general (see Attachment).

For the alternative we consider a change $\delta_m \boldsymbol{\theta} = (1,1)^T$ at point $k^* = 200$. In Figure 6.16 we can see DRL and DPC charts for all three procedures considered and previously mentioned error distributions. The charts confirm the robustness of L_1 and Huber procedures and the inappropriateness of L_2 procedure for heavy-tailed/contaminated distributions. It also shows that the Huber procedure is not much worse than the L_2 even for the normal errors. L_1 procedure is a bit worse in this situation. When the change occurs in just one coordinate the procedures lose power to a small degree. The results are almost the same also for the FLT kernel LRV estimator. (See figures attached.)

6.3 CAPM

In this section we focus on the monitoring procedure for the Capital Asset Pricing Model. As the situation there is similar to the multivariate location model, the presentation of results is analogical. Some results mainly in a form of a numerical summary can be also find in Chochola et al. [2013], where the Bartlett kernel was used. Based on the experience with the LRV estimators from the previous section, we use here the adaptive Quadratic Spectral one instead. Also FLT with $\Lambda_m = 4$ was tried, figures for which can be found attached.

We consider $d = 2$ dimensional model (4.3) with $\boldsymbol{\alpha}^0 = \boldsymbol{\beta}^0 = (1,1)^T$, where both the random errors as well as the market portfolio can be dependent. Similarly as in the previous section the random errors form either vector autoregression (6.3) where we use an autoregression AR(1) model with the same coefficient ρ for modelling the market portfolios demeaned returns $\widetilde{r}_{iM}$'s. Or we consider independent components of heavy-tailed/contaminated distributions for random errors. In this case we used i.i.d. standard normal distribution for the market portfolio returns. Notation in figures is based on the distribution of random errors, which according to previous governs also the one of market portfolio returns. We use the central value of the tuning parameter $\gamma = 0.25$ in all figures.

Figure 6.17 is a direct analogue of Figure 6.15. Also the conclusions are the same and thus are not discussed here. Use the FLT kernel instead of the adaptive QS one for the estimation of LRV causes worse results of the procedures in respect of keeping the nominal level. Especially when m is small.

Figure 6.18 presents the alternative, when $\boldsymbol{\alpha}^1 = \mathbf{0}$ and $\delta_m \boldsymbol{\beta}^1 = (1,1)^T$, i.e. a change in both components of the portfolio beta occurs at $k^* = 200$. We see that the robust (L_1 and Huber) procedures are again reasonably sensitive, whereas the L_2 procedure

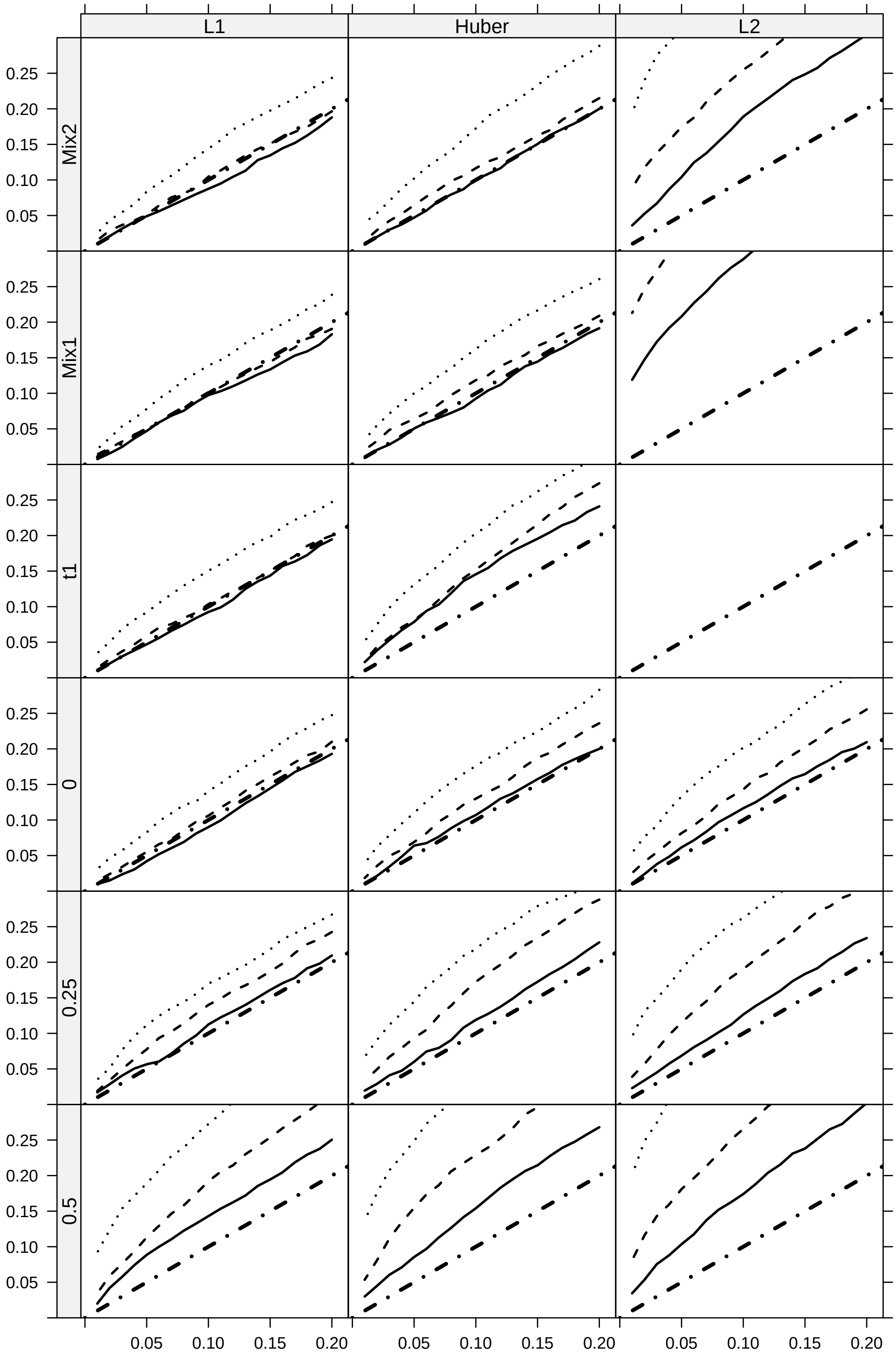

Figure 6.17: SPC for different procedures with adaptive QS estimator under H_0, different distribution of errors in each panel, $\gamma = 0.25$.
m : 80 - dotted, 200 - dashed, 400 - solid.

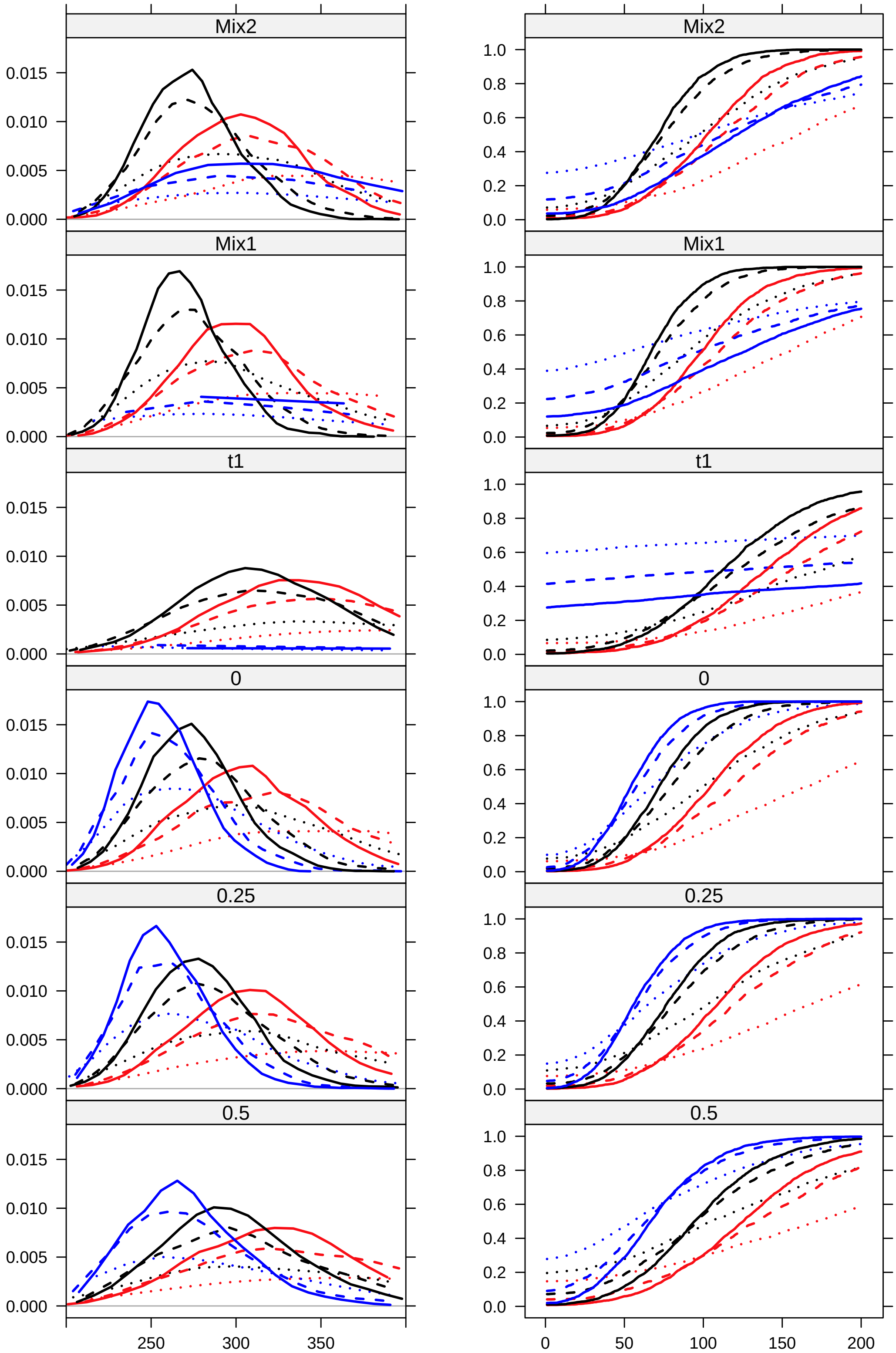

Figure 6.18: DRL (on the left) and DPC (on the right) of the procedures for different distributions, $k^* = 200$, $\gamma = 0.25$, $\delta_m \boldsymbol{\beta}_1 = (1, 1)^T$, $\boldsymbol{\alpha}_1 = \mathbf{0}$.
Procedure: Huber - black, L_1 - red, L_2 - blue; m : 80 - dotted, 200 - dashed, 400 - solid.

fails to detect the change soon (or at all in case of t_1 distribution). For the normal errors the Huber procedure performs not much worse than the L_2 and thus it can be recommended. In case when also the intercept changes, the procedures perform slightly worse. A further bit worse is a situation when only one of the portfolio betas changes. And for the sake of completeness we discuss also the situation when only the intercept changes. According to Remark 4.2 the limit distribution of the test statistic is not sensitive to this change, which was confirmed by the simulation as well. The results for the FLT kernel are similar. Figures for all the above mentioned alternatives can be found in Attachment.

Finally, as an illustration of a possible application, we investigated a data set of MSCI Global Sector Indices (net prices) that can serve as a benchmark to conduct relative valuations of sectors, industry groups and industries across countries and regions. Three sector indices – NDWUCSTA-World Consumer Staples (food, beverages, tobacco, prescription drugs and household products), NDWUFNCL-World Financials, and NDWUHC-World Health Care have been studied, and we chose NDDUWI MSCI World Index (a weighted index designed to measure the equity market performance of 24 developed country market indices) to represent the market portfolio. *

We have a sample of data from 29/12/2000 to 29/03/2011. The data from the period 31/12/2004 to 01/12/2006 (of length $m = 500$) were examined by a retrospective test of Section 5.3. Since the test did not reject the null hypothesis of no change, we used this data set as the (stable) historical period for our monitoring procedure. The length of the monitoring period is $2m = 1000$, that is, the monitoring terminates on 01/10/2010. So, critical values were chosen for $T = 2$.

Since, as described earlier, the Huber-type procedure provides a good combination of efficiency and robustness, we only present the results for this type of monitoring here. Also, as we did not know where to expect the possible change, we used $\gamma = 0.25$ as a compromise between detecting an early or late change. The Bartlett kernel variance estimator with bandwidth of $\Lambda_m = 4$ was used. Moreover, we considered the portfolio including all three indices as well as all pairwise combinations.

In Figure 6.19 the values of the test statistics are shown together with the critical values; a solid line indicates the critical value for $T = \infty$ (i.e., for the open-end monitoring) and a dashed line the one for $T = 2$ (closed-end monitoring). A solid vertical line marks the date 01/08/2007, i.e., the date when the sub-prime mortgage crisis approximately started. The figures, particularly the first three ones, demonstrate the high sensitivity of the portfolio risk with respect to the financial sector.

*Source: Bloomberg, 2011,
http://www.msci.com/products/indices/tools/tickers/bb_eod/bloomberg_tickers_eod_sector.html

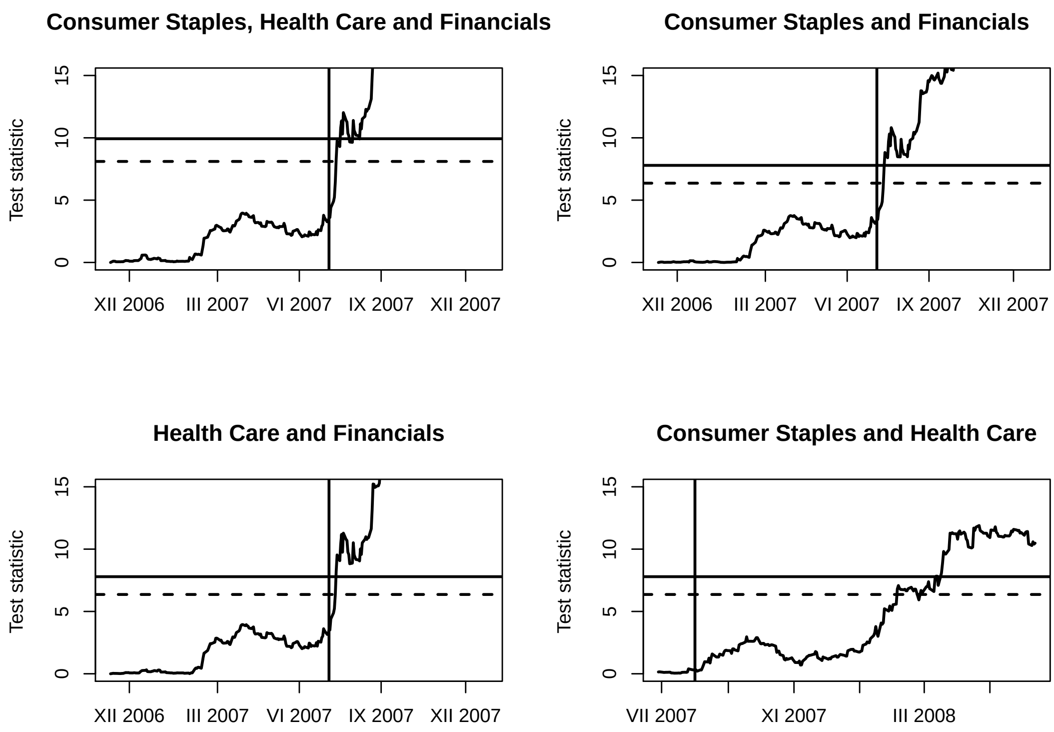

Figure 6.19: Test statistics and critical values for different combinations of indices.

Chapter 7

Critical Values

In this short chapter we gather the information about critical values of the procedures proposed in the thesis. We show how the critical values can be obtained and also present tables with these values available to use.

7.1 Online Monitoring

As it has been already written in Section 1.3 the critical values have to be chosen such that the requirements on the type I and type II error rates, i.e., (1.6) and (1.7) are satisfied. For approximation of the critical values $c_{m,T}(\alpha)$ introduced in (1.5), we can use the limit distribution derived in particular theorems.

For the univariate case Theorem 2.1 gives us that the critical value can be approximated by value $c_T(\alpha, \gamma)$ such that

$$P\left(\sup_{0\le t\le T/(T+1)} \frac{|W(t)|}{t^\gamma} > c_T(\alpha,\gamma)\right) = \alpha, \tag{7.1}$$

where the dependence on $\gamma \in [0, 1/2)$ comes from the boundary function $q_\gamma(\cdot)$.

By the time rescalling property of Wiener process

$$\sup_{0\le x\le T/(T+1)} \frac{|W(x)|}{x^\gamma} \stackrel{\mathcal{D}}{=} \sup_{0\le t\le 1} \frac{\left|W\left(t\frac{T}{T+1}\right)\right|}{\left(t\frac{T}{T+1}\right)^\gamma} \stackrel{\mathcal{D}}{=} \left(\frac{T}{T+1}\right)^{1/2-\gamma} \sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma} \tag{7.2}$$

and thus we can further consider only the functional

$$\sup_{0\le t\le 1} \frac{|W(t)|}{t^\gamma} \tag{7.3}$$

which corresponds to the open-end monitoring procedure. We will denote the critical values for this functional as $c_\infty(\alpha, \gamma)$. The relation between both critical values is by

(7.2) easy to see, namely we have

$$c_T(\alpha,\gamma) = \Big(\frac{T}{T+1}\Big)^{1/2-\gamma} c_\infty(\alpha,\gamma). \tag{7.4}$$

The problem is that an explicit formula for the distribution function of the functional (7.3) is known only when $\gamma = 0$. Otherwise one has to use simulations. However these simulations were already performed and the critical values are published in Horváth et al. [2004] (Table 1). For convenience we present them also in Table 7.1 for various test levels α and tuning parameters γ.

$\gamma \setminus \alpha$	0.1	0.05	0.025	0.01
0.00	1.9497	2.2365	2.4948	2.7912
0.15	2.0273	2.2996	2.5475	2.8516
0.25	2.1060	2.3860	2.6396	2.9445
0.35	2.2433	2.5050	2.7394	3.0475
0.45	2.5437	2.7992	3.0144	3.3015
0.49	2.8259	3.0722	3.2944	3.5705

Table 7.1: Simulated critical values $c_\infty(\alpha,\gamma)$.

To illustrate the influence of the closed-end monitoring we evaluate the mutliplicative factor $\left(\frac{T}{T+1}\right)^{1/2-\gamma}$ of (7.4) for $T = 1, 2, 5, 10$ and some γ previously considered. These are presented in Table 7.2. We can see that for $T = 1$ or 2 the influence is quite significant.

$\gamma \setminus T$	1	2	5	10
0	0.7071	0.8165	0.9129	0.9535
0.25	0.8409	0.9036	0.9554	0.9765
0.45	0.9659	0.9799	0.9909	0.9952

Table 7.2: Multiplicative factors for $c_T(\cdot,\gamma)$.

For the multivariate and the CAPM case the situation is similar. The asymptotic distribution of the detectors derived in Theorems 3.1 and 4.1 is the same and the functional corresponding to open-end procedure there is

$$\sup_{0\le t\le 1} \frac{\sum_{j=1}^d W_j^2(t)}{t^{2\gamma}}.$$

We denote its critical values by $c_\infty^{(d)}(\alpha,\gamma)$. These are published in Chochola et al. [2013] for $d = 2,\ldots,5$. For $d = 1$ we can easily obtain them from univariate case since $c_\infty^{(1)}(\alpha,\gamma) = c_\infty^2(\alpha,\gamma)$. The critical values were obtained from simulations again as a respective empirical quantile of suprema of the functional given above. This supremum has been approximated by a maximum over a grid of 25,000 equidistant points, and 100,000 repetitions have been run in total. We present them in Table 7.3 .

d	$\alpha \setminus \gamma$	0	0.15	0.25	0.4	0.45	0.49
	10%	5.83300	6.16964	6.54486	7.79693	8.90706	10.97680
2	5%	7.27319	7.62029	8.01801	9.24979	10.38189	12.51981
	1%	10.47212	10.81526	11.18947	12.41796	13.58373	16.08758
	10%	7.55347	7.91567	8.33422	9.69223	10.89566	13.24342
3	5%	9.15817	9.51428	9.92618	11.27827	12.47845	14.93875
	1%	12.64423	12.97544	13.35888	14.71475	15.93770	18.61511
	10%	9.15704	9.54268	9.96759	11.40482	12.68321	15.28504
4	5%	10.89252	11.26607	11.67221	13.12474	14.41193	17.05890
	1%	14.65064	15.00585	15.43069	16.88893	18.13029	20.88200
	10%	10.63242	11.04519	11.48214	12.97519	14.35397	17.13813
5	5%	12.47376	12.87663	13.31469	14.80208	16.16445	19.02006
	1%	16.43966	16.84611	17.32441	18.86821	20.13233	23.11929

Table 7.3: Simulated critical values of $c_\infty^{(d)}(\alpha, \gamma)$.

Critical values for a closed-end monitoring $c_T^{(d)}(\alpha, \gamma)$ can be again easily obtained from $c_\infty^{(d)}(\alpha, \gamma)$ as

$$c_T^{(d)}(\alpha, \gamma) = c_\infty^{(d)}(\alpha, \gamma)\Big(\frac{T}{T+1}\Big)^{1-2\gamma},$$

where the derivation is similar to one in (7.2).

7.2 Retrospective Analysis

For the univariate model the limit distribution (5.6) of the test statistic is well known, and thus we focus on the multivariate and CAPM models. By Theorem 5.3 is it necessary to find critical values for

$$\sup_{0<t<1} \sum_{j=1}^{d} B_j^2(t), \tag{7.5}$$

where $\{B_j(t),\, t \in (0,1)\},\ j = 1, \dots, d$ are independent Brownian bridges. The simulated critical values for this functional are presented in Table 7.4 for $d = 1, 2, 3, 4, 5$ and nominal levels $\alpha = 10\%, 5\%, 1\%$.

The supremum of the functional given in (7.5) has been approximated by a maximum over a grid of 50,000 equidistant points, and 100,000 repetitions have been run. Thus these critical values refine those of Lee et al. [2003], where the supremum was approximated only on 1000 points.

$\alpha \setminus d$	1	2	3	4	5
10%	1.49260	2.10796	2.62212	3.07204	3.50604
5%	1.83855	2.50356	3.04211	3.52956	3.98640
1%	2.64916	3.36212	3.98668	4.51394	5.02544

Table 7.4: Simulated critical values for functional (7.5).

Chapter 8

Conclusion

We proposed robust monitoring procedures for dependent and possibly multivariate data. Critical values were derived from the asymptotic distribution of the test statistics. Simulation studies showed that the asymptotic approximation works well, provided that we have reasonably long period of stable historical data. We also commented on retrospective change-point procedures, that allow one to verify this stability in a robust way.

The simulation studies also showed that a key ingredient of the test detectors is an estimator of the (long-run) variance (matrix). We therefore examined a class of kernel variance estimators, especially the topic of a proper bandwidth choice. It was shown that the adaptive choice of the bandwidth can significantly improve the monitoring procedures. The question of variance estimators is usually not discussed in details in a change-point literature, but it was shown to be quite important.

A generalization of the proposed methods to different models can be further studied. For example to a multivariate regression model as an extension of the CAPM model of Chapter 4. Or to consider high-dimensional (functional) data, which is challenging but rather difficult. Another possibility is to focus on MOSUM type procedures instead of CUSUM ones, which could provide faster detection of the change.

Appendix A

Some Useful Results

Following two lemmas adopted from [Hušková and Marušiaková, 2012] provide useful inequalities for α-mixing sequences.

Lemma I. *Let $\{e_i\}_i$ be a strictly stationary α-mixing sequence with coefficients $\{\alpha(i)\}_i$. Denote by $\mathcal{F}_j^k$ the σ-fields generated by $\{e_i\}_{j\le i\le k}$. Let Z_1 and Z_2 be measurable w.r.t. $\mathcal{F}_1^k$ and $\mathcal{F}_{k+n}^{\infty}$ $(n \ge 1)$, respectively. Then,*

$$|\mathrm{E}(Z_1Z_2) - \mathrm{E}(Z_1)\,\mathrm{E}(Z_2)| \le 12\,(\alpha(n))^{1/s}\,(\mathrm{E}\,|Z_1|^p)^{1/p}\,(\mathrm{E}\,|Z_2|^q)^{1/q} \tag{A.1}$$

for all $1 \le p, q, s \le \infty$ with $1/p + 1/q + 1/s = 1$.

Proof. Proof of (A.1) can be found, e.g., in Ibragimov [1962] and Davydov [1970]. □

Lemma II. *Let $\{e_i\}_i$ be a strictly stationary α-mixing sequence with coefficients $\{\alpha(i)\}_i$. Let g_n be a measurable function such that $\mathrm{E}\ g_n(e_i) = 0$.*

Then, for any $\xi > 0$

$$\mathrm{E}\left|\sum_{i=1}^{n} g_n(e_i)\right|^2 \le Dn\left(\mathrm{E}\,|g_n(e_1)|^{2+\xi}\right)^{2/(2+\xi)}\sum_{j=1}^{\infty}(\alpha(j))^{\xi/(2+\xi)} \tag{A.2}$$

and

$$\mathrm{E}\left(\max_{1\le j\le n} b_j\left|\sum_{i=1}^{j} g_n(e_i)\right|\right)^2 \le$$

$$\le D(\log(2n))^2\left(\sum_{j=1}^{n} b_j^2\right)\left(\mathrm{E}\,|g_n(e_1)|^{2+\xi}\right)^{2/(2+\xi)}\sum_{j=1}^{\infty}(\alpha(j))^{\xi/(2+\xi)} \tag{A.3}$$

for any $b_1 \ge \ldots \ge b_n > 0$ and some $D > 0$.

Proof. The assertion (A.2) follows from the proof of Theorem 1 in Yokoyama [1980]. The assertion (A.3) is a consequence of Theorem B.4 in Kirch [2006] and the assertion (A.2). □

Now we focus on a weak convergence of vector valued processes. Theory is presented for example in sections 27.7 and 29.5 of Davidson [1994] and is an obvious generalization of one dimensional case. We will denote by $C^d[0,T]$ the space of continuous d-dimensional vector functions defined on interval $[0,T]$ for some $T > 0$. Analogously we denote $D^d[0,T]$ the space of d- dimensional vector cadlag functions defined on interval $[0,T]$. Weak convergence in these spaces is denoted as $\xrightarrow{\mathcal{D}^d[0,T]}$, $\xrightarrow{\mathcal{C}^d[0,T]}$ respectively.

Following lemma gives a functional Cramer-Wold device, which allows one to transform the vector case to the univariate case for which standard methods can be applied.

Lemma III. *Let $\boldsymbol{X}_n \in D^d[0,T]$ be an d-vector of cadlag functions. Then $\boldsymbol{X}_n \xrightarrow{\mathcal{D}^d[0,T]} \boldsymbol{X}$, where $P(\boldsymbol{X} \in C^d[0,T]) = 1$, if and only if $\boldsymbol{\lambda}^T \boldsymbol{X}_n \xrightarrow{\mathcal{D}[0,T]} \boldsymbol{\lambda}^T \boldsymbol{X}$ for every fixed $\boldsymbol{\lambda}$ with $\boldsymbol{\lambda}^T \boldsymbol{\lambda} = 1$.*

Proof. Can be found in Davidson [1994], Theorem 29.16. □

Note that the lemma requires that the weak limit is almost surely continuous. As we apply the lemma only to the situation where $\boldsymbol{X} = \boldsymbol{W}$ is a d-dimensional Wiener process, the requirement is always fulfilled.

Bibliography

Anderson, T. (1958). *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons, New York, 1st edition.

Anděl, J. (1985). *Matematická statistika.* SNTL - Nakladatelství technické literatury.

Andrews, D. W. K. (1984). Nonstrong mixing autoregressive processes. *J. Appl. Probab.*, 21:930–934.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica*, 59:817–858.

Andrews, D. W. K. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61(4):821–856.

Antoch, J. and Hušková, M. (1989). Some m-tests for detection of a change in linear model. In *Proceedings of the Fourth Prague Symposium on Asymptotic Statistics*, pages 123–136.

Antoch, J., Hušková, M., and Daniela, J. (2002). Off-line statistical process control. In *Multivariate Total Quality Control*, Contributions to Statistics, pages 1–86. Physica-Verlag HD.

Antoch, J., Hušková, M., and Prášková, Z. (1997). Effect of dependency on statistics for determination of change. *Journal of Statistical Planning and Inference*, 60:291–310.

Aue, A., Hörmann, S., Horváth, L., Hušková, M., and Steinebach, J. G. (2012). Sequential testing for the stability of high-frequency portfolio betas. *Econometric Theory*, 28(4):804–837.

Aue, A., Horváth, L., Hušková, M., and Kokoszka, P. (2006). Change-point monitoring in linear models. *Econometrics Journal*, 9(3):373–403.

Aue, A., Horváth, L., Kokoszka, P., and Steinebach, J. (2008). Monitoring shifts in mean: Asymptotic normality of stopping times. *Test*, 17(3):515–530.

Bahadur, R. R. (1966). A note on quantiles in large samples. *Ann. Math. Statist.*, 37:577–580.

Billingsley, P. (1968). *Convergence of Probability Measures.* John Willey & Sons, New York, 1st edition.

Bradley, R. C. (2005). Basic properties of strong mixing conditions. a survey and some open questions. *Probability surveys*, 2(107-44):37.

Caporale, T. (2012). Time varying capm betas and banking sector risk. *Economics Letters*, 115(2):293–295.

Chochola, O. (2008). Sequential monitoring for change in scale. *Kybernetika*, 44:717–730.

Chochola, O., Hušková, M., Prášková, Z., and Steinebach, J. G. (2013). Robust monitoring of capm portfolio betas. *J. Multivariate Analysis*, 115:374–395.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64:1045–1065.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley, Chichester.

Csörgő, M. and Horváth, L. (1993). *Weighted Approximation in Probability and Statistics*. John Wiley & Sons, Chichester.

Davidson, J. (1994). *Stochastic Limit Theory*. Oxford University Press.

Davydov, Y. (1970). The invariance principle for stationary processes. *Theory of Probability and Its Applications*, 15:487–498.

Doukhan, P. (1994). *Mixing: properties and examples*, volume 85 of *Lecture Notes in Statistics*. Springer, New York.

Doukhan, P. and Louhichi, S. (1999). A new weak dependence condition and applications to moment inequalities. *Stochastic Processes and their Applications*, 84(2):313–342.

Genton, M. G. and Ronchetti, E. (2008). Robust prediction of beta. In *Computational Methods in Financial Engineering*, pages 147–161. Springer.

Ghysels, E. (1998). On stable factor structures in the pricing of risk: Do time-varying betas help or hurt? *The Journal of Finance*, 53(2):549–573.

He, X. and Shao, Q. M. (1996). A general bahadur representation of m-estimators and its application to linear regression with nonstochastic design. *The Annals of Statistics*, 24(6):2608–2630.

Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *Annals of Statistics*, pages 1845–1884.

Horváth, L., Hušková, M., Kokoszka, P., and Steinebach, J. (2004). Monitoring changes in linear models. *J. Stat. Plann. Inference*, 126:225–251.

Huber, P. (1981). *Robust Statistics*. Wiley, Chichester.

Huber, P. and Ronchetti, E. M. (2009). *Robust Statistics*. Wiley, Chichester.

Huber, P. J. (1964). Robust estimation of a location parameter. *Annals of Mathematical Statistics*, 35:73–101.

Hušková, M. and Antoch, J. (2001). M-estimators of structural changes in regression models. *Tatra Mt. Math. Publ.*, 22:197—208.

Hušková, M. and Chochola, O. (2010). Simple sequential procedures for change in distribution. In *Nonparametrics and Robustness in Modern Statistical Inference and Time Series Analysis: A Festschrift in honor of Professor Jana Jurečková*, volume 7, pages 95–104. Institute of Mathematical Statistics.

Hušková, M. and Koubková, A. (2005). Monitoring jump changes in linear models. *Journal of Statistical Research*.

Hušková, M. and Kirch, C. (2010). A note on studentized confidence intervals for the change-point. *Computational Statistics*, 25(2):269–289.

Hušková, M. and Marušiaková, M. (2012). M-procedures for detection of changes for dependent observations. *Comunications in Statistics*, 41(7):1032–1050.

Ibragimov, I. (1959). Some limit theorems for stochastic processes stationary in the strict sense. *Doklady Akademii Nauk SSSR (in Russian)*, 125:711—714.

Ibragimov, I. (1962). Some limit theorems for stationary processes. *Theory of Probability and Its Applications*, 7:349–382.

Ibragimov, I. and Linik, Y. (1971). *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff.

Jurečková, J. and Sen, P. K. (1996). *Robust Statistical Procedures: Asymptotics and Interrelations*. John Wiley & Sons, Inc.

Kirch, C. (2006). *Resampling methods for the change analysis of dependent data*. PhD thesis, Köln University. Referees: J. Steinebach (Köln), W. Wefelmayer (Köln), M. Hušková (Prague).

Koubková, A. (2006). *Sequential Change-Point Analysis*. PhD thesis, Matematicko-fyzikální fakulta, UK.

Lee, S., Ha, J., Na, O., and Na, S. (2003). The cusum test for parameter change in time series models. *Scandinavian Journal of Statistics*, 30(4):781–796.

Leisch, F., Hornik, K., and Kuan, C. M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16:835–854.

Lin, Z. and Lu, C. (2010). *Limit Theory for Mixing Dependent Random Variables*. Mathematics and Its Applications. Springer.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. *The review of economics and statistics*, 47(1):13–37.

Martin, R. D. and Simin, T. T. (2003). Outlier-resistant estimates of beta. *Financial Analysts Journal*, pages 56–69.

Merton, R. C. (1973). An intertemporal capital asset pricing model. *Econometrica*, 41(5):867–887.

Page, E. S. (1954). Continuous inspection schemes. *Biometrika*, 41:100–115.

Politis, D. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics*, 15(4):517–533.

Prášková, Z. and Chochola, O. (To appear, 2013). M-procedures for detection of a change under weak dependence. *Journal of Statistical Planning and Inference*.

Rosenblatt, M. (1956). A central limit theorem and a strong mixing condition. *Proceedings of the National Academy of Sciences of the United States of America*, 42(1):43.

Rosenblatt, M. (1971). *Markov Processes: Structure and Asymptotic Behavior*. Springer-Verlag., Berlin.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442.

Siegmund, D. (1985). *Sequential analysis*. Springer Series in Statistics. Springer-Verlag, New York. Tests and confidence intervals.

Wald, A. (1947). *Sequential Analysis*. John Wiley and Sons.

Withers, C. (1981). Conditions for linear processes to be strong-mixing. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 57:477–480.

Wu, W. (2007). m-estimation of linear models with dependent errors. *Annals of Statistics*, pages 495–521.

Yokoyama, R. (1980). Moment bounds for stationary mixing sequences. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 52:45–57.

Zeileis, A. (2004). Econometric computing with hc and hac covariance matrix estimators. *Journal of Statistical Software*, 11(10):1–17.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, 20:99–121.