Asymptotic Methods in Change-Point Analysis

Inaugural-Dissertation

zur
Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität zu Köln
vorgelegt von

Stefan Fremdt
aus Hadamar

Köln 2012

1. Berichterstatter: Prof. Dr. Josef G. Steinebach (Köln)
2. Berichterstatter: Prof. Dr. Wolfgang Wefelmeyer (Köln)
3. Berichterstatter: Prof. Dr. Lajos Horváth (Salt Lake City)

Tag der mündlichen Prüfung: 06.Dezember 2012

Zusammenfassung

Im Fokus dieser Arbeit stehen zwei Teilgebiete der Statistik, die Change-Point-Analyse und die Analyse funktionaler Daten, sowie die Schnittmenge dieser Gebiete, die sich mit der Aufdeckung von Strukturbrüchen in funktionalen stochastischen Modellen befasst. Die behandelten Fragestellungen aus der (skalaren) Change-Point-Analyse resultieren aus Kritik an bereits entwickelten Verfahren, denen mangelnde Stabilität bezüglich des Eintrittszeitpunktes eines Strukturbruchs vorgeworfen wurde. Als mögliche Antwort auf diese Kritik werden im Rahmen eines linearen Modells sequentielle Verfahren vorgestellt, die auf der sogenannten *Page CUSUM* basieren. Um die gewünschten Eigenschaften dieser Verfahren auch theoretisch zu belegen, wird die asymptotische Verteilung der Verzögerung beim Erkennen eines Strukturbruchs hergeleitet.

Der Begriff *Analyse funktionaler Daten* steht stellvertretend für den Teilbereich der Statistik, der *Funktionen* (in der Regel auf einem kompakten Intervall) als *Datenpunkte* betrachtet. Beispiele hierfür sind Temperaturkurven oder der Verlauf eines Aktienkurses an einem Handelstag. Um Statistik auf solchen funktionalen Stichproben betreiben zu können, stellen Techniken zur Dimensionsreduktion ein unerlässliches Hilfsmittel dar. Die in dieser Arbeit präsentierten Verfahren basieren auf der sogenannten Hauptkomponentenanalyse und verdeutlichen wie diese zur Konstruktion von Zweistichproben-, sowie von Change-Point-Tests verwendet werden können. Insbesondere wird das Problem der angemessenen Wahl der Dimension des Bildraumes der im Rahmen der Hauptkomponentenanalyse verwendeten Projektion in die Konstruktion der Testverfahren einbezogen.

Abstract

This thesis is focussed on two areas of statistics, change-point analysis and functional data analysis, and the intersection of these two areas, i.e., the detection of structural breaks in functional stochastic models. The considered problems from (scalar) change-point analysis result from criticism of already existing sequential change-point procedures. The subject of this criticism was a lack of stability of these procedures regarding the time of occurrence of a change. As a possible solution to this criticism sequential methods are presented in this thesis in the framework of a linear regression model on the basis of the so-called *Page CUSUM*. To prove the desired properties of these procedures theoretically the asymptotic distribution of the delay time in the detection of structural breaks is derived in the special case of a location model.

The notion *functional data analysis* represents an area of statistics that considers *functions* (in general defined on a compact interval) as *data points*. Examples for such data are temperature curves or the path of a stock price on one trading day. To derive statistical procedures for this class of data dimension reduction techniques play a key role. The methods presented in this thesis are based on one of those techniques, the functional principal component analysis. They illustrate the construction of two-sample tests as well as change-point tests exploiting the properties of these functional principal components. In particular the problem of an adequate choice of the dimension of the space to project on in order to reduce the dimension is addressed and included in the construction of the respective testing procedures.

To Lena,
my family
and those who are like family to me

Contents

Introduction

Technical advances and developments in computer sciences do not only influence and change our personal life in many aspects, they also change the infrastructure of scientific research. In all scientific disciplines an immensely increasing amount of data is collected, stored and made publically available through data bases that organize this vast amount of information. But not only the number and size of data sets is growing. The density and frequency of collected data and the diversity of these data sets is increasing as well.
For statistics this has various ramifications. One could even say that these developments change the face of modern statistics and bring with them new challenges for statistical research. Along with these challenges for statistics come such challenges for probability theory. The changing nature of data causes the development of novel stochastic models which open new fields to probability theory. In these models the need for statistical methods to investigate the data implies the need for adequate probability theoretical results which build the basis for the construction of these statistical methods.
In the sequel of this introduction we want to go more into detail how these ramifications influence the fields of change-point analysis and functional data analysis which form the scope of this thesis and thereby give a motivation for the results presented here. Since the references to specific literature can be found in the articles in the main part of this thesis we will restrict ourselves to references on more general and mainly introductory literature. In particular we would like to highlight some PhD theses which provide thorough reviews of certain important aspects of the areas of research relevant to this work.

Change-point analysis

The aim of change-point analysis is the detection of structural changes in stochastic models in general. Besides the situation where a data set is examined a posteriori and it should be determined statistically whether a change in the stochastic model underlying these data occurred, it is often of interest to monitor data on-line and decide with every new observation whether a change occurred.
The roots of this field go back to the work of W.A. Shewhart (starting in the 1920s) and E.S. Page (starting in the 1950) and were motivated by problems arising in quality control. Based on their results a wide theory with applications in most sciences developed. Differences in the respective problem settings lead to several lines of research that differ in the criteria underlying the construction of a decision rule. While the average run length (ARL) is a popular criterion in one of these lines of research another approach focusses on controlling the type one error and designs the testing procedure in a Neyman-Pearson fashion. In this thesis the latter approach will be used as a starting point. For an overview of these two lines of research and a comparison of the two approaches we refer to Koubková (2006).
With regard to the increasing sizes of data sets asymptotic statistical methods gain importance in many fields of applications. As implied by the title of this thesis all methods

presented here are of asymptotic nature. The monograph of Csörgő and Horváth (1997) is an indispensable reference in this context.
Many change-point procedures are built on invariance principles that allow to approximate the behavior of the partial sum of random varialbles $X_i,\ i = 1, 2, \ldots$, i.e.,

$$S_n = \sum_{i=1}^{n} X_i, \quad \text{for } n = 1, 2, \ldots \tag{1}$$

by a Wiener process. For an overview of results on independent, identically distributed random variables we refer to Aue (2003). But especially in the development of statistical procedures for dependent random variables such invariance principles play a key role. Examples for invariance principles for different dependence concepts are provided in Schmitz (2011), Wu (2007) or Berkes et al. (2011) just to name a few.
The statistics or detectors (which is the term used in sequential change-point analysis) given as a partial sum of (usually centered) random variables like (1) are called cumulative sum (CUSUM) statistics or detectors. In a sequential setup these procedures work well in so-called "early change scenarios" but perform weaker the later a change occurs. Since all observations of the monitoring period are used to calculate the cumulative sum it is intuitively clear that all observations before the change "disturb" the detection because they do not contribute to the drift that causes a reaction of the detector. From a theoretical point of view this is confirmed by results on the asymptotic distribution of the corresponding stopping times which could be given only under such early change scenarios (cf., e.g., Aue and Horváth (2004), Aue et al. (2009b) and Černíková et al. (2011)). This property has been the subject of critcism and alternative approaches like weighted cumulative sum or moving sum detectors have been proposed (cf. Kühn (2007)).
However the expression "CUSUM" was first introduced in the context of the work of E.S. Page (cf. Page (1954)) for a detector of the type

$$T_n = S_n - \min_{0 \le i \le n} S_i, \quad \text{with } S_0 = 0 \text{ and } S_j \text{ defined in (1) for } j \ge 1. \tag{2}$$

The idea behind this approach is to take out a part of the observations that is not likely to contribute to the drift introduced by the change and therefore achieve a higher "robustness" regarding the time of change. This idea initiated the work on this thesis and finally led to the results that can be found in the first two articles of the main section.
To highlight the wide applicability of the described methodology we conclude this section on change-point analysis by a list of some of the most famous fields of application.

EXAMPLE 1. *Quality control: As mentioned before one of the first applications of sequential or on-line change-point procedures was quality control where from the results of the production process it should be determined whether the production facility is still working correctly.*

EXAMPLE 2. *Climate changes: The influence of industrialization and environmental pollution on our climate is not only discussed in the scientific society but can as well be found in the media almost permanently. The need for statistical methods to investigate a possible change is therefore fairly obvious.*

EXAMPLE 3. *Success of environmental/health protection measures: After the introduction of measures to protect the environment and the health of people like the prohibition of certain vehicles in centers of large cities it is of interest to investigate if and when a consequence of these measures in terms of an improvement of, e.g., air pollution levels can be detected.*

EXAMPLE 4. *Changes in the parameters of economic/econometric models induced by a shock in the market: Not only the recent developments at the stock markets worldwide and their influence on politics and the citizens of many countries have shown that prizes of assets and risks derived from stochastic models have to be handled with care when shocks occur in the market. In particular for models which are supposed to include and transfer possible changes in their input to their output it is often not obvious whether these models remain valid after such a shock.*

EXAMPLE 5. *Medicine: Change-point methods can be applied to monitor the health conditions of a patient. In this case an alarm system is needed that calls the attention of the medical staff in case of an unusual behaviour of some indicator for the critical health condition.*

FUNCTIONAL DATA ANALYSIS

The field of functional data analysis is probably one of the fields of statistics that gained most importance with regard to the aforementioned technical developments and the resulting capability to store high-dimensional data sets. While before many data sets had to be reduced to key figures (e.g., daily, monthly ore yearly averages) because of the restrictions in the storage of data, nowadays storage is a minor issue. The challenge in statistics is therefore to develop powerful tools to analyze these high-dimensional data sets and exploit as much information inherent in the data as possible.
The term "functional data" can be described quite easily from a mathematical point of view. In the scope of this thesis (we can and will assume without loss of generality that) *functional data* are realizations of a random element in $L^2[0,1]$, the space of measurable, real-valued, square integrable functions on the interval $[0,1]$, equipped with the Borel σ-algebra. Yet from this definition it is not quite obvious where to find such data in practical applications and to see why this type of data gained importance in recent years. So to demonstrate the practical meaning of the term "functional data" we go back to some of the examples given above. Temperature measurements (cf. Example 2) or air pollution levels (cf. Example 3) are good examples for data that are obviously suitable to be modelled as functional data, i.e., to be viewed as a curve (e.g., on a daily basis). The same holds true for prices of assets which are traded in a high frequency at the market (cf. Example 4).

Pictures of continuous paths of stock market indices or the prizes of stocks can be seen on the news and in newspapers every day. But these are just some examples of a variety of applications. Besides those applications where the random process driving the data can obviously be modelled as such a function there exist many applications where the high dimension and nature of the data justifies the modelling as functional objects. For more examples of both types of applications we refer to the introductory parts of Horváth and Kokoszka (2012), Ramsay and Silverman (2005) and Ramsay et al. (2009).
From these applications it is evident that many statistical problems from the univariate or multivariate setting like, e.g., the change-point problem or the related two-sample problem that can be found in this thesis transfer to the functional setting. One concrete example where this is quite obvious is the problem of determining statistically whether there exists a difference in the mean of certain subsets of a data set that transfers to the problem of determining whether there exists a difference in the mean *function* of such subsets of the corresponding functional data set.
Since the space $L^2[0,1]$ is of infinite dimension most statistical approaches are based on dimension reduction techniques to derive statistical methodology in a finite dimensional subspace and make use of existing multivariate results from statistics as well as probability theory. The dimension reduction technique that probably plays the most important role is the functional principal component analysis. The idea of this technique is to project the data onto a subspace that is spanned by those functions of a certain basis that explain a major part of the variation of the data with the intention to capture the most important information for the statistical analysis. The results presented in this thesis also focus on functional principal components and show the construction of statistical methods using a fixed number of principal components as well as methods where the number may depend on the sample size. In particular the latter is of interest since the choice of the number of functional principal components was often carried out based on data-driven rules of thumb and was often criticized in this respect.

To conclude this introduction we will now give brief summaries of the articles of this thesis with the intent to connect the contents of the single articles with the motivational introduction given above.

Summaries of the articles included in this thesis

The thesis consists of four articles where the first two are not concerned with functional data but the development of procedures based on Page's CUSUM and the derivation of their asymptotic properties. The last two articles then consist of results on functional data illustrating the impact of functional principal component techniques on the development of statistical methods in this area.

Page's sequential procedure for change-point detection in time series regression
By Stefan Fremdt

In this article sequential open-end procedures to detect an abrupt change in the regression parameters of a linear model are presented. The assumptions of the model allow for certain time series dependencies for both regressors and error terms. The construction of the detectors is based on the idea of Page (1954) yet the design of the testing procedure as first crossing time of this detector and a given boundary function is novel. Besides a procedure that is built directly from the regression residuals further procedures are presented that are built from the squares of the regression residuals and allow for weaker assumptions on the change. To establish a stopping rule critical values are derived from the asymptotic distribution of these detectors under the null hypothesis of no change in the model parameters. The asymptotics are always with regard to the length of a training period prior to the beginning of the monitoring. This training period is assumed to have constant model parameters and the data from this period are used to estimate the necessary parameters. The proofs of the asymptotic behaviour of these procedures under the null hypothesis rely heavily on an invariance principle with certain rates for the error terms of the model. To show the validity of the proposed procedures the asymptotic consistency under the alternative is proven. In the empirical part of the article a simulation study underlines the properties of the constructed procedures. Finally a data set of portfolio theoretical background is used to show the applicability of the results. In this application a change in the model parameters in the context of the so-called "subprime-crisis" is detected.

Asymptotic distribution of the delay time in Page's sequential procedure
By Stefan Fremdt

While criteria like the average run length only provide a key figure to assess and compare change-point procedures, results on the asymptotic distribution of the stopping time of a change-point procedure provide far more information. In the present article the asymptotic distribution of the delay time of the procedure (built from the model residuals) from the previous article is derived in the special case of the so-called location model. Depending on the change-point different limit distributions are obtained which can be used as a benchmark to compare this procedure to other detection procedures. The presented results extend similar results for related change-point procedures in different aspects and theoretically confirm the desired properties described above. A simulation study illustrates the convergence empirically and provides information about the speed of convergence for the different limit distributions.

Testing the equality of covariance operators in functional samples
By Stefan Fremdt, Lajos Horváth, Piotr Kokoszka and Josef G. Steinebach

Besides the mean function the covariance structure of functional data represents one of the most interesting subjects to be investigated statistically. The main result of this article is a two-sample test for the equality of the covariance operator. Like already mentioned this test is based on the functional principal component analysis. In this case the analysis is carried out for the asymptotic empirical covariance operator of the overall sample. After the construction of a suitable test statistic a multivariate central limit theorem yields a chi square distribution as limit distribution under the null hypothesis. Again a simulation study illustrates the finite sample behaviour of the procedure. In a data example the test is applied to egg-laying trajectories of a certain fruit fly species where it is of interest whether the egg-laying behaviour of this species depends on the life span of the fly.

Functional data analysis with increasing number of projections
By Stefan Fremdt, Lajos Horváth, Piotr Kokoszka and Josef G. Steinebach

In functional data analysis many methods are based on the projection of the data in the direction of a certain number of functional principal components. Yet the choice of this number was mostly left to the practitioner. Data-driven rules of thumb were suggested but the need to develop procedures that handle this problem on a solid theoretical basis is apparent. In this article a uniform normal approximation for the partial sum process of the projections is presented that can be used as a starting point for the construction of a variety of statistical procedures. Exemplary change-point procedures to test the constancy of the mean function in a given sample and a two-sample test for the equality of the mean functions are developed. These change-point procedures are constructed on the basis of a limit theorem for a two-parameter process derived from the aforementioned partial sum process that is proven in the scope of the article. In the empirical part the performance of the presented procedures is investigated in a simulation study. Here the focus is directed on a change-point procedure that uses the convergence of the two-parameter process in its full force. Finally this change-point test is applied to a data set of yearly temperature curves from a weather station in Melbourne, Australia.

Some remarks on the organization of the dissertation:

Since the dissertation is of cumulative form and hence a collection of articles, the main part of the thesis consists of these articles. They will be given in the order the work on the respective project was initiated. Thus the order is not intended to assess the contribution of those results. After the articles a discussion of the combined results of the articles concludes the dissertation and gives an outlook on prospective research and new challenges arising from the obtained results. The references given in an article can be found at the end of this article. The supplementary references used outside the articles (i.e., in this introduction and the concluding discussion) can be found at the end of the thesis.

Acknowledgements. Like many others I want to use this occasion to thank those people who supported me during the work on this thesis. Besides the support that is clearly related to work I also want to express my thanks to my family and friends. They helped me to remember that sometimes, to make a step forward it can be necessary to first take a step back and look at a problem from a different perspective.
Evidently the person I want to thank first is my supervisor Prof. Dr. Josef G. Steinebach. Not only did he give me the opportunity to write this thesis, he supported me in all my plans and projects and never gave me a reason to doubt this support. His experience and knowledge in our field and the research environment he created encouraged my own work and helped me in critical moments to take the next step. Special thanks also go to Prof. Dr. Lajos Horváth who is partly responsible for my growing addiction to coffee but also played an important role in finding a direction for my research. I am grateful for his advice, for letting me share his deep insight in our field (not only) during our joint projects and many "fruitful" and funny conversations. Besides these two persons I would like to mention a number of other people I met in the course of my work who supported me in one way or the other. I thank Prof. Dr. Piotr Kokoszka and Prof. Dr. Alexander Aue for the cooperation in our joint projects, Prof. Dr. Siegfried Hörmann, Prof. Dr. Marie Hušková and Prof. Dr. Zuzana Prášková for the opportunity to visit them and for showing so much interest in my work. Furthermore I thank the stochastics group at University of Cologne who accompanied me for the last three years, in particular Prof. Dr. Wolfgang Wefelmeyer who agreed to be a referee for this thesis.
At last (but of course not at least) I am deeply grateful to the people who are closest to me for their patience and for always being there for me. Thank you!

PAGE'S SEQUENTIAL PROCEDURE FOR CHANGE-POINT DETECTION IN TIME SERIES REGRESSION

By Stefan Fremdt[‡]

[‡] *University of Cologne*

Abstract

Cumulative sum (CUSUM) procedures have been applied for the sequential detection of structural breaks in stochastic models in a variety of different settings. Yet their performance depends strongly on the time of change and is best under early-change scenarios. For later changes their finite sample behaviour is rather questionable. We therefore propose modified CUSUM procedures for the detection of abrupt changes in the regression parameter of multiple time series regression models that show a higher stability regarding the time of change than ordinary CUSUM procedures. The asymptotic distributions of the test statistics and the consistency of the procedures are provided. In a simulation study it is shown that the proposed procedures behave well in finite samples. Finally the procedures are applied to a set of capital asset pricing data related to the Fama-French extension of the capital asset pricing model.

Keywords: CUSUM, Linear model, Change-point, Sequential test, Asymptotic distribution, Invariance principle, CAPM, Fama-French model.

AMS subject classification: Primary 62J05; secondary 62L99

1 Introduction

The recent worldwide economical developments have shown again that shocks in financial markets can lead to mispricing of assets and risks due to structural changes in the underlying valuation models. As a consequence, there is a need to reliably monitor the validity of these models. Many of the widely used approaches to pricing of assets are based on regression models describing the linear relationship between the asset price and factors that explain a major part of its variation. Examples for these approaches include the famous and still widely applied capital asset pricing model (CAPM) of Sharpe (1964) and Lintner

(1965) and its extension proposed by Fama and French (1993) which will be investigated in Section 4. In contrast to the one-factor CAPM this multifactor extension of Fama and French (1993) uses two factors in addition to the market excess returns to explain a higher proportion of the variation of the asset price.

In the literature the change-point problem for linear models has been discussed extensively. While most of the contributions are made from an a-posteriori point of view (we refer to, e.g., Bai (1997), Perron (2006) and Csörgő and Horváth (1997)), recently the sequential or on-line change-point detection has received more and more attention. Antoch and Jarušková (2002) give a bibliographical overview of the field of on-line statistical process control. The basis for this work is given in the articles of Chu et al. (1996), Horváth et al. (2004) and Aue et al. (2006b) who suggest cumulative sum (CUSUM) procedures in different stochastic models. CUSUM procedures work best for relatively early changes but show a slower reaction the later the change occurs. Aue et al. (2009) provided the asymptotic normality of the suitably normalized stopping time of the CUSUM procedure in a similar setting as will be considered in this work but only in a relatively small range after the start of the monitoring. The procedures that will be developed here found on an idea of Page (1954) and should give a higher stability towards the time of change. Other approaches that tackle this task are so called moving sum (MOSUM) procedures that were studied by, e.g., Aue et al. (2008) and Chu et al. (1995). Their drawback is a strong dependence on the choice of the parameters, in particular the right choice of the window size by the statistician.

For the applicability to financial problem settings we want to explicitly allow certain dependencies, i.e. we will include many of the commonly applied time series models for the error terms as well as for the regressors in our setting. Other contributions assuming dependencies are given by, e.g., Schmitz and Steinebach (2010) who considered strongly mixing error terms in a linear model or Hušková et al. (2007) who studied autoregressive time series in a closed-end setting.

The paper is organized as follows. In Section 2 the linear model and the underlying assumptions are introduced. Section 3 contains the definition of the detectors and stopping times as well as the results about the asymptotic distribution under the null hypothesis and the asymptotic consistency of the procedures. In Section 4 we will present a simulation study and the results of an application of the procedures to the aforementioned Fama-French model. We conclude the paper with the proofs of Section 3 in Section 5.

2 Model description and assumptions

Consider the linear model:

$$y_i = \mathbf{x}_i^T \boldsymbol{\beta}_i + \varepsilon_i, \quad 1 \leq i < \infty, \tag{2.1}$$

where $\mathbf{x}_i$ is a $p \times 1$ random vector and $\boldsymbol{\beta}_i \in \mathbb{R}^p$.
We assume that for the first m observations the so-called "noncontamination assumption" (cf. Chu et al. (1996)) holds, i.e.

$$\boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad 1 \leq i \leq m. \tag{2.2}$$

As mentioned before the constancy of the regression parameters $\boldsymbol{\beta}_i$ in time should be tested which leads to the null hypothesis

$$H_0: \quad \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad i = m+1, m+2, \ldots.$$

We consider alternatives of one abrupt change in the regression parameter at an unknown change-point, i.e.

$$\begin{aligned} H_A: \quad & \text{there is } k^* \geq 1 \text{ such that } \boldsymbol{\beta}_i = \boldsymbol{\beta}_0, \quad m < i < m + k^* \\ & \text{and } \boldsymbol{\beta}_i = \boldsymbol{\beta}_*, \quad i = m + k^*, m + k^* + 1, \ldots \quad \text{with } \boldsymbol{\Delta} = \boldsymbol{\beta}_* - \boldsymbol{\beta}_0 \neq \mathbf{0}. \end{aligned}$$

The detection procedures will consist of stopping times $\tau(m)$ (to be defined in detail in Section 3 of this article) chosen in such a way that under the null hypothesis:

$$\lim_{m \to \infty} P(\tau(m) < \infty) = \alpha, \quad 0 < \alpha < 1 \tag{2.3}$$

and under the alternative

$$\lim_{m \to \infty} P(\tau(m) < \infty) = 1. \tag{2.4}$$

We assume the following conditions on the regressors and the error terms

$\{\mathbf{x}_i\}$ be a stationary sequence. **(A.1)**

$\mathbf{x}_i^T = (1, x_{2i}, \ldots, x_{pi}), \quad 1 \leq i < \infty,$ **(A.2)**

There exist a p-dimensional vector $\boldsymbol{d} = (d_1, \ldots, d_p)^T$ and constants $K > 0,\ \nu > 2$ such that

$$\mathrm{E}\left|\sum_{i=1}^{k}(x_{i,j} - d_j)\right|^{\nu} \leq K\, k^{\nu/2}, \quad 1 \leq j \leq p. \tag{A.3}$$

$\{\varepsilon_i,\ 1 \le i < \infty\}$ and $\{\mathbf{x}_i,\ 1 \le i < \infty\}$ are independent. **(A.4)**

For every m there are a constant $\sigma > 0$ and independent Wiener processes **(A.5)**
$\{W_{1,m}(t) : t \ge 0\}$ and $\{W_{0,m}(t) : t \ge 0\}$ such that

$$\sup_{1\le k<\infty} \frac{1}{k^\xi} \left| \sum_{i=m+1}^{m+k} \varepsilon_i - \sigma W_{1,m}(k) \right| = \mathcal{O}_P(1) \quad (m \to \infty) \tag{2.5}$$

and

$$\sum_{\ell=1}^{m} \varepsilon_\ell - \sigma W_{2,m}(m) = \mathcal{O}_P\left(m^\xi\right) \quad (m \to \infty), \tag{2.6}$$

with some $\xi < 1/2$.

The above stated assumptions on the regressors and error terms are satisfied for a variety of important stochastic models. For examples we refer to Aue et al. (2009) who showed that **(A.1)** and **(A.3)** are satisfied for, e.g., i.i.d. sequences, linear processes or augmented GARCH sequences. The latter were introduced by Duan (1997) and include most of the conditionally heteroskedastic models used in practice. For a collection of examples belonging to this class we suggest the papers of Aue et al. (2006a) and Carrasco and Chen (2002). Concerning the error terms Aue et al. (2006b) provided the proof of **(A.5)** again for augmented GARCH sequences under appropriate assumptions, Aue and Horváth (2004) give further examples, besides the i.i.d. case, including martingale difference sequences and stationary mixing sequences.

All procedures treated in this work are based on the behaviour of the residuals of the model

$$\hat{\varepsilon}_i = y_i - \mathbf{x}_i^T \widehat{\boldsymbol{\beta}}_m, \quad i = 1, 2, \ldots,$$

where $\widehat{\boldsymbol{\beta}}_m$ denotes the OLSE for $\boldsymbol{\beta}$ from the data $(y_1, \mathbf{x}_1), \ldots, (y_m, \mathbf{x}_m)$, i.e.

$$\widehat{\boldsymbol{\beta}}_m = \left(\sum_{1\le i\le m} \mathbf{x}_i \mathbf{x}_i^T \right)^{-1} \sum_{1\le j\le m} \mathbf{x}_j y_j.$$

In the sequel we will by $\hat{\sigma}_m$ denote a weakly consistent estimator for the parameter σ from Assumption **(A.5)**. The estimation of this parameter will be discussed later in detail.

3 Sequential testing procedures and asymptotic results

Many sequential detection procedures in the literature are constructed as first passage times of a so called detector over a certain boundary function. For example Horváth et al. (2004) proposed as a detector the (ordinary) CUSUM of the residuals, i.e.

$$\widehat{Q}(m,k) = \sum_{m<i\leq m+k} \hat{\varepsilon}_i, \quad k = 1,2,\ldots, \quad \text{and} \quad \widehat{Q}(m,0) = 0,$$

and as a boundary function

$$h_{\alpha,\gamma}(m,k) = c\, g(m,k) = c\, m^{1/2}\left(1+\frac{k}{m}\right)\left(\frac{k}{k+m}\right)^{\gamma}, \tag{3.7}$$

with

$$0 \leq \gamma < 1/2 \tag{3.8}$$

and $c = c(\alpha,\gamma)$ such that (2.3) holds. The first procedure we want to introduce goes back to an idea of Page (1954) and we define the detector

$$\widehat{Q}_P(m,k) = \max_{0\leq i\leq k}\left|\widehat{Q}(m,k) - \widehat{Q}(m,i)\right| = \max\left\{\widehat{Q}_P^u(m,k), \widehat{Q}_P^d(m,k)\right\}, \tag{3.9}$$

where

$$\widehat{Q}_P^u(m,k) = \widehat{Q}(m,k) - \min_{0\leq i\leq k}\widehat{Q}(m,i) \quad \text{and}$$
$$\widehat{Q}_P^d(m,k) = \max_{0\leq i\leq k}\widehat{Q}(m,i) - \widehat{Q}(m,k).$$

The corresponding stopping time is then given by

$$\tau_{\alpha,\gamma}^{\text{Page}}(m) = \inf\left\{k \geq 1 : \widehat{Q}_P(m,k) > h_{\alpha,\gamma}(m,k)\right\}$$

where $\inf\emptyset = \infty$ and the constant $c = c(\alpha,\gamma)$ in the definition of $h_{\alpha,\gamma}$ can be derived from Theorem 3.1 below.

THEOREM 3.1. *Assume that* (2.2), **(A.1)** – **(A.5)** *and* (3.8) *hold. Then under the null hypothesis we have, for* $c \in \mathbb{R}$ *and a Wiener process* $\{W(t) : t \geq 0\}$,

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

Page (1954) proposed a detector of the type $\widehat{Q}_P^u(m,k)$ for one-sided change-in-the-mean alternatives, in the case of a linear model this detector is appropriate for alternatives with $\boldsymbol{\Delta}^T\boldsymbol{d} > 0$, where the vector $\boldsymbol{d}$ was introduced in **(A.3)**. The corresponding asymptotic result under the null hypothesis for these one-sided detectors is given in Theorem 3.2.

THEOREM 3.2. *Assume that* (2.2), **(A.1)** – **(A.5)** *and* (3.8) *hold. Then under the null hypothesis we have, for* $c \in \mathbb{R}$ *and a Wiener process* $\{W(t) : t \geq 0\}$,

$$\begin{aligned} &\lim_{m\to\infty} P\left(\frac{1}{\hat{\sigma}_m}\sup_{1\leq k<\infty}\frac{\widehat{Q}_P^u(m,k)}{g(m,k)} \leq c\right) \\ &= \lim_{m\to\infty} P\left(\frac{1}{\hat{\sigma}_m}\sup_{1\leq k<\infty}\frac{\widehat{Q}_P^d(m,k)}{g(m,k)} \leq c\right) \\ &= P\left(\sup_{0<t<1}\frac{1}{t^\gamma}\left(W(t) - \inf_{0\leq s\leq t}\frac{1-t}{1-s}W(s)\right) \leq c\right). \end{aligned}$$

From this result again the critical value $c(\alpha,\gamma)$ can be derived for the two one-sided detectors. We will denote this critical value by $c_1 = c_1(\alpha,\gamma)$ and for the two-sided detector by $c_2 = c_2(\alpha,\gamma)$. Under the alternative hypothesis the detectors diverge as the following theorem shows.

THEOREM 3.3. *Assume that* (2.2), **(A.1)** – **(A.5)** *and* (3.8) *hold.*

a) Then under H_A *and if* $\mathbf{d}^T\boldsymbol{\Delta} > 0$ *we have*

$$\frac{1}{\hat{\sigma}_m}\sup_{1\leq k<\infty}\frac{\widehat{Q}_P^u(m,k)}{g(m,k)} \xrightarrow{P} \infty \quad \text{as } m\to\infty,$$

b) Then under H_A *and if* $\mathbf{d}^T\boldsymbol{\Delta} < 0$ *we have*

$$\frac{1}{\hat{\sigma}_m}\sup_{1\leq k<\infty}\frac{\widehat{Q}_P^d(m,k)}{g(m,k)} \xrightarrow{P} \infty \quad \text{as } m\to\infty,$$

c) Then under H_A *and if* $\mathbf{d}^T\boldsymbol{\Delta} \neq 0$ *we have*

$$\frac{1}{\hat{\sigma}_m}\sup_{1\leq k<\infty}\frac{\widehat{Q}_P(m,k)}{g(m,k)} \xrightarrow{P} \infty \quad \text{as } m\to\infty.$$

Theorem 3.3 gives a sufficient condition that guarantees (2.4). In Section 4 tables with simulated critical values for selected values of α and γ can be found for the functionals

$$\sup_{0<t<1} \frac{1}{t^\gamma}\left(W(t) - \inf_{0\le s\le t} \frac{1-t}{1-s} W(s)\right) \qquad \text{and} \qquad \sup_{0<t<1}\sup_{0\le s\le t} \frac{1}{t^\gamma}\left| W(t) - \frac{1-t}{1-s}W(s)\right|.$$

The additional assumptions on the amount of change (i.e. $\boldsymbol{d}^T\boldsymbol{\Delta} \gtrless 0$ resp. $\boldsymbol{d}^T\boldsymbol{\Delta} \neq 0$) for the above developed procedures that guarantee their consistency are quite restrictive. Yet under additional assumptions on the error terms we can modify the presented procedures which allows to drop these assumptions on the amount of change. In this context we want to refer to the work of Hušková and Koubková (2005) who with the same intention developed monitoring procedures based on quadratic forms of weighted cumulative sums. We define the detectors based on the sum of squares of the residuals

$$S_P(m,k) = \max_{0\le i\le k} |S_R(m,k) - S_R(m,i)| \quad \text{and} \quad S_P^u(m,k) = \max_{0\le i\le k}\left(S_R(m,k) - S_R(m,i)\right),$$

where

$$S_R(m,k) = \sum_{i=m+1}^{m+k} \hat\varepsilon_i^2 - \frac{k}{m}\sum_{\ell=1}^{m} \hat\varepsilon_\ell^2, \quad k = 1, 2, \ldots.$$

Aue et al. (2006b) showed a similar result based on the squared prediction errors using the additional assumptions

$$E\varepsilon_i^2 = \sigma^2, \quad 0 < \kappa = E\varepsilon_i^4 < \infty \quad (i \ge 1), \tag{3.10}$$

$$\eta^2 = \operatorname{Var}(\varepsilon_0^2) + 2\sum_{i=1}^{\infty} \operatorname{Cov}(\varepsilon_0^2, \varepsilon_i^2) > 0. \tag{3.11}$$

Furthermore they assumed that for every m there exist independent Wiener processes

$$\{W_{3,m}(t) : t \ge 0\} \quad \text{and} \quad \{W_{4,m}(t) : t \ge 0\}$$

such that

$$\sup_{1\le k<\infty} \frac{1}{k^\zeta}\left|\sum_{i=m+1}^{m+k} (\varepsilon_i^2 - \sigma^2) - \eta W_{3,m}(k)\right| = \mathcal{O}_P(1) \quad (m\to\infty) \tag{3.12}$$

and

$$\sum_{i=1}^{m} (\varepsilon_i^2 - \sigma^2) - \eta W_{4,m}(m) = \mathcal{O}_P\left(m^\zeta\right) \quad (m\to\infty) \tag{3.13}$$

with some $\zeta < 1/2$ and η from (3.11). Combining the techniques of Aue et al. (2006b) and from the proof of Theorem 3.1 it is obvious that similar asymptotic results hold for these procedures:

THEOREM 3.4. *Assume that* (2.2), **(A.1)** – **(A.4)**, (3.8) *and* (3.10) – (3.13) *hold. Then under the null hypothesis we have, for a real number c and a Wiener process $\{W(t) : t \geq 0\}$,*

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

and

$$\lim_{m\to\infty} P\left(\frac{1}{\eta}\sup_{1\leq k<\infty}\frac{S_P(m,k)}{g(m,k)} \leq c\right) = P\left(\sup_{0<t<1}\sup_{0\leq s\leq t}\frac{1}{t^\gamma}\left|W(t) - \frac{1-t}{1-s}W(s)\right| \leq c\right).$$

The parameter η in the statement of Theorem 3.4 can be replaced by a weakly consistent estimator $\hat{\eta}_m$. Aue et al. (2006b) pointed out that the Bartlett estimator $\hat{\eta}^2_{B,m}$ for η^2 under the conditions of Theorem 3.4 satisfies $\hat{\eta}^2_{B,m} \xrightarrow{P} \eta^2$ and can therefore be applied in the general setting of this section. The same arguments hold for the estimation of σ. However it should be noted that the quality of the estimators affects the finite sample behaviour of the procedures. This will be discussed in Section 4.
Under the alternative hypothesis without additional assumptions on the amount of the change we again have the desired divergence.

THEOREM 3.5. *Assume that* (2.2), **(A.1)** – **(A.4)**, (3.8) *and* (3.10) – (3.13) *hold. Then under H_A we have*

$$\frac{1}{\hat{\eta}_m}\sup_{1\leq k<\infty}\frac{|S_R(m,k)|}{g(m,k)} \xrightarrow{P} \infty \quad \text{as } m\to\infty$$

and

$$\frac{1}{\hat{\eta}_m}\sup_{1\leq k<\infty}\frac{S_P(m,k)}{g(m,k)} \xrightarrow{P} \infty \quad \text{as } m\to\infty.$$

Analogous results to those of Theorems 3.4 (with the corresponding limit distributions from Theorem 3.2) and 3.5 hold for the detectors S_P^u (and S_R). However as we will see in Section 4 these show a poorer finite sample behaviour than the detectors S_P and $|S_R|$.
One drawback of the detectors S_P, $|S_R|$, S_P^u and S_R is that the assumption on the existence

of a constant σ is crucial to the testing procedure. It can be seen easily that, due to its construction, the procedure is also sensitive towards changes in σ, i.e. in case of constant $\boldsymbol{\beta}_i$ but a change in $\sigma(=\sigma_i)$ the testing procedure would decide that there has been a change in the $\boldsymbol{\beta}_i$ with probability one. This sensitivity exists as well for the further introduced procedures, although in a weaker sense, i.e. in the derivation of the critical values which is also strongly dependent on the assumption of a constant σ. But since in general practitioners are concerned with the validity of their underlying model, the detection of a switch in the regime, including $\boldsymbol{\beta}_i$ as well as σ, is of great interest to them.

4 SIMULATIONS AND APPLICATION TO ASSET PRICING DATA

In this section the results of a simulation study are presented that was performed to confirm the theoretical results from Section 3. Furthermore it should show that the proposed monitoring procedures have the desired properties. With regard to the application to the Fama-French model and its financial context the carried out simulations will focus on GARCH regressors. We will first consider the asymptotic results from Section 3 and provide the empirical sizes under the null hypothesis. A comparison of the detection properties of the different procedures in finite samples concludes the simulation study and highlights the advantages of the newly developed sequential tests. The last part of this section will then contain the results of an application of our monitoring procedures to a data set made publically available by Kenneth R. French on his website (cf. French (2011)).
To establish (2.3) for the suggested procedures it is necessary to determine the critical values from the definition of $h_{\alpha,\gamma}$ in (3.7) using the statements of Theorems 3.1, 3.2 and 3.4. The critical values $c_1(\gamma,\alpha)$ and $c_2(\gamma,\alpha)$ for the functionals

$$\sup_{0<t<1} \frac{1}{t^\gamma}\left(W(t) - \inf_{0\le s\le t} \frac{1-t}{1-s} W(s)\right) \qquad \text{and} \qquad \sup_{0<t<1}\sup_{0\le s\le t} \frac{1}{t^\gamma}\left|W(t) - \frac{1-t}{1-s}W(s)\right|,$$

for selected values of α and γ, can be found in Table 1 and Table 2, respectively. These were simulated with 100,000 replications of an approximation of a Wiener process generated on a grid of 100,000 points. Horváth et al. (2004) provided the simulated critical values for the functional $\sup_{0<t<1}|W(t)|/t^\gamma$. For $\gamma = 0$ we calculated these critical values numerically using the series representation

$$P\left(\sup_{0<t<1}|W(t)|\leq c\right)=\frac{4}{\pi}\sum_{k=0}^{\infty}\frac{(-1)^k}{2k+1}\exp\left(-\pi^2(2k+1)^2/8c^2\right),$$

from, e.g., Csörgő and Révész (1981), Theorem 1.5.1, to find:

α	0.010	0.025	0.050	0.100	0.250
$c(0,\alpha)$	2.8070	2.4977	2.2414	1.9600	1.5341

4.1 Simulation results

The simulations were performed for a selection of the above mentioned models satisfying our assumptions (cf. Aue et al. (2009), Aue and Horváth (2004)), but since all gave similar results, we only present the results for our model (2.1) with $p = 2$, $x_{2,i}$ according to a GARCH(1,1) model and independent normally distributed errors ε_i with $\sigma^2 = 0.5$ (in this specification $\sigma = \eta$ to achieve a better comparability of the procedures based on ordinary and squared residuals under the alternative). We followed Aue et al. (2009) and chose the specification of the GARCH(1,1) model as

$$x_{2,i} = d_2 + \bar{\sigma}_i z_i, \text{ with } \bar{\sigma} \text{ given as solution of } \bar{\sigma}_i^2 = \bar{\omega} + \bar{\alpha} z_{i-1}^2 + \bar{\beta}\bar{\sigma}_{i-1}^2,$$

where $\{z_i\}$ are iid standard normally distributed and $(\bar{\omega}, \bar{\alpha}, \bar{\beta}) = (0.5, 0.2, 0.3)$. From the decomposition (5.24) in the proof of Theorem 3.3 and a similar decomposition for the procedure based on the squared residuals we find that for this model the drift in case of a change is determined by $\boldsymbol{d}^T\boldsymbol{\Delta}$ for the ordinary residuals and for the squared residuals (asymptotically) via $\Delta_2^2 + (\boldsymbol{d}^T\boldsymbol{\Delta})^2$. For the simulations we chose $d_2 = 1$. Due to the uncorrelated error terms in this model the OLSE for the parameter σ from Assumption **(A.5)**, i.e., $\sqrt{\hat{\sigma}_m^2} = \left(\frac{1}{m-p}\sum_{i=1}^m\left(\hat{\varepsilon}_i - \frac{1}{m}\sum_{\ell=1}^m \hat{\varepsilon}_\ell\right)^2\right)^{1/2}$, and the corresponding estimator for η can be utilized. As mentioned above in the general setting of this paper the Bartlett estimator is a consistent estimator in the case of correlated error terms. However simulations have shown that due to a slower convergence of the estimator, size distortions can be observed under the null hypothesis. Consequently larger training samples are needed to achieve satisfying results.

The length of the training period m was chosen as $m = 100, 200, 500$ and 1000, the number of replications as 5000. For the tuning parameter γ the values were set to

	α				
γ	0.010	0.025	0.050	0.100	0.250
0.00	2.5955	2.2564	1.9897	1.6924	1.2474
0.15	2.6632	2.3341	2.0757	1.7915	1.3671
0.25	2.7372	2.4206	2.1686	1.8992	1.4887
0.35	2.8691	2.5684	2.3273	2.0757	1.6817
0.45	3.1712	2.9224	2.6976	2.4592	2.0932
0.49	3.5385	3.2791	3.0640	2.8225	2.4391

Table 1: Critical values $c_1 = c_1(\gamma, \alpha)$ simulated on a grid of 100,000 points with 100,000 replications.

	α				
γ	0.010	0.025	0.050	0.100	0.250
0.00	2.8262	2.5188	2.2599	1.9914	1.5918
0.15	2.8925	2.5925	2.3416	2.0803	1.6976
0.25	2.9638	2.6707	2.4296	2.1758	1.8063
0.35	3.0857	2.8041	2.5758	2.3339	1.9839
0.45	3.3817	3.1259	2.9241	2.7002	2.3685
0.49	3.7357	3.4903	3.2848	3.0603	2.7178

Table 2: Critical values $c_2 = c_2(\gamma, \alpha)$ simulated on a grid of 100,000 points with 100,000 replications.

	m	$\gamma = 0$ $\alpha = 0.05$	$\gamma = 0$ $\alpha = 0.1$	$\gamma = 0.25$ $\alpha = 0.05$	$\gamma = 0.25$ $\alpha = 0.1$	$\gamma = 0.49$ $\alpha = 0.05$	$\gamma = 0.49$ $\alpha = 0.1$
$\widehat{Q}_P$	100	0.0298	0.0698	0.0390	0.0802	0.0168	0.0334
	200	0.0294	0.0646	0.0364	0.0760	0.0194	0.0382
	500	0.0286	0.0682	0.0368	0.0812	0.0248	0.0470
	1000	0.0300	0.0704	0.0408	0.0842	0.0272	0.0554
$\widehat{Q}_P^u$	100	0.0354	0.0770	0.0438	0.0818	0.0166	0.0364
	200	0.0338	0.0720	0.0418	0.0806	0.0194	0.0382
	500	0.0342	0.0752	0.0430	0.0880	0.0256	0.0466
	1000	0.0350	0.0766	0.0432	0.0852	0.0256	0.0510

Table 3: Empirical sizes of the Page CUSUM procedures for 5000 replications with a monitoring horizon of $N = 5m$.

$\gamma = 0.00, 0.25, 0.49$.

Table 3 shows the empirical sizes of the testing procedures based on the detectors $\widehat{Q}_P$ and $\widehat{Q}_P^u$ under the null hypothesis with $\boldsymbol{\beta}_0 = (1, 1)^T$ taking $N = 5m$ observations after the end of the training period. It can be seen that for all parameter combinations the sizes remain conservative for short as well as long training periods. A similar behaviour was observed for the procedures based on the ordinary CUSUM and the corresponding results are therefore ommitted here.

The conservative nature of the empirical sizes from Table 3 cannot be found for the procedures based on the squared residuals. In Table 4 the corresponding empirical sizes are displayed which show a reasonable behaviour for small values of γ. With increasing γ the size of the training period has to increase as well to find satisfactory results. This can again be explained by the estimation error for the parameter η and the higher sensitivity of the boundary functions at the beginning of the monitoring for larger values of γ. For γ close to 1/2 the empirical sizes exceed the significance levels even for the larger sample sizes. This effect of a slower convergence should be taken into account by practitioners choosing the value of γ and an adaptation of the procedure to include the variation of the estimator for small samples may be considered. The detectors $\widehat{S}_P$ and $|\widehat{S}_R|$ show a nicer behaviour

		$\gamma = 0$		$\gamma = 0.25$		$\gamma = 0.49$	
	m	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$	$\alpha = 0.05$	$\alpha = 0.1$
$\widehat{S}_P$	100	0.0890	0.1358	0.1072	0.1624	0.1130	0.1408
	200	0.0598	0.0976	0.0718	0.1210	0.0912	0.1190
	500	0.0416	0.0804	0.0548	0.0998	0.0826	0.1152
	1000	0.0392	0.0794	0.0520	0.0958	0.0856	0.1178
$\lvert\widehat{S}_R\rvert$	100	0.0826	0.1246	0.1014	0.1504	0.1122	0.1408
	200	0.0550	0.0888	0.0676	0.1118	0.0898	0.1176
	500	0.0396	0.0756	0.0526	0.0916	0.0784	0.1112
	1000	0.0358	0.0748	0.0488	0.0908	0.0788	0.1152
$\widehat{S}_P^u$	100	0.1292	0.1932	0.1570	0.2172	0.1404	0.1836
	200	0.0898	0.1516	0.1142	0.1814	0.1174	0.1560
	500	0.0676	0.1222	0.0852	0.1416	0.1124	0.1514
	1000	0.0604	0.1094	0.0760	0.1326	0.1116	0.1578
$\widehat{S}_R$	100	0.1196	0.1832	0.1452	0.2082	0.1410	0.1812
	200	0.0812	0.1426	0.1066	0.1702	0.1162	0.1574
	500	0.0636	0.1178	0.0792	0.1334	0.1100	0.1492
	1000	0.0566	0.1016	0.0708	0.1222	0.1074	0.1520

Table 4: Empirical sizes of the procedures based on the squared residuals for 5000 replications with a monitoring horizon of $N = 5m$.

for small samples compared to $\widehat{S}_P^u$ and $\widehat{S}_R$ (which once more is due to the estimation error mentioned above). On the other hand we will see later that these procedures provide better behaviour regarding the speed of detection.

To investigate the behaviour of the proposed procedures under the alternative hypothesis extensive simulations were performed for a collection of different parameter settings. We will therefore again only give a selection of the obtained results. Since we are interested mainly in the comparison of the speed of detection of the Page CUSUM procedures with the ordinary CUSUM procedures we will comment only briefly on the power properties of the proposed procedures. The question whether a change is detected by these procedures in this open-end setting is not as interesting with regard to the comparison of ordinary and Page CUSUM. To explain this we again refer to the construction of the procedures. Since the drift induced by a change is similar for both types of procedures and the boundary functions only differ by a constant with an infinite monitoring horizon the power will as well be similar for both types of procedures. The results of our simulations confirm this and in this matter we refer to the literature on ordinary CUSUM procedures. We will therefore continue with the comparison of the speed of detection.

Changes occurring at $k^* = 1, m, 5m$ were considered and the monitoring was terminated at the latest after $N = k^* + 2000$ observations (which guarantees the detection of the change in all cases). The model setting under the null hypothesis described above was used and with regard to Theorems 3.3 and 3.5 we chose two types of changes, $\boldsymbol{\Delta}_1 = (0, 0.5)^T$ and $\boldsymbol{\Delta}_2 = (-0.8, 0.8)^T$, and will denote the corresponding alternative hypothesis by H_1 and H_2. With the specification of H_1 the above mentioned drift terms for ordinary and squared residuals are equal and a better comparability of these procedures is achieved. H_2 was chosen to satisfy $\boldsymbol{d}^T\boldsymbol{\Delta} = 0$ and therefore shows that the procedures based on squared residuals perform well in this case while the procedures based on ordinary residuals are not able to detect the change. However the differences in the performance and applicability of the testing procedures based on ordinary residuals and those based on squared residuals should also be discussed briefly. We want to make clear that the performance strongly depends on the amount of change. For example due to their construction it can be seen from the respective drift terms that the procedures based on quadratic residuals show a slower reaction under slight changes with $\boldsymbol{d}^T\boldsymbol{\Delta} \neq 0$ than the procedures based on ordinary residuals whereas under larger changes for the same reason the opposite is true. In addition the influence of the parameters σ and η on the drift has to be taken into account. Depending on the application in practice a combination of the two procedures may be considered to balance the advantages and disadvantages of the two types of procedures.

We now want to illustrate that the procedures based on the Page CUSUM show a higher stability regarding the time of change than those based on the ordinary CUSUM. Additionally the influence of the tuning parameter γ on the speed of detection should be examined. Figures 1 and 2 show density estimations of the delay times (excluding false alarms) under the alternative H_1 for $\alpha = 0.1$. In Figure 1 a training period of length $m = 200$ was used, in Figure 2 the length was set to $m = 1000$. The rows correspond from top to bottom to very early ($k^* = 1$), intermediate ($k^* = m$) and late ($k^* = 5m$) changes. The left columns show the density estimates for $\widehat{Q}_P$ (red) and $|\widehat{Q}|$ (blue), the right columns show the estimates for $\widehat{S}_P$ (red) and $|\widehat{S}_R|$ (blue), in both columns for the different values of γ. Tables containing the five number summaries of the data used for the density estimation can be found in the appendix.
The density estimates show clearly that for a change immediately after the end of the training period, as could be expected, there is only a slight difference between the procedures based on Page's CUSUM and those based on the ordinary CUSUM. In this case a choice of γ close to $1/2$ delivers the best results. For intermediate changes it is already obvious that the Page CUSUM procedures show a better behaviour than the ordinary CUSUM procedures for both ordinary and squared residuals. This effect is getting stronger the later the change happens as can be seen in the bottom rows. For intermediate changes a choice of $\gamma = 0.25$ gave the best results, for late changes $\gamma = 0$ is the appropriate choice. This observation which reflects the intention of the parameter γ has already been discussed in, e.g., Horváth et al. (2004).
As mentioned before the procedures based on ordinary residuals are not applicable under the alternative H_2. We will therefore only present the density estimates for the procedures based on squared residuals which can be found in Figure 3. The obtained results are similar to the results under H_1 regarding the comparison of Page and ordinary cumulative sums for all sample sizes. Therefore we only present these for $m = 1000$ and $\gamma = 0$ (where a reasonable behaviour under the null hypothesis for all detectors was observed). The density estimates show that the detectors $\widehat{S}_R$ and $\widehat{S}_P^u$ detect changes faster than $|\widehat{S}_R|$ and $\widehat{S}_P$ but due to the slower convergence to the asymptotic distribution under the null hypothesis (cf. Table 4) their application on the basis of smaller training periods is not recommended.
As a conclusion of this small simulation study we find that the proposed procedures in

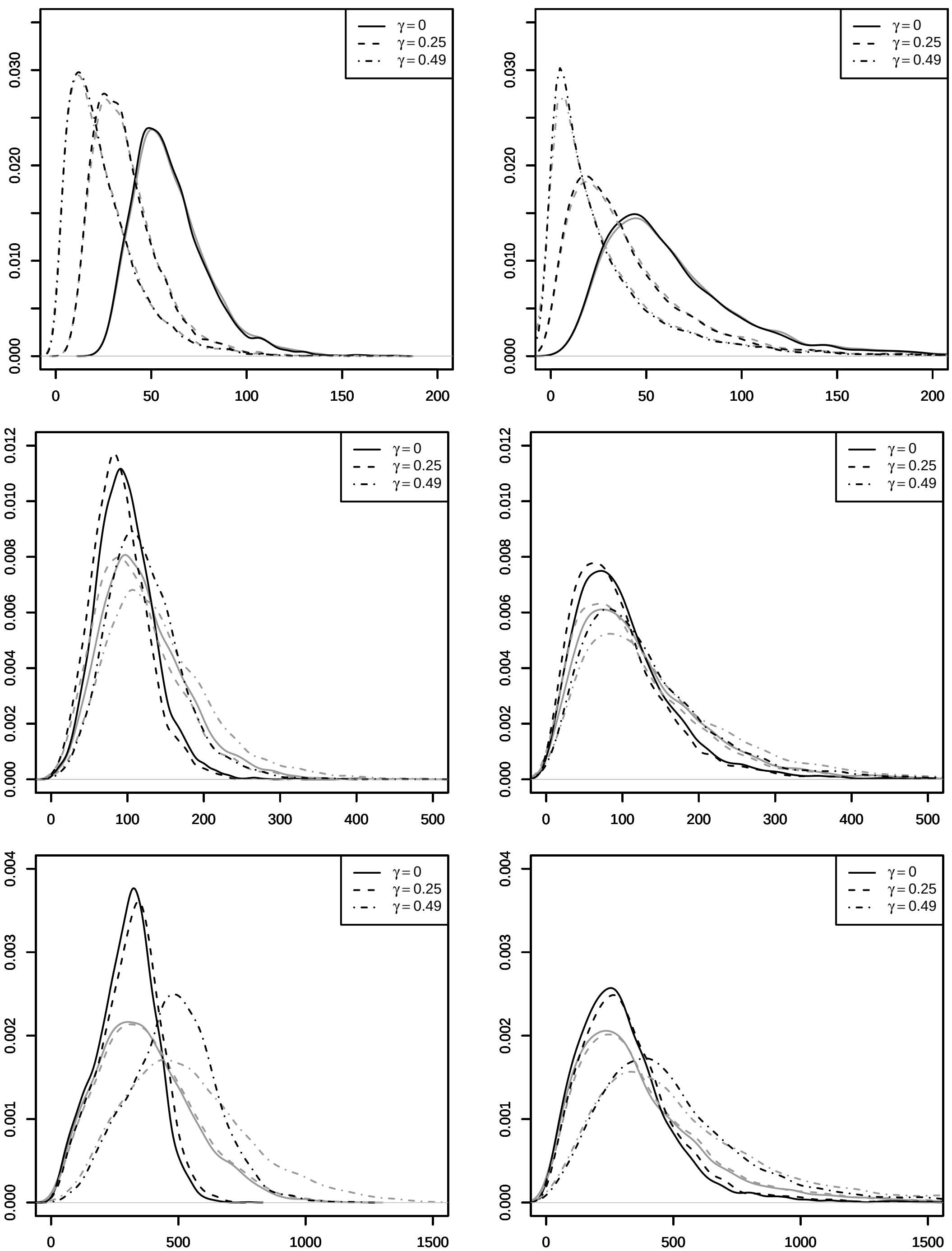


Figure 1: Estimated density plots for the delay times under H_1 for $m = 100$ and $\gamma = 0.00, 0.25, 0.49$. Black lines represent Page CUSUM procedures, gray lines ordinary CUSUM procedures. The left column shows the densities of $\widehat{Q}_P$ and $|\widehat{Q}|$, the right column of $\widehat{S}_P$ and $|\widehat{S}_R|$. The rows from top to bottom represent early ($k^* = 1$), intermediate ($k^* = m$) and late ($k^* = 5m$) changes.

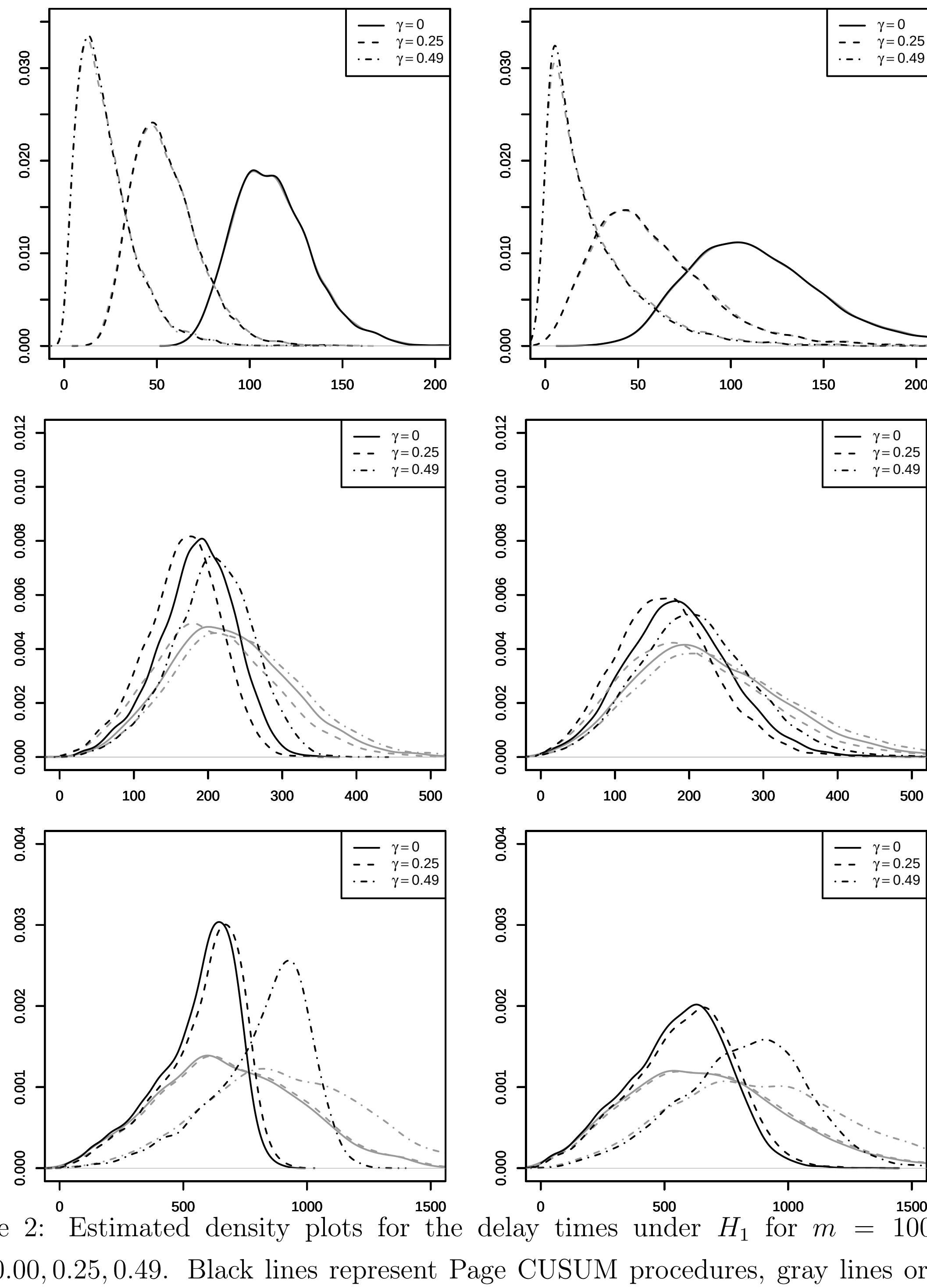


Figure 2: Estimated density plots for the delay times under H_1 for $m = 1000$ and $\gamma = 0.00, 0.25, 0.49$. Black lines represent Page CUSUM procedures, gray lines ordinary CUSUM procedures. The left column shows the densities of $\widehat{Q}_P$ and $|\widehat{Q}|$, the right column of $\widehat{S}_P$ and $|\widehat{S}_R|$. The rows from top to bottom represent early ($k^* = 1$), intermediate ($k^* = m$) and late ($k^* = 5m$) changes.

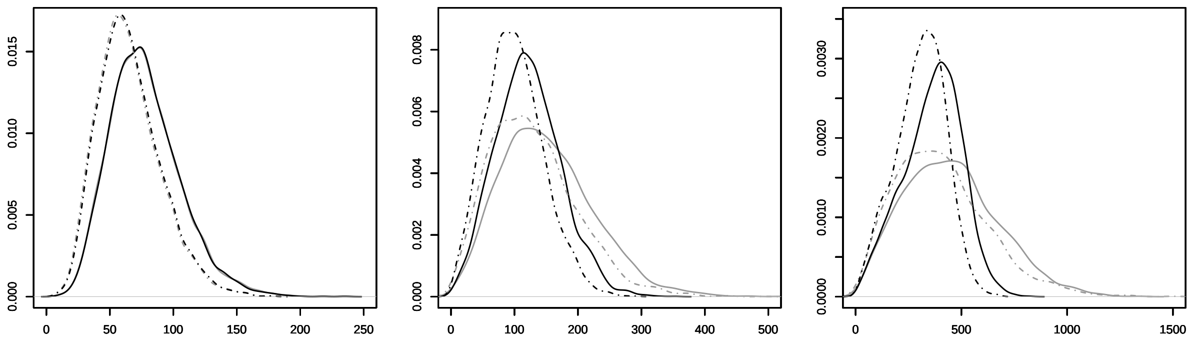


Figure 3: Estimated density plots for the delay times under H_2 for $m = 1000$ and $\gamma = 0.00$ for the procedures based on squared residuals. Black lines represent Page CUSUM procedures, gray lines ordinary CUSUM procedures. Solid lines correspond to the procedures $\widehat{S}_P$ and $|\widehat{S}_R|$, dashed lines correspond to $\widehat{S}_P^u$ and $\widehat{S}_R$. The columns from left to right again represent early ($k^* = 1$), intermediate ($k^* = m$) and late ($k^* = 5m$) changes.

early-change scenarios show a similar behaviour to ordinary CUSUM procedures, yet their advantage lies in the behaviour in scenarios that include a later change. In this case the Page CUSUM procedures detect changes faster and therefore overall show a higher stability regarding the time of change. The procedures based on squared residuals need stronger moment assumptions but they work in contrast to the procedures based on ordinary residuals even under orthogonal changes. The Page CUSUM shows for these a similar behaviour and can therefore be recommended. Nevertheless the procedures based on ordinary residuals in general detect small, non-orthogonal changes faster and can therefore still be of great use in practice.

4.2 Data application: The Fama-French asset pricing model

In this subsection we first want to describe briefly the asset pricing model of Fama and French (1993) that by the introduction of additional factors to the capital asset pricing model of Sharpe (1964) and Lintner (1965) is trying to explain a higher proportion of the variation in the prices of asset portfolios. We will then apply the monitoring procedures introduced in Section 3 to a data set consisting of daily data for 25 asset portfolios considered by Fama and French (1993) and the corresponding factors in the context of the economic crisis from the years 2007 and 2008. This analysis is not intended to give new insights in the economic context of asset pricing nor should it assess the model itself (for this we refer to, e.g., Fama and French (1996), Kothari et al. (1995) or MacKinlay (1995)), it should rather make aware that in times of great disturbance at the markets one should handle with care the data derived by models of this type. In the context of testing asset pricing models for the constancy of their parameters we want to refer, e.g., to Garcia and Ghysels (1998) or Aue et al. (2011).

Fama and French (1993) investigated the influence of risk factors besides the market excess return on an empirical basis to explain the cross-section of average returns. As a consequence they formulated the three-factor model for the excess return of a portfolio i via

$$R_i - R_f = \alpha_i + b_i(R_M - R_f) + s_i\mathrm{SMB} + h_i\mathrm{HML} + \varepsilon_i, \tag{4.14}$$

where R_f is the one month Treasury bill rate, R_M is the return on the market (calculated as the value-weight return on all NYSE, AMEX and NASDAQ stocks), SMB and HML are the so called size and book-to-market factors. For a complete description of the derivation of these factors and how they are calculated we refer to Fama and French (1993), Fama and French (1996) and the website of Kenneth R. French (cf. French (2011)) where the underlying data set can also be found. The data were monitored for the time period January 15, 2004, to June 30, 2011. As responses of this regression model we will consider 25 portfolios formed according to a categorization by size and book-to-market; for the construction of these portfolios we again refer to Fama and French (1996). Fama and French (1993) claim that the excess returns of these portfolios over the market are well explained by (4.14). For our concerns the categorization underlying the construction is not of importance, we will consequently denote the portfolios by Portfolio 1 – 25. In Figure 4 the time series plots of the responses $R_1 - R_f$ as well as of the regressors $R_M - R_f$, SMB and HML for the period January 15, 2004, to January 4, 2010, can be seen, showing obviously conditionally heteroskedastic patterns. The stopping times of detectors $\widehat{Q}_P$, $|\widehat{Q}|$, $\widehat{S}_P$, $|\widehat{S}_R|$, $\widehat{S}_P^u$ and $\widehat{S}_R$ are displayed in Table 5 using for the training period a length of $m = 700$ (i.e. until October 24, 2006, which is a relatively stable period at the markets), $\alpha = 0.1$ and $\gamma = 0.25$. In all but five cases a change is detected, the times of detection considering all detectors lie between August 2007 and March 2009 and thus in the time of the crisis, in many cases the changes are even detected quite early in the crisis. Regarding the strong reaction of the detectors based on squared residuals one should keep in mind the sensitivity of these procedures towards changes in the model parameter σ which cannot be ruled out in this context and should be examined separately. To compare the procedures based on the ordinary CUSUM with those based on the Page CUSUM the same effects that were found in the simulations are evident in the results for this data set. The Page procedures (especially the detectors built from the squared residuals) in general detect a change earlier than the ordinary CUSUM detectors, in none of the cases an ordinary CUSUM detector reacted earlier than the corresponding Page detector.

5 Proofs

5.1 Proof of Theorem 3.1

Because of the similarity of the arguments in the proofs of Theorems 3.1 and 3.2 we only provide the proof of Theorem 3.1. The proof is based on a stepwise approximation of the detector $\widehat{Q}_P(m,k)$ from (3.9) via

$$Q_P(m,k) = \max_{0\le i\le k} |Q(m,k) - Q(m,i)|\,, \text{ where} \tag{5.15}$$

$$Q(m,k) = \sum_{m<i\le m+k} \varepsilon_i - k\bar{\varepsilon}_m, \quad k=1,2,\dots, \quad \text{and} \quad \bar{\varepsilon}_m = \frac{1}{m}\sum_{\ell=1}^{m} \varepsilon_\ell,$$

in the first step and for every m via the following functional of independent Wiener processes $\{W_{1,m}(t) : t \ge 0\}$ and $\{W_{0,m}(t) : t \ge 0\}$ in the second step:

$$W_P(m,k) = \max_{0\le i\le k} \left| W_{1,m}(k) - W_{1,m}(i) - \frac{k-i}{m} W_{0,m}(m) \right|. \tag{5.16}$$

If not stated otherwise the asymptotics in the proofs are always assuming $m \to \infty$.

Lemma 5.1. *If the conditions of Theorem 3.1 are satisfied then*

$$\sup_{1\le k<\infty} \frac{1}{g(m,k)} \left| \widehat{Q}_P(m,k) - Q_P(m,k) \right| = o_P(1),$$

where Q_P was defined in (5.15).

Proof: We have

$$\left| \max_{0\le i\le k} \left| \widehat{Q}(m,k) - \widehat{Q}(m,i) \right| - \max_{0\le i\le k} |Q(m,k) - Q(m,i)| \right|$$

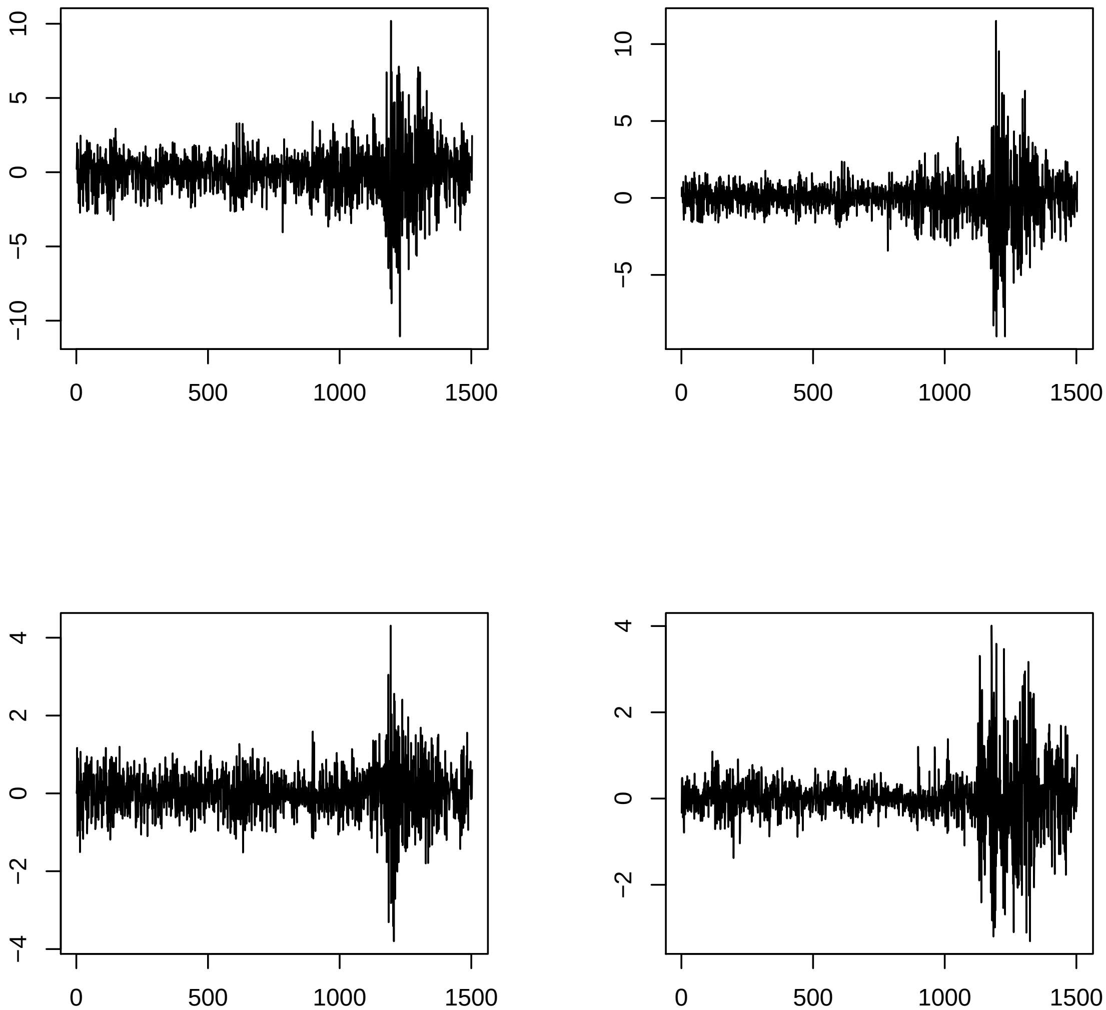


Figure 4: Time series plot of the excess returns of Portfolio 1 (upper left panel), market excess return (upper right panel), size factor (lower left panel) and book-to-market factor (lower right panel) for the time period January 15, 2004 to January 4, 2010.

Portf. No.	$\widehat{Q}_P$	$\lvert\widehat{Q}\rvert$	$\widehat{S}_P$	$\lvert\widehat{S}_R\rvert$	$\widehat{S}_P^u$	$\widehat{S}_R$
1	07/10/2008	07/10/2008	28/07/2008	18/09/2008	24/03/2008	15/09/2008
2	06/10/2008	07/10/2008	14/02/2008	01/04/2008	11/01/2008	18/03/2008
3	06/10/2008	07/10/2008	26/11/2007	27/12/2007	07/11/2007	06/12/2007
4	07/10/2008	07/10/2008	06/03/2008	24/03/2008	15/01/2008	11/03/2008
5	04/1/2008	15/01/2008	14/12/2007	11/01/2008	28/11/2007	27/12/2007
6	30/06/2011	30/06/2011	18/09/2008	07/10/2008	17/09/2008	03/10/2008
7	30/06/2011	30/06/2011	11/03/2008	16/09/2008	01/02/2008	29/07/2008
8	19/11/2008	19/11/2008	27/12/2007	11/03/2008	11/12/2007	18/03/2008
9	19/11/2008	19/11/2008	23/01/2008	08/07/2008	04/01/2008	18/03/2008
10	27/10/2008	27/10/2008	17/01/2008	18/03/2008	21/12/2007	11/03/2008
11	24/10/2008	19/11/2008	15/01/2008	29/02/2008	04/01/2008	31/01/2008
12	19/11/2008	20/11/2008	29/02/2008	01/04/2008	31/01/2008	18/03/2008
13	20/11/2008	05/03/2009	08/01/2008	10/03/2008	11/12/2007	05/02/2008
14	09/10/2008	27/10/2008	11/03/2008	09/09/2008	29/02/2008	16/04/2008
15	27/10/2008	19/11/2008	22/07/2008	24/07/2008	16/07/2008	22/07/2008
16	24/10/2008	20/11/2008	04/02/2008	11/03/2008	17/01/2008	14/02/2008
17	07/10/2008	09/10/2008	17/01/2008	29/02/2008	04/01/2008	01/02/2008
18	15/07/2008	24/07/2008	29/08/2007	13/11/2007	17/08/2007	12/11/2007
19	06/10/2008	07/10/2008	12/11/2007	28/11/2007	18/09/2007	19/11/2007
20	10/03/2008	27/06/2008	17/12/2007	28/01/2008	26/11/2007	17/01/2008
21	09/10/2008	09/10/2008	28/08/2007	01/11/2007	15/08/2007	18/09/2007
22	20/11/2008	30/06/2011	09/08/2007	31/08/2007	09/08/2007	28/08/2007
23	06/10/2008	06/10/2008	14/08/2007	29/08/2007	09/08/2007	17/08/2007
24	17/09/2008	07/10/2008	06/08/2007	09/08/2007	03/08/2007	09/08/2007
25	10/10/2008	27/10/2008	07/11/2007	26/11/2007	18/09/2007	13/11/2007

Table 5: Stopping times of the procedures for the 25 Fama-French Portfolios explained by their three-factor model (Dates given as dd/mm/yyyy).

$$\leq \left|\widehat{Q}(m,k) - Q(m,k)\right| + \max_{0\leq i\leq k}\left|\widehat{Q}(m,i) - Q(m,i)\right|.$$

Now because $g(m,k)$ increases monotonically in k

$$\begin{aligned}
&\sup_{1\leq k<\infty}\frac{1}{g(m,k)}\left|\widehat{Q}_P(m,k) - Q_P(m,k)\right| \\
\leq &\sup_{1\leq k<\infty}\frac{1}{g(m,k)}\left|\widehat{Q}(m,k) - Q(m,k)\right| + \sup_{1\leq k<\infty}\frac{1}{g(m,k)}\max_{0\leq i\leq k}\left|\widehat{Q}(m,i) - Q(m,i)\right| \\
\leq &\sup_{1\leq k<\infty}\frac{1}{g(m,k)}\left|\widehat{Q}(m,k) - Q(m,k)\right| + \sup_{1\leq k<\infty}\max_{0\leq i\leq k}\frac{1}{g(m,i)}\left|\widehat{Q}(m,i) - Q(m,i)\right| \\
=&2\sup_{1\leq k<\infty}\frac{1}{g(m,k)}\left|\widehat{Q}(m,k) - Q(m,k)\right|.
\end{aligned}$$

It is therefore sufficient to show

$$\sup_{1\leq k<\infty}\frac{1}{g(m,k)}\left|\widehat{Q}(m,k) - Q(m,k)\right| = o_P(1).$$

Using the identities

$$\widehat{Q}(m,k) = \sum_{i=m+1}^{m+k}\varepsilon_i - \sum_{i=m+1}^{m+k}\mathbf{x}_i^T(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0)$$

and

$$0 = \sum_{\ell=1}^{m}\hat{\varepsilon}_\ell = \sum_{\ell=1}^{m}\varepsilon_\ell - \sum_{i=1}^{m}\mathbf{x}_i^T(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0), \tag{5.17}$$

with $\boldsymbol{d}$ from **(A.3)** we get

$$\left|\widehat{Q}(m,k) - Q(m,k)\right| = \left|\left(\frac{k}{m}\sum_{i=1}^{m}(\mathbf{x}_i - \boldsymbol{d})^T - \sum_{i=m+1}^{m+k}(\mathbf{x}_i - \boldsymbol{d})^T\right)(\widehat{\boldsymbol{\beta}}_m - \boldsymbol{\beta}_0)\right|.$$

In (5.17) the first equality follows from the definition of $\hat{\varepsilon}_i$ and **(A.2)**.
We consider first the term $\sum_{i=1}^{m}(\mathbf{x}_i - \boldsymbol{d})$. By Markov's inequality and **(A.3)** it is clear that we can find $\rho < 1/2$ such that

$$\left|\sum_{i=1}^{m}(\mathbf{x}_i - \boldsymbol{d})\right| = \mathcal{O}_P\left(m^\rho\right). \tag{5.18}$$

The same arguments used to show (5.18) and the Borel-Cantelli Lemma combined with the stationarity of the regressors yield for the term $\sum_{i=m+1}^{m+k}(\mathbf{x}_i - \boldsymbol{d})$ that there exists $\delta > 0$ such that

$$\left| \sum_{i=m+1}^{m+k} (\mathbf{x}_i - \boldsymbol{d}) \right| = \mathcal{O}\left(k^{1-\delta}\right) \text{ a.s., as } k \to \infty, \text{ uniformly in } m. \tag{5.19}$$

The $\sqrt{m}$-consistency of $\widehat{\boldsymbol{\beta}}_m$ together with

$$\lim_{m\to\infty} \sup_{1\le k<\infty} \frac{km^{\rho-1} + k^{1-\delta}}{\sqrt{m}g(m,k)} = 0$$

as well as (5.18) and (5.19) conclude the proof of Lemma 5.1. □

LEMMA 5.2. *If the conditions of Theorem 3.1 are satisfied then for each m there are two independent Wiener processes $\{W_{1,m}(t) : t \ge 0\}$, $\{W_{0,m}(t) : t \ge 0\}$ such that*

$$\sup_{1\le k<\infty} \frac{1}{g(m,k)} \left| Q_P(m,k) - \sigma W_P(m,k) \right| = o_P(1),$$

where Q_P and W_P were defined in (5.15) *and* (5.16)*, respectively.*

Proof: Similar estimations as in the proof of Lemma 5.1 give us

$$\begin{aligned}
&\left| Q_P(m,k) - \sigma W_P(m,k) \right| \\
&\le \max_{0\le i\le k} \left| \left| \sum_{j=m+i+1}^{m+k} \varepsilon_j - (k-i)\bar{\varepsilon}_m \right| - \sigma \left| W_{1,m}(k) - W_{1,m}(i) - \frac{k-i}{m} W_{0,m}(m) \right| \right| \\
&\le \left| \sum_{j=m+1}^{m+k} \varepsilon_j - \sigma W_{1,m}(k) \right| + \max_{0\le i\le k} \left| \sum_{j=m+1}^{m+i} \varepsilon_j - \sigma W_{1,m}(i) \right| + \frac{k}{m} \left| \sum_{\ell=1}^{m} \varepsilon_\ell - \sigma W_{0,m}(m) \right|
\end{aligned}$$

and hence with assumption **(A.5)**

$$\begin{aligned}
&\sup_{1\le k<\infty} \frac{1}{g(m,k)} \left| Q_P(m,k) - \sigma W_P(m,k) \right| \\
&\le \sup_{1\le k<\infty} \frac{1}{g(m,k)} \left| \sum_{j=m+1}^{m+k} \varepsilon_j - \sigma W_{1,m}(k) \right| \\
&\quad + \sup_{1\le k<\infty} \frac{1}{g(m,k)} \max_{0\le i\le k} \left| \sum_{j=m+1}^{m+i} \varepsilon_j - \sigma W_{1,m}(i) \right| \\
&\quad + \sup_{1\le k<\infty} \frac{1}{g(m,k)} \frac{k}{m} \left| \sum_{\ell=1}^{m} \varepsilon_\ell - \sigma W_{0,m}(m) \right|
\end{aligned}$$

$$= \mathcal{O}_P(1) \sup_{1\le k<\infty} \frac{k^{\xi}}{g(m,k)} + \mathcal{O}_P(1) \sup_{1\le k<\infty} \frac{km^{\xi-1}}{g(m,k)}$$
$$= o_P(1),$$

where the last equality was shown in the proof of Lemma 3 of Aue et al. (2006b). □

PROOF OF THEOREM 3.1

The distribution of $\{(W_{1,m}(t), W_{0,m}(t)) : t \ge 0\}$ does not depend on m and therefore the index can be omitted, i.e. we write $\{(W_1(t), W_0(t)) : t \ge 0\}$ instead. Due to the scaling property of the Wiener process we have

$$\sup_{1\le k<\infty} \frac{W_P(m,k)}{g(m,k)} \overset{\mathcal{D}}{=} \sup_{1\le k<\infty} \max_{0\le i\le k} \frac{|W_1(k/m) - W_1(i/m) - ((k-i)/m)W_0(1)|}{(1+k/m)\,(k/(k+m))^{\gamma}}$$

and define

$$R_P(m,k) = \max_{0\le i\le k} \frac{|W_1(k/m) - W_1(i/m) - ((k-i)/m)W_0(1)|}{(1+k/m)\,(k/(k+m))^{\gamma}}.$$

Furthermore we define

$$u(t) = (1+t)\,(t/(1+t))^{\gamma}$$

and the following functionals of the Wiener processes W_0 and W_1

$$R_P(t) = \frac{1}{u(t)} \sup_{0\le s\le t} |W_1(t) - W_1(s) - (t-s)W_0(1)|$$
$$\overline{R}_P(m,t) = \frac{1}{u(t/m)} \sup_{0\le s\le t} |W_1(t/m) - W_1(s/m) - ((s-t)/m)W_0(1)|\,,$$
$$\widetilde{R}_P(m,\lceil t\rceil) = \frac{1}{u(\lceil t\rceil/m)} \sup_{0\le s\le \lceil t\rceil} |W_1(\lceil t\rceil/m) - W_1(\lceil s\rceil/m) - ((\lceil t\rceil - \lceil s\rceil)/m)W_0(1)|\,.$$

We note that

$$\sup_{1\le k<\infty} R_P(m,k) = \sup_{0<t<\infty} \widetilde{R}_P(m,\lceil t\rceil).$$

The next step is to show:

$$\sup_{1\le k<\infty} R_P(m,k) \overset{(m\to\infty)}{\longrightarrow} \sup_{0<t<\infty} R_P(t) \quad \text{a.s.}. \tag{5.20}$$

We divide the proof of (5.20) into two steps and show

(i) For any $T > 0$:

$$\max_{1\le k\le mT} R_P(m,k) \overset{(m\to\infty)}{\longrightarrow} \sup_{0<t\le T} R_P(t) \quad \text{a.s.}$$

and

(ii) For almost every $\omega \in \Omega$ there exists a positive integer $T = T(\omega)$ such that

$$\sup_{mT\le k<\infty} R_P(m,k) \overset{(m\to\infty)}{\longrightarrow} \sup_{T\le t<\infty} R_P(t).$$

The first claim follows directly because of the a.s. continuity of $R_P(t)$ on $[0,T]$ (with $R_P(0) = 0$). For the second claim we get for any $T > 0$

$$\begin{aligned}
&\sup_{mT\le t<\infty} \left|\widetilde{R}_P(m,\lceil t\rceil) - \overline{R}_P(m,t)\right| \\
\le &\sup_{mT\le t<\infty}\left|\frac{W_1(\lceil t\rceil/m)}{u(\lceil t\rceil/m)} - \frac{W_1(t/m)}{u(t/m)}\right| + \sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{W_1(\lceil s\rceil/m)}{u(\lceil t\rceil/m)} - \frac{W_1(s/m)}{u(t/m)}\right| \\
&+ \sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{t-s}{u(t/m)} - \frac{\lceil t\rceil - \lceil s\rceil}{u(\lceil t\rceil/m)}\right| \frac{|W_0(1)|}{m} \\
\le &\, 2\sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{W_1(\lceil s\rceil/m)}{u(\lceil t\rceil/m)} - \frac{W_1(s/m)}{u(t/m)}\right| + \sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{t-s}{u(t/m)} - \frac{\lceil t\rceil - \lceil s\rceil}{u(\lceil t\rceil/m)}\right| \frac{|W_0(1)|}{m} \\
= &\, 2A_1 + A_2.
\end{aligned}$$

For A_1 we have for any $T > 0$

$$\begin{aligned}
&\sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{W_1(\lceil s\rceil/m) - W_1(s/m)}{u(\lceil t\rceil/m)} - \frac{W_1(s/m)}{u(t/m)} + \frac{W_1(s/m)}{u(\lceil t\rceil/m)}\right| \\
\le &\sup_{mT\le t<\infty}\sup_{0\le s\le t}\left|\frac{W_1(\lceil s\rceil/m) - W_1(s/m)}{u(\lceil t\rceil/m)}\right| \\
&+ \sup_{mT\le t<\infty}\sup_{0\le s\le t}|W_1(s/m)|\left|\frac{1}{u(\lceil t\rceil/m)} - \frac{1}{u(t/m)}\right| \\
= &\sup_{mT\le t<\infty}\sup_{0\le s\le t} A_3(t,s) + \sup_{mT\le t<\infty}\sup_{0\le s\le t} A_4(t,s).
\end{aligned}$$

By Theorem 1.2.1 of Csörgő and Révész (1981) for all $\varepsilon > 0$ there exists $T(\omega) > 0$ independent of m such that

$$\sup_{mT\le t<\infty}\sup_{0\le s\le t}\sup_{0\le r\le 1}\frac{|W_1((s+r)/m) - W_1(s/m)|}{u(t/m)} < \varepsilon \quad \text{a.s..}$$

Consequently for almost every $\omega \in \Omega$ there exists $T_1(\omega) > 0$ providing

$$\begin{aligned}
&\sup_{mT_1 \le t < \infty} \sup_{0 \le s \le t} A_3(t,s) \\
&\le \sup_{mT_1 \le t < \infty} \sup_{0 \le s \le t} \frac{|W_1(\lceil s \rceil/m) - W_1(s/m)|}{u(t/m)} \\
&\le \sup_{mT_1 \le t < \infty} \sup_{0 \le s \le t} \sup_{0 \le r \le 1} \frac{\left|W_1\left(\frac{s+r}{m}\right) - W_1(s/m)\right|}{u(t/m)} < \frac{\varepsilon}{8}.
\end{aligned}$$

For A_4 with Theorem 1.3.1* of Csörgő and Révész (1981) we get similarly that for almost every $\omega \in \Omega$ there exists $T_2(\omega) > 0$ (again independent of m) such that

$$\sup_{mT_2 \le t < \infty} \sup_{0 \le s \le t} A_4 \le \sup_{mT_2 \le t < \infty} \sup_{0 \le s \le t} \left| \frac{W_1(s/m)}{u(t/m)} \right| < \frac{\varepsilon}{8}.$$

For A_2 we find that for any $T > 0$:

$$\begin{aligned}
&\sup_{mT \le t < \infty} \sup_{0 \le s \le t} \left| \frac{\lceil t \rceil - \lceil s \rceil}{u(\lceil t \rceil/m)} - \frac{t-s}{u(t/m)} \right| \frac{|W_0(1)|}{m} \\
&\le \sup_{mT \le t < \infty} \sup_{0 \le s \le t} \left| \frac{\lceil t \rceil - t - (\lceil s \rceil - s)}{u(\lceil t \rceil/m)} + (t-s)\left(\frac{1}{u(\lceil t \rceil/m)} - \frac{1}{u(t/m)}\right) \right| \frac{|W_0(1)|}{m} \\
&\le \sup_{mT \le t < \infty} \frac{1}{m} \frac{|W_0(1)|}{u(\lceil t \rceil/m)} + \sup_{mT \le t < \infty} \frac{t}{m} \left| \frac{1}{u(t/m)} - \frac{1}{u(\lceil t \rceil/m)} \right| |W_0(1)| \\
&= A_5.
\end{aligned}$$

From (3.8) and because

$$(m\, u(\lceil t \rceil/m))^{-1} \le (m\, u(t/m))^{-1}, \tag{5.21}$$

we have

$$\begin{aligned}
\left| \frac{t/m}{u(t/m)} - \frac{t/m}{u(\lceil t \rceil/m)} \right| &\le \left(\frac{t}{t+m}\right)^{1-\gamma} - \frac{t}{(t+m+1)\left((t+1)/(t+m+1)\right)^{\gamma}} \\
&\le 1 - \left(\frac{t+m}{t+m+1}\right)^{1-\gamma} \left(\frac{t}{t+1}\right)^{\gamma}.
\end{aligned} \tag{5.22}$$

This follows since the right-hand sides of (5.21) and (5.22) are both monotonically decreasing in t. Now we find

$$A_5 \le \frac{|W_0(1)|}{m(1+T)\left(T/(T+1)\right)^{\gamma}} + \left[1 - \left(\frac{T+1}{T+1+1/m}\right)^{1-\gamma} \left(\frac{T}{T+1/m}\right)^{\gamma}\right] |W_0(1)|$$

$$\leq |W_0(1)| \left\{ (1+T)^{\gamma-1} T^{\gamma} + \left[1 - \left(\frac{T+1}{T+2} \right)^{1-\gamma} \left(\frac{T}{T+1} \right)^{\gamma} \right] \right\}.$$

Therefore for almost every $\omega \in \Omega$ there exists $T_3(\omega) > 0$ and independent of m such that:

$$\sup_{mT_3 \leq t < \infty} \sup_{0 \leq s \leq t} \left| \frac{t-s}{u(t/m)} - \frac{\lceil t \rceil - \lceil s \rceil}{u(\lceil t \rceil / m)} \right| \frac{|W_0(1)|}{m} < \frac{\varepsilon}{2}.$$

Finally we have for almost every $\omega \in \Omega$ and with $T := \max(T_1, T_2, T_3)$ that

$$\left| \sup_{mT \leq k < \infty} R_P(m,k) - \sup_{T \leq t < \infty} R_P(t) \right| < \varepsilon,$$

since it is clear that $\sup_{mT \leq t < \infty} \overline{R}_P(m,t) = \sup_{T \leq t < \infty} R_P(t)$ for every m. Putting these together we get

$$\sup_{1 \leq k < \infty} R_P(m,k) \overset{(m \to \infty)}{\longrightarrow} \sup_{0 < t < \infty} R_P(t) \quad \text{a.s.}$$

and thus

$$\sup_{1 \leq k < \infty} \frac{1}{g(m,k)} \max_{0 \leq i \leq k} \left| W_{1,m}(k) - W_{1,m}(i) - \frac{k-i}{m} W_{0,m}(m) \right| \xrightarrow{\mathcal{D}} \sup_{0 < t < \infty} R_P(t).$$

By computing the covariance functions it can be shown that

$$\{W_1(t) - tW_0(1),\ 0 \leq t < \infty\} \overset{\mathcal{D}}{=} \{(1+t) W(t/(1+t)),\ 0 \leq t < \infty\},$$

where $\{W(t),\ 0 \leq t < \infty\}$ is a Wiener process (cf. Horváth et al. (2004)). We conclude

$$\begin{aligned} \sup_{0 < t < \infty} R_P(t) &\overset{\mathcal{D}}{=} \sup_{0 < t < \infty} \sup_{0 \leq s \leq t} \frac{|(1+t)W(t/(1+t)) - (1+s)W(s/(1+s))|}{(1+t)(t/(1+t))^{\gamma}} \\ &= \sup_{0 < t < \infty} \sup_{0 \leq s \leq t} \frac{|W(t/(1+t)) - ((1+s)/(1+t))W(s/(1+s))|}{(t/(1+t))^{\gamma}} \\ &= \sup_{0 < t < 1} \sup_{0 \leq s \leq t} \frac{1}{t^{\gamma}} |W(t) - ((1-t)/(1-s))W(s)|. \end{aligned}$$

The weak consistency of the estimator $\hat{\sigma}_m$ completes the proof. □

5.2 Proof of Theorem 3.3

We only prove part a) of Theorem 3.3, parts b) and c) then follow immediately. Since

$$\min_{0 \leq i \leq k} \sum_{j=m+1}^{m+i} \hat{\varepsilon}_j \leq 0,$$

we have

$$\widehat{Q}(m,k) = \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i \quad \leq \quad \sum_{i=m+1}^{m+k} \hat{\varepsilon}_i - \min_{0\leq i\leq k} \sum_{j=m+1}^{m+i} \hat{\varepsilon}_j = \widehat{Q}_P^u(m,k).$$

Consequently it is sufficient to show that under H_A and $\boldsymbol{d}^T\boldsymbol{\Delta} > 0$ we have

$$\frac{1}{\hat{\sigma}_m} \sup_{1\leq k<\infty} \frac{\widehat{Q}(m,k)}{g(m,k)} \xrightarrow{P} \infty. \tag{5.23}$$

To show this we expand $\widehat{Q}(m,k)$ for $k \geq k^*$ to

$$\widehat{Q}(m,k) = \sum_{i=m+1}^{m+k} \varepsilon_i + \left(\sum_{i=m+1}^{m+k} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_m) + \left(\sum_{i=m+k^*}^{m+k} (\mathbf{x}_i - \boldsymbol{d})\right)^T \boldsymbol{\Delta} + (k-k^*+1)\boldsymbol{d}^T\boldsymbol{\Delta}. \tag{5.24}$$

From the proof of Theorem 3.1 we get

$$\sup_{1\leq k<\infty} \frac{1}{g(m,k)} \left| \sum_{i=m+1}^{m+k} \varepsilon_i + \left(\sum_{i=m+1}^{m+k} \mathbf{x}_i\right)^T (\boldsymbol{\beta}_0 - \widehat{\boldsymbol{\beta}}_m) \right| = \mathcal{O}_P(1).$$

Now because of (5.19) from the proof of Lemma 5.1 we obtain

$$\left(\sum_{i=m+k^*}^{m+k} (\mathbf{x}_i - \boldsymbol{d})\right)^T \boldsymbol{\Delta} = o(k-k^*) \quad \text{as } k \to \infty, \qquad \text{a.s., uniformly in } m.$$

As a consequence the drift term $\sup_{1\leq k<\infty} (k-k^*+1)\boldsymbol{d}^T\boldsymbol{\Delta}/g(m,k)$ is the dominating term and it is clearly diverging as $m \to \infty$. □

$k^* = 1$			$m = 200$					$m = 1000$		
$\gamma = 0.00$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	15	39	48	60	159	48	84	96	110	191
$\lvert\widehat{Q}\rvert$	14	39	48	60	159	48	84	95	109	189
$\widehat{Q}_P^u$	11	31	39	49	142	38	70	80	92	179
$\widehat{Q}^u$	11	31	39	49	143	38	68	79	91	177
$\widehat{S}_P$	3	32	46	68	558	22	76	96	120	256
$\lvert\widehat{S}_R\rvert$	3	32	47	69	558	22	75	96	120	263
$\widehat{S}_P^u$	3	25	38	56	449	11	62	80	102	251
$\widehat{S}_R$	3	25	38	57	451	11	61	80	101	251
$\gamma = 0.25$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	4	20	28	38	144	12	35	44	56	146
$\lvert\widehat{Q}\rvert$	4	20	28	39	144	11	34	44	55	148
$\widehat{Q}_P^u$	4	16	22	31	108	9	28	37	47	135
$\widehat{Q}^u$	3	15	22	31	107	8	27	36	46	135
$\widehat{S}_P$	1	14	26	44	558	1	28	44	64	216
$\lvert\widehat{S}_R\rvert$	1	15	27	46	558	1	29	44	64	217
$\widehat{S}_P^u$	1	11	21	37	451	1	22	36	54	202
$\widehat{S}_R$	1	11	22	38	451	1	22	35	53	196
$\gamma = 0.49$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	1	9	17	29	148	1	9	16	26	135
$\lvert\widehat{Q}\rvert$	1	9	17	28	148	1	9	16	26	133
$\widehat{Q}_P^u$	1	8	14	24	137	1	8	13	22	124
$\widehat{Q}^u$	1	7	14	24	144	1	7	13	22	117
$\widehat{S}_P$	1	6	15	33	1888	1	6	14	30	211
$\lvert\widehat{S}_R\rvert$	1	6	16	35	1460	1	6	15	32	200
$\widehat{S}_P^u$	1	5	13	29	1269	1	5	12	27	197
$\widehat{S}_R$	1	5	13	30	838	1	5	13	28	190

Table 6: Five number summary under H_1 with an early-change $k^* = 1$ for $\alpha = 0.1$.

$k^* = m$	$m = 200$					$m = 1000$				
$\gamma = 0.00$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	1	59	79	101	289	4	126	159	189	318
$\lvert\widehat{Q}\rvert$	1	66	95	131	443	9	143	193	248	476
$\widehat{Q}_P^u$	1	45	63	82	261	1	98	129	157	263
$\widehat{Q}^u$	2	51	78	110	450	1	113	160	214	558
$\widehat{S}_P$	1	46	75	112	2200	2	114	156	200	467
$\lvert\widehat{S}_R\rvert$	1	53	90	141	2200	6	130	188	258	767
$\widehat{S}_P^u$	1	36	60	91	1593	1	90	127	165	403
$\widehat{S}_R$	1	41	74	117	1595	1	103	157	222	650
$\gamma = 0.25$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	2	54	73	96	289	3	113	145	175	305
$\lvert\widehat{Q}\rvert$	1	61	89	124	423	2	129	177	231	451
$\widehat{Q}_P^u$	1	43	62	80	290	1	93	124	150	256
$\widehat{Q}^u$	1	49	75	108	464	1	107	153	206	556
$\widehat{S}_P$	1	43	71	107	2200	3	104	143	186	436
$\lvert\widehat{S}_R\rvert$	1	49	85	136	2200	3	118	175	242	731
$\widehat{S}_P^u$	1	36	60	91	1644	1	87	122	159	380
$\widehat{S}_R$	1	41	73	117	1726	1	99	151	215	648
$\gamma = 0.49$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	3	77	103	133	490	14	153	189	223	386
$\lvert\widehat{Q}\rvert$	2	81	115	160	658	15	162	214	273	526
$\widehat{Q}_P^u$	2	67	91	119	430	6	136	170	202	342
$\widehat{Q}^u$	1	71	103	144	569	7	144	196	254	631
$\widehat{S}_P$	2	63	100	155	2200	1	140	187	237	641
$\lvert\widehat{S}_R\rvert$	1	67	111	181	2200	5	151	212	288	890
$\widehat{S}_P^u$	1	55	90	138	2200	2	125	169	217	522
$\widehat{S}_R$	1	59	100	163	2200	3	135	194	267	826

Table 7: Five number summary under H_1 with $k^* = m$ for $\alpha = 0.1$.

$k^* = 5m$	$m = 200$					$m = 1000$				
$\gamma = 0.00$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	2	182	256	313	550	11	382	506	581	785
$\lvert\widehat{Q}\rvert$	2	191	293	412	933	6	404	577	777	1436
$\widehat{Q}_P^u$	3	152	212	259	489	5	319	425	490	666
$\widehat{Q}^u$	2	158	252	366	1199	2	340	503	700	1826
$\widehat{S}_P$	2	150	234	331	3000	6	352	492	601	1074
$\lvert\widehat{S}_R\rvert$	1	158	263	418	3000	1	369	566	777	1817
$\widehat{S}_P^u$	1	126	198	278	1634	13	296	414	511	1027
$\widehat{S}_R$	1	133	225	366	3000	5	311	491	696	1938
$\gamma = 0.25$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	7	199	276	335	608	8	411	537	613	826
$\lvert\widehat{Q}\rvert$	3	205	310	432	1035	8	427	601	801	1471
$\widehat{Q}_P^u$	2	172	236	288	604	2	357	466	533	713
$\widehat{Q}^u$	2	176	273	391	1248	9	372	536	737	1857
$\widehat{S}_P$	3	166	255	361	3000	11	384	523	638	1228
$\lvert\widehat{S}_R\rvert$	1	172	280	443	3000	15	395	593	808	1895
$\widehat{S}_P^u$	2	142	221	311	3000	15	333	456	555	1123
$\widehat{S}_R$	3	149	247	395	3000	3	343	527	733	1977
$\gamma = 0.49$	min	1^{st}Q	med	3^{rd}Q	max	min	1^{st}Q	med	3^{rd}Q	max
$\widehat{Q}_P$	1	326	430	526	1089	34	638	785	876	1203
$\lvert\widehat{Q}\rvert$	23	315	450	610	1577	17	622	819	1049	1898
$\widehat{Q}_P^u$	12	292	386	469	977	30	578	715	802	1085
$\widehat{Q}^u$	13	278	402	553	1679	19	558	746	975	7000
$\widehat{S}_P$	11	268	399	574	3000	36	595	766	912	1584
$\lvert\widehat{S}_R\rvert$	3	256	407	641	3000	22	580	808	1059	7000
$\widehat{S}_P^u$	1	239	356	508	3000	6	537	700	833	1468
$\widehat{S}_R$	6	226	363	575	3000	3	516	736	977	7000

Table 8: Five number summary under H_1 with $k^* = 5m$ for $\alpha = 0.1$.

		min	1^{st}Q	med	3^{rd}Q	max
$k^* = 1$	$\widehat{S}_P$	10	58	75	94	235
	$\lvert\widehat{S}_R\rvert$	10	58	75	94	235
	$\widehat{S}_P^u$	8	48	63	80	178
	$\widehat{S}_R$	8	47	62	79	178
$k^* = m$	$\widehat{S}_P$	2	86	119	155	353
	$\lvert\widehat{S}_R\rvert$	2	98	144	196	537
	$\widehat{S}_P^u$	1	68	98	129	293
	$\widehat{S}_R$	1	78	121	169	493
$k^* = 5m$	$\widehat{S}_P$	4	271	376	461	823
	$\lvert\widehat{S}_R\rvert$	4	287	435	590	1345
	$\widehat{S}_P^u$	7	228	317	392	676
	$\widehat{S}_R$	6	243	380	532	1526

Table 9: Five number summary under H_2 for $\alpha = 0.1$, $\gamma = 0.00$ and $m = 1000$.

REFERENCES

J. Antoch and D. Jarušková. On-line statistical process control. In *Multivariate Total Quality Control, Foundations and Recent Advances. Eds: Lauro, C., Antoch, J., and Vinzi, V.E.* Physica, Heidelberg, 2002.

A. Aue, I. Berkes, and L. Horváth. Strong approximation for the sums of squares of augmented GARCH sequences. *Bernoulli*, 12(4):583, 2006a.

A. Aue and L. Horváth. Delay time in sequential detection of change. *Statistics & Probability Letters*, 67(3):221–231, 2004.

A. Aue, L. Horváth, M. Hušková, and P. Kokoszka. Change-point monitoring in linear models. *Econometrics Journal*, 9(3):373–403, 2006b.

A. Aue, L. Horváth, M. Kühn, and J. Steinebach. On the reaction time of moving sum detectors. *Journal of Statistical Planning and Inference*, 142(8):2271–2288, 2012.

A. Aue, L. Horváth, and M.L. Reimherr. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, 149(2):174 – 190, 2009.

A. Aue, S. Hörmann, L. Horváth, M. Hušková, and J.G. Steinebach. Sequential testing for the stability of high frequency portfolio betas. *Econometric Theory*, 28(4):804-837, 2012.

J. Bai. Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, 79(4):551–563, 1997.

M. Carrasco and X. Chen. Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory*, 18(1):17–39, 2002.

C.S.J. Chu, K. Hornik, and C.M. Kuan. Mosum tests for parameter constancy. *Biometrika*, 82(3):603, 1995.

C.S.J. Chu, M. Stinchcombe, and H. White. Monitoring structural change. *Econometrica*, 64(5):1045–1065, 1996.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis.* Wiley, New York, 1997.

M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics*. Academic Press, 1981.

J.C. Duan. Augmented GARCH (p,q) process and its diffusion limit. *Journal of Econometrics*, 79(1):97–127, 1997.

E.F. Fama and K.R. French. Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):3–56, 1993.

E.F. Fama and K.R. French. Multifactor explanations of asset pricing anomalies. *Journal of Finance*, 51(1):55–84, 1996.

K.R. French. Kenneth R. French - Data Library, July 2011. `http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data\_library.html`.

R. Garcia and E. Ghysels. Structural change and asset pricing in emerging markets. *Journal of International Money and Finance*, 17(3):455–473, 1998.

L. Horváth, M. Hušková, P. Kokoszka, and J. Steinebach. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225–251, 2004.

M. Hušková and A. Koubková. Monitoring jump changes in linear models. *Journal of Statistical Research*, 39(2):51–70, 2005.

M. Hušková, Z. Prášková, and J. Steinebach. On the detection of changes in autoregressive time series I. Asymptotics. *Journal of Statistical Planning and Inference*, 137(4):1243–1259, 2007.

S.P. Kothari, J. Shanken, and R.G. Sloan. Another look at the cross-section of expected stock returns. *Journal of Finance*, 50(1):185–224, 1995.

J. Lintner. 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, 1965.

A.C. MacKinlay. Multifactor models do not explain deviations from the CAPM. *Journal of Financial Economics*, 38(1):3–28, May 1995.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

P. Perron. Dealing with structural breaks. *Palgrave Handbook of Econometrics*, 1:278–352, 2006.

A. Schmitz and J.G. Steinebach. A note on the monitoring of changes in linear models with dependent errors. *Dependence in Probability and Statistics, Lecture Notes in Statistics*, 200:159–174, 2010.

W.F. Sharpe. Capital asset prices: A theory of market equilibrium under conditions of risk. *Journal of Finance*, 19(3):425–442, 1964.

ASYMPTOTIC DISTRIBUTION OF THE DELAY TIME IN PAGE'S SEQUENTIAL PROCEDURE

By Stefan Fremdt[‡]

[‡] *University of Cologne*

Abstract

In this paper the asymptotic distribution of the stopping time in Page's sequential CUSUM procedure is presented. Page's CUSUM is introduced as a detector for changes in the mean of observations satisfying a weak invariance principle. The main result on the stopping time derived from this procedure extends a series of results on the asymptotic normality of stopping times of CUSUM-type procedures. In contrast to the result presented here, the asymptotic normality in these papers holds only in limited range of possible change locations and change sizes. The theoretical results are then illustrated by a small simulation study.

Keywords: CUSUM, Delay time, Asymptotic distribution, Location model, Change-point, Sequential test, Invariance principle.

AMS subject classification: Primary 62L99; secondary 62G20

1 Introduction

In sequential change-point analysis a comparison between different types of detection procedures in many cases is based on their average delay times. Even optimality criteria that were introduced for these procedures are referring to the expected value of the delay time. Until now however only few contributions were made regarding the asymptotic distribution of the stopping times, providing more information about the behaviour of the stopping times than results on the average behaviour. Concerning the aforementioned optimality criteria for Page's CUSUM we refer to the work of Lorden (1971). The monograph of Basseville and Nikiforov (1993) gives an extensive overview of the contributions made since the introduction of the CUSUM procedure by Page (1954).

In a time series regression model Fremdt (2012) proposed as a stopping time the first-passage time of Page's CUSUM of the residuals over a boundary function introduced by

Horváth et al. (2004). Based on this procedure we will consider changes in the mean in the so-called location model which, as a special case, is included in the time series regression model of Fremdt (2012) and investigate the limit distribution of the corresponding stopping time. Ordinary CUSUM procedures are defined as the partial sum of, e.g., the residuals from the beginning of the monitoring to the present. These procedures have been studied in the literature extensively, we refer to, e.g., Horváth et al. (2004), Horvath et al. (2007) or Aue et al. (2006b). Results on the asymptotic distribution, in particular the asymptotic normality, of this ordinary CUSUM procedure were given for the location model by Aue (2003) and Aue and Horváth (2004), an extension to a linear regression model was then provided by Aue et al. (2009). To prove these results on the asymptotic normality of the ordinary CUSUM detector strong conditions on the time of change as well as on the magnitude of the change had to be imposed. In this context we also want to mention the work of Hušková and Koubková (2005), who introduced monitoring procedures based on quadratic forms of weighted cumulative sums, and Černíková et al. (2011), who showed the asymptotic normality of the corresponding stopping time under assumptions on the time of change similar to those used in Aue and Horváth (2004) and Aue et al. (2009).

Building on the work of Aue and Horváth (2004), we will derive the asymptotic distribution of the Page CUSUM procedure while relaxing the strong conditions on the time and magnitude of the change and show hereby that it is more robust to the location and the size of the change than ordinary CUSUM procedures. The corresponding limit distributions are novel in this context and provide a classification of the behaviour of the stopping time according to the specifications of the change.

The paper is organized as follows. In Section 2 we will introduce our model settings and assumptions and formulate our main result, Section 3 then contains the results of a small simulation study. We will conclude this work with the proof of our main result from Section 2 in Section 4.

2 Asymptotic distribution of the stopping times

Aue and Horváth (2004) investigated in a sequential setup the asymptotic normality of the CUSUM stopping time in the case of the so-called location model with the alternative hypothesis of a change in the mean for relatively early changes. They used the "noncontamination assumption" introduced by Chu et al. (1996) which is standard for these types of problems. This assumption guarantees the constancy of the model in an initial training period of length m. If not stated otherwise the asymptotics considered in this paper are always with respect to $m \to \infty$ (which is therefore omitted). The characterization of a change-point k^* as "relatively early" is then also with respect to the length of this training period. A formulation of this in mathematical terms will follow later on. In this work we will derive the asymptotic distribution of the stopping time based on Page's CUSUM detector in this location model which is given via

$$X_i = \begin{cases} \mu + \varepsilon_i, & i = 1, \dots, m + k^* - 1, \\ \mu + \varepsilon_i + \Delta_m, & i = m + k^*, m + k^* + 1, \dots, \end{cases} \tag{2.1}$$

where μ and Δ_m are real numbers and $1 \leq k^* < \infty$ denotes the unknown time of change. We will restrict ourselves to the case $\Delta_m \geq 0$ and only formulate the statements for the hypotheses

$$H_0: \quad \Delta_m = 0 \qquad \text{and} \qquad H_A: \quad \Delta_m > 0. \tag{2.2}$$

The respective statements for the alternatives $\Delta_m < 0$ and $\Delta_m \neq 0$ follow analogously using the corresponding detectors given in Fremdt (2012).

Following Aue and Horváth (2004) we assume that the error terms $\{\varepsilon_i\}_{i \in 1,2,\dots}$ satisfy Assumptions

$$\left| \sum_{i=1}^{m} \varepsilon_i \right| = \mathcal{O}_P\left(\sqrt{m}\right), \tag{A1}$$

There is a sequence of Wiener processes $\{W_m(t) : t \geq 0\}_{m \geq 1}$ and a positive constant σ such that

$$\sup_{\frac{1}{m} \leq t < \infty} \frac{1}{(mt)^{1/\nu}} \left| \sum_{i=m+1}^{m+mt} \varepsilon_i - \sigma W_m(mt) \right| = \mathcal{O}_P(1) \quad \text{with some } \nu > 2. \tag{A2}$$

Furthermore we need the following assumptions on Δ_m and k^*:

$$\text{there exists a } \theta > 0 \text{ such that } k^* = \lfloor \theta m^\beta \rfloor \text{ with } 0 \leq \beta < 1 \tag{A3}$$

$$\sqrt{m}\Delta_m \longrightarrow \infty, \tag{A4}$$

$$\Delta_m = \mathcal{O}(1). \tag{A5}$$

Examples for sequences of random variables satisfying Assumptions (A1) and (A2) are given in Aue and Horváth (2004), besides i.i.d. sequences including martingale difference sequences and certain stationary mixing sequences. Aue et al. (2006b) showed that the class of augmented GARCH processes which were introduced by Duan (1997) and include most of the conditionally heteroskedastic time series models applied to describe financial time series also satisfies Assumptions (A1) and (A2). A selection of GARCH models that are included in this class can be found in Aue et al. (2006a). For the location model Aue and Horváth (2004) defined the CUSUM detector of the (centered) X_i

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

and as a corresponding stopping time

$$\tau_m = \min\{k \geq 1 :\ Q(m,k) \geq \tilde{c} g(m,k)\}, \tag{2.3}$$

where

$$g(m,k) = \sqrt{m}\,(1 + k/m)\,(k/(k+m))^{\gamma} \quad \text{for } \gamma \in [0, 1/2) \tag{2.4}$$

and $\tilde{c} = \tilde{c}(\alpha, \gamma)$ is a critical constant derived from the asymptotic distribution of the detector under the null hypothesis. They showed that for the stopping time τ_m under the more restrictive *local change* assumption $\Delta_m \to 0$ and for early change alternatives, i.e.,

$$k^* = \mathcal{O}\left(m^{\beta}\right) \quad \text{with some } 0 \leq \beta < \left(\frac{\frac{1}{2} - \gamma}{1 - \gamma}\right)^2,$$

one can find (deterministic) sequences a_m and b_m such that $(\tau_m - a_m)/b_m$ is asymptotically normal.

Our aim is now to show a similar result for the Page CUSUM detector under the assumptions stated above, including in particular *fixed change* alternatives, and extend it to changes up to an order of m^{β} with β arbitraryly close to 1 (see Assumption (A3)). We define the detector based on Page's CUSUM for the one-sided hypotheses as

$$S(m,k) = Q(m,k) - \min_{0 \leq i \leq k} Q(m,i)$$

and the corresponding stopping time with g from (2.4) as

$$\tau_m^{\text{Page}} = \min\{k \geq 1 :\ S(m,k) \geq c g(m,k)\},$$

where for a given confidence level $\alpha \in (0,1)$ according to Fremdt (2012) the critical value $c = c(\alpha,\gamma)$ can be derived from choosing z_α such that

$$\lim_{m\to\infty} P\left(\frac{1}{\hat{\sigma}_m}\sup_{1\le k<\infty}\frac{S(m,k)}{g(m,k)} > z_\alpha\right) = P\left(\sup_{0<t<1}\frac{1}{t^\gamma}\left[W(t) - \inf_{0\le s\le t}\left(\frac{1-t}{1-s}W(s)\right)\right] > z_\alpha\right) = \alpha.$$

A table containing simulated versions of the critical values $c(\alpha,\gamma)$ for a selection of values for γ and α can be found in Fremdt (2012). From Assumption (A3) we have a given order of the change-point k^* in terms of m depending on the exponent β. As we will see later on the asymptotic distribution of the stopping times depends crucially on the decay of the sequence Δ_m, which is implicitly allowed via Assumptions (A4) and (A5). This dependence can be expressed in terms of the asymptotic behaviour of the quantities $\Delta_m m^{\gamma-1/2} k^{*1-\gamma}$ and due to Assumption (A3) consequently $\Delta_m m^{\beta(1-\gamma)-1/2+\gamma}$. We distinguish the following three cases:

$$m^{\beta(1-\gamma)-1/2+\gamma}\Delta_m \longrightarrow 0, \tag{I}$$

$$m^{\beta(1-\gamma)-1/2+\gamma}\Delta_m \longrightarrow \tilde{c}_1 \in (0,\infty), \tag{II}$$

$$m^{\beta(1-\gamma)-1/2+\gamma}\Delta_m \longrightarrow \infty. \tag{III}$$

Remark 2.1. *a) We note that independently of the asymptotic behaviour of* Δ_m *(only assuming* (A4) *and* (A5)*) we have* (I) *for*

$$0 \le \beta < \frac{\frac{1}{2}-\gamma}{1-\gamma}. \tag{Ia}$$

Therefore under (Ia) *without additional knowledge of the exact time and amount of change using the following results one can, e.g., derive confidence intervals for the stopping times.*

b) Under (II) *because of* (A3)

$$\Delta_m m^{\gamma-1/2} k^{*1-\gamma} \longrightarrow \theta^{1-\gamma}\tilde{c}_1 = c_1 \in (0,\infty). \tag{2.5}$$

To state our main result we first introduce the distribution function $\overline{\Psi}$ depending on the given case (I), (II) or (III). Under (II) denote by d_1 the unique solution of

$$d_1 = 1 - \frac{c}{c_1} d_1^{1-\gamma}. \tag{2.6}$$

For all real x let

$$\overline{\Psi}(x) = \begin{cases} \Phi(x), & \text{under (I)}, \\ P\left(\sup_{d_1<t<1} W(t) \leq x\right), & \text{under(II)}, \\ P\left(\sup_{0<t<1} W(t) \leq x\right) = \begin{cases} 0, & x < 0, \\ 2\Phi(x) - 1, & x \geq 0, \end{cases} & \text{under (III)}, \end{cases}$$

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

THEOREM 2.2. *Let $\{X_n\}_{n\geq 1}$ be a sequence of random variables according to* (2.1) *such that* (A1) – (A5) *are satisfied and let $\gamma \in [0, 1/2)$. Then for all real x under H_A*

$$\lim_{m\to\infty} P\left(\frac{\tau_m^{\text{Page}} - a_m}{b_m} \leq x\right) = 1 - \overline{\Psi}(-x) = \Psi(x) \tag{2.7}$$

where a_m is the unique solution of

$$a_m = \left(\frac{cm^{1/2-\gamma}}{\Delta_m} + \frac{k^*}{a_m^\gamma}\right)^{\frac{1}{1-\gamma}}$$

and

$$b_m = \sigma\sqrt{a_m}\Delta_m^{-1}\left(1 - \gamma\left(1 - \frac{k^*}{a_m}\right)\right)^{-1}.$$

3 A SMALL SIMULATION STUDY

In this section we want to present the results of a small simulation study to illustrate the theoretical result from Section 2. The simulations were carried out for various types of sequences $\{\varepsilon_i\}_{i=1,2,\ldots}$ all leading to similar results. We will therefore only present results for $\mu = 0$ using a GARCH(1,1) sequence

$$\varepsilon_i = \sigma_i z_i, \quad \sigma_i = \overline{\omega} + \overline{\alpha} z_{i-1}^2 + \overline{\beta}\sigma_{i-1}^2, \tag{3.8}$$

where $\{z_i\}_{i=1,2,\ldots}$ are i.i.d. standard normally distributed and the parameters were specified as

$$\overline{\omega} = 0.5,\ \overline{\alpha} = 0.2 \text{ and } \overline{\beta} = 0.3, \tag{3.9}$$

which implies (unconditional) unit variance. For all presented results a fixed change alternative with $\Delta_m = 1$ was considered which implies that the behaviour of $m^{\beta(1-\gamma)-1/2+\gamma}\Delta_m$ and hereby the determination of the corresponding case (I) – (III) depends only on the exponent which will be denoted by η. The cases hence correspond to $\eta \lesseqgtr 0$. In all presented figures the density of the limit distribution is plotted as a solid line and denoted by $\Psi^{(\mathrm{I})}$, $\Psi^{(\mathrm{II})}_{d_1}$ and $\Psi^{(\mathrm{III})}$, resp..

Figures 1–3 show the estimated density plots for constant changes, i.e., $\beta = 0$ in (A3), choosing $k^* = \theta = 1, 100$ and 200. These belong to case (I) for all values of γ and therefore have the standard normal distribution as a limit. For $k^* = 1$ a fast and clear convergence against the standard normal can be found for $\gamma = 0.00$ and $\gamma = 0.25$. For $\gamma = 0.45$ a deviation is visible that can be explained by a slower convergence due to the transition from case (I) to case (II) (and (III)) in $\eta = 0$ and thus for $\beta = 1/11$. E.g., for $m = 100{,}000$ we have $m^{1/11} = 2.848$, a fast convergence against the standard normal can consequently hardly be expected. The influence of the parameter θ from Assumption (A3) which leads to a bias in the limiting behaviour is shown in Figures 2 and 3. While for $\gamma = 0.00$ the convergence against the standard normal distribution can be seen nicely, for the larger values of γ it is not obvious. The explanation for this is again the influence of γ on η, e.g., we have for $m = 100{,}000$ that $k_1^* = 200m^0$ and $k_2^* = m^{0.460206}$ are two possible alternatives with different limit distributions for larger values of γ. The parameter θ was used in earlier works to justify the assumption of early changes, the results from Figures 1–3 show however that this argument has to be handled with caution.

Figure 4 shows the density plots for $k^* = \lfloor m^{0.45} \rfloor$ which implies $\eta < 0$ for $\gamma = 0.00$ and $\eta > 0$ for $\gamma = 0.25$ and $\gamma = 0.45$. For $\gamma = 0.00$ again the convergence against the standard normal distribution is obvious, for $\gamma = 0.25$ and $\gamma = 0.45$ the convergence away from the standard normal distribution towards $\Psi^{(\mathrm{III})}$ is also clearly visible. Finally Figure 5 treats the cases (II) and (III) for all values of γ. The plots in the left column show case (II), i.e., under our model specification we have $\eta = 0$, corresponding to β taking values $\beta = 1/2, 1/3$ and $1/11$, respectively. The limiting densities show the dependence on d_1 and therefore implicitly on γ. The model setting implies $c_1 = 1$, consequently d_1 takes the values:

γ	0.00	0.25	0.45
$d_1(\gamma)$	0.3714	0.1887	0.1051

In case (III) we considered a change $k^* = \lfloor m^{0.75} \rfloor$ and we find that according to the theoretical results only little influence of γ can be seen and the convergence against the

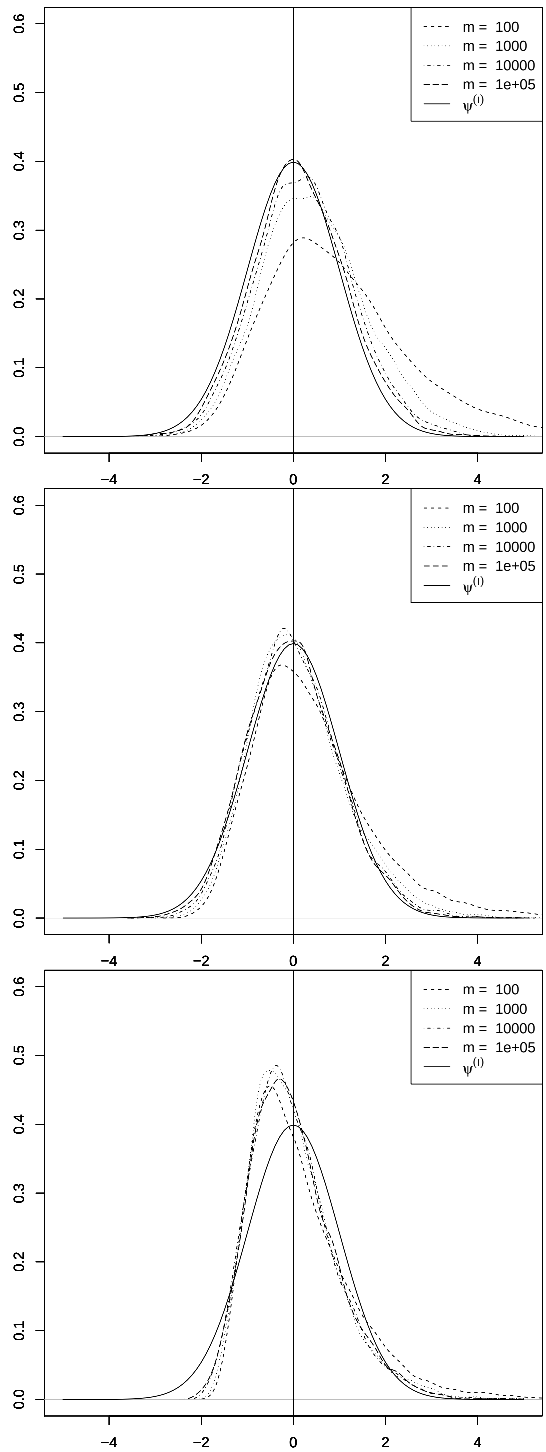


Figure 1: Estimated density plots for the standardized stopping times τ_m^{Page} for a GARCH(1,1) sequence according to (3.8) and (3.9) for $k^* = 1$. The figures from top to bottom correspond to the tuning parameter $\gamma = 0.00, 0.25$ and 0.45.

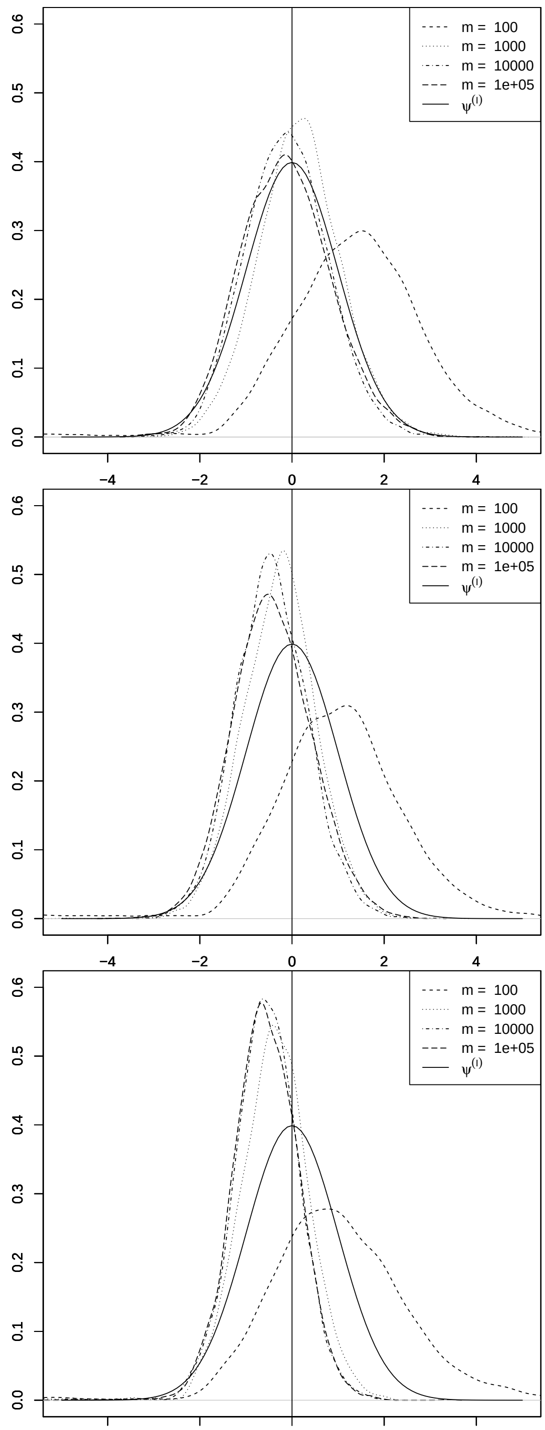


Figure 2: Estimated density plots for the standardized stopping times τ_m^{Page} for a GARCH(1,1) sequence according to (3.8) and (3.9) for $k^* = 100$. The figures from top to bottom correspond to the tuning parameter $\gamma = 0.00, 0.25$ and 0.45.

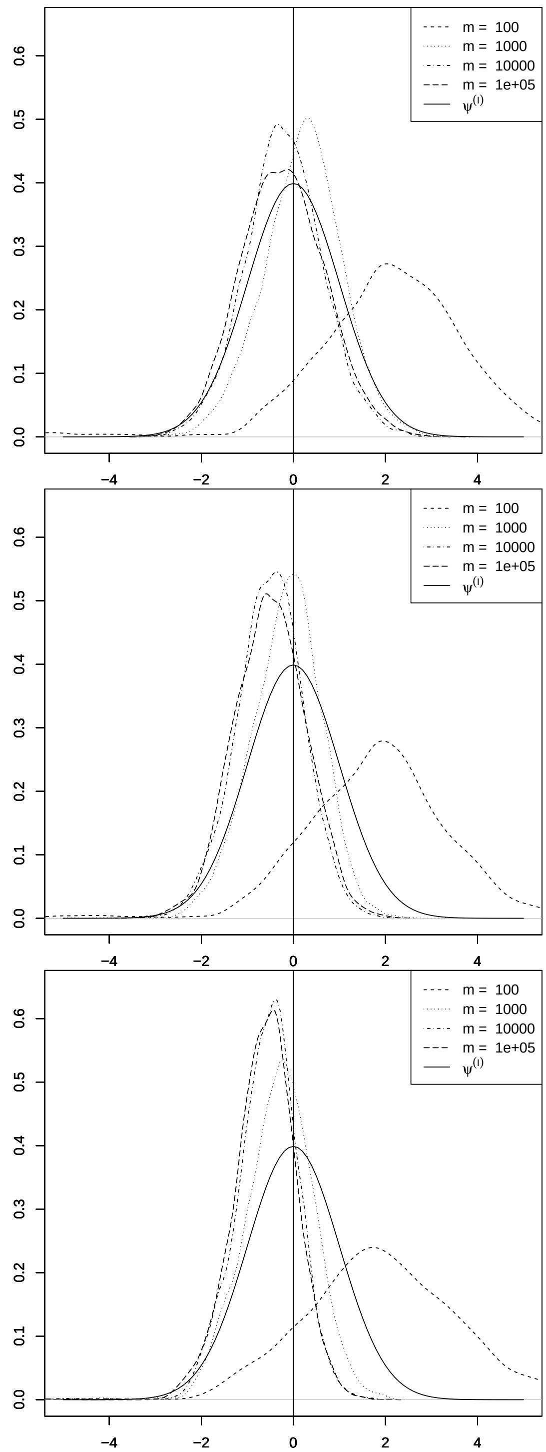


Figure 3: Estimated density plots for the standardized stopping times τ_m^{Page} for a GARCH(1,1) sequence according to (3.8) and (3.9) for $k^* = 200$. The figures from top to bottom correspond to the tuning parameter $\gamma = 0.00, 0.25$ and 0.45.

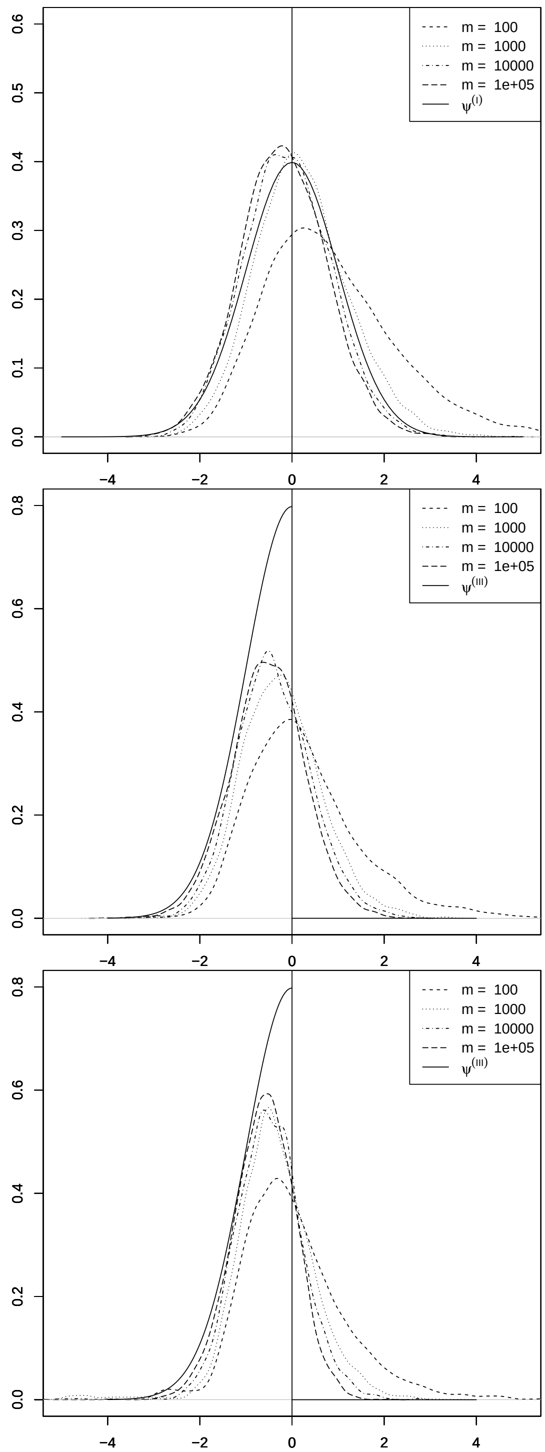


Figure 4: Estimated density plots for the standardized stopping times τ_m^{Page} for a GARCH(1,1) sequence according to (3.8) and (3.9). The change-point was set to $k^* = \lfloor m^{0.45} \rfloor$, the figures from top to bottom correspond to the tuning parameter $\gamma = 0.00, 0.25$ and 0.45 .

limit distribution $\Psi^{(\mathrm{III})}$ is again obvious.

4 Proof of Theorem 2.2

To prove Theorem 2.2, we adopt the method of Aue and Horváth (2004) or Černíková et al. (2011), that is finding a sequence $N = N(m,x)$ such that:

$$P(\tau_m^{\mathrm{Page}} > N) = P\left(\max_{1\le k\le N} \frac{S(m,k)}{c\,g(m,k)} \le 1\right) \longrightarrow \overline{\Psi}(x) \quad \text{for all real } x. \tag{4.10}$$

As can be seen in the proofs this sequence can be chosen as

$$N = N(m,x) = \left(\frac{c\,m^{1/2-\gamma}}{\Delta_m} + \frac{k^*}{a_m^\gamma} - \sigma x \frac{a_m^{1/2-\gamma}(1-\gamma)}{\Delta_m(1-\gamma(1-\frac{k^*}{a_m}))}\right)^{1/(1-\gamma)}. \tag{4.11}$$

Before we start with the proof of Theorem 2.2 we give some facts that will be useful in the proofs:

Remark 4.1. *It is obvious that $Q(m,k)$ can be rewritten as*

$$Q(m,k) = \sum_{i=m+1}^{m+k} \varepsilon_i - \frac{k}{m}\sum_{i=1}^{m} \varepsilon_i + \Delta_m(k-k^*+1)I_{\{k\ge k^*\}} \quad \text{for } k = 1,2,\ldots, \tag{4.12}$$

and consequently with $\overline{\varepsilon}_m = \frac{1}{m}\sum_{\ell=1}^m \varepsilon_\ell$ we have

$$S(m,k) \le \left|\sum_{i=m+1}^{m+k} \varepsilon_i\right| + k\,|\overline{\varepsilon}_m| + \left|\Delta_m(k-k^*+1)I_{\{k\ge k^*\}}\right| + \left|\min_{0\le i\le k} Q(m,i)\right| \tag{4.13}$$

and

$$\frac{\left|\min\limits_{0\le i\le k} Q(m,i)\right|}{g(m,k)} \le \max_{0\le i\le k} \frac{\left|\sum\limits_{j=m+1}^{m+i} \varepsilon_j\right|}{g(m,i)} + \max_{0\le i\le k} \frac{i\,|\overline{\varepsilon}_m|}{g(m,i)} + \max_{0\le i\le k} \frac{\left|\Delta_m(i-k^*+1)I_{\{i\ge k^*\}}\right|}{g(m,i)}. \tag{4.14}$$

Proposition 4.2. *Introducing the notation $r_m \approx s_m$ for $r_m = s_m(1+o(1))$ we get under the assumptions of Theorem 2.2 that*

$$a_m \approx \begin{cases} \left(\frac{c\,m^{1/2-\gamma}}{\Delta_m}\right)^{\frac{1}{1-\gamma}}, & \textit{under } (\mathrm{I}),\\ d_2\,k^*, & \textit{under } (\mathrm{II}) \textit{ with } d_2 = \left(\frac{c}{c_1} + d_1^\gamma\right)^{\frac{1}{1-\gamma}} \textit{ and}\\ k^*, & \textit{under } (\mathrm{III}). \end{cases}$$

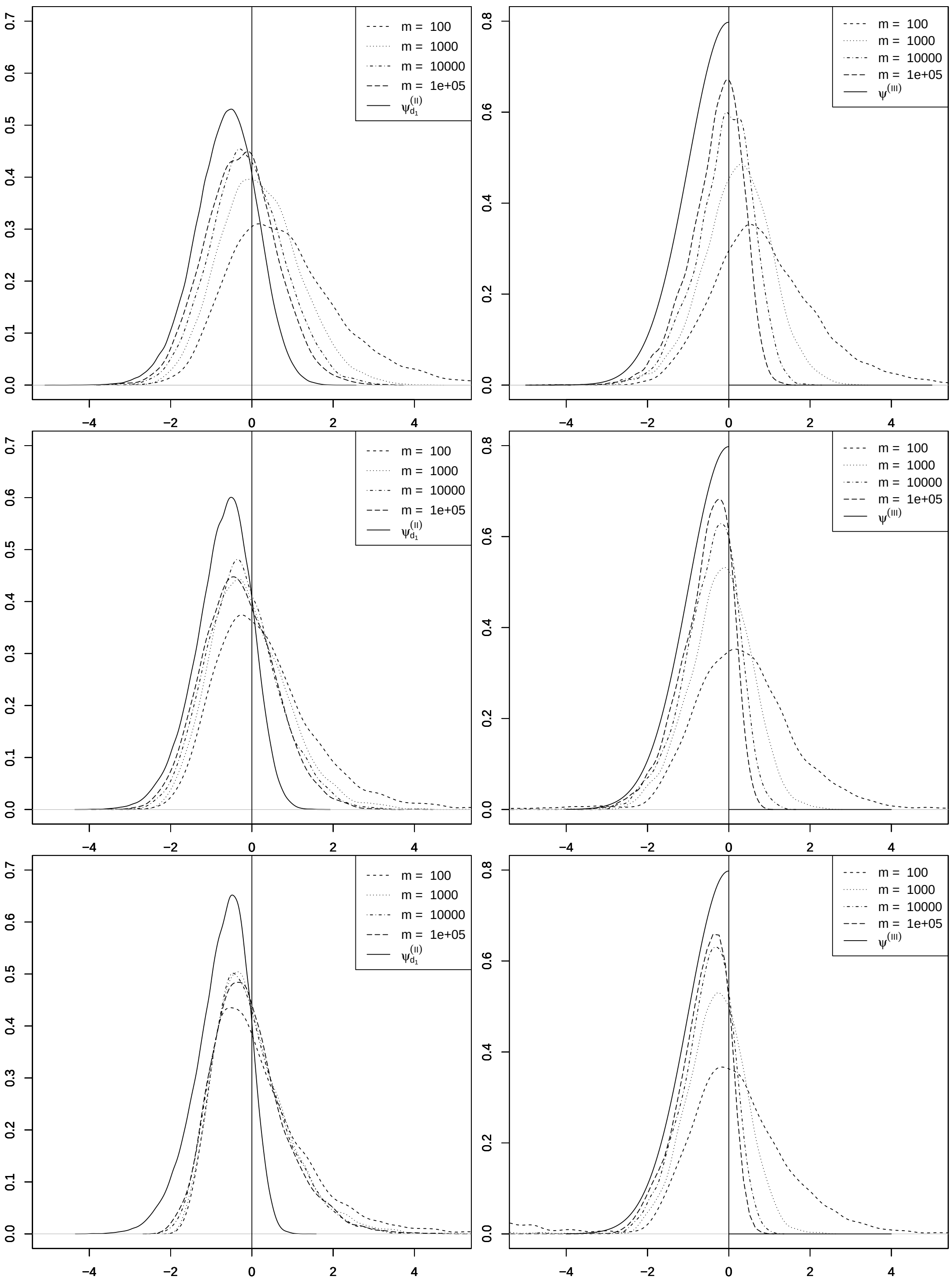


Figure 5: Estimated density plots for the standardized stopping times τ_m^{Page} for a GARCH(1,1) sequence according to (3.8) and (3.9). The rows from top to bottom correspond to the tuning parameter $\gamma = 0.00, 0.25$ and 0.45, the left column corresponds to case (II) with $\Delta_m = 1$ and $\beta = (1/2 - \gamma)/(1 - \gamma)$, the right column to $k^* = \lfloor m^{0.75} \rfloor$ and thus case (III) for all values of γ.

PROOF: From the definition of a_m it follows obviously that this definition is equivalent to

$$a_m = \frac{c\,m^{1/2-\gamma}}{\Delta_m} a_m^\gamma + k^* \tag{4.15}$$

which yields $a_m \geq k^*$ and thus $\frac{k^*}{a_m} \leq 1$. Now under (I) the assertion follows directly by

$$a_m = \frac{c\,m^{1/2-\gamma}}{\Delta_m} a_m^\gamma \left(1 + \frac{\Delta_m {k^*}^{1-\gamma}}{c\,m^{1/2-\gamma}} \left(\frac{k^*}{a_m}\right)^\gamma\right) = \left(\frac{c\,m^{1/2-\gamma}}{\Delta_m}\right)^{\frac{1}{1-\gamma}} (1 + o\,(1)).$$

Under (II) we have $\lim_{m\to\infty} k^*/a_m = d_1$ (cf. Lemma 4.3 a) (iv)) and with (2.5) consequently

$$a_m = k^* \left(c\,\Delta_m^{-1} m^{1/2-\gamma} {k^*}^{\gamma-1} + \left(\frac{k^*}{a_m}\right)^\gamma\right)^{\frac{1}{1-\gamma}} \approx k^* \left(\frac{c}{c_1} + d_1^\gamma\right)^{\frac{1}{1-\gamma}} = d_2 k^*$$

Finally under (III) we have from the definition of a_m

$$\frac{a_m}{k^*} = \left(\left(\frac{k^*}{a_m}\right)^\gamma + c\,\Delta_m^{-1} m^{1/2-\gamma} {k^*}^{\gamma-1}\right)^{\frac{1}{1-\gamma}} = \mathcal{O}\,(1)\,,$$

leading to

$$a_m^{1-\gamma} = \frac{k^*}{a_m^\gamma} \left(1 + \frac{c\,m^{1/2-\gamma}}{\Delta_m {k^*}^{1-\gamma}} \left(\frac{a_m}{k^*}\right)^\gamma\right) = \frac{k^*}{a_m^\gamma} (1 + o\,(1))\,,$$

and thus concluding the proof. □

LEMMA 4.3. *Let* $\gamma \in [0, 1/2)$ *and let* (A3) – (A5) *be satisfied. Then*

a) *(i)* $a_m/m \longrightarrow 0$,

(ii) $\sqrt{a_m}\Delta_m \longrightarrow \infty$,

(iii) $k^*/m \longrightarrow 0$,

(iv) $k^*/a_m \longrightarrow \begin{cases} 0, & \textit{under (I)}, \\ d_1 \in (0,1), & \textit{under (II) and} \\ 1, & \textit{under (III)}. \end{cases}$

b) $N/a_m \longrightarrow 1$ *and consequently the statements of part a) still hold after substitution of* a_m *by* N.

c) Furthermore

$$\lim_{m\to\infty}\frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\left(c-\frac{\Delta_m}{m^{1/2-\gamma}}\left(N^{1-\gamma}-\frac{k^*}{N^\gamma}\right)\right)=x \quad \textit{for all real } x. \tag{4.16}$$

PROOF: a) Part (iii) follows directly from Assumption (A1).
(i) Under (I) with Proposition 4.2 and Assumption (A4)

$$\frac{a_m}{m}\approx\left(\frac{c}{\sqrt{m}\Delta_m}\right)^{\frac{1}{1-\gamma}}\longrightarrow 0,$$

under (II) and (III) we get the result again with Proposition 4.2 and part (iii) of this Lemma from

$$\frac{a_m}{m}\approx\begin{cases} d_2k^*/m, & \text{under (II) and}\\ k^*/m & \text{under (III).}\end{cases}$$

(ii) Because $a_m\geq\left(c\,\Delta_m^{-1}m^{1/2-\gamma}\right)^{\frac{1}{1-\gamma}}$ we have by Assumption (A4)

$$\sqrt{a_m}\Delta_m\geq\left(\left(c\,\Delta_m^{-1}m^{1/2-\gamma}\right)\Delta_m^{2(1-\gamma)}\right)^{\frac{1}{2(1-\gamma)}}=c^{\frac{1}{2(1-\gamma)}}\left(\sqrt{m}\Delta_m\right)^{\frac{1/2-\gamma}{1-\gamma}}\longrightarrow\infty$$

(iv) Under (I) and (III) the result follows directly from Proposition 4.2 and its proof, under (II) consider

$$\frac{a_m}{k^*}=\frac{c\,m^{1/2-\gamma}}{\Delta_m{k^*}^{1-\gamma}}\left(\frac{a_m}{k^*}\right)^\gamma+1=\frac{c}{c_1}\left(\frac{a_m}{k^*}\right)^\gamma+1+o\left(1\right).$$

Now it can easily be seen that because of the definition of a_m the term a_m/k^* solving the equation above converges towards a real number $d_1^{-1}\in(1,\infty)$ and hence $k^*/a_m\longrightarrow d_1\in(0,\infty)$. Hence we find d_1 as the solution of

$$d_1=\lim_{m\to\infty}k^*/a_m=1-\lim_{m\to\infty}c\,\Delta_m^{-1}m^{1/2-\gamma}{k^*}^{\gamma-1}\left(k^*/a_m\right)^{1-\gamma}=1-\left(c/c_1\right)d_1^{1-\gamma}.$$

b) It is enough to consider

$$\frac{N^{1-\gamma}}{a_m^{1-\gamma}}=1-\sigma x\frac{a_m^{1/2-\gamma}(1-\gamma)}{a_m^{1-\gamma}\Delta_m\left(1-\gamma\left(1-k^*/a_m\right)\right)}=1-\sigma x\frac{1-\gamma}{\sqrt{a_m}\Delta_m\left(1-\gamma\left(1-k^*/a_m\right)\right)}.$$

But aside from Lemma 4.3 a) (ii) giving us $\sqrt{a_m}\Delta_m\longrightarrow\infty$ we have

$$1-\gamma\left(1-k^*/a_m\right)\longrightarrow\begin{cases}1-\gamma>0, & \text{under (I) and (I),}\\ 1-\gamma(1-d_1)>0, & \text{under (II),}\\ 1, & \text{under (III),}\end{cases}$$

which yields the desired result.
c) To ease the notation we first introduce

$$u_\gamma(s,t) = \frac{1-\gamma}{1-\gamma\,(1-s/t)}. \tag{4.17}$$

By inserting the definition of N in $N^{1-\gamma}$, from (4.11), we get

$$\begin{aligned}
&(N/m)^{\gamma-1/2}\left(c-\Delta_m m^{\gamma-1/2}\left(N^{1-\gamma}-k^*/N^\gamma\right)\right)\\
=&\frac{\Delta_m k^*}{\sqrt{N}}\left(1-\left(N^{1-\gamma}/a_m^{1-\gamma}\right)^{\frac{\gamma}{1-\gamma}}\right)+\sigma x\,(a_m/N)^{1/2-\gamma}\,u_\gamma(k^*,a_m)\\
=&\frac{\Delta_m k^*}{\sqrt{N}}\left(1^{\frac{\gamma}{1-\gamma}}-\left(1+\frac{N^{1-\gamma}-a_m^{1-\gamma}}{a_m^{1-\gamma}}\right)^{\frac{\gamma}{1-\gamma}}\right)+\sigma x\,(a_m/N)^{1/2-\gamma}\,u_\gamma(k^*,a_m)\\
=&A_1,
\end{aligned}$$

so by the mean value theorem we can find ξ_m between 1 and $(N/a_m)^{1-\gamma}$ (which satisfies $\xi_m \to 1$ because of part b)) such that

$$\begin{aligned}
A_1 =&\sigma x\,(a_m/N)^{1/2-\gamma}\,u_\gamma(k^*,a_m)+\frac{\Delta_m k^*}{\sqrt{N}a_m^{1-\gamma}}\left(\sigma x\frac{a_m^{1/2-\gamma}}{\Delta_m}u_\gamma(k^*,a_m)\right)\frac{\gamma}{1-\gamma}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\\
=&\sigma x\,(a_m/N)^{1/2-\gamma}\,u_\gamma(k^*,a_m)\left(1+\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\frac{a_m^\gamma}{N^\gamma}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\right)\\
=&\sigma x\left(\frac{a_m}{N}\right)^{1/2-\gamma}\left(1+\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\frac{a_m^\gamma}{N^\gamma}\xi_m^{\frac{2\gamma-1}{1-\gamma}}\right)\left(1-\frac{\gamma}{1-\gamma}\frac{k^*}{a_m}\right)^{-1} \quad \longrightarrow \sigma x,
\end{aligned}$$

which completes the proof of Lemma 4.3. □

To prove Theorem 2.2 we formulate a set of lemmas containing stepwise approximations of the detector that finally give us the desired asymptotics. For these first steps we follow again the outline of the proofs in Aue and Horváth (2004). The first step of the proof is to show that the observations before the change-point do not have an impact on the asymptotics under the alternative.

LEMMA 4.4. *Let* $\gamma \in [0,1/2)$. *If* (A1) – (A5) *hold, then*

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\left(\max_{1\le k<k^*}\frac{S(m,k)}{g(m,k)}-\frac{\Delta_m(N-k^*)}{\sqrt{m}\,(N/m)^\gamma}\right)\xrightarrow{P}-\infty.$$

PROOF: First we note that $g(m,k) = m^{1/2-\gamma}\left(1+k/m\right)^{1-\gamma} k^{\gamma} \geq m^{1/2-\gamma}k^{\gamma}$. Because the indicator function in (4.12) equals zero for $1 \leq k < k^*$ and because of (4.13) it is enough to consider

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\leq k<k^*} \frac{1}{g(m,k)} \left| \sum_{i=m+1}^{m+k} \varepsilon_i \right| + \left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\leq k<k^*} \frac{k\,|\bar{\varepsilon}_m|}{g(m,k)} \\
&+ \left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\leq k<k^*} \frac{1}{g(m,k)} \left| \min_{0\leq i\leq k} Q(m,i) \right| + \left(\frac{N}{m}\right)^{\gamma-1/2} \frac{\Delta_m(N-k^*)}{\sqrt{m}\,(N/m)^{\gamma}} \\
&= A_2 + A_3 + A_4 + A_5
\end{aligned}$$

We will first show that all but the deterministic term A_5 are stochastically bounded and therefore they do not contribute to the asymptotics. Then it is sufficient to show the divergence of A_5 to prove the lemma. We begin with the term A_2 and replace the partial sum of the error terms by a Wiener process and have with Lemma 4.3 a)(iv) and b)

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

We note that

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\leq k<k^*} \frac{W_m(k)}{g(m,k)} \leq \sup_{0<t\leq k^*} \frac{W_m(t)}{\sqrt{N}\,(t/N)^{\gamma}} \stackrel{\mathcal{D}}{=} \sup_{0<t\leq k^*/N} \frac{W(t)}{t^{\gamma}} = \mathcal{O}_P(1),$$

where the equality in distribution comes from the scaling property of the Wiener process. For A_3 Lemma 4.3 a) (iii), (iv) and Assumption (A1) yield

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\leq k<k^*} \frac{k\,|\bar{\varepsilon}_m|}{g(m,k)} \leq \frac{k^{*1-\gamma}}{mN^{1/2-\gamma}} \left| \sum_{\ell=1}^{m} \varepsilon_\ell \right| = \mathcal{O}_P\left((k^*/N)^{1/2-\gamma} (k^*/m)^{1/2} \right) = o_P(1).$$

For A_4 it follows by (4.14) that

$$\max_{1\leq k<k^*} \frac{1}{g(m,k)} \left| \min_{0\leq i\leq k} Q(m,i) \right| \leq \max_{1\leq k<k^*} \frac{1}{g(m,k)} \left| \sum_{j=m+1}^{m+k} \varepsilon_j \right| + \max_{1\leq k<k^*} \frac{k\,|\bar{\varepsilon}_m|}{g(m,k)} = \mathcal{O}_P(1).$$

Thus we only have to consider the deterministic term

$$A_5 = \left(\frac{N}{m}\right)^{\gamma-1/2} \frac{\Delta_m(N-k^*)}{\sqrt{m}\,(N/m)^{\gamma}} = \frac{\Delta_m}{N^{1/2-\gamma}} \left(N^{1-\gamma} - \frac{k^*}{N^{\gamma}} \right) = \Delta_m \sqrt{N} \left(1 - \frac{k^*}{N} \right).$$

It is obvious that the right hand side under (I) and (II) tends to infinity. Under (III) we have with u_γ from (4.17) that

$$\frac{\Delta_m}{N^{1/2-\gamma}}\left(N^{1-\gamma}-\frac{k^*}{N^\gamma}\right)=\frac{\Delta_m}{N^{1/2-\gamma}}\left(\frac{c\,m^{1/2-\gamma}}{\Delta_m}-\sigma x\frac{a_m^{1/2-\gamma}}{\Delta_m}u_\gamma(k^*,N)+\frac{k^*}{N^\gamma}\left(\left(\frac{N}{a_m}\right)^\gamma-1\right)\right)$$
$$=c\left(\frac{m}{N}\right)^{1/2-\gamma}-\sigma x\left(\frac{a_m}{N}\right)^{1/2-\gamma}u_\gamma(k^*,N)+\frac{\Delta_m k^*}{\sqrt{N}}\left(\left(\frac{N}{a_m}\right)^\gamma-1\right).$$

From Lemma 4.3 a) (i) it follows directly that the first term diverges, i.e., $c\,(m/N)^{1/2-\gamma}\longrightarrow\infty$, for the second term by Lemma 4.3 a) (iv) and b) it is clear that this term is bounded. The third term can be treated analogously to the proof of Lemma 4.3 c) applying the mean value theorem:

$$\begin{aligned}&\frac{\Delta_m k^*}{\sqrt{N}}\left(\left(1+\frac{N^{1-\gamma}-a_m^{1-\gamma}}{a_m^{1-\gamma}}\right)^{\gamma/(1-\gamma)}-1\right)\\
=&\frac{\Delta_m k^*}{\sqrt{N}a_m^{1-\gamma}}\left(-\sigma x\frac{a_m^{1/2-\gamma}}{\Delta_m}u_\gamma(k^*,N)\right)\frac{\gamma}{1-\gamma}\xi_m^{(2\gamma-1)/(1-\gamma)}\\
=&-\sigma x\,u_\gamma(k^*,N)\frac{k^*}{\sqrt{Na_m}}\xi_m^{(2\gamma-1)/(1-\gamma)}\\
=&\mathcal{O}(1).\end{aligned}$$

This gives us the desired result. □

The next step is an approximation of our detector by functionals of a sequence of Wiener processes. To ease the notation we define

$$W_D(m,j)=\sigma W_m(j)+(j-k^*+1)\Delta_m I_{\{j\geq k^*\}} \tag{4.18}$$

and

$$W_S(m,k)=W_D(m,k)-\min_{0\leq i\leq k}W_D(m,i). \tag{4.19}$$

LEMMA 4.5. *Let* $\gamma\in[0,1/2)$ *and Assumptions* (A1) – (A5) *hold. Then*

$$\left(\frac{N}{m}\right)^{\gamma-1/2}\max_{k^*\leq k\leq N}\frac{1}{g(m,k)}\left|S(m,k)-W_S(m,k)\right|=o_P(1).$$

Proof: The deterministic terms cancel out thus using

$$\begin{aligned}
&\max_{k^*\le k\le N} \frac{1}{g(m,k)} \left| \min_{0\le i\le k} W_D(m,i) - \min_{0\le i\le k} Q(m,i) \right| \\
&\le \max_{k^*\le k\le N} \frac{1}{g(m,k)} \max_{0\le i\le k} \left| \sigma W_m(i) - \left(\sum_{j=m+1}^{m+i} \varepsilon_j - \frac{i}{m} \sum_{\ell=1}^{m} \varepsilon_\ell \right) \right| \\
&\le \max_{1\le k\le N} \frac{1}{g(m,k)} \left(\left| \sum_{j=m+1}^{m+k} \varepsilon_j - \sigma W_m(k) \right| + \frac{k}{m} \left| \sum_{\ell=1}^{m} \varepsilon_\ell \right| \right)
\end{aligned}$$

it is sufficient to consider

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\le k\le N} \frac{1}{g(m,k)} \left| \sum_{j=m+1}^{m+k} \varepsilon_j - \sigma W_m(k) \right| = A_6,$$

for which with Assumption (A2) and because $y^{1/\nu-\gamma}$ is monotone we have

$$\begin{aligned}
A_6 =& \mathcal{O}_P(1) \max_{1\le k\le N} \frac{k^{1/\nu}}{\sqrt{N}\,(k/N)^\gamma} = \mathcal{O}_P(1) N^{\gamma-1/2} \max\left\{1, N^{1/\nu-\gamma}\right\} \\
=& \mathcal{O}_P(1) \max\left\{N^{\gamma-1/2}, N^{1/\nu-1/2}\right\} = \mathcal{O}_P(1) o(1) \\
=& o_P(1),
\end{aligned}$$

and

$$\begin{aligned}
\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{1\le k\le N} \frac{k}{m} \left| \sum_{\ell=1}^{m} \varepsilon_\ell \right| \Big/ g(m,k) \le& \left(\frac{N}{m}\right)^{\gamma-1/2} \frac{1}{m} \left| \sum_{\ell=1}^{m} \varepsilon_\ell \right| \max_{1\le k\le N} \frac{k^{1-\gamma}}{m^{1/2-\gamma}} \\
=& \frac{\sqrt{N}}{m} \mathcal{O}_P\left(\sqrt{m}\right) = \mathcal{O}_P\left(\sqrt{N/m}\right) \\
=& o_P(1).
\end{aligned}$$

□

The boundary function $g(m,k)$ can be replaced by an asymptotically equivalent function that simplifies the coming calculations.

Lemma 4.6. *Let* $\gamma \in [0, 1/2)$ *and Assumptions* (A1) – (A5) *hold and define*

$$h(m,k) = \left| \frac{1}{g(m,k)} - \frac{1}{\sqrt{m}\,(k/m)^\gamma} \right|.$$

Then

$$\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \left|W_S(m,k)\right| h(m,k) = o_P(1).$$

PROOF: Aue and Horváth (2004) showed that

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

For the deterministic term $(k-k^*+1)\Delta_m$ we get

$$\begin{aligned}
&\left(\frac{N}{m}\right)^{\gamma-1/2} \max_{k^*\le k\le N} \Delta_m(k-k^*+1)h(m,k)\\
=&\frac{\Delta_m}{N^{1/2-\gamma}} \max_{k^*\le k\le N} \frac{k-k^*+1}{k^\gamma}\left(1^{1-\gamma}-\left(\frac{m}{k+m}\right)^{1-\gamma}\right)\\
=&\frac{\Delta_m}{N^{1/2-\gamma}} \max_{k^*\le k\le N} k^{1-\gamma}\left(1^{1-\gamma}-\left(\frac{m}{k+m}\right)^{1-\gamma}\right)\\
=&A_7.
\end{aligned}$$

By application of the mean value theorem for every $k^*\le k\le N$ we can find a real number $\xi_{m,k}$ satisfying $m/(m+k)<\xi_{m,k}<1$ such that

$$A_7 = \frac{\Delta_m}{N^{1/2-\gamma}} \max_{k^*\le k\le N} \frac{k-k^*+1}{k^\gamma}\frac{k}{k+m}(1-\gamma)\xi_{m,k}^{-\gamma}.$$

Because $\xi_{m,k}^{-\gamma}\le (m/(m+k))^{-\gamma}$ and $(k-k^*+1)\left(k/(k+m)\right)^{1-\gamma}$ is strictly increasing in k we have

$$\begin{aligned}
A_7 \le&\frac{\Delta_m}{N^{1/2-\gamma}m^\gamma} \max_{k^*\le k\le N}(k-k^*+1)\left(\frac{k}{k+m}\right)\\
=&\frac{\Delta_m}{N^{1/2-\gamma}}(N-k^*+1)\left(\frac{N}{N+m}\right)^{1-\gamma}\\
\le&\frac{\Delta_m\sqrt{N}(N-k^*+1)}{m}\\
=&\frac{\Delta_m\sqrt{N}(N-k^*)}{m}+\frac{\Delta_m\sqrt{N}}{m}.
\end{aligned}$$

Here the last term clearly tends to 0 and for the first term we get by Lemma 4.3 b) and from (4.15)

$$\frac{\Delta_m\sqrt{N}(N-k^*)}{m} \approx \frac{\Delta_m\sqrt{a_m}(a_m-k^*)}{m} = \frac{\Delta_m\sqrt{a_m}}{m}\frac{c\,m^{1/2-\gamma}a_m^\gamma}{\Delta_m} = c\left(\frac{a_m}{m}\right)^{1/2+\gamma} = o(1),$$

where the last equality comes from Lemma 4.3 a) (i). For the minimum term the claim follows with the same arguments and this is completing the proof. □

Before we can state the next lemma we again have to introduce some notation. For $\delta \in (0,1)$ we define

$$N_\delta = (1-\delta)N, \quad \overline{N} = N - k^* - 1 \quad \text{and} \quad \overline{N}_\delta = (1-\delta)\overline{N}.$$

Lemma 4.7. *Let $\gamma \in [0, 1/2)$ and (A1) – (A5) hold. Then for every $\delta \in (0,1)$*

a) $\lim_{m\to\infty} P\left(\max_{k^* \le k \le N} \frac{W_S(m,k)}{\sqrt{m}(k/m)^\gamma} = \max_{N_\delta \le k \le N} \frac{W_S(m,k)}{\sqrt{m}(k/m)^\gamma} \right) = 1,$

b) $\lim_{m\to\infty} P\left(\max_{k^* \le j \le \overline{N}} \frac{\sigma W_m(j) + \Delta_m j}{\sqrt{N}} = \max_{\overline{N}_\delta \le j \le \overline{N}} \frac{\sigma W_m(j) + \Delta_m j}{\sqrt{N}} \right) = 1.$

Proof: a) We note (cf. Aue and Horváth (2004)) that

$$\max_{k^* \le k \le N} \frac{|W(k)|}{\sqrt{m}\,(k/m)^\gamma} = \mathcal{O}_P\left(\left(\frac{N}{m} \right)^{1/2-\gamma} \right) = o_P\left(\frac{\Delta_m N^{1-\gamma}}{m^{1/2-\gamma}} \right). \tag{4.20}$$

Now it can be seen easily that this result also holds true for the extended range $0 \le k \le N$. Then

$$\begin{aligned} &P\left(\max_{k^* \le k \le N} \frac{W_S(m,k)}{\sqrt{m}\,(k/m)^\gamma} > \max_{N_\delta \le k \le N} \frac{W_S(m,k)}{\sqrt{m}\,(k/m)^\gamma} \right) \\ &= P\left(\bigcup_{k=k^*}^{\lfloor N_\delta \rfloor} \bigcap_{\ell=N_\delta}^{N} \left\{ k^{-\gamma} W_S(m,k) > \ell^{-\gamma} W_S(m,\ell) \right\} \right) \\ &\le P\left(\bigcup_{k=k^*}^{\lfloor N_\delta \rfloor} \left\{ k^{-\gamma} W_S(m,k) > N^{-\gamma} W_S(m,N) \right\} \right). \end{aligned}$$

We can rewrite

$$k^{-\gamma} W_S(m,k) > N^{-\gamma} W_S(m,N)$$

as

$$(k/N)^{-\gamma}\left(W_D(m,k) - \min_{0\le i\le k} W_D(m,i) \right) - \sigma W_m(N) + \min_{0\le i\le N} W_D(m,i) > \Delta_m (N - k^* + 1), \tag{4.21}$$

where the term on the right can be replaced by $\Delta_m N$ to get an upper bound for the probability of (4.21). Because for $k \in [k^*, N_\delta)$

$$\left(\frac{k}{N}\right)^{-\gamma} \min_{0\le i\le k} \left(\sigma W_m(i) + (i-k^*+1)\Delta_m I_{\{i\ge k^*\}}\right) \ge \min_{0\le i\le k} \frac{\sigma W_m(i)}{(i/N)^\gamma} \ge -\max_{0\le i\le N} \frac{\sigma(-W_m(i))}{(i/N)^\gamma},$$

it follows that

$$P\left(\bigcup_{k=k^*}^{\lfloor N_\delta\rfloor} \left\{k^{-\gamma} W_S(m,k) > N^{-\gamma} W_S(m,N)\right\}\right)$$

$$\le P\left(\max_{k^*\le k<N_\delta}\left(\frac{\sigma W_m(k)}{(k/N)^\gamma\,\Delta_m N}\right) + \max_{k^*\le k<N_\delta}\left(\frac{k-k^*+1}{(k/N)^\gamma\,N}\right) + \max_{0\le i\le N}\left(\frac{\sigma(-W_m(i))}{\Delta_m N\,(i/N)^\gamma}\right) - \sigma\frac{W_m(N)}{\Delta_m N} > 1\right).$$

Now (4.20) yields

$$\max_{k^*\le k<N_\delta} \frac{\sigma W_m(k)}{(k/N)^\gamma\,\Delta_m N} = o_P(1),$$
$$\max_{0\le i\le N} \frac{\sigma W_m(i)}{\Delta_m N\,(i/N)^\gamma} = o_P(1),$$
$$\sigma\frac{W_m(N)}{\Delta_m N} = o_P(1).$$

Thus with

$$\max_{k^*\le k<N_\delta} \frac{k-k^*+1}{(k/N)^\gamma\,N} \le \frac{N_\delta-k^*+1}{(N_\delta/N)^\gamma\,N} = (1-\delta)^{1-\gamma} - \frac{k^*}{(1-\delta)^\gamma N} + \frac{1}{(1-\delta)^\gamma N} < 1,$$

for large enough m, we have

$$P\left(\max_{k^*\le k<N_\delta}\left(\frac{\sigma W_m(k)}{(k/N)^\gamma\,\Delta_m N}\right) + \max_{k^*\le k<N_\delta}\left(\frac{k-k^*+1}{(k/N)^\gamma\,N}\right) + \max_{0\le i\le N}\left(\frac{\sigma(-W_m(i))}{\Delta_m N\,(i/N)^\gamma}\right) - \sigma\frac{W_m(N)}{\Delta_m N} > 1\right) \longrightarrow 0.$$

b) Similarly we get

$$P\left(\max_{k^*\le j\le \overline{N}} \frac{\sigma W_m(j)+\Delta_m j}{\sqrt{N}} > \max_{\overline{N}_\delta\le j\le \overline{N}} \frac{\sigma W_m(j)+\Delta_m j}{\sqrt{N}}\right)$$

$$\leq P\left(\bigcup_{j=k^*}^{\lfloor \overline{N}_\delta \rfloor}\left\{\sigma W_m(j)+\Delta_m j > \sigma W_m(\overline{N})+\Delta_m \overline{N}\right\}\right)$$
$$\leq P\left(\max_{k^*\leq j<\overline{N}_\delta}\frac{\sigma W_m(j)}{\Delta_m \overline{N}}+\max_{k^*\leq j<\overline{N}_\delta}\frac{j}{\overline{N}}-\frac{\sigma W_m(\overline{N})}{\sqrt{\overline{N}}}\frac{1}{\Delta_m\sqrt{\overline{N}}}>1\right).$$

We first consider the term $\Delta_m\sqrt{\overline{N}} = \Delta_m\sqrt{N}\sqrt{1-(k^*+1)/N}$. Under (I) and (II) it is obvious that $\Delta_m\sqrt{\overline{N}} \longrightarrow \infty$, under (III) with Lemma 4.3 b) and (4.15)

$$\Delta_m\sqrt{\overline{N}} \approx \Delta_m\sqrt{a_m-k^*-1} = \Delta_m\sqrt{\frac{cm^{1/2-\gamma}}{\Delta_m}a_m^\gamma - 1} = \sqrt{c\Delta_m m^{1/2}\left(\frac{a_m}{m}\right)^\gamma - \Delta_m^2}.$$

Because of Assumption (A3) it is enough to consider

$$\Delta_m m^{1/2}\left(a_m/m\right)^\gamma \approx \Delta_m m^{1/2-\gamma+\gamma\beta} = \Delta_m m^{\beta(1-\gamma)-1/2+\gamma} m^{2(\gamma\beta+1/2-\gamma)-\beta},$$

but since $2(\gamma\beta+1/2-\gamma)-\beta \geq 0$ under $\gamma < 1/2$ is equivalent to $\beta \leq 1$ we have $\Delta_m\sqrt{\overline{N}} \longrightarrow \infty$. This together with $\overline{N}^{-1/2}W_m(\overline{N}) = \mathcal{O}_P(1)$ gives us

$$\frac{W_m(\overline{N})}{\sqrt{\overline{N}}}\frac{1}{\Delta_m\sqrt{\overline{N}}} = o_P(1).$$

The rest of the proof follows analogously to part a) of this proof. □

LEMMA 4.8. *Let $\gamma \in [0,1/2)$. If* (A1) – (A5) *are satisfied, then*

$$P\left(\max_{k^*\leq k\leq N}\frac{W_S(m,k)}{\sqrt{m}\,(k/m)^\gamma}\leq c\right)\longrightarrow \overline{\Psi}(x) \quad \textit{for all real } x.$$

PROOF: Application of Lemma 4.7 a) and then letting $\delta \downarrow 0$ yields

$$\begin{aligned}
&\lim_{m\to\infty} P\left(\max_{k^*\leq k\leq N}\frac{W_S(m,k)}{\sqrt{m}\,(k/m)^\gamma}\leq c\right)\\
&=\lim_{m\to\infty} P\left(\max_{N_\delta\leq k\leq N}\frac{W_S(m,k)}{\sqrt{m}\,(k/m)^\gamma}\leq c\right)\\
&=\lim_{m\to\infty} P\left(\frac{1}{\sqrt{N}}\left(W_D(m,N)-\min_{0\leq i\leq N}W_D(m,i)\right)\leq c\left(\frac{N}{m}\right)^{\gamma-1/2}\right). \qquad (4.22)
\end{aligned}$$

But by replacing $N-i$ with j and denoting $\widetilde{N} = N-k^*+1$ we get

$$\frac{1}{\sqrt{N}}\left(W_D(m,N)-\min_{0\leq i\leq N}W_D(m,i)\right)$$

$$
\begin{aligned}
&= \max_{0\le i\le N} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(i)) + \Delta_m \left(\widetilde{N} - (i-k^*+1)I_{\{i\ge k^*\}}\right)\right)\\
&= \max_{0\le j\le N} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(N-j)) + \Delta_m \left(\widetilde{N} - (\widetilde{N}-j)I_{\{N-k^*\ge j\}}\right)\right)\\
&= \max\Bigg\{ \max_{0\le j\le N-k^*} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(N-j)) + \Delta_m j\right),\\
&\qquad \max_{N-k^*<j\le N} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(N-j)) + \Delta_m \widetilde{N}\right)\Bigg\}.
\end{aligned}
$$

Because of the time reversibility of the Wiener process (cf. Borodin and Salminen (2002)) we can find a sequence of Wiener processes $\{W_{1,m}(t), t\ge 0\}_{m=1,2,\ldots}$ such that

$$
\begin{aligned}
&\max\Bigg\{ \max_{0\le j\le N-k^*} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(N-j)) + \Delta_m j\right),\\
&\qquad \max_{N-k^*<j\le N} \frac{1}{\sqrt{N}} \left(\sigma(W_m(N)-W_m(N-j)) + \Delta_m \widetilde{N}\right)\Bigg\}\\
&\overset{\mathcal{D}}{=} \max\Bigg\{ \max_{0\le j\le N-k^*} \frac{1}{\sqrt{N}} \left(\sigma W_{1,m}(j) + \Delta_m j\right),\\
&\qquad \max_{N-k^*<j\le N} \frac{1}{\sqrt{N}} \left(\sigma W_{1,m}(j) + \Delta_m \widetilde{N}\right)\Bigg\}.
\end{aligned}
$$

Applying Lemma 4.7 b) and again letting $\delta \downarrow 0$ we see that the first term in the outer maximum is taking its maximum arbitrarily close to $N-k^*$ and hence can be omitted and we can proceed with (4.22) to have

$$
\begin{aligned}
&\lim_{m\to\infty} P\left(\frac{1}{\sqrt{N}}\left(W_D(m,N) - \min_{0\le i\le N} W_D(m,i)\right) \le c\left(\frac{N}{m}\right)^{\gamma-1/2}\right)\\
&= \lim_{m\to\infty} P\left(\max_{N-k^*\le j\le N} \frac{1}{\sqrt{N}}\left(\sigma W_{1,m}(j) + \Delta_m\widetilde{N}\right) \le c\left(\frac{N}{m}\right)^{\gamma-1/2}\right)\\
&= A_8
\end{aligned}
$$

Due to the scaling property of the Wiener process again we can find a sequence $\{W_{2,m}(t), t\ge 0\}_{m=1,2,\ldots}$ of Wiener processes such that

$$
A_8 = \lim_{m\to\infty} P\left(\max_{1-\frac{k^*}{N}\le \frac{j}{N}\le 1} W_{2,m}(j/N) \le \frac{1}{\sigma}\left(\frac{N}{m}\right)^{\gamma-1/2}\left(c - \frac{\Delta_m(N-k^*)}{m^{1/2-\gamma}N^{\gamma}}\right)\right)
$$

$$=\overline{\Psi}(x),$$

where the last equation follows from Lemma 4.3 a) (iv), c) and Slutsky's Lemma. □

Proof of Theorem 2.2:

The major part of the proof is just a combination of the preceding lemmas. With u_γ from (4.17) the rest follows along the lines of the proof of Theorem 1.1 of Aue and Horváth (2004):

$$\begin{aligned}
\Psi(x) &= 1-\overline{\Psi}(-x) \\
&= 1-\lim_{m\to\infty} P\left(\tau_m^{\text{Page}} > N(m,-x)\right) \\
&= \lim_{m\to\infty} P\left(\tau_m^{\text{Page}} \le N(m,-x)\right) \\
&= \lim_{m\to\infty} P\left(\left(\tau_m^{\text{Page}}\right)^{1-\gamma} \le (N(m,-x))^{1-\gamma}\right) \\
&= \lim_{m\to\infty} P\left(\left(\tau_m^{\text{Page}}\right)^{1-\gamma} - a_m^{1-\gamma} \le \sigma x \frac{a_m^{1/2-\gamma}}{\Delta_m} u_\gamma(k^*, a_m)\right) \\
&= \lim_{m\to\infty} P\left(\left(\tau_m^{\text{Page}}\right)^{1-\gamma} - a_m^{1-\gamma} \le \sigma x \frac{a_m^{1/2-\gamma}}{\Delta_m} u_\gamma(k^*, a_m)\right) \\
&= \lim_{m\to\infty} P\left(\frac{a_m^\gamma}{1-\gamma} \frac{\left(\tau_m^{\text{Page}}\right)^{1-\gamma} - a_m^{1-\gamma}}{b_m} \le \sigma x \frac{\sqrt{a_m}}{\Delta_m\left(1-\gamma\left(1-\frac{k^*}{a_m}\right)\right)} b_m^{-1}\right) \\
&= \lim_{m\to\infty} P\left(\frac{\tau_m^{\text{Page}} - a_m}{b_m} \le x\right),
\end{aligned}$$

where the last equation follows because with the same arguments as in Aue and Horváth (2004) it can be shown that

$$\frac{\tau_m^{\text{Page}} - a_m}{b_m} \quad \text{and} \quad \frac{a_m^\gamma}{1-\gamma} \frac{\left(\tau_m^{\text{Page}}\right)^{1-\gamma} - a_m^{1-\gamma}}{b_m}$$

have the same limit distribution. □

Acknowledgements. I thank Alexander Aue, Lajos Horváth, and Josef G. Steinebach for the fruitful discussions throughout the work on this paper.

References

A. Aue. *Sequential Change-Point Analysis based on Invariance Principles*. PhD thesis, University of Cologne, 2003.

A. Aue, I. Berkes, and L. Horváth. Strong approximation for the sums of squares of augmented GARCH sequences. *Bernoulli*, 12(4):583, 2006a.

A. Aue and L. Horváth. Delay time in sequential detection of change. *Statist. Probab. Lett.*, 67(3):221–231, 2004.

A. Aue, L. Horváth, M. Hušková, and P. Kokoszka. Change-point monitoring in linear models. *Econom. J.*, 9(3):373–403, 2006b.

A. Aue, L. Horváth, and M.L. Reimherr. Delay times of sequential procedures for multiple time series regression models. *J. Econometrics*, 149(2):174 – 190, 2009.

M. Basseville and I. Nikiforov. *Detection of abrupt Changes: Theory and Application*. 1993.

A.N. Borodin and P. Salminen. *Handbook of Brownian Motion: Facts and Formulae*. Birkhäuser, 2002.

A. Černíková, M. Hušková, Z. Prášková, and J. Steinebach. Delay time in monitoring jump changes in linear models. *Statistics*, 2011. doi:10.1017/S0266466611000673.

C.S.J. Chu, M. Stinchcombe, and H. White. Monitoring structural change. *Econometrica*, 64(5):1045–1065, 1996.

J.C. Duan. Augmented GARCH (p,q) process and its diffusion limit. *J. Econometrics*, 79 (1):97–127, 1997.

S. Fremdt. Page's sequential procedure for change-point detection in time series regression. *Preprint, University of Cologne*, 2012.

L. Horváth, M. Hušková, P. Kokoszka, and J. Steinebach. Monitoring changes in linear models. *J. Statist. Plann. Inference*, 126(1):225–251, 2004.

L. Horvath, P. Kokoszka, and J. Steinebach. On sequential detection of parameter changes in linear regression. *Statist. Probab. Lett.*, 77(9):885–895, 2007.

M. Hušková and A. Koubková. Monitoring jump changes in linear models. *J. Statist. Res.*, 39(2):51–70, 2005.

G. Lorden. Procedures for reacting to a change in distribution. *Ann. Math. Stat.*, 42(6): 1897–1908, 1971.

E. S. Page. Continuous inspection schemes. *Biometrika*, 41(1/2):100–115, 1954.

TESTING THE EQUALITY OF COVARIANCE OPERATORS IN FUNCTIONAL SAMPLES

By Stefan Fremdt[‡], Lajos Horváth[†], Piotr Kokoszka[§] and Josef G. Steinebach[‡]

[‡]*University of Cologne,* [†]*University of Utah and* [§]*Colorado State University*

Abstract

We propose a nonparametric test for the equality of the covariance structures in two functional samples. The test statistic has a chi-square asymptotic distribution with a known number of degrees of freedom, which depends on the level of dimension reduction needed to represent the data. Detailed analysis of the asymptotic properties is developed. Finite sample performance is examined by a simulation study and an application to egg–laying curves of fruit flies.

Keywords: Asymptotic distribution, Covariance operator, Functional data, Quadratic forms, Two sample problem.

AMS subject classification: Primary 62G10; secondary 62G20, 62H15

1 Introduction

The last decade has seen increasing interest in methods of functional data analysis which offer novel and effective tools for dealing with problems where curves can naturally be viewed as data objects. The books by Ramsay and Silverman (2005) and Ramsay *et al.* (2009) offer comprehensive introductions to the subject, the collection Ferraty and Romain (2011) reviews some recent developments focusing on advances in the relevant theory, while the monographs of Bosq (2000), Ferraty and Vieu (2006) and Horváth and Kokoszka (2012) develop the field in several important directions. Despite the emergence of many alternative ways of looking at functional data, and many dimension reduction approaches, the functional principal components (FPC's) still remain the most important starting point for many functional data analysis procedures, Reiss and Ogden (2007), Gervini (2008), Yao and Müller (2010), Gabrys *et al.* (2010) are just a handful of illustrative references. The

FPC's are the eigenfunctions of the covariance operator. This paper focuses on testing if the covariance operators of two functional samples are equal. By the Karhunen-Loève expansion, this is equivalent to testing if both samples have the same set of FPC's. Benko *et al.* (2009) developed bootstrap procedures for testing the equality of specific FPC's. Panaretos *et al.* (2010) proposed a test of the type we consider, but assuming that the curves have a Gaussian distribution. The main result of Panaretos *et al.* (2010) follows as a corollary of our more general approach (Theorem 3.2). A generalization to non-Gaussian data was discussed in Panaretos *et al.* (2010) and Panaretos *et al.* (2011). For some recent work confer also Boente *et al.* (2011) who studied a related approach together with a corresponding bootstrap procedure.

Despite their importance, two sample problems for functional data received relatively little attention. In addition to the work of Benko *et al.* (2009) and Panaretos *et al.* (2010), the relevant references are Horváth *et al.* (2009) and Horváth *et al.* (2012) who focus, respectively, on the regression kernels in functional linear models and the mean of functional data exhibiting temporal dependence. For a recent contribution see also Gaines *et al.* (2011), who use a likelihood ratio-type approach for testing the equality of two covariance operators. Clearly, if some population parameters of two functional samples are different, estimating them using the pooled sample may lead to spurious conclusions. Due to the importance of the FPC's, a relatively simple and nonparametric procedure for testing the equality of the covariance operators is called for.

The remainder of this paper is organized as follows. Section 2 sets out the notation and definitions. The construction of the test statistic and its asymptotic properties are developed in Section 3. Section 4 reports the results of a simulation study and illustrates the procedure by application to egg-laying curves of Mediterranean fruit flies. The proofs of the asymptotic results of Section 3 are given in Section 5.

2 Preliminaries

Let $X_1, X_2, \ldots, X_N$ be independent, identically distributed random variables with values in $L_2[0,1]$, the Hilbert space of square-integrable $\mathbb{R}$-valued functions on $[0,1]$, and set $EX_i(t) = \mu(t)$ and $cov(X_i(t), X_i(s)) = C(t,s)$. We assume that another sample $X_1^*, X_2^*, \ldots X_M^*$ is also available and let $\mu^*(t) = EX_i^*(t)$ and $C^*(t,s) = cov(X_i^*(t), X_i^*(s))$

for $t, s \in [0,1]$. We wish to test the null hypothesis

$$H_0: \; C = C^*$$

against the alternative H_A that H_0 does not hold.
A crucial assumption considering the asymptotics of our test procedure will be that

$$\Theta_{N,M} = \frac{N}{M+N} \to \Theta \in (0,1) \quad \text{as} \quad N, M \to \infty. \tag{1}$$

For the construction of our test procedure we will use an estimate of the asymptotic pooled covariance operator $\mathfrak{R}$ of the two given samples (cf. (4)) which is defined by the kernel

$$R(t,s) = \Theta C(t,s) + (1-\Theta) C^*(t,s).$$

In the case of samples $\{X_i\}$ and $\{X_j^*\}$ of *Gaussian* random functions, the latter approach has successfully been applied by Panaretos *et al.* (2010) to construct an asymptotic test for checking the equality of two covariance operators (see also Panaretos *et al.* (2011)).
Denote by $(\lambda_1, \varphi_1), (\lambda_2, \varphi_2), \ldots$ the eigenvalue/eigenfunction pairs of $\mathfrak{R}$, which are defined by

$$\lambda_k \varphi_k(t) = \mathfrak{R} \varphi_k(t) = \int_0^1 R(t,s) \varphi_k(s) ds, \quad t \in [0,1], \quad 1 \le k < \infty \; . \tag{2}$$

Throughout this paper we assume

$$\lambda_1 > \lambda_2 > \ldots > \lambda_p > \lambda_{p+1}, \tag{3}$$

i.e. there exist at least p distinct (positive) eigenvalues. Under assumption (3), we can uniquely (up to signs) choose $\varphi_1, \ldots, \varphi_p$ satisfying (2), if we require $\|\varphi_i\| = 1$, where $\|\cdot\|$ always denotes the L_2-norm, e.g., for $x \in L_2([0,1])$,

$$\|x\| = \left(\int_0^1 x^2(t) \, dt \right)^{1/2} .$$

Thus, under (3), $\{\varphi_i, 1 \le i \le p\}$ is an orthonormal system that can be extended to an orthonormal basis $\{\varphi_i, 1 \le i < \infty\}$.
If H_0 holds, then (λ_i, φ_i), $1 \le i < \infty$, are also the eigenvalues/eigenfunctions of the covariance operators $\mathfrak{C}$ of the first and $\mathfrak{C}^*$ of the second sample. To construct a test statistic which converges under H_0, we can therefore pool the two samples, as explained in Section 3.

3 The test and the asymptotic results

Along the lines of Panaretos *et al.* (2010), our procedure is also based on projecting the observations onto a suitably chosen finite-dimensional space. To define this space, introduce the empirical pooled covariance operator $\widehat{\mathfrak{R}}_{N,M}$ defined by the kernel

$$\widehat{R}_{N,M}(t,s) = \frac{1}{N+M}\left\{ \sum_{k=1}^{N}(X_k(t)-\overline{X}_N(t))(X_k(s)-\overline{X}_N(s)) + \sum_{k=1}^{M}(X_k^*(t)-\overline{X}_M^*(t))(X_k^*(s)-\overline{X}_M^*(s)) \right\}, \tag{4}$$

where

$$\overline{X}_N(t) = \frac{1}{N}\sum_{k=1}^{N} X_k(t) \quad \text{and} \quad \overline{X}_M^*(t) = \frac{1}{M}\sum_{k=1}^{M} X_k^*(t)$$

are the sample mean functions. Let $(\widehat{\lambda}_i, \widehat{\varphi}_i)$ denote the eigenvalues/eigenfunctions of $\widehat{\mathfrak{R}}_{N,M}$, i.e.

$$\widehat{\lambda}_i\widehat{\varphi}_i(t) = \widehat{\mathfrak{R}}_{N,M}\widehat{\varphi}_i(t) = \int_0^1 \widehat{R}_{N,M}(t,s)\widehat{\varphi}_i(s)ds, \quad t\in[0,1], \quad 1\le i\le N+M,$$

with $\widehat{\lambda}_1 \ge \widehat{\lambda}_2 \ge \dots$. We can and will assume that the $\widehat{\varphi}_i$ form an orthonormal system. We consider the projections

$$\widehat{a}_k(i) = \langle X_k - \overline{X}_N, \widehat{\varphi}_i\rangle = \int_0^1 (X_k(t)-\overline{X}_N(t))\widehat{\varphi}_i(t)dt \tag{5}$$

and

$$\widehat{a}_k^*(j) = \langle X_k^* - \overline{X}_M^*, \widehat{\varphi}_j\rangle = \int_0^1 \left(X_k^*(t)-\overline{X}_M^*(t)\right)\widehat{\varphi}_j(t)dt, \tag{6}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product of two elements of the Hilbert space $L_2[0,1]$. To test H_0, we compare the matrices $\widehat{\Delta}_N$ and $\widehat{\Delta}_M^*$ with entries

$$\widehat{\Delta}_N(i,j) = \frac{1}{N}\sum_{k=1}^{N}\widehat{a}_k(i)\widehat{a}_k(j), \quad 1\le i,j\le p,$$

and

$$\widehat{\Delta}_M^*(i,j) = \frac{1}{M}\sum_{k=1}^{M}\widehat{a}_k^*(i)\widehat{a}_k^*(j), \quad 1\le i,j\le p.$$

We note that $\widehat{\Delta}_N(i,j) - \widehat{\Delta}^*_M(i,j)$ is the projection of $\widehat{C}_N(t,s) - \widehat{C}^*_M(t,s)$ in the direction of $\widehat{\varphi}_i(t)\widehat{\varphi}_j(s)$, where

$$\widehat{C}_N(t,s) = \frac{1}{N}\sum_{k=1}^{N}\left(X_k(t) - \overline{X}_N(t)\right)\left(X_k(s) - \overline{X}_N(s)\right)$$

and

$$\widehat{C}^*_M(t,s) = \frac{1}{M}\sum_{k=1}^{M}\left(X^*_k(t) - \overline{X}^*_M(t)\right)\left(X^*_k(s) - \overline{X}^*_M(s)\right)$$

are the empirical covariances of the two samples.
We create the vector $\widehat{\boldsymbol{\xi}}_{N,M}$ from the columns below the diagonal of $\widehat{\Delta}_N - \widehat{\Delta}^*_M$ as follows:

$$\widehat{\boldsymbol{\xi}}_{N,M} = \operatorname{vech}\left(\widehat{\Delta}_N - \widehat{\Delta}^*_M\right) = \begin{pmatrix} \widehat{\Delta}_N(1,1) - \widehat{\Delta}^*_M(1,1) \\ \widehat{\Delta}_N(2,1) - \widehat{\Delta}^*_M(2,1) \\ \vdots \\ \widehat{\Delta}_N(p,p) - \widehat{\Delta}^*_M(p,p) \end{pmatrix}. \tag{7}$$

For the properties of the vech operator we refer to Abadir and Magnus (2005).
Next we estimate the asymptotic covariance matrix of $(MN/(N+M))^{1/2}\widehat{\boldsymbol{\xi}}_{N,M}$. Note that, in general, this estimate differs from the one which was used in the Gaussian case (cf. Panaretos *et al.* (2010) and Theorem 3.2 below). Let

$$\begin{aligned}\widehat{L}_{N,M}(k,k') = (1-\Theta_{N,M})&\left\{\frac{1}{N}\sum_{\ell=1}^{N}\widehat{a}_\ell(i)\widehat{a}_\ell(j)\widehat{a}_\ell(i')\widehat{a}_\ell(j') - \left\langle\widehat{\mathfrak{C}}_N\widehat{\varphi}_i,\widehat{\varphi}_j\right\rangle\left\langle\widehat{\mathfrak{C}}_N\widehat{\varphi}_{i'},\widehat{\varphi}_{j'}\right\rangle\right\} \\ +\Theta_{N,M}&\left\{\frac{1}{M}\sum_{\ell=1}^{M}\widehat{a}^*_\ell(i)\widehat{a}^*_\ell(j)\widehat{a}^*_\ell(i')\widehat{a}^*_\ell(j') - \left\langle\widehat{\mathfrak{C}}^*_M\widehat{\varphi}_i,\widehat{\varphi}_j\right\rangle\left\langle\widehat{\mathfrak{C}}^*_M\widehat{\varphi}_{i'},\widehat{\varphi}_{j'}\right\rangle\right\},\end{aligned}$$

where i,j,i',j' depend on k,k' (see below), and $\widehat{\mathfrak{C}}_N$ ($\widehat{\mathfrak{C}}^*_M$) is interpreted as an operator with $\widehat{\mathfrak{C}}_N$ defined as

$$\widehat{\mathfrak{C}}_N\widehat{\varphi}_i = \int_0^1 \widehat{C}_N(t,s)\widehat{\varphi}_i(s)ds.$$

(An analogous definition holds for $\widehat{\mathfrak{C}}^*_M$.) From this definition it follows that

$$\left\langle\widehat{\mathfrak{C}}_N\widehat{\varphi}_i,\widehat{\varphi}_j\right\rangle = \frac{1}{N}\sum_{l=1}^{N}\widehat{a}_\ell(i)\widehat{a}_\ell(j).$$

There are other ways to estimate the asymptotic covariance matrix. We note that one can use $\widehat{L}^*_{N,M}(k,k')$ instead of $\widehat{L}_{N,M}(k,k')$, where $\widehat{L}^*_{N,M}(k,k')$ is defined like $\widehat{L}_{N,M}(k,k')$, but

$\langle \widehat{\mathfrak{C}}_N \widehat{\varphi}_i, \widehat{\varphi}_j \rangle$ and $\langle \widehat{\mathfrak{C}}^*_M \widehat{\varphi}_i, \widehat{\varphi}_j \rangle$ are replaced with 0 if $i \neq j$ and $\widehat{\lambda}_i$ if $i = j$. In the same spirit, $\langle \widehat{\mathfrak{C}}_N \widehat{\varphi}_{i'}, \widehat{\varphi}_{j'} \rangle$ and $\langle \widehat{\mathfrak{C}}^*_M \widehat{\varphi}_{i'}, \widehat{\varphi}_{j'} \rangle$ are replaced with 0 for $i' \neq j'$ and $\widehat{\lambda}_{i'}$ if $i' = j'$.
The index (i, j) is computed from k in the following way: Let

$$k' = \frac{p(p+1)}{2} - k + 1, \quad i' = p - i + 1, \quad \text{and} \quad j' = p - j + 1. \tag{8}$$

We look at an upper triangle matrix $(a_{i',j'})$. Then, for column j', we have that $(j'-1)j'/2 < k \leq j'(j'+1)/2$. Thus $j' = \left\lceil \sqrt{2k' + \frac{1}{4}} - \frac{1}{2} \right\rceil$ and $i' = k' - (j'-1)j'/2$, where $\lceil r \rceil = \min\{k \in \mathbb{Z} : k \geq r\}$ for $r \in \mathbb{R}$. Consequently, the index (i, j) can be computed from k via

$$j = p - \left\lceil \sqrt{p(p+1) - 2k + \frac{9}{4}} - \frac{1}{2} \right\rceil + 1 \quad \text{and} \quad i = k + p - p \cdot j + \frac{j(j-1)}{2}. \tag{9}$$

With the above notation, we can formulate the main result of this paper in the non-Gaussian case. The latter case has briefly been mentioned (without any mathematical details) in the concluding remarks of Panaretos *et al.* (2010) (see also Panaretos *et al.* (2011)).

THEOREM 3.1. *We assume that H_0, (1) and (3) hold, and*

$$\int_0^1 E(X_1(t))^4 dt < \infty, \quad \int_0^1 E(X_1^*(t))^4 dt < \infty. \tag{10}$$

Then

$$\frac{NM}{N+M} \widehat{\boldsymbol{\xi}}^T_{N,M} \widehat{L}^{-1}_{N,M} \widehat{\boldsymbol{\xi}}_{N,M} \xrightarrow{\mathcal{D}} \chi^2_{p(p+1)/2}, \quad \textit{as} \quad N, M \to \infty,$$

where $\chi^2_{p(p+1)/2}$ stands for a χ^2 random variable with $p(p+1)/2$ degrees of freedom.

Theorem 3.1 implies that the null hypothesis is rejected if the test statistic

$$\widehat{T}_1 = \frac{NM}{N+M} \widehat{\boldsymbol{\xi}}^T_{N,M} \widehat{L}^{-1}_{N,M} \widehat{\boldsymbol{\xi}}_{N,M}$$

exceeds a critical quantile of the chi–square distribution with $p(p+1)/2$ degrees of freedom. If both samples are Gaussian random processes, the quadratic form $\widehat{\boldsymbol{\xi}}^T_{N,M} \widehat{L}^{-1}_{N,M} \widehat{\boldsymbol{\xi}}_{N,M}$ can be replaced with the normalized sum of the squares of $\widehat{\Delta}_{N,M}(i,j) - \widehat{\Delta}^*_{N,M}(i,j)$, as stated in the following theorem (cf. Panaretos *et al.* (2010)).

Theorem 3.2. *If X_1, X_1^* are Gaussian processes and the conditions of Theorem 3.1 are satisfied, then, as $N, M \to \infty$,*

$$\widehat{T}_2 = \frac{NM}{N+M} \sum_{1 \le i,j \le p} \frac{1}{2} \frac{\left(\widehat{\Delta}_{N,M}(i,j) - \widehat{\Delta}^*_{N,M}(i,j)\right)^2}{\widehat{\lambda}_i \widehat{\lambda}_j} \xrightarrow{\mathcal{D}} \chi^2_{p(p+1)/2}.$$

Observe that the statistic $\widehat{T}_2$ can be written as

$$\widehat{T}_2 = \frac{NM}{N+M} \left\{ \sum_{1 \le i < j \le p} \frac{\left(\widehat{\Delta}_{N,M}(i,j) - \widehat{\Delta}^*_{N,M}(i,j)\right)^2}{\widehat{\lambda}_i \widehat{\lambda}_j} + \sum_{i=1}^{p} \frac{\left(\widehat{\Delta}_{N,M}(i,i) - \widehat{\Delta}^*_{N,M}(i,i)\right)^2}{2\widehat{\lambda}_i^2} \right\}.$$

Next we discuss the asymptotic consistency of the testing procedure based on Theorem 3.1. Analogously to the definition of $\widehat{\boldsymbol{\xi}}_{N,M}$ we define the vector $\boldsymbol{\xi} = (\xi(1), \ldots, \xi(p(p+1)/2))$ using the columns of the matrix

$$\boldsymbol{D} = \left(\int_0^1 \int_0^1 (C(t,s) - C^*(t,s)) \varphi_i(t) \varphi_j(s) dt\, ds \right)_{i,j=1,\ldots,p} \tag{11}$$

instead of $\widehat{\Delta}_N - \widehat{\Delta}^*_M$, i.e.

$$\boldsymbol{\xi} = \text{vech}(\boldsymbol{D}).$$

Theorem 3.3. *We assume that H_A, (1), (3) and (10) hold. Then there exist random variables $\widehat{h}_1 = \widehat{h}_1(N,M), \ldots, \widehat{h}_{p(p+1)/2} = \widehat{h}_{p(p+1)/2}(N,M)$, taking values in $\{-1, 1\}$ such that, as $N, M \to \infty$,*

$$\max_{1 \le i \le p(p+1)/2} \left| \widehat{\xi}_{N,M}(i) - \widehat{h}_i \xi(i) \right| = o_P(1) \tag{12}$$

and therefore

$$\left| \widehat{\boldsymbol{\xi}}_{N,M} \right| \xrightarrow{P} |\boldsymbol{\xi}|, \tag{13}$$

where $|\cdot|$ denotes the Euclidean norm. If $\boldsymbol{\xi} \neq \boldsymbol{0}$ and the p largest eigenvalues of C and C^* are positive, we also have

$$\widehat{T}_1 \xrightarrow{P} \infty, \quad \text{as} \quad N, M \to \infty. \tag{14}$$

The assumption that the p largest eigenvalues of C and C^* are positive implies that the random functions $X_i,\ i = 1, \ldots, N$, and $X_j^*,\ j = 1, \ldots, M$, are not included in a $(p-1)$-dimensional subspace.

REMARK 3.4. *The application of the test requires the selection of the number p of the empirical FPC's to be used. A rule of thumb is to choose p so that the first p empirical FPC's in each sample (i.e. those calculated as the eigenfunctions of $\widehat{C}_N$ and $\widehat{C}^*_M$) explain about 85–90% of the variance in each sample. Choosing p too large generally negatively affects the finite sample performance of tests of this type, and for this reason we do not study asymptotics as p tends to infinity. It is often illustrative to apply the test for a range of the values of p; each p specifies a level of relevance of differences in the curves or kernels. A good practical approach is to look at the Karhunen–Loève approximations of the curves in both samples, and choose p which gives approximation errors that can be considered unimportant. Cross validation has also been suggested in the literature without investigating its properties in detail. For a more formal discussion of this selection, confer also Section 3.3 in Panaretos et al. (2010).*

4 A SIMULATION STUDY AND AN APPLICATION

We first describe the results of a simulations study designed to evaluate finite sample properties of the tests based on the statistics $\widehat{T}_1$ and $\widehat{T}_2$. The emphasis is on verifying the advantage of a nonparametric procedure, i.e., to see the "robustness" to the violation of the assumption of normality. We simulated Gaussian curves as Brownian motions and Brownian bridges, and non-Gaussian curves via

$$X(t) = A\sin(\pi t) + B\sin(2\pi t) + C\sin(4\pi t), \tag{15}$$

where $A = 5Y_1,\ B = 3Y_2,\ C = Y_3$, and Y_1, Y_2, Y_3 are independent t_5-distributed random variables (similarly $X^*(t)$ for the second sample). All curves were simulated at 1000 equidistant points in the interval $[0,1]$, and transformed into functional data objects using the Fourier basis with 49 basis functions. For each data generating process we used one thousand replications.

Table 1 displays the empirical sizes for non-Gaussian data. The test based on $\widehat{T}_2$ has severely inflated size, due to the violation of the assumption of normality. As documented in Panaretos *et al.* (2010), and confirmed by our own simulations, this test has very good

Table 1: Empirical sizes of the tests based on statistics $\widehat{T}_1$ and $\widehat{T}_2$ for non-Gaussian data. The curves in each sample were generated according to (15).

	$p = 2$					
	$\widehat{T}_1$			$\widehat{T}_2$		
Sample Sizes	1%	5%	10%	1%	5%	10%
$N = M = 100$	0.005	0.028	0.061	0.152	0.275	0.380
$N = M = 200$	0.003	0.021	0.058	0.163	0.314	0.402
$N = M = 1000$	0.002	0.021	0.056	0.190	0.313	0.426
	$p = 3$					
	$\widehat{T}_1$			$\widehat{T}_2$		
Sample Sizes	1%	5%	10%	1%	5%	10%
$N = M = 100$	0.004	0.028	0.065	0.167	0.332	0.434
$N = M = 200$	0.004	0.024	0.064	0.194	0.338	0.423
$N = M = 1000$	0.004	0.028	0.070	0.240	0.384	0.484

Table 2: Power of the test based on statistic $\widehat{T}_1$ for non-Gaussian data. The curves in the equally sized samples were generated according to (15) in the first sample and as a scaled version of (15) in the second sample, i.e. $X^*(t) = cX(t)$.

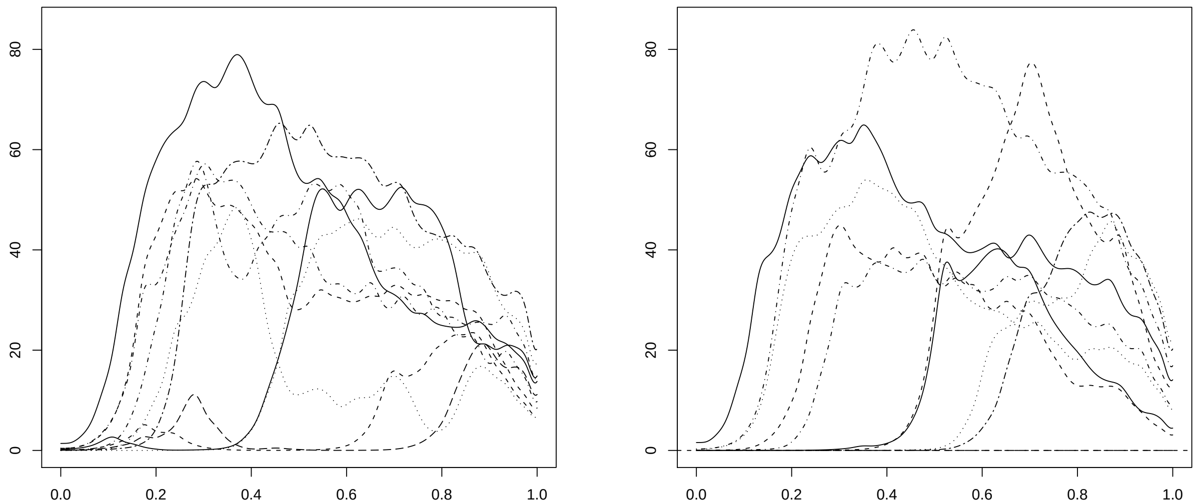
	$c = 0.8$						$c = 0.9$					
	$p = 2$			$p = 3$			$p = 2$			$p = 3$		
N,M	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
200	0.061	0.289	0.491	0.033	0.181	0.351	0.007	0.041	0.106	0.006	0.037	0.089
500	0.532	0.811	0.917	0.485	0.804	0.896	0.032	0.159	0.274	0.017	0.116	0.231
1000	0.947	0.986	0.993	0.965	0.997	0.998	0.113	0.360	0.519	0.099	0.327	0.490

	$c = 1.2$						$c = 1.1$					
	$p = 2$			$p = 3$			$p = 2$			$p = 3$		
N,M	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
200	0.017	0.160	0.295	0.015	0.124	0.231	0.001	0.041	0.099	0.004	0.034	0.083
500	0.244	0.570	0.718	0.186	0.524	0.689	0.009	0.096	0.222	0.010	0.079	0.165
1000	0.762	0.913	0.952	0.760	0.956	0.980	0.091	0.295	0.461	0.056	0.226	0.396

empirical size when the data are Gaussian. The test based on $\widehat{T}_1$ is conservative, especially for smaller sample sizes. This is true for both Gaussian and non-Gaussian data; there is not much difference in the empirical size of this test for different data generating processes.

Table 2 gives an example of the empirical power of the test based on statistic $\widehat{T}_1$. The test was carried out for two equally sized samples of 200, 500 and 1000 realizations, respectively, of (15) for the first sample and scaled versions of (15), i.e. $X^*(t) = cX(t)$, for the second sample. The results are displayed for a selection of values for the scaling parameter c. It can be seen that in all cases the power increases with the sample size. As can be expected the convergence of the power towards 1 improves for larger deviations ($c \neq 1$) from the null hypothesis. Since, due to the inflated size of the test based on $\widehat{T}_2$ in the non-Gaussian case (cf. Table 1), its power is (misleadingly) higher than that of the test based on $\widehat{T}_1$, and

Figure 1: Ten randomly selected smoothed egg-laying curves of short-lived medflies (left panel), and ten such curves for long–lived medflies (right panel).



thus will not be displayed here. We also studied a Monte Carlo version of the test based on the statistic $\widehat{T}_3 = NM(N+M)^{-1}\widehat{\boldsymbol{\xi}}_{N,M}^T\widehat{\boldsymbol{\xi}}_{N,M}$, and found that its finite sample properties were similar to those of the test based on $\widehat{T}_1$.

We now describe the results of the application of both tests to an interesting data set consisting of egg–laying trajectories of Mediterranean fruit flies (medflies). The data were kindly made available to us by Hans–Georg Müller. This data set has been extensively studied in biological and statistical literature, see Müller and Stadtmüller (2005) and references therein. We consider 534 egg-laying curves of medflies who lived at least 34 days, but we only consider the egg–laying activities on the first 30 days. We examined two versions of these egg-laying curves. The curves are scaled such that the functions in either version are defined on the interval $[0, 1]$. Version 1 curves (denoted $X_i(t)$) are the absolute counts of eggs laid by fly i on day $\lfloor 30t \rfloor$. Version 2 curves (denoted $Y_i(t)$) are the counts of eggs laid by fly i on day $\lfloor 30t \rfloor$ *relative* to the total number of eggs laid in the lifetime of fly i. The 534 flies are classified into long-lived, i.e. those who lived 44 days or longer, and short-lived, i.e. those who died before the end of the 43rd day

Figure 2: Ten randomly selected smoothed egg-laying curves of short-lived medflies (left panel), and ten such curves for long–lived medflies (right panel), relative to the number of eggs laid in the fly's lifetime.

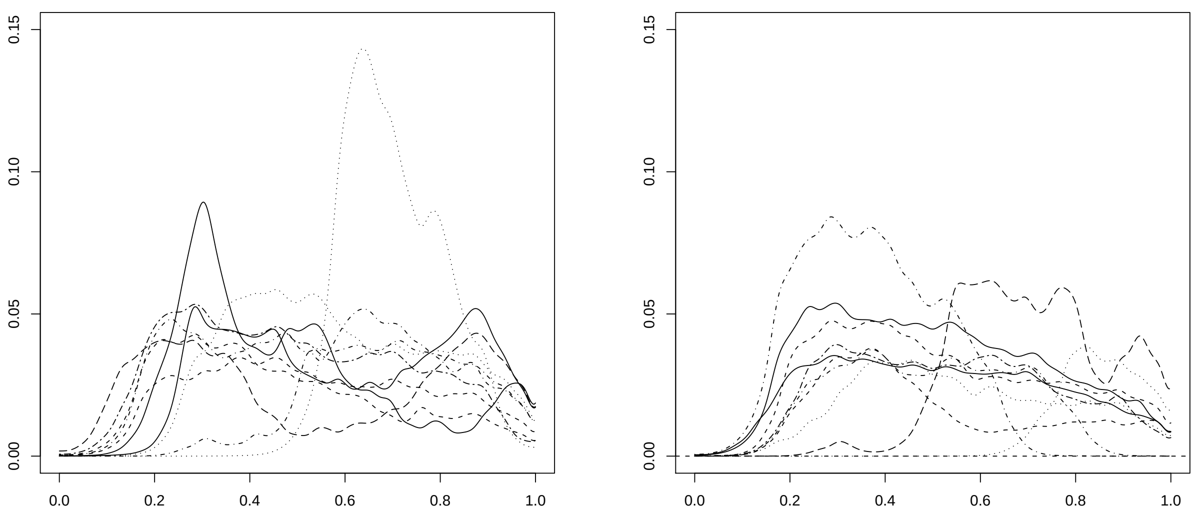


Table 3: P–values (in percent) of the test based on statistics $\widehat{T}_1$ and $\widehat{T}_2$ applied to absolute medfly data. Here f_p denotes the fraction of the sample variance explained by the first p FPCs, i.e. $f_p = (\sum_{k=1}^{p} \hat{\lambda}_k)/(\sum_{k=1}^{N+M} \hat{\lambda}_k)$.

P-values								
p	2	3	4	5	6	7	8	9
$\widehat{T}_1$	82.70	36.22	30.59	63.84	37.71	39.03	33.77	34.77
$\widehat{T}_2$	0.54	0.13	0.11	0.12	0.02	0.00	0.00	0.00
f_p	72.93	78.36	81.87	83.94	85.62	87.08	88.49	89.72

after birth. In the data set, there are 256 short-lived, and 278 long-lived flies. This classification naturally defines two samples: *Sample 1:* the egg-laying curves $\{X_i(t)$ (resp. $Y_i(t)),\ 0 \le t \le 1,\ i = 1, 2, \ldots, 256\}$ of the short-lived flies. *Sample 2:* the egg-laying curves $\{X_j^*(t)$ (resp. $Y_j^*(t)),\ 0 < t \le 30,\ j = 1, 2, \ldots, 278\}$ of the long-lived flies. The egg-laying curves are very irregular; Figure 1 shows ten (smoothed) curves of short- and long-lived flies for version 1, Figure 2 shows ten (smoothed) curves for version 2 (both using a B-spline basis for the representation).

Table 3 shows the P–values for the absolute egg-laying counts (version 1). For the statistic $\widehat{T}_1$ the null hypothesis cannot be rejected irrespective of the choice of p. For the statistic $\widehat{T}_2$, the result of the test varies depending on the choice of p. As explained in Section 3, the usual recommendation is to use the values of p which explain 85 to 90 percent of the variance For such values of p, $\widehat{T}_2$ leads to a clear rejection. Since this test has however overinflated size, we conclude that there is little evidence that the covariance structures of version 1 curves for long– and short–lived flies are different. For the version 2 curves, the statistic $\widehat{T}_2$ yields P–values equal to zero (in machine precision), potentially indicating that the covariance structures for the short– and long–lived flies are different. The assumption of a normal distribution is however questionable, as the QQ-plots in Figure 3 show. These QQ-plots are constructed for the inner products $\langle Y_i, e_k \rangle$ and $\langle Y_i^*, e_k \rangle$, where the Y_i are the curves from one of the samples (we cannot pool the data to construct QQ-plots because we test if the stochastic structures are different), and e_k is the kth element of the Fourier basis. The normality of a functional sample implies the normality of all projections onto a complete orthonormal system. For $\langle X_i, e_k \rangle$, the QQ-plots show a strong deviation from a

Figure 3: Normal QQ–plots for the scores of the version 2 medfly data with respect to the first two Fourier basis functions. Left – sample 1, Right – sample 2.

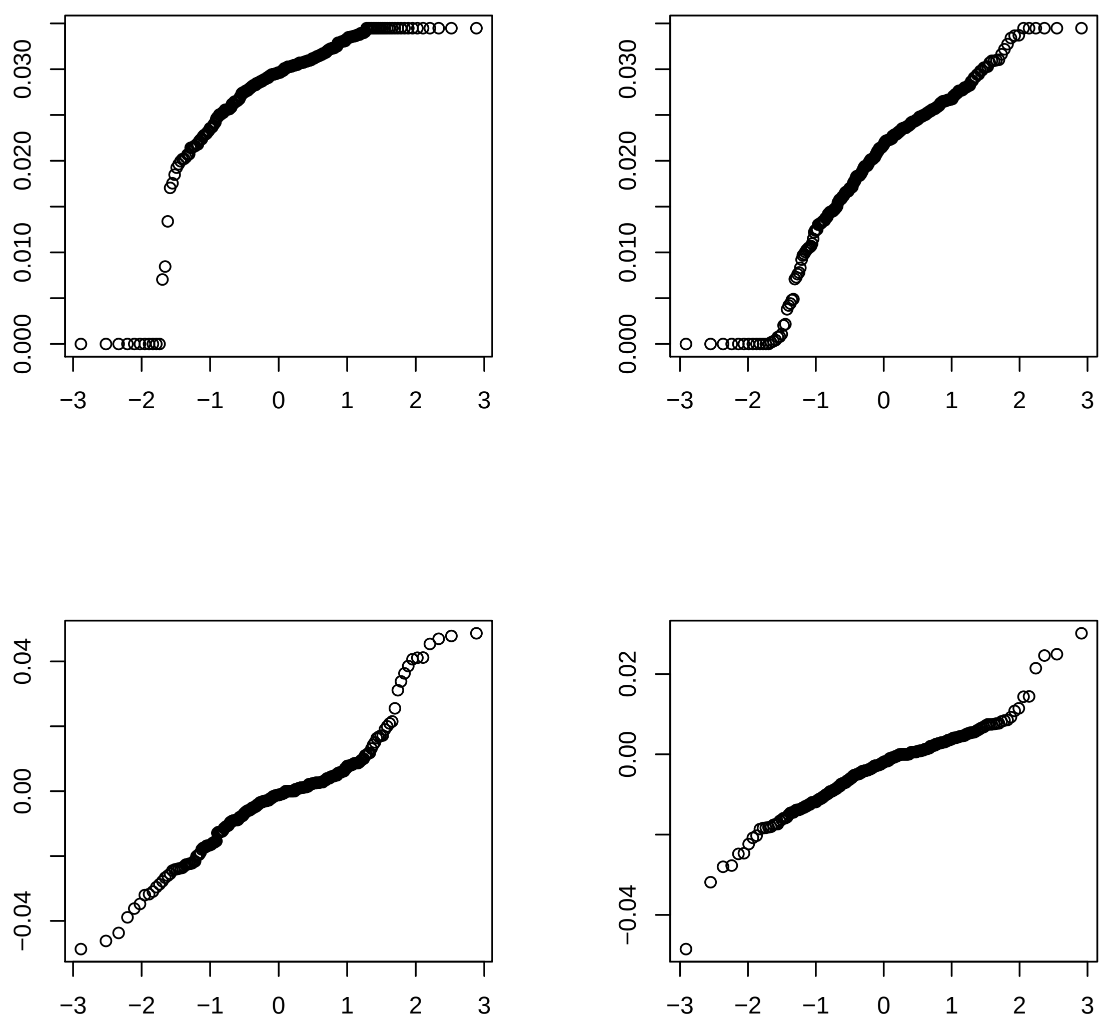


Table 4: P–values (in percent) of the test based on statistics $\widehat{T}_1$ applied to relative medfly data; f_p denotes the fraction of the sample variance explained by the first p FPCs, i.e. $f_p = (\sum_{k=1}^{p} \hat{\lambda}_k)/(\sum_{k=1}^{N+M} \hat{\lambda}_k)$.

P-values							
p	2	3	4	5	6	7	8
$\widehat{T}_1$	0.14	0.06	0.33	1.50	3.79	4.53	10.28
f_p	33.99	44.08	52.72	59.04	65.08	70.40	75.29
p	9	10	11	12	13	14	15
$\widehat{T}_1$	5.51	2.78	5.32	3.21	1.78	6.28	3.80
f_p	79.91	83.72	86.58	89.02	91.34	93.30	95.03

straight line for some projections. Almost all projections $\langle Y_i, e_k \rangle$ have QQ-plots indicating a strong deviation from normality. It is therefore important to apply the nonparametric test based on the statistic $\widehat{T}_1$. The corresponding P–values for version 2 are displayed in Table 4. For most values of p, these P–values indicate the rejection of H_0. Many of them hover around the 5 percent level, but since the test is conservative, we can with confidence view them as favoring H_A.

The above application confirms the properties of the statistics established through the simulation study. It shows that while there is little evidence that the covariance structures for the absolute counts are different, there is strong evidence that they are different for relative counts.

5 Proofs of the results of Section 3

The proof of Theorem 3.1 follows from several lemmas, which we establish first. We can and will assume without loss of generality that $\mu(t) = \mu^*(t) = 0$ for all $t \in [0, 1]$.

We will use the identity

$$\frac{1}{N^{1/2}} \sum_{k=1}^{N} \left(X_k(t) - \overline{X}_N(t)\right) \left(X_k(s) - \overline{X}_N(s)\right) \tag{16}$$

$$= \frac{1}{N^{1/2}} \sum_{k=1}^{N} X_k(t)X_k(s) - N^{1/2}\overline{X}_N(t)\overline{X}_N(s),$$

and an analogous identity for the second sample.
Our first lemma establishes bounds in probability which will often be used in the proofs.

LEMMA 5.1. *Under the assumptions of Theorem 3.1, as $N, M \to \infty$,*

$$\left\| N^{-1/2} \sum_{k=1}^{N} \{X_k(t)X_k(s) - C(t,s)\} \right\| = O_P(1), \tag{17}$$

$$\left\| N^{1/2}\overline{X}_N(t) \right\| = O_P(1), \tag{18}$$

and

$$\left\| M^{-1/2} \sum_{k=1}^{M} \{X_k^*(t)X_k^*(s) - C^*(t,s)\} \right\| = O_P(1), \tag{19}$$

$$\left\| M^{1/2}\overline{X}_M^*(t) \right\| = O_P(1), \tag{20}$$

where here and in the sequel the notation $\|\cdot\|$ is also used for the corresponding norm in $L_2([0,1]^2)$.

Proof: These are classical estimates and can easily be obtained by a straightforward calculation of the second moments. Note, for example, that

$$E \int_0^1 \int_0^1 \Big[\frac{1}{N^{1/2}} \sum_{k=1}^{N} \{X_k(t)X_k(s) - C(t,s)\} \Big]^2 dt\, ds = \int_0^1 \int_0^1 E\{X_1(t)X_1(s) - C(t,s)\}^2 \, dt\, ds,$$

so, by Markov's inequality, we have

$$\left\| \frac{1}{N^{1/2}} \sum_{k=1}^{N} \{X_k(t)X_k(s) - C(t,s)\} \right\|^2 = O_P(1).$$

Similar arguments yield (18) – (20). Confer also Dauxois *et al.* (1982) for an early reference.

□

The next lemma shows that the estimation of the mean functions, cf. the definition of the projections $\widehat{a}_k(i)$ and $\widehat{a}_k^*(j)$ in (5) and (6), has an asymptotically negligible effect.

LEMMA 5.2. *Under the assumptions of Theorem 3.1, for all $1 \le i, j \le p$, as $N, M \to \infty$,*

$$N^{1/2}\widehat{\Delta}_N(i,j) = \frac{1}{N^{1/2}} \sum_{k=1}^{N} \langle X_k, \widehat{\varphi}_i \rangle \langle X_k, \widehat{\varphi}_j \rangle + O_P\left(N^{-1/2}\right)$$

and

$$M^{1/2}\widehat{\Delta}^*_M(i,j) = \frac{1}{M^{1/2}} \sum_{k=1}^{M} \langle X^*_k, \widehat{\varphi}_i \rangle \langle X^*_k, \widehat{\varphi}_j \rangle + O_P\left(M^{-1/2}\right).$$

PROOF: Using (16) and (18) we have by the Cauchy-Schwarz inequality,

$$\begin{aligned}
&\left| \int_0^1 \int_0^1 N^{1/2} \overline{X}_N(t) \overline{X}_N(s) \widehat{\varphi}_i(t) \widehat{\varphi}_j(s) dt\, ds \right| \\
&= N^{-1/2} \left| \int_0^1 N^{1/2} \overline{X}_N(t) \widehat{\varphi}_i(t) dt \right| \left| \int_0^1 N^{1/2} \overline{X}_N(s) \widehat{\varphi}_j(s) ds \right| \\
&\le N^{-1/2} \left(\int_0^1 \left(N^{1/2}\overline{X}_N(t)\right)^2 dt \int_0^1 \widehat{\varphi}_i^2(t) dt \right)^{1/2} \left(\int_0^1 \left(N^{1/2}\overline{X}_N(s)\right)^2 ds \int_0^1 \widehat{\varphi}_j^2(s) ds \right)^{1/2} \\
&= N^{-1/2} \int_0^1 \left(N^{1/2}\overline{X}_N(t)\right)^2 dt \\
&= O_P\left(N^{-1/2}\right).
\end{aligned}$$

The second part can be proven in the same way. □

We now state bounds on the distances between the estimated and the population eigenvalues and eigenfunctions. These bounds are true under the null hypothesis, and extend the corresponding one sample bounds.

LEMMA 5.3. *If the conditions of Theorem 3.1 are satisfied, then, as $N, M \to \infty$,*

$$\max_{1 \le i \le p} |\widehat{\lambda}_i - \lambda_i| = O_P\left((N+M)^{-1/2}\right)$$

and

$$\max_{1 \le i \le p} \|\widehat{\varphi}_i - \widehat{c}_i \varphi_i\| = O_P\left((N+M)^{-1/2}\right),$$

where

$$\widehat{c}_i = \widehat{c}_i(N,M) = \operatorname{sign}(\langle \widehat{\varphi}_i, \varphi_i \rangle).$$

PROOF: These estimates are also well-known (cf., e.g., Bosq (2000), Lemma 4.3 and assertion (4.43), or Horváth and Kokoszka (2012), Lemmas 2.2 – 2.3). Note that the first rate above is independent of p, whereas the second one may actually depend on the projection dimension p. □

Lemma 5.3 now allows us to replace the estimated eigenfunctions by their population counterparts. The random signs $\widehat{c}_i$ must appear in the formulation of Lemma 5.4, but they cancel in the subsequent results.

LEMMA 5.4. *If the conditions of Theorem 3.1 are satisfied, then, for all* $1 \le i, j \le p$, *as* $N, M \to \infty$,

$$\left(\frac{NM}{N+M}\right)^{1/2}\left(\widehat{\Delta}_N(i,j) - \widehat{\Delta}^*_M(i,j)\right)$$
$$= \left(\frac{NM}{N+M}\right)^{1/2}\left\{\frac{1}{N}\sum_{k=1}^{N}\langle X_k, \widehat{c}_i\varphi_i\rangle\langle X_k, \widehat{c}_j\varphi_j\rangle - \frac{1}{M}\sum_{k=1}^{M}\langle X^*_k, \widehat{c}_i\varphi_i\rangle\langle X^*_k, \widehat{c}_j\varphi_j\rangle\right\} + o_P(1).$$

PROOF: We write

$$\frac{1}{N}\sum_{k=1}^{N}\langle X_k, \widehat{\varphi}_i\rangle\langle X_k, \widehat{\varphi}_j\rangle - \int_0^1\int_0^1 C(t,s)\widehat{\varphi}_i(t)\widehat{\varphi}_j(s)dt\ ds$$
$$= N^{1/2}\int_0^1\int_0^1\left\{\frac{1}{N^{1/2}}\sum_{k=1}^{N}\left(X_k(t)X_k(s) - C(t,s)\right)\right\}\widehat{\varphi}_i(t)\widehat{\varphi}_j(s)dt\,ds.$$

Using Lemmas 5.1 – 5.3 we get

$$\left|\int_0^1\int_0^1\left\{\frac{1}{N^{1/2}}\sum_{k=1}^{N}\left(X_k(t)X_k(s) - C(t,s)\right)\right\}\left(\widehat{\varphi}_i(t)\widehat{\varphi}_j(s) - \widehat{c}_i\varphi_i(t)\widehat{c}_j\varphi_j(s)\right)dt\,ds\right|$$
$$= \left|\int_0^1\int_0^1\left\{\frac{1}{N^{1/2}}\sum_{k=1}^{N}\left(X_k(t)X_k(s) - C(t,s)\right)\right\}\right.$$
$$\left.\times\left\{\left(\widehat{\varphi}_i(t) - \widehat{c}_i\varphi_i(t)\right)\widehat{\varphi}_j(s) + \widehat{c}_i\varphi_i(t)\left(\widehat{\varphi}_j(s) - \widehat{c}_j\varphi_j(s)\right)\right\}dt\,ds\right|$$
$$\le \left(\int_0^1\int_0^1\left\{\frac{1}{N^{1/2}}\sum_{k=1}^{N}\left(X_k(t)X_k(s) - C(t,s)\right)\right\}^2 dt\,ds\right.$$
$$\left.\times\int_0^1\int_0^1\left(\widehat{\varphi}_i(t) - \widehat{c}_i\varphi_i(t)\right)^2\widehat{\varphi}_j^2(s)dt\,ds\right)^{1/2}$$
$$+\left(\int_0^1\int_0^1\left\{\frac{1}{N^{1/2}}\sum_{k=1}^{N}\left(X_k(t)X_k(s) - C(t,s)\right)\right\}^2 dt\,ds\right.$$
$$\left.\times\int_0^1\int_0^1\varphi_i^2(t)\left(\widehat{\varphi}_j(s) - \widehat{c}_j\varphi_j(s)\right)^2 dt\,ds\right)^{1/2}$$

$$= \left\| \frac{1}{N^{1/2}} \sum_{k=1}^{N} \big(X_k(t) X_k(s) - C(t,s) \big) \right\| \left\{ \, \|\widehat{\varphi}_i - \widehat{c}_i \varphi_i\| + \|\widehat{\varphi}_j - \widehat{c}_j \varphi_j\| \, \right\}$$
$$= o_P(1).$$

Similar arguments give that

$$\left| \int_0^1 \int_0^1 \left\{ \frac{1}{M^{1/2}} \sum_{k=1}^{M} \big(X_k^*(t) X_k^*(s) - C^*(t,s) \big) \right\} \big\{ \widehat{\varphi}_i(t) \widehat{\varphi}_j(s) - \widehat{c}_i \varphi_i(t) \widehat{c}_j \varphi_j(s) \big\} dt\, ds \right| = o_P(1).$$

Since $C = C^*$, the lemma is proven. □

The previous lemmas isolated the main terms in the differences $\widehat{\Delta}_N(i,j) - \widehat{\Delta}^*_M(i,j)$. The following lemma describes the limits of these main terms (without the random signs).

Lemma 5.5. *If the conditions of Theorem 3.1 are satisfied, then, as $N, M \to \infty$,*

$$\{ \Delta_{N,M}(i,j), 1 \le i,j \le p \} \ \xrightarrow{\mathcal{D}} \ \{ \Delta(i,j), 1 \le i,j \le p \},$$

where

$$\Delta_{N,M}(i,j) = \left(\frac{NM}{N+M} \right)^{1/2} \left\{ \frac{1}{N} \sum_{k=1}^{N} \langle X_k, \varphi_i \rangle \langle X_k, \varphi_j \rangle - \frac{1}{M} \sum_{k=1}^{M} \langle X_k^*, \varphi_i \rangle \langle X_k^*, \varphi_j \rangle \right\},$$

and $\{ \Delta(i,j), 1 \le i,j \le p \}$ is a Gaussian matrix with $E\Delta(i,j) = 0$ and

$$\begin{aligned} E\Delta(i,j)\Delta(i',j') = (1-\Theta) \big\{ & E\big(\langle X_1, \varphi_i \rangle \langle X_1, \varphi_j \rangle \langle X_1, \varphi_{i'} \rangle \langle X_1, \varphi_{j'} \rangle \big) \\ & - E\big(\langle X_1, \varphi_i \rangle \langle X_1, \varphi_j \rangle \big) E\big(\langle X_1, \varphi_{i'} \rangle \langle X_1, \varphi_{j'} \rangle \big) \big\} \\ + \Theta \big\{ & E\big(\langle X_1^*, \varphi_i \rangle \langle X_1^*, \varphi_j \rangle \langle X_1^*, \varphi_{i'} \rangle \langle X_1^*, \varphi_{j'} \rangle \big) \\ & - E\big(\langle X_1^*, \varphi_i \rangle \langle X_1^*, \varphi_j \rangle \big) E\big(\langle X_1^*, \varphi_{i'} \rangle \langle X_1^*, \varphi_{j'} \rangle \big) \big\}. \end{aligned}$$

Proof: First we note that

$$E \langle X_1, \varphi_i \rangle \langle X_1, \varphi_j \rangle = E \langle X_1^*, \varphi_{i'} \rangle \langle X_1^*, \varphi_{j'} \rangle = \begin{cases} 0 & \text{if } \ i \ne j, \\ \lambda_i & \text{if } \ i = j. \end{cases}$$

Since $E\big(\langle X_1, \varphi_i \rangle \langle X_1, \varphi_j \rangle \big)^2 < \infty$ and $E\big(\langle X_1^*, \varphi_i \rangle \rangle X_1^*, \varphi_j \rangle \big)^2 < \infty$, the multivariate central limit theorem implies the result. □

Finally, we need an asymptotic approximation to the covariances $\widehat{L}_{N,M}(k,k')$. Let

$$L_{N,M}(k,k') = (1-\Theta_{N,M})\left\{\frac{1}{N}\sum_{\ell=1}^{N} a_\ell(i)a_\ell(j)a_\ell(i')a_\ell(j') - \big\langle\widehat{\mathfrak{C}}_N\widehat{\varphi}_i,\widehat{\varphi}_j\big\rangle\big\langle\widehat{\mathfrak{C}}_N\widehat{\varphi}_{i'},\widehat{\varphi}_{j'}\big\rangle\right\}$$
$$+\Theta_{N,M}\left\{\frac{1}{M}\sum_{\ell=1}^{M} a^*_\ell(i)a^*_\ell(j)a^*_\ell(i')a^*_\ell(j') - \big\langle\widehat{\mathfrak{C}}^*_M\widehat{\varphi}_i,\widehat{\varphi}_j\big\rangle\big\langle\widehat{\mathfrak{C}}^*_M\widehat{\varphi}_{i'},\widehat{\varphi}_{j'}\big\rangle\right\},$$

where

$$a_\ell(i) = \langle X_\ell,\varphi_i\rangle \quad \text{and} \quad a^*_\ell(i) = \langle X^*_\ell,\varphi_i\rangle,$$

and i, j, i', j' are determined from k and k' as in (8) and (9).

LEMMA 5.6. *If the conditions of Theorem 3.1 are satisfied, then for all* $1 \le k,k' \le p(p+1)/2$,

$$\widehat{L}_{N,M}(k,k') - \widehat{c}_i\widehat{c}_j\widehat{c}_{i'}\widehat{c}_{j'}L_{N,M}(k,k') = o_P(1) \quad as \quad N,M\to\infty,$$

where (i,j) *and* (i',j') *are determined from* k *and* k' *as in* (8) *and* (9).

PROOF: The result follows from Lemma 5.3 along the lines of the proof of Lemma 5.4 □

Proof of Theorem 3.1. According to Lemmas 5.2 and Lemmas 5.4 – 5.6, the asymptotic distribution of $\widehat{\boldsymbol{\xi}}^T_{N.M}\widehat{L}^{-1}_{N,M}\widehat{\boldsymbol{\xi}}_{N,M}$ does not depend on the signs $\widehat{c}_1,\ldots,\widehat{c}_p$, so it is sufficient to prove the result for $\widehat{c}_1 = \ldots = \widehat{c}_p = 1$. The law of large numbers yields that

$$L_{N,M}(k,k') \xrightarrow{P} L(k,k'), \tag{21}$$

where

$$L(k,k') = (1-\Theta)\left\{E\big(a_1(i)a_1(j)a_1(i')a_1(j')\big) - E\big(a_1(i)a_1(j)a_1(i')a_1(j')\big)\right\}$$
$$+\Theta\left\{E\big(a^*_1(i)a^*_1(j)a^*_1(i')a^*_1(j')\big) - E\big(a^*_1(i)a^*_1(j)a^*_1(i')a^*_1(j')\big)\right\}. \tag{22}$$

The result then follows from Lemmas 5.2, 5.4 and 5.5 □

Proof of Theorem 3.2. In the case of Gaussian observations, $\Delta(i,j)$, $1\le i\le j\le p$, are independent normal random variables with mean 0 and

$$E\Delta^2(i,j) = \begin{cases} \lambda_i\lambda_j & \text{if} \quad i\ne j,\\ 2\lambda_i^2 & \text{if} \quad i=j.\end{cases}$$

Now the result follows from Lemmas 5.1 – 5.5. For more details we refer to Panaretos *et al.* (2010). □

Proof of Theorem 3.3. First we observe that by the law of large numbers we have

$$\int_0^1 \int_0^1 (\widehat{R}_{N,M}(t,s) - R(t,s))^2 dt\, ds = o_P(1).$$

Hence using the result in section VI.1. of Gohberg *et al.* (1990) we get that

$$\max_{1 \le i \le p} \left| \widehat{\lambda}_i - \lambda_i \right| = o_P(1) \tag{23}$$

and

$$\max_{1 \le i \le p} \| \widehat{\varphi}_i - \widehat{c}_i \varphi_i \| = o_P(1), \tag{24}$$

where $\widehat{c}_i = \widehat{c}_i(N,M) = \operatorname{sign}(\langle \widehat{\varphi}_i, \varphi_i \rangle)$. Relations (23) and (24) show that Lemma 5.3 remains true. It follows from the law of large numbers and (24) that for all $1 \le i, j \le p$

$$\begin{aligned}
&\left| \widehat{\Delta}_N(i,j) - \widehat{\Delta}^*_M(i,j) - \widehat{c}_i \widehat{c}_j \int_0^1 \int_0^1 \left(C(t,s) - C^*(t,s)\right) \varphi_i(t)\varphi_j(s) dt\, ds \right| \\
&= \left| \int_0^1 \int_0^1 \left(\widehat{C}_N(t,s) - \widehat{C}^*_M(t,s) \right) \widehat{\varphi}_i(t)\widehat{\varphi}_j(s) dt\, ds - \widehat{c}_i \widehat{c}_j \int_0^1 \int_0^1 \left(C(t,s) - C^*(t,s)\right) \varphi_i(t)\varphi_j(s) dt\, ds \right| \\
&\le \left| \int_0^1 \int_0^1 \left(\widehat{C}_N(t,s) - C(t,s) - \left(\widehat{C}^*_M(t,s) - C^*(t,s) \right) \right) \widehat{\varphi}_i(t)\widehat{\varphi}_j(s) dt\, ds \right| \\
&\quad + \left| \int_0^1 \int_0^1 \left(C(t,s) - C^*(t,s)\right) \left(\widehat{\varphi}_i(t)\widehat{\varphi}_j(s) - \widehat{c}_i \varphi_i(t) \widehat{c}_i \varphi_j(s) \right) dt\, ds \right| \\
&\le \left\| \widehat{C}_N - C \right\| + \left\| \widehat{C}^*_M - C^* \right\| + \| C - C^* \| \, \| \widehat{\varphi}_i \widehat{\varphi}_j - \widehat{c}_i \varphi_i \widehat{c}_i \varphi_j \| \\
&= o_P(1),
\end{aligned}$$

where the fact that $\|\varphi_i\| = 1 = \|\widehat{\varphi}_i\|$ was used. Hence the proof of (12) is complete. It is also clear that (12) implies (13).

Next we observe that Lemma 5.6 and (21) remain true under the alternative. Now by some lengthy calculations it can be verified that L given in (22) is positive definite so that (14) follows from (13). □

References

Abadir, K. M., and Magnus, J.R. (2005). *Matrix algebra.* Cambridge University Press, New York.

Benko, M., Härdle, W., and A. Kneip (2009). Common functional principal components. *Ann. Statist.*, **37**, 1-34.

Boente, G., Rodriguez, D., and Sued, M. (2011). Testing the equality of covariance operators. In *Recent advances in functional data analysis and related topics* (ed F. Ferraty), 49-53, Physica-Verlag.

Bosq, D. (2000). *Linear processes in function spaces.* Springer, New York.

Dauxois, J., Pousse, A., and Romain, Y. (1982). Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference. *J. Multiv. Analysis*, **12**, 136-154.

Ferraty, F., and Romain, Y., eds, (2011). *The Oxford handbook of functional data analysis.* Oxford University Press.

Ferraty, F., and Vieu, P. (2006). *Nonparametric functional data analysis: Theory and practice* Springer, New York.

Gabrys, R., Horváth, L., and Kokoszka, P. (2010). Tests for error correlation in the functional linear model. *J. Amer. Statist. Assoc.*, **105**, 1113-1125.

Gaines, G., Kaphle, K., and Ruymgaart, F. (2011). Application of a delta-method for random operators to testing equality of two covariance operators. *Math. Meth. Statist.*, **20**, 232-245.

Gervini, D. (2008). Robust functional estimation using the spatial median and spherical principal components. *Biometrika*, **95**, 587-600.

Gohberg, I., Goldberg, S., and Kaashoek, M.A. (1990). *Classes of linear operators.* Operator theory: Advances and applications, **49**, Birkhäuser, Basel.

Horváth, L., and Kokoszka, P. (2012). *Inference for functional data with applications.* Springer Series in Statistics. Springer, New York.

Horváth, L., Kokoszka, P., and Reeder, R. (2012). Estimation of the mean of functional time series and a two sample problem. *J. Royal Stat. Soc., Ser. B*, doi: 10.1111/j.1467-9868.2012.01032.x.

Horváth, L., Kokoszka, P., and Reimherr, M. (2009). Two sample inference in functional linear models. *Canad. J. Statist.*, **37**, 571-591.

Müller, H-G., and Stadtmüller, U. (2005). Generalized functional linear models. *Ann. Statist.*, **33**, 774-805.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2010). Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *J. Amer. Statist. Assoc.*, **105**, 670-682.

Panaretos, V. M., Kraus, D., and Maddocks, J. H. (2011). Second-order inference for functional data with application to DNA minicircles. In *Recent advances in functional data analysis and related topics* (ed F. Ferraty), 245-250, Physica-Verlag.

Ramsay, J., Hooker, G., and Graves, S. (2009). *Functional data analysis with R and MATLAB*. Springer, New York.

Ramsay, J. O., and Silverman, B. W. (2005). *Functional data analysis*. Springer, New York.

Reiss, P. T., and Ogden, R. T. (2007). Functional principal component regression and functional partial least squares. *J. Amer. Statist. Assoc.*, **102**, 984-996.

Yao, F., and Müller, H-G. (2010). Functional quadratic regression. *Biometrika*, **97**, 49-64.

FUNCTIONAL DATA ANALYSIS WITH INCREASING NUMBER OF PROJECTIONS

By Stefan Fremdt[‡], Lajos Horváth[†], Piotr Kokoszka[§] and Josef G. Steinebach[‡]

[‡]*University of Cologne,* [†]*University of Utah and* [§]*Colorado State University*

Abstract

Functional principal components (FPC's) provide the most important and most extensively used tool for dimension reduction and inference for functional data. The selection of the number, d, of the FPC's to be used in a specific procedure has attracted a fair amount of attention, and a number of reasonably effective approaches exist. Intuitively, they assume that the functional data can be sufficiently well approximated by a projection onto a finite dimensional subspace, and the error resulting from such an approximation does not impact the conclusions. This has been shown to be a very effective approach, but it is desirable to understand the behavior of many inferential procedures by considering the projections on subspaces spanned by an increasing number of the FPC's. Such an approach reflects more fully the infinite dimensional nature of functional data, and allows to derive procedures which are fairly insensitive to the selection of d. This is accomplished by considering limits as $d \to \infty$ with the sample size.

We propose a specific framework in which we let $d \to \infty$ by deriving a normal approximation for the partial sum process

$$\sum_{j=1}^{\lfloor du \rfloor} \sum_{i=1}^{\lfloor Nx \rfloor} \xi_{i,j}, \quad 0 \le u \le 1, \quad 0 \le x \le 1,$$

where N is the sample size and $\xi_{i,j}$ is the score of the ith function with respect to the jth FPC. Our approximation can be used to derive statistics that use segments of observations and segments of the FPC's. We apply our general results to derive two inferential procedures for the mean function: a change point test and a two sample test. In addition to the asymptotic theory, the tests are assessed through a small simulation study and a data example.

Keywords: Functional data, change in mean, increasing dimension, normal approximation, principal components.

1 Introduction

Functional data analysis has grown into a comprehensive and useful field of statistics which provides a convenient framework to handle some high–dimensional data structures, including curves and images. The monograph of Ramsay and Silverman (2005) has done a lot to introduce its ideas to the statistics community and beyond. Several other monographs and thousands of papers followed. This paper focuses on a specific aspect of the mathematical foundations of functional data analysis, which is however of fairly central importance. We first describe the contribution of this paper in broad terms, and provide some more detailed background and discussion in the latter part of this section.

Perhaps the most important, and definitely the most commonly used, tool for dimension reduction of functional data is the principal component analysis. Suppose we observe a sample of functions, $X_1, X_2, \ldots, X_N$, and denote by

$$\hat{\eta}_{i,j} = \int \big(X_i(t) - \bar{X}_N(t)\big)\, \hat{v}_j(t)dt, \qquad i = 1, 2, \ldots, N, \quad j = 1, 2, \ldots, d,$$

the scores of the X_i with respect to the estimated functional principal components $\hat{v}_j$. The scores $\hat{\eta}_{i,j}$ depend on two variables i and j, and to reflect the infinite–dimensional nature of the data, it may be desirable to consider asymptotics in which both N and d increase. This paper establishes results that allow us to study the two–dimensional partial sum process

$$\sum_{j=1}^{\lfloor du \rfloor} \sum_{i=1}^{\lfloor Nx \rfloor} \int (X_i(t) - \mu_X(t))\, v_j(t)dt, \quad 0 \le u \le 1, \quad 0 \le x \le 1.$$

More specifically, we derive a uniform normal approximation and apply it to two problems related to testing the null hypothesis that all observed curves have the same mean function. We obtain new test statistics in which the number of the functional principal components, d, increases slowly with the sample size N. We hope that our general approach will be used to derive similar results in other settings.

Statistical procedures for functional data which use functional principal components (FPC's) often depend on the number d of the components used to compute various statistics. The selection of an optimal d has received a fair deal of attention. Commonly used approaches include the cumulative variance method, the scree plot, and several forms of cross–validation and pseudo information criteria. By now, most of these approaches are implemented in several R packages and in the Matlab package PACE. A related direction of research has focused on the identification of the dimension d assuming that the functional data actually live in a finite–dimensional space of this dimension, see Hall and Vial (2006) and Bathia et al. (2010). The research presented in this paper is concerned with functional data which cannot be reduced to finite–dimensional data in an obvious and easy way. Such data are

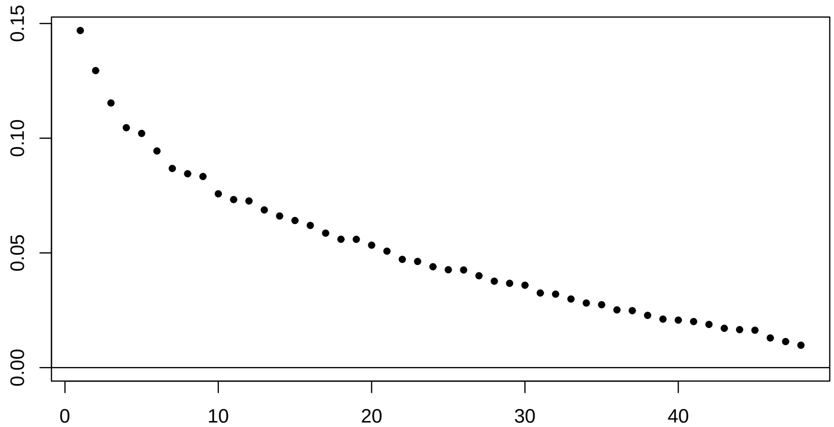


Figure 1: Melbourne temperature data: eigenvalues $\hat{\lambda}_2, \ldots, \hat{\lambda}_{49}$.

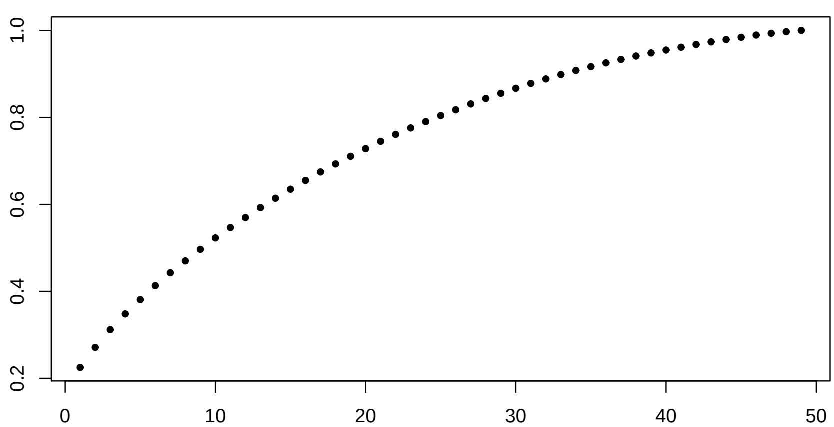


Figure 2: Melbourne temperature data: percentage of variance explained by the first k eigenvalues, i.e. $f_k = \sum_{i=1}^{k} \hat{\lambda}_i / \sum_{j=1}^{N} \hat{\lambda}_j, k = 1, 2, \ldots, 49$.

typically characterized by a slow decay of the eigenvalues of the empirical covariance operator. Figure 1 shows the eigenvalues of the empirical covariance operator of the annual temperature curves obtained over the period 1856–2011 in Melbourne, Australia, while Figure 2 shows the cumulative variance plot for the same data set. It is seen that the eigenfunctions decay at a slow rate, and neither their visual inspection nor the analysis of cumulative variance provide a clear guidance on how to select d. This data set is analyzed in greater detail in Section 5.

In situations when the choice of d is difficult, two approaches seem reasonable. In the first approach, one can apply a test using several values of d in a reasonable range. If the conclusion does not depend on d, we can be confident that it is correct. This approach has been used in applied research, see Gromenko et al. (2012) for a recent analysis of this type. The second approach, would be to let d increase with the sample size N, and derive a test statistic based on the limit. In a sense, the second approach is a formalization of the first one because if a limit as $d \to \infty$ exists, then the conclusions should not depend on the choice of d, if it is reasonably large. In the FDA community there is a well grounded intuition that d should increase much slower than N, so asymptotically large d need not be very large in practice. It is also known that the rate at which d increases should depend on the manner in which the eigenvalues decay. We obtain specific conditions that formalize this intuition in the framework we consider. In more specific settings, contributions in this directions were made by Cardot et al. (2003) and Panaretos et al. (2010). The work of Cardot et al. (2003) is more closely related to our research: as part of the justification of their testing procedure, they establish conditions under which a limiting chi–square distribution with d degrees of freedom can be approximated by a normal distribution as $d = d(N) \to \infty$. Panaretos et al. (2010) are concerned with a test of the equality of the covariance operators in two samples of Gaussian curves. In the supplemental material, they derive asymptotics in which d is allowed to increase with the sample size. Our theory is geared toward testing the equality of mean functions, but we do not assume the normality of the functional observations, so we cannot use arguments that use the equivalence of independence and zero covariances. We develop a new technique based on the estimation of the Prokhorov–Lévy distance between the underlying processes and the corresponding normal partial sums.

The paper is organized as follows. In Section 2, we set the framework and state a general normal approximation result in Theorem 2.1. This result is then used in Sections 3 and 4 to derive, respectively, change–point and two–sample tests based on an increasing number of FPC's. Section 5 contains a small simulation study and an application to the annual Melbourne temperature curves. All proofs are collected in the appendices.

2 Uniform normal approximation

We consider functional observations $X_i(t),\ t \in \mathcal{I},\ i = 1, 2, \ldots, N$, defined over a compact interval $\mathcal{I}$. We can and shall assume without loss of generality that $\mathcal{I} = [0, 1]$. Throughout the paper, we use the notation $\int = \int_0^1$ and

$$\langle f, g \rangle = \int f(t)g(t)dt, \quad \|f\|^2 = \langle f, f \rangle.$$

All functions we consider will be elements of the Hilbert space L^2 of square integrable functions on $[0, 1]$.

In the testing problems that motivate this research, under the null hypothesis, the observations follow the model

$$X_i(t) = \mu(t) + Z_i(t), \quad 1 \le i \le N, \tag{2.1}$$

where $EZ_i(t) = 0$ and $\mu(t)$ is the common mean. We impose the following standard assumptions.

ASSUMPTION 2.1. *$Z_1, Z_2, \ldots, Z_N$ are independent and identically distributed.*

ASSUMPTION 2.2. *$\int \mu^2(t)dt < \infty$ and $E\|Z_1\|^2 < \infty$.*

Under these assumptions, the covariance function

$$\mathfrak{c}(t, s) = EZ_1(t)Z_1(s),$$

is square integrable on the unit square and therefore it has the representation

$$\mathfrak{c}(t, s) = \sum_{k=1}^{\infty} \lambda_k v_k(t) v_k(s),$$

where $\lambda_1 \ge \lambda_2 \ge \ldots$ are the eigenvalues and $v_1, v_2, \ldots$ are the orthonormal eigenfunctions of the covariance operator, i.e. they satisfy the integral equation

$$\lambda_j v_j(t) = \int \mathfrak{c}(t, s) v_j(s) ds. \tag{2.2}$$

One of the most important dimension reduction techniques of functional data analysis is to project the observations $X_1(t), \ldots, X_N(t)$ onto the space spanned by $v_1, \ldots, v_d$, the eigenfunctions associated with the d largest eigenvalues. Since the covariance function $\mathfrak{c}$, and therefore $v_1, \ldots, v_d$, are unknown, we use the empirical eigenfunctions $\hat{v}_1, \ldots, \hat{v}_d$ and eigenvalues $\hat{\lambda}_1 \ge \hat{\lambda}_2 \ge \ldots \ge \hat{\lambda}_d$ defined by

$$\hat{\lambda}_j \hat{v}_j(t) = \int \hat{\mathfrak{c}}_N(t, s) \hat{v}_j(s) ds, \tag{2.3}$$

where

$$\hat{\mathfrak{c}}_N(t,s) = \frac{1}{N}\sum_{i=1}^{N}\left(X_i(t)-\bar{X}_N(t)\right)\left(X_i(s)-\bar{X}_N(s)\right)$$

with $\bar{X}_N(t) = N^{-1}\sum_{i=1}^{N} X_i(t)$.
In this section, we require only two more assumptions, namely

ASSUMPTION 2.3. $\lambda_1 > \lambda_2 > \ldots$

ASSUMPTION 2.4. $E\|Z_1\|^3 < \infty$.

Assumption 2.3 is needed to ensure that the FPC's v_j are uniquely defined. In Theorem 2.1 it could, of course, be replaced by requiring only that the first d eigenvalues are positive and different, but since in the applications we let $d \to \infty$, we just assume that all eigenvalues are positive and distinct. If $\lambda_{d^*+1} = 0$ for some d^*, then the observations are in the linear span of $v_1, \ldots, v_{d^*}$, i.e. they are elements of a d^*–dimensional space, so in this case we cannot consider $d = d(N) \to \infty$. Assumption 2.3 means that the observations are in an infinite–dimensional space. Assumption 2.4 is weaker than the usual assumption $E\|Z_1\|^4 < \infty$. As will be seen in the proofs, subtle arguments of the probability theory in Banach spaces are needed to dispense with the fourth moment.
To state the main result of this section, define

$$\boldsymbol{\xi}_i = (\xi_{i,1}, \ldots, \xi_{i,d})^T \quad \text{and} \quad \xi_{i,j} = \lambda_j^{-1/2}\langle Z_i, v_j\rangle, \;\; 1 \le i \le N, \; 1 \le j \le d,$$

where $\cdot^T$ denotes the transpose of vectors and matrices. Set

$$S_{j,N}(x) = \frac{1}{N^{1/2}}\sum_{i=1}^{\lfloor Nx\rfloor}\xi_{i,j}, \qquad 0 \le x \le 1, \quad 1 \le j \le d. \tag{2.4}$$

We now provide an approximation for the partial sum processes $S_{j,N}(x)$ defined in (2.4) with suitably constructed Wiener processes (standard Brownian motions).

THEOREM 2.1. *If Assumptions 2.1, 2.3 and 2.4 hold, then for every N we can define independent Wiener processes $W_{1,N}, \ldots, W_{d,N}$ such that*

$$P\left\{\max_{1\le j\le d}\sup_{0\le x\le 1}|S_{j,N}(x) - W_{j,N}(x)| \ge N^{1/2-1/80}\right\} \tag{2.5}$$

$$\le c_* N^{-1/80}\left\{d^{1/12}\left(\sum_{\ell=1}^{d} 1/\lambda_\ell\right)^{1/8} + \sum_{j=1}^{d} 1/\lambda_j^{3/2}\right\},$$

where c_* only depends on λ_1 and $E\|Z_1\|^3$.

The constant 1/80 in (2.5) is not crucial, it is a result of our calculations. Theorem 2.1 is related to the results of Einmahl (1987, 1989) who obtained strong approximations for partial sums of independent and identically distributed random vectors with zero mean and with identity covariance matrix. In our setting, for any fixed d, the covariance matrix is not the identity, but this is not the central difficulty. The main value of Theorem 2.1 stems from the fact that it shows how the rate of the approximation depends on d; no such information is contained in the work of Einmahl (1987, 1989), who did not need to consider the dependence on d. The explicit dependence of the right hand side of (2.5) on d is crucial in the applications presented in the following sections in which the dimension of the projection space depends on the sample size N.

Very broadly speaking, Theorem 2.1 implies that in all reasonable statistics based on averaging the scores, even in those based on an increasing number of FPC's, the partial sums of scores can be replaced by Wiener processes to obtain a limit distribution. The right hand side of (2.5) allows us to derive assumptions on the eigenvalues required to obtain a specific result. Replacing the unobservable scores $\xi_{i,j}$ by the sample scores $\hat{\eta}_{i,j}$ is relatively easy. We will illustrate these ideas in Sections 3 and 4.

3 CHANGE–POINT DETECTION

Over the past four decades, the investigation of the asymptotic properties of partial sum processes has to a large extent been motivated by change–point detection procedures, and this is the most natural application of Theorem 2.1. The research on the change–point problem in various contexts is very extensive, some aspects of the asymptotic theory are presented in Csörgő and Horváth (1997). Detection of a change in the mean function was studied by Berkes et al. (2009) who considered a procedure in which the number of the FPC's, d, was fixed, and the asymptotic distribution of the test statistic depended on d. We show in this section that it is possible to derive tests with a standard normal limiting distribution by allowing the d to depend on the sample size N.

We want to test whether the mean of the observations remained the same during the observation period, i.e. we test the null hypothesis

$$H_0 : \quad EX_1(\cdot) = EX_2(\cdot) = \cdots = EX_N(\cdot)$$

("=" means equality in L^2). Under the null hypothesis, the X_i follow model (2.1) in which $\mu(\cdot)$ is an unknown common mean function under H_0. The alternative hypothesis is

$$H_A : \text{there is } \; k^* \in [1, 2, \ldots, N) \; \text{ such that}$$
$$EX_1(\cdot) = \cdots = EX_{k^*}(\cdot) \neq EX_{k^*+1}(\cdot) = \cdots = EX_N(\cdot).$$

Under H_A the mean changes at an unknown time k^*.
To derive a new class of tests, we introduce the process

$$\hat{Z}_N(u,x) = \frac{1}{d^{1/2}} \sum_{j=1}^{\lfloor du \rfloor} \left\{ \frac{1}{N} \left[\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N) \right]^2 - x(1-x) \right\}, \quad 0 \le u, x \le 1,$$

where

$$\hat{S}_j(k) = \frac{1}{\hat{\lambda}_j^{1/2}} \sum_{i=1}^{k} \hat{\eta}_{i,j}.$$

The process $\hat{Z}_N(u,x)$ contains the cumulative sums $\hat{S}_j(\lfloor Nx \rfloor) - x\hat{S}_j(N)$ which measure the deviation of the partial sums from their "trend" under H_0, and a correction term $x(1-x)$ needed to ensure convergence as $d \to \infty$.
To obtain a limit which does not depend on any unknown quantities, we need to impose assumptions on the rate at which $d = d(N)$ increases with N. Intuitively, the assumptions below state that d is much smaller than the sample size N, the d largest eigenvalues are not too small, and that the difference between the consecutive eigenvalues tends to zero slowly. Very broadly speaking, these assumptions mean that the distribution of the observations must sufficiently fill the whole infinite–dimensional space L^2.

ASSUMPTION 3.1. $d = d(N) \to \infty$

ASSUMPTION 3.2. $(d \log N)^{1/2} N^{-1/80} \to 0,$

ASSUMPTION 3.3. $d^{1/12} N^{-1/80} \left(\sum_{j=1}^{d} 1/\lambda_j \right)^{1/8} \to 0.$

ASSUMPTION 3.4. $N^{-1/80} \sum_{j=1}^{d} 1/\lambda_j^{3/2} \to 0.$

ASSUMPTION 3.5.

$$\frac{1}{d^{1/2} N^{1/3}} \sum_{j=1}^{d} \frac{1}{\lambda_j \zeta_j} \to 0,$$

where $\zeta_1 = \lambda_2 - \lambda_1,\ \zeta_j = \min(\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1}),\ j \ge 2.$

With these preparations, we can state the main result of this section.

THEOREM 3.1. *If Assumptions 2.1–2.3 and 3.1–3.5 are satisfied, then*

$$\hat{Z}_N(u,x) \ \to \ \Gamma(u,x) \ \ in \ \ \mathcal{D}[0,1]^2,$$

where $\Gamma(u,x)$ *is a mean zero Gaussian process with*

$$E[\Gamma(u,x)\Gamma(v,y)] = 2ux^2(1-y)^2, \quad 0 \le u \le v \le 1, \quad 0 \le x \le y \le 1.$$

One can verify by computing the covariance functions that

$$\{\Gamma(u,x),\ 0\le u,x\le 1\} \stackrel{\mathcal{D}}{=} \{\sqrt{2}(1-x)^2 W(u,x^2/(1-x)^2),\ 0\le u,x\le 1\}, \tag{3.1}$$

where $\{W(v,y),v,y\ge 0\}$ is a bivariate Wiener process, i.e. $W(v,y)$ is a Gaussian process with $EW(v,y)=0$ and $E[W(v,y)W(v',y')]=\min(v,v')\min(y,y')$. Representation (3.1) means that continuous functionals of the process $\Gamma(\cdot,\cdot)$ can be simulated with arbitrary precision, so Monte Carlo tests can be used. It is however possible to obtain a number of simple asymptotic tests by examining closer the structure of the process $\Gamma(\cdot,\cdot)$. We list some of them in Corollary 3.1, and we will see in Section 5 that the Cramér-von-Mises type tests have very good finite sample properties. Let B denote a Brownian bridge and define

$$\mu_0 = E\left(\sup_{0\le x\le 1} B^2(x)\right) \quad \text{and} \quad \sigma_0^2 = \operatorname{var}\left(\sup_{0\le x\le 1} B^2(x)\right).$$

Corollary 3.1. *If the assumptions of Theorem 3.1 are satisfied, then*

$$\frac{1}{d^{1/2}\sigma_0}\left\{\sum_{j=1}^d \sup_{0\le x\le 1}\frac{1}{N}\left(\hat S_j(\lfloor Nx\rfloor)-x\hat S_j(N)\right)^2 - d\mu_0\right\} \stackrel{\mathcal{D}}{\to} N(0,1), \tag{3.2}$$

$$\frac{1}{(d/45)^{1/2}}\left\{\sum_{j=1}^d \frac{1}{N}\int(\hat S_j(\lfloor Nx\rfloor)-x\hat S_j(N))^2dx - \frac{d}{6}\right\} \stackrel{\mathcal{D}}{\to} N(0,1), \tag{3.3}$$

$$\frac{1}{(d/8)^{1/2}}\left\{\sup_{0\le x\le 1}\sum_{j=1}^d \frac{1}{N}(\hat S_j(\lfloor Nx\rfloor)-x\hat S_j(N))^2 - \frac{d}{4}\right\} \stackrel{\mathcal{D}}{\to} N(0,1), \tag{3.4}$$

where $N(0,1)$ stands for a standard normal random variable.

We conclude this section with two examples which show that Assumptions 3.2–3.5 hold under both power law and exponential decay of the eigenvalues.

Example 3.1. *If the eigenvalues satisfy*

$$\lambda_j = \frac{c_1}{(j-c_2)^\alpha} + o\left(\frac{1}{j^{\alpha+1}}\right), \quad \textit{as} \quad j\to\infty,$$

with some $c_1>0,\ 0\le c_2<1$ and $\alpha>0$, then Assumptions 3.2–3.5 hold if $d/(\log N)^\beta\to 0$ with some $\beta>0$.

Example 3.2. *If the eigenvalues satisfy*

$$\lambda_j = c_0 e^{-\alpha j} + o(e^{-\alpha j}), \quad \textit{as} \quad j\to\infty,$$

with some $c_0>0$ and $\alpha>0$, then Assumptions 3.2–3.5 hold if $d/(\log\log N)^\beta\to 0$ with some $\beta>0$.

4 Two–sample problem

The two–sample problem for functional data was perhaps first discussed in depth by Benko et al. (2009) who were motivated by a problem related to implied volatility curves. It has recently attracted a fair amount of attention motivated by problems arising in space physics, see Horváth et al. (2009), genetics, see Panaretos et al. (2010), and finance, see Horváth et al. (2012). The above list does not include many other important contributions. In its simplest, but most important form, it is about testing if curves obtained from two populations have the same mean functions. The most direct approach, developed into a bootstrap procedure by Benko et al. (2009), is to look at the norm of the difference of the estimated mean functions. In this section, we show that the normal approximation of Section 2 leads to an asymptotic test whose limit distribution is standard normal.

Suppose we have two random samples of functions: $X_1, \ldots, X_N$ and $Y_1, \ldots, Y_M$. We assume the X sample satisfies (2.1) and Assumptions 2.1, 2.2 and 2.4. Similarly, the Y sample is a location model given by

$$Y_i(t) = \mu_*(t) + Q_i(t), \quad 1 \le i \le M, \tag{4.1}$$

where $\mu_*(t)$ is the common mean of the Y sample and $EQ_i(t) = 0$. As in the case of the X sample, the Y sample satisfies the following conditions:

Assumption 4.1. *$Q_1, Q_2, \ldots, Q_M$ are independent and identically distributed.*

Assumption 4.2. *$\int \mu_*^2(t)dt < \infty$ and $E\|Q_1\|^3 < \infty$.*

Assumption 4.2 yields that

$$\mathfrak{c}_*(t,s) = EQ_1(t)Q_1(s)$$

is a square integrable function on the unit square.

In this section we are interested in testing the null hypothesis

$$H_0^* : \quad \mu(\cdot) = \mu_*(\cdot).$$

The statistical inference to test H_0 is based on the difference $\bar{X}_N - \bar{Y}_M$, where $\bar{X}_N$ and $\bar{Y}_M$ denote the sample means. We assume

Assumption 4.3.

$$\frac{N}{M} = \lambda + O(N^{-1/4}) \quad \textit{as} \quad \min(M,N) \to \infty$$

with some $0 < \lambda < \infty$.

Now we define the pooled covariance function

$$\mathfrak{c}_P(t,s) = \mathfrak{c}(t,s) + \lambda \mathfrak{c}_*(t,s).$$

Since $\mathfrak{c}_P(t,s)$ is a positive–definite, symmetric, square integrable function, there are real numbers $\kappa_1 \geq \kappa_2 \geq \ldots$ and orthonormal functions $u_1, u_2, \ldots$ satisfying

$$\kappa_i u_i(t) = \int \mathfrak{c}_P(t,s) u_i(s) ds, \quad i = 1, 2, \ldots.$$

We wish to project $\bar{X}_N - \bar{Y}_M$ into the space spanned by $u_1, \ldots, u_d$, where $d = d(N) \to \infty$, so similarly to Assumption 2.3 we require

Assumption 4.4. $\kappa_1 > \kappa_2 > \kappa_3 > \ldots$

Assumption 4.5.

$$N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^{d} 1/\kappa_\ell \right)^{3/8} \to 0.$$

Our test statistic is

$$D_{N,M} = \sum_{i=1}^{d} N \langle \bar{X}_N - \bar{Y}_M, u_i \rangle^2 / \kappa_i.$$

As in Section 3, we need additional assumptions balancing the rate of growth of $d = d(N)$ and the rate of decay of the κ_ℓ and the differences between them.

Assumption 4.6.

$$\frac{1}{d^{1/2} N^{1/4}} \sum_{\ell=1}^{d} \frac{1}{\kappa_\ell^2} \to 0 \quad \textit{and} \quad \frac{1}{d^{1/2} N^{1/4}} \sum_{\ell=1}^{d} \frac{1}{\kappa_\ell \iota_\ell} \to 0,$$

where $\iota_1 = \kappa_2 - \kappa_1,\ \iota_\ell = \min(\iota_{\ell-1} - \iota_\ell, \iota_\ell - \iota_{\ell+1}),\ \ell \geq 2.$

Since $u_1, u_2, \ldots$ are unknown, we replace them with the corresponding empirical eigenfunctions $\hat{u}_1, \hat{u}_2, \ldots$ defined by the integral operator

$$\hat{\kappa}_i \hat{u}_i(t) = \int \hat{\mathfrak{c}}_P(t,s) \hat{u}_i(s) ds, \quad i = 1, 2, \ldots,$$

where $\hat{\kappa}_1 \geq \hat{\kappa}_2 \geq \ldots$ and

$$\hat{\mathfrak{c}}_P(t,s) = \hat{\mathfrak{c}}_N(t,s) + \frac{N}{M} \hat{\mathfrak{c}}_{*M}(t,s),$$

α	0.01	0.05	0.10
	0.109256	0.0726292	0.0578267

Table 5.1: Critical values for the distribution of (5.3).

with

$$\hat{\mathfrak{c}}_{*M}(t,s) = \frac{1}{M}\sum_{\ell=1}^{M}(Y_\ell(t) - \bar{Y}_M(t))(Y_\ell(s) - \bar{Y}_M(s)).$$

The empirical version of $D_{N,M}$ is

$$\widehat{D}_{N,M} = \sum_{i=1}^{d} N\langle \bar{X}_N - \bar{Y}_M, \hat{u}_i\rangle^2/\hat{\kappa}_i.$$

THEOREM 4.1. *If H_0^*, Assumptions 2.1, 2.2 and 4.1–4.6 hold, then*

$$(2d)^{-1/2}(\widehat{D}_{N,M} - d) \quad \xrightarrow{\mathcal{D}} \quad N(0,1),$$

where $N(0,1)$ stands for a standard normal random variable.

5 A SMALL SIMULATION STUDY AND A DATA EXAMPLE

The main contribution of this paper lies in the statistical theory, but it is of interest to check if the new tests derived in Sections 3 and 4 perform well in finite samples. We report the results for the test based on Theorem 3.1 in some detail, as it utilizes the convergence of the two–parameter process in full force, and such an approach has not been used before. We also comment on the tests based on Corollary 3.1 and Theorem 4.1. We conclude this section with an illustrative data example.

The simulated data which satisfy the null hypotheses of Sections 3 and 4 are generated as independent Brownian motions on the interval $[0,1]$. We generate them by using iid normal increments on 1,000 equispaced points in $[0,1]$ (random walk approximation). (Example 3.1 shows that for the Brownian motion the assumptions of Theorem 3.1 are satisfied.) Alternatives are obtained by adding the curve $at(1-t)$ after a change–point or to the observations in the second sample. The parameter a regulates the size of the change or the difference in the means in two samples.

Many tests can be obtained from Theorem 3.1 by applying functionals continuous on $\mathcal{D}[0,1]^2$. It is not our objective to provide a systematic comparison, we consider only the

$N = 100$

d	$\alpha = 0.05$ $\hat{p}$	$\alpha = 0.05$ $[a, b]$	$\alpha = 0.1$ $\hat{p}$	$\alpha = 0.1$ $[a, b]$
2	0.047	[0.0360,0.0580]	0.058	[0.0458,0.0702]
3	0.056	[0.0440,0.0680]	0.074	[0.0604,0.0876]
4	0.060	[0.0476,0.0724]	0.081	[0.0668,0.0952]
5	0.059	[0.0467,0.0713]	0.089	[0.0742,0.1038]
6	0.057	[0.0449,0.0691]	0.089	[0.0742,0.1038]
7	0.056	[0.0440,0.0680]	0.089	[0.0742,0.1038]
8	0.059	[0.0467,0.0713]	0.091	[0.0760,0.1060]
9	0.051	[0.0396,0.0624]	0.090	[0.0751,0.1049]
10	0.050	[0.0387,0.0613]	0.082	[0.0677,0.0963]
11	0.054	[0.0422,0.0658]	0.083	[0.0687,0.0973]
12	0.057	[0.0449,0.0691]	0.079	[0.0650,0.0930]
13	0.059	[0.0467,0.0713]	0.075	[0.0613,0.0887]
14	0.057	[0.0449,0.0691]	0.076	[0.0622,0.0898]
15	0.056	[0.0440,0.0680]	0.075	[0.0613,0.0887]

$N = 200$

d	$\alpha = 0.05$ $\hat{p}$	$\alpha = 0.05$ $[a, b]$	$\alpha = 0.1$ $\hat{p}$	$\alpha = 0.1$ $[a, b]$
2	0.039	[0.0289,0.0491]	0.055	[0.0431,0.0669]
3	0.048	[0.0369,0.0591]	0.070	[0.0567,0.0833]
4	0.049	[0.0378,0.0602]	0.075	[0.0613,0.0887]
5	0.053	[0.0413,0.0647]	0.076	[0.0622,0.0898]
6	0.057	[0.0449,0.0691]	0.085	[0.0705,0.0995
7	0.057	[0.0449,0.0691]	0.085	[0.0705,0.0995]
8	0.053	[0.0413,0.0647]	0.085	[0.0705,0.0995]
9	0.051	[0.0396,0.0624]	0.083	[0.0687,0.0973]
10	0.051	[0.0378,0.0602]	0.081	[0.0668,0.0952]
11	0.054	[0.0496,0.0624]	0.083	[0.0687,0.0973]
12	0.052	[0.0405,0.0635]	0.086	[0.0714,0.1006]
13	0.050	[0.0387,0.0613]	0.087	[0.0723,0.1017]
14	0.054	[0.0422,0.0658]	0.086	[0.0714,0.1006]
15	0.052	[0.0405,0.0635]	0.079	[0.0650,0.0930]

Table 5.2: Empirical sizes and 90% confidence intervals for the probability of rejection for the change–point test based on convergence (5.1).

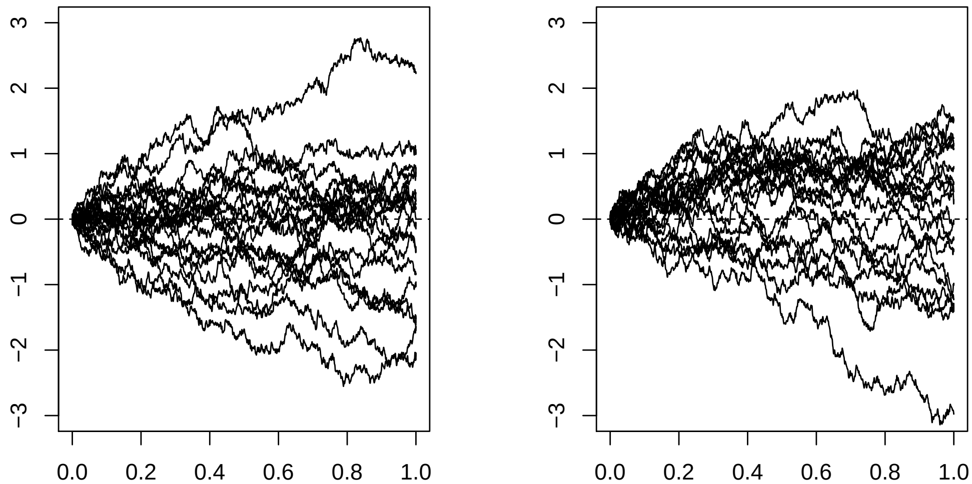


Figure 3: Left panel: 20 realizations of the Brownian motion; Right panel: independent 20 realizations of the Brownian motion with the curve $at(1-t), a = 1.5$ added.

test based on the weak convergence

$$\int_0^1 \int_0^1 \hat{Z}_N^2(u,x)dudx \;\to\; \int_0^1 \int_0^1 \Gamma^2(u,x)dudx. \tag{5.1}$$

To compute the critical values, we use the following representation of the limit

$$\int_0^1 \int_0^1 \Gamma^2(u,x)dudx \overset{\mathcal{D}}{=} \sum_{1\le k,\ell<\infty} \lambda_k \nu_\ell N_{k,\ell}^2. \tag{5.2}$$

In (5.2), the $\lambda_k = (\pi(k-1/2))^{-2}$ are the eigenvalues of the Wiener process, the ν_ℓ are the eigenvalues of the covariance operator with kernel $2(\min(s,t)-st)^2$, and $\{N_{k,\ell}\}$ is an array of independent standard normal random variables. The critical values were determined for a truncated version of the right–hand side of (5.2) with truncation level 49, i.e. for

$$\sum_{1\le k,\ell\le 49} \lambda_k \nu_\ell N_{k,\ell}^2. \tag{5.3}$$

Since the eigenvalues ν_ℓ are difficult to determine explicitly, they were calculated numerically using the `R` package `fda`, cf.Ramsay et al. (2009). The simulated critical values based on 100,000 replications of (5.3) are provided in Table 5.1.

Table 5.2 shows the empirical sizes $\hat{p}$, i.e. the fraction of rejections, as well as asymptotic 90% confidence intervals

$$\left[\hat{p} - 1.654\sqrt{\frac{\hat{p}(1-\hat{p})}{R}},\;\; \hat{p} + 1.654\sqrt{\frac{\hat{p}(1-\hat{p})}{R}}\right]. \tag{5.4}$$

$N = 100$

	$a = 1$		$a = 1.5$	
d	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
2	0.168	0.192	0.356	0.398
3	0.456	0.517	0.819	0.851
4	0.501	0.564	0.843	0.875
5	0.496	0.564	0.855	0.887
6	0.481	0.552	0.847	0.883
7	0.473	0.543	0.843	0.881
8	0.465	0.530	0.834	0.874
9	0.461	0.519	0.823	0.870
10	0.453	0.504	0.812	0.859
11	0.441	0.501	0.802	0.853
12	0.431	0.496	0.793	0.844
13	0.420	0.484	0.791	0.834
14	0.400	0.472	0.782	0.822
15	0.388	0.467	0.767	0.817

$N = 200$

	$a = 1$		$a = 1.5$	
d	$\alpha = 0.05$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.10$
2	0.327	0.370	0.620	0.660
3	0.784	0.814	0.984	0.991
4	0.808	0.849	0.988	0.994
5	0.823	0.860	0.992	0.994
6	0.825	0.863	0.991	0.996
7	0.819	0.864	0.992	0.994
8	0.814	0.859	0.990	0.994
9	0.802	0.846	0.990	0.993
10	0.791	0.837	0.990	0.993
11	0.766	0.830	0.988	0.992
12	0.754	0.821	0.987	0.992
13	0.740	0.800	0.987	0.991
14	0.734	0.794	0.987	0.991
15	0.726	0.787	0.986	0.990

Table 5.3: Power of the test based on convergence (5.1). The change–point is at $k^* = \lfloor N/2 \rfloor$.

for the probability p of rejection. The entries are based on $R = 1,000$ replications. The table shows that the test based on convergence (5.1) has correct empirical size at the 5% level and is a bit too conservative at the 10% level. However even at the 10% level the empirical sizes for $d \geq 3$ are not significantly different; they all fall into each others 90% confidence intervals. This illustrates the main point that for the tests that use the asymptotics with $d \to \infty$ developed in the paper, selecting d is not essential; every sufficiently large d gives the same conclusion on the significance.

The empirical power of the test is reported in Table 5.3. Again, for $d \geq 3$, the power remains statistically the same. We note that the change in mean equal to the function $at(1-t)$ with $a = 1.5$ is fairly small if the "noise curves" are Brownian motions. This is illustrated in Figure 3 which shows 20 Brownian motions in the left panel and another independent sample of 20 Brownian motions with the curve $at(1-t), a = 1.5$ added. If one knows that this curve was added, one can discern it in the plot in the right panel, but the difference would have been much less obvious if individual curves were observed, as in the change–point setting relevant to Table 5.3.

Regarding Corollary 3.1, we found out that the test based on convergence (3.3) has empirical size only slightly higher than nominal (about 1% at 5% level). For $d \geq 3$, the empirical size does not depend on d. The test based on (3.4) severely overrejects for $N = 100$, and we do not recommend it. The test based on Theorem 4.1 overrejects by about 2% at the 5% level, and by about 1% at the 10% level. The power of the test is above 95% for $N, M = 100$ and $a = 1.0$, and practically 100% for larger a or N, M. For $d \geq 2$, the rejection probabilities do not depend on d.

Change–point analysis of annual temperature profiles. The goal of this section is to illustrate the application of the change–point test based on convergence (5.1). Change–point analysis is an important field of statistics with a large number of applications, the recent monographs of Chen and Gupta (2011) and Basseville et al. (2012) provide numerous references. The change–point problem in the context of functional data has also received some attention, we refer to Horváth and Kokoszka (2012) for the references, Aston and Kirch (2012) report some most recent research.

The data set we study consists of 156 years (1856-2011) of minimum daily temperatures in Melbourne. These data are available at `www.bom.gov.au` (the Australian Bureau of Meteorology website). The original data can be viewed as 156 curves with 365 measurements on each curve. We converted them to functional objects in `R` using 49 Fourier basis functions. Five consecutive functions are shown in Figure 4. It is important to emphasize the difference between the data we use and the Canadian temperature data made popular by the books of Ramsay and Silverman (2005) and Ramsay et al. (2009). The Canadian temperature curves are the curves at 35 locations in Canada obtained by averaging an-

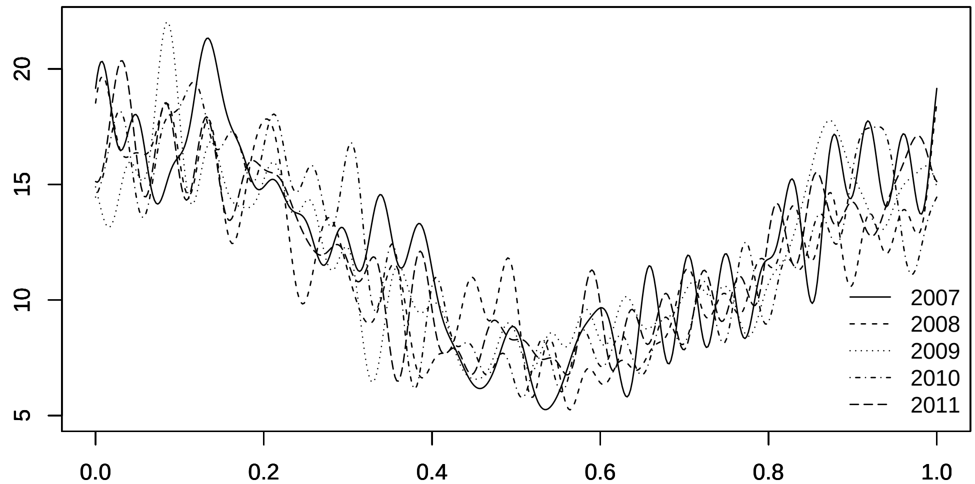


Figure 4: Five annual temperature curves represented as functional objects.

nual temperature over forty years. Since each such curve is an average of forty curves like those shown in Figure 4, those curves are much smoother, and the first two FPC's are sufficient to describe their variability. Even after smoothing with 49 Fourier functions, the annual temperature curves exhibit noticeable year to year variability, and a larger number of FPC's is needed to capture it, see Table 5.4. The goals of our analysis are also different from those of Ramsay and Silverman (2005). We are interested in detecting a change in the mean function using a sequence of noisy curves; the examples in Ramsay and Silverman (2005) used the averaged curves to describe static regression type dependencies between climatic variables.

The analysis proceeds through the usual binary segmentation procedure. The test is first applied to the whole data set. If the P–value is small, the change–point is estimated as

$$\hat{\theta}_N = \inf\{k : I_N(k) = \sup_{1\le j\le N} I_N(j)\},$$

where

$$I_N(\ell) = \frac{1}{d^2}\sum_{i=1}^{d-1}\left(\sum_{j=1}^{i}\left\{\frac{1}{N}\left[\hat{S}_j(\ell) - \frac{\ell}{N}\hat{S}_j(N)\right]^2 - \frac{\ell}{N}\left(\frac{N-\ell}{N}\right)\right\}\right)^2.$$

(I_N is a discretization of $\hat{Z}_N$.) The test is then applied to the two segments, and the procedure continues until no change–points are detected. In practice, a procedure of this type detects only a few change–points (four in our case), so the problems of multiple testing are not an issue. We applied the test using many values of d, and we were pleased to see that the final segmentation does not depend on d. Table 5.5 shows the outcome. The

k	1	2	3	4	5	6	7	8
$\hat{\lambda}_k$	0.7151	0.1469	0.1295	0.1154	0.1046	0.1021	0.0944	0.0868
f_p	0.2248	0.2711	0.3118	0.3480	0.3809	0.4130	0.4427	0.4700
k	9	10	11	12	13	14	15	16
$\hat{\lambda}_k$	0.0845	0.0833	0.0758	0.0732	0.0726	0.0687	0.0661	0.0641
f_p	0.4966	0.5228	0.5466	0.5696	0.5925	0.6141	0.6349	0.6550
k	17	18	19	20	21	22	23	24
$\hat{\lambda}_k$	0.0620	0.0586	0.0559	0.0559	0.0534	0.0508	0.0472	0.0463
f_p	0.6745	0.6930	0.7105	0.7281	0.7449	0.7609	0.7757	0.7903
k	25	26	27	28	29	30	31	32
$\hat{\lambda}_k$	0.0440	0.0427	0.0426	0.0400	0.0377	0.0367	0.0359	0.0325
f_p	0.8041	0.8175	0.8309	0.8435	0.8553	0.8669	0.8782	0.8884
k	33	34	35	36	37	38	39	40
$\hat{\lambda}_k$	0.0320	0.0299	0.0281	0.0274	0.0252	0.0248	0.0228	0.0211
f_p	0.8985	0.9079	0.9167	0.9253	0.9332	0.9410	0.9482	0.9548
k	41	42	43	44	45	46	47	48
$\hat{\lambda}_k$	0.0207	0.0201	0.0188	0.0171	0.0166	0.0163	0.0129	0.0114
f_p	0.9614	0.9677	0.9736	0.9790	0.9842	0.9893	0.9934	0.9969

Table 5.4: Eigenvalues and percentage of variance explained by the first k eigenvalues, i.e. $f_k = \sum_{i=1}^{k} \hat{\lambda}_i / \sum_{j=1}^{N} \hat{\lambda}_j$, for $k = 1, 2, \ldots, 49$.

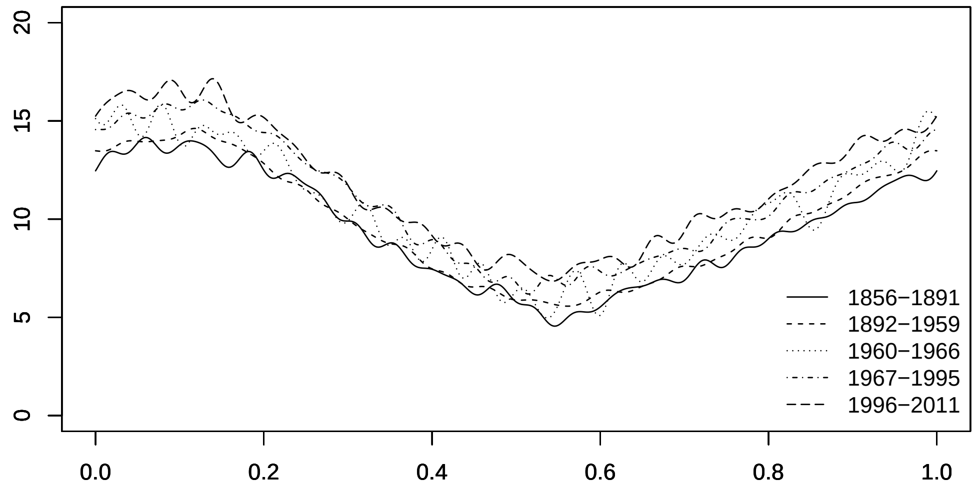


Figure 5: Average temperature functions in the estimated partition segments.

estimated change–points are the years 1892, 1960, 1967, 1996. It is clear that the change–point model is not an exact climatological model for the evolution of annual temperature curves, but it is popular in climate studies, see e.g. Gallagher et al. (2012), as it allows us to attach statistical significance to conclusions and provides periods of approximately constant mean temperature profiles. In this light, the weak evidence for a change–point in 1967 could be viewed as indicating an accelerated change in the period 1960–1995. The estimated mean temperature curves over the segments of approximately constant mean are shown in Figure 5. An increasing pattern of the mean temperature is seen; the mean curve shifted upwards by about two degrees Celsius over the last 150 years. This could be due to the conjectured global temperature increase or the urbanization of the Melbourne area, or a combination of both. A discussion of such issues is however beyond the intended scope of this paper.

A Proof of Theorem 2.1.

We start with some elementary properties of the projections $\xi_{i,j}$. Let $|\cdot|$ denote the Euclidean norm of vectors.

Lemma A.1. *If Assumptions 2.1, 2.3 and 2.4 hold, then*

$$E\boldsymbol{\xi}_1 = \mathbf{0}, \tag{A.1}$$

$$E\boldsymbol{\xi}_1\boldsymbol{\xi}_1^T = \mathbf{I}_d, \tag{A.2}$$

It.	Segment	Estimated change–point	P-value $d = 3$	$d = 4$	$d = 5$	$d = 6$
1	1856-2011	1960	0.0000	0.0000	0.0000	0.0000
2	1856-1959	1892	0.0000	0.0000	0.0000	0.0000
3	1856-1891	—	0.1865	0.2323	0.3524	0.4822
4	1892-1959	—	0.9522	0.9690	0.9256	0.6561
5	1960-2011	1996	0.0000	0.0000	0.0000	0.0000
6	1960-1995	1967	0.0013	0.0011	0.0025	0.0017
7	1960-1966	—	0.9568	0.9549	0.9818	0.9935
8	1967-1995	—	0.2927	0.4305	0.1786	0.1348
9	1996-2011	—	0.4285	0.5345	0.6413	0.7365
It.	Segment	Estimated change–point	P-value $d = 7$	$d = 8$	$d = 9$	$d = 10$
1	1856-2011	1960	0.0000	0.0000	0.0000	0.0000
2	1856-1959	1892	0.0000	0.0000	0.0000	0.0000
3	1856-1891	—	0.4235	0.4325	0.4901	0.5667
4	1892-1959	—	0.4646	0.4348	0.4696	0.5068
5	1960-2011	1996	0.0000	0.0000	0.0000	0.0000
6	1960-1995	1967	0.0026	0.0038	0.0058	0.0067
7	1960-1966	—	0.9992	—	—	—
8	1967-1995	—	0.1245	0.0690	0.0571	0.0586
9	1996-2011	—	0.8243	0.9118	0.9618	0.9779

Table 5.5: Segmentation procedure of the data into periods with constant mean function

where $\mathbf{I}_d$ *is the* $d \times d$ *identity matrix. Moreover,*

$$E|\boldsymbol{\xi}_1|^3 \leq E\|Z_1\|^3 \left(\sum_{j=1}^{d} 1/\lambda_j \right)^{3/2} \tag{A.3}$$

and for all $1 \leq j \leq d$

$$E|\xi_{1,j}|^3 \leq E\|Z_1\|^3/\lambda_j^{3/2}. \tag{A.4}$$

Proof: Since $EZ_1(t) = 0$, the relation in (A.1) is obvious. The orthonormal functions v_k and v_ℓ satisfy (2.2), so we get

$$E\xi_{i,k}\xi_{i,\ell} = \frac{1}{(\lambda_k \lambda_\ell)^{1/2}} \iint \mathfrak{c}(t,s) v_k(s) v_\ell(s) dt ds = \begin{cases} 0, & \text{if } k \neq \ell \\ 1, & \text{if } k = \ell, \end{cases}$$

proving (A.2). Using the definition of the Euclidean norm and the Cauchy–Schwarz inequality we conclude

$$|\boldsymbol{\xi}_1|^3 = \left(\sum_{j=1}^{d} \langle Z_1, v_j \rangle^2 / \lambda_j \right)^{3/2} \leq \left(\sum_{j=1}^{d} \|Z_1\|^2 \|v_j\|^2 / \lambda_j \right)^{3/2} = \|Z_1\|^3 \left(\sum_{j=1}^{d} 1/\lambda_j \right)^{3/2},$$

since $\|v_j\| = 1$. Taking the expected value of the equation above we obtain (A.3). Clearly,

$$E|\xi_{1,j}|^3 = \lambda_j^{-3/2} E|\langle Z_1, v_j \rangle|^3 \leq \lambda_j^{-3/2} E\|Z_1\|^3.$$

□

The next lemma plays a central role in the proof of Theorem 2.1.

Lemma A.2. *If Assumptions 2.1, 2.3 and 2.4 hold, then for all* n *we can define independent identically distributed standard normal vectors* $\boldsymbol{\gamma}_1, \ldots, \boldsymbol{\gamma}_n$ *in* R^d *such that*

$$P\left\{ \left| \sum_{i=1}^{n} \boldsymbol{\xi}_i - \sum_{i=1}^{n} \boldsymbol{\gamma}_i \right| \geq c n^{3/8} d^{1/4} (E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4} \right\} \leq c n^{-1/8} d^{1/4} (E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4},$$

where c *is an absolute constant.*

Proof: The result is a consequence of Theorem 6.4.1 on p. 207 of Senatov (1998) and the corollary to Theorem 11 in Strassen (1965). □

We note that

$$(E|\boldsymbol{\xi}_1|^3 + E|\boldsymbol{\gamma}_1|^3)^{1/4} \le (E|\boldsymbol{\xi}_1|^3)^{1/4} + (E|\boldsymbol{\gamma}_1|^3)^{1/4}. \tag{A.5}$$

Also, since $|\boldsymbol{\gamma}_1|^2$ is the sum of the squares of d independent standard normal random variables, Minkowski's inequality implies

$$E|\boldsymbol{\gamma}_1|^3 \le c_1 d^{3/2}, \tag{A.6}$$

with some constant c_1, and clearly

$$d^{3/2} \le \lambda_1^{3/2}\left(\sum_{\ell=1}^{d} 1/\lambda_\ell\right)^{3/2}. \tag{A.7}$$

Combining Lemma A.2 with (A.5)–(A.7), we conclude that

$$P\left\{\left|\sum_{i=1}^{n}\boldsymbol{\xi}_i - \sum_{i=1}^{n}\boldsymbol{\gamma}_i\right| \ge c_2 n^{3/8} d^{1/4}\left(\sum_{j=1}^{d} 1/\lambda_j\right)^{3/8}\right\} \le c_2 n^{-1/8} d^{1/4}\left(\sum_{j=1}^{d} 1/\lambda_j\right)^{3/8}, \tag{A.8}$$

where c_2 does not depend on d.

In the next lemma we provide an upper bound for the variance of $\sum_i^n(\xi_{i,j} - \gamma_{i,j})$, where $\boldsymbol{\gamma}_i = (\gamma_{i,1}, \ldots, \gamma_{i,d})^T$ is defined in Lemma A.2.

LEMMA A.3. *If Assumptions 2.1, 2.3 and 2.4 hold, then for any $1 \le j \le d$ we get*

$$E\left(\sum_{i=1}^{n}\xi_{i,j} - \sum_{i=1}^{n}\gamma_{i,j}\right)^2 \le c_3 n^{23/24}\frac{1}{\lambda_j}\left(d^{1/4}\left(\sum_{\ell=1}^{d} 1/\lambda_\ell\right)^{3/8}\right)^{1/3},$$

where c_3 does not depend on d.

PROOF: Let

$$U_n(j) = n^{-1/2}\sum_{i=1}^{n}(\xi_{i,j} - \gamma_{i,j}) \quad \text{and} \quad r_n = c_2 n^{-1/8} d^{1/4}\left(\sum_{\ell=1}^{d} 1/\lambda_\ell\right)^{3/8}.$$

First we write

$$\begin{aligned} EU_n^2(j) &= E[U_n^2(j) I\{|U_n(j)| \le r_n\}] + E[U_n^2(j) I\{|U_n(j)| > r_n\}] \\ &\le r_n^2 + \frac{2}{n}E\left[\left(\sum_{i=1}^{n}\xi_{i,j}\right)^2 I\{|U_n(j)| > r_n\}\right] + \frac{2}{n}E\left[\left(\sum_{i=1}^{n}\gamma_{i,j}\right)^2 I\{|U_n(j)| > r_n\}\right]. \end{aligned}$$

Using Hölder's inequality we get that

$$E\Big[\Big(\sum_{i=1}^{n}\xi_{i,j}\Big)^2 I\{|U_n(j)|>r_n\}\Big] \le E\Big[\Big|\sum_{i=1}^{n}\xi_{i,j}\Big|^3\Big]^{2/3}\Big[P\{|U_n(j)|>r_n\}\Big]^{1/3}$$
$$\le E\Big[\Big|\sum_{i=1}^{n}\xi_{i,j}\Big|^3\Big]^{2/3} r_n^{1/3}$$

by (A.8). Applying now Rosenthal's inequality (cf. Petrov (1995), p. 59) we obtain

$$E\Big|\sum_{i=1}^{n}\xi_{i,j}\Big|^3 \le c_4\Big\{\sum_{i=1}^{n}E|\xi_{i,j}|^3 + \Big(\sum_{i=1}^{n}E\xi_{i,j}^2\Big)^{3/2}\Big\},$$

where c_4 is an absolute constant. Hence

$$E\Big|\sum_{i=1}^{n}\xi_{i,j}\Big|^3 \le c_5\{n\lambda_j^{-3/2}+n^{3/2}\} \le c_6(n/\lambda_j)^{3/2}$$

and therefore

$$E\Big[\Big(\sum_{i=1}^{n}\xi_{i,j}\Big)^2 I\{|U_n(j)|>r_n\}\Big] \le c_7(n/\lambda_j)r_n^{1/3}$$
$$\le c_8 n^{23/24}\frac{1}{\lambda_j}\Big(d^{1/4}\Big(\sum_{\ell=1}^{d}1/\lambda_\ell\Big)^{3/8}\Big)^{1/3}.$$

Following the previous arguments one can show that

$$E\Big[\Big(\sum_{i=1}^{n}\gamma_{i,j}\Big)^2 I\{|U_n(j)|>r_n\}\Big] \le c_9 n^{23/24}\frac{1}{\lambda_j}\Big(d^{1/4}\Big(\sum_{\ell=1}^{d}1/\lambda_\ell\Big)^{3/8}\Big)^{1/3}.$$

The constants c_8 and c_9 do not depend on d. Since in view of Assumption 3.3, nr_n^2 is smaller than the latter rates, this completes the proof of Lemma A.3. □

Proof of Theorem 2.1. We use a blocking argument to construct a Wiener process which is close to the partial sums $\sum_{1\le i\le k}\xi_{i,j}, 1\le k\le N, 1\le j\le d$. Let K be the length of the blocks to be chosen later. Let $M=\lfloor N/K\rfloor$. For $k=\ell M, 1\le \ell\le K$ we write

$$\sum_{i=1}^{k}\xi_{i,j} = \sum_{v=1}^{\ell}\Big(\sum_{i=(v-1)M+1}^{vM}\xi_{i,j}\Big).$$

Using the $\gamma_{i,j}$'s, the independent standard normal random variables constructed in Lemma A.2, we define

$$W_j(k) = \sum_{i=1}^{k} \gamma_{i,j}, \quad 1 \le j \le d,\ 1 \le k \le N. \tag{A.9}$$

By Lemma A.3 we get for any $0 < \delta < 1/2$ and $1 \le j \le d$ via Kolmogorov's inequality (cf. Petrov (1995)), p. 54)

$$\begin{aligned}
&P\left\{ \max_{1\le \ell \le K} \left| \sum_{i=1}^{\ell M} \xi_{i,j} - W_j(\ell M) \right| \ge N^{1/2-\delta} \right\} \qquad\qquad (A.10)\\
&\quad = P\left\{ \max_{1\le \ell \le K} \left| \sum_{v=1}^{\ell} \left(\sum_{i=(v-1)M+1}^{vM} (\xi_{i,j} - \gamma_{i,j}) \right) \right| \ge N^{1/2-\delta} \right\}\\
&\quad \le \frac{1}{N^{1-2\delta}} \sum_{v=1}^{K} E\left(\sum_{i=(v-1)M+1}^{vM} (\xi_{i,j} - \gamma_{i,j}) \right)^2\\
&\quad \le \frac{c_3}{N^{1-2\delta}} K M^{23/24} \frac{1}{\lambda_j} \left(d^{1/4} \left(\sum_{\ell=1}^{d} 1/\lambda_\ell \right)^{3/8} \right)^{1/3}\\
&\quad \le c_3 N^{2\delta - 1/24} K^{1/24} \frac{1}{\lambda_j} \left(d^{1/4} \left(\sum_{\ell=1}^{d} 1/\lambda_\ell \right)^{3/8} \right)^{1/3}.
\end{aligned}$$

One can define independent Wiener processes (standard Brownian motions) $W_j(x), x \ge 0, 1 \le j \le d$ such that (A.9) holds. We obtained approximations for the partial sums of the $\xi_{i,j}$'s at the points $k = \ell M, 1 \le \ell \le K$. Next we show that neither the partial sums of the $\xi_{i,j}$'s nor the Wiener processes $W_j(x)$ can oscillate too much between ℓM and $(\ell+1)M$. Using again Rosenthal's inequality (cf. Petrov (1995), p. 59) we obtain for all $1 \le j \le d$ that

$$\begin{aligned}
E\left| \sum_{i=1}^{M} \xi_{i,j} \right|^3 &\le c_{10} \left\{ \sum_{i=1}^{M} E|\xi_{i,j}|^3 + \left(\sum_{i=1}^{M} E\xi_{i,j}^2 \right)^{3/2} \right\} \qquad (A.11)\\
&\le c_{11} \{ M/\lambda_j^{3/2} + M^{3/2} \}\\
&\le c_{11} (1 + \lambda_1^{3/2}) (M/\lambda_j)^{3/2}
\end{aligned}$$

on account of Lemma A.1. Combining the Marcinkiewicz–Zygmund inequality (cf. Petrov (1995), p. 82) with (A.11) we conclude

$$E\left(\max_{1 \le h \le M} \left| \sum_{i=1}^{h} \xi_{i,j} \right| \right)^3 \le c_{12} (M/\lambda_j)^{3/2}. \tag{A.12}$$

Applying (A.12) we get

$$P\left\{\max_{0\le\ell\le K+1}\max_{1\le h\le M}\left|\sum_{i=1}^{\ell M}\xi_{i,j}-\sum_{i=1}^{\ell M+h}\xi_{i,j}\right|\ge N^{1/2-\delta}\right\} \tag{A.13}$$

$$\le (K+2)P\left\{\max_{1\le h\le M}\left|\sum_{i=1}^{h}\xi_{i,j}\right|> N^{1/2-\delta}\right\}$$

$$\le \frac{c_{13}}{N^{3/2-3\delta}}K(M/\lambda_j)^{3/2}$$

$$\le c_{13}N^{3\delta}K^{-1/2}\lambda_j^{-3/2}.$$

Lemma 1.2.1 of Csörgő and Révész (1981) yields

$$P\left\{\max_{0\le\ell\le K}\sup_{|h|\le M}|W_j(\ell M)-W_j(\ell M+h)|\ge c_{14}M^{1/2}(\log N)^{1/2}\right\}\le\frac{c_{15}}{N^2}. \tag{A.14}$$

Now choosing $\delta=1/80$ and $K=\lfloor N^{\beta}\rfloor$ with $\beta=1/10$, it follows from (A.10), (A.13) and (A.14) for all $1\le j\le d$ that

$$P\left\{\sup_{0\le y\le N}\left|\sum_{1\le i\le y}\xi_{i,j}-W_j(y)\right|> N^{1/2-\delta}\right\} \tag{A.15}$$

$$\le c_{15}N^{-\delta}\left\{\frac{1}{\lambda_j}\left(d^{1/4}\left(\sum_{\ell=1}^{d}1/\lambda_\ell\right)^{3/8}\right)^{1/3}+\frac{1}{\lambda_j^{3/2}}\right\}.$$

The result now follows from (A.15) with $W_{j,N}(x)=N^{-1/2}W_j(Nx), 0\le x\le 1$. □

B Proofs of the results of Section 3.

We first investigate the weak convergence of the process

$$Z_N(u,x)=\frac{1}{d^{1/2}}\sum_{j=1}^{\lfloor du\rfloor}\left\{(S_{j,N}(x)-xS_{j,N}(1))^2-x(1-x)\right\},\quad 0\le u,x\le 1,$$

with $S_{j,N}(x)$ given by (2.4). The difference between $\hat{Z}_N(u,x)$ and $Z_N(u,x)$ is that $\hat{Z}_N$ is computed from the empirical projections $\hat{v}_1,\dots,\hat{v}_d$, while Z_N is based on the unknown population eigenfunctions $v_1,\dots,v_d$.

Theorem B.1. *If Assumptions 2.1, 2.3, 2.4 and 3.1–3.4 hold, then*

$$Z_N(u,x)\quad\to\quad\Gamma(u,x)\ \ \text{in}\ \ \mathcal{D}[0,1]^2,$$

where the Gaussian process $\Gamma(u,x)$ *is defined in Theorem 3.1.*

To prove Theorem B.1, we need several lemmas and some additional notation.
Let

$$V_{j,N}(x) = S_{j,N}(x) - xS_{j,N}(1) \quad \text{and} \quad B_{j,N}(x) = W_{j,N}(x) - xW_{j,N}(1),$$

where $S_{j,N}$ is defined in (2.4) and the $W_{j,N}$'s are the Wiener processes of Theorem 2.1. It follows from the definition that for each N the processes $B_{j,N}, 1 \le j \le d$, are independent Brownian bridges.

LEMMA B.2. *If Assumptions 2.1, 2.3 and 2.4 hold, then*

$$P\left\{\sup_{0\le x\le 1}\sum_{j=1}^{d}\left|V_{j,N}^2(x) - B_{j,N}^2(x)\right| \ge 20dN^{-1/80}(\log N)^{1/2}\right\}$$
$$\le c_* N^{-1/80}\left\{d^{1/12}\left(\sum_{\ell=1}^{d} 1/\lambda_\ell\right)^{1/8} + \sum_{j=1}^{d} 1/\lambda_j^{3/2}\right\} + c_{**}dN^{-2},$$

where c_* and c_{**} only depend on λ_1 and $E\|Z_1\|^3$.

PROOF: First we write

$$V_{j,N}^2(x) - B_{j,N}^2(x) = (V_{j,N}(x) - B_{j,N}(x))^2 + 2B_{j,N}(x)(V_{j,N}(x) - B_{j,N}(x)).$$

Since the $B_{j,N}$'s are Brownian bridges, the distribution of the supremum functional of the Brownian bridge (cf. Csörgő and Révész (1981)) gives

$$P\left\{\max_{1\le j\le d}\sup_{0\le x\le 1}|B_{j,N}(x)| \ge 4(\log N)^{1/2}\right\} \le c_{**}\frac{d}{N^2},$$

where c_{**} is an absolute constant. Now the result follows immediately from Theorem 2.1. □

Now we prove the weak convergence of the partial sums of the squares of independent Brownian bridges. Let $B_1, B_2, \ldots, B_d$ be independent Brownian bridges.

LEMMA B.3. *As $d \to \infty$, we have that*

$$\frac{1}{d^{1/2}}\sum_{j=1}^{\lfloor du\rfloor}\left(B_j^2(x) - x(1-x)\right) \to \Gamma(u,x) \quad \text{in } \mathcal{D}[0,1]^2,$$

where the Gaussian process $\Gamma(u,x)$ is defined in Theorem 3.1.

Proof: The proof is based on Theorem 2 of Hahn (1978). Let B denote a Brownian bridge and $\theta_1 = \sup_{0\le t\le 1}|B(t)|$. It is clear that $E\theta_1^m < \infty$ for all $m \ge 1$. According to Garsia (1970), there is a random variable θ_2 such that $E\theta_2^m < \infty$ for all $m \ge 1$ and

$$|B(t) - B(s)| \le \theta_2(|t-s|\log(1/|t-s|))^{1/2}, \quad 0 \le t, s \le 1.$$

Let $V(t) = B^2(t) - t(1-t)$. We note

$$|V(t) - V(s)| \le 2\theta_1\theta_2(|t-s|\log(1/|t-s|))^{1/2} + |t-s|.$$

Thus we get

$$E(V(t) - V(s))^2 \le c_{16}|t-s|\log(1/|t-s|) \quad \text{for all } \ 0 \le t, s \le 1 \tag{B.1}$$

and

$$E[(V(t) - V(z))^2(V(z) - V(s))^2] \le c_{17}(|t-s|\log(1/|t-s|))^2 \tag{B.2}$$

for all $0 \le s \le z \le t \le 1$. The estimates in (B.1) and (B.2) yield that the conditions of Theorem 2 of Hahn (1978) are satisfied, completing the proof Lemma B.3. □

Proof of Theorem B.1. It follows immediately from Lemmas B.2 and B.3. □

The transition from Theorem B.1 to Theorem 3.1 is based on the following lemma, in which the norm is the Hilbert–Schmidt norm.

Lemma B.4. *If Assumptions 2.1, 2.2 and 2.3 hold, then*

$$|\lambda_j - \hat{\lambda}_j| \le \|\mathfrak{c} - \hat{\mathfrak{c}}\| \tag{B.3}$$

and

$$\|v_j - \hat{c}_j\hat{v}_j\| \le \frac{2\sqrt{2}}{\zeta_j}\|\mathfrak{c} - \hat{\mathfrak{c}}\|, \tag{B.4}$$

where $\hat{c}_j = \operatorname{sign}(\langle \hat{v}_j, v_j\rangle)$ *are random signs, and* $\zeta_1, \zeta_2, \ldots$ *are defined in Assumption 3.5.*

Proof: Inequality (B.3) can be deduced from the general results presented in Section VI.1 of Gohberg et al. (1990) or in Dunford and Schwartz (1988). These results are presented in a convenient form in Lemma 2.2 in Horváth and Kokoszka (2012). Finally Lemma 2.3 in Horváth and Kokoszka (2012) gives (B.4). □

Proof of Theorem 3.1. Introducing

$$U_N(x) = U_N(x,t) = \frac{1}{N^{1/2}}\left\{\sum_{i=1}^{\lfloor Nx \rfloor} Z_i(t) - x\sum_{i=1}^{N} Z_i(t)\right\}$$

we can write

$$\hat{Z}_N(u,x) = \frac{1}{d^{1/2}}\sum_{j=1}^{\lfloor du \rfloor}\left\{\frac{1}{\hat{\lambda}_j}\langle U_N(x),\hat{v}_j\rangle^2 - x(1-x)\right\}.$$

Elementary arguments give

$$\sum_{j=1}^{\lfloor du \rfloor}\frac{1}{\hat{\lambda}_j}\langle U_N(x),\hat{v}_j\rangle^2 = \sum_{j=1}^{\lfloor du \rfloor}\frac{1}{\lambda_j}\langle U_N(x),\hat{c}_j v_j\rangle^2 + \sum_{j=1}^{\lfloor du \rfloor}\left\{\frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j}\right\}\langle U_N(x),\hat{v}_j\rangle^2$$
$$+\sum_{j=1}^{\lfloor du \rfloor}\frac{1}{\lambda_j}(\langle U_N(x),\hat{v}_j\rangle^2 - \langle U_N(x),\hat{c}_j v_j\rangle^2).$$

By the Cauchy–Schwarz inequality we have

$$\frac{1}{d^{1/2}}\sum_{j=1}^{d}\left|\frac{1}{\hat{\lambda}_j} - \frac{1}{\lambda_j}\right|\langle U_N(x),\hat{v}_j\rangle^2 \le \|U_N(x)\|^2\frac{1}{d^{1/2}}\sum_{j=1}^{d}\frac{|\lambda_j - \hat{\lambda}_j|}{\hat{\lambda}_j\lambda_j} \tag{B.5}$$

and since $|a^2 - b^2| = |a+b||a-b|$,

$$\frac{1}{d^{1/2}}\sum_{j=1}^{d}\frac{1}{\lambda_j}(\langle U_N(x),\hat{v}_j\rangle^2 - \langle U_N(x) - \hat{c}_j v_j\rangle^2) \le \|U_N(x)\|^2\frac{2}{d^{1/2}}\sum_{j=1}^{d}\frac{1}{\lambda_j}\|\hat{v}_j - \hat{c}_j v_j\|^2. \tag{B.6}$$

It follows from the results of Kuelbs (1973) (for a shorter proof we refer to Theorem 6.3 in Horváth and Kokoszka (2012)) that

$$\sup_{0\le x\le 1}\|U_N(x)\|^2 = O_P(1).$$

Due to Assumption 2.4 we can use a Marcinkiewicz–Zygmund type law of large numbers for sums of independent and identically distributed random functions in Banach spaces (cf., e.g., Woyczynski (1978) or Howell and Taylor (1980)) to conclude

$$\|\mathfrak{c} - \hat{\mathfrak{c}}\| = O_P(N^{-1/3}).$$

Assumption 3.4 gives that $N^{-1/120}/\lambda_d \to 0$ and therefore by Lemma B.4

$$\max_{1\le i\le d}\frac{\lambda_i}{\hat{\lambda}_i} = O_P(1).$$

So by Lemma B.4 and (B.5) we have

$$\begin{aligned}\frac{1}{d^{1/2}}\sum_{j=1}^{d}\left|\frac{1}{\hat\lambda_j}-\frac{1}{\lambda_j}\right|\langle U_N(x),\hat v_j\rangle^2 &= O_P(1)\frac{1}{d^{1/2}N^{1/3}}\sum_{i=1}^{d}1/\lambda_i^2\\ &= O_P(1)\frac{d^{1/2}}{N^{1/3}}\frac{1}{\lambda_d^2}\\ &= O_P(1)\frac{N^{1/80}}{N^{1/3}}N^{1/60}\\ &= o_P(1)\end{aligned}$$

on account of Assumptions 3.2 and 3.4. Similarly, (B.6) and Assumption 3.5 yield

$$\frac{1}{d^{1/2}}\sum_{j=1}^{d}\frac{1}{\lambda_j}\langle U_N(x),\hat v_j-\hat c_j v_j\rangle^2 = O_P(1)\frac{1}{d^{1/2}N^{1/3}}\sum_{j=1}^{d}\frac{1}{\lambda_j\zeta_j}=o_P(1). \tag{B.7}$$

Theorem 3.1 now follows from Theorem B.1. □

Proof of Corollary 3.1. By Lemma B.2 and (B.7), relation (3.2) is proven if we show that

$$\frac{1}{d^{1/2}\sigma_0}\left\{\sum_{i=1}^{d}\sup_{0\le x\le 1}B_i^2(x)-d\kappa_0\right\}\ \stackrel{\mathcal{D}}{\to}\ N(0,1), \tag{B.8}$$

where $B_1, B_2, \ldots, B_d$ are independent Brownian bridges. Clearly, (B.8) is an immediate consequence of the central limit theorem. Similarly, to establish (3.3), we need to show only that

$$\frac{1}{(d/45)^{1/2}}\left\{\sum_{i=1}^{d}\int B_i^2(x)dx-\frac{d}{6}\right\}\ \stackrel{\mathcal{D}}{\to}\ N(0,1).$$

The above result is known, see Remark 2.1 in Aue et al. (2009). The same argument can be used to prove (3.4). □

C PROOFS OF THE RESULTS OF SECTION 4.

We note that under the null hypothesis $\bar X_N-\bar Y_M=\bar Z_N-\bar Q_M$. Define

$$F_{N,M}=\sum_{j=1}^{N}Z_j-\frac{N}{M}\sum_{j=1}^{M}Q_j.$$

The proof of Theorem 4.1 is based on Lemma A.2, we need to write $F_{N,M}$ as a single sum of independent identically distributed random processes and an additional small remainder term. Let K be an integer and define the integers $R = \lfloor N/K \rfloor$ and $L = \lfloor M/K \rfloor$. Next we define

$$A_i = \sum_{\ell=R(i-1)+1}^{iR} Z_\ell - \sum_{\ell=L(i-1)+1}^{iL} \frac{N}{M} Q_\ell, \quad i = 1, 2, \ldots, K.$$

Clearly,

$$F_{N,M} = \sum_{i=1}^{K} A_i + \tilde{A},$$

where

$$\tilde{A} = \sum_{\ell=KR+1}^{N} Z_\ell - \frac{N}{M} \sum_{\ell=KL+1}^{M} Q_\ell.$$

We will show first if v is a function with $\|v\| = 1$, then for every n

$$E \left| \sum_{\ell=1}^{n} \langle Z_\ell, v \rangle \right|^3 \leq c_1 n^{3/2} \tag{C.1}$$

and

$$E \left| \sum_{\ell=1}^{n} \langle Q_\ell, v \rangle \right|^3 \leq c_2 n^{3/2}, \tag{C.2}$$

where c_1 and c_2 only depends on $E\|Z_1\|^3$ and $E\|Q_1\|^3$, respectively. Using Rosenthal's inequality (cf. Petrov (1995), p. 59) we get

$$E \left| \sum_{\ell=1}^{n} \langle Z_\ell, v \rangle \right|^3 \leq c_3 \left\{ nE|\langle Z_1, v \rangle|^3 + (nE\langle Z_1, v \rangle^2)^{3/2} \right\},$$

where c_3 is an absolute constant. It is easy to see that

$$|\langle Z_1, v \rangle| \leq \|Z_1\|,$$

which implies (C.1). The same argument can be used to prove (C.2).
Next we define the function

$$\mathfrak{c}_{N,M}(t,s) = \mathfrak{c}(t,s) + \frac{N^2 L}{M^2 R} \mathfrak{c}_*(t,s).$$

It is clear that $\mathfrak{c}_{N,M}$ is a covariance function and therefore we can find $\bar{\kappa}_1 = \bar{\kappa}_1(N,M) \geq \bar{\kappa}_2 = \bar{\kappa}_2(N,M) \geq \ldots$ and orthonormal functions $\bar{u}_1(t) = \bar{u}_1(N,M), \bar{u}_2(t) = \bar{u}_2(N,M), \ldots$ satisfying

$$\bar{\kappa}_i \bar{u}_i(t) = \int \mathfrak{c}_{N,M}(t,s) \bar{u}_i(s) ds, \quad 1 \leq i < \infty.$$

Now we define the vector

$$\boldsymbol{\psi}_i = (\langle A_i, \bar{u}_1\rangle/(R\bar{\kappa}_1)^{1/2}, \langle A_i, \bar{u}_2\rangle/(R\bar{\kappa}_2)^{1/2}, \ldots, \langle A_i, \bar{u}_d\rangle/(R\bar{\kappa}_d)^{1/2})^T, \quad 1 \le i \le K.$$

It is easy to see that $\boldsymbol{\psi}_i$, $1 \le i \le K$, are independent and identically distributed random vectors with mean $\mathbf{0}$ and $E\boldsymbol{\psi}_1\boldsymbol{\psi}_1^T = \mathbf{I}_d$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix. Also, (C.1) and (C.2) imply that

$$E|\boldsymbol{\psi}_1| \le c_4 \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/2},$$

where c_4 only depends on $E\|Z_1\|^3$ and $E\|Q_1\|^3$. Using Lemma A.2 we obtain similarly to (A.8) that there are independent standard normal random vectors $\boldsymbol{\gamma}_i = \boldsymbol{\gamma}_i(N, M), 1 \le i \le K$, in R^d such that

$$P\left\{\left|\sum_{i=1}^{K} \boldsymbol{\psi}_i - \sum_{i=1}^{K} \boldsymbol{\gamma}_i\right| \ge c_5 K^{3/8} d^{1/4}\left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/8}\right\} \tag{C.3}$$

$$\le c_5 K^{-1/8} d^{1/4} \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/8},$$

where c_5 does not depend on d. Let

$$\tilde{\boldsymbol{\psi}} = (\langle \tilde{A}, \bar{u}_1\rangle/\sqrt{\bar{\kappa}_1}, \langle \tilde{A}, \bar{u}_2\rangle/\sqrt{\bar{\kappa}_2}, \ldots, \langle \tilde{A}, \bar{u}_d\rangle/\sqrt{\bar{\kappa}_d})^T.$$

It follows from (C.1) and (C.2) that with some constant c_6, not depending on d we have

$$E|\tilde{\boldsymbol{\psi}}|^3 \le c_6 K^{3/2} \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/2}$$

and therefore by Markov's inequality for every $x > 0$

$$P\left\{N^{-1/2}|\tilde{\boldsymbol{\psi}}| > x\right\} \le c_7 \frac{K^{3/2}}{x^3 N^{3/2}} \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/2}. \tag{C.4}$$

Let

$$\boldsymbol{\kappa}_{N,M} = (\langle F_{N,M}, \bar{u}_1\rangle/\sqrt{\bar{\kappa}_1}, \langle F_{N,M}, \bar{u}_2\rangle/\sqrt{\bar{\kappa}_2}, \ldots, \langle F_{N,M}, \bar{u}_d\rangle/\sqrt{\bar{\kappa}_d})^T.$$

Next we choose $K = \lfloor N^{3/4} \rfloor$ in (C.3), (C.4) and $x = K^{-1/8}(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell)^{3/8}$ in (C.4) to conclude that there is $\boldsymbol{\gamma}_{N,M}$, a standard normal random vector in R^d such that

$$P\left\{\left|\frac{1}{\sqrt{N^*}}\boldsymbol{\kappa}_{N,M} - \boldsymbol{\gamma}_{N,M}\right| \ge c_8 N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell\right)^{3/8}\right\} \tag{C.5}$$

$$\leq c_8 N^{-3/32} d^{1/4} \left(\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell \right)^{3/8},$$

where $N^* = \lfloor N/\lfloor N^{3/4} \rfloor \rfloor \lfloor N^{3/4} \rfloor$. Using the definitions of $\mathfrak{c}_P$ and $\mathfrak{c}_{N,M}$, together with Assumption 4.3, we conclude

$$\|\mathfrak{c}_P - \mathfrak{c}_{N,M}\| = O(N^{-1/4}), \tag{C.6}$$

so by Lemma 2.3 of Horváth and Kokoszka (2012), cf. Lemma B.4, we have

$$|\kappa_i - \bar{\kappa}_i| \leq c_9 \|\mathfrak{c}_P - \mathfrak{c}_{N,M}\| = O(N^{-1/4}). \tag{C.7}$$

Using Assumption 4.5 we conclude that

$$\sum_{\ell=1}^{d} 1/\bar{\kappa}_\ell = O\left(\sum_{\ell=1}^{d} 1/\kappa_\ell \right).$$

Hence it follows from (C.5) and Assumption 4.5 that

$$\frac{1}{N} |\boldsymbol{\kappa}_{N,M}|^2 - \frac{N^*}{N} |\boldsymbol{\gamma}_{N,M}|^2 = o_P(d^{1/2}).$$

Since $|\boldsymbol{\gamma}_{N,M}|^2$ is a χ^2 random variable with d degrees of freedom, Assumption 4.5 yields that

$$\left| \frac{N^*}{N} - 1 \right| |\boldsymbol{\gamma}_{N,M}|^2 = o_P(d^{1/2}).$$

It is well known that $(|\boldsymbol{\gamma}_{N,M}|^2 - d)/(2d)^{1/2}$ converges in distribution to a standard normal random variable, and therefore

$$\frac{1}{\sqrt{2d}} \left\{ \frac{1}{N} |\boldsymbol{\kappa}_{N,M}|^2 - d \right\} \stackrel{\mathcal{D}}{\to} N(0,1),$$

where $N(0,1)$ stands for a standard normal random variable.

The difference between $|\boldsymbol{\kappa}_{N,M}|^2/N$ and $\widehat{D}_{N,M}$ is that the projections are done into the direction of different functions ($\bar{u}_i$'s and $\hat{u}_i$'s, respectively) and the normalizations ($\bar{\kappa}_i$'s and $\hat{\kappa}_i$'s, respectively) are also different. However, using the Marcinkiewicz–Zygmund law of large numbers in a Banach space together with (C.6) and Assumption 4.5, we obtain that

$$\|\hat{\mathfrak{c}}_P - \mathfrak{c}_{N,M}\| = O_P(N^{-1/4}).$$

Hence, in view of (C.7), also

$$\sup_i |\hat{\kappa}_i - \bar{\kappa}_i| = O_P(N^{-1/4}),$$

and there are random signs $\hat{d}_i$ such that

$$\sup_i \left(\sum_{\ell=1}^{i} 1/\iota_\ell\right)^{-1} \|\hat{u}_i - \hat{d}_i \bar{u}_i\| = O_P(N^{-1/4}).$$

So repeating the arguments used in the proof of Theorem 3.1, we get

$$\left|\widehat{D}_{N,M} - \frac{1}{N}|\boldsymbol{\kappa}_{N,M}|^2\right| = o_P(d^{1/2}),$$

completing the proof.

References

J. A. D. Aston and C. Kirch. Estimation of the distribution of change–points with application to fMRI data. *The Annals of Applied Statistics*, 2012. Forthcoming.

A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37:4046–4087, 2009.

M. Basseville, I. V. Nikifirov, and A. Tartakovsky. *Sequential Analysis: Hypothesis Testing and Change–Point Detection*. Chapman & Hall/CRC, 2012.

N. Bathia, Q. Yao, and F. Ziegelmann. Identifying the finite dimensionality of curve time series. *The Annals of Statistics*, 38:3353–3386, 2010.

M. Benko, W. Härdle, and A. Kneip. Common functional principal components. *The Annals of Statistics*, 37:1–34, 2009.

I. Berkes, R. Gabrys, L. Horváth, and P. Kokoszka. Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*, 71:927–946, 2009.

H. Cardot, F. Ferraty, A. Mas, and P. Sarda. Testing hypothesis in the functional linear model. *Scandinavian Journal of Statistics*, 30:241–255, 2003.

J. Chen and A. K. Gupta. *Parametric Statistical Change Point Analysis: With Applications to Genetics, Medicine, and Finance*. Birkhäuser, 2011.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis*. Wiley, New York, 1997.

M. Csörgő and P. Révész. *Strong Approximations in Probability and Statistics*. Academic Press, New York, 1981.

N. Dunford and J. T. Schwartz. *Linear Operators, Parts I and II*. Wiley, 1988.

U. Einmahl. Strong invariance principles for partial sums of independent random vectors. *The Annals of Probability*, 15:1419–1440, 1987.

U. Einmahl. Extension of results of Komlós, Major and Tusnady to the multivariate case. *Journal of Multivariate Analysis*, 28:20–68, 1989.

C. Gallagher, R. Lund, and M. Robbins. Changepoint detection in daily precipitation data. *Environmetrics*, 23:407–419, 2012.

A. M. Garsia. Continuity properties of Gaussian processes with multidimensional time parameter. In *Proceedings of the* 6^{th} *Berkeley Symp. Math. Stat. Probab.*, volume 2, pages 369–374. University of California Press, 1970.

I. Gohberg, S. Golberg, and M. A. Kaashoek. *Classes of Linear Operators*, volume 49 of *Operator Theory: Advances and Applications*. Birkhäuser, 1990.

O. Gromenko, P. Kokoszka, L. Zhu, and J. Sojka. Estimation and testing for spatially indexed curves with application to ionospheric and magnetic field trends. *The Annals of Applied Statistics*, 6:669–696, 2012.

M. G. Hahn. Central limit theorems in $D[0,1]$. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 44:89–101, 1978.

P. Hall and C. Vial. Assessing the finite dimensionality of functional data. *Journal of the Royal Statistical Society (B)*, 68:689–705, 2006.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

L. Horváth, P. Kokoszka, and M. Reimherr. Two sample inference in functional linear models. *Canadian Journal of Statistics*, 37:571–591, 2009.

L. Horváth, P. Kokoszka, and R. Reeder. Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, doi: 10.1111/j.1467-9868.2012.01032.x, 2012.

J. O. Howell and R. L. Taylor. Marcinkiewicz–Zygmund weak laws of large numbers for unconditional random elements in banach spaces. In J. Kuelbs, editor, *Probability in Banach Spaces. III. Proceedings of the Third International Conference held at Tufts University, Medford, Mass.*, pages 219–230. Springer, 1980.

J. Kuelbs. The invariance principle for Banach space valued random variables. *Journal of Multivariate Analysis*, 3:161–172, 1973.

V. M. Panaretos, D. Kraus, and J. H. Maddocks. Second-order comparison of Gaussian random functions and the geometry of DNA minicircles. *Journal of the American Statistical Association*, 105:670–682, 2010.

V. V. Petrov. *Limit Theorems of Probability Theory: Sequences of Independent Random Varianles*. Clarendon Press, 1995.

J. Ramsay, G. Hooker, and S. Graves. *Functional Data Analysis with R and MATLAB*. Springer, 2009.

J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer, 2005.

V. V. Senatov. *Normal Approximation: New Results, Methods and Problems*. VSP, 1998.

V. Strassen. The existence of probability measures with given marginals. *The Annals of Mathematical Statistics*, 36:423–439, 1965.

W. Woyczynski. Geometry and martingales in Banach spaces. II. independent increments. In J. Kuelbs, editor, *Probability on Banach Spaces*, pages 267–517. Marcel Dekker, 1978.

Discussion

The results presented in the preceding articles deal with problems from different areas of change-point analysis and from the related two-sample inference. Aside from the results themselves the intersection and combination of various of these areas and their particular characteristics raise a multitude of interesting questions and bring forward suggestions for further research.

Change-point procedures are presented in an a posteriori setting as well as in a sequential open-end setting. Explicitly univariate and functional data were considered, implicitly multivariate methodology was applied to cope with the functional nature of the data. The model characteristics subject to change (or difference in the two-sample inference) are the mean, the regression parameters of a linear model, the mean function and the covariance structure of the data. This list does not only show the great diversity of change-point analysis it also indicates many yet unexplored research topics in a rapidly growing and developing field.

In this section we want to connect the different concepts and approaches of the articles and discuss in detail the several results with special respect to their impact in the context of the general scope of this thesis. In particular the relevance of change-point procedures in functional models should be highlighted by showing the need to develop statistical methods to investigate and monitor the validity of functional models in the expanding fields of their applications.

The first two articles Fremdt (2012b) and Fremdt (2012a) consider a so-called AMOC (at most one change) model in a sequential open-end setting. In this open-end setting the model parameters are unknown and a training period (where constancy of the model parameters is assumed) is used to estimate the so-called "in control situation". The monitoring then begins with the end of the training period and is assumed to go on until a change is detected. In contrast to a closed-end setting where a fixed monitoring horizon is part of the design of the procedure, in the open-end setting the monitoring goes on infinitely long unless the procedure detects a change.

The two articles focus on the development and assessment of change-point detection procedures based on an idea of Page (1954). The procedure introduced by Page (1954) has been investigated extensively in the context of control charts and statistical process control, mostly using a constant threshold. However the approach of Fremdt (2012b) for this detector uses a threshold function that guarantees the control of the error of first kind, the "false alarm rate", asymptotically and is therefore novel. The approach entails a number of interesting features that can be transferred to related problem settings or related approaches for the construction of sequential change-point procedures.

The detectors introduced in Fremdt (2012b) are built from the residuals of a linear model using cumulative sums of the residuals as well as cumulative sums of the squares of the residuals. The main steps to show the validity of the procedures are then approximations for the sum (sum of squares, resp.) of residuals with the sum (sum of squares) of the error

terms and invariance principles for these. In principle these two approximations are the key steps to translate this approach into the context of other time series models like, e.g., ARMA (autoregressive moving average), where residuals can be calculated similarly. The success of the implementation depending on the rates of convergence in the estimation and in the invariance principle. In the context of the linear model in Fremdt (2012b) the two approaches, residuals versus squared residuals, show different advantages and disadvantages. While especially for moderate changes methods based on residuals perform better, they cannot detect certain orthogonal changes. Procedures based on squared residuals have different asymptotic properties like a different constant in the invariance principle or differences in the drift term in case of a change. However one of their advantages is that due to their quadratic construction even orthogonal changes can be detected. With respect to a transfer of the approach to different model settings the procedures based on squared residuals gain importance since the same argument as in the context of the linear model applies yet with a higher impact. As an example in case of an ARMA-model procedures based on the CUSUM of the residuals do not have asymptotic power one in case of a change in the autoregressive and/or moving average parameters as well as in the variance of the error sequence. They are in this case only suitable to detect changes in the mean of such time series.
However, most importantly it can be noted that the arguments used in the proofs in Fremdt (2012b) can be applied to adapt the approach to these important types of time series models.

The basic idea behind the given procedure is to compare the parameter estimates calculated from the data of the training period to parameter estimates calculated from the data of the monitoring period. Since only the data after the occurrence of a change will cause the difference of estimators to increase, but the time of change is unknown, estimates are calculated from the data of all subintervals of the monitoring period with the present as right end point. The maximum value of these differences is then taken as detector.
Many statistical procedures are based on this principle of comparing parameter estimates. In particular in sequential change-point analysis such procedures have been suggested in different contexts, mostly using all observations of the monitoring period to calculate the reference estimate. For an example in the very similar setting of a linear model we refer to Hušková and Koubková (2005) who propose a detector calculated as a quadratic form of partial sums of weighted residuals which can be represented as quadratic form of the differences of the estimators. This approach can be generalized to multivariate and functional problem settings and can therefore be used as a starting point to construct sequential change-point methods with similar properties to those introduced in Fremdt (2012b). Of course an implementation of such procedures depends strongly on the derivation of the behaviour of the detector whithout the occurrence of a structural break which is not a trivial matter as can be seen in the proofs of the corresponding results in Fremdt (2012b).

Further interesting aspects that motivate future investigations are, e.g., the restriction of

the presented procedures to a finite monitoring horizon, i.e., a closed-end setting. This would simplify the proof of the results on the behaviour under the null hypothesis of no change and allow for different choices of the threshold function. In general an adaptation of the weights in the procedures could be considered that takes the length of the subinterval of the monitoring procedure into account.
In the scope of the article the ordinary CUSUM procedure of Horváth et al. (2004) was used as a reference for a comparison of the approaches. An extension of this comparison to other approaches like moving sum detectors or exponentially weighted moving averages would be of interest. For a discussion of these approaches we refer to Kühn (2007). However conclusions from such comparisons have to be handled with care since the performance of these procedures depends strongly on the underlying model specification.
Another important aspect to be discussed in future research is the small sample behaviour of the procedures particularly of those based on squared residuals. While the procedures based on the cumulative sum of residuals show a satisfactory behaviour under the null hypothesis in small samples this is not the case for the procedures based on squared residuals. A possible approach for an improvement in this respect are resampling methods like permutation tests or bootstrap. We refer to Kirch (2006) and her subsequent work for a thorough discussion of this topic.

The result on the asymptotic distribution of the delay time in Fremdt (2012a) provides important information on the dynamics driving the procedure introduced in Fremdt (2012b) under the alternative in the special case of the so-called location model. While often criteria like average run length or (stochastic) bounds for the delay time are used to assess change-point procedures the asymptotic distribution of the delay time provides far more information than these criteria. E.g., upper bounds are not suitable to compare two procedures like ordinary CUSUM (cf. Aue and Horváth (2004)) and Page CUSUM since these have similar worst case behaviour and consequently similar upper bounds. Yet the distributions of the delay times differ when a change occurs not too early.
The result in Fremdt (2012a) is extending the result of Aue and Horváth (2004) for an ordinary CUSUM procedure in various aspects. The assumption on the size of the change could be relaxed from local alternatives (where the size of the change tends to zero as the length of the training period increases) to the assumption of boundedness of the size of change. This assumption includes not only the local alternatives but also fixed alternatives. However the more important extension regards the range for the change-point for which the asymptotic distribution can be derived. This range could be extended from an order of $m^{1/4}$ (in the best case) to an order of m (in all cases) for the right end point.
Furthermore it is important to note that the methodology applied in this particular model can again be used as a starting point to derive similar results for Page's CUSUM procedure in different models like, e.g., the linear model of Fremdt (2012b). In more complex models the derivation of such results requires additional considerations since the variation of additional stochastic quantities has to be taken into account in the determination of the

drift introduced by a change in the model. In the case of a linear model to obtain a similar result the variation of the regressors has to be considered resulting in an additional term in the drift. For a related approach for the ordinary CUSUM of residuals we refer to Aue et al. (2009b).
Since the drift in the location model is deterministic many calculations are simplified compared to models where stochastic factors influence this drift. In models like, e.g., the ARMA model the drift only converges in probability to a constant. This leads among other effects to a slower convergence for the distribution of the delay time. However the presented methodology is applicable in these cases as well.
Interesting challenges for future research are the translation of this approach in the context of models like, e.g. the linear model or the ARMA model mentioned above. In addition the adaptation of this result to a closed-end approach are of interest as well as a comparison to methods like MOSUM or EWMA in the light of this theoretic result.

The two latter articles present results from functional data analysis and therefore primarily focus on different problems than the first two articles. In univariate statistical analysis the key steps to the development of procedures like those presented in Fremdt (2012b) have already been performed and there exists a vast theory around them. The field of functional data analysis is comparatively young and consequently the statistical theory is still at a different stage in its development. The functional nature of the data has to be taken into account and a sophisticated methodology is needed to cope with this and allow for the construction of effective statistical tools to analyze functional data sets in practice.
In the introduction the increasing number of applications for functional data analytic methods on account of the technical developments in collection and storage of data has already been discussed. Along with this increasing number and diversity of data sets come newly developed functional models fit to describe the dynamics and dependencies of the data. Among them are many concepts from univariate or multivariate statistics that can be translated into the functional context. For examples like functional linear models (in different forms) or the functional autoregressive model we refer to Horváth and Kokoszka (2012). Yet many other models can be found in the literature, e.g., generalized functional linear models introduced by Müller and Stadtmüller (2005) or a functional ARCH model introduced by Hörmann et al. (2012).
But this variety of models also strengthens the call for statistical methodology to investigate the model's characteristics. The scope of this statistical inference for functional data is like in the univariate and multivariate context also focussed on estimation and testing problems.
As mentioned above many of the introduced models include certain dependence structures. However the first step in the development of statistical procedures is almost always carried out under the assumptions of independence and identical distribution of the data. As this is also the case in Fremdt *et al.* (2012) and Fremdt *et al.* (2012) an extension of the presented methods allowing for certain dependencies in the data is a natural step in further

research.

In Fremdt *et al.* (2012) a two-sample test for the equality of the covariance operator in functional samples is developed. The covariance structure is probably one of the most interesting characteristics of data in general. It has been studied extensively in the multivariate case (in the context of change-point analysis we refer to, e.g., Aue et al. (2009a) and Wied et al. (2012)). Yet for functional data such problems have mainly been investigated under normality assumptions, i.e. for Gaussian processes (for a discussion of this we refer again to Horváth and Kokoszka (2012)). The test presented in Fremdt *et al.* (2012) extends these results by dropping the normality assumption and it is shown empirically that under a violation of this assumption the methods developed for Gaussian data are inadequate.
The applied methodology to construct a suitable test statistic is a good example how exploiting the properties of functional principal components can help to develop statistical tools for functional data. In particular it is shown how multivariate asymptotic results can be used to obtain asymptotic distributions for the test statistic which are easily applicable for a practitioner. In this special case a multivariate central limit theorem is used to obtain a χ^2-distribution for the test statistic and consequently the corresponding critical values are well-known.
However since the choice of the number of principal components is left to the statistician the criticism mentioned in the introduction is justified in this case as well. It should be noted that from a practical point of view even without theoretical justification the rules of thumb proposed for the choice of this number in general deliver acceptable results. Nevertheless a solid theoretical basis for this matter is doubtless of great importance. A starting point for such a theoretical basis is given in Fremdt *et al.* (2012).

The main result in Fremdt *et al.* (2012) is a uniform normal approximation for the partial sum process of the functional principal component scores. This result provides various opportunities for the construction of statistics that show a certain robustness with respect to the choice of the number of principal components. This is illustrated in the context of change-point detection (in an a posteriori setting) in the mean of functional data and in the context of a two-sample problem as well regarding the equality of mean functions. For the change-point problem the Gaussian limit of a two-parameter partial sum process constructed from the scores is presented that can be used to derive statistics directly or from functionals of this process. In particular a Cramér-von-Mises type statistic should be highlighted that uses the full force of the asymptotics of the two-parameter process. In the empirical part of the article it is shown that this procedure behaves nicely in finite samples including applications like the Australian weather data investigated in the scope of the article. The proceeding in the construction of a statistic for the two-sample problem using the uniform normal approximation is illustrated by deriving a statistic from the scores of the differences of the empirical mean functions. As limit distribution of this statistic the standard normal distribution is obtained.
The application of this uniform normal approximation in these two important statistical

settings indicates the impact of the result. It also shows that the result can be used as a basis for the construction of statistical procedures in other areas of functional data analysis.

Finally with respect to future research the different concepts presented in the respective articles should be brought together. The different problems were motivated in the scope of the single articles nevertheless the concepts carry over to the intersection of their frameworks. What is meant by this is that sequential change-point procedures are as well needed in functional frameworks to monitor the validity of the multitude of already existing functional models (and those that will doubtlessly be defined in the future). First work in this direction has been conducted (e.g., cf. Aue et al. (2012) and Aue et al. (2012+)), yet the problems described to motivate the introduction of Page's procedure in the univariate setting translate to the functional change-point problem. It would therefore be desirable to develop similar approaches using Page's idea for detectors based on differences of estimators in a similar fashion to the one described earlier. As functional principal components represent one of the most effective tools in the development of such procedures an incorporation of the results from Fremdt *et al.* (2012) would certainly increase the value of such procedures. However the challenges hidden behind this easily formulated idea should not be underestimated.

A concluding remark should be directed towards the applications presented in the scope of this thesis. They come from economics, biology and climatology and indicate the wide field the methods described and developed here can be applied to. And still these only represent some of the many fields of application for the procedures developed in the field of change-point analysis. They are the motivation to always push forward our research and guarantee that mathematical statistics and probability theory stay a dynamic discipline always keeping interesting challenges at hand.

Supplementary References

A. Aue. *Sequential Change-Point Analysis based on Invariance Principles.* PhD thesis, University of Cologne, 2003.

A. Aue and L. Horváth. Delay time in sequential detection of change. *Statistics & Probability Letters*, 67(3):221–231, 2004.

A. Aue, S. Hörmann, L. Horváth, and M. Reimherr. Break detection in the covariance structure of multivariate time series models. *Annals of Statistics*, 37(6B):4046–4087, 2009a.

A. Aue, L. Horváth, and M.L. Reimherr. Delay times of sequential procedures for multiple time series regression models. *Journal of Econometrics*, 149(2):174–190, 2009b.

A. Aue, S. Hörmann, L. Horváth, M. Hušková, and J.G. Steinebach. Sequential testing for the stability of high frequency portfolio betas. *Econometric Theory*, 28(4):804–837, 2012.

A. Aue, S. Hörmann, L. Horváth, and M. Hušková. Sequential stability tests for functional linear models. *Preprint, University of California Davis, Université Libre de Bruxelles, University of Utah, Charles University in Prague*, 2012+.

I. Berkes, S. Hörmann, and J. Schauer. Split invariance principles for stationary processes. *Annals of Probability*, 39(6):2441–2473, 2011.

M. Csörgő and L. Horváth. *Limit Theorems in Change-Point Analysis.* Wiley, New York, 1997.

S. Fremdt. Asymptotic distribution of the delay time in Page's sequential procedure. *Preprint, University of Cologne*, 2012a.

S. Fremdt. Page's sequential procedure for change-point detection in time series regression. *Preprint, University of Cologne*, 2012b.

S. Fremdt, L. Horváth, P. Kokoszka, and J.G. Steinebach. Testing the equality of covariance operators in functional samples. *Scandinavian Journal of Statistics*, 2012a. doi: 10.1111/j.1467-9469.2012.00796.x.

S. Fremdt, L. Horváth, P. Kokoszka, and J.G. Steinebach. Functional data analysis with increasing number of projections. *Preprint, University of Cologne, University of Utah, Colorado State University*, 2012b.

S. Hörmann, L. Horváth, and R. Reeder. A functional version of the ARCH model. *Econometric Theory*, 2012. doi: 10.1017/S0266466612000345.

L. Horváth and P. Kokoszka. *Inference for Functional Data with Applications*. Springer, 2012.

L. Horváth, M. Hušková, P. Kokoszka, and J. Steinebach. Monitoring changes in linear models. *Journal of Statistical Planning and Inference*, 126(1):225–251, 2004.

M. Hušková and A. Koubková. Monitoring Jump Changes in Linear Models. *Journal of Statistical Research*, 39(2):51–70, 2005.

C. Kirch. *Resampling Methods for the Change Analysis of Dependent Data*. PhD thesis, University of Cologne, 2006.

A. Koubková. *Sequential Change-Point Analysis*. PhD thesis, Charles University in Prague, 2006.

M. Kühn. *Sequential Change-Point Analysis based on Weighted Moving Averages*. PhD thesis, University of Cologne, 2007.

H.-G. Müller and U. Stadtmüller. Generalized functional linear models. *Annals of Statistics*, 33(2):774–805, 2005.

E. S. Page. Continuous Inspection Schemes. *Biometrika*, 41(1/2):100–115, 1954.

J.O. Ramsay and B.W. Silverman. *Functional data analysis*. Springer, 2005.

J.O. Ramsay, G. Hooker, and S Graves. *Functional data analysis with R and MATLAB*. Springer, 2009.

A. Schmitz. *Limit Theorems in Change-Point Analysis for Dependent Data*. PhD thesis, University of Cologne, 2011.

A. Černíková, M. Hušková, Z. Prášková, and J. Steinebach. Delay time in monitoring jump changes in linear models. *Statistics*, 2011. doi: 10.1080/02331888.2011.577895.

D. Wied, W. Krämer, and H. Dehling. Testing for a change in correlation at an unknown point in time using an extended functional delta method. *Econometric Theory*, 28(3): 570–589, 2012.

W.B. Wu. Strong invariance principles for dependent random variables. *Annals of Probability*, 35(6):2294–2320, 2007.

Erklärung zum Eigenanteil

Zum Eigenanteil an den Artikeln dieser Dissertation mit mehreren Autoren ist anzumerken, dass in der Mathematik eine Rangfolge der Autoren unüblich ist. Für eine Angabe des Eigenanteils in Prozent gilt das Gleiche, da dieser durch die enormen Unterschiede in den für das Zustandekommen der Publikation notwendigen Arbeiten (z.B. Entwicklung der mathematischen Theorie, Durchführung von Simulationsstudien, Auswertung von Daten) nur sehr schwer zu beziffern ist. Der Eigenanteil wird deshalb hier separat für den theoretischen sowie den empirischen Anteil der Artikel angegeben.
Bei den beiden in dieser Dissertation enthaltenen Artikeln, die gemeinsam mit Prof. Dr. Horváth, Prof. Dr. Kokoszka und Prof. Dr. Steinebach erstellt wurden, sind die empirischen Teile der Artikel, d.h. die Implementierung und Durchführung der Simulationsstudie und die Bearbeitung der anwendungsbezogenen Daten, vollständig durch meine Leistung und lediglich unter Rücksprache mit Prof. Dr. Kokoszka entstanden. Beim theoretischen Teil der beiden Artikel liegt der Eigenanteil bei ca. 25%, wobei dieser in *Functional data analysis with increasing number of projections* im Vergleich etwas höher anzusiedeln ist.

Köln, im Oktober 2012

(Stefan Fremdt)

Erklärung

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit – einschließlich Tabellen, Karten und Abbildungen –, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie – abgesehen von unten angegebenen Teilpublikationen – noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. Dr. Josef G. Steinebach betreut worden.

Köln, im Oktober 2012

(Stefan Fremdt)

Teilpublikationen

Page's sequential procedure for change-point detection in time series regression.
Econometric Theory (2012), eingereicht (unter Revision).

Asymptotic distribution of the delay time in Page's sequential procedure.
J. Statist. Plann. Inference (2012), eingereicht.

Testing the equality of covariance operators in functional samples.
Scand. J. Stat. (2012). (mit L. Horváth, P. Kokoszka und J.G. Steinebach).
doi:10.1111/j.1467-9469.2012.00796.x.

Functional data analysis with increasing number of projections.
Ann. Statist. (2012), eingereicht. (mit L. Horváth, P. Kokoszka und J.G. Steinebach).