Goodness-of-Fit and Change-Point Tests for Functional Data

Robertas Gabrys
Utah State University

Follow this and additional works at: https://digitalcommons.usu.edu/etd

Recommended Citation
https://digitalcommons.usu.edu/etd/658

This Dissertation is brought to you for free and open access by the Graduate Studies at DigitalCommons@USU. It has been accepted for inclusion in All Graduate Theses and Dissertations by an authorized administrator of DigitalCommons@USU. For more information, please contact digitalcommons@usu.edu.
GOODNESS-OF-FIT AND CHANGE-POINT TESTS FOR FUNCTIONAL DATA

by

Robertas Gabrys

A dissertation submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Mathematical Sciences

Approved:

Dr. Piotr S. Kokoszka
Major Professor

Dr. Daniel C. Coster
Committee Member

Dr. Richard D. Cutler
Committee Member

Dr. John R. Stevens
Committee Member

Dr. Lie Zhu
Committee Member

Dr. Byron R. Burnham
Dean of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2010
ABSTRACT

Goodness-of-Fit and Change-Point Tests for Functional Data

by

Robertas Gabrys, Doctor of Philosophy
Utah State University, 2010

Major Professor: Dr. Piotr S. Kokoszka
Department: Mathematics and Statistics

A test for independence and identical distribution of functional observations is proposed in this thesis. To reduce dimension, curves are projected on the most important functional principal components. Then a test statistic based on lagged cross-covariances of the resulting vectors is constructed. We show that this dimension reduction step introduces asymptotically negligible terms, i.e. the projections behave asymptotically as iid vector-valued observations. A complete asymptotic theory based on correlations of random matrices, functional principal component expansions, and Hilbert space techniques is developed. The test statistic has \( \chi^2 \) asymptotic null distribution.

Two inferential tests for error correlation in the functional linear model are put forward. To construct them, finite dimensional residuals are computed in two different ways, and then their autocorrelations are suitably defined. From these autocorrelation matrices, two quadratic forms are constructed whose limiting distributions are chi-squared with known numbers of degrees of freedom (different for the two forms).
A test for detecting a change point in the mean of functional observations is developed. The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution. A comprehensive asymptotic theory for the estimation of a change–point in the mean function of functional observations is developed.

The procedures developed in this thesis can be readily computed using the R package \texttt{fda}. All theoretical insights obtained in this thesis are confirmed by simulations and illustrated by real life-data examples.
This work is dedicated to my dear mother, Aldona, and my beloved grandparents, Kostancija and Vaclovas.
ACKNOWLEDGMENTS

I would first like to thank my family, my mother Aldona and my grandparents Kostancija and Vaclavas. Your love and invaluable support has always kept me on the crest of the wave. Thank you for believing in me.

My sincerest thanks are due to my advisor, Dr. Piotr Kokoszka, for setting a fine example of academic excellence, sound scholarship, and supportive guidance. Your gift of time, encouragement, inexhaustible enthusiasm, and wise advice kept me constantly motivated. I know I would not have made it this far without you.

To the co-authors of my papers: Lajos Horvath (University of Utah), Alex Aue (University of California at Davis), Istvan Berkes (Graz University of Technology). Thank you for your contributions and experience. I have learned so much from all of you. I look forward to working with you in the future.

Many thanks to my committee members, Dr. Daniel C. Coster, Dr. Richard D. Cutler, Dr. John R. Stevens, and Dr. Lee Zhu, for their agreement to work with me and sit on my committee. All of you have taken the time out of your very busy schedules to answer my questions, and provide your wise advice.

Dr. Adele Cutler, thank you for your help and support. You always made yourself available to answer my questions, discuss problems, and give valuable advice.

To the department of Mathematics and Computer Science at Vilnius University: thank you for the invaluable knowledge and skills I obtained while pursuing my undergraduate and master’s degrees. Dr. Alfredas Račkauskas, thank you for your comprehensive support. Dr. Remigijus Leipus, thank you for your encouragement and enthusiasm. The firm foundation provided by Vilnius University played a crucial role in my success as a Ph.D. student.
I would also like to thank my friends for their support, encouragement, enthusiasm, insights, and for making this time in my life a little less stressful. But most of all for their friendship. They are Vilma Januleviciute, Dr. Vaidotas Zemlys, Dr. Linas Zaleskis, Dr. Aldona Skucaite, Dr. Agnieszka Jach, Lewis Weight, Dr. Inga Maslova, Dr. Andrea Bruder, Daniel Useche, Aneta N. Skoberla, Dr. Aleksandra Mikosz, Abbass al Shariff, Mohamad El Hamoui, Magathi Jayaram, and Brian Abel.

Finally, I would like to acknowledge Aldona Januleviciene, the person who showed me that mathematics is really fascinating and interesting. You used cut apples and other visual techniques to teach a confused 7\textsuperscript{th}-grade high school pupil how to add fractions (5\textsuperscript{th}-grade material). Mathematics seemed to me like a dark forest I was too intimidated to enter. Thank you for being the light that guided me through that forest.

Robertas Gabrys
## CONTENTS

<table>
<thead>
<tr>
<th>ABSTRACT</th>
<th>iii</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENTS</td>
<td>vi</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>x</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>xii</td>
</tr>
<tr>
<td>1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>2 PORTMANTEAU TEST OF INDEPENDENCE FOR FUNCTIONAL OBSERVATIONS</td>
<td>11</td>
</tr>
<tr>
<td>2.1 Introduction</td>
<td>11</td>
</tr>
<tr>
<td>2.2 Assumptions and main results</td>
<td>13</td>
</tr>
<tr>
<td>2.3 Finite sample performance</td>
<td>19</td>
</tr>
<tr>
<td>2.4 Application to credit card transactions and diurnal geomagnetic variation</td>
<td>23</td>
</tr>
<tr>
<td>2.5 Proofs of Theorems 2.1 and 2.2</td>
<td>28</td>
</tr>
<tr>
<td>2.6 Auxiliary Lemmas for H-valued random elements</td>
<td>33</td>
</tr>
<tr>
<td>2.7 Limit theory for sample autocovariance matrices</td>
<td>34</td>
</tr>
<tr>
<td>3 TESTS FOR ERROR CORRELATION IN THE FUNCTIONAL LINEAR MODEL</td>
<td>39</td>
</tr>
<tr>
<td>3.1 Introduction</td>
<td>39</td>
</tr>
<tr>
<td>3.2 Preliminaries</td>
<td>43</td>
</tr>
<tr>
<td>3.3 Least squares estimation</td>
<td>47</td>
</tr>
<tr>
<td>3.4 Testing the independence of model errors</td>
<td>51</td>
</tr>
<tr>
<td>3.5 Asymptotic theory</td>
<td>56</td>
</tr>
<tr>
<td>3.6 A simulation study</td>
<td>58</td>
</tr>
<tr>
<td>3.7 Application to space physics and high-frequency financial data</td>
<td>61</td>
</tr>
<tr>
<td>3.8 Proof of Theorem 3.3</td>
<td>84</td>
</tr>
<tr>
<td>3.9 Proof of Theorem 3.4</td>
<td>98</td>
</tr>
<tr>
<td>3.10 Proofs of Theorems 3.5 and 3.6</td>
<td>104</td>
</tr>
<tr>
<td>4 DETECTING CHANGES IN THE MEAN OF FUNCTIONAL OBSERVATIONS</td>
<td>112</td>
</tr>
<tr>
<td>4.1 Introduction</td>
<td>112</td>
</tr>
<tr>
<td>4.2 Notation and assumptions</td>
<td>114</td>
</tr>
<tr>
<td>4.3 Detection procedure</td>
<td>118</td>
</tr>
<tr>
<td>4.4 Finite sample performance and application to temperature data</td>
<td>124</td>
</tr>
</tbody>
</table>
4.4.1 Gaussian processes ........................................ 124
4.4.2 Temperature data .......................................... 126
4.5 Proof of Theorems 4.1 and 4.2 .............................. 135
  4.5.1 Proof of Theorems 4.1 ................................. 138
  4.5.2 Proof of Theorem 4.2 ................................. 140

5 ESTIMATION OF A CHANGE-POINT IN THE MEAN FUNCTION OF FUNCTIONAL DATA ................................. 144
  5.1 Introduction ................................................. 144
  5.2 Change-point estimator and its limit distribution ......... 148
  5.3 Finite sample behavior .................................... 152
  5.4 Proofs ......................................................... 157
    5.4.1 Preliminary calculations ............................... 157
    5.4.2 Proof of Theorem 5.3 ................................. 159
    5.4.3 Proof of Theorem 5.4 ................................. 164
    5.4.4 Verification of equation (5.4.4) ...................... 169
    5.4.5 Verification of equation (5.4.7) ...................... 171

6 SUMMARY AND CONCLUSIONS .................................. 174
REFERENCES ....................................................... 176
APPENDICES .......................................................... 185
CURRICULUM VITAE .................................................. 206
# List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 Empirical size (in percent) of the test using Fourier basis. The simulated observations are Brownian bridges.</td>
<td>19</td>
</tr>
<tr>
<td>2.2 Empirical power of the test using Fourier basis. The observations follow the ARH(1) model (2.3.11) with Gaussian kernel with $</td>
<td></td>
</tr>
<tr>
<td>2.3 Empirical power of the test using Fourier basis. The observations follow the ARH(1) model (2.3.11) with Gaussian kernel with $</td>
<td></td>
</tr>
<tr>
<td>2.4 P-values for the functional ARH(1) residuals of the credit card data $X_n$.</td>
<td>24</td>
</tr>
<tr>
<td>2.5 P-values for the magnetometer data split by season.</td>
<td>27</td>
</tr>
<tr>
<td>3.1 Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB$.</td>
<td>62</td>
</tr>
<tr>
<td>3.2 Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB_{t_5}$.</td>
<td>63</td>
</tr>
<tr>
<td>3.3 Method I: Empirical size for independent regressor functions: $X_n = BB_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j\pi t)$, $n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0,1)$.</td>
<td>64</td>
</tr>
<tr>
<td>3.4 Method I: Empirical size for dependent predictors: $X \sim ARH(1)$ with the BB innovations, $\psi =$Gaussian, $\varepsilon = BB$.</td>
<td>65</td>
</tr>
<tr>
<td>3.5 Method II: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB$.</td>
<td>66</td>
</tr>
<tr>
<td>3.6 Method II: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB_{t_5}$.</td>
<td>67</td>
</tr>
<tr>
<td>3.7 Method II: Empirical size for independent regressor functions: $X_n = BB_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j\pi t)$, $n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0,1)$.</td>
<td>68</td>
</tr>
<tr>
<td>3.8 Method II: Empirical size for dependent predictor functions: $X \sim ARH(1)$ with the BB innovations, $\psi =$Gaussian, $\varepsilon = BB$.</td>
<td>69</td>
</tr>
</tbody>
</table>
3.9 Method I: Empirical power for independent predictors: \( X = BB, \varepsilon \sim ARH(1) \) with the BB innovations. ................................. 70

3.10 Method I: Empirical power for independent predictors: \( X = BB, \varepsilon \sim ARH(1) \) with the BB\(_t\) innovations. ................................. 71

3.11 Method I: Empirical power for dependent predictor functions: \( X \sim ARH(1) \) with the BB innovations, \( \varepsilon \sim ARH(1) \) with the BB innovations. 72

3.12 Method II: Empirical power for independent predictors: \( X = BB, \varepsilon \sim ARH(1) \) with the BB innovations. ................................. 73

3.13 Method II: Empirical power for independent predictors: \( X = BB, \varepsilon \sim ARH(1) \) with the BB\(_t\) innovations. ................................. 74

3.14 Method II: Empirical power for dependent predictors: \( X \sim ARH(1) \) with the BB innovations; \( \varepsilon = BB \sim ARH(1) \) with the BB innovations. 75

3.15 Isolated substorms data. P-values in percent. ................................. 78

3.16 P-values, in percent, for the FCAPM (3.7.21) in which the regressors are the intra–daily cumulative returns on the Standard & Poor’s 100 index, and the responses are such returns on the Exxon–Mobil stock. 85

4.1 Simulated critical values of the distribution of \( K_d \). ....................... 125

4.2 Empirical size (in percent) of the test using the B-spline basis. ............ 127

4.3 Empirical power (in percent) of the test using B-spline basis. Change point at \( k^* = [n/2] \). ....................................................... 128

4.4 Segmentation procedure of the data into periods with constant mean function. ................................................... 130

4.5 Summary and comparison of segmentation. Beginning and end of data period in bold. ..................................................... 133

4.6 Empirical size of the test for models derived from the temperature data. 134

4.7 Empirical power of the test for change-point models derived from temperature data (FDA approach). ................................. 136

4.8 Empirical power of the test for change-point models derived from temperature data (MDA approach). ................................. 137

5.1 Summary statistics for the change-point estimator. The change-point processes were generated by combining BB and \( t + BB \) for three different locations of the change-point \( \tau^* \). We used \( d = 2 \) and \( d = 3 \) (in parenthesis). ................................. 154
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.</td>
<td>8</td>
</tr>
<tr>
<td>1.2</td>
<td>Share prices of the Standard &amp; Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM). Dashed lines separate years.</td>
<td>9</td>
</tr>
<tr>
<td>1.3</td>
<td>Intra-daily cumulative returns on 10 consecutive days for the Standard &amp; Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM).</td>
<td>10</td>
</tr>
<tr>
<td>2.1</td>
<td>The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.</td>
<td>14</td>
</tr>
<tr>
<td>2.2</td>
<td>Three weeks of centered time series of ({X_n(t_i)}) derived from credit card transaction data. The vertical dotted lines separate days.</td>
<td>25</td>
</tr>
<tr>
<td>2.3</td>
<td>Two functional observations (X_n) derived from the credit card transactions (left–most panel) together with smooths obtained by projection on 40 and 80 Fourier basis functions.</td>
<td>25</td>
</tr>
<tr>
<td>2.4</td>
<td>Top: horizontal intensity (nT) measured at Honolulu 30/3/2001 - 13/4/2001 with the straight lines connecting first and last measurements in each day. Bottom: the same after subtracting the lines.</td>
<td>27</td>
</tr>
<tr>
<td>3.1</td>
<td>Diagnostic plots of Chiu and Müller (2007) for a synthetic data set simulated according to model (3.1.1) in which the errors (\varepsilon_n) follow the functional autoregressive model of Bosq (2000).</td>
<td>42</td>
</tr>
<tr>
<td>3.2</td>
<td>Magnetometer data on 10 consecutive days (separated by vertical dashed lines) recorded at College, Alaska (CMO) and Honolulu, Hawaii, (HON).</td>
<td>77</td>
</tr>
<tr>
<td>3.3</td>
<td>Magnetometer data on 10 chronologically arranged isolated substorm days recorded at College, Alaska (CMO), Honolulu, Hawaii, (HON) and Boulder, Colorado (BOU).</td>
<td>79</td>
</tr>
<tr>
<td>3.4</td>
<td>Intra-daily cumulative returns on 10 consecutive days for the Standard &amp; Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM).</td>
<td>82</td>
</tr>
</tbody>
</table>
3.5 Share prices of the Standard & Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM). Dashed lines separate years. 83

4.1 Daily temperatures in 1916 with monthly averages and functional object obtained by smoothing with B-splines. 131

4.2 Graph of mean functions of last four segments. 132

5.1 Estimated density of $\|\delta_n\|^2_2 (\hat{k}_n^* - k^*)$ for the process obtained combining $BM$ and $t_{n^{0.05}} + BM$. 155

5.2 Estimated density of $\|\delta_n\|^2_2 (\hat{k}_n^* - k^*)$ for the process obtained combining $BB$ and $\sin(t)^{n^{0.45}} + BB$. 156
CHAPTER 1
INTRODUCTION

Functional data analysis (FDA) is a relatively new direction in statistics that refers to the analysis of data which consist of observed functions or curves recorded at a finite subset of some interval. Due to technological advances in measurement devices and increasing computational power, high frequency data is becoming an important subject of current statistical research. By high frequency data, we understand measurements or observations available on finite grid of points in an interval. One way to take this into account is to consider the data as an observation of the continuous random variable $X(t), t \in T$. A random variable $X = \{X(t), t \in T\}$ takes values in an infinite dimensional space (or function space). An observation of $X$ is referred to as a functional data point.

While FDA and the multivariate data analysis share many common principles, the infinite-dimensional nature of the functional data presents many new challenges that are absent in the traditional multivariate analysis. On the other hand, traditional statistical methods often fail when we deal with functional data. One of the advantages of FDA over classical statistics is that the time points are not required to be equally spaced, they can vary from one subject to another. FDA does not assume that an observation at one time point in the interval is independent of that at another point within the same functional datum. It might be assumed to be independent from one functional datum to another, but not necessarily to be independent of observed values at distinct time points within the same functional datum.

To clarify the notation, consider $X(t) \in L^2[0, 1]$, i.e. we view a random curve
\{X(t), \; t \in T\} as a random element of \(L^2[0, 1]\) equipped with the Borel \(\sigma\)-algebra. Suppose the stochastic process \(\{X(t), \; t \in T\}\) is defined on a common probability space \((\Omega, \mathcal{F}, P)\). For fixed \(t \in T\), \(X(t)\) is a measurable map from \(\Omega\) to \(\mathbb{R}\). For fixed \(\omega \in \Omega\), the function \(X(\omega, \cdot)\) is a sample path of the stochastic process. For fixed \(\omega\) a sample path \(X(\omega, \cdot)\) is an equivalent class of functions that function space, for example \(L^2[0, 1]\) Hilbert function space. Since functions in the space \(L^2\) can be expressed in terms of basis functions generating the space and moreover since the space is separable Hilbert space, each function in the space can be expressed as a countable linear combination of basis functions. Further explanation could be found in Lee (2004). Suppose \(\{\phi_k\}\) is a set of basis functions of \(L^2\), then for fixed \(\omega\) there exist a unique sequence of numbers \(c_1, c_2, \ldots \in l_2\) such that

\[(1.0.1) \quad X(t) = \sum_{k=1}^{\infty} c_k \phi_k(t).\]

In (1.0.1) the stochastic process \(X(t, \omega)\) is decomposed into two parts, namely \(c_k\) and \(\phi_k\). Note that the randomness is reflected only in the coefficients \(c_k = c_k(\omega)\).

The typical FDA tends to include the following steps. Further details could be found in Ramsay (2005).

- First of all, the raw data are collected, cleaned and organized. Functional observations are observed only at discrete sampling points values \(t_j\), and these may or may not be equally spaced. There may be replications of each function referred to as an observation \(X_i\) with different argument values \(t_{ij}\).

- The next step is to convert functional observations to functional form, i.e. the raw data for observation \(X_i\) are used to define a function \(X_i(t)\) that can be evaluated at all values of \(t\) over some interval. In order to do this a functional
basis should be chosen. A basis is a collection of basic functions whose linear combination defines the actual functional observation:

\[ X_i(t) = \sum_{k=1}^{K} c_{ik} \phi_k(t), \quad i = 1, \ldots, N. \]  

When basis functions \( \phi_k(t) \) are specified, then the conversion of the data into functional data objects involves computing and storing the coefficients, \( c_{ik} \) of the expansion (1.0.2) into a coefficient matrix. It should be pointed out that this step usually involves dimension reduction and initial smoothing. Most commonly used bases include:

- the Fourier basis: typically used for stable, periodic data, i.e. data without specific local features;
- the B-spline or Wavelet basis: typically used for non-periodic data;
- the exponential basis, a set of exponential functions, \( e^{\lambda_k t} \), each with a different rate parameter \( \lambda_k \);
- the polynomial basis, consisting of the powers of \( t \): \( 1, t, t^2, \ldots \);
- the power basis, consisting of a sequence of possibly non-integer powers, including negative powers of an argument \( t \) that is usually required to be positive.

- Next a variety of preliminary displays is performed and summary statistics are computed.
• The functions may also need to be registered or aligned in order to have important features found in each curve that occur at roughly the same argument values.

• Exploratory analyses are carried out on the registered data. The main techniques include principal component analysis (PCA), canonical correlation analysis (CCA), and principal differential analysis (PDA).

• Models are constructed for the data. The models may be in the form of a functional linear model, functional time series model or in the form of a differential equation.

• The models are evaluated using graphical checks, calculating various summary statistics and carrying formal tests of significance.

For the last decade functional principal component (FPC) analysis has been the main tool to analyze the variability of functional data. As its classical counterpart, functional PCA is mainly used to identify a few orthogonal functions that most efficiently describe the variability of functional observations. The inherent complexity of functional data analysis, as a distinctly infinite-dimensional and infinite-parameter (or non parametric) problem, means that principal component methods assume a greater importance in FDA than in more traditional, finite dimensional settings. There is often no practical opportunity for estimating, in a meaningful way, the “distribution” of a random function. Both the representation of such a distribution, and the slow convergence rates of estimators, throw up obstacles which seem insurmountable in many cases. Considerations of this type indicate that the properties of the principal component functions are often going to be of greater importance than properties of the distribution itself. For example, it will be of greater interest to assess peaks and
troughs in a principal component function, than to look for extrema in the “density” of the distribution. The principal component functions appear explicitly in the Karhunen-Loève representation of $X(t)$, where they are weighted by scalar random variables. Further details could be found in Hall and Vial (2006).

Functional variable can be a random surface or image, a vector of random curves or any other more complicated infinite dimensional mathematical object. Examples in which the collected data are curves arise in a variety of fields of applied sciences including environmetrics, chemometrics, biometrics, medicine, econometrics, physics, etc. To illustrate the notion of functional data we present a few data sets that motivated the research presented in this thesis. They are extensively analyzed in the dissertation.

The currents flowing in the magnetosphere and ionosphere (M-I) form a complicated multiscale geosystem that contains the temporal scales from seconds to days. A magnetometer is an instrument that measures the three components of the magnetic field at a location where it is placed. There are over 100 magnetic observatories located on the surface of the Earth, and most of them have digital magnetometers. These magnetometers record the strength and direction of the field every five seconds, but the magnetic field exists at any moment of time, so it is natural to think of a magnetogram as an approximation to a continuous record. The raw magnetometer data are cleaned and reported as averages over one minute intervals. Such averages were used to produce Figure 1.1. Thus $7 \times 24 \times 60 = 10,080$ values (of one component of the field) were used to draw Figure 1.1. The dotted vertical lines separate days in Universal Time (UT). It is natural to view a curve defined over one UT day as a single observation because one of the main sources influencing the shape of the record is the daily rotation of the Earth. When an observatory faces the Sun, it records the magnetic field generated by wind currents flowing in the ionosphere which are
driven mostly by solar heating. Thus, Figure 1.1 shows seven consecutive functional observations.

In contrast to the magnetic field, the price of an asset exists only when the asset is traded. A great deal of financial research has been done using the closing daily price, i.e. the price in the last transaction of a trading day. However many assets are traded so frequently that one can practically think of a price curve that is defined at any moment of time. Intra–daily prices of financial instruments are well-known to have properties quite different than those of daily or monthly closing prices and also offer new interesting research perspectives. It is natural to choose one trading day as the underlying time interval. To illustrate, consider the S&P 100 index and its major component, the Exxon Mobil Corporation (currently it contributes 6.78% to this index). The price processes over the period of about 8 years are shown in Figure 1.2. For these data, $P_n(t_j)$ is the price on day $n$ at tick $t_j$ (time of trade). For such data, it is not appropriate to define returns by looking at price movements between the ticks because that would lead to very noisy trajectories for FDA methods, based on the FPC’s, are not appropriate; Johnstone and Lu (2009) explain why principal components cannot be meaningfully estimated for such data. Instead, we adopt the following definition of the intra-daily cumulative returns. Suppose $P_n(t_j)$, $n = 1,\ldots,N$, $j = 1,\ldots,m$, is the price of a financial asset at time $t_j$ on day $n$. We call the functions

$$r_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)], \quad j = 2,\ldots,m, \quad n = 1,\ldots,N,$$

the intra-daily cumulative returns. Figure 1.3 shows intra-daily cumulative returns on 10 consecutive days for the Standard & Poor’s 100 index and the Exxon Mobil Corporation. These returns have now an appearance amenable to smoothing via
FPC’s.

The dissertation consists of several chapters, which represent individual manuscripts on goodness of fit and change point tests for functional observations. Chapter 2 introduces a test for independence and identical distribution of functional observations. Two inferential tests for error correlation in the functional linear model are put forward in Chapter 3. Chapter 4 deals with the novel test for detecting a change point in the mean of functional observations, while Chapter 5 develops and discusses a comprehensive asymptotic theory for the estimation of a change–point in the mean function of functional observations.
Fig. 1.1: The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.
Fig. 1.2: Share prices of the Standard & Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM). Dashed lines separate years.
Fig. 1.3: Intra-daily cumulative returns on 10 consecutive days for the Standard & Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM).
CHAPTER 2
PORTMANTEAU TEST OF INDEPENDENCE FOR FUNCTIONAL OBSERVATIONS

2.1 Introduction

The last two decades have seen the emergence of new technology allowing the collection and storage of time series of finely sampled observations. Every single trade on a speculative asset is recorded and stored, and so e.g. minute by minute values of an asset are available, resulting in 390 observations in a six and half hour trading day rather than one closing price. Similar examples abound. Data of this type can be conveniently viewed as functional observations, e.g. we treat the curve built up of 390 minute by minute values of an asset as a one single observation.

Most inferential tools of Functional Data Analysis, see Ramsay and Silverman (2005), rely on the assumption of iid functional observations. In designed experiments (see e.g. Müller and Stadtmüller (2005), this assumption can be ensured, but in observational data derived from time series it requires a verification. In traditional (nonfunctional) time series analysis, tests of independence and residual–based diagnostic tests derived from them play a fundamental role. In this paper we propose a simple portmanteau test of independence for functional observations. In addition to its direct applicability, this test, with the underlying theory and numerical work presented here, will likely form the basis for residual–based specification tests for various functional time series models. While there are many such tests for real– and vector–valued observations, see Hosking (1980) and Li (1981) among others, no methodology is yet available for functional data.
The idea of the test is simple and intuitively appealing. The functional observations \( X_n(t), \ t \in [0, T], \ n = 1, 2, \ldots, N, \) are approximated by the first \( p \) terms of the principal component expansion

\[
X_n(t) \approx \sum_{k=1}^{p} X_{kn}v_k(t), \quad n = 1, 2, \ldots, N.
\]

The \( v_k(t) \) are the principal modes of variation (principal components, PC’s), and the \( X_{kn} \) are the random weights (scores) corresponding to each individual curve. Viewing, for the sake of an intuitive argument, the \( v_k(t) \) as deterministic curves, the iid assumption for the curves \( X_n(\cdot) \) implies this assumption for the random vectors \([X_1n, \ldots, X_{pn}]'\). To test it, the method proposed by Chitturi (1976) can be used. In reality, the \( v_k(t) \) must be replaced by estimated PC’s. This transition is not trivial because the estimated PC’s depend on all observations. Moreover, the difference between the population and sample PC’s is of the order \( N^{-1/2} \), and so the limit distribution of some statistics may contain an extra term. Our test statistic is based on products of scores, and we show that the effect of estimation of the PC’s is asymptotically negligible. We also show by simulations that it has a small effect on the finite sample performance of the test.

An example of data which motivated this research is shown in Figure 2.1. About a hundred terrestrial geomagnetic observatories form a network, INTERMAGNET, designed to monitor and understand the behavior of currents of charged particles flowing in the magnetosphere and ionosphere. Modern digital magnetometers record three components of the magnetic field in five second resolution, but the data available at INTERMAGNET’s website (http://www.intermagnet.org) or CD’s consist of one minute averages (1440 data points per day per component per observatory). The recent availability of these large data sets may lead to new insights if appropriate statistical tools are developed. In this respect, Functional Data Analysis stands out with its unique strength: physicists naturally view magnetogram records as curves
reflecting a continuous change in the structure of the various magnetic fields. In Figure 2.1, the daily variation caused by the rotation of the Earth is prominent. Notice however that the periodic pattern is not strong, as the current system is often nonlinearly disturbed by solar energy flows. For this reason, tools of traditional time series analysis based on stationarity, seasonality and polynomial trends are not suitable.

The paper is organized as follows. In Section 2.2, we formulate the test procedure together with mathematical assumptions and theorems establishing its asymptotic validity. The proofs of the theorems of Section 2.2 are presented in Section 2.5. Sections 4.4 and 2.4 are devoted, correspondingly, to the study of the finite sample performance of the test and its application to two types of functional data: credit card sales and geomagnetic records. Section 2.6 contains some lemmas on Hilbert space valued random elements which are used in Section 2.5, and may be useful in other similar contexts. In Section 2.7, we develop the required theory for random matrices.

2.2 Assumptions and main results

We first state the assumptions and introduce some notation. Our main result establishing limit null distribution is stated in Theorem 2.1. The consistency against the popular functional AR(1) model is established in Theorem 2.2.

We observe random functions \( \{X_n(t), \ t \in [0,1], \ n = 1, 2, \ldots N \} \) and want to test

\[ H_0 : \text{the } X_n(\cdot) \text{ are independent and identically distributed} \]

versus

\[ H_A : H_0 \text{ does not hold.} \]

We assume that the \( X_n \) are measurable elements of the space \( L^2[0,1) \) in which the norm is defined by \( ||X||^2 = \int_0^1 X^2(t)dt \).
Fig. 2.1: The horizontal component of the magnetic field measured in one minute resolution at Honolulu magnetic observatory from 1/1/2001 00:00 UT to 1/7/2001 24:00 UT.
For theoretical convenience we assume that the \( X_n \) have mean zero. We explain in Section 2.4 how to center the data.

Our theory is valid under the assumption of finite fourth moments:

\[
(2.2.2) \quad E[|X_n|^4] = E\left[\int_0^1 X_n^2(t) dt\right]^2 < \infty.
\]

If the \( X_n \) form a strictly stationary sequence (as is the case under \( H_0 \)), we denote by \( X \) a random element with the distribution of the \( X_n \) and define the covariance operator

\[
C(x) = E[\langle X, x \rangle X], \quad x \in L^2[0,1).
\]

The empirical covariance operator is defined by

\[
C_N(x) = \frac{1}{N} \sum_{n=1}^N \langle X_n, x \rangle X_n, \quad x \in L^2[0,1).
\]

The eigenelements of \( C \) are defined by

\[
C(v_j) = \lambda_j v_j, \quad j \geq 1.
\]

The eigenfunctions \( v_j \) form an orthonormal basis of \( L^2[0,1) \). We assume \( \lambda_1 \geq \lambda_2 \geq \lambda_3, \ldots \). The empirical eigenelements are defined by

\[
C_N(v_{jN}) = \lambda_{jN} v_{jN}, \quad j = 1, 2, \ldots, N.
\]

where we assume \( \lambda_{1N} \geq \lambda_{2N} \geq \ldots \geq \lambda_{NN} \). Since the operators \( C \) and \( C_N \) are symmetric and nonnegative definite, the eigenvalues \( \lambda_j \) and \( \lambda_{jN} \) are nonnegative.
If the eigenspace corresponding to the eigenvalue $\lambda_k$ is one dimensional, then formula (4.49) on p. 106 of Bosq (2000) implies that

\[(2.2.3) \quad \limsup_{N \to \infty} NE||v_k - v_kN||^2 < \infty.\]

(Note that $n$ before the expected value is missing in formula (4.49) of Bosq (2000), cf. formulas (4.17) and (4.44) of that monograph.)

To establish the null distribution of the test statistic $Q_N^F$ defined below, we merely require the following assumption:

**Assumption 2.2.** The functional observations $X_n$ are mean zero in $L^2[0,1]$, and (2.2.2) and (3.2.3) for $k \leq p$ hold.

The number $p$ (of principal components) appears in the statements of our main results.

We approximate the $X_n(t)$ by

$X_n^F(t) = \sum_{k=1}^{p} X_{kn}^F v_k N(t),$

where

\[(2.2.4) \quad X_{kn}^F := \int_0^1 X_n^F(t) v_{kN}(t) dt = \int_0^1 X_n(t) v_{kN}(t) dt.\]

The $X_n(t)$ are thus the projections of the functional observations on the subspace spanned by the largest $p$ empirical principal components, see Chapter 8 of Ramsay and Silverman (2005).

We will work with the random vectors

\[(2.2.5) \quad X_n^F = [X_{1n}^F, X_{2n}^F, \ldots, X_{pn}^F]'.\]
and analogously defined (unobservable) vectors

\[(2.2.6)\quad X_n = [X_{1n}, X_{2n}, \ldots, X_{pn}]', \]

where

\[(2.2.7)\quad X_{kn} = \int_0^1 X_n(t)v_k(t)dt.\]

Under \(H_0\), the \(X_n\) are iid mean zero random vectors in \(R^p\) for which we denote

\[v(i, j) = E[X_{it}X_{jt}], \quad V = [v(i, j)]_{i,j=1,\ldots,p}.\]

The matrix \(V\) is thus the \(p \times p\) covariance matrix. By \(C_h\) we denote the sample autocovariance matrix whose entries are

\[c_h(k, l) = \frac{1}{N} \sum_{t=1}^{N-h} X_{kt}X_{l,t+h}, \quad 0 \leq h < N.\]

Denote by \(r_{f,h}(i, j)\) and \(r_{b,h}(i, j)\) the \((i, j)\) entries of \(C_0^{-1}C_h\) and \(C_hC_0^{-1}\), respectively and introduce the statistic

\[(2.2.8)\quad Q_N = N \sum_{h=1}^H \sum_{i,j=1}^p r_{f,h}(i, j)r_{b,h}(i, j).\]

Theorem 2.11 shows that under \(H_0\), \(Q_N \xrightarrow{d} \chi^2_{p^2H} \).

Analogously to the way the statistic \(Q_N\) (2.2.8) is constructed from the vectors \(X_n, n = 1, \ldots, N\), we construct the statistic \(Q^F_N\) from the vectors \(X^F_n, n = 1, \ldots, N\).

The following theorem establishes the limit null distribution of the test statistic \(Q^F_N\).

**Theorem 2.1.** Under \(H_0\), if Assumption 2.2 holds, then \(Q^F_N \xrightarrow{d} \chi^2_{p^2H} \).
Theorem 2.1 is proven in Section 2.5.

Out of many possible directional alternatives, we focus on the ARH(1) (H stands for Hilbert Space) model of Bosq (2000) which has been used in several interesting applications [see e.g. Laukaitis and Račkauskas (2002), Fernández de Castro et al. (2005) and Kargin and Onatski (2008)]. It introduces serial correlation analogous to the usual AR(1) model.

Suppose then that

\[ X_{n+1} = \Psi X_n + \varepsilon_n \]

with iid mean zero innovations \( \varepsilon_n \in L^2[0,1] \). We assume that the operator \( \Psi \) is bounded, the solution to equations (2.2.9) is stationary and Assumption 2.2 holds. These conditions are implied by very mild assumptions on \( \Psi \), see Chapters 3 and 4 of Bosq (2000).

It is easy to check that the vectors (2.2.6) form a stationary VAR(1) process:

\[ X_{k,n+1} = \sum_{i=1}^{p} \psi_{ki} X_{in} + \varepsilon_{k,n+1}, \quad k = 1, 2, \ldots, p, \]

where \( \psi_{ik} = \langle v_i, \Psi v_k \rangle \), \( \varepsilon_{k,n} = \langle \varepsilon_n, v_k \rangle \). In the vector form

\[ X_{n+1} = \Psi X_n + e_{n+1}. \]

If the operator \( \Psi \) is not zero, there are \( i, k \geq 1 \) such that \( \psi_{ik} \neq 0 \).

The following theorem establishes the consistency against the ARH(1) model (2.2.9).

**Theorem 2.2.** Suppose the functional observations \( X_n \) follow a stationary solution to equations (2.2.9), Assumption 2.2 holds, and \( p \) is so large that the \( p \times p \) matrix \( \Psi \) in (2.2.10) is not zero. Then \( Q_N^F \xrightarrow{P} \infty \).
Table 2.1: Empirical size (in percent) of the test using Fourier basis. The simulated observations are Brownian bridges.

<table>
<thead>
<tr>
<th>Lag</th>
<th>p=3</th>
<th></th>
<th></th>
<th>p=4</th>
<th></th>
<th></th>
<th>p=5</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>N=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>7.7</td>
<td>2.5</td>
<td>0.6</td>
<td>7.4</td>
<td>2.8</td>
<td>0.3</td>
<td>7.9</td>
<td>3.5</td>
<td>0.4</td>
</tr>
<tr>
<td>3</td>
<td>6.8</td>
<td>2.5</td>
<td>0.3</td>
<td>6.7</td>
<td>3.3</td>
<td>0.6</td>
<td>4.9</td>
<td>2.0</td>
<td>0.3</td>
</tr>
<tr>
<td>5</td>
<td>4.9</td>
<td>2.0</td>
<td>0.0</td>
<td>3.6</td>
<td>1.4</td>
<td>0.2</td>
<td>4.0</td>
<td>1.7</td>
<td>0.2</td>
</tr>
<tr>
<td>N=100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.0</td>
<td>5.1</td>
<td>0.4</td>
<td>8.9</td>
<td>3.9</td>
<td>0.6</td>
<td>10.0</td>
<td>3.9</td>
<td>0.8</td>
</tr>
<tr>
<td>3</td>
<td>8.1</td>
<td>3.5</td>
<td>0.6</td>
<td>8.3</td>
<td>4.0</td>
<td>0.9</td>
<td>7.5</td>
<td>3.2</td>
<td>0.4</td>
</tr>
<tr>
<td>5</td>
<td>8.8</td>
<td>3.6</td>
<td>0.6</td>
<td>6.6</td>
<td>2.7</td>
<td>0.3</td>
<td>6.7</td>
<td>2.4</td>
<td>0.3</td>
</tr>
<tr>
<td>N=300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>9.8</td>
<td>4.6</td>
<td>1.2</td>
<td>9.4</td>
<td>4.0</td>
<td>0.9</td>
<td>10.1</td>
<td>4.7</td>
<td>0.6</td>
</tr>
<tr>
<td>3</td>
<td>9.3</td>
<td>4.8</td>
<td>1.0</td>
<td>9.1</td>
<td>4.7</td>
<td>0.9</td>
<td>10.0</td>
<td>5.4</td>
<td>0.8</td>
</tr>
<tr>
<td>5</td>
<td>7.2</td>
<td>3.7</td>
<td>1.0</td>
<td>8.2</td>
<td>3.8</td>
<td>0.7</td>
<td>10.6</td>
<td>5.5</td>
<td>1.2</td>
</tr>
</tbody>
</table>

Theorem 2.2 is proven in Section 2.5. Other departures from the null can be treated in a similar manner. The idea is that under the null the sample autocovariances of the $X^F_{kn}$ at a positive lag converge in distribution when multiplied by $\sqrt{N}$, whereas under any reasonable alternative, these autocovariances tend in probability to some constants. This could be used to establish consistency against local alternatives, but this theoretical investigation is not pursued here.

2.3 Finite sample performance

In this section we investigate the finite sample properties of the test using some generic models and sample sizes typical of applications discussed in Section 2.4.

To investigate the empirical size, we generated independent trajectories of the standard Brownian motion (BM) on $[0,1]$ and the standard Brownian bridge (BB). This was done by transforming cumulative sums of independent normal variables computed on a grid of $m$ equispaced points in $[0,1]$. We used values of $m$ ranging from 10 to 1440, and found that the empirical size basically does not depend on $m$.
To compute the principal components $v_{kN}$ and the corresponding eigenvalues using the R package \texttt{fda}, the functional data must be represented (smoothed) using a specified number of functions from a basis. We worked with Fourier and B splines functional bases and used 8, 16, and 80 basis functions. All results are based on one thousand replications.

Table 2.1 shows empirical sizes for the Brownian bridge and and the Fourier basis for several values of the lag $H = 1, 3, 5$, the number of principal components $p = 16, 80$ and sample sizes $N = 50, 100, 300$. The standard errors in this table are between 0.5 and 1 percent. In most cases, the empirical sizes are within two standard errors of the nominal size, and the size improves somewhat with increasing $N$. The same is true for the BM and B splines; no systematic dependence on the type of data or basis is seen, which accords with the nonparametric nature of the test. Of course, since the CLT is used to establish the asymptotic validity of the test, results are likely to be worse for nonnormal data, but a detailed empirical study is beyond the intended scope of this paper.

In a power study, we focus on the ARH(1) model (2.2.9), which can be more explicitly written as:

\begin{equation}
X_n(t) = \int_0^1 \psi(t, s)X_{n-1}(s)ds + \varepsilon_n(t), \quad t \in [0, 1], \quad n = 1, 2, \ldots, N.
\end{equation}

A sufficient condition for the assumptions of Theorem 2.2 to hold is

\begin{equation}
||\Psi||^2 = \int_0^1 \int_0^1 \psi^2(t, s)dt\,ds < 1.
\end{equation}

The norm in (2.3.12) is the Hilbert–Schmidt norm.
Table 2.2: Empirical power of the test using Fourier basis. The observations follow the ARH(1) model (2.3.11) with Gaussian kernel with $||\Psi|| = 0.5$ and iid standard Brownian motion innovations.

<table>
<thead>
<tr>
<th>Lag</th>
<th>p=3</th>
<th>p=4</th>
<th>p=5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>N=50</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>44.7</td>
<td>33.8</td>
<td>17.7</td>
</tr>
<tr>
<td>3</td>
<td>35.2</td>
<td>27.0</td>
<td>13.3</td>
</tr>
<tr>
<td>5</td>
<td>26.7</td>
<td>20.0</td>
<td>11.0</td>
</tr>
<tr>
<td>N=100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>71.2</td>
<td>64.2</td>
<td>51.4</td>
</tr>
<tr>
<td>3</td>
<td>67.9</td>
<td>61.0</td>
<td>44.9</td>
</tr>
<tr>
<td>5</td>
<td>62.3</td>
<td>54.6</td>
<td>38.6</td>
</tr>
<tr>
<td>N=300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>98.7</td>
<td>98.2</td>
<td>96.7</td>
</tr>
<tr>
<td>3</td>
<td>97.6</td>
<td>97.1</td>
<td>95.5</td>
</tr>
<tr>
<td>5</td>
<td>96.8</td>
<td>95.9</td>
<td>92.8</td>
</tr>
</tbody>
</table>

In our study, the innovations $\varepsilon_n$ in (2.3.11) are either standard BM’s or BB’s. We used two kernel functions: the Gaussian kernel

$$\psi(t, s) = C \exp\left\{\frac{t^2 + s^2}{2}\right\}, \quad (t, s) \in [0, 1]^2,$$

and Wiener kernel

$$\psi(t, s) = C \min(s, t), \quad (t, s) \in [0, 1]^2.$$

The constants $C$ were chosen so that $||\Psi|| = 0.3, 0.5, 0.7$. We used both Fourier and B spline basis.

The power against this alternative is expected to increase rapidly with $N$, as the test statistic is proportional to $N$. This is clearly seen in Table 2.2. The power also increases with $||\Psi||$; for $||\Psi|| = 0.7$ and the Gaussian kernel, it is practically 100% for $N = 100$ and all choices of other parameters.

There are two less trivial observations. The power is highest for lag $H = 1$. 
Table 2.3: Empirical power of the test using Fourier basis. The observations follow the ARH(1) model (2.3.11) with Gaussian kernel with $||\Psi|| = 0.5$ and iid standard Brownian bridge innovations.

<table>
<thead>
<tr>
<th>Lag</th>
<th>p=3</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>p=4</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>p=5</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>N=50</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>98.3</td>
<td>97.0</td>
<td>92.1</td>
<td></td>
<td>98.4</td>
<td>96.4</td>
<td>87.6</td>
<td></td>
<td>99.1</td>
<td>97.3</td>
<td>87.6</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>95.2</td>
<td>90.3</td>
<td>77.4</td>
<td></td>
<td>92.1</td>
<td>86.2</td>
<td>69.6</td>
<td></td>
<td>89.9</td>
<td>85.1</td>
<td>63.2</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>86.9</td>
<td>80.2</td>
<td>61.7</td>
<td></td>
<td>78.5</td>
<td>71.7</td>
<td>51.4</td>
<td></td>
<td>75.2</td>
<td>63.9</td>
<td>40.4</td>
</tr>
<tr>
<td></td>
<td>N=100</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>100</td>
<td>100</td>
<td>99.9</td>
<td></td>
<td>100</td>
<td>100</td>
<td>99.7</td>
<td></td>
<td>100</td>
<td>99.9</td>
<td>99.8</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>99.9</td>
<td>99.3</td>
<td>98.7</td>
<td></td>
<td>99.9</td>
<td>99.8</td>
<td>98.6</td>
<td></td>
<td>99.7</td>
<td>99.5</td>
<td>97.8</td>
</tr>
<tr>
<td></td>
<td>N=300</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
<td></td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>

This is because for the ARH(1) process the “correlation” between $X_n$ and $X_{n-1}$ is largest at this lag. By increasing the maximum lag $H$, the value of $Q^F_N$ generally increases, but the critical value increases too (df increases by $p^2$ for a unit increase in $H$). The power also depends on how the action of the operator $\Psi$ is “aligned” with the eigenfunctions $v_k$. If the inner products $\langle v_i, \Psi v_k \rangle$ are large in absolute value, the power is high. Thus, with all other parameters being the same, the power in Table 2.3 is greater than in Table 2.2 because of the different covariance structure of the Brownian bridge and the Brownian motion. In all cases, the power for the Wiener kernel is slightly lower than for the Gaussian kernel.

A full set of size and power tables is available upon request. The empirical sizes for the two processes we simulated are generally fairly close to nominal sizes and are not much affected by the choice of $H$ and $p$. Power against the ARH(1) model is very good if $H = 1$ is used.
2.4 Application to credit card transactions and diurnal geomagnetic variation

In this section, we apply our test to two data sets which naturally lend themselves to functional data analysis. The first data set, studied in Laukaitis and Račkauskas (2002), consists of the number of transactions with credit cards issued by Vilnius Bank, Lithuania. The second, is a daily geomagnetic variation, a similar data set was studied by Xu and Kamide (2004).

Suppose \( D_n(t_i) \) is the number of credit card transactions in day \( n \), \( n = 1, \ldots, 200 \), \((03/11/2000 - 10/02/2001)\) between times \( t_{i-1} \) and \( t_i \), where \( t_i - t_{i-1} = 8 \text{ min} \), \( i = 1, \ldots, 128 \). For our analysis, the transactions were normalized to have time stamps in interval \([0, 1]\), which thus corresponds to one day. To remove the weekly periodicity, we work with the differences \( X_n(t_i) = D_n(t_i) - D_{n-7}(t_i) \), \( n = 1, 2, \ldots, 193 \). Figure 2.2 displays the first three weeks of these data. A characteristic pattern of an AR(1) process with clusters of positive and negative observations is clearly seen. Two consecutive days are shown in the left–most panel of Figure 2.3 together with functional objects obtained by smoothing with 40 and 80 Fourier basis functions. As expected, the test rejects the null hypothesis at 1% level for both smooths, and all lag values \( 1 \leq H \leq 5 \) and the number of principal components equal to 4, 5, 10 and 20.

We estimated the ARH(1) model (2.3.11) using the function \texttt{linmod} from the \texttt{R} package \texttt{fda}, see Malfait and Ramsay (2003) and Ramsay and Silverman (2002, 2005). Table 2.4 displays the P-values which support this model choice, see Laukaitis and Račkauskas (2002). Note how starting with \( p = 2 \) or \( p = 3 \) the P-values increase and approach 100%. This accords with the findings of Section 4.4, and is caused by the fact that the dependence is captured by only a few most important principal components, and increasing \( p \) does not change the sampling distribution of \( Q_N^F \) very much, but shifts the limiting distribution to the right. We note that, in analogy with well–known results for real–valued time series, see e.g. Ljung and Box (1978) and references therein, the number of degrees of freedom in the asymptotic distribution
Table 2.4: P-values for the functional ARH(1) residuals of the credit card data $X_n$.

<table>
<thead>
<tr>
<th>Lag, $H$</th>
<th>p=1</th>
<th>p=2</th>
<th>p=3</th>
<th>p=4</th>
<th>p=5</th>
<th>p=6</th>
<th>p=7</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF=40</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>69.54</td>
<td>22.03</td>
<td>13.60</td>
<td>46.29</td>
<td>80.35</td>
<td>96.70</td>
<td>99.20</td>
</tr>
<tr>
<td>2</td>
<td>35.57</td>
<td>38.28</td>
<td>7.75</td>
<td>47.16</td>
<td>64.92</td>
<td>95.00</td>
<td>99.04</td>
</tr>
<tr>
<td>3</td>
<td>54.44</td>
<td>53.63</td>
<td>25.28</td>
<td>52.61</td>
<td>71.33</td>
<td>86.84</td>
<td>94.93</td>
</tr>
<tr>
<td>BF=80</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>57.42</td>
<td>18.35</td>
<td>53.30</td>
<td>89.90</td>
<td>88.33</td>
<td>95.40</td>
<td>99.19</td>
</tr>
<tr>
<td>2</td>
<td>35.97</td>
<td>23.25</td>
<td>23.83</td>
<td>45.07</td>
<td>55.79</td>
<td>46.39</td>
<td>70.65</td>
</tr>
<tr>
<td>3</td>
<td>36.16</td>
<td>36.02</td>
<td>26.79</td>
<td>30.21</td>
<td>56.81</td>
<td>34.51</td>
<td>47.00</td>
</tr>
</tbody>
</table>

of the statistic $Q^F_N$ computed from residuals of the ARH(1) model is likely to be less than $p^2H$. This question is a subject of on–going research, but even with this caveat, Table 2.4 gives strong support to the ARH(1) model.

We now turn to the ground-based magnetogram records which reflect the variations of the currents flowing in the Magnetosphere/Ionosphere. These data are used to understand the structure of this important complex geosystem. Since we present here merely an illustration of our procedure, we focus only on the horizontal intensity measured at Honolulu in 2001. The horizontal (H) intensity is the component of the magnetic field tangent to the Earth’s surface and pointing toward the magnetic North; its variation best reflects the changes in the large currents flowing in the magnetic equatorial plane. The top panel of Figure 2.4 shows two weeks of these data. Xu and Kamide (2004) used the H–component measured at Beijing in 2001 in order to understand the statistical structure of the daily variation and associate it with known or conjectured currents. Following Xu and Kamide (2004), we subtracted the linear change over a day to obtain the curves like those showed in the bottom panel of Figure 2.4. After centering over a period under study, we obtain the functional observations we work with. The analysis was conducted using Fourier base functions.
Fig. 2.2: Three weeks of centered time series of \{X_n(t_i)\} derived from credit card transaction data. The vertical dotted lines separate days.

Fig. 2.3: Two functional observations \(X_n\) derived from the credit card transactions (left-most panel) together with smooths obtained by projection on 40 and 80 Fourier basis functions.
We note that the issue of separating the daily variation from larger disturbances caused by magnetic storms is a complex one, and is the subject of on-going geophysical research, see Jach et al. (2006) for a recent contribution. For example, one can question whether what we see in the second day in the bottom panel of Figure 2.4 is an unusually large daily variation or an unremoved signature of a magnetic storm. This paper is however not concerned with such issues.

Testing one year magnetometer data with lags $H = 1, 2, 3$ and different numbers of principal components $p = 3, 4, 5$, yields P-values very close to zero. This indicates that while principal component analysis, advocated by Xu and Kamide (2004), may be a useful exploratory tool to study daily variation over the whole year, one must be careful when using any inferential tools based on it, as they typically require a simple random sample, see e.g. Section 5.2 of Seber (1984). We also applied the test to smaller subsets of data roughly corresponding to boreal Spring and Summer. The P-values reported in Table 2.5 show that the transformed data can to a reasonable approximation be viewed as a functional simple random sample. The discrepancy in the outcome of the test when applied to the whole year and to a season is probably due to the annual change of the position of the Honolulu observatory relative to the Sun whose energy drives the convective currents mainly responsible for the daily variation.

The two examples discussed in this section show that our test can detect departures from the assumption of independence (credit card data) or from the assumption of identical distribution (magnetometer data), and confirm both assumptions when they are expected to hold. In our examples, the results of the test do not depend much on the choice of the functional basis.
Table 2.5: P-values for the magnetometer data split by season.

<table>
<thead>
<tr>
<th>Lag</th>
<th>Feb, Mar, Apr, May</th>
<th>Jun, Jul, Aug, Sep</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>p=4</td>
<td>p=5</td>
</tr>
<tr>
<td>1</td>
<td>13.44</td>
<td>6.51</td>
</tr>
<tr>
<td>3</td>
<td>3.37</td>
<td>2.99</td>
</tr>
</tbody>
</table>

Fig. 2.4: Top: horizontal intensity (nT) measured at Honolulu 30/3/2001 - 13/4/2001 with the straight lines connecting first and last measurements in each day. Bottom: the same after subtracting the lines.
2.5 Proofs of Theorems 2.1 and 2.2

Proof of Theorem 2.1. By Theorem 2.11, it is enough to show that \( Q_F^N - Q_N \xrightarrow{P} 0 \).

By (2.2.8), this will follow if we show that for \( h \geq 1 \)

\[
\text{(2.5.13)} \quad N^{1/2}(C^F_h - C_h) \xrightarrow{P} 0
\]

and \( C^F_0 - C_0 \xrightarrow{P} 0 \). We will verify that (2.5.13) holds for all \( h \geq 0 \). Recall that

\[
c_h(k, l) = \frac{1}{N} \sum_{n=1}^{N-h} X_{kn} X_{l,n+h}; \quad c^F_h(k, l) = \frac{1}{N} \sum_{n=1}^{N-h} X^F_{kn} X^F_{l,n+h}.
\]

Therefore \( c^F_h(k, l) - c_h(k, l) = M_1 + M_2, \) where

\[
M_1 = \frac{1}{N} \sum_{n=1}^{N-h} (X_{kn} - X^F_{kn}) X_{l,n+h}; \quad M_2 = \frac{1}{N} \sum_{n=1}^{N-h} X^F_{kn} (X_{l,n+h} - X^F_{l,n+h}).
\]

We will first show that \( N^{1/2} M_1 \xrightarrow{P} 0 \). Observe that

\[
N^{1/2} M_1 = N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, v_k - v_{kN} \rangle \langle X_{n+h}, v_l \rangle
\]

\[
= \left\langle N^{-1/2} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l \rangle X_n, v_k - v_{kN} \right\rangle = \langle S_N, Y_N \rangle,
\]

where

\[
S_N := N^{-1/2} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l \rangle X_n; \quad Y_N = v_k - v_{kN}
\]

Note that by (3.2.3),

\[
E|\langle S_N, Y_N \rangle| \leq E[||S_N|| ||Y_N||] \leq (E[||S_N||^2])^{1/2}(E[||Y_N||^2])^{1/2} = O(N^{-1/2})(E[||S_N||^2])^{1/2}.
\]
To show that $N^{1/2}M_1 \xrightarrow{P} 0$, it thus remains to verify that $E\|S_N\|^2$ is bounded. Notice that

\[ E\|S_N\|^2 = N^{-1}E\| \sum_{n=1}^{N-h} \langle X_{n+h}, v_l \rangle X_n \|^2 \]

\[ = N^{-1}E \sum_{m,n=1}^{N-h} \langle X_{m+h}, v_l \rangle \langle X_{n+h}, v_l \rangle \langle X_m, X_n \rangle \]

\[ = N^{-1} \sum_{n=1}^{N-h} E[\langle X_{n+h}, v_l \rangle]^2 E\|X_n\|^2 \leq [E\|X_n\|^2]^2. \]

To show that $N^{1/2}M_2 \xrightarrow{P} 0$, decompose $M_2$ as $M_2 = M_{21} + M_{22}$, where

\[ M_{21} = \frac{1}{N} \sum_{n=1}^{N-h} \langle X_n, v_k \rangle \langle X_{n+h}, v_l - v_{lN} \rangle; \]

\[ M_{22} = \frac{1}{N} \sum_{n=1}^{N-h} \langle X_n, v_{kN} - v_k \rangle \langle X_{n+h}, v_l - v_{lN} \rangle. \]

By the argument developed for $M_1$, $N^{1/2}M_{21} \xrightarrow{P} 0$, so we must show $N^{1/2}M_{22} \xrightarrow{P} 0$. This follows from Lemma 2.7. \(\blacksquare\)

**Proof of Theorem 2.2.** We first state Lemmas 2.3 and 2.4 which form two critical building blocks of the proof.

**Lemma 2.3.** Suppose the vectors $X_n = [X_{1n}, X_{2n}, \ldots, X_{pn}]'$ follow a stationary vector AR(1) process $X_{n+1} = \Psi X_n + e_{n+1}$. The errors $e_n$ are iid mean zero with finite variance and $e_{n+1}$ is independent of $X_n$. Then,

\[ \sum_{i,j=1}^{p} r_{f,1}(i, j) r_{b,1}(i, j) \xrightarrow{P} \text{tr}[\Psi V \Psi' V^{-1}], \]

where $V$ is the covariance matrix of the vector $X_n$. 

Proof. Observe that
\[
C_1 = \frac{1}{N} \sum_{n=1}^{N-1} X_n X'_{n+1} = \frac{1}{N} \sum_{n=1}^{N-1} X_n [\Psi X_n + e_{n+1}]'
\]
\[
= \frac{1}{N} \sum_{n=1}^{N} X_n X' \Psi' + o_P(1) + \frac{1}{N} \sum_{n=1}^{N-1} X_n e_{n+1}' = V \Psi' + o_P(1),
\]
by the ergodic theorem. Consequently,
\[
C_0^{-1} C_1 \xrightarrow{P} \Psi'; \quad C_1 C_0^{-1} \xrightarrow{P} V \Psi V^{-1}
\]
and so \( r_{f,1}(i, j) \xrightarrow{P} \psi_{ji} \) and \( r_{b,1}(i, j) \xrightarrow{P} \{V \Psi V^{-1}\}_{ij} \). Therefore
\[
\sum_{i,j=1}^{p} r_{f,1}(i, j) r_{b,1}(i, j) \xrightarrow{P} \sum_{j=1}^{p} \sum_{i=1}^{p} \psi_{ji} [V \Psi V^{-1}]_{ij} = \sum_{j=1}^{p} [V \Psi V^{-1}]_{jj} = \text{tr}[V \Psi V^{-1}]. \quad \blacksquare
\]

Lemma 2.4. If \( V \) is a symmetric positive definite \( p \times p \) matrix and \( \Psi \) is a nonzero matrix of the same dimension, then \( \text{tr}[V \Psi V^{-1}] > 0 \).

Proof. To get a feel why this result is true, suppose \( p = 2 \) and \( V \) is diagonal, i.e.
\[
V = \begin{bmatrix}
\lambda_1 & 0 \\
0 & \lambda_2
\end{bmatrix}, \quad \Psi = \begin{bmatrix}
\psi_{11} & \psi_{12} \\
\psi_{21} & \psi_{22}
\end{bmatrix}.
\]

Then
\[
\Psi V \Psi V^{-1} = \begin{bmatrix}
\psi_{11} \lambda_1 & \psi_{12} \lambda_2 \\
\psi_{21} \lambda_1 & \psi_{22} \lambda_2
\end{bmatrix} \begin{bmatrix}
\psi_{11} \lambda_1^{-1} & \psi_{21} \lambda_2^{-1} \\
\psi_{12} \lambda_1^{-1} & \psi_{22} \lambda_2^{-1}
\end{bmatrix}
\]
\[
= \begin{bmatrix}
\psi_{11}^2 + \psi_{12}^2 \lambda_1 \lambda_2^{-1} & \psi_{11} \psi_{21} \lambda_1 \lambda_2^{-1} + \psi_{12} \psi_{22} \\
\psi_{21} \psi_{11} + \psi_{22} \psi_{12} \lambda_2 \lambda_1^{-1} & \psi_{21}^2 \lambda_1 \lambda_2^{-1} + \psi_{22}^2
\end{bmatrix}.
\]
Since \( \lambda_1 \) and \( \lambda_2 \) are positive, the trace is positive if one of the \( \psi_{jk} \) is positive.
For arbitrary symmetric positive definite $V$, there is an orthonormal matrix $P$ such that $V = P \Lambda P'$, where $\Lambda$ is diagonal, and its diagonal consists of positive eigenvalues of $V$. Therefore

$$\Psi V \Psi' V^{-1} = \Psi P \Lambda P' (P \Lambda P')^{-1} = \Psi P \Lambda P' \Psi' \Lambda^{-1} P'.$$

Since $\text{tr}(AB) = \text{tr}(BA)$, setting $A = \Psi P \Lambda P' \Psi' \Lambda^{-1}$ and $B = P'$, we obtain

$$\text{tr}[\Psi V \Psi' V^{-1}] = \text{tr}[P' \Psi P \Lambda P' \Psi' \Lambda^{-1}] = \text{tr}[\Phi \Lambda \Phi' \Lambda^{-1}],$$

where $\Phi = P' \Psi P$. Since $\Psi = P \Phi P'$, $\Phi$ is nonzero. Direct verification shows that the $j$th diagonal entry of $\Phi V \Phi' V^{-1}$ is $\sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j^{-1}$, and so

$$\text{(2.5.14)} \quad \text{tr}[\Psi V \Psi' V^{-1}] = \sum_{j=1}^{p} \sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j^{-1} > 0. \blacksquare$$

**Lemma 2.5.** If $V$ is a symmetric nonnegative definite $p \times p$ matrix and $\Psi$ is any matrix of the same dimension, then $\text{tr}[\Psi V \Psi' V] \geq 0$.

**Proof.** Proceeding exactly as in the proof of Lemma 2.4, we obtain

$$\text{tr}[\Psi V \Psi' V] = \sum_{j=1}^{p} \sum_{k=1}^{p} \phi_{jk}^2 \lambda_k \lambda_j \geq 0. \blacksquare$$

We now present the remainder of the proof of Theorem 2.2. Direct verification shows that

$$\text{(2.5.15)} \quad \sum_{i,j=1}^{p} r_{f,h}^F (i,j) r_{b,h}^F (i,j) = \text{tr} \left\{ [C_h^F]' [C_0^F]^{-1} C_h^F [C_0^F]^{-1} \right\}.$$
If $[C_0^F]^{-1}$ exists, it is positive definite, so by (2.5.14) and Lemma 2.5,

$$\sum_{i,j=1}^{p} r_{f,h}^F(i,j)r_{b,h}^F(i,j) \geq 0.$$  

By Lemmas 2.3 and 2.4,

$$\sum_{i,j=1}^{p} r_{f,1}(i,j)r_{b,1}(i,j) \rightarrow q > 0.$$  

It thus suffices to show that

$$(2.5.16) \quad \sum_{i,j=1}^{p} \left[ r_{f,1}^F(i,j)r_{b,1}^F(i,j) - r_{f,1}(i,j)r_{b,1}(i,j) \right] \rightarrow 0.$$  

Relation (2.5.16) will follow from $C_h^F - C_h \rightarrow 0$.

In the remainder of the proof we use the notation introduced in the proof of Theorem 2.1. We must show that $M_1 \rightarrow 0$ and $M_2 \rightarrow 0$. We will display the argument only for $M_1$. Observe that

$$M_1 = \left< N^{-1} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l \rangle X_n, v_k - v_{kN} \right>$$

By (3.2.3), $||v_k - v_{kN}|| \rightarrow 0$. Since,

$$E||N^{-1} \sum_{n=1}^{N-h} \langle X_{n+h}, v_l \rangle X_n|| \leq E|| \langle X_{n+h}, v_l \rangle X_n || \leq E||X_n||^2,$$

it follows that $M_1 \rightarrow 0$. $\blacksquare$
2.6 Auxiliary Lemmas for $H$-valued random elements

Consider the empirical lag-$h$ autocovariance operator

\[(2.6.17)\quad C_{N,h}(x) = \frac{1}{N} \sum_{n=1}^{N-h} \langle X_n, x \rangle X_{n+h}.\]

Recall that the Hilbert-Schmidt norm of a Hilbert-Schmidt operator $S$ is defined by

\[||S||_S^2 = \sum_{j=1}^\infty ||S(e_j)||^2,\]

where $\{e_1, e_2, \ldots\}$ is any orthonormal basis.

**Lemma 2.6.** Suppose the $X_i$ are iid random elements in a separable Hilbert space with $E||X_0||^2 < \infty$, then

\[E||C_{N,h}||_S^2 = \frac{N - h}{N^2} \left(E||X_0||^2\right)^2.\]

**Proof.** Observe that

\[||C_{N,h}||_S^2 = \sum_{j=1}^\infty ||C_{N,h}(e_j)||^2\]

\[= \sum_{j=1}^\infty \left(\frac{1}{N} \sum_{n=1}^{N-h} \langle X_n, e_j \rangle X_{n+h}, \frac{1}{N} \sum_{m=1}^{N-h} \langle X_m, e_j \rangle X_{m+h}\right)\]

\[= \sum_{j=1}^\infty \frac{1}{N^2} \sum_{m,n=1}^{N-h} \langle X_m, e_j \rangle \langle X_n, e_j \rangle \langle X_{m+h}, X_{n+h}\rangle .\]

It follows from the independence of the $X_n$ that

\[E||C_{N,h}||_S^2 = \frac{1}{N^2} \sum_{n=1}^{N-h} \sum_{j=1}^\infty E[(X_n, e_j)]^2 E[(X_{n+h}, X_{n+h})]^2\]

\[= E||X_0||^2 \frac{1}{N^2} \sum_{n=1}^{N-h} E\left[\sum_{j=1}^\infty (X_n, e_j)^2\right] = \left[E||X_0||^2\right]^2 \frac{N - h}{N^2}. \]

\[\blacksquare\]
Lemma 2.7. Suppose $X_n, Z_N, Y_N$ are random elements in a separable Hilbert space. We assume

\begin{equation}
E|\|Y_N\||^2 = O(N^{-1}), \quad E|\|Z_N\||^2 = O(N^{-1});
\end{equation}

(2.6.18)

\begin{equation}
X_n \sim iid, \quad E|\|X_n\||^2 < \infty.
\end{equation}

(2.6.19)

Then

\[ N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, Y_N \rangle \langle X_{n+h}, Z_N \rangle \xrightarrow{P} 0.\]

Proof. Observe that

\[ N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, Y_N \rangle \langle X_{n+h}, Z_N \rangle = \langle C_{N,h}(Y_N), N^{1/2}Z_N \rangle, \]

with the operator $C_{N,h}$ defined in (3.9.34). Since $P(N^{1/2}|Z_N| > C) \leq C^{-2}NE|Z_N|^2$, $N^{1/2}|Z_N| = O_P(1)$. Thus it remains to verify that $C_{N,h}(Y_N) \xrightarrow{P} 0$. Since the Hilbert–Schmidt norm is not less than the uniform operator norm $\| \cdot \|_\mathcal{L}$, see Bosq (2000), p. 35, we obtain from Lemma 2.6:

\[ E|\|C_{N,h}(Y_N)\|| \leq E[|\|C_{N,h}\|_\mathcal{L}|Y_N||] \leq E[|\|C_{N,h}\|_S|Y_N||] \leq (E|\|C_{N,h}\|_S^2)^{1/2} (E|Y_N|^2)^{1/2} = O(N^{-1/2})O(N^{-1/2}) = O(N^{-1}). \]

2.7 Limit theory for sample autocovariance matrices

Detailed proofs of Theorems 2.8, 2.10, 2.11, Remark 2.9 and all displayed formulas leading to them are omitted to conserve space, but are available upon request.

Consider random vectors $X_1, \ldots, X_N$, where $X_t = [X_{1t}, X_{2t}, \ldots, X_{pt}]$. We assume
that the \( X_t, \ t = 1, 2, \ldots \) are iid mean zero with finite variance and denote

\[
v(i, j) = E[X_{it}X_{jt}], \quad \mathbf{V} = [v(i, j)]_{i,j=1,...,p}.
\]

By \( \mathbf{C}_h \) we denote the sample autocovariance matrix with entries

\[
c_h(k, l) = \frac{1}{N} \sum_{t=1}^{N-h} X_{kt}X_{l,t+h}, \quad h \geq 0.
\]

We first establish the joint asymptotic distribution of \( \mathbf{C}_h, \ h = 1, 2, \ldots, H \). We use the asymptotic normality of \( h \)-independent sequences, see Theorem 6.4.2 of Brockwell and Davis (1991), and the Wald–Cramer device. Let \( \mathbf{Z}_0 \) and \( \mathbf{Z}_h \) be matrices with jointly Gaussian entries \( Z_0(k,l), Z_h(k,l), \ k, l = 1, \ldots, p \), with mean zero and covariances

(2.7.20) \[ E[Z_0(k,l)Z_0(i,j)] = \eta(k,l,i,j) - v(i,j)v(k,l); \]

(2.7.21) \[ E[Z_h(k,l)Z_h(i,j)] = v(k,i)v(l,j) \quad (h \geq 1), \]

where \( \eta(k,l,i,j) = E[X_{kt}X_{lt}X_{it}X_{jt}] \).

**Theorem 2.8.** If the \( X_t \) are iid with finite fourth moment, then

\[
N^{1/2} [\mathbf{C}_0 - \mathbf{V}, \mathbf{C}_1, \ldots, \mathbf{C}_H] \overset{d}{\to} [\mathbf{Z}_0, \mathbf{Z}_1, \ldots, \mathbf{Z}_H],
\]

where the \( \mathbf{Z}_h, \ h = 0, 1, \ldots, H, \) are independent mean zero Gaussian matrices with covariances (2.7.20) and (2.7.21).

A critical ingredient of the derivation of the asymptotic distribution of the test statistic \( Q_N \) is the understanding of the asymptotic distribution of \( \mathbf{C}_0^{-1} \). Let \( u(k,l) \)
be the \((k, l)\)-entry of \(V^{-1}\). Calculations involving derivatives of products of matrices and the delta method lead to the following result:

\[(2.7.22) \quad N^{1/2}(C_0^{-1} - V^{-1}) \xrightarrow{d} Y_0,\]

where \(Y_0\) is a mean zero Gaussian matrix with \((i, j)\)-entry

\[(2.7.23) \quad Y_0(i, j) = -\sum_{k, l=1}^{p} u(i, k)u(l, j)Z_0(k, l).\]

**Remark 2.9.** Observe that

\[
E[Y_0(i, j)Y_0(\alpha, \beta)] = \sum_{k, l=1}^{p} \sum_{\kappa, \lambda=1}^{p} u(i, k)u(l, j)u(\alpha, \kappa)u(\lambda, \beta)E[Z_0(k, l)Z_0(\kappa, \lambda)].
\]

The covariances \(E[Z_0(k, l)Z_0(\kappa, \lambda)]\) are given in (2.7.20), and it is seen that they do not imply, in general, the covariances in formula (4.4) of Chitturi (1976), which is true only if the process \(X_t\) is Gaussian.

We now find the limit of \(N^{1/2}C_0^{-1}C_h\), \(h \geq 1\). Further calculations using the delta method applied to the matrix \([C_0 - V, C_1, \ldots, C_H]'\) lead to the following theorem:

**Theorem 2.10.** If the \(X_t\) are iid with finite fourth moment, then

\[(2.7.24) \quad N^{1/2}C_0^{-1}[C_1, \ldots, C_H] \xrightarrow{d} V^{-1}[Z_1, \ldots, Z_H],\]

where the \(Z_h\), \(h = 0, 1, \ldots, H\), are independent mean zero Gaussian matrices with covariances (2.7.21).
Denote by \( r_{f,h}(i, j) \) and \( r_{b,h}(i, j) \) the \((i, j)\) entries of \( C_0^{-1} C_h \) and \( C_h C_0^{-1} \), respectively. Introduce the statistic

\[
Q_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{p} r_{f,h}(i, j) r_{b,h}(i, j).
\]

**Theorem 2.11.** If the \( X_t \) are iid with finite fourth moment, then \( Q_N \xrightarrow{d} \chi^2_{p^2 H} \).

**Proof.** Similarly to (2.7.24), it can be verified that

\[
N^{1/2} [C_1, \ldots, C_H] C_0^{-1} \xrightarrow{d} [Z_1, \ldots, Z_H] V^{-1},
\]

and that convergence (2.7.24) and (2.7.26) are joint. Since the matrices \([C_0^{-1} C_h, C_h C_0^{-1}]\) are asymptotically independent, it suffices to verify that

\[
N \sum_{i,j=1}^{p} r_{f,h}(i, j) r_{b,h}(i, j) \xrightarrow{d} \chi^2_{p^2}.
\]

To lighten the notation, in the remainder of the proof we suppress the index \( h \) (the limit distributions do not depend on \( h \)). Denote by \( \rho_f(i, j) \) and \( \rho_b(i, j) \), respectively, the entries of matrices \( V^{-1} Z \) and \( Z V^{-1} \). By (2.7.24) and (2.7.26), it suffices to show that

\[
\sum_{i,j=1}^{p} \rho_f(i, j) \rho_b(i, j) \xrightarrow{d} \chi^2_{p^2}.
\]

Denote by \( \tilde{Z} \) the column vector of length \( p^2 \) obtained by expanding the matrix \( Z \) row by row. Then the covariance matrix of \( \tilde{Z} \) is the \( p^2 \times p^2 \) matrix \( V \otimes V \). By formula (23) on p. 600 of Anderson (1984), its inverse is \( (V \otimes V)^{-1} = V^{-1} \otimes V^{-1} = U \otimes U \). It thus follows from theorem 3.3.3 of Anderson (1984) that

\[
\tilde{Z}'(U \otimes U)\tilde{Z} \xrightarrow{d} \chi^2_{p^2}.
\]
It remains to show that the LHS of (2.7.28) is equal to the LHS of (2.7.29). The entry $Z(i, k)$ of the vector $\tilde{Z}'$ multiplies the row $u(i, \cdot)u(k, \cdot)$ of $U \otimes U$; the entry $Z(j, l)$ of $\tilde{Z}$ multiplies the column $u(\cdot, j)u(\cdot, l)$. Consequently,

\[
\tilde{Z}'(U \otimes U)\tilde{Z} = \sum_{i,j,k,l=1}^{p} u(i, j)u(k, l)Z(i, k)Z(j, l)
\]

\[
= \sum_{i,l=1}^{p} \sum_{j=1}^{p} u(i, j)Z(j, l) \sum_{k=1}^{p} Z(i, k)u(k, l) = \sum_{i,l=1}^{p} \rho_f(i, l)\rho_b(i, l),
\]

completing the proof. \blacksquare
CHAPTER 3
TESTS FOR ERROR CORRELATION IN THE FUNCTIONAL LINEAR MODEL\textsuperscript{1}

3.1 Introduction
The last decade has seen the emergence of the functional data analysis (FDA) as a useful area of statistics which provides convenient and informative tools for the analysis of data objects of large dimension. The influential book of Ramsay and Silverman (2005) provides compelling examples of the usefulness of this approach. Functional data arise in many contexts. This paper is motivated by our work with data obtained from very precise measurements at fine temporal grids which arise in engineering, physical sciences and finance. At the other end of the spectrum are sparse data measured with error which are transformed into curves via procedures that involve smoothing. Such data arise, for example, in longitudinal studies on human subjects or in biology, and wherever frequent, precise measurements are not feasible. Our methodology and theory are applicable to such data after they have been appropriately transformed into functional curves. Many such procedures are now available.


\textsuperscript{1}COAUTHORED BY R. GABRYS, L. HORVÁTH, AND P. KOKOSZKA. REPRODUCED BY PERMISSION FROM JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION, FORTHCOMING, 2010.
For any statistical model, it is important to evaluate its suitability for particular data. In the context of the multivariate linear regression, well established approaches exist, but for the functional linear model, only the paper of Chiou and Müller (2007) addresses the diagnostics in any depth. These authors emphasize the role of the functional residuals \( \hat{\varepsilon}_i(t) = \hat{Y}_i(t) - Y_i(t) \), where the \( Y_i(t) \) are the response curves, and the \( \hat{Y}_i(t) \) are the fitted curves, and propose a number of graphical tools, akin to the usual residual plots, which offer a fast and convenient way of assessing the goodness of fit. They also propose a test statistic based on Cook’s distance introduced in Cook (1977) or Cook and Weisberg (1982), whose null distribution can be computed by randomizing a binning scheme.

We propose two goodness–of–fit tests aimed at detecting serial correlation in the error functions \( \varepsilon_n(t) \) in the fully functional model

\[
Y_n(t) = \int \psi(t, s)X_n(s)ds + \varepsilon_n(t), \quad n = 1, 2, \ldots, N.
\]  

The assumption of iid \( \varepsilon_n \) underlies all inferential procedures for model (3.1.1) proposed to date. As in the multivariate regression, error correlation affects various variance estimates, and, consequently, confidence regions and distributions of test statistics. In particular, prediction based on LS estimation is no longer optimal. In the context of scalar data, these facts are well–known and go back at least to Cochrane and Orcutt (1949). If functional error correlation is detected, currently available inferential procedures cannot be used. At this point, no inferential procedures for the functional linear model with correlated errors are available, and it is hoped that this paper will motivate research in this direction. For scalar data, the relevant research is very extensive, so we mention only the influential papers of Sacks and Ylvisaker (1966) and Rao and Griliches (1969), and refer to textbook treatments in Chapters 9 and 10 of Seber and Lee (2003), Chapter 8 of Hamilton (1989) and Section 13.5 of Bowerman and O’Connell (1990). The general idea is that when dependence in errors is detected, it must be modeled, and inference must be suitably adjusted.
The methodology of Chiou and Müller (2007) was not designed to detect error correlation, and can leave it undetected. Figure 3.1 shows diagnostic plots of Chiou and Müller (2007) obtained for synthetic data that follow a functional linear model with highly correlated errors. These plots exhibit almost ideal football shapes. It is equally easy to construct examples in which our methodology fails to detect departures from model (3.1.1), but the graphs of Chiou and Müller (2007) immediately show it. The simplest such example is given by \( Y_n(t) = X_n^2(t) + \varepsilon_n(t) \) with iid \( \varepsilon_n \). Thus, the methods we propose are complimentary tools designed to test the validity of specification (3.1.1) with iid errors against the alternative of correlation in the errors.

Despite a complex asymptotic theory, the null distribution of both test statistics we propose is asymptotically chi–squared, which turns out to be a good approximation in finite samples. The test statistics are relatively easy to compute, an \( \text{R} \) code is available upon request. They can be viewed as nontrivial refinements of the ideas of Durbin and Watson (1950, 1951, 1971), see also Chatfield (1998) and Section 10.4.4 of Seber and Lee (2003), who introduced tests for serial correlation in the standard linear regression. Their statistics are functions of sample autocorrelations of the residuals, but their asymptotic distributions depend on the distribution of the regressors, and so various additional steps and rough approximations are required, see Thiel and Nagar (1961) and Thiel (1965), among others. To overcome these difficulties, Schmoyer (1994) proposed permutation tests based on quadratic forms of the residuals. We appropriately define residual autocorrelations, and their quadratic forms (not the quadratic forms of the residuals as in Schmoyer (1994), in such a way that the asymptotic distribution is the standard chi–squared distribution.

The complexity of the requisite asymptotic theory is due to the fact that in order to construct a computable test statistic, finite dimensional objects reflecting the relevant properties of the infinite dimensional unobservable errors \( \varepsilon_n(t) \) must be constructed. In the standard regression setting, the explanatory variables live in a finite
Fig. 3.1: Diagnostic plots of Chiou and Müller (2007) for a synthetic data set simulated according to model (3.1.1) in which the errors $\varepsilon_n$ follow the functional autoregressive model of Bosq (2000).
dimensional Euclidean space with a fixed (standard) basis, and the residuals reflect the effect of parameter estimation. In the functional setting, before any estimation can be undertaken, the dimension of the data must be reduced, typically by projecting on an “optimal” finite dimensional subspace. This projection operation introduces an error. Next, the “optimal subspace” must be estimated, and this introduces another error. Finally, estimation of the kernel $\psi(\cdot, \cdot)$ introduces still another error.

The two methods proposed in this paper start with two ways of defining the residuals. Method I uses projections of all curves on the functional principal components of the regressors, and so is closer to the standard regression in that one common basis is used. This approach is also useful for testing the stability of model (3.1.1) against a change point alternative, see Horváth et al. (2009a). Method II uses two bases: the eigenfunctions of the covariance operators of the regressors and of the responses.

The remainder of the paper is organized as follows. Section 3.2 introduces the assumptions and the notation. Section 3.3 develops the setting for the least squares estimation needed define the residuals used in Method I. After these preliminaries, both tests are described in Section 3.4, with the asymptotic theory presented in Section 3.5. The finite sample performance is evaluated in Section 3.6 through a simulation study, and further examined in Section 3.7 by applying both methods to magnetometer and financial data. All proofs are collected in Sections 3.8, 3.9 and 3.10.

3.2 Preliminaries

We denote by $L^2$ the space of square integrable functions on the unit interval, and by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$ the usual inner product and the norm it generates.

The usual conditions imposed on model (3.1.1) are collected in the following assumption:

**Assumption 3.2.** The errors $\varepsilon_n$ are independent identically distributed mean zero elements of $L^2$ satisfying $E\|\varepsilon_n\|^4 < \infty$. The covariates $X_n$ are independent identically
distributed mean zero elements of $L^2$ satisfying $E\|X_n\|^4 < \infty$. The sequences $\{X_n\}$ and $\{\varepsilon_n\}$ are independent.

For data collected sequentially over time, the regressors $X_n$ need not be independent. We formalize the notion of dependence in functional observations using the notion of $L^4$-approximability advocated in other contexts by Hörmann (2008), Berkes et al. (2009b), Aue et al. (2009), and used for functional data by Hörmann and Kokoszka (2010) and Aue et al. (2010). We now list the assumptions we need to establish the asymptotic theory. For ease of reference, we repeat some conditions contained in Assumption 3.2; the weak dependence of the $\{X_n\}$ is quantified in Conditions (A2) and (A5). Assumption 3.2 will be needed to state intermediate results.

(A1) The $\varepsilon_n$ are independent, identically distributed with $E\varepsilon_n = 0$ and $E\|\varepsilon_n\|^4 < \infty$.
(A2) Each $X_n$ admits the representation

$$X_n = g(\alpha_n, \alpha_{n-1}, \ldots),$$

in which the $\alpha_k$ are independent, identically distributed elements of a measurable space $S$, and $g : S^\infty \to L^2$ is a measurable function.
(A3) The sequences $\{\varepsilon_n\}$ and $\{\alpha_n\}$ are independent.
(A4) $EX_n = 0, \ E\|X_n\|^4 < \infty$.
(A5) There are $c_0 > 0$ and $\kappa > 2$ such that

$$\left( E\|X_n - X_n^{(k)}\|^4 \right)^{1/4} \leq c_0 k^{-\kappa},$$

where

$$X_n^{(k)} = g(\alpha_n, \alpha_{n-1}, \ldots, \alpha_{n-k+1}, \alpha_n^{(k)}, \alpha_{n-k}^{(k)}, \alpha_{n-k-1}^{(k)}, \ldots),$$

and where the $\alpha_{k}^{(k)}$ are independent copies of $\alpha_0$.

Condition (A2) means that the sequence $\{X_n\}$ admits a causal representation
known as a Bernoulli shift. It follows from (A2) that \( \{X_n\} \) is stationary and ergodic. The structure of the function \( g(\cdot) \) is not important, it can be a linear or a highly non-linear function. What matters is that according to (A5), \( \{X_n\} \) is weakly dependent, as it can be approximated with sequences of \( k \)-dependent variables, and the approximation improves as \( k \) increases. Several examples of functional sequences satisfying (A2), (A4) and (A5) can be found in H"{o}rmann and Kokoszka (2010) and Aue et al. (2010). They include functional linear, bilinear and conditionally heteroscedastic processes.

We denote by \( C \) the covariance operator of the \( X_i \) defined by \( C(x) = E[\langle X, x \rangle X] \), \( x \in L^2 \), where \( X \) has the same distribution as the \( X_i \). By \( \lambda_k \) and \( v_k \), we denote, correspondingly, the eigenvalues and the eigenfunctions of \( C \). The corresponding objects for the \( Y_i \) are denoted \( \Gamma, \gamma_k, u_k \), so that

\[
C(v_k) = \lambda_k v_k, \quad X_n = \sum_{i=1}^{\infty} \xi_{ni} v_i, \quad \xi_{ni} = \langle v_i, X_n \rangle ;
\]

\[
\Gamma(u_k) = \gamma_k u_k, \quad Y_n = \sum_{j=1}^{\infty} \zeta_{nj} u_j, \quad \zeta_{nj} = \langle u_j, Y_n \rangle .
\]

In practice, we must replace the population eigenfunctions and eigenvalues by their empirical counterparts \( \hat{\lambda}_k, \hat{v}_k, \hat{\gamma}_k, \hat{u}_k \) defined as the eigenelements of the empirical covariance operators (we assume \( EX_n(t) = 0 \))

\[
\hat{C}(x) = N^{-1} \sum_{n=1}^{N} \langle X_n, x \rangle X_n, \quad x \in L^2,
\]

and analogously defined \( \hat{\Gamma} \). The empirical scores are also denoted with the "hat", i.e. by \( \hat{\xi}_{ni} \) and \( \hat{\zeta}_{nj} \). We often refer to the \( v_i, u_j \) as the functional principal components (FPC’s), and to the \( \hat{v}_i, \hat{u}_j \) as the empirical functional principal components (EFPC’s).

To state the alternative, we must impose dependence conditions on the \( \varepsilon_n \). We use the same conditions that we imposed on the \( X_n \), because then the asymptotic
arguments under $H_A$ can use the results derived for the $X_n$ under $H_0$. Specifically, we introduce the following assumptions:

(B1) $E\varepsilon_n = 0$ and $E||\varepsilon_n||^4 < \infty$.

(B2) Each $\varepsilon_n$ admits the representation

$$\varepsilon_n = h(u_n, u_{n-1}, \ldots),$$

in which the $u_k$ are independent, identically distributed elements of a measurable space $S$, and $h : S^\infty \to L^2$ is a measurable function.

(B3) The sequences $\{u_n\}$ and $\{\alpha_n\}$ are independent.

(B4) There are $c_0 > 0$ and $\kappa > 2$ such that

$$\left( E||\varepsilon_n - \varepsilon_n^{(k)}||^4 \right)^{1/4} \leq c_0 k^{-\kappa},$$

where

$$\varepsilon_n^{(k)} = h(u_n, u_{n-1}, \ldots, u_{n-k+1}, u_{n-k}^{(k)}, u_{n-k-1}^{(k)}, \ldots),$$

and where the $u^{(k)}_k$ are independent copies of $u_0$.

The tests proposed in Section 3.4 detect dependence which manifests itself in a correlation between $\varepsilon_n$ and $\varepsilon_n+h$ for at least one $h$. Following Bosq (2000), we say that $\varepsilon_n$ and $\varepsilon_{n+h}$ are uncorrelated if $E[\langle \varepsilon_n, x \rangle \langle \varepsilon_{n+h}, y \rangle] = 0$ for all $x, y \in L^2$. If $\{e_j\}$ is any orthonormal basis in $L^2$, this is equivalent to $E[\langle \varepsilon_n, e_i \rangle \langle \varepsilon_{n+h}, e_j \rangle] = 0$ for all $i, j$. The two methods introduced in Section 3.4 detect the alternatives with $e_i = v_i$ (Method I) and $e_i = u_i$ (Method II). These methods test for correlation up to lag $H$, and use the FPC $v_i, i \leq p$, and $u_i, i \leq q$.

With this background, we can state the null and alternative hypotheses as follows.

$H_0$: Model (3.1.1) holds together with Assumptions (A1)–(A5).

The key assumption is (A1), i.e. the independence of the $\varepsilon_n$. 

H_{A,I}: Model (3.1.1) holds together with Assumptions, (A2), (A4), (A5), (B1)–(B4), and $E[(\varepsilon_0, v_i) (\varepsilon_h, v_j)] \neq 0$ for some $1 \leq h \leq H$ and $1 \leq i, j \leq p$.

H_{A,II}: Model (3.1.1) holds together with Assumptions, (A2), (A4), (A5), (B1)–(B4), and $E[(\varepsilon_0, u_i) (\varepsilon_h, u_j)] \neq 0$ for some $1 \leq h \leq H$ and $1 \leq i, j \leq q$.

Note that the $u_i$ are well defined under the alternative, because (A2), (A4), (A5) and (B1)–(B4) imply that the $Y_n$ form a stationary sequence.

In the proofs, we will often use the following result established in Hörmann and Kokoszka (2010) and Aue et al. (2010). In Theorem 3.1, and in the following, we set

$$\hat{c}_j = \text{sign}((v_j, \hat{v}_j)).$$

**Theorem 3.1.** Suppose Assumptions (A2), (A4) and (A5) hold, and

$$(3.2.2) \quad \lambda_1 > \lambda_2 > \ldots > \lambda_p > \lambda_{p+1}.$$  

Then, for each $1 \leq j \leq p$,

$$(3.2.3) \quad \limsup_{N \to \infty} NE \left[ ||\hat{c}_j v_j - v_j||^2 \right] < \infty, \quad \limsup_{N \to \infty} NE \left[ |\lambda_j - \hat{\lambda}_{j}|^2 \right] < \infty.$$  

### 3.3 Least squares estimation

In this section we show how model (3.1.1) can be cast into a standard estimable form. The idea is different from the usual approaches, e.g. in Ramsay and Silverman (2005) and Yao et al. (2005b), so a detailed exposition is necessary. The goal is to obtain clearly defined residuals which can be used to construct a goodness–of–fit test. This section carefully explains the three steps involved in the construction of the residuals in the setting of model (3.1.1). The idea is that the curves are represented by their coordinates with respect to the FPC’s of the $X_n$, e.g. $Y_{nk} = \langle Y_n, v_k \rangle$ is the projection of the $n$th response onto the $k$th largest FPC. A formal linear model for
these coordinates is constructed and estimated by least squares. This formal model
does not however satisfy the usual assumptions due to the effect of the projection of
infinite dimensional curves on a finite dimensional subspace, and so its asymptotic
analysis is delicate.

Since the $v_k$ form a basis in $L^2([0,1])$, the products $v_i(t)v_j(s)$ form a basis in
$L^2([0,1] \times [0,1])$. Thus, if $\psi(\cdot,\cdot)$ is a Hilbert–Schmidt kernel, then

$$(3.3.4) \quad \psi(t,s) = \sum_{i,j=1}^{\infty} \psi_{ij} v_i(t)v_j(s),$$

where $\psi_{ij} = \iint \psi(t,s)v_i(t)v_j(s)dtds$. Therefore,

$$\int \psi(t,s)X_n(s)ds = \sum_{i,j=1}^{\infty} \psi_{ij} v_i(t) \langle X_n, v_j \rangle.$$ 

Hence, for any $1 \leq k \leq p$, we have

$$(3.3.5) \quad Y_{nk} = \sum_{j=1}^{p} \psi_{kj} \xi_{nj} + e_{nk} + \eta_{nk},$$

where

$$Y_{nk} = \langle Y_n, v_k \rangle, \quad \xi_{nj} = \langle X_n, v_j \rangle, \quad e_{nk} = \langle \varepsilon_n, v_k \rangle,$$

and where

$$\eta_{nk} = \sum_{j=p+1}^{\infty} \psi_{kj} \langle X_n, v_j \rangle.$$ 

We combine the errors $e_{nk}$ and $\eta_{nk}$ by setting

$$\delta_{nk} = e_{nk} + \eta_{nk}.$$ 

Note that the $\delta_{nk}$ are no longer iid.
Setting

\[ X_n = [\xi_{n1}, \ldots, \xi_{np}]^T \quad Y_n = [Y_{n1}, \ldots, Y_{np}]^T \quad \delta_n = [\delta_{n1}, \ldots, \delta_{np}]^T, \]

\[ \psi = [\psi_{11}, \ldots, \psi_{1p}, \psi_{21}, \ldots, \psi_{2p}, \ldots, \psi_{p1}, \ldots, \psi_{pp}]^T, \]

we rewrite (3.3.5) as

\[ Y_n = Z_n \psi + \delta_n, \quad n = 1, 2, \ldots, N, \]

where each \( Z_n \) is a \( p \times p^2 \) matrix

\[
Z_n = \begin{bmatrix}
X_n^T & 0_p^T & \cdots & 0_p^T \\
0_p^T & X_n^T & \cdots & 0_p^T \\
\vdots & \vdots & \ddots & \vdots \\
0_p^T & 0_p^T & \cdots & X_n^T
\end{bmatrix}
\]

with \( 0_p = [0, \ldots, 0]^T \).

Finally, defining the \( Np \times 1 \) vectors \( \mathbf{Y} \) and \( \delta \) and the \( Np \times p^2 \) matrix \( \mathbf{Z} \) by

\[
\mathbf{Y} = \begin{bmatrix}
\mathbf{Y}_1 \\
\mathbf{Y}_2 \\
\vdots \\
\mathbf{Y}_N
\end{bmatrix}, \quad \delta = \begin{bmatrix}
\delta_1 \\
\delta_2 \\
\vdots \\
\delta_N
\end{bmatrix}, \quad \mathbf{Z} = \begin{bmatrix}
\mathbf{Z}_1 \\
\mathbf{Z}_2 \\
\vdots \\
\mathbf{Z}_N
\end{bmatrix}
\]

we obtain the following linear model

\[
(3.3.6) \quad \mathbf{Y} = \mathbf{Z} \psi + \delta.
\]

Note that (3.3.6) is not a standard linear model. Firstly, the design matrix \( \mathbf{Z} \) is
random. Secondly, $\mathbf{Z}$ and $\delta$ are not independent. The error term $\delta$ in (3.3.6) consists of two parts: the projections of the $\varepsilon_n$, and the remainder of an infinite sum. Thus, while (3.3.6) looks like the standard linear model, the existing asymptotic results do not apply to it, and a new asymptotic analysis involving the interplay of the various approximation errors is needed. Representation (3.3.6) leads to the formal “least squares estimator” for $\psi$ is

\begin{equation}
\hat{\psi} = (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \mathbf{Y} = \psi + (\mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T \delta.
\end{equation}

which cannot be computed because the $v_k$ must be replaced by the $\hat{v}_k$.

Now we turn to the effect of replacing the $v_k$ by the $\hat{v}_k$. Projecting onto the $\hat{v}_k$, we are “estimating” the \textit{random} vector

\begin{equation}
\tilde{\psi} = \left[ \hat{c}_1 \psi_{11} \hat{c}_1, \ldots, \hat{c}_1 \psi_{1p} \hat{c}_p, \ldots, \hat{c}_p \psi_{p1} \hat{c}_1, \ldots, \hat{c}_p \psi_{pp} \hat{c}_p \right]^T.
\end{equation}

with the “estimator”

\[ \tilde{\psi}^\wedge = (\mathbf{Z}^T \hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^T \hat{\mathbf{Y}} \]

obtained by replacing the $v_k$ by the $\hat{v}_k$ in (3.3.7). It will be convenient to associate this vector of dimension $p^2$ with the $p \times p$ matrix

\begin{equation}
\tilde{\Psi}_p^\wedge = \begin{bmatrix}
\tilde{\psi}_{11}^\wedge & \tilde{\psi}_{12}^\wedge & \cdots & \tilde{\psi}_{1p}^\wedge \\
\tilde{\psi}_{21}^\wedge & \tilde{\psi}_{22}^\wedge & \cdots & \tilde{\psi}_{2p}^\wedge \\
\vdots & \vdots & \ddots & \vdots \\
\tilde{\psi}_{p1}^\wedge & \tilde{\psi}_{p2}^\wedge & \cdots & \tilde{\psi}_{pp}^\wedge 
\end{bmatrix}.
\end{equation}

It can be shown that if the regularity conditions of Hall and Hosseini-Nasab (2006) hold, then

\begin{equation}
N^{1/2}(\tilde{\psi}^\wedge - \tilde{\psi}) = [\tilde{\mathbf{C}} \otimes \tilde{\mathbf{C}}] N^{1/2}(\hat{\psi} - \psi) + \mathbf{Q}^{-1}(R_{N1} + R_{N2}) + o_P(1),
\end{equation}
where

\[
\hat{C} = \begin{bmatrix}
\hat{c}_1 & 0 & \cdots & 0 \\
0 & \hat{c}_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \hat{c}_p
\end{bmatrix}, \\
Q = I_p \otimes \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_p
\end{bmatrix},
\]

and where \( \otimes \) denotes the Kronecker product of two matrices. The terms \( R_{N1} \) and \( R_{N2} \) are linear functionals of \( N^{-1/2} \sum_{n=1}^{N} X_n(t) \) and \( N^{-1/2} \sum_{n=1}^{N} \{ X_n(t)X_n(s) - E[X_n(t)X_n(s)] \} \).

The limits of \( N^{1/2}(\hat{\psi} - \psi) \), \( R_{N1} \) and \( R_{N2} \) are thus jointly Gaussian, but the asymptotic normality of \( \tilde{\psi}^\wedge - \tilde{\psi} \) does not follow due to the random signs \( \hat{c}_j \). It does however follow from (3.3.10) that \( N^{1/2}(\tilde{\psi}^\wedge - \tilde{\psi}) = O_P(1) \), and this relation does not require the regularity assumptions of Hall and Hosseini-Nasab (2006). The rate \( N^{1/2} \) is optimal, i.e. if \( a_N/N^{1/2} \to \infty \), then \( a_N(\tilde{\psi}^\wedge - \tilde{\psi}) \overset{P}{\to} \infty \). This is exactly the result that will be used in the following, and we state it here as lemma 3.2. We need the following additional assumption.

**Assumption 3.3.** The coefficients \( \psi_{ij} \) of the kernel \( \psi(\cdot, \cdot) \) satisfy \( \sum_{i,j=1}^{\infty} |\psi_{ij}| < \infty \).

**Lemma 3.2.** If Assumptions (A1)–(A5) and 3.3 hold, then \( \tilde{\psi}^\wedge - \tilde{\psi} = O_P(N^{-1/2}) \).

The proof of lemma 3.2 is fairly technical and is developed in Aue et al. (2010).

Relation (3.3.10) shows that replacing the \( v_k \) by the \( \hat{v}_k \) changes the asymptotic distribution. While the limiting distribution of \( \tilde{\psi}^\wedge \) is complex and cannot be used directly, this estimator itself can be used to construct a feasible goodness–of–fit test.

### 3.4 Testing the independence of model errors

We propose two test statistics, (3.4.16) and (3.4.19), which can be used to test the assumption that the errors \( \varepsilon_n \) in (3.1.1) are iid functions in \( L^2 \). These statistics arise from two different ways of defining finite dimensional vectors of residuals. Method I builds on the ideas presented in Section 3.3, the residuals are derived using the
estimator $\hat{\psi}$ obtained by projecting both the $Y_n$ and the $X_n$ on the $\hat{v}_i$, the functional principal components of the regressors. Method II uses two projections. As before, the $X_n$ are projected on the $\hat{v}_i$, but the $Y_n$ are projected on the $\hat{u}_i$. Thus, as in Yao et al. (2005b), we approximate $\psi(\cdot, \cdot)$ by

\begin{equation}
\hat{\psi}_{pq}(t, s) = \sum_{j=1}^{q} \sum_{i=1}^{p} \hat{\lambda}_i^{-1} \hat{\sigma}_{ij} \hat{u}_j(t) \hat{v}_i(s) \quad \hat{\sigma}_{ij} = N^{-1} \sum_{n=1}^{N} \langle X_n, \hat{v}_i \rangle \langle Y_n, \hat{u}_j \rangle.
\end{equation}

Method I emphasizes the role of the regressors $X_n$, and is, in a very loose sense, analogous to the plot of the residuals against the independent variable in a straight line regression. Method II emphasizes the role of the responses, and is somewhat analogous to the plot of the residuals against the fitted values. Both statistics have the form $\sum_{h=1}^{H} \tilde{r}_h T \tilde{\Sigma}^{-1} \tilde{r}_h$, where $\tilde{r}_h$ are vectorized covariance matrices of appropriately constructed residuals, and $\tilde{\Sigma}$ is a suitably constructed matrix which approximates the covariance matrix of the residuals, which are asymptotically iid. As in all procedures of this type, the P-values are computed for a range of values of $H$, typically $H \leq 5$ or $H \leq 10$. The main difficulty, and a central contribution of this paper, is in deriving explicit formulas for the $\tilde{r}_h$ and $\tilde{\Sigma}$ and showing that the test statistics converge to the $\chi^2$ distribution despite a very complex structure of the residuals in the fully functional linear model.

**Method I.** Recall the definition of the matrix $\tilde{\Psi}_p^\wedge$ (3.3.9) whose $(i, j)$ entry approximates $\hat{c}_i \psi_{ij} \hat{c}_j$, and define also $p \times 1$ vectors

$$ \hat{Y}_n = [\hat{Y}_{n1}, \hat{Y}_{n2}, \ldots, \hat{Y}_{np}]^T, \quad \hat{Y}_{nk} = \langle Y_n, \hat{v}_k \rangle; $$

$$ \hat{X}_n = [\hat{\xi}_{n1}, \hat{\xi}_{n2}, \ldots, \hat{\xi}_{np}]^T, \quad \hat{\xi}_{nk} = \langle X_n, \hat{v}_k \rangle. $$

The fitted vectors are then

\begin{equation}
\tilde{Y}_n^\wedge = \tilde{\Psi}_p^\wedge \hat{X}_n, \quad n = 1, 2, \ldots, N;
\end{equation}
and the residuals are $R_n = Y_n - \hat{Y}_n$. For $0 \leq h < N$, define the sample autocovariance matrices of these residuals as

$$
(3.4.14) \quad V_h = N^{-1} \sum_{n=1}^{N-h} R_n R_{n+h}^T.
$$

Finally, by $\text{vec}(V_h)$ denote the column vectors of dimension $p^2$ obtained by stacking the columns of the matrices $V_h$ on top of each other starting from the left. Next, define

$$
\hat{M}_0 = \left[ \frac{1}{N} \sum_{n=1}^{N} e_{nk}^\wedge e_{nk'}^\wedge, \ 1 \leq k, k' \leq p \right],
$$

and

$$
(3.4.15) \quad \hat{M} = \hat{M}_0 \otimes \hat{M}_0.
$$

With this notation in place, we can define the test statistic

$$
(3.4.16) \quad Q_N^\wedge = N \sum_{h=1}^{H} [\text{vec}(V_h)]^T \hat{M}^{-1} \text{vec}(V_h).
$$

Properties of the Kronecker product, $\otimes$, give simplified formulae for $Q_N^\wedge$. Since $\hat{M}^{-1} = \hat{M}_0^{-1} \otimes \hat{M}_0^{-1}$, see Horn and Johnson (1991) p. 244, by Problem 25 on p. 252 of Horn and Johnson (1991), we have

$$
Q_N^\wedge = N \sum_{h=1}^{H} \text{tr} \left[ \hat{M}_0^{-1} V_h^T \hat{M}_0^{-1} V_h \right].
$$
Denoting by \( \hat{m}_{f,h}(i,j) \) and \( \hat{m}_{b,h}(i,j) \) the \((i,j)\) entries, respectively, of \( \hat{M}^{-1}V_h \) and \( V_h\hat{M}^{-1} \), we can write according to the definition of the trace

\[
\hat{Q}_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{p} \hat{m}_{f,h}(i,j)\hat{m}_{b,h}(i,j).
\]

The null hypothesis is rejected if \( \hat{Q}_N \) exceeds an upper quantile of the chi–squared distribution with \( p^2 H \) degrees of freedom, see Theorem 3.3.

**Method II.** Equation (3.1.1) can be rewritten as

\[
\sum_{j=1}^{\infty} \zeta_{nj}u_j = \sum_{i=1}^{\infty} \xi_{ni}\Psi(v_i) + \varepsilon_n,
\]

where \( \Psi \) is the Hilbert–Schmidt operator with kernel \( \psi(\cdot, \cdot) \). To define the residuals, we replace the infinite sums in (3.4.17) by finite sums, the unobservable \( u_j, v_i \) with the \( \hat{u}_j, \hat{v}_i \), and \( \Psi \) with the estimator \( \hat{\Psi}_{pq} \) with kernel (3.4.12). This leads to the equation

\[
\sum_{j=1}^{q} \hat{\zeta}_{nj}\hat{u}_j = \sum_{i=1}^{p} \hat{\xi}_{ni}\hat{\Psi}_{pq}(\hat{v}_i) + \hat{z}_n,
\]

where, similarly as in Section 3.3, \( \hat{z}_n \) contains the \( \varepsilon_n \), the effect of replacing the infinite sums with finite ones, and the effect of the estimation of the eigenfunctions. Method II is based on the residuals defined by

\[
\hat{z}_n = \hat{z}_n(p,q) = \sum_{j=1}^{q} \hat{\zeta}_{nj}\hat{u}_j - \sum_{i=1}^{p} \hat{\xi}_{ni}\hat{\Psi}_{pq}(\hat{v}_i)
\]

Since \( \hat{\Psi}_{pq}(\hat{v}_i) = \sum_{j=1}^{q} \hat{\lambda}_i^{-1}\hat{\sigma}_{ij}\hat{u}_j(t) \), we see that

\[
\hat{z}_n = \sum_{j=1}^{q} \left( \hat{\zeta}_{nj} - \sum_{i=1}^{p} \hat{\xi}_{ni}\hat{\lambda}_i^{-1}\hat{\sigma}_{ij} \right) \hat{u}_j(t).
\]
Next define
\[ \hat{Z}_{nj} := \langle \hat{u}_j, \hat{z}_n \rangle = \hat{\xi}_{nj} - \sum_{i=1}^{p} \hat{\xi}_{ni} \hat{\lambda}^{-1}_i \hat{\sigma}_{ij}. \]

and denote by \( \hat{C}_h \) the \( q \times q \) autocovariance matrix with entries
\[ \hat{c}_h(k, \ell) = \frac{1}{N} \sum_{n=1}^{N-h} \left( \hat{Z}_{nk} - \hat{\mu}_Z(k) \right) \left( \hat{Z}_{n+h,\ell} - \hat{\mu}_Z(\ell) \right), \]

where \( \hat{\mu}_Z(k) = N^{-1} \sum_{n=1}^{N} \hat{Z}_{nk} \). Finally denote by \( \hat{r}_{f,h}(i, j) \) and \( \hat{r}_{b,h}(i, j) \) the \((i, j)\) entries, respectively, of \( \hat{C}_0^{-1} \hat{C}_h \) and \( \hat{C}_h \hat{C}_0^{-1} \).

The null hypothesis is rejected if the statistic

(3.4.19) \[ \hat{Q}_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{q} \hat{r}_{f,h}(i, j) \hat{r}_{b,h}(i, j) \]

exceeds an upper quantile of the chi–squared distribution with \( q^2 H \) degrees of freedom, see Theorem 3.4.

Repeating the arguments in the discussion of Method I, we get the following equivalent expressions for \( \hat{Q}_N \):

\[ \hat{Q}_N = N \sum_{h=1}^{H} \text{tr} \left[ \hat{C}_0^{-1} \hat{C}_h^T \hat{C}_0^{-1} \hat{C}_h \right] \]

and
\[ \hat{Q}_N = N \sum_{h=1}^{H} [\text{vec}(\hat{C}_h)]^T [\hat{C}_0 \otimes \hat{C}_0]^{-1} [\text{vec}(\hat{C}_h)]. \]

Both methods require the selection of \( p \) and \( q \) (Method I, only of \( p \)). We recommend the popular method based on the cumulative percentage of total variability (CPV) calculated as
\[ CPV(p) = \frac{\sum_{k=1}^{p} \hat{\lambda}_k}{\sum_{k=1}^{\infty} \hat{\lambda}_k}, \]

with a corresponding formula for the \( q \). The numbers of eigenfunctions, \( p \) and \( q \), are
chosen as the smallest numbers, $p$ and $q$, such that $CPV(p) \geq 0.85$ and $CPV(q) \geq 0.85$. Other approaches are available as well, including the scree graph, the pseudo-AIC criterion, BIC, cross-validation, etc. All these methods are implemented in the Matlab PACE package developed at the University of California at Davis.

### 3.5 Asymptotic theory

The exact asymptotic $\chi^2$ distributions are obtained only under Assumption 3.2 which, in particular, requires that the $X_n$ be iid. Under Assumption (A1)–(A5), these $\chi^2$ distributions provide only approximations to the true limit distributions. The approximations are however very good, as the simulations in Section 3.6 show; size and power for dependent $X_n$ are the same as for iid $X_n$, within the standard error. Thus, to understand the asymptotic properties of the tests, we first consider their behavior under Assumption 3.2.

Method I is based on the following theorem which is proven in Section 3.8.

**Theorem 3.3.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. Then the statistics $Q_N^{\wedge}$ converges to the $\chi^2$–distribution with $p^2H$ degrees of freedom.

Method II is based on Theorem 3.4 which is proven in Section 3.9. It is analogous to Theorem 1 of Gabrys and Kokoszka (2007), but the observations are replaced by residuals (3.4.18), so a more delicate proof is required.

**Theorem 3.4.** Suppose Assumption 3.2 and condition (3.2.2) hold. Then statistic (3.4.19) converges in distribution to a chi–squared random variable with $q^2H$ degrees of freedom.

We now turn to the case of dependent regressors $X_n$. We focus on Method I. Similar results can be developed to justify the use of Method II, except that the $u_j$
will also be involved. The case of dependent regressors involves the $p \times p$ matrices $bD_h$ with entries

$$D_h(i, j) = \sum_{\ell=p+1}^{\infty} \sum_{k=p+1}^{\infty} \iint v_\ell(s)e_h(s, t)v_k(t)dsdt, \quad 1 \leq i, j \leq p,$$

where

$$e_h(s, t) = E[X_0(s)X_h(t)].$$

**Theorem 3.5.** Suppose Assumptions (A1)–(A5), Assumption 3.3 and condition (3.2.2) hold. Then, for any $h > 0$,

$$N^{-1/2}V_h = N^{-1/2} [\hat{c}_i \hat{c}_j V^*_h(i, j), \ 1 \leq i, j \leq p] + R_{N,p}(h) + o_P(1).$$

The matrices $V^*_h = [V^*_h(i, j), \ 1 \leq i, j \leq p], \ 1 \leq h \leq H$, are jointly asymptotically normal. More precisely,

$$N^{-1/2} \{ \text{vec}(V^*_h - N\hat{b}D_h), \ 1 \leq h \leq H \} \overset{d}{\to} \{ Z_1, Z_2, \ldots, Z_H \},$$

where the $p^2$–dimensional vectors $Z_h$ are iid normal, and coincide with the limits of $N^{-1/2} \text{vec}(V_h)$, if the $X_n$ are independent.

For any $r > 0$, the terms $R_{N,p}(h)$ satisfy,

$$\lim \sup_{p \to \infty} \lim_{N \to \infty} P \{ ||R_{N,p}(h)|| > r \} = 0.$$ (3.5.20)

Theorem 3.5, proven in Section 3.10, justifies using Method I for weakly dependent $X_n$, provided $p$ is so large that the first $p$ FPC $v_k$ explain a large percentage of variance of the $X_n$. To understand why, first notice that $|D_h(i, j)| \leq (\lambda_\ell \lambda_k)_{1/2}$, and since $k, \ell > p$, the eigenvalues $\lambda_\ell, \lambda_k$ are negligible, as for functional data sets encountered in practice the graph of the $\lambda_k$ approaches zero very rapidly. The exact form of $R_{N,p}(h)$ can be reconstructed from matrices $\hat{K}_p, \hat{F}_p, \hat{G}_p$ appearing in Lemmas
3.19–3.21. If $E[X_0(u)X_h(v)] = 0$, all these matrices (and the matrices $bD_h$) vanish. If the $X_n$ are dependent, these matrices do not vanish, but are negligibly small because they all involve coefficients $\psi_{jk}$ with at least one index greater than $p$ multiplied by factors of order $O_P(N^{-1/2})$. In (3.5.20), the limit of $p$ increasing to infinity should not be interpreted literally, but again merely indicates that $p$ is so large that the first $p$ FPC $v_k$ explain a large percentage of variance of the $X_n$.

Our last theorem states conditions under which the test is consistent. The interpretation of the limit as $p \to \infty$ is the same as above. Theorem 3.6 states that for such $p$ and sufficiently large $N$ the test will reject with large probability if $\varepsilon_n$ and $\varepsilon_{n+h}$ are correlated in the subspace spanned by $\{v_i, 1 \leq i \leq p\}$.

**Theorem 3.6.** Suppose Assumptions (B1)–(B4), (A2), (A4), (A5), Assumption 3.3 and condition (3.2.2) hold. Then, for all $R > 0$,

$$\lim_{p \to \infty} \liminf_{N \to \infty} P\{Q_N^h > R\} = 1,$$

provided $E[\langle \varepsilon_0, v_i \rangle \langle \varepsilon_h, v_j \rangle] \neq 0$, for some $1 \leq h \leq H$ and $1 \leq i, j \leq p$.

### 3.6 A simulation study

In this section we report the results of a simulation study performed to assess the empirical size and power of the proposed tests (Method I and Method II) for small to moderate sample sizes. Simulations are based on model (3.1.1). The sample size $N$ takes values ranging from 50 to 500. Both independent and dependent covariate functions are considered. The simulation runs have 1,000 replications each. The simulations are done in the R language, using the fda package.

For the noise component independent trajectories of the Brownian bridge (BB) and the Brownian motion (BM) are generated by transforming cumulative sums of independent normal random variables computed on a grid of 1,000 equispaced points in $[0, 1]$. In order to evaluate the effect of non Gaussian errors on the finite sample
performance, for the noise component we also simulated $t_5$ and uniform BB and BM ($BB_{t_5}$, $BB_U$, $BM_{t_5}$, and $BM_U$) by generating $t_5$ and uniform, instead of normal increments. We also generate errors using Karhunen–Loève expansions

$$
\varepsilon_n(t) = \sum_{j=1}^{5} \vartheta_{nj} j^{-1/2} \sin(j \pi t),
$$

with the iid $\vartheta_{nj}$ distributed according to the normal, $t_5$ and uniform distributions.

We report simulation results obtained using B-spline bases with 20 basis functions, which are suitable for the processes we consider. We also performed the simulations using the Fourier basis and found that they are not significantly different.

To determine the number of principal components ($p$ for $X_n$ and $q$ for $Y_n$), the cumulative percentage of total variability (CPV) is used as described in Section 3.4.

Three different kernel functions in (3.1.1) are considered: the Gaussian kernel

$$
\psi(t, s) = \exp \left\{ \frac{t^2 + s^2}{2} \right\},
$$

the Wiener kernel

$$
\psi(t, s) = \min(t, s),
$$

and the Parabolic kernel

$$
\psi(t, s) = -4 \left[ (t + 1/2)^2 + (s + 1/2)^2 \right] + 2.
$$

The first set of runs under $H_0$ is performed to determine whether for finite sample sizes the procedures achieve nominal 10%, 5%, and 1% levels of significance deduced from the asymptotic distribution. The covariates in (3.1.1) for both methods are either iid BB or BM, or follow the ARH(1) model of Bosq (2000), which has been extensively used to model weak dependence in functional time series data. To simulate the ARH(1) $X_n$ we used the kernels of the three types above, but multiplied by a
constant $K$, so that their Hilbert–Schmidt norm is 0.5. Thus, the dependent regressors follow the model

$$X_n(t) = K \int_0^1 \psi_X(t, s) X_{n-1}(s) ds + \alpha_n(t),$$

where the $\alpha_n$ are iid BB, BM, BB$_{t_5}$, BB$_U$, BM$_{t_5}$, or BM$_U$.

The empirical rejection rates are collected in Tables 3.1 through 3.8: Method I: Tables 3.1 through 3.4 and Method II: Tables 3.5 through 3.8. The tables show that Method I tends to be more conservative and slightly underestimates the nominal levels while Method II tends to overestimate them, especially for $H = 5$. For samples of size 200 or larger, the procedures achieve significance levels close to the true nominal levels. The tables show that the empirical sizes do not depend on whether the BB or the BM was used, nor whether regressors are iid or dependent, nor on the shape of the kernel. These sizes do not deteriorate if errors are not Gaussian either. This shows that the empirical size of both methods is robust to the form of the kernel, to moderate dependence in the regressors, and to departures from normality in the errors.

For the power simulations, we consider model (3.1.1) with the Gaussian kernel and $\varepsilon_n \sim \text{ARH}(1)$, i.e.

$$\varepsilon_n(t) = K \int_0^1 \psi_\varepsilon(t, s) \varepsilon_{n-1}(s) ds + u_n(t),$$

where $\psi_\varepsilon(t, s)$ is Gaussian, Wiener or Parabolic and $K$ is chosen so that the Hilbert–Schmidt norm of the above ARH(1) operator is 0.5 and the $u_n(t)$ are iid BB, BM, BB$_{t_5}$, BB$_U$, BM$_{t_5}$, or BM$_U$. Empirical power for all sample sizes considered in the simulation study and for all departures from the model assumptions is summarized in a series of tables: Method I: Tables 3.9 through 3.11, Method II: Tables 3.12 through 3.14. To conserve space results are presented for $\psi = \text{Gaussian}$ and $\psi_\varepsilon = \psi_X = \text{Gaussian}$, Wiener and Parabolic. For Method I, $\varepsilon_n = \text{BB}$ gives slightly higher power than using the BM. For sample sizes $N = 50$ and 100 Method II dominates Method
I, but starting with samples of 200 or larger both methods give very high power for both Gaussian and non-Gaussian innovations. Simulations show that the power is not affected on whether regressors are iid or dependent. From the tables, we observe that the power is highest for lag $H = 1$, especially for smaller samples, because the errors follow the ARH(1) process.

3.7 Application to space physics and high–frequency financial data

We now illustrate the application of the tests on functional data sets arising in space physics and finance.

**Application to Magnetometer data.** Electrical currents flowing in the magnetosphere-ionosphere (M-I) form a complex multiscale system in which a number of individual currents connect and influence each other. Among the various observational means, the global network of ground-based magnetometers stands out with unique strengths of global spacial coverage and real time fine resolution temporal coverage. About a hundred terrestrial geomagnetic observatories form a network, INTERMAGNET, designed to monitor the variations of the M-I current system. Digital magnetometers record three components of the magnetic field in five second resolution, but the INTERMAGNET’s data we use consist of one minute averages, i.e. 1440 data points per day per component per observatory. Due to the daily rotation of the Earth, we split magnetometer records into days, and treat each daily curve as a single functional observation. We consider the Horizontal (H) component of the magnetic field, lying in the Earth’s tangent plane and pointing toward the magnetic North. It most directly reflects the variation of the M-I currents we wish to study.

The problem that motivated the examples in this section is that of the association between the auroral (high latitude) electrical currents and the currents flowing at mid– and low latitudes.
Table 3.1: Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$\psi = \text{Gaussian}$</th>
<th>$\psi = \text{Wiener}$</th>
<th>$\psi = \text{Parabolic}$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$p = 3$</td>
<td>$p = 3$</td>
<td>$p = 3$</td>
</tr>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>6.7 2.5 0.1</td>
<td>5.8 3.2 0.3</td>
<td>7.4 3.7 0.1</td>
</tr>
<tr>
<td>100</td>
<td>7.4 3.7 0.7</td>
<td>9.5 4.4 0.8</td>
<td>8.9 3.8 0.6</td>
</tr>
<tr>
<td>200</td>
<td>9.8 4.6 0.9</td>
<td>8.9 4.2 0.4</td>
<td>9.0 4.1 0.5</td>
</tr>
<tr>
<td>300</td>
<td>9.3 4.8 1.2</td>
<td>10.0 5.1 0.5</td>
<td>8.1 3.5 0.7</td>
</tr>
<tr>
<td>500</td>
<td>8.8 5.2 1.1</td>
<td>9.8 5.3 1.1</td>
<td>9.6 4.9 1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.3 2.5 0.1</td>
<td>5.6 2.1 0.5</td>
<td>6.0 3.4 0.2</td>
</tr>
<tr>
<td>100</td>
<td>7.6 3.7 0.5</td>
<td>6.9 3.6 0.6</td>
<td>6.4 3.3 0.5</td>
</tr>
<tr>
<td>200</td>
<td>8.7 4.6 0.6</td>
<td>6.4 3.2 0.7</td>
<td>8.0 3.3 0.8</td>
</tr>
<tr>
<td>300</td>
<td>7.6 3.5 0.7</td>
<td>9.5 4.2 1.2</td>
<td>9.5 4.8 0.5</td>
</tr>
<tr>
<td>500</td>
<td>9.8 4.6 1.4</td>
<td>9.1 3.9 0.9</td>
<td>9.2 4.9 0.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.6 0.9 0.1</td>
<td>3.5 1.1 0.1</td>
<td>4.1 1.4 0.1</td>
</tr>
<tr>
<td>100</td>
<td>6.5 3.7 0.8</td>
<td>5.9 3.0 0.6</td>
<td>4.8 1.9 0.1</td>
</tr>
<tr>
<td>200</td>
<td>8.5 4.4 1.3</td>
<td>7.5 3.7 0.8</td>
<td>7.4 3.3 0.2</td>
</tr>
<tr>
<td>300</td>
<td>7.6 4.0 0.6</td>
<td>9.9 4.7 1.0</td>
<td>7.6 2.8 0.3</td>
</tr>
<tr>
<td>500</td>
<td>10.1 4.6 1.0</td>
<td>9.8 4.4 1.1</td>
<td>7.9 3.6 0.3</td>
</tr>
</tbody>
</table>
Table 3.2: Method I: Empirical size for independent predictors: $X = BB$, $\varepsilon = BB_t$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi =$ Gaussian</td>
<td>$\psi =$ Wiener</td>
<td>$\psi =$ Parabolic</td>
</tr>
<tr>
<td></td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7.4 3.4 0.2</td>
<td>6.4 2.0 0.0</td>
<td>6.4 2.5 0.2</td>
</tr>
<tr>
<td>100</td>
<td>8.7 4.2 0.3</td>
<td>5.8 2.8 0.6</td>
<td>9.1 4.3 1.0</td>
</tr>
<tr>
<td>200</td>
<td>8.2 3.2 0.7</td>
<td>9.5 4.2 0.8</td>
<td>8.5 4.1 0.9</td>
</tr>
<tr>
<td>300</td>
<td>8.8 4.0 0.5</td>
<td>9.3 4.7 0.6</td>
<td>9.2 5.6 1.1</td>
</tr>
<tr>
<td>500</td>
<td>8.6 4.0 0.4</td>
<td>11.0 5.6 1.2</td>
<td>8.8 3.8 1.2</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.1 1.7 0.2</td>
<td>4.4 1.4 0.3</td>
<td>4.3 1.2 0.3</td>
</tr>
<tr>
<td>100</td>
<td>7.0 3.1 0.8</td>
<td>6.8 2.4 0.2</td>
<td>5.8 3.3 0.2</td>
</tr>
<tr>
<td>200</td>
<td>7.3 3.4 1.0</td>
<td>11.0 5.6 1.2</td>
<td>8.0 3.5 0.7</td>
</tr>
<tr>
<td>300</td>
<td>10.1 5.0 0.5</td>
<td>8.9 3.4 1.0</td>
<td>9.8 4.0 0.7</td>
</tr>
<tr>
<td>500</td>
<td>10.8 6.4 1.1</td>
<td>9.2 5.7 1.2</td>
<td>10.2 5.5 1.2</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>3.8 0.7 0.0</td>
<td>4.4 2.5 1.1</td>
<td>3.5 1.5 0.4</td>
</tr>
<tr>
<td>100</td>
<td>5.4 2.4 0.3</td>
<td>4.8 1.9 0.2</td>
<td>5.1 2.7 0.8</td>
</tr>
<tr>
<td>200</td>
<td>10.2 4.5 1.0</td>
<td>6.9 3.8 0.8</td>
<td>7.1 3.4 0.5</td>
</tr>
<tr>
<td>300</td>
<td>10.1 5.1 1.2</td>
<td>9.4 4.2 0.7</td>
<td>9.3 4.7 0.9</td>
</tr>
<tr>
<td>500</td>
<td>10.2 5.1 1.3</td>
<td>10.3 5.1 1.3</td>
<td>8.6 3.7 1.0</td>
</tr>
</tbody>
</table>
Table 3.3: Method I: Empirical size for independent regressor functions: $X_n = BB_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j \pi t)$, $n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0, 1)$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi =$ Gaussian</td>
<td>$\psi =$ Wiener</td>
<td>$\psi =$ Parabolic</td>
<td></td>
</tr>
<tr>
<td>$10%$</td>
<td>$5%$</td>
<td>$1%$</td>
<td>$10%$</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>6.5</td>
<td>1.9</td>
<td>0.2</td>
</tr>
<tr>
<td>100</td>
<td>9.1</td>
<td>4.6</td>
<td>0.7</td>
</tr>
<tr>
<td>200</td>
<td>9.5</td>
<td>4.8</td>
<td>1.3</td>
</tr>
<tr>
<td>300</td>
<td>8.1</td>
<td>3.6</td>
<td>1.2</td>
</tr>
<tr>
<td>500</td>
<td>9.3</td>
<td>4.4</td>
<td>0.7</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.0</td>
<td>1.2</td>
<td>0.0</td>
</tr>
<tr>
<td>100</td>
<td>6.9</td>
<td>3.2</td>
<td>1.0</td>
</tr>
<tr>
<td>200</td>
<td>10.0</td>
<td>5.5</td>
<td>0.8</td>
</tr>
<tr>
<td>300</td>
<td>10.1</td>
<td>4.7</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>7.9</td>
<td>4.2</td>
<td>1.0</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>2.9</td>
<td>1.4</td>
<td>0.1</td>
</tr>
<tr>
<td>100</td>
<td>5.5</td>
<td>2.2</td>
<td>0.3</td>
</tr>
<tr>
<td>200</td>
<td>9.3</td>
<td>4.6</td>
<td>0.5</td>
</tr>
<tr>
<td>300</td>
<td>7.1</td>
<td>3.3</td>
<td>0.7</td>
</tr>
<tr>
<td>500</td>
<td>9.9</td>
<td>5.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Table 3.4: Method I: Empirical size for dependent predictors: $X \sim ARH(1)$ with the BB innovations, $\psi =$Gaussian, $\varepsilon = BB$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_X =$ Gaussian</td>
<td>$\psi_X =$ Wiener</td>
<td>$\psi_X =$ Parabolic</td>
</tr>
<tr>
<td>$H = 1$</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>50</td>
<td>8.4</td>
<td>3.9</td>
<td>0.3</td>
</tr>
<tr>
<td>100</td>
<td>8.9</td>
<td>4.4</td>
<td>0.7</td>
</tr>
<tr>
<td>200</td>
<td>10.2</td>
<td>4.7</td>
<td>0.9</td>
</tr>
<tr>
<td>300</td>
<td>9.2</td>
<td>4.9</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>10.5</td>
<td>5.2</td>
<td>1.4</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.4</td>
<td>2.2</td>
<td>0.3</td>
</tr>
<tr>
<td>100</td>
<td>6.6</td>
<td>3.1</td>
<td>0.3</td>
</tr>
<tr>
<td>200</td>
<td>7.8</td>
<td>3.1</td>
<td>0.5</td>
</tr>
<tr>
<td>300</td>
<td>8.2</td>
<td>4.8</td>
<td>0.7</td>
</tr>
<tr>
<td>500</td>
<td>11.4</td>
<td>5.3</td>
<td>1.5</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>4.2</td>
<td>1.8</td>
<td>0.1</td>
</tr>
<tr>
<td>100</td>
<td>7.2</td>
<td>3.2</td>
<td>0.6</td>
</tr>
<tr>
<td>200</td>
<td>7.6</td>
<td>2.8</td>
<td>0.9</td>
</tr>
<tr>
<td>300</td>
<td>8.3</td>
<td>4.2</td>
<td>0.6</td>
</tr>
<tr>
<td>500</td>
<td>10.7</td>
<td>5.8</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Table 3.5: Method II: Empirical size for independent predictors: $X = BB, \varepsilon = BB$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3, q = 2$</th>
<th>$p = 3, q = 3$</th>
<th>$p = 3, q = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi =$ Gaussian</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$\psi =$ Wiener</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$\psi =$ Parabolic</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7.9</td>
<td>3.7</td>
<td>0.4</td>
</tr>
<tr>
<td>100</td>
<td>10.6</td>
<td>5.2</td>
<td>1.4</td>
</tr>
<tr>
<td>200</td>
<td>8.9</td>
<td>4.4</td>
<td>0.9</td>
</tr>
<tr>
<td>300</td>
<td>8.7</td>
<td>4.4</td>
<td>0.5</td>
</tr>
<tr>
<td>500</td>
<td>8.8</td>
<td>4.2</td>
<td>1.1</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10.7</td>
<td>5.3</td>
<td>0.9</td>
</tr>
<tr>
<td>100</td>
<td>9.9</td>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>200</td>
<td>9.6</td>
<td>4.8</td>
<td>0.9</td>
</tr>
<tr>
<td>300</td>
<td>11.0</td>
<td>5.1</td>
<td>1.1</td>
</tr>
<tr>
<td>500</td>
<td>11.1</td>
<td>6.8</td>
<td>1.3</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10.4</td>
<td>5.7</td>
<td>1.1</td>
</tr>
<tr>
<td>100</td>
<td>11.3</td>
<td>5.3</td>
<td>1.1</td>
</tr>
<tr>
<td>200</td>
<td>11.3</td>
<td>5.7</td>
<td>1.1</td>
</tr>
<tr>
<td>300</td>
<td>9.4</td>
<td>4.9</td>
<td>0.5</td>
</tr>
<tr>
<td>500</td>
<td>12.1</td>
<td>6.8</td>
<td>1.2</td>
</tr>
</tbody>
</table>
Table 3.6: Method II: Empirical size for independent predictors: $X = BB, \varepsilon = BB_t$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3, q = 2$</th>
<th>$p = 3, q = 3$</th>
<th>$p = 3, q = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\psi = \text{Gaussian}$</td>
<td>$\psi = \text{Wiener}$</td>
<td>$\psi = \text{Parabolic}$</td>
<td></td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>8.3</td>
<td>3.8</td>
<td>0.6</td>
</tr>
<tr>
<td>100</td>
<td>9.6</td>
<td>3.7</td>
<td>0.8</td>
</tr>
<tr>
<td>200</td>
<td>8.1</td>
<td>3.9</td>
<td>1.0</td>
</tr>
<tr>
<td>300</td>
<td>10.7</td>
<td>5.5</td>
<td>1.5</td>
</tr>
<tr>
<td>500</td>
<td>11.6</td>
<td>5.6</td>
<td>1.3</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>8.6</td>
<td>3.4</td>
<td>0.5</td>
</tr>
<tr>
<td>100</td>
<td>9.7</td>
<td>4.9</td>
<td>0.6</td>
</tr>
<tr>
<td>200</td>
<td>8.9</td>
<td>5.6</td>
<td>1.3</td>
</tr>
<tr>
<td>300</td>
<td>11.</td>
<td>5.6</td>
<td>1.0</td>
</tr>
<tr>
<td>500</td>
<td>10.</td>
<td>6.3</td>
<td>0.8</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>10.9</td>
<td>5.7</td>
<td>1.9</td>
</tr>
<tr>
<td>100</td>
<td>10.6</td>
<td>6.0</td>
<td>1.4</td>
</tr>
<tr>
<td>200</td>
<td>10.6</td>
<td>6.2</td>
<td>1.2</td>
</tr>
<tr>
<td>300</td>
<td>10.5</td>
<td>5.5</td>
<td>1.0</td>
</tr>
<tr>
<td>500</td>
<td>10.6</td>
<td>5.0</td>
<td>0.9</td>
</tr>
</tbody>
</table>
Table 3.7: Method II: Empirical size for independent regressor functions: $X_n = B B_n$, $\varepsilon_n = \sum_{j=1}^{5} \vartheta_{nj} \cdot j^{-1/2} \cdot \sin(j \pi t), n = 1, \ldots, N$, $\vartheta_{nj} \sim N(0,1)$.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$, $q = 4$</th>
<th>$p = 3$, $q = 4$</th>
<th>$p = 3$, $q = 4$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi$ = Gaussian</td>
<td>$\psi$ = Wiener</td>
<td>$\psi$ = Parabolic</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>7.2</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>100</td>
<td>9.3</td>
<td>4.6</td>
<td>0.5</td>
</tr>
<tr>
<td>200</td>
<td>10.1</td>
<td>4.3</td>
<td>0.5</td>
</tr>
<tr>
<td>300</td>
<td>9.2</td>
<td>4.1</td>
<td>0.6</td>
</tr>
<tr>
<td>500</td>
<td>9.3</td>
<td>4.3</td>
<td>0.7</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>8.3</td>
<td>3.8</td>
<td>0.5</td>
</tr>
<tr>
<td>100</td>
<td>10.0</td>
<td>5.2</td>
<td>0.9</td>
</tr>
<tr>
<td>200</td>
<td>9.5</td>
<td>4.0</td>
<td>0.9</td>
</tr>
<tr>
<td>300</td>
<td>7.9</td>
<td>4.0</td>
<td>0.8</td>
</tr>
<tr>
<td>500</td>
<td>9.6</td>
<td>4.9</td>
<td>1.1</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>13.7</td>
<td>7.0</td>
<td>2.3</td>
</tr>
<tr>
<td>100</td>
<td>12.7</td>
<td>6.2</td>
<td>1.1</td>
</tr>
<tr>
<td>200</td>
<td>10.7</td>
<td>5.1</td>
<td>0.8</td>
</tr>
<tr>
<td>300</td>
<td>10.1</td>
<td>4.5</td>
<td>1.0</td>
</tr>
<tr>
<td>500</td>
<td>9.5</td>
<td>4.7</td>
<td>0.6</td>
</tr>
</tbody>
</table>
Table 3.8: Method II: Empirical size for dependent predictor functions: \( X \sim ARH(1) \) with the BB innovations, \( \psi = \text{Gaussian}, \varepsilon = BB \).

<table>
<thead>
<tr>
<th>Sample size</th>
<th>( p = 3 )</th>
<th>( p = 3 )</th>
<th>( p = 3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \psi_X = \text{Gaussian} )</td>
<td>( \psi_X = \text{Wiener} )</td>
<td>( \psi_X = \text{Parabolic} )</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>( H = 1 )</td>
<td>50</td>
<td>9.2</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>10.4</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>9.5</td>
<td>4.8</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>10.1</td>
<td>4.1</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.0</td>
<td>4.2</td>
</tr>
<tr>
<td>( H = 2 )</td>
<td>50</td>
<td>9.9</td>
<td>5.2</td>
</tr>
<tr>
<td></td>
<td>100</td>
<td>10.5</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>200</td>
<td>11.4</td>
<td>4.6</td>
</tr>
<tr>
<td></td>
<td>300</td>
<td>10.7</td>
<td>5.5</td>
</tr>
<tr>
<td></td>
<td>500</td>
<td>9.0</td>
<td>4.1</td>
</tr>
</tbody>
</table>
Table 3.9: Method I: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th></th>
<th>$p = 3$</th>
<th></th>
<th>$p = 3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_\varepsilon = $ Gaussian</td>
<td>$\psi_\varepsilon = $ Wiener</td>
<td>$\psi_\varepsilon = $ Parabolic</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$H = 1$</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>50</td>
<td>84.3</td>
<td>75.3</td>
<td>46.0</td>
<td>53.7</td>
<td>36.7</td>
<td>12.9</td>
</tr>
<tr>
<td>100</td>
<td>99.9</td>
<td>99.7</td>
<td>98.0</td>
<td>96.1</td>
<td>92.2</td>
<td>76.3</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.7</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>50</td>
<td>60.4</td>
<td>46.2</td>
<td>24.0</td>
<td>35.7</td>
<td>25.3</td>
<td>9.5</td>
</tr>
<tr>
<td>100</td>
<td>97.9</td>
<td>96.9</td>
<td>88.9</td>
<td>83.9</td>
<td>75.7</td>
<td>54.8</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
<td>99.5</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>50</td>
<td>43.2</td>
<td>32.4</td>
<td>15.3</td>
<td>24.5</td>
<td>16.1</td>
<td>6.3</td>
</tr>
<tr>
<td>100</td>
<td>94.6</td>
<td>90.5</td>
<td>75.6</td>
<td>72.4</td>
<td>61.5</td>
<td>42.2</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.8</td>
<td>99.2</td>
<td>98.0</td>
<td>94.4</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.10: Method I: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB$t_5$ innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_x = \text{Gaussian}$</td>
<td>$\psi_x = \text{Wiener}$</td>
<td>$\psi_x = \text{Parabolic}$</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>85.1</td>
<td>73.6</td>
<td>46.1</td>
</tr>
<tr>
<td>100</td>
<td>99.7</td>
<td>99.7</td>
<td>98.0</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>60.7</td>
<td>47.6</td>
<td>24.1</td>
</tr>
<tr>
<td>100</td>
<td>98.7</td>
<td>96.5</td>
<td>88.8</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>40.8</td>
<td>29.8</td>
<td>13.8</td>
</tr>
<tr>
<td>100</td>
<td>95.0</td>
<td>91.1</td>
<td>76.6</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.11: Method I: Empirical power for dependent predictor functions: $X \sim ARH(1)$ with the BB innovations, $\varepsilon \sim ARH(1)$ with the BB innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_\varepsilon = \psi_X = \text{Gaussian}$</td>
<td>$\psi_\varepsilon = \psi_X = \text{Wiener}$</td>
<td>$\psi_\varepsilon = \psi_X = \text{Parabolic}$</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>79.2</td>
<td>68.6</td>
<td>40.1</td>
</tr>
<tr>
<td>100</td>
<td>99.9</td>
<td>99.6</td>
<td>97.9</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>53.8</td>
<td>40.7</td>
<td>19.8</td>
</tr>
<tr>
<td>100</td>
<td>98.0</td>
<td>95.7</td>
<td>87.2</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>41.2</td>
<td>27.9</td>
<td>12.3</td>
</tr>
<tr>
<td>100</td>
<td>95.1</td>
<td>90.3</td>
<td>76.4</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.12: Method II: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3, q = 2$</th>
<th>$p = 3, q = 3$</th>
<th>$p = 3, q = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_\varepsilon = $ Gaussian</td>
<td>$\psi_\varepsilon = $ Wiener</td>
<td>$\psi_\varepsilon = $ Parabolic</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>86.1</td>
<td>78.6</td>
<td>57.1</td>
</tr>
<tr>
<td>100</td>
<td>99.2</td>
<td>98.6</td>
<td>95.4</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>74.4</td>
<td>63.4</td>
<td>43.1</td>
</tr>
<tr>
<td>100</td>
<td>97.6</td>
<td>94.4</td>
<td>89.3</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>66.3</td>
<td>55.9</td>
<td>34.3</td>
</tr>
<tr>
<td>100</td>
<td>95.4</td>
<td>92.6</td>
<td>82.2</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>99.6</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.13: Method II: Empirical power for independent predictors: $X = BB$, $\varepsilon \sim ARH(1)$ with the BB$_{t_5}$ innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3, q = 2$</th>
<th>$p = 3, q = 3$</th>
<th>$p = 3, q = 2$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_x = \text{Gaussian}$</td>
<td>$\psi_x = \text{Wiener}$</td>
<td>$\psi_x = \text{Parabolic}$</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>83.2</td>
<td>72.3</td>
<td>50.9</td>
</tr>
<tr>
<td>100</td>
<td>99.4</td>
<td>97.3</td>
<td>93.1</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>70.8</td>
<td>59.8</td>
<td>36.9</td>
</tr>
<tr>
<td>100</td>
<td>95.2</td>
<td>92.6</td>
<td>83.9</td>
</tr>
<tr>
<td>200</td>
<td>99.9</td>
<td>99.9</td>
<td>99.9</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>50</td>
<td>62.4</td>
<td>51.2</td>
<td>31.7</td>
</tr>
<tr>
<td>100</td>
<td>93.8</td>
<td>88.0</td>
<td>75.1</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>99.0</td>
<td>99.4</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
Table 3.14: Method II: Empirical power for dependent predictors: $X \sim ARH(1)$ with the BB innovations; $\varepsilon = BB \sim ARH(1)$ with the BB innovations.

<table>
<thead>
<tr>
<th>Sample size</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
<th>$p = 3$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\psi_\varepsilon = \psi_X = \text{Gaussian}$</td>
<td>$\psi_\varepsilon = \psi_X = \text{Wiener}$</td>
<td>$\psi_\varepsilon = \psi_X = \text{Parabolic}$</td>
</tr>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$H = 1$</td>
<td>86.0</td>
<td>77.5</td>
<td>57.4</td>
</tr>
<tr>
<td>50</td>
<td>99.7</td>
<td>98.9</td>
<td>95.5</td>
</tr>
<tr>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 3$</td>
<td>73.8</td>
<td>61.1</td>
<td>40.4</td>
</tr>
<tr>
<td>50</td>
<td>96.8</td>
<td>95.3</td>
<td>90.1</td>
</tr>
<tr>
<td>100</td>
<td>99.9</td>
<td>99.9</td>
<td>99.8</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$H = 5$</td>
<td>65.8</td>
<td>56.2</td>
<td>36.3</td>
</tr>
<tr>
<td>50</td>
<td>95.0</td>
<td>91.7</td>
<td>83.3</td>
</tr>
<tr>
<td>100</td>
<td>99.9</td>
<td>99.8</td>
<td>99.5</td>
</tr>
<tr>
<td>200</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>300</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>500</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
This problem was studied in Maslova et al. (2010), who provide more extensive references to the relevant space physics literature, and to a smaller extent in Kokoszka et al. (2008) and Horváth et al. (2009b). The problem has been cast into the setting of the functional linear model (3.1.1) in which the $X_n$ are centered high-latitude records and $Y_n$ are centered mid- or low-latitude magnetometer records. We consider two settings 1) consecutive days, 2) non-consecutive days on which disturbances known as substorms occur. For consecutive days, we expect the rejection of the null hypothesis as there is a visible dependence of the responses from one day to the next, see the bottom panel of Figure 3.2. The low latitude curves, like those measured at Honolulu, exhibit changes on scales of several days. The high latitude curves exhibit much shorter dependence essentially confined to one day. This is because the auroral electrojets change on a scale of about 4 hours. In setting 2, the answer is less clear: the substorm days are chronologically arranged, but substorms may be separated by several days, and after each substorm the auroral current system resets itself to a quiet state.

To apply the tests, we converted the data to functional objects using 20 spline basis functions, and computed the EFPC’s $\hat{v}_k$ and $\hat{u}_j$. For low latitude magnetometer data, 2 or 3 FPC’s are needed to explain $87 - 89$, or $92 - 94$, percent of variability while for high latitude stations to explain $88 - 91$ percent of variability we need $8 - 9$ FPC’s.

Setting 1 (consecutive days): We applied both methods to pairs $(X_n, Y_n)$ in which the $X_n$ are daily records at College, Alaska, and the $Y_n$ are the corresponding records at six equatorial stations. Ten such pairs are shown in Figure 3.2. The samples consisted of all days in 2001, and of about 90 days corresponding to the four seasons. For all six stations and for the whole year the p-values were smaller than $10^{-12}$. For the four seasons, all p-values, except two, were smaller than 2%. The higher p-values for the samples restricted to 90 days, are likely due to a smaller seasonal effect (the structure of the M-I system in the northern hemisphere changes with season).
Fig. 3.2: Magnetometer data on 10 consecutive days (separated by vertical dashed lines) recorded at College, Alaska (CMO) and Honolulu, Hawaii, (HON).
Table 3.15: Isolated substorms data. P-values in percent.

<table>
<thead>
<tr>
<th>Method</th>
<th>Response</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HON</td>
</tr>
<tr>
<td>I</td>
<td>9.80</td>
</tr>
<tr>
<td>II</td>
<td>6.57</td>
</tr>
</tbody>
</table>

We conclude that it is not appropriate to use model (3.1.1) with iid errors to study the interaction of high- and low latitude currents when the data are derived from consecutive days.

Setting 2 (substorm days): We now focus on two samples studied in Maslova et al. (2010). They are derived from 37 days on which isolated substorms were recorded at College, Alaska (CMO). A substorm is classified as an isolated substorm, if it is followed by 2 quiet days. There were only 37 isolated substorms in 2001, data for 10 such days are shown in Figure 3.3. The first sample consists of 37 pairs \((X_n, Y_n)\), where \(X_n\) is the curve of the \(n\)th isolated storm recorded at CMO, and \(Y_n\) is the curve recorded on the same UT day at Honolulu, Hawaii, (HON). The second sample is constructed in the same way, except that \(Y_n\) is the curve recorded at Boulder, Colorado (BOU). The Boulder observatory is located in geomagnetic midlatitude, i.e. roughly half way between the magnetic north pole and the magnetic equator. Honolulu is located very close to the magnetic equator.

The p-values for both methods and the two samples are listed in Table 3.15. For Honolulu, both tests indicate the suitability of model (3.1.1) with iid errors. For Boulder, the picture is less clear. The acceptance by Method I may be due to the small sample size \((N = 37)\). The simulations in Section 3.6 show that for \(N = 50\) this method has the power of about 50\% at the nominal level of 5\%. On the other hand, Method II has the tendency to overreject. The sample with the Boulder records as responses confirms the general behavior of the two methods observed in Section 3.6, and emphasizes that it is useful to apply both of them to obtain more reliable conclusions. From the space physics perspective, midlatitude records are very difficult to
Fig. 3.3: Magnetometer data on 10 chronologically arranged isolated substorm days recorded at College, Alaska (CMO), Honolulu, Hawaii, (HON) and Boulder, Colorado (BOU).
interpret because they combine features of high latitude events (exceptionally strong auroras have been seen as far south as Virginia) and those of low latitude and field aligned currents.

We also applied the tests to samples in which the regressors are curves on days on which different types of substorms occurred, and the responses are the corresponding curves at low altitude stations. The general conclusion is that for substorm days, the errors in model (3.1.1) can be assumed iid if the period under consideration is not longer than a few months. For longer periods, seasonal trends cause differences in distribution.

**Application to intradaily returns.** Perhaps the best known application of linear regression to financial data is the celebrated Capital Asset Pricing Model (CAMP), see e.g. Chapter 5 of Campbell et al. (1997). In its simplest form, it is defined by

\[ r_n = \alpha + \beta r_{m,n} + \varepsilon_n, \]

where

\[ r_n = 100(\ln P_n - \ln P_{n-1}) \approx 100 \frac{P_n - P_{n-1}}{P_{n-1}} \]

is the return, in percent, over a unit of time on a specific asset, e.g. a stock of a corporation, and \( r_{m,n} \) is the analogously defined return on a relevant market index. The unit of time can be can be day, month or year.

In this section we work with intra–daily price data, which are known to have properties quite different than those of daily or monthly closing prices, see e.g. Chapter 5 of Tsay (2005), Guillaume et al. (1997), and Andersen and Bollerslev (1997a, 1997b) also offer interesting perspectives. For these data, \( P_n(t_j) \) is the price on day \( n \) at tick \( t_j \) (time of trade); we do not discuss issues related to the bid–ask spread, which are not relevant to what follows. For such data, it is not appropriate to define returns by looking at price movements between the ticks because that would lead to very noisy trajectories for which the methods discussed in this paper, based
on the FPC's, are not appropriate; Johnstone and Lu (2009) explain why principal components cannot be meaningfully estimated for such data. Instead, we adopt the following definition.

**Definition 1** Suppose $P_n(t_j), n = 1, \ldots, N, j = 1, \ldots, m$, is the price of a financial asset at time $t_j$ on day $n$. We call the functions

$$r_n(t_j) = 100[\ln P_n(t_j) - \ln P_n(t_1)], \quad j = 2, \ldots, m, \quad n = 1, \ldots, N,$$

the *intra-daily cumulative returns*.

Figure 3.4 shows intra-daily cumulative returns on 10 consecutive days for the Standard & Poor's 100 index and the Exxon Mobil Corporation. These returns have an appearance amenable to smoothing via FPC's.

We propose an extension of the CAPM to such return by postulating that

$$(3.7.21) \quad r_n(t) = \alpha(t) + \int \beta(t, s)r_{m,n}(s)ds + \varepsilon_n(t), \quad t \in [0, 1],$$

where the interval $[0, 1]$ is the rescaled trading period (in our examples, 9:30 to 16:00 EST). We refer to model (3.7.21) as the functional CAPM (FCAPM). As far as we know, this model has not been considered in the financial literature, but just as for the classical CAPM, it is designed to evaluate the extent to which intradaily market returns determine the intra–daily returns on a specific asset. It is not our goal in this example to systematically estimate the parameters in (3.7.21) and compare them for various assets and markets, we merely want to use the methods developed in this paper to see if this model can be assumed to hold for some well–known asset. With this goal in mind, we considered FCAPM for S&P 100 and its major component, the Exxon Mobil Corporation (currently it contributes 6.78% to this index). The price processes over the period of about 8 years are shown in Figure 3.5. The functional observations are however not these processes, but the cumulative intra–daily returns, examples of which are shown in Figure 3.4.
Fig. 3.4: Intra-daily cumulative returns on 10 consecutive days for the Standard & Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM).
Fig. 3.5: Share prices of the Standard & Poor’s 100 index (SP) and the Exxon–Mobil Corporation (XOM). Dashed lines separate years.
After some initial data cleaning and preprocessing steps, we could compute the p-values for any period within the time stretch shown in Figure 3.5. The p-values for calendar years, the sample size \( N \) is equal to about 250, are reported in Table 3.16. In this example, both methods lead to the same conclusions, which match the well-known macroeconomic background. The tests do not indicate departures from the FCAMP model, except in 2002, the year between September 11 attacks and the invasion of Iraq, and in 2006 and 2007, the years preceding the collapse of 2008 in which oil prices were growing at a much faster rate than then the rest of the economy.

In the above examples we tested the correlation of errors in model (3.1.1). A special case of this model is the historical functional model of Malfait and Ramsay (2003), i.e. model (3.1.1) with \( \psi(t, s) = \beta(s, t)I_H(s, t) \), where \( \beta(\cdot, \cdot) \) is an arbitrary Hilbert–Schmidt kernel and \( I_H(\cdot, \cdot) \) is the indicator function of the set \( H = \{(s, t) : 0 \leq s \leq t \leq 1\} \). This model requires that \( Y_n(t) \) depends only on the values of \( Y_n(s) \) for \( s \leq t \), i.e. it postulates temporal causality within the pairs of curves. Our approach cannot be readily extended to test for error correlation in the historical model of Malfait and Ramsay (2003) because it uses series expansions of a general kernel \( \psi(t, s) \), and the restriction that the kernel vanishes in the complement of \( H \) does not translate to any obvious restrictions on the coefficients of these expansions. We note however that the magnetometer data are obtained at locations with different local times, and for space physics applications the dependence between the shapes of the daily curves is of importance. Temporal causality for financial data is often not assumed as asset values reflect both historical returns and expectations of future market conditions.

### 3.8 Proof of Theorem 3.3

Relation (3.3.5) can be rewritten as

\[
Y_n = \Psi_p X_n + bDg_n,
\]
Table 3.16: P–values, in percent, for the FCAPM (3.7.21) in which the regressors are the intra–daily cumulative returns on the Standard & Poor’s 100 index, and the responses are such returns on the Exxon–Mobil stock.

<table>
<thead>
<tr>
<th>Year</th>
<th>Method I</th>
<th>Method II</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td>46.30</td>
<td>55.65</td>
</tr>
<tr>
<td>2001</td>
<td>43.23</td>
<td>56.25</td>
</tr>
<tr>
<td>2002</td>
<td>0.72</td>
<td>0.59</td>
</tr>
<tr>
<td>2003</td>
<td>22.99</td>
<td>27.19</td>
</tr>
<tr>
<td>2004</td>
<td>83.05</td>
<td>68.52</td>
</tr>
<tr>
<td>2005</td>
<td>21.45</td>
<td>23.67</td>
</tr>
<tr>
<td>2006</td>
<td>2.91</td>
<td>3.04</td>
</tr>
<tr>
<td>2007</td>
<td>0.78</td>
<td>0.72</td>
</tr>
</tbody>
</table>

where

$$\Psi_p = \begin{bmatrix}
\psi_{11} & \psi_{12} & \cdots & \psi_{1p} \\
\psi_{21} & \psi_{22} & \cdots & \psi_{2p} \\
\vdots & \vdots & \ddots & \vdots \\
\psi_{p1} & \psi_{p2} & \cdots & \psi_{pp}
\end{bmatrix}.$$  

The vectors $$Y_n, X_n, bDg_n$$ are defined in Section 3.3 as the projections on the FPC’s $$v_1, v_2, \ldots, v_p$$. Lemma 3.8 establishes an analog of (3.8.22) if these FPC’s are replaced by the EFPC’s $$\hat{v}_1, \hat{v}_2, \ldots, \hat{v}_p$$. These replacement introduces additional terms generically denoted with the letter $$\gamma$$. First we prove Lemma 3.7 which leads to a decomposition analogous to (3.3.5).

**Lemma 3.7.** If relation (3.1.1) holds with a Hilbert–Schmidt kernel $$\psi(\cdot, \cdot)$$, then

$$Y_n(t) = \int \left( \sum_{i,j=1}^{p} \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) \hat{v}_j(s) \right) X_n(s) ds + \Delta_n(t),$$

where

$$\Delta_n(t) = \varepsilon_n(t) + \eta_n(t) + \gamma_n(t).$$
The terms \( \eta_n(t) \) and \( \gamma_n(t) \) are defined as follows:

\[
\eta_n(t) = \eta_{n1}(t) + \eta_{n2}(t);
\]

\[
\eta_{n1}(t) = \int \left( \sum_{i=p+1}^{\infty} \sum_{j=1}^{\infty} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds,
\]

\[
\eta_{n2}(t) = \int \left( \sum_{i=1}^{p} \sum_{j=p+1}^{\infty} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds.
\]

\[
\gamma_n(t) = \gamma_{n1}(t) + \gamma_{n2}(t);
\]

\[
\gamma_{n1}(t) = \int \sum_{i,j}^{p} \hat{c}_i \psi_{ij} [\hat{c}_i v_i(t) - \hat{v}_i(t)] v_j(s) X_n(s) ds,
\]

\[
\gamma_{n2}(t) = \int \sum_{i,j}^{p} \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) [\hat{c}_j v_j(s) - \hat{v}_j(s)] X_n(s) ds.
\]

**Proof.** Observe that by (3.3.4),

\[
\int \psi(t, s) X_n(s) ds = \int \left( \sum_{i,j}^{\infty} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds
\]

\[
= \int \left( \sum_{i,j}^{p} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds + \eta_n(t).
\]

Thus model (3.1.1) can be written as

\[
Y_n(t) = \int \left( \sum_{i,j}^{p} \psi_{ij} v_i(t) v_j(s) \right) X_n(s) ds + \eta_n(t) + \varepsilon_n(t)
\]

To take into account the effect of the estimation of the \( v_k \), we will use the decomposition

\[
\psi_{ij} v_i(t) v_j(s) = \hat{c}_i \psi_{ij} \hat{c}_j [\hat{c}_i v_i(t)] (\hat{c}_j v_j(s))
\]

\[
\quad = \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) \hat{v}_j(s)
\]
which allows us to rewrite (3.1.1) as

\[
Y_n(t) = \int \left( \sum_{i,j=1}^{p} \hat{c}_i \psi_{ij} \hat{c}_j \hat{v}_i(t) \hat{v}_j(s) \right) X_n(s) ds + \Delta_n(t). \]

To state lemma 3.8, we introduce the vectors

\[
\hat{Y}_n = [\hat{Y}_n1, \hat{Y}_n2, \ldots, \hat{Y}_np]^T, \quad \hat{Y}_{nk} = \langle Y_n, \hat{v}_k \rangle;
\]

\[
\hat{X}_n = [\hat{\xi}_n1, \hat{\xi}_n2, \ldots, \hat{\xi}_np]^T, \quad \hat{\xi}_{nk} = \langle X_n, \hat{v}_k \rangle;
\]

\[
\hat{\Delta}_n = [\hat{\Delta}_n1, \hat{\Delta}_n2, \ldots, \hat{\Delta}_np]^T, \quad \hat{\Delta}_{nk} = \langle \Delta_n, \hat{v}_k \rangle.
\]

Projecting relation (3.1.1) onto \( \hat{v}_k \), we obtain by Lemma 3.7,

\[
\langle Y_n, \hat{v}_k \rangle = \sum_{j=1}^{p} \hat{c}_k \psi_{kj} \hat{c}_j \langle X_n, \hat{v}_j \rangle + \langle \Delta_n, \hat{v}_k \rangle, \quad 1 \leq k \leq p,
\]

from which the following lemma follows.

**Lemma 3.8.** If relation (3.1.1) holds with a Hilbert–Schmidt kernel \( \psi(\cdot, \cdot) \), then

\[
\hat{Y}_n = \tilde{\Psi}_p \hat{X}_n + \hat{D}_n, \quad n = 1, 2, \ldots N,
\]

where \( \tilde{\Psi}_p \) is the \( p \times p \) matrix with entries \( \hat{c}_k \psi_{kj} \hat{c}_j, \ k, j = 1, 2, \ldots p \).

To find the asymptotic distribution of the matrices \( \mathbf{V}_h \), we establish several lemmas. Each of them removes terms which are asymptotically negligible, and in the process the leading terms are identified. Our first lemma shows that, asymptotically,
in the definition of $V_h$, the residuals

\[(3.8.23) \quad R_n = \hat{Y}_n - \tilde{Y}_n = (\tilde{\Psi}_p - \hat{\Psi}_p)\hat{X}_n + \hat{\Delta}_n.\]

can be replaced by the “errors” $\hat{\Delta}_n$. The essential element of the proof is the relation $\tilde{\Psi}_p - \hat{\Psi}_p = O_P(N^{-1/2})$ stated in lemma 3.2.

**Lemma 3.9.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. Then, for any fixed $h > 0$,

$$\left\| V_h - N^{-1} \sum_{n=1}^{N-h} \hat{\Delta}_n \hat{\Delta}_{n+h}^T \right\| = O_P(N^{-1}).$$

**Proof.** By (3.8.23) and (3.4.14),

$$V_h = N^{-1} \sum_{n=1}^{N-h} [(\tilde{\Psi}_p - \hat{\Psi}_p)\hat{X}_n + \hat{\Delta}_n][(\tilde{\Psi}_p - \hat{\Psi}_p)\hat{X}_{n+h} + \hat{\Delta}_{n+h}]^T.$$  

Denoting, $\hat{C}_h = N^{-1} \sum_{n=1}^{N-h} \hat{X}_n\hat{X}_{n+h}^T$, we thus obtain

$$V_h = (\tilde{\Psi}_p - \hat{\Psi}_p)\hat{C}_h(\tilde{\Psi}_p - \hat{\Psi}_p)^T + (\tilde{\Psi}_p - \hat{\Psi}_p)N^{-1} \sum_{n=1}^{N-h} \hat{X}_n\hat{\Delta}_{n+h}^T$$

$$+ N^{-1} \sum_{n=1}^{N-h} \hat{\Delta}_n\hat{X}_{n+h}(\tilde{\Psi}_p - \hat{\Psi}_p)^T + N^{-1} \sum_{n=1}^{N-h} \hat{\Delta}_n\hat{\Delta}_{n+h}^T.$$  

By the CLT for $h$–dependent vectors, $\hat{C}_h = O_P(1)$, so the first term satisfies

$$\quad (\tilde{\Psi}_p - \hat{\Psi}_p)\hat{C}_h(\tilde{\Psi}_p - \hat{\Psi}_p)^T = O_P(N^{-1/2}N^{-1/2}) = O_P(N^{-1}).$$

To deal with the remaining three terms, we use the decomposition of Lemma 3.7. It is enough to bound the coordinates of each of the resulting terms. Since $\Delta_n = \varepsilon_n + \eta_{n1} + \eta_{n2} + \gamma_{n1} + \gamma_{n2}$, we need to establish bounds for $2 \times 5 = 10$ terms,
but these bounds fall only to a few categories, so we will only deal with some typical cases.

Starting with the decomposition of $\hat{X}_n \hat{\Delta}^T_{n+h}$, observe that

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \varepsilon_{n+h}, \hat{v}_j \rangle = \iint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \varepsilon_{n+h}(s) \right) \hat{v}_i(t) \hat{v}_j(s) dt ds.$$ 

The terms $X_n(t) \varepsilon_{n+h}(s)$ are iid elements of the Hilbert space $L^2([0,1] \times [0,1])$, so by the CLT in a Hilbert space, see e.g. Section 2.3 of Bosq (2000),

$$\iint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \varepsilon_{n+h}(s) dt ds \right)^2 = O_P(1).$$

Since the $\hat{v}_j$ have unit norm, $\iint (\hat{v}_i(t) \hat{v}_j(s))^2 dt ds = 1$. It therefore follows from the Cauchy–Schwarz inequality that

$$\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \varepsilon_{n+h}, \hat{v}_j \rangle = O_P(N^{1/2}).$$

Thus, the $\varepsilon_n$ contribute to $(\hat{\Psi}_p - \hat{\Psi}^\wedge_p) N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \hat{\Delta}^T_{n+h}$ a term of the order $O_P(N^{-1/2}N^{-1}N^{1/2}) = O_P(N^{-1})$, as required.

We now turn to the contribution of the $\eta_{n,1}$. As above, we have

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle = \iint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \eta_{n+h,1}(s) \right) \hat{v}_i(t) \hat{v}_j(s) dt ds$$

$$= \iint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \int \left( \sum_{k=p+1}^{\infty} \sum_{\ell=1}^{\infty} \psi_{k\ell} v_k(s) v_{\ell}(u) \right) X_{n+h}(u) du \right) \hat{v}_i(t) \hat{v}_j(s) dt ds$$

$$= \int \left[ \iint N_h(t, u) R_p(t, u) dt du \right] v_k(s) \hat{v}_j(s) ds,$$
where
\[ N_h(t, u) = N^{-1/2} \sum_{n=1}^{N-h} X_n(t)X_{n+h}(u) \]

and
\[ R_p(t, u) = \sum_{\ell=1}^{\infty} \sum_{k=p+1}^{\infty} \psi_{k\ell}v_{\ell}(u)\hat{v}_k(t). \]

By the CLT for \( m \)-dependent elements in a Hilbert space (follows e.g. from Theorem 2.17 of Bosq (2000)), \( N_h(\cdot, \cdot) \) is \( O_P(1) \) in \( L^2([0,1] \times [0,1]) \), so
\[ \iint N_h^2(t, u)dtdu = O_P(1). \]

A direct verification using Assumption 3.3 shows that also
\[ \iint R_p^2(t, u)dtdu = O_P(1). \]

Thus, by the Cauchy–Schwarz inequality, we obtain that
\[ \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle = O_P(N^{1/2}), \]

and this again implies that the \( \eta_{n1} \) make a contribution of the same order as the \( \varepsilon_n \).

The same argument applies to the \( \eta_{n2} \).

We now turn to the contribution of the \( \gamma_{n1} \), the same argument applies to the \( \gamma_{n2} \). Observe that, similarly as for the \( \eta_{n1} \),

(3.8.24)
\[ N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = \iint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t)\gamma_{n+h,1}(s) \right) \hat{v}_i(t)\hat{v}_j(s)dtds \]
\[ = \iint \left[ \iint N_h(t, u) \sum_{k,\ell=1}^p \hat{c}_k\psi_{k\ell}v_{\ell}(u)\hat{v}_k(t)dtdu \right] [\hat{c}_kv_k(s) - \hat{v}_k(s)]\hat{v}_j(s)ds \]
Clearly,
\[ \int \int \left( \sum_{k,l=1}^{p} \hat{c}_k \psi_{k,l}(u) \dot{v}_l(t) \right)^2 \, dt \, du = O_P(1), \]

By Theorem 3.1,
\[ (3.8.25) \quad \left\{ \int [\hat{c}_k v_k(s) - \hat{v}_k(s)]^2 ds \right\}^{1/2} = O_P(1). \]

We thus obtain
\[ (3.8.26) \quad \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = O_P(1), \]
so the contribution of \( \gamma_n \) is smaller than that of \( \varepsilon_n \) and \( \eta_n \).

To summarize, we have proven that
\[ \left( \tilde{\Psi}_p - \tilde{\Psi}_p^\wedge \right)^T N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \hat{\Delta}_{n+h}^T = O_P(N^{-1}). \]

The term \( N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \hat{\Delta}_{n+h}^T (\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)^T \) can be dealt with in a fully analogous way.

By Lemma 3.7, the errors \( \hat{\Delta}_n \) can be decomposed as follows
\[ \hat{\Delta}_n = \hat{e}_n + \hat{\eta}_n + \hat{\gamma}_n, \]
with the coordinates obtained by projecting the functions \( \varepsilon_n, \eta_n, \gamma_n \) onto the EFPC’s \( \hat{v}_j \). For example,
\[ \hat{\eta}_n = [\langle \eta_n, \hat{v}_1 \rangle, \langle \eta_n, \hat{v}_2 \rangle, \ldots, \langle \eta_n, \hat{v}_p \rangle]^T. \]

Lemma 3.10 shows that the vectors \( \hat{\gamma}_n \) do not contribute to the asymptotic distribution of \( \mathbf{V}_h \). This is essentially due to the fact that by Theorem 3.1, the difference between \( \hat{v}_j \) and \( \hat{c}_j v_j \) is of the order \( O_P(N^{-1/2}) \). For the same reason, in the
definition of $\hat{e}_n$ and $\hat{\eta}_n$, the $\hat{v}_j$ can be replaced by the $\hat{c}_j v_j$, as stated in Lemma 3.11. Lemma 3.11 can be proven in a similar way as Lemma 3.10, so we present only the proof of Lemma 3.10.

**Lemma 3.10.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. Then, for any fixed $h > 0$,

$$\left\| V_h - N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n] [\hat{e}_{n+h} + \hat{\eta}_{n+h}]^T \right\| = O_P(N^{-1}).$$

**Lemma 3.11.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. Then, for any fixed $h > 0$,

$$\left\| V_h - N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n] [\hat{e}_{n+h} + \hat{\eta}_{n+h} + \hat{\gamma}_n]^T \right\| = O_P(N^{-1}),$$

where

$$\hat{e}_n = \hat{c}_1 \langle \varepsilon_n, v_1 \rangle, \hat{c}_2 \langle \varepsilon_n, v_2 \rangle, \ldots, \hat{c}_p \langle \varepsilon_n, v_p \rangle]^T$$

and

$$\hat{\eta}_n = \hat{c}_1 \langle \eta_n, v_1 \rangle, \hat{c}_2 \langle \eta_n, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_n, v_p \rangle]^T = [\hat{c}_1 \langle \eta_{n2}, v_1 \rangle, \hat{c}_2 \langle \eta_{n2}, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_{n2}, v_p \rangle]^T.$$

**Proof of Lemma 3.10.** In light of Lemma 3.9, we must show that the norm of difference between

$$N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n] [\hat{e}_n + \hat{\eta}_n]^T$$

and

$$N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n + \hat{\gamma}_n] [\hat{e}_n + \hat{\eta}_n + \hat{\gamma}_n]^T$$

is $O_P(N^{-1})$. 

Writing $\hat{\eta}_n = \hat{\eta}_{n1} + \hat{\eta}_{n2}$ and $\hat{\gamma}_n = \hat{\gamma}_{n1} + \hat{\gamma}_{n2}$, we see that this difference consists of 20 terms which involve multiplication by $\hat{\gamma}_{n1}$ or $\hat{\gamma}_{n2}$. For example, analogously to (3.8.24), the term involving $\varepsilon_n$ and $\gamma_{n+h,1}$ has coordinates

$$N^{-1} \sum_{n=1}^{N-h} \langle \varepsilon_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle$$

$$= N^{-1/2} \int \left[ \int N_{\varepsilon,h}(t, u) \sum_{k, \ell=1}^{p} \hat{c}_k \psi_k \psi_{\ell}(u) \hat{v}_i(t) dt du \right] \left[ \hat{c}_k v_k(s) - \hat{v}_k(s) \right] \hat{v}_j(s) ds,$$

where

$$N_{\varepsilon,h}(t, u) = N^{-1/2} \sum_{n=1}^{N-h} \varepsilon_n(t) X_{n+h}(u).$$

By the argument leading to (3.8.26) (in particular by (3.8.25)),

$$N^{-1} \sum_{n=1}^{N-h} \langle \varepsilon_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = O_P(N^{-1}).$$

The other terms can be bounded using similar arguments. The key point is that by (3.8.25), all these terms are $N^{1/2}$ times smaller than the other terms appearing in the decomposition of $N^{-1} \sum_{n=1}^{N-h} \hat{\Delta}_n \hat{\Delta}_n^T$. $\blacksquare$

No more terms can be dropped. The asymptotic approximation to $V_h$ thus involves linear functionals of the following processes.

$$R_{N,h}^{(1)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_n(t) \varepsilon_{n+h}(s),$$

$$R_{N,h}^{(2)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_n(t) X_{n+h}(s),$$

$$R_{N,h}^{(3)} = N^{-1/2} \sum_{n=1}^{N} \varepsilon_{n+h}(t) X_n(s),$$
Lemma 3.12, which follows directly for the CLT in the space $L^2([0, 1] \times [0, 1])$ and the calculation of the covariances, summarizes the asymptotic behavior of the processes $R_{N,h}^{(i)}$.

**Lemma 3.12.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. Then

$$\left\{ R_{N,h}^{(i)}, 1 \leq i \leq 4, 1 \leq h \leq H \right\} \xrightarrow{d} \left\{ \Gamma_h^{(i)}, 1 \leq i \leq 4, 1 \leq h \leq H \right\},$$

where the $\Gamma_h^{(i)}$ are $L^2([0, 1] \times [0, 1])$–valued jointly Gaussian process such that the processes $\left\{ \Gamma_h^{(i)}, 1 \leq i \leq 4 \right\}$ are independent and identically distributed.

According to Lemmas 3.11 and 3.12, if

$$(3.8.27) \quad \hat{c}_1 = \hat{c}_2 = \ldots = \hat{c}_p = 1,$$

then

$$N^{1/2} \{ V_h, 1 \leq h \leq H \} \xrightarrow{d} \{ T_h, 1 \leq h \leq H \},$$

where the $T_h, 1 \leq h \leq H$, are independent identically distributed normal random matrices. This is because the limit distribution of the $V_h$ is determined by the random processes $R_{N,h}^{(i)}, 1 \leq i \leq 4, 1 \leq h \leq H$ which are uncorrelated for every fixed $N$. Since their joint limit is multivariate normal, the asymptotic independence of the $R_{N,h}^{(i)}, 1 \leq i \leq 4, 1 \leq h \leq H$ follows. This yields the asymptotic independence of $V_1, \ldots V_H$. Their asymptotic covariances can be computed using Lemma 3.7. After lengthy but straightforward calculations, the following lemma is established

**Lemma 3.13.** Suppose Assumptions 3.2 and 3.3 and condition (3.2.2) hold. If $$(3.8.27) \text{ holds},$$ then for any fixed $h > 0$,

$$N \text{ Cov}(V_h(k, \ell), V_h(k', \ell')) \to a(k, \ell; k', \ell'),$$
where

$$a(k, \ell; k', \ell')$$

$$= r_2(k, k') r_2(\ell, \ell') + r_2(k, k') r_1(\ell, \ell') + r_2(\ell, \ell') r_1(k, k') + r_1(k, k') r_1(\ell, \ell'),$$

with

$$r_1(\ell, \ell') = \sum_{j=p+1}^{\infty} \lambda_j \psi_{\ell j} \psi_{\ell' j}$$

and

$$r_2(k, k') = \int\int E[\varepsilon_1(t) \varepsilon_1(s)] v_k(t) v_{k'}(s) dt ds.$$
\[ a.s. \ E \left[ \left( \langle \varepsilon_n, v_k \rangle + \sum_{j=p+1}^{\infty} \psi_{kj} \langle X_n, v_j \rangle \right) \left( \langle \varepsilon_n, v_{k'} \rangle + \sum_{j=p+1}^{\infty} \psi_{k'j} \langle X_n, v_j \rangle \right) \right] = r_1(k, k') + r_2(k, k'). \]

Therefore, defining,

\[ \hat{a}(k, k', \ell, \ell') = \left( \frac{1}{N} \sum_{n=1}^{N} e_{nk}' e_{nk} \right) \left( \frac{1}{N} \sum_{n=1}^{N} e_{n\ell}' e_{n\ell} \right), \]

we see that

\[ (3.8.28) \quad \hat{a}(k, k', \ell, \ell') \approx \hat{c}_k \hat{c}_{k'} \hat{c}_{\ell} \hat{c}_{\ell'} a(k, k', \ell, \ell'). \]

By Lemma 3.13, under (3.8.27), the asymptotic covariance matrix of \( N^{1/2} \text{vec}(V_h) \) is a \( p^2 \times p^2 \) matrix

\[ M = [ A(i, j), \ 1 \leq i, j \leq p ], \]

where

\[ A(i, j) = [ a(\ell, i, k, j), \ 1 \leq \ell, k \leq p ]. \]

By (3.8.28), an estimator of \( M \) is

\[ \hat{M} = [ \hat{M}(i, j), \ 1 \leq i, j \leq p ], \]

where

\[ \hat{M}(i, j) = [ \hat{a}(\ell, i, k, j), \ 1 \leq \ell, k \leq p ]. \]

Direct verification shows that \( \hat{M} \) can be written in the form (3.4.15), which is convenient for coding.

As seen from (3.8.28), it cannot be guaranteed that the matrix \( \hat{M} \) will be close to the matrix \( M \) because of the unknown signs \( \hat{c}_i \). However, as will be seen in the proof of Theorem 3.3, statistic (3.4.16) does not depend on these signs.
Proof of Theorem 3.3. By Lemmas 3.9 and 3.10,

\[ \text{vec}(V_h) = \text{vec} \left( N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n][\hat{e}_{n+h} + \hat{\eta}_{n+h}]^T \right) + O_P(N^{-1}). \]

The arguments used in the proof of Lemma 3.9 show that

\[ \text{vec} \left( N^{-1} \sum_{n=1}^{N-h} [e_n + \eta_n][e_{n+h} + \eta_{n+h}]^T \right) = [\hat{\mathbf{C}} \otimes \hat{\mathbf{C}}] \text{vec} \left( N^{-1} \sum_{n=1}^{N-h} [e_n + \eta_n][e_{n+h} + \eta_{n+h}]^T \right) + o_P(1), \]

where the matrix \( \hat{\mathbf{C}} \) is defined by (3.3.11), and where

\[ e_n = [(\varepsilon_n, v_1), (\varepsilon_n, v_2), \ldots, (\varepsilon_n, v_p)]^T; \]

\[ \eta_n = [(\eta_n, v_1), (\eta_n, v_2), \ldots, (\eta_n, v_p)]^T. \]

Similar arguments also show that

\[ \hat{\mathbf{M}} = [\hat{\mathbf{C}} \otimes \hat{\mathbf{C}}] \mathbf{M} [\hat{\mathbf{C}} \otimes \hat{\mathbf{C}}] + o_P(1). \]

Since \([\hat{\mathbf{C}} \otimes \hat{\mathbf{C}}]^T[\hat{\mathbf{C}} \otimes \hat{\mathbf{C}}]\) is the \( p^2 \times p^2 \) identity matrix, we obtain by Lemma 3.12 that

\[ Q_N^\wedge = N \sum_{h=1}^H \left\{ \text{vec} \left( N^{-1} \sum_{n=1}^{N-h} [e_n + \eta_n][e_{n+h} + \eta_{n+h}]^T \right) \right\}^T \]

\[ \mathbf{M}^{-1} \left[ \text{vec} \left( N^{-1} \sum_{n=1}^{N-h} [e_n + \eta_n][e_{n+h} + \eta_{n+h}]^T \right) \right]^T + o_P(1). \]

In particular, we see that the asymptotic distribution of \( Q_N^\wedge \) does not depend on the signs \( \hat{c}_1, \hat{c}_2, \ldots, \hat{c}_p \) (the same argument shows that \( Q_N^\wedge \) itself does not depend on these
signs), so we may assume that they are all equal to 1. The claim then follows form Lemmas 3.12 and 3.13.

3.9 Proof of Theorem 3.4

We use expansions with respect to the $v_i$ as well as to the $u_j$, so we replace (3.3.4) by

\[
\psi(t, s) = \sum_{i,j=1}^{\infty} \lambda_i^{-1} \sigma_{ij} u_j(t) v_i(s)
\]

which leads to

\[
\zeta_{nj} = \sum_{i=1}^{\infty} \xi_{ni} \lambda_i^{-1} \sigma_{ij} + \varepsilon_{nj}, \quad 1 \leq j \leq q, \quad 1 \leq n \leq N.
\]

where

\[
\zeta_{nj} = \langle u_j, Y_n \rangle, \quad \xi_{ni} = \langle v_i, X_n \rangle, \quad \varepsilon_{nj} = \langle u_j, \varepsilon_n \rangle,
\]

\[
\sigma_{ij} = E[\xi_{ni} \zeta_{nj}], \quad \lambda_i = E\xi_{ni}^2.
\]

Introducing

\[
Z_n = [Z_{n1}, Z_{n2}, \ldots, Z_{nq}]^T, \quad 1 \leq n \leq N,
\]

\[
Z_{nj} := \zeta_{nj} - \sum_{i=1}^{p} \xi_{ni} \lambda_i^{-1} \sigma_{ij}, \quad 1 \leq j \leq q,
\]

the vector of nonobservable residuals, by (3.9.30), we have

\[
Z_{nj} = \varepsilon_{nj} + r_{nj}, \quad r_{nj} = \sum_{i=p+1}^{\infty} \xi_{ni} \lambda_i^{-1} \sigma_{ij}, \quad 1 \leq j \leq q.
\]
Define by $C_h$ the $q \times q$ autocovariance matrix with entries
\[ c_h(k, \ell) = \frac{1}{N} \sum_{n=1}^{N-h} Z_{nk} Z_{n+h, \ell} \]

Analogously to the statistics $\hat{Q}_N$ of Theorem 3.4, define
\[ Q_N = N \sum_{h=1}^{H} \sum_{i,j=1}^{q} r_{f,h}(i, j) r_{b,h}(i, j), \]

Where $r_{f,h}(i, j)$ and $r_{b,h}(i, j)$ the $(i, j)$ entries, respectively, of $C_0^{-1}C_h$ and $C_hC_0^{-1}$.

Our first result is the limit distribution of $Q_N$.

**Lemma 3.14.** Under the assumptions of Theorem 3.4, $Q_N$ converges in distribution to the $\chi^2$-distribution with $q^2H$ degrees of freedom.

**Proof.** Since the vectors $Z_n$ with coordinates (3.9.31) are independent and identically distributed with zero mean and finite fourth moment, the result follows from Theorem B.3 of Gabrys and Kokoszka (2007).

We must now show that $\hat{Q}_N - Q_N \xrightarrow{P} 0$. To do this, we must use Theorem 3.1, which in turn requires that $Q_N$ and $\hat{Q}_N$ be invariant to the signs of the EFPC’s. This property is established in the following lemma.

**Lemma 3.15.** Set $c_i = \text{sign}(\langle v_i, \hat{v}_i \rangle)$ and $d_i = \text{sign}(\langle u_i, \hat{u}_i \rangle)$. (i) The value of $Q_N$ does not change if each $v_i$ is replaced by $c_i v_i$ and each $u_i$ by $d_i u_i$. (ii) The value of $\hat{Q}_N$ does not change if each $\hat{v}_i$ is replaced by $c_i \hat{v}_i$ and each $\hat{u}_i$ by $d_i \hat{u}_i$.

**Proof.** We will proof statement (i), the arguments for statement (ii) is the same, “hats” have to be added to all formulas. Denote all quantities obtained by using $c_i v_i$ and $d_i u_i$ in place of $v_i$ and $u_i$ with a prime $'$. Then
\[ Z'_{jn} = d_j \zeta_{nj} - \sum_{i=1}^{p} c_i \xi_m \lambda_i^{-1} c_i \sigma_{ij} d_j = d_j Z_{nj}. \]
Therefore,
\[ c_h'(k, \ell) = \frac{1}{N} \sum_{n=1}^{N-h} d_k Z_{nk} d_\ell Z_{n\ell} = d_k d_\ell c_h(k, \ell). \]

Denoting \( D = \text{diag}(d_1, d_2, \ldots, d_q) \), we thus have \( C'_h = D C_h D \). Direct verification shows that \( Q_N = \text{tr}[C'_h C_0^{-1} C_h C_0^{-1}] \). Consequently,
\[ Q'_N = \text{tr}[D C'_h D D^{-1} D^{-1} D C_h D D^{-1} D^{-1}] = \text{tr}[D C_h^T C_0^{-1} C_h C_0^{-1} D^{-1}]. \]

Since \( \text{tr}(AB) = \text{tr}(BA) \), it follows that \( Q'_N = Q_N \). 

Lemma 3.14 and Lemma 3.15 show that to verify the convergence \( \hat{Q}_N - Q_N \overset{P}{\to} 0 \), we may replace \( \hat{v}_i \) by \( c_i \hat{v}_i \) and \( \hat{u}_i \) by \( d_i \hat{u}_i \) in the definition of \( \hat{Q}_N \) to ensure that the differences \( c_i \hat{v}_i - v_i \) and \( d_i \hat{u}_i - u_i \) are small. The residuals \( \hat{Z}_{nj} \) can be expressed as follows
\[ \hat{Z}_{nj} = d_j Z_{nj} + (\hat{Z}_{nj} - Z_{nj}) + D_n(p), \]
\[ D_n(p) = \sum_{i=1}^{p} \left[ \hat{\lambda}_i^{-1} \hat{\sigma}_{i j} \hat{\xi}_{ni} - \lambda_i^{-1} d_j c_i \sigma_{ij} c_i \xi_{ni} \right]. \]

In the following, we use the notation
\[ v_{iN} = c_i \hat{v}_i, \quad u_{iN} = d_i \hat{u}_i \]
and replace \( \hat{v}_i \) and \( \hat{u}_i \), respectively, by \( v_{iN} \) and \( u_{iN} \) in all definitions.

The following Lemma forms part of the proof of Theorem 1 of Gabrys and Kokoszka (2007).

**Lemma 3.16.** Under the assumptions of Theorem 3.4,
\[ \frac{1}{N} \sum_{n=1}^{N} \xi_{nk}(\xi_{n\ell} - \hat{\xi}_{n\ell}) = O_P(N^{-1/2}) \]
and for $h \geq 1$,
\[
\frac{1}{\sqrt{N}} \sum_{n=1}^{N-h} \hat{\xi}_{nk}(\xi_{n+h,\ell} - \hat{\xi}_{n+h,\ell}) = O_P(N^{-1}).
\]

Analogous statements hold for the scores of the $Y_n$.

We will also use the following bounds.

**Lemma 3.17.** Under the assumptions of Theorem 3.4,
\[
\hat{\lambda}_j - \lambda_j = O_P(N^{-1/2}), \quad \hat{\sigma}_{ij} - c_id_j\sigma_{ij} = O_P(N^{-1/2}).
\]

**Proof.** The relation $\hat{\lambda}_j - \lambda_j = O_P(N^{-1/2})$ follows from Theorem 3.1. To establish the second relation, we start with the decomposition
\[
\hat{\sigma}_{ij} - c_id_j\sigma_{ij} = \frac{1}{N} \sum_{n=1}^{N} (\hat{\xi}_{ni}\hat{\xi}_{nj} - \xi_{ni}\xi_{nj}) + \frac{1}{N} \sum_{n=1}^{N} (\xi_{ni}\xi_{nj} - E[\xi_{ni}\xi_{nj}]).
\]
The second term is $O_P(N^{-1/2})$ by the central limit theorem. The first term is further decomposed as
\[
\frac{1}{N} \sum_{n=1}^{N} (\hat{\xi}_{ni}\hat{\xi}_{nj} - \xi_{ni}\xi_{nj}) = \frac{1}{N} \sum_{n=1}^{N} \hat{\xi}_{ni}(\hat{\xi}_{nj} - \xi_{nj}) + \frac{1}{N} \sum_{n=1}^{N} \xi_{nj}(\hat{\xi}_{ni} - \xi_{ni}).
\]
We will show that $N^{-1/2} \sum_{n=1}^{N} \xi_{nj}(\hat{\xi}_{ni} - \xi_{ni})$ is bounded in probability, the other term is dealt with in a similar way. Observe that
\[
N^{-1/2} \sum_{n=1}^{N} \xi_{nj}(\hat{\xi}_{ni} - \xi_{ni}) = N^{-1/2} \sum_{n=1}^{N} \langle Y_n, u_j \rangle \langle X_n, v_iF - v_{i} \rangle
\]
\[
= \left\langle N^{-1/2} \sum_{n=1}^{N} \langle Y_n, u_j \rangle X_n, v_iF - v_{i} \right\rangle.
\]
By the strong law of large numbers in a Hilbert space, the norm of $N^{-1/2} \sum_{n=1}^{N} \langle Y_n, u_j \rangle X_n$ is $O_P(N^{1/2})$, and by Theorem 3.1, $\|v_iF - v_{i}\| = O_P(N^{-1/2}).$
Now we are ready to prove Lemma 3.18 which completes the proof of Theorem 3.4.

**Lemma 3.18.** If the assumptions of Theorem 3.4 holds, then

\[
\hat{C}_0 - C_0 = O_P(N^{-1/2})
\]

and for \( h \geq 1, \)

\[
\hat{C}_h - C_h = O_P(N^{-1})
\]

**Proof.** We will use a modified definition of the autocovariances \( \hat{c}_h(k, \ell) \), namely

\[
\hat{c}_h(k, \ell) = \frac{1}{N} \sum_{n=1}^{N-h} \hat{Z}_{nk} \hat{Z}_{n+h,\ell},
\]

the sample means \( \hat{\mu}_Z(k) \) add terms of the order \( O_P(N^{-1}) \) to the \( \hat{c}_h(k, \ell) \).

To prove relations (3.9.33) and (3.9.34), we decompose \( \hat{C}_h - C_h \) into a number of terms, and use Lemmas 3.16 and 3.17 to show that these terms are of an appropriate order in probability. Observe that for \( h \geq 0, \)

\[
\hat{C}_h - C_h = \frac{1}{N} \sum_{h=1}^{N-h} \hat{Z}_{nk}(\hat{Z}_{n+h,\ell} - Z_{n+h,\ell}) + \frac{1}{N} \sum_{h=1}^{N-h} Z_{n+h,\ell}(\hat{Z}_{nk} - Z_{nk}) =: M_1 + M_2.
\]

In the following, we consider only the first term, \( M_1 \), the same tools apply to \( M_2 \). We decompose \( M_1 \) as

\[
M_1 = M_{11} + M_{12} + M_{13} + M_{14},
\]

where

\[
M_{11} = \frac{1}{N} \sum_{n=1}^{N-h} \hat{\zeta}_{nk}(\hat{\zeta}_{n+h,\ell} - \zeta_{n+h,\ell});
\]
\[ M_{12} = \sum_{j=1}^{p} \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{nk} (\xi_{n+h,j} \lambda_j^{-1} \sigma_{j\ell} - \hat{\xi}_{n+h,j} \hat{\lambda}_j^{-1} \hat{\sigma}_{j\ell}) =: \sum_{j=1}^{p} M_{12j}; \]

\[ M_{13} = -\sum_{i=1}^{p} \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{ni} \hat{\lambda}_i^{-1} \hat{\sigma}_{ik} (\hat{\xi}_{n+h,\ell} - \zeta_{n+h,\ell}) =: -\sum_{i=1}^{p} M_{13i}; \]

\[ M_{14} = -\sum_{j=1}^{p} \sum_{i=1}^{p} \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{nk} \hat{\lambda}_i^{-1} \hat{\sigma}_{ik} (\xi_{n+h,j} \lambda_j^{-1} \sigma_{j\ell} - \hat{\xi}_{n+h,j} \hat{\lambda}_j^{-1} \hat{\sigma}_{j\ell}) =: -\sum_{i=1}^{p} \sum_{j=1}^{p} M_{14ij}. \]

The term \( M_{11} \) is of correct order by Lemma 3.16.

Each term \( M_{12j} \) can be decomposed as

\[ M_{12j} = M_{12j1} + M_{12j2}, \]

where

\[ M_{12j1} = \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{nk} \xi_{n+h,j} (\lambda_j^{-1} \sigma_{j\ell} - \hat{\lambda}_j^{-1} \hat{\sigma}_{j\ell}); \]

\[ M_{12j2} = \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{nk} \hat{\lambda}_i^{-1} \hat{\sigma}_{ik} (\xi_{n+h,j} - \hat{\xi}_{n+h,j}). \]

By Lemma 3.17, \( M_{12j1} = O_P(N^{-1/2})N^{-1} \sum_{n=1}^{N-h} \zeta_{nk} \xi_{n+h,j} \). If \( h = 0 \), \( N^{-1} \sum_{n=1}^{N} \zeta_{nk} \xi_{n,j} = O_P(1) \) by the law of large numbers. If \( h \geq 1 \), \( N^{-1} \sum_{n=1}^{N-h} \zeta_{nk} \xi_{n+h,j} = O_P(N^{-1/2}) \), by the central limit theorem. The term \( M_{12j2} \) has the same rates. If \( h = 0 \), by the law of large numbers in a Hilbert space,

\[ \frac{1}{N} \sum_{n=1}^{N} \zeta_{nk} (\xi_{n,j} - \hat{\xi}_{n,j}) = \left\langle \frac{1}{N} \sum_{n=1}^{N} \langle u_k, Y_n \rangle X_n, v_j - v_jN \right\rangle = O_P(1)O_P(N^{-1/2}). \]

If \( h \geq 1 \), by the central limit theorem in a Hilbert space,

\[ \frac{1}{N} \sum_{n=1}^{N-h} \zeta_{nk} (\xi_{n+h,j} - \hat{\xi}_{n+h,j}) = \left\langle \frac{1}{N} \sum_{n=1}^{N-h} \langle u_k, Y_n \rangle X_{n+h}, v_j - v_jN \right\rangle = O_P(N^{-1/2})O_P(N^{-1/2}). \]

We conclude that

\[ M_{12} = O_P(N^{-1/2}), \quad \text{if } h = 0; \quad M_{12} = O_P(N^{-1}), \quad \text{if } h \geq 1. \]
The same technique shows that for $\alpha = 3$ and $\alpha = 4$

$$M_{1\alpha} = O_P(N^{-1/2}), \quad \text{if } h = 0; \quad M_{1\alpha} = O_P(N^{-1}), \quad \text{if } h \geq 1.$$ □

### 3.10 Proofs of Theorems 3.5 and 3.6

We closely follow the plan of the proof of Theorem 3.3. The decomposition in Lemma 3.7 and lemma 3.8 clearly hold for dependent $X_n$.

To formulate our first lemma, we introduce the $p \times p$ matrix

$$\hat{K}_p = \left[ \hat{c}_i \hat{c}_j \sum_{k=p+1}^{\infty} \int \int \psi_{jk}v_k(u)e_h(t,u)v_j(t)dtdu, \quad 1 \leq i,j \leq p \right].$$

**Lemma 3.19.** Under the assumptions of Theorem 3.5, for any fixed $h > 0$,

$$\left\| V_h - \left[ N^{-1} \sum_{n=1}^{N-h} \Delta_n \Delta_{n+h}^T + (\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)\hat{K}_p + \hat{K}_p^T(\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)^T \right] \right\| = O_P(N^{-1}).$$

**Proof.** By (3.8.23) and (3.4.14),

$$V_h = N^{-1} \sum_{n=1}^{N-h} ((\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)\hat{X}_n + \Delta_n)((\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)\hat{X}_{n+h} + \Delta_{n+h})^T.$$

Denoting, $\hat{C}_h = N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \hat{X}_{n+h}^T$, we thus obtain

$$V_h = (\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)\hat{C}_h(\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)^T + (\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \Delta_{n+h}^T$$

$$+ N^{-1} \sum_{n=1}^{N-h} \Delta_n \hat{X}_{n+h}^T(\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)^T + N^{-1} \sum_{n=1}^{N-h} \Delta_n \Delta_{n+h}^T.$$

By the ergodic theorem, $\hat{C}_h = O_P(1)$, so the first term satisfies

$$(\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)\hat{C}_h(\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge)^T = O_P(N^{-1/2} N^{-1/2}) = O_P(N^{-1}).$$
To deal with the remaining three terms, we use the decomposition of Lemma 3.7. Starting with the decomposition of \( \hat{X}_n^{\Delta T} n+h \), observe that

\[
N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \varepsilon_{n+h}, \hat{v}_j \rangle = \int \int \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \varepsilon_{n+h}(s) \right) \hat{v}_i(t) \hat{v}_j(s) dt ds.
\]

It is verified in Aue et al. (2010) that

\[
\int \int \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \varepsilon_{n+h}(s) dt ds \right)^2 = O_p(1).
\]

Since the \( \hat{v}_j \) have unit norm, it follows from the Cauchy-Schwarz inequality that

\[
\sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \varepsilon_{n+h}, \hat{v}_j \rangle = O_p(N^{1/2}).
\]

Thus, the \( \varepsilon_n \) contribute to \( (\tilde{\Psi}_p - \Psi^p) N^{-1} \sum_{n=1}^{N-h} \hat{X}_n^{\Delta T} n+h \) a term of the order \( O_p(N^{-1/2} N^{-1} N^{1/2}) = O_p(N^{-1}) \).

We now turn to the contribution of the \( \eta_{n,1} \). As above, we have

\[
N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle = \int \int \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \eta_{n+h,1}(s) \right) \hat{v}_i(t) \hat{v}_j(s) dt ds
\]

\[
= \int \int \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t) \int \left( \sum_{k=p+1}^{\infty} \sum_{\ell=1}^{\infty} \psi_{k\ell} v_k(s) v_{\ell}(u) \right) X_{n+h}(u) du \right) \hat{v}_i(t) \hat{v}_j(s) dt ds.
\]

Setting

\[
N_h(t, u) = N^{-1/2} \sum_{n=1}^{N-h} [X_n(t) X_{n+h}(u) - e_h(t, u)]
\]

and

\[
R_p(s, u) = \sum_{\ell=1}^{\infty} \sum_{k=p+1}^{\infty} \psi_{k\ell} v_k(u) \hat{v}_i(s),
\]
we thus obtain
\[
N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \eta_{n+h,1}, \hat{v}_j \rangle
\]
\[
= \iiint N_h(t,u) \hat{v}_j(s) R_p(s,u) \hat{v}_i(t) dtdsdu
\]
\[
+ \frac{N-h}{N^{1/2}} \iiint e_h(t,u) \hat{v}_j(s) R_p(s,u) \hat{v}_i(t) dtdsdu.
\]
By the Cauchy–Schwarz inequality, we have
\[
\left| \iiint N_h(t,u) \hat{v}_j(s) R_p(s,u) \hat{v}_i(t) dtdsdu \right|
\]
\[
\leq \left( \iiint N_h^2(t,u) \hat{v}_j^2(s) dtdsdu \right)^{1/2} \left( \iiint R_p^2(s,u) \hat{v}_i^2(t) dtdsdu \right)^{1/2}.
\]
Aue et al. (2010) verified that
\[
(3.10.35) \quad \iiint N_h^2(t,u) dtdu = O_P(1).
\]
A direct verification using Assumption 3.3 shows that also
\[
\iiint R_p^2(t,u) dtdu = O_P(1).
\]
Hence
\[
\left| \iiint N_h(t,u) \hat{v}_j(s) R_p(s,u) \hat{v}_i(t) dtdsdu \right| = O_P(1).
\]
Using (3.2.3), we conclude that
\[
\iiint e_h(t,u) \hat{v}_j(s) R_p(s,u) \hat{v}_i(t) dtdsdu
\]
\[
= \iiint e_h(t,u) v_j(s) R_p(s,u) v_i(t) dtdsdu + O_P(N^{-1/2}).
\]
Since $j \leq p$, $\int R_p(s, u)v_j(s)ds = 0$, leading to

\[(3.10.36)\]

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \hat{\eta}_{n+h,1}, \hat{v}_j \rangle = O_P(N^{1/2}).$$

Repeating the arguments leading to (3), we obtain that

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \hat{\eta}_{n+h,2}, \hat{v}_j \rangle$$

$$= (N - h)\hat{c}_i \hat{c}_j \iiint e_h(t, u)v_j(s) \sum_{\ell=1}^{p} \sum_{k=p+1}^{\infty} \psi_{\ell k} v_\ell(s)v_k(u)v_i(t)dsdtdu + O_P(N^{1/2}).$$

By orthogonality of the $v_i$,

$$\int \sum_{\ell=1}^{p} \sum_{k=p+1}^{\infty} \psi_{\ell k} v_\ell(s)v_k(u)v_j(s)ds = \sum_{k=p+1}^{\infty} \psi_{j k} v_k(u).$$

We now turn to the contribution of the $\gamma_{n1}$, the same argument applies to the $\gamma_{n2}$. Observe that, similarly as for the $\eta_{n1}$,

\[(3.10.37)\]

$$N^{-1/2} \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle = \iiint \left( N^{-1/2} \sum_{n=1}^{N-h} X_n(t)\gamma_{n+h,1}(s) \right) \hat{v}_i(t)\hat{v}_j(s)dtds$$

$$= \int \left[ \iiint N_h(t, u) \sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k \ell} v_\ell(u)\hat{v}_i(t)dtdu \right] [\hat{c}_k v_k(s) - \hat{v}_k(s)]\hat{v}_j(s)ds$$

$$+ \frac{N - h}{N^{1/2}} \int \left[ \iiint e_h(t, u) \sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k \ell} v_\ell(u)\hat{v}_i(t)dtdu \right] [\hat{c}_k v_k(s) - \hat{v}_k(s)]\hat{v}_j(s)ds.$$ Clearly,

$$\iiint \left( \sum_{k,\ell=1}^{p} \hat{c}_k \psi_{k \ell} v_\ell(u)\hat{v}_i(t) \right)^2 dtdu = O_P(1).$$

\[\]
By (3.2.3),
\[
\left\{ \int [\hat{c}_k v_k(s) - \hat{v}_k(s)]^2 ds \right\}^{1/2} = O_P(N^{-1/2})
\]
and
\[
\iint\int c_h(t, u) \sum_{k, \ell=1}^p \hat{c}_k \psi_k \nu_\ell(u) \hat{v}_i(t) [\hat{c}_k v_k(s) - \hat{v}_k(s)] \hat{v}_j(s) dtduds = O_P(N^{-1/2}).
\]

We thus obtain
\[
(3.10.38) \sum_{n=1}^{N-h} \langle X_n, \hat{v}_i \rangle \langle \gamma_{n+h, 1}, \hat{v}_j \rangle = O_P(N^{1/2}),
\]
with the same bound holding for \( \gamma_{n+h, 2} \) in place of \( \gamma_{n+h, 1} \).

To summarize, we have expanded, as required, the term \((\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge) N^{-1} \sum_{n=1}^{N-h} \hat{X}_n \hat{\Delta}_n^T\).

The term \( N^{-1} \sum_{n=1}^{N-h} \hat{\Delta}_n \hat{X}_{n+h}^T (\tilde{\Psi}_p - \tilde{\Psi}_p^\wedge) \) can be dealt with in a fully analogous way.  

By Lemma 3.7, the errors \( \hat{\Delta}_n \) can be decomposed as follows
\[
\hat{\Delta}_n = \hat{e}_n + \hat{\eta}_n + \hat{\gamma}_n,
\]
with the coordinates obtained by projecting the functions \( \varepsilon_n, \eta_n, \gamma_n \) onto the EFPC's \( \hat{v}_j \). For example,
\[
\hat{\eta}_n = [\langle \eta_n, \hat{v}_1 \rangle, \langle \eta_n, \hat{v}_2 \rangle, \ldots, \langle \eta_n, \hat{v}_p \rangle]^T.
\]
Lemma 3.20 shows that $\hat{\gamma}_n$ contributes a drift term to the asymptotic distribution of $V_h$. To formulate it, we introduce the a $p \times p$ matrix $\hat{F}_p$ with entries

\[
\hat{F}_p(i, j) = \sum_{m, r=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{i\ell} \psi_{mr} \int \int v_{\ell}(s) v_{r}(z) e_h(s, z) ds dz \int \hat{c}_m [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_j(u) du \\
+ \sum_{m=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{i\ell} \psi_{mj} \hat{c}_m \int \int v_{\ell}(s) e_h(s, z) [\hat{c}_j v_j(s) - \hat{v}_j(s)] ds dz \\
+ \sum_{m, r=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{j\ell} \psi_{mr} \int \int v_{\ell}(s) v_{r}(z) e_h(s, z) ds dz \int \hat{c}_m [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_i(u) du \\
+ \sum_{m=1}^{p} \sum_{\ell=1}^{\infty} \psi_{j\ell} \psi_{mj} \hat{c}_m \int \int v_{\ell}(s) e_h(s, z) [\hat{c}_j v_j(s) - \hat{v}_j(s)] ds dz.
\]

Lemma 3.20. Under the assumptions of Theorem 3.5, for any fixed $h > 0$,

\[
\left| \left| N^{-1} \sum_{n=1}^{N-h} \Delta_n \Delta_{n+h}^T - \left\{ N^{-1} \sum_{n=1}^{N-h} [\hat{e}_n + \hat{\eta}_n] [\hat{e}_{n+h} + \hat{\eta}_{n+h}]^T + \hat{F}_p \right\} \right| \right| = O_P(N^{-1}).
\]

Proof. Following the proof of Lemma 3.19, one can verify that

\[
\left| \left| \sum_{n=1}^{N-h} \hat{e}_n \hat{\eta}_{n+h}^T \right| \right| = O_P(1), \quad \left| \left| \sum_{n=1}^{N-h} \hat{\eta}_n \hat{e}_{n+h}^T \right| \right| = O_P(1)
\]

and

\[
\left| \left| \sum_{n=1}^{N-h} \hat{\eta}_n \hat{\gamma}_{n+h}^T \right| \right| = O_P(1), \quad \left| \left| \sum_{n=1}^{N-h} \hat{\gamma}_n \hat{\eta}_{n+h,1}^T \right| \right| = O_P(1).
\]

Using (3.2.3), (3.10.35), and the orthonormality of the $v_i$, we get

\[
\sum_{n=1}^{N-h} \langle \eta_{n,2}, \hat{v}_i \rangle \langle \gamma_{n+h,1}, \hat{v}_j \rangle
\]

\[
= \hat{c}_i \hat{c}_j N \sum_{m, r=1}^{p} \sum_{\ell=p+1}^{\infty} \psi_{i\ell} \psi_{mr} \int \int v_{\ell}(s) v_{r}(z) e_h(s, z) ds dz \hat{c}_m \psi_{mr} [\hat{c}_m v_m(u) - \hat{v}_m(u)] v_j(u) du + O_P(1)
\]
and
\[ \sum_{n=1}^{N-h} \langle \eta_n, \hat{\nu}_i \rangle \langle \gamma_{n+h}, \hat{\nu}_j \rangle \]
= \hat{c}_i \hat{c}_j N \sum_{m=1}^{p} \sum_{l=p+1}^{\infty} \psi_l \psi_{mj} \iint v_t(s)v_r(z)e_h(s, z)dsdz \hat{c}_m \psi_{mr} [\hat{c}_j v_j(s) - \hat{v}_j(s)]dsdz + O_P(1).

The remaining two terms, \( \sum_{n=1}^{N-h} \langle \gamma_n, \hat{\nu}_i \rangle \langle \eta_{n+h}, \hat{\nu}_j \rangle, \alpha = 1, 2 \), can be handled in the same way.

To formulate the next lemma, we introduce the matrix \( \hat{G}_p \) whose \((i, j)\) entry is
\[
\hat{G}_p(i, j) = \hat{c}_j \sum_{k=1}^{p} \sum_{\ell, r=1}^{p} \psi_k \psi_{jr} \iint v_k(t)[\hat{v}_i(t) - \hat{c}_i v_i(t)]v_\ell(s) e_h(s, z)v_r(z)dt ds dz + \hat{c}_i \sum_{m=1}^{p} \sum_{\ell, r=1}^{p} \psi_{i\ell} \psi_{mr} \iint v_\ell(s)e_h(s, z)v_m(u)[\hat{v}_j(u) - \hat{c}_j v_j(u)]v_r(z)du ds dz.
\]

Recall also that
\[ \tilde{e}_n = [\hat{c}_1 \langle \varepsilon_n, v_1 \rangle, \hat{c}_2 \langle \varepsilon_n, v_2 \rangle, \ldots, \hat{c}_p \langle \varepsilon_n, v_p \rangle]^T \]
and
\[ \tilde{\eta}_n = [\hat{c}_1 \langle \eta_{n,2}, v_1 \rangle, \hat{c}_2 \langle \eta_{n,2}, v_2 \rangle, \ldots, \hat{c}_p \langle \eta_{n,2}, v_p \rangle]^T. \]

**Lemma 3.21.** Under the assumptions of Theorem 3.5, for any fixed \( h > 0 \),
\[ \left\| N^{-1} \sum_{n=1}^{N-h} [\tilde{e}_n + \tilde{\eta}_n][\tilde{e}_{n+h} + \tilde{\eta}_{n+h}]^T - \left\{ N^{-1} \sum_{n=1}^{N-h} [\tilde{e}_n + \tilde{\eta}_n][\tilde{e}_{n+h} + \tilde{\eta}_{n+h}]^T + \hat{G}_p \right\} \right\| = O_P(N^{-1}). \]

Lemma 3.21 can be established along the lines of the proof of Lemma 3.20.

**Proof of Theorem 3.6.** The proof of Theorem 3.5 shows that
\[ \left\| V_h - \left[ \hat{c}_i \hat{c}_j \left( V_h^{(e)}(i, j) + D_h(i, j) \right) \right] v_j \leq i, j \leq p \right\| \rightarrow 0, \]
where
\[ V_h^{(e)}(i,j) = E[\langle \varepsilon_0, v_i \rangle \langle \varepsilon_h, v_j \rangle]. \]

Thus, but the definition of the statistic \( Q_N^\wedge \), the consistency is established if for some \( 1 \leq h \leq H \),
\[ [\text{vec}(V_h^{(e)}) + D_h]^T \{ \hat{M}_0 \otimes \hat{M}_0 \}^{-1} \text{vec}(V_h^{(e)}) + D_h \neq 0. \]

The matrix \( D_h \) converges to zero, as \( p \to \infty \). It can be shown, as in the proof of Theorem 3.5, that by imposing the same dependence conditions on the \( \varepsilon_n \) as on the \( X_n \), we have \( \hat{c}_i \hat{c}_j \hat{M}_0(i,j) \overset{p}{\to} M_0(i,j) \), where \( M_0(i,j) \) is the limit of the empirical covariances \( N^{-1} \sum_{n=1}^{N} e_n^{\wedge} e_n^{\wedge} \), and so is positive definite.

Thus, the test is consistent if for some \( 1 \leq h \leq H \),
\[ [\text{vec}(V_h^{(e)})]^T \{ \hat{M}_0 \otimes \hat{M}_0 \}^{-1} \text{vec}(V_h^{(e)}) \neq 0, \]
that is, if for some \( 1 \leq h \leq H \) and \( 1 \leq i, j \leq p \), \( E[\langle \varepsilon_0, v_i \rangle \langle \varepsilon_h, v_j \rangle] \neq 0. \)
CHAPTER 4
DETECTING CHANGES IN THE MEAN OF FUNCTIONAL OBSERVATIONS\(^1\)

4.1 Introduction

Functional data analysis (FDA) has been enjoying increased popularity over the last decade due to its applicability to problems which are difficult to cast into a framework of scalar or vector observations. Even if such standard approaches are available, the functional approach often leads to a more natural and parsimonious description of the data, and to more accurate inference and prediction results. The monograph of Ramsay and Silverman (2005) has become a standard reference to the ideas and tools of FDA. To name a few recent applications of FDA which illustrate its advantages alluded to above, we cite Antoniadis and Sapatinas (2003), Fernández de Castro et al. (2005), Müller and Stadtmüller (2005), Yao et al. (2005b), and Glendinning and Fleet (2007).

A main tool of FDA is the principal component analysis (PCA). It represents the functional observations \(X_i(t), \ t \in \mathcal{T}, \ i = 1, 2, \ldots, N,\) in the form \(X_i(t) = \mu(t) + \sum_{1 \leq \ell < \infty} \eta_{i,\ell} v_{\ell}(t),\) where \(\mu\) is the mean, \(v_{\ell}\) are the eigenfunctions of the covariance operator, and the \(\eta_{i,\ell}\) are the scores. The set \(\mathcal{T}\) can be interpreted as a time or a spatial domain, the methodology we develop requires merely that it be a compact subset of a Euclidean space. To perform the functional PCA, the functional mean \(\mu(t),\) approximated by the sample mean of the \(X_i(t),\) is first subtracted from the data. The first principal component \(v_1(t)\) is then interpreted as the main pattern of deviation of the observations from the mean \(\mu(t),\) or equivalently, as the direction in a function space of the largest variability away from the mean function. The subsequent

eigenfunction define analogous directions orthogonal to the previous eigenfunctions.

This interpretation and inferential procedures based on it assume that the mean function $\mu(t)$ is the same for all values of $i$. If, in fact, the mean changes at some index(es) $i$, the results of PCA are confounded by the change(s). Issues of this type are most likely to emerge if the data are collected sequentially over time. Applications we have in mind abound in climatology, environmental science and economics; detecting and locating changes in mean can be interpreted, for example, as climate shifts, a baseline change in a pollution level, or a shift in a long-term rate of growth.

It is thus useful to develop a methodology for the detection of changes in the mean of functional observations that is both easy to apply and justified by an clear large sample argument. We propose a significance test for testing the null hypothesis of a constant functional mean against the alternative of a changing mean. We also show how to locate the change points if the null hypothesis is rejected. Our methodology is readily implemented using the R package fda. The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution going back to the work of Kiefer (1959).

The problem of detecting a change in the mean of a sequence Banach space valued random elements has recently been approached from a theoretical angle by Račkauskas and Suquet (2006). Motivated by detecting an epidemic change (the mean changes and then returns to its original value), Račkauskas and Suquet (2006) proposed an interesting statistic based on increasingly fine dyadic partitions of the index interval, and derived its limit, which is nonstandard.

The change point problem has been extensively studied in the multivariate setting starting with Srivastava and Worsley (1986), while the work of Horváth et al. (1999) is most closely related to the present paper. Different multivariate settings with further references are discussed in Lavielle and Teyssière (2006), Zamba and Hawkins (2006), and Qu and Perron (2007), among others.

Returning to the functional setting, a somewhat related problem has recently
been studied by Benko et al. (2009) who considered two populations, admitting the PCA’s:
\[
X_{i,p}(t) = \mu_p(t) + \sum_{1 \leq \ell < \infty} \eta_{i,p,\ell} v_{p,\ell}(t), \quad p = 1, 2.
\]
Benko et al. (2009) developed a bootstrap test for checking if the elements of the two decompositions (including the means) are the same. Earlier, Laukaitis and Račkauskas (2005) considered the model
\[
X_{i,g}(t) = \mu_g(t) + \varepsilon_{i,g}(t), \quad g = 1, 2, \ldots, G,
\]
with innovations \(\varepsilon_{i,g}\) and group means \(\mu_g\), and tested \(H_0 : E\mu_1(t) = \ldots = E\mu_G(t)\).

Other contributions in this direction include Cuevas et al. (2004), Delicado (2007), and Ferraty et al. (2007). In these settings, it is known which population or group each observation belongs to. In our setting, we do not have any partition of the data into several sets with possibly different means. The change can occur at any point, and we want to test if it occurs or not.

The paper is organized as follows. In Section 4.2, we introduce the required notation and assumptions, and recall several results which will be used in the following sections. Section 4.3 describes the proposed methodology and contains theorems which provide its asymptotic justification. The finite sample performance is investigated in Section 4.4, which also contains an illustrative application to the detection of changes in mean patterns of annual temperatures. The proofs of the Theorems of Section 4.3 are presented in Section 4.5.

### 4.2 Notation and assumptions

We consider functional observations \(X_i(t), \quad t \in \mathcal{T}, \quad i = 1, 2, \ldots, N\), defined over a compact set \(\mathcal{T}\). We assume that the \(X_i\) are independent, and we want to test if their
mean remains constant in $i$. Thus we test the null hypothesis

$$H_0 : \ E X_1(t) = E X_2(t) = \ldots = E X_N(t), \ t \in \mathcal{I}.$$ 

Note that under $H_0$, we do not specify the value of the common mean.

Under the alternative, $H_0$ does not hold. The test we construct, has particularly good power against the alternative in which the data can be divided into several consecutive segments, and the mean is constant within each segment, but changes from segment to segment. The simplest case of only two segments (one change point) is specified in Assumption 4.2.

Under the null hypothesis, we can represent each functional observation as

$$X_i(t) = \mu(t) + Y_i(t), \ E Y_i(t) = 0.$$  

The following assumption specifies conditions on $\mu(\cdot)$ and the errors $Y_i(\cdot)$ needed to establish the asymptotic distribution of the test statistic.

In the following, unless indicated otherwise, all integrals denote integration over the set $\mathcal{I}$.

**Assumption 4.2.** The mean $\mu(\cdot)$ is in $L^2(\mathcal{I})$. The errors $Y_i(\cdot)$ are iid mean zero random elements of $L^2(\mathcal{I})$ which satisfy

$$E||Y_i||^2 = \int E Y_i^2(t) dt < \infty.$$ 

Their covariance function

$$c(t, s) = E[Y_i(t)Y_i(s)] \quad t, s \in \mathcal{I}$$

is square integrable, i.e. is in $L^2(\mathcal{I} \times \mathcal{I})$. 
Assumption 4.2 implies the following expansions, see e.g. Chapter 4 of Indritz (1963):

\[(4.2.4)\quad c(t, s) = \sum_{1 \leq k < \infty} \lambda_k v_k(t)v_k(s)\]

and

\[(4.2.5)\quad Y_i(t) = \sum_{1 \leq \ell < \infty} \sqrt{\lambda_\ell} \xi_{i,\ell} v_\ell(t),\]

where \(\lambda_k\) and \(v_k\) are, respectively, the eigenvalues and eigenfunctions of the covariance operator, defined by

\[(4.2.6)\quad \int c(t, s)v_\ell(s)ds = \lambda_\ell v_\ell(t), \quad \ell = 1, 2, \ldots\]

The sequences \(\{\xi_{i,\ell}, \ell = 1, 2, \ldots\}\) are independent, and within each sequence the \(\xi_{i,\ell}\) are uncorrelated with mean zero and unit variance. The infinite sum in (4.2.5) converges in \(L^2(\mathcal{T})\) with probability one. Recall also that \(v_\ell, \ell = 1, 2, \ldots,\) form an orthonormal basis in \(L^2(\mathcal{T})\), and all \(\lambda_\ell\) are nonnegative.

In practice, we work with estimated eigenelements defined by

\[(4.2.7)\quad \int \hat{c}(t, s)\hat{v}_\ell(s)ds = \hat{\lambda}_\ell \hat{v}_\ell(t), \quad \ell = 1, 2, \ldots,\]

where

\[
\hat{c}(t, s) = \frac{1}{N} \sum_{1 \leq i \leq N} (X_i(t) - \bar{X}_N(t))(X_i(s) - \bar{X}_N(s)) \quad \text{and} \quad \bar{X}_N(t) = \frac{1}{N} \sum_{1 \leq i \leq N} X_i(t).
\]

To control the distance between the estimated and the population eigenelements, we need the following assumptions:
Assumption 4.3. The eigenvalues $\lambda_\ell$ satisfy, for some $d > 0$

\[
\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1}.
\]

Assumption 4.4. The $Y_i$ in (4.2.1) satisfy

\[
E\|Y_i\|^4 = \int EY_i^4(t)dt < \infty.
\]

The results of Dauxois et al. (1982) and Bosq (2000) then imply that for each $k \leq d$:

\[
\limsup_{N \to \infty} NE \left[ ||\hat{c}_k v_k - \hat{v}_k||^2 \right] < \infty, \quad \limsup_{N \to \infty} NE \left[ |\lambda_k - \hat{\lambda}_k|^2 \right] < \infty,
\]

where $\hat{c}_k = \text{sign} \int_T v_k(t)\hat{v}_k(t)dt$. The random sign $\hat{c}_k$ is included because the $v_k$ and $\hat{v}_k$ are defined up to a sign, and since $v_k$ is unknown, it is impossible to ensure that $\int_T v_k(t)\hat{v}_k(t)dt \geq 0$.

We establish the consistency of the test under the alternative of one change point formalized in Assumption 4.5. A similar argument can be developed if there are several change points, but the technical complications then obscure the main idea explained in Sections 4.3 and 4.5.2 (in particular the functions (4.2.10) and (4.3.18) would need to be modified). The more general case is studied empirically in Section 4.4.

Assumption 4.5. The observations follow the model

\[
X_i(t) = \begin{cases} 
\mu_1(t) + Y_i(t), & 1 \leq i \leq k^*, \\
\mu_2(t) + Y_i(t), & k^* < i \leq N,
\end{cases}
\]
in which the $Y_i$ satisfy Assumption 4.2, the mean functions $\mu_1$ and $\mu_2$ are in $L^2(\mathcal{I})$, and

$$k^* = \lfloor n\theta \rfloor$$

for some $0 < \theta < 1$.

We will see in the proof of Theorem 4.2 that under Assumption 4.5 the sample covariances of the functional observations converge to the function

$$\tilde{c}(t, s) = c(t, s) + \theta(1 - \theta)(\mu_1(t) - \mu_2(t))(\mu_1(s) - \mu_2(s)).$$

This is a symmetric, square integrable function, and it is easy to see that for any $x, y \in L^2(\mathcal{I})$,

$$\int \int \tilde{c}(t, s)x(t)x(s)dtds \geq 0,$$

so $\tilde{c}(t, s)$ is a covariance function. Consequently, it has orthonormal eigenfunctions $w_k$ and nonnegative eigenvalues $\gamma_k$ satisfying

$$\int \tilde{c}(t, s)w_k(s)ds = \gamma_kw_k(t).$$

The quantities $\tilde{c}(t, s)$, $w_k$ and $\gamma_k$ are used in Section 4.3 to describe the distribution of the test statistic under the alternative of one change point.

### 4.3 Detection procedure

To explain the idea of the test procedure, denote

$$\hat{\mu}_k(t) = \frac{1}{\bar{k}} \sum_{1 \leq i \leq \bar{k}} X_i(t), \quad \bar{\mu}_k(t) = \frac{1}{N - \bar{k}} \sum_{\bar{k} < i \leq N} X_i(t).$$

If the mean is constant, the difference $\Delta_k(t) = \hat{\mu}_k(t) - \bar{\mu}_k(t)$ is small for all $1 \leq k < N$ and all $t \in \mathcal{I}$. However, $\Delta_k(t)$ can become large due to chance variability if $k$ is close
to 1 or to \( N \). It is therefore usual to work with the sequence

\[
P_k(t) = \sum_{1 \leq i \leq k} X_i(t) - \frac{k}{N} \sum_{1 \leq i \leq N} X_i(t) = \frac{k(N - k)}{N} [\hat{\mu}_k(t) - \bar{\mu}_k(t)]
\]

in which the variability at the end points is attenuated by a parabolic weight function. If the mean changes, the difference \( P_k(t) \) is large for some values of \( k \) and of \( t \). Since the observations are in an infinite dimensional domain, we work with the projections of the functions \( P_k(\cdot) \) on the principal components of the data. These projections can be expressed in terms of functional scores which can be easily computed using the R package \texttt{fda}.

Consider thus the scores corresponding the largest \( d \) eigenvalues:

\[
\hat{\eta}_{i,\ell} = \int [X_i(t) - \bar{X}_N(t)] \hat{v}_\ell(t) dt, \quad i = 1, 2, \ldots, N, \quad \ell = 1, 2, \ldots, d.
\]

Observe that the value of \( P_k(t) \) does not change if the \( X_i(t) \) are replaced by \( X_i(t) - \bar{X}_N(t) \). Consequently, setting \( \ell = [Nx], \ x \in (0, 1) \), we obtain

\[
\int \left\{ \sum_{1 \leq i \leq Nx} X_i(t) - \left[ \frac{Nx}{N} \sum_{1 \leq i \leq N} X_i(t) \right] \right\} \hat{v}_\ell(t) dt = \sum_{1 \leq i \leq Nx} \hat{\eta}_{i,\ell} - \left[ \frac{Nx}{N} \sum_{1 \leq i \leq N} \hat{\eta}_{i,\ell} \right].
\]

Identity (4.3.13) shows that functional scores can be used for testing the constancy of the mean function.

The following theorem can be used to derive a number of test statistics. To state it, introduce the vectors

\[
\hat{\beta}_i = [\hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,d}]^T, \quad i = 1, 2, \ldots, N,
\]
the covariance matrix

\[ \hat{\Sigma}_d = \begin{bmatrix} \hat{\lambda}_1 & 0 & \cdots & 0 \\ 0 & \hat{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \hat{\lambda}_d \end{bmatrix}. \]

and denote by \( B_1(\cdot), \ldots, B_d(\cdot) \) independent standard Brownian bridges.

**Theorem 4.1.** Suppose Assumptions 4.2, 4.3, 4.3 hold. Then, under \( H_0 \),

\[
\frac{1}{N} \left[ \sum_{1 \leq i \leq Nx} \hat{\beta}_i - x \sum_{1 \leq i \leq N} \hat{\beta}_i \right]^T \hat{\Sigma}_d^{-1} \left[ \sum_{1 \leq i \leq Nx} \hat{\beta}_i - x \sum_{1 \leq i \leq N} \hat{\beta}_i \right] \xrightarrow{d} \sum_{1 \leq \ell \leq d} B^2_\ell(x) \ (0 \leq x \leq 1),
\]

in the Skorokhod space \( D[0,1] \).

Theorem 4.1 is proved in Section 4.5.

To see how it can be used to derive test statistics, denote

\[
T_N(x) = \frac{1}{N} \sum_{\ell=1}^d \hat{\lambda}_\ell^{-1} \left( \sum_{1 \leq i \leq Nx} \hat{\eta}_{\ell,i} - x \sum_{1 \leq i \leq N} \hat{\eta}_{\ell,i} \right)^2.
\]

By Theorem 4.1, \( U(T_N) \xrightarrow{d} U(\sum_{1 \leq \ell \leq d} B^2_\ell(\cdot)) \), for any continuous functional \( U : D[0,1] \to \mathbb{R} \). Applying integral or max functionals, or their weighted versions, leads to useful statistics. In this paper, we focus on the integral of the squared function, i.e. the Cramer-von-Mises functional, which is known to produce effective tests [this functional was also selected in a different context by Bugni et al. (2006)]. Thus, we consider the convergence

\[
\int_0^1 T_N(x) \, dx \xrightarrow{d} \int_0^1 \sum_{1 \leq \ell \leq d} B^2_\ell(x) \, dx,
\]

which can be rewritten as

\[
S_{N,d} := \frac{1}{N^2} \sum_{\ell=1}^d \hat{\lambda}_\ell^{-1} \sum_{k=1}^N \left( \sum_{1 \leq i \leq k} \hat{\eta}_{\ell,i} - \frac{k}{N} \sum_{1 \leq i \leq N} \hat{\eta}_{\ell,i} \right)^2 \xrightarrow{d} \int_0^1 \sum_{1 \leq \ell \leq d} B^2_\ell(x) \, dx.
\]
The distribution of the random variable

\[ K_d = \int_0^1 \sum_{1 \leq t \leq d} B_t^2(x) dx \]

was derived by Kiefer (1959). Denoting by \( c_d(\alpha) \) its \((1-\alpha)\)th quantile, the test rejects \( H_0 \) if \( S_{N,d} > c_d(\alpha) \). The critical values \( c_d(\alpha) \) are presented in Table 4.1.

A multivariate analog of statistic (4.3.15) considered is Horváth et al. (1999)

\[ M_{N,d} = \frac{1}{N^2} \sum_{k=1}^{N} \left( \frac{k}{N} \frac{N-k}{N} \right)^2 \Delta(k) \hat{D}_d^{-1} \Delta^T(k), \]

where \( \Delta(k) \) is the difference of the mean vectors (of dimension \( d \)) computed from the first \( k \) and the last \( N-k \) data vectors, and \( \hat{D}_d \) is the \( d \times d \) matrix of estimated residual vectors. If \( d \) is large, the inverse of \( \hat{D}_d \) is unstable. In statistic (4.3.15), this inverse is “replaced” by inverses of the \( d \) largest eigenvalues \( \hat{\lambda}_\ell \), and the whole statistic is properly “diagonalized” so that only the most important variability of the data is considered, while the high dimensional noise is ignored.

We now turn to the behavior of the test under the alternative. We will show that it is consistent, i.e. \( S_{N,d} \xrightarrow{P} \infty \). In fact, we can obtain the rate of divergence: under \( H_A \), \( S_{n,d} \) grows linearly with \( N \). We formulate these results under the assumption of one change point. Under Assumption 4.5, for \( 1 \leq k \leq d \), introduce the functions

\[ g_k(x) = \begin{cases} 
  x(1-\theta) \int (\mu_1(t) - \mu_2(t)) w_k(t) dt, & 0 < x \leq \theta \\
  \theta(1-x) \int (\mu_1(t) - \mu_2(t)) w_k(t) dt, & \theta < x < 1.
\end{cases} \]

**Theorem 4.2.** Under Assumption 4.2,

\[ \sup_{0 \leq x \leq 1} \left| N^{-1} T_N - g^T(x) \Sigma^* g(x) \right| = o_P(1), \]
where
\[ g(x) = [g_1(x), \ldots, g_d(x)]^T; \quad \Sigma^* = \begin{bmatrix}
\frac{1}{\gamma_1} & 0 & \cdots & 0 \\
0 & \frac{1}{\gamma_2} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \frac{1}{\gamma_d}
\end{bmatrix}. \]

Theorem 4.2 is proved in Section 4.5.

It follows that the test statistic (4.3.15) satisfies the law of large numbers under
the alternative, i.e.
\[ \frac{1}{N} S_{N,d} \xrightarrow{P} \sum_{1 \leq k \leq d} \frac{1}{\gamma_k} \int_0^1 g_k^2(x) dx. \]

If \( \int_0^1 g_k^2(x) dx > 0 \) for some \( 1 \leq k \leq d \), then \( S_{N,d} \xrightarrow{P} \infty \).

To understand when the test is consistent, introduce the jump function \( \Delta(t) = \mu_1(t) - \mu_2(t) \). By (4.3.18), the condition \( \int_0^1 g_k^2(x) dx > 0 \) is equivalent to \( \int_0^1 \Delta(s) w_k(s) ds \neq 0 \). Thus the test will have no power, if

\[ \int_0^1 \Delta(s) w_k(s) ds = 0, \quad \text{for all } 1 \leq k \leq d. \]  \hspace{1cm} (4.3.19)

By (4.2.10) and (4.2.11), (4.3.19) is equivalent to

\[ \int_0^1 c(t, s) w_k(s) ds = \gamma_k w_k(t), \quad \text{for all } 1 \leq k \leq d. \]  \hspace{1cm} (4.3.20)

Comparing to (4.2.6), we see that condition (4.3.19) means that, up to a sign, the
\( w_k, \gamma_k \) are equal to \( v_k, \lambda_k \), for \( 1 \leq k \leq d \). This lead us to the following lemma.

**Lemma 4.3.** If Assumption 4.5 holds, and the jump function \( \Delta(t) = \mu_1(t) - \mu_2(t) \) is
not orthogonal to the subspace spanned by the first \( d \) eigenfunctions of the covariance
kernel \( c(t, s) \) (4.2.3), then \( S_{N,d} \xrightarrow{P} \infty \), as \( N \to \infty \).

To estimate the change point, we plot the function \( T_N(x) \) (4.3.14) against \( 0 \leq x \leq 1 \), and estimate \( \theta \) by the value of \( x \) which maximizes \( T_N(x) \). The intuition behind
this estimator is clear from (4.3.14) and (4.3.13). To ensure uniqueness, we formally define this estimator as

\[(4.3.21) \quad \hat{\theta}_N = \inf \left\{ x : T_N(x) = \sup_{0 \leq y \leq 1} T_N(y) \right\}.\]

Its weak consistency is established in the following lemma

**Lemma 4.4.** If the assumptions of Lemma 4.3 hold, then \( \hat{\theta}_N \xrightarrow{P} \theta. \)

**Proof.** The argument \( x \) maximizing \( T_n(x) \), clearly maximizes \( A_N(x) = N^{-1}T_N(x) \). Theorem 4.2 states that \( \sup_{0 \leq x \leq 1} |A_N(x) - A(x)| \xrightarrow{P} 0 \), where

\[A(x) = g^T(x)\Sigma^*g(x) = \begin{cases} x(1 - \theta)A, & 0 \leq x \leq \theta \\ \theta(1 - x)A, & \theta < x < 1, \end{cases}\]

with

\[A = \sum_{1 \leq \ell \leq d} \frac{1}{\gamma_\ell} \left( \int \Delta(t)w_\ell(t)dt \right)^2.\]

Under the assumptions of lemma 4.3, \( A > 0 \), and it is easy to verify that \( A(x) \) has then a unique maximum at \( x = \theta \). \( \blacksquare \)

An important aspect of the procedure is the choice of the number \( d \) of the eigenfunctions \( v_k \). This issue is common to all FDA procedures using functional PCA, and several approaches have been proposed. These include an adaptation of the *scree plot* of Cattell (1966), see Kokoszka et al. (2008), the cumulative percentage variance approach used in Section 4.4.2, the pseudo AIC and the cross-validation, see Yao et al. (2005). All these methods are implemented in the MATLAB PACE package developed at the University of California at Davis. A general recommendation for the cumulative percentage variance method is to use \( d \) which explains 85% of the variance. This choice is suitable in the setting of Section 4.4.2, where \( d = 8 \) explains 84% of the variance.
4.4 Finite sample performance and application to temperature data

In this section, we report the results of a simulation study that examines the finite sample performance of the test. Recall that the test rejects if \( S_{N,d} \) (4.3.15) exceeds the \((1 - \alpha)\)th quantile of \( K_d \) (4.3.16). For \( d \leq 5 \), these quantiles were computed by Kiefer (1959) using a series expansion of the CDF of \( K_d \). Horváth et al. (1999) used these expansions to find the critical values for \( d = 12 \) and noticed that the critical values obtained by simulating \( K_d \) by discretizing the integral are slightly different, but actually lead to more accurate tests. To cover a fuller range of the \( d \) values, Table 4.1 gives simulated critical values for \( d = 1, \ldots, 30 \), computed by discretizing the integral over 1,000 points and running 100,000 replications.

The simulation study consists of two parts. First we use standard Gaussian processes as the errors \( Y_i \) and a number of rather arbitrary mean functions \( \mu \). This part assesses the test in some generic cases analogous to assuming a normal distribution of scalar observations. In the second part, we use mean functions and errors derived from monthly temperature data. No assumptions on the marginal distribution of the \( Y_i \)'s or the shape of the \( \mu \)'s are made. This part assesses the test in a specific, practically relevant setting.

4.4.1 Gaussian processes

To investigate the empirical size, without loss of generality, \( \mu(t) \) was chosen to be equal to zero and two different cases of \( Y_i(t) \) were considered, namely the trajectories of the standard Brownian motion (BM), and the Brownian bridge (BB). These processes were generated by transforming cumulative sums of independent normal variables computed on a grid of \( 10^3 \) equispaced points in \([0, 1]\). Following Ramsay and Silverman (2005) (Chapter 3) discrete trajectories were converted to functional observations (functional objects in \( \mathbb{R} \)) using B-spline and Fourier bases and various numbers of basis functions. No systematic dependence either on the type of the basis or on the number of basis functions was found.
Table 4.1: Simulated critical values of the distribution of $K_d$.

<table>
<thead>
<tr>
<th>Nominal size</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>10%</td>
<td>0.345165</td>
</tr>
<tr>
<td>5%</td>
<td>0.460496</td>
</tr>
<tr>
<td>1%</td>
<td>0.740138</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td>10%</td>
<td>1.690773</td>
</tr>
<tr>
<td>5%</td>
<td>1.895557</td>
</tr>
<tr>
<td>1%</td>
<td>2.342252</td>
</tr>
<tr>
<td></td>
<td>13</td>
</tr>
<tr>
<td>10%</td>
<td>2.884214</td>
</tr>
<tr>
<td>5%</td>
<td>3.147604</td>
</tr>
<tr>
<td>1%</td>
<td>3.708033</td>
</tr>
<tr>
<td></td>
<td>19</td>
</tr>
<tr>
<td>10%</td>
<td>4.024313</td>
</tr>
<tr>
<td>5%</td>
<td>4.327286</td>
</tr>
<tr>
<td>1%</td>
<td>4.974172</td>
</tr>
<tr>
<td></td>
<td>25</td>
</tr>
<tr>
<td>10%</td>
<td>5.159057</td>
</tr>
<tr>
<td>5%</td>
<td>5.495721</td>
</tr>
<tr>
<td>1%</td>
<td>6.203718</td>
</tr>
</tbody>
</table>
The results reported in this section were obtained using B-spline basis with 800 basis functions. We used a wide spectrum of $N$ and $d$, but to conserve space, we present the results for $N = 50, 150, 200, 300, 500$ and $d = 1, 2, 3, 4$. All empirical rejection rates are based on 1,000 replications.

Table 4.2 shows the empirical sizes based on critical values reported in Table 4.1. The empirical sizes are fairly stable. Except for a very few cases of small sample sizes, all deviations from the nominal significance levels do not exceed two standard errors computed using the normal approximation $\sqrt{p(1-p)/R}$, where $p$ is a nominal level and $R$ the number of repetitions. Table 4.2 shows that for these Gaussian processes, the empirical size does not depend appreciably either on $n$ or on $d$.

In the power study, several cases that violate the null were considered. We report the power for $k^* = \lfloor N/2 \rfloor$. Several other values of $k^*$ were also considered, and only a small loss of power was observed for $N/4 < k^* \leq 3N/4$. A few different mean functions $\mu$ before and after change were used, namely $\mu_i(t) = 0, t, t^2, \sqrt{t}, e^t, \sin(t), \cos(t), i = 1, 2$, for instance $\mu_1(t) = t$ and $\mu_2(t) = \cos(t)$, etc.

Table 4.3 presents selected results of the power study. It shows that the test has overall good power. For small samples, $N \leq 100$, in cases where the BB was used the power is slightly higher than for those with the BM. Nonetheless, for $N \geq 150$ the power approaches 100% for both processes and all choices of other parameters. The power decreases as the number of principal components $d$ increases. This can be explained as follows: the critical values of $S_{N,d}$ increase with $d$, but the change point is mainly captured by a few initial leading principal components explaining the major part of the variance.

4.4.2 Temperature data

The goal of this section is twofold: to investigate the performance of the test in a real world setting, and to demonstrate the advantages of the functional approach for high-dimensional data.
Table 4.2: Empirical size (in percent) of the test using the B-spline basis.

<table>
<thead>
<tr>
<th>Process</th>
<th>$d=1$</th>
<th></th>
<th>$d=2$</th>
<th></th>
<th>$d=3$</th>
<th></th>
<th>$d=4$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>10.3</td>
<td>4.6</td>
<td>0.1</td>
<td>9.9</td>
<td>4.8</td>
<td>0.7</td>
<td>8.4</td>
<td>3.3</td>
</tr>
<tr>
<td>BB</td>
<td>11.2</td>
<td>5.5</td>
<td>0.8</td>
<td>10.6</td>
<td>4.9</td>
<td>1.1</td>
<td>8.4</td>
<td>4.0</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>12.2</td>
<td>5.6</td>
<td>1.3</td>
<td>9.8</td>
<td>5.6</td>
<td>0.9</td>
<td>9.3</td>
<td>4.6</td>
</tr>
<tr>
<td>BB</td>
<td>12.4</td>
<td>5.7</td>
<td>0.7</td>
<td>10.2</td>
<td>4.2</td>
<td>0.6</td>
<td>9.9</td>
<td>4.6</td>
</tr>
<tr>
<td>$N = 150$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>10.8</td>
<td>5.7</td>
<td>1.3</td>
<td>9.7</td>
<td>4.6</td>
<td>1.2</td>
<td>11.8</td>
<td>6.2</td>
</tr>
<tr>
<td>BB</td>
<td>10.5</td>
<td>5.0</td>
<td>1.2</td>
<td>9.8</td>
<td>4.4</td>
<td>1.1</td>
<td>10.4</td>
<td>6.2</td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>9.7</td>
<td>5.4</td>
<td>0.8</td>
<td>9.2</td>
<td>4.3</td>
<td>0.7</td>
<td>9.3</td>
<td>5.8</td>
</tr>
<tr>
<td>BB</td>
<td>9.2</td>
<td>5.1</td>
<td>0.8</td>
<td>10.8</td>
<td>5.6</td>
<td>1.2</td>
<td>10.0</td>
<td>5.2</td>
</tr>
<tr>
<td>$N = 300$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>10.3</td>
<td>5.2</td>
<td>1.5</td>
<td>11.1</td>
<td>6.1</td>
<td>0.6</td>
<td>10.1</td>
<td>4.5</td>
</tr>
<tr>
<td>BB</td>
<td>10.4</td>
<td>5.6</td>
<td>1.1</td>
<td>9.4</td>
<td>4.8</td>
<td>0.9</td>
<td>9.9</td>
<td>4.1</td>
</tr>
<tr>
<td>$N = 500$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM</td>
<td>11.6</td>
<td>6.3</td>
<td>1.3</td>
<td>10.6</td>
<td>6.9</td>
<td>1.5</td>
<td>10.9</td>
<td>5.7</td>
</tr>
<tr>
<td>BB</td>
<td>11.7</td>
<td>5.1</td>
<td>1.3</td>
<td>9.7</td>
<td>5.8</td>
<td>1.4</td>
<td>10.3</td>
<td>5.3</td>
</tr>
</tbody>
</table>
Table 4.3: Empirical power (in percent) of the test using B-spline basis. Change point at $k^* = [n/2]$.

<table>
<thead>
<tr>
<th>Process</th>
<th>$d=1$</th>
<th></th>
<th>$d=2$</th>
<th></th>
<th>$d=3$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
<td>10%</td>
<td>5%</td>
<td>1%</td>
</tr>
<tr>
<td>$N = 50$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM; BM + sin(t)</td>
<td>81.5</td>
<td>70.8</td>
<td>43.7</td>
<td>72.6</td>
<td>60.0</td>
<td>33.2</td>
</tr>
<tr>
<td>BM; BM + t</td>
<td>88.4</td>
<td>78.0</td>
<td>54.1</td>
<td>84.7</td>
<td>74.0</td>
<td>45.4</td>
</tr>
<tr>
<td>BB; BB + sin(t)</td>
<td>99.8</td>
<td>99.4</td>
<td>97.4</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>BB; BB + t</td>
<td>99.9</td>
<td>99.8</td>
<td>98.9</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>$N = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM; BM + sin(t)</td>
<td>97.4</td>
<td>95.3</td>
<td>86.3</td>
<td>96.4</td>
<td>91.0</td>
<td>76.5</td>
</tr>
<tr>
<td>BM; BM + t</td>
<td>99.0</td>
<td>97.5</td>
<td>91.2</td>
<td>98.7</td>
<td>97.1</td>
<td>87.6</td>
</tr>
<tr>
<td>BB; BB + sin(t)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>BB; BB + t</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$N = 150$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM; BM + sin(t)</td>
<td>99.9</td>
<td>99.5</td>
<td>96.6</td>
<td>99.6</td>
<td>98.6</td>
<td>95.1</td>
</tr>
<tr>
<td>BM; BM + t</td>
<td>100</td>
<td>99.8</td>
<td>98.7</td>
<td>99.8</td>
<td>99.7</td>
<td>98.8</td>
</tr>
<tr>
<td>BB; BB + sin(t)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>BB; BB + t</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>$N = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BM; BM + sin(t)</td>
<td>100</td>
<td>99.9</td>
<td>99.1</td>
<td>100</td>
<td>99.8</td>
<td>99.0</td>
</tr>
<tr>
<td>BM; BM + t</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>99.9</td>
</tr>
<tr>
<td>BB; BB + sin(t)</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td>BB; BB + t</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
<td>100</td>
</tr>
</tbody>
</table>
The data consists of 228 years (1780 to 2007) of average daily temperatures in central England. The original data can thus be viewed as 228 curves with 365 measurements on each curve. These data were converted to functional objects in R using 12 B-spline basis functions. Multivariate observations were obtained as in Horváth et al. (1999) by computing monthly averages resulting in 228 vectors of dimension $d = 12$. (We could not even compute statistics (4.3.17) for vectors of dimension 365 because R reported that $\hat{D}$ was singular.) These two procedures are illustrated in Figure 4.1. Even though we used 12 B-splines and 12 averages, the resulting data look quite different, especially in spring and fall, when the temperatures change most rapidly. Gregorian months form a somewhat arbitrary fixed partition of the data, while the splines adapt to their shapes which differ from year to year.

To compute statistic (4.3.15), we used $d = 8$ eigenfunctions which explain 84% of variability. If the test indicates a change, we estimate it by the estimator $\hat{\theta}_N$ (4.3.21). This divides the data set into two subsets. The procedure is then repeated for each subset until periods of constant mean functions are obtained. We proceed in exactly the same manner using statistic (4.3.17). We refer to these procedures, respectively, as FDA and MDA approaches. The resulting segmentations are shown in Tables 4.4 and 4.5.

The functional approach identified two more change point, 1850 and 1992, which roughly correspond to the beginning of mass industrialization and the advent of rapid global warming. The multivariate approach “almost” identified these change points with the P-values in iterations 4 and 5 being just above the significance level of 5%. This may indicate that the functional method has better power, perhaps due to its greater flexibility in capturing the shape of the data. This conjecture is investigated below. Figure 4.2 shows average temperatures in the last four segments, and clearly illustrates the warming trend.
Table 4.4: Segmentation procedure of the data into periods with constant mean function.

<table>
<thead>
<tr>
<th>Iteration</th>
<th>Segment</th>
<th>Decision</th>
<th>$S_{n,d}$</th>
<th>P-value</th>
<th>Estimated change point</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>England temperatures ($d = 8$) (FDA approach)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1780 - 2007</td>
<td>Reject</td>
<td>8.020593</td>
<td>0.00000</td>
<td>1926</td>
</tr>
<tr>
<td>2</td>
<td>1780 - 1925</td>
<td>Reject</td>
<td>3.252796</td>
<td>0.00088</td>
<td>1808</td>
</tr>
<tr>
<td>3</td>
<td>1780 - 1807</td>
<td>Accept</td>
<td>0.888690</td>
<td>0.87404</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1808 - 1925</td>
<td>Reject</td>
<td>2.351132</td>
<td>0.02322</td>
<td>1850</td>
</tr>
<tr>
<td>5</td>
<td>1808 - 1849</td>
<td>Accept</td>
<td>0.890845</td>
<td>0.87242</td>
<td>-</td>
</tr>
<tr>
<td>6</td>
<td>1850 - 1925</td>
<td>Accept</td>
<td>1.364934</td>
<td>0.41087</td>
<td>-</td>
</tr>
<tr>
<td>7</td>
<td>1926 - 2007</td>
<td>Reject</td>
<td>2.311151</td>
<td>0.02643</td>
<td>1993</td>
</tr>
<tr>
<td>8</td>
<td>1926 - 1992</td>
<td>Accept</td>
<td>0.927639</td>
<td>0.84289</td>
<td>-</td>
</tr>
<tr>
<td>9</td>
<td>1993 - 2007</td>
<td>Accept</td>
<td>1.626515</td>
<td>0.21655</td>
<td>-</td>
</tr>
<tr>
<td><strong>England temperatures ($d = 12$) (MDA approach)</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>1780 - 2007</td>
<td>Reject</td>
<td>7.971031</td>
<td>0.00000</td>
<td>1926</td>
</tr>
<tr>
<td>2</td>
<td>1780 - 1925</td>
<td>Reject</td>
<td>3.576543</td>
<td>0.00764</td>
<td>1815</td>
</tr>
<tr>
<td>3</td>
<td>1780 - 1814</td>
<td>Accept</td>
<td>1.534223</td>
<td>0.81790</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>1815 - 1925</td>
<td>Accept</td>
<td>2.813596</td>
<td>0.07171</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>1926 - 2007</td>
<td>Accept</td>
<td>2.744801</td>
<td>0.08662</td>
<td>-</td>
</tr>
</tbody>
</table>
Fig. 4.1: Daily temperatures in 1916 with monthly averages and functional object obtained by smoothing with $B$-splines.
Fig. 4.2: Graph of mean functions of last four segments.
Table 4.5: Summary and comparison of segmentation. Beginning and end of data period in bold.

<table>
<thead>
<tr>
<th>Approach</th>
<th>Change points</th>
</tr>
</thead>
<tbody>
<tr>
<td>FDA</td>
<td>1780 1808 1850 1926 1992 2007</td>
</tr>
<tr>
<td>MDA</td>
<td>1780 1815 1926 2007</td>
</tr>
</tbody>
</table>

The analysis presented above assumes a simple functional change point model for the daily temperatures. Obviously, one cannot realistically believe that the mean curves change abruptly in one year, this is merely a modeling assumption useful in identifying patterns of change in mean temperature curves. Well-established alternative modeling approaches have been used to study the variability of temperatures. For example, Hosking (1984) fitted a fractionally differenced ARMA(1,1) model to the series of annual average temperatures in central England in 1659–1976. It is generally very difficult to determine on purely statistical grounds if a change–point or a long–range dependent model is more suitable for any particular finite length record, see Berkes et al. (2006) and Jach and Kokoszka (2008) for recent methodology, discussion and references. It is often more useful to choose a modeling methodology which depends on specific goals, and this is the approach we use. One way of checking an approximate adequacy of our model is to check if the residuals obtained after subtracting the mean in each segment are approximately independent and identically distributed. This can be done by applying the test developed by Gabrys and Kokoszka (2007) which is a functional analog of the well–known test of Hosking (1980) and Li and McLeod (1981) [see also Hosking (1981, 1989)]. The P-value of 8% indicates the acceptance of the hypothesis that the residuals are iid.

Keeping these caveats in mind, we use the partitions obtained above to generate realistic synthetic data with and without change–points. We use them to evaluate and compare the size and power properties of the FDA and MDA tests, and to validate our findings. We compute the residuals of every observation in a constant mean segment by subtracting the average of the segment, i.e. \( \hat{Y}_{is} = X_{is} - \hat{\mu}_s \), where \( s = 1, \ldots, S \).
Table 4.6: Empirical size of the test for models derived from the temperature data.

<table>
<thead>
<tr>
<th>Segment</th>
<th>Number of functions</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
<th>10%</th>
<th>5%</th>
<th>1%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Case I</td>
<td></td>
<td></td>
<td>Case II</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FDA approach ($d = 8$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1780 - 1807 ($\Delta_1$)</td>
<td>28</td>
<td>8.0</td>
<td>3.0</td>
<td>0.1</td>
<td>7.6</td>
<td>2.5</td>
<td>0.2</td>
</tr>
<tr>
<td>1808 - 1849 ($\Delta_2$)</td>
<td>42</td>
<td>9.5</td>
<td>3.9</td>
<td>0.4</td>
<td>9.7</td>
<td>4.1</td>
<td>0.4</td>
</tr>
<tr>
<td>1850 - 1925 ($\Delta_3$)</td>
<td>76</td>
<td>10.0</td>
<td>4.7</td>
<td>0.7</td>
<td>10.2</td>
<td>4.3</td>
<td>0.7</td>
</tr>
<tr>
<td>1926 - 1992 ($\Delta_4$)</td>
<td>66</td>
<td>8.8</td>
<td>3.7</td>
<td>0.8</td>
<td>9.2</td>
<td>4.1</td>
<td>1.0</td>
</tr>
<tr>
<td>1993 - 2007 ($\Delta_5$)</td>
<td>16</td>
<td>3.8</td>
<td>0.3</td>
<td>0.0</td>
<td>3.3</td>
<td>0.1</td>
<td>0.0</td>
</tr>
<tr>
<td>MDA approach ($d = 12$)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1780 - 1807 ($\Delta_1$)</td>
<td>28</td>
<td>3.0</td>
<td>0.5</td>
<td>0.0</td>
<td>2.8</td>
<td>0.4</td>
<td>0.0</td>
</tr>
<tr>
<td>1808 - 1849 ($\Delta_2$)</td>
<td>42</td>
<td>5.3</td>
<td>2.3</td>
<td>0.1</td>
<td>5.4</td>
<td>1.3</td>
<td>0.0</td>
</tr>
<tr>
<td>1850 - 1925 ($\Delta_3$)</td>
<td>76</td>
<td>6.9</td>
<td>1.9</td>
<td>0.0</td>
<td>9.1</td>
<td>4.2</td>
<td>0.6</td>
</tr>
<tr>
<td>1926 - 1992 ($\Delta_4$)</td>
<td>66</td>
<td>7.9</td>
<td>3.3</td>
<td>0.5</td>
<td>7.4</td>
<td>2.7</td>
<td>0.2</td>
</tr>
<tr>
<td>1993 - 2007 ($\Delta_5$)</td>
<td>16</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

denotes the segment, and $i = 1, \ldots, I_s$ indexes observations in the $s$th segment. The $\hat{Y}_{is}$ are functional residuals, and their average in each segment is clearly the zero function.

To assess the empirical size, we simulate “temperature-like” data by considering two cases. *Case I*: for every constant mean segment $s$, we produce synthetic observations by adding to its mean function $\hat{\mu}_s$ errors drawn from the empirical distribution of the residuals of that segment, i.e. synthetic (bootstrap) observations in the $s$th segment are generated via $X_{is}^* = \hat{\mu}_s + \hat{Y}_{i^*_s}$, where $i^*_s$ indicates that $\hat{Y}_{i^*_s}$ is obtained by drawing with replacement from $\{\hat{Y}_{1s}, \ldots, \hat{Y}_{Is}\}$. *Case II*: We compute residuals in each segment and pool them together. We use this larger set of residuals to create new observations by adding to the average of a segment the errors drawn with replacement from that pool of residuals. For each segment, we generate 1000 of these bootstrap sequences. Table 4.6 shows the the resulting empirical sizes. As the sample size increases, the FDA rejection rates approach nominal sizes, while the MDA test is much more conservative. For the 1993–2007 segment, the size is not reported because the matrix $D$ was (numerically) singular for most bootstrap replications.
We next investigate the power. Three cases are considered. Case I: For each segment, we produce synthetic observations using the bootstrap procedure and sampling residuals from a corresponding period. This means that the errors in each segment come from possibly different distributions. Case II: We pool together two, three, four, or five sets of residuals (depending on how many constant mean segments we consider) and sample from that pool to produce new observations. This means that the errors in each segment come from the same distribution. Case III: We slightly modify Case II by combining all residuals from all segments into one population and use it to produce new observations. In both Case II and Case III, the theoretical assumptions of Section 4.2 are satisfied, cf. Assumption 4.2, i.e. the means change, but the errors come from the same population. Table 4.7 shows the power of the test for FDA approach and Table 4.8 presents results of discrete MDA method. As seen in Table 4.7, the differences between the three cases are of the order of the chance error. Table 4.7 shows that the test has excellent power, even in small samples, both for single and multiple change points. As for the Gaussian processes, power is slightly higher if there is a change point around the middle of the sample. Comparing Tables 4.7 and 4.8, it is seen that in FDA approach dominates the MDA approach. There are a handful of cases, indicated with *, when MDA performed better, but their frequency and the difference size suggests that this may be attributable to the chance error.

4.5 Proof of Theorems 4.1 and 4.2

A key element of the proofs in bound (4.5.26), which follows from a functional central limit theorem in a Hilbert space. A result of this type is needed because the observations $X_i(\cdot)$ are elements of a Hilbert space, and to detect a change point, we must monitor the growth of the partial sums $\sum_{1 \leq i \leq Nx} X_i(t)$ which are a function of $0 < x < 1$ (and of $t \in \mathcal{T}$).

Lemma 4.5 is particularly noteworthy because it shows that the eigenvalues and the eigenfunctions also converge under the alternative.
Table 4.7: Empirical power of the test for change-point models derived from temperature data (FDA approach).

<table>
<thead>
<tr>
<th>Segment</th>
<th>Sample size</th>
<th>Change point(s) $\theta$</th>
<th>Nominal level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10% 5% 1%</td>
</tr>
<tr>
<td>England ($d = 8$) (FDA approach)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2$</td>
<td>70</td>
<td>.41</td>
<td>85.6 76.8 49.7</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_3$</td>
<td>104</td>
<td>.28</td>
<td>86.2 75.8 47.4</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_4$</td>
<td>94</td>
<td>.31</td>
<td>100 100 98.7</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_5$</td>
<td>44</td>
<td>.66</td>
<td>100 99.9 93.4</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_3$</td>
<td>118</td>
<td>.36</td>
<td>87.9* 78.5 52.8</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_4$</td>
<td>108</td>
<td>.40</td>
<td>99.7 99.0 95.6</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_5$</td>
<td>58</td>
<td>.74</td>
<td>99.2 97.8 86.3</td>
</tr>
<tr>
<td>$\Delta_3, \Delta_4$</td>
<td>142</td>
<td>.54</td>
<td>99.9* 99.5* 99.1</td>
</tr>
<tr>
<td>$\Delta_3, \Delta_5$</td>
<td>92</td>
<td>.84</td>
<td>99.1 96.7 82.9</td>
</tr>
<tr>
<td>$\Delta_4, \Delta_5$</td>
<td>82</td>
<td>.82</td>
<td>93.0 85.0 58.8</td>
</tr>
</tbody>
</table>
Table 4.8: Empirical power of the test for change-point models derived from temperature data (MDA approach).

<table>
<thead>
<tr>
<th>Segment</th>
<th>Sample size</th>
<th>Change point(s) $\theta$</th>
<th>Nominal level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Case I</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>10%</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2$</td>
<td>70</td>
<td>.41</td>
<td>82.9</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_3$</td>
<td>104</td>
<td>.28</td>
<td>79.7</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_4$</td>
<td>94</td>
<td>.31</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_5$</td>
<td>44</td>
<td>.66</td>
<td>98.4</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_3$</td>
<td>118</td>
<td>.36</td>
<td>88.3</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_4$</td>
<td>108</td>
<td>.40</td>
<td>97.3</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_5$</td>
<td>58</td>
<td>.74</td>
<td>93.9</td>
</tr>
<tr>
<td>$\Delta_3, \Delta_4$</td>
<td>142</td>
<td>.54</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_3, \Delta_5$</td>
<td>92</td>
<td>.84</td>
<td>98.2</td>
</tr>
<tr>
<td>$\Delta_4, \Delta_5$</td>
<td>82</td>
<td>.82</td>
<td>78.4</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_3$</td>
<td>146</td>
<td>.20 .49</td>
<td>97.5</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_4$</td>
<td>136</td>
<td>.21 .52</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_5$</td>
<td>86</td>
<td>.34 .83</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_3, \Delta_4$</td>
<td>184</td>
<td>.23 .65</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_3, \Delta_5$</td>
<td>134</td>
<td>.32 .89</td>
<td>99.8</td>
</tr>
<tr>
<td>$\Delta_3, \Delta_4, \Delta_5$</td>
<td>158</td>
<td>.49 .91</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_3, \Delta_4$</td>
<td>212</td>
<td>.14 .33 .69</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_3, \Delta_5$</td>
<td>162</td>
<td>.18 .44 .91</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_2, \Delta_3, \Delta_4, \Delta_5$</td>
<td>200</td>
<td>.22 .60 .93</td>
<td>100</td>
</tr>
<tr>
<td>$\Delta_1, \Delta_2, \Delta_3, \Delta_4, \Delta_5$</td>
<td>228</td>
<td>.13 .31 .64 .93</td>
<td>100</td>
</tr>
</tbody>
</table>
4.5.1 Proof of Theorems 4.1

We will work with the unobservable projections

$$\tilde{\beta}_{i,k} = \int Y_i(t)\hat{v}_k(t)dt, \quad \beta_{i,k} = \int Y_i(t)v_k(t)dt, \quad \beta^*_{i,k} = \hat{c}_k\beta_{i,k}$$

and the vectors

$$\beta_{i} = [\beta_{i,1}, \ldots, \beta_{i,d}]^T, \quad \beta^*_{i} = [\beta^*_{i,1}, \ldots, \beta^*_{i,d}]^T, \quad 1 \leq i \leq N.$$

Since the $Y_i$ are iid functions with mean zero, the $\beta_i$ are iid mean zero vectors in $R^d$. A simple calculation using the orthonormality of the $v_k$ shows that each $\beta_i$ has a diagonal covariance matrix

$$\Sigma_d = \begin{bmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_d
\end{bmatrix}$$

The functional central limit theorem, thus implies that

$$N^{-1/2} \sum_{1 \leq i \leq Nx} \beta_{i} \overset{d}{\to} \Delta_d(x) \quad (0 \leq x \leq 1),$$

where the convergence is in the Skorokhod space $D^d[0,1]$. The process $\{\Delta_d(x), \ 0 \leq x \leq 1\}$ takes values in $R^d$, has zero mean and covariance matrix $\Sigma_d$. Convergence (4.5.22) implies in turn that

$$\frac{1}{N}\left[ \sum_{1 \leq i \leq Nx} \beta_{i} - x \sum_{1 \leq i \leq N} \beta_{i} \right]^T \Sigma_d^{-1} \left[ \sum_{1 \leq i \leq Nx} \beta_{i} - x \sum_{1 \leq i \leq N} \beta_{i} \right] \overset{d}{\to} \sum_{1 \leq i \leq d} B^2_i(x)$$

in the Skorokhod space $D[0,1]$. 

The matrix $\Sigma_d$ is estimated by $\hat{\Sigma}_d$. By (4.2.8) and Assumption 4.2, $\hat{\Sigma}_d^{-1} \xrightarrow{P} \Sigma_d^{-1}$, so (4.5.23) yields

\[ (4.5.24) \quad \frac{1}{N} \left[ \sum_{1 \leq i \leq Nx} \beta_i - x \sum_{1 \leq i \leq N} \beta_i \right]^T \hat{\Sigma}_d^{-1} \left[ \sum_{1 \leq i \leq Nx} \beta_i - x \sum_{1 \leq i \leq N} \beta_i \right] \xrightarrow{d} \sum_{1 \leq i \leq d} B_i^2(x). \]

Note that

\[ \sum_{1 \leq i \leq Nx} \beta^*_{i,k} - x \sum_{1 \leq i \leq N} \beta^*_{i,k} = \hat{c}_k \left( \sum_{1 \leq i \leq Nx} \beta_{i,k} - x \sum_{1 \leq i \leq N} \beta_{i,k} \right). \]

Since $\hat{c}_k^2 = 1$, we can replace the $\beta_i$ in (4.5.24) by the $\beta^*_i$, and obtain

\[ (4.5.25) \quad \frac{1}{N} \left[ \sum_{1 \leq i \leq Nx} \beta^*_i - x \sum_{1 \leq i \leq N} \beta^*_i \right]^T \hat{\Sigma}_d^{-1} \left[ \sum_{1 \leq i \leq Nx} \beta^*_i - x \sum_{1 \leq i \leq N} \beta^*_i \right] \xrightarrow{d} \sum_{1 \leq i \leq d} B_i^2(x). \]

We now turn to the effect of replacing the $\beta^*_{i,k}$ by $\hat{\beta}_{i,k}$. Observe that

\[ \sup_{0 < x < 1} \left| N^{-1/2} \sum_{1 \leq i \leq Nx} \beta^*_{i,k} - N^{-1/2} \sum_{1 \leq i \leq N} \hat{\beta}_{i,k} \right| \]

\[ = \sup_{0 < x < 1} \left| \int \left( N^{-1/2} \sum_{1 \leq i \leq Nx} Y_i(t) \right) (\hat{c}_k v_k(t) - \hat{v}_k(t)) \, dt \right| \]

\[ \leq \sup_{0 < x < 1} \left[ \int \left( N^{-1/2} \sum_{1 \leq i \leq Nx} Y_i(t) \right)^2 \, dt \right]^{1/2} \left[ \int (\hat{c}_k v_k(t) - \hat{v}_k(t))^2 \, dt \right]^{1/2}. \]

The first factor is bounded in probability, i.e.

\[ (4.5.26) \quad \sup_{0 < x < 1} \int \left( N^{-1/2} \sum_{1 \leq i \leq Nx} Y_i(t) \right)^2 \, dt = O_P(1). \]
Relation (4.5.26) follows from the weak convergence in $D([0, 1], L^2(\mathcal{F}))$ of the partial sum process $\sum_{1 \leq i \leq Nx} Y_i, \ x \in [0, 1]$, see e.g. Kuelbs (1973).

Combining (4.5.26) and (4.2.8), we obtain

$$\sup_{0 < x < 1} \left| N^{-1/2} \sum_{1 \leq i \leq Nx} \beta_{i,k}^*- N^{-1/2} \sum_{1 \leq i \leq Nx} \hat{\beta}_{i,k} \right| \xrightarrow{P} 0,$$

which in turn implies that

$$(4.5.27) \left\| \left[ \sum_{1 \leq i \leq Nx} \beta_i^* - x \sum_{1 \leq i \leq N} \beta_i^* \right] - \left[ \sum_{1 \leq i \leq Nx} \hat{\beta}_i - x \sum_{1 \leq i \leq N} \hat{\beta}_i \right] \right\| = o_P(N^{-1/2}),$$

where the norm is the Euclidean norm in $R^d$. Relations (4.5.25) and (4.5.27) yield the claim in Theorem 4.1. ■

4.5.2 Proof of Theorem 4.2

Theorem 4.2 follows from relation (4.5.31) and Lemma 4.6. To establish them, we need the following Lemma.

Lemma 4.5. Under assumption 4.2, for every $1 \leq k \leq d$, as $N \to \infty$,

$$(4.5.28) \hat{\lambda}_k \xrightarrow{P} \gamma_k,$$

$$(4.5.29) \int [\hat{v}_k(t) - \hat{c}_kw_k(t)]^2 dt \xrightarrow{P} 0,$$

where $\hat{v}_k, \hat{\lambda}_k$ are defined by (4.2.7), $w_k, \gamma_k$ by (4.2.11) and $\hat{c}_k = \text{sign} \int_{\mathcal{T}} v_k(t)\hat{v}_k(t)dt$.

Proof. It is easy to see that

$$X_N(t) = Y_N(t) + \frac{k^*}{N} \mu_1(t) + \frac{N - k^*}{N} \mu_2(t)$$
and, denoting $\Delta(t) = \mu_1(t) - \mu_2(t)$,

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

$$= \frac{1}{N} \sum_{1 \leq i \leq k^*} \left( Y_i(t) - \bar{Y}_N(t) + \mu_1(t) - \frac{k^*}{N} \mu_1(t) - \frac{N - k^*}{N} \mu_2(t) \right)$$

$$\times \left( Y_i(s) - \bar{Y}_N(s) + \mu_1(s) - \frac{k^*}{N} \mu_1(s) - \frac{N - k^*}{N} \mu_2(s) \right)$$

$$+ \frac{1}{N} \sum_{k^* < i \leq N} \left( Y_i(t) - \bar{Y}_N(t) + \mu_2(t) - \frac{k^*}{N} \mu_1(t) - \frac{N - k^*}{N} \mu_2(t) \right)$$

$$\times \left( Y_i(s) - \bar{Y}_N(s) + \mu_2(s) - \frac{k^*}{N} \mu_1(s) - \frac{N - k^*}{N} \mu_2(s) \right)$$

$$= \frac{1}{N} \sum_{1 \leq i \leq k^*} \left( Y_i(t) - \bar{Y}_N(t) + \frac{N - k^*}{N} \Delta(t) \right) \left( Y_i(s) - \bar{Y}_N(s) + \frac{N - k^*}{N} \Delta(s) \right)$$

$$+ \frac{1}{N} \sum_{k^* < i \leq N} \left( Y_i(t) - \bar{Y}_N(t) - \frac{k^*}{N} \Delta(t) \right) \left( Y_i(s) - \bar{Y}_N(s) - \frac{k^*}{N} \Delta(s) \right).$$

Rearranging terms, we obtain

$$\hat{c}_N(t, s) = \frac{1}{N} \sum_{i=1}^{N} (Y_i(t) - \bar{Y}_N(t))(Y_i(s) - \bar{Y}_N(s)) + \frac{k^*}{N} \left( 1 - \frac{k^*}{N} \right) \Delta(t) \Delta(s) + r_N(t, s),$$

where

$$r_N(t, s) = \left( 1 - \frac{k^*}{N} \right) \frac{1}{N} \sum_{1 \leq i \leq k^*} \left[ (Y_i(t) - \bar{Y}_N(t)) \Delta(s) + (Y_i(s) - \bar{Y}_N(s)) \Delta(t) \right]$$

$$+ \frac{k^*}{N} \frac{1}{N} \sum_{k^* < i \leq N} \left[ (Y_i(t) - \bar{Y}_N(t)) \Delta(s) + (Y_i(s) - \bar{Y}_N(s)) \Delta(t) \right].$$

Using the law of large numbers for independent, identically distributed Hilbert space valued random variables (see e.g. Theorem 2.4 of Bosq (2000)), we obtain $\int_{\tau} \int_{\tau} r_N^2(t, s) dt ds \xrightarrow{P}$
\(0\) and

\[(4.5.30)\]  
\[\int \int [\hat{c}_N(t, s) - \tilde{c}_N(t, s)]^2 P \to 0.\]

Hence Lemmas 4.2 and 4.3 of Bosq (2000) imply, respectively, (4.5.28) and (4.5.29).

As an immediate corollary to (4.5.28), we obtain

\[(4.5.31)\]  
\[\hat{\Sigma}^{-1} P \to \Sigma^*.\]

**Lemma 4.6.** Under Assumption 4.2,

\[
\sup_{0 \leq x \leq 1} \left| \frac{1}{N} \left[ \sum_{1 \leq i \leq N_x} \hat{\eta}_{i,k} - x \sum_{1 \leq i \leq N} \hat{\eta}_{i,k} \right] - c_k g_k(x) \right| = o_P(1),
\]

with the functions \(g_k\) defined by (4.3.18).

**Proof.** Denote

\[
\hat{g}_k(x) = \frac{1}{N} \left[ \sum_{1 \leq i \leq N_x} \hat{\eta}_{i,k} - x \sum_{1 \leq i \leq N} \hat{\eta}_{i,k} \right], \quad x \in [0, 1],
\]

and observe that

\[
\hat{\eta}_{i,k} = \int Y_i(t) \hat{v}_k(t) dt + \int \mu_1(t) \hat{v}_k(t) dt - \int \bar{X}_N(t) \hat{v}_k(t) dt, \quad \text{if } 1 \leq i \leq k^*
\]

and

\[
\hat{\eta}_{i,k} = \int Y_i(t) \hat{v}_k(t) dt + \int \mu_2(t) \hat{v}_k(t) dt - \int \bar{X}_N(t) \hat{v}_k(t) dt, \quad \text{if } k^* < i \leq N.
\]
We will use the relation

\[(4.5.32) \quad \sup_{0 < x < 1} \left| \sum_{1 \leq i \leq N x} Y_i(t) \hat{v}_k(t) dt \right| = O_P(N^{1/2}),\]

which follows from (4.5.26).

Suppose first that $0 < x \leq \theta$. Then, by (4.5.32) and (4.5.29), uniformly in $x \in [0, 1],$

\[\hat{g}_k(x) = x(1 - \theta) \left[ \int \mu_1(t) \hat{v}_k(t) dt - \int \mu_2(t) \hat{v}_k(t) dt \right] + o_P(N^{-1/2})\]

\[= x(1 - \theta) \hat{c}_k \left[ \int \mu_1(t) w_k(t) dt - \int \mu_2(t) w_k(t) dt \right] + o_P(1).\]

If $x > \theta$, then, uniformly in $x \in [0, 1],$

\[\hat{g}_k(x) = \theta(1 - x) \hat{c}_k \left[ \int \mu_1(t) w_k(t) dt - \int \mu_2(t) w_k(t) dt \right] + o_P(1). \]
CHAPTER 5
ESTIMATION OF A CHANGE–POINT IN THE MEAN FUNCTION OF
FUNCTIONAL DATA

5.1 Introduction

Functional data analysis (FDA) has been enjoying increased attention over the last decade due to its applicability to problems which are difficult to cast into a framework of scalar or vector observations. Even if such standard approaches are available, the functional approach often leads to a more natural and parsimonious description of the data, and to more accurate inference and prediction results, see, for example, Antoniadis and Sapatinas (2003, 2007), Chiou et al. (2004), Fernández de Castro et al. (2005), Laukaitis and Račkauskas (2005), Müller and Stadtmüller (2005), Yao et al. (2005), and Glendinning and Fleet (2007). Both inferential and exploratory tools of FDA can however be severely biased if the stochastic structure of the data changes at some unknown point within the sample. In the scalar context, this issue has received considerable attention, see Cobb (1978), Inclán and Tiao (1994), Davis et al. (1995), Antoch et al. (1997), Garcia and Ghysels (1998), Horváth et al. (1999), Kokoszka and Leipus (2000), among many others.

The most important change that can occur in the functional context is the change of the mean function. This paper investigates large sample properties of an estimator of such a change–point. We consider both the case of a fixed size change and a contiguous change whose size approaches zero as the sample size increases. Specifically, we assume that the functional observations $X_1, \ldots, X_n$ are defined on a compact set
and follow the model

\[(5.1.1) \quad X_i = \mu + \Delta \mathbb{1}\{i > k^*\} + Y_i, \quad i = 1, \ldots, n,\]

where \(\mu\) and \(\Delta \neq 0\) are unknown, square integrable and deterministic functions over \(\mathcal{T}\), and \(Y_1, \ldots, Y_n\) are independent, identically distributed zero mean random elements of \(L^2(\mathcal{T})\) with covariance function

\[K(s, t) = \mathbb{E}[Y_1(s)Y_1(t)], \quad s, t \in \mathcal{T},\]

satisfying \(\mathbb{E}[\|Y_1\|^2] = \int_{\mathcal{T}} \mathbb{E}[Y_1^2(t)]dt < \infty\). The unknown integer \(k^* \in \{1, \ldots, n\}\) is called the change-point. We assume that

\[(5.1.2) \quad k^* = [\theta n] \quad \text{with some fixed } \theta \in (0, 1].\]

Model (5.1.1) describes a sequence of functional observations which suffer from a mean change if \(k^* < n\) or, equivalently, if \(\theta < 1\). The corresponding hypothesis testing problem

\[H_0 : k^* = n \quad \text{vs} \quad H_A : k^* < n\]

has been addressed in Berkes et al. (2009a). To explain their results and present our contribution, we must state several consequences of the assumptions made so far. First, Mercer’s theorem (see Chapter 4 of Indritz, 1963) implies that, under the null hypothesis, there is a spectral decomposition for the covariance operator \(K(s, t)\), namely

\[K(s, t) = \sum_{\ell=1}^{\infty} \lambda_\ell \varphi_\ell(s)\varphi_\ell(t), \quad s, t \in \mathcal{T},\]

where \(\lambda_\ell\) and \(\varphi_\ell\) denote the eigenvalues and eigenfunctions of \(K(s, t)\), respectively. These can be obtained as the solutions of the equation system \(\int_{\mathcal{T}} K(s, t)\varphi_\ell(t)dt = \lambda_\ell \varphi_\ell(s)\) with \(s, t \in \mathcal{T}\). Since the eigenfunctions form a complete orthonormal basis in
$\mathcal{L}^2(\mathcal{T})$ and all eigenvalues of $K(s, t)$ are nonnegative, they lead to the the Karhunen–Loève representation (in $\mathcal{L}^2(\mathcal{T})$, not pointwise in $t \in \mathcal{T}$)

$$Y_i(t) = \sum_{\ell=1}^{\infty} \sqrt{\lambda_\ell} \rho_{i,\ell} \varphi_\ell(t), \quad t \in \mathcal{T}, \ i = 1, \ldots, n,$$

where $\sqrt{\lambda_\ell} \rho_{i,\ell} = \int_{\mathcal{T}} Y_i(t) \varphi_\ell(t) dt$ is called the $\ell$th functional principal component score. It is also implied that the sequences $(\rho_{i,\ell})_{\ell \geq 1}$ consist of uncorrelated random variables with zero mean and unit variance and that, for $i \neq j$, $(\rho_{i,\ell})_{\ell \geq 1}$ and $(\rho_{j,\ell})_{\ell \geq 1}$ are independent.

For the statistical analysis, the population eigenvalues and eigenfunctions have to be replaced by their estimated versions. These are based on the estimated covariance operator

(5.1.3)  $$\hat{K}(s, t) = \frac{1}{n} \sum_{i=1}^{n} [X_i(s) - \bar{X}_n(s)][X_i(t) - \bar{X}_n(t)],$$

where $\bar{X}_n = n^{-1}(X_1 + \ldots + X_n)$. From this, estimated eigenvalues $\hat{\lambda}_\ell$ and eigenfunctions $\hat{\varphi}_\ell$ can then be derived as the solutions of the equations

$$\int_{\mathcal{T}} \hat{K}(s, t) \hat{\varphi}_\ell(t) dt = \hat{\lambda}_\ell \hat{\varphi}_\ell(s).$$

We make the assumption that, for some fixed $\theta > 0$,

(5.1.4)  $$\lambda_1 > \lambda_2 > \ldots > \lambda_d > \lambda_{d+1} \geq 0,$$

which together with the assumption of finite fourth moment of the $Y_i$ guarantees that the estimated and population eigenvalues and eigenfunctions are sufficiently close under $H_0$, see Chapter 4 of Bosq (2000) and Dauxois et al. (1982).

The hypothesis test for $H_0$ versus $H_A$ in Berkes et al. (2009a) is based on the projection of the functions $\bar{X}_{[nx]} - \bar{X}_n$, $x \in (0, 1)$, on the space spanned by the first
$d$ estimated eigenfunctions $\hat{\phi}_1, \ldots, \hat{\phi}_d$. The corresponding estimated scores are
\[
\hat{\eta}_{i,\ell} = \int T [X_i(t) - \bar{X}_n(t)] \hat{\phi}_\ell(t) dt.
\]
Berkes et al. (2009a) introduced the test statistic
\[
S_{n,d} = \frac{1}{n^2} \sum_{\ell=1}^{d} \frac{1}{\lambda_\ell} \sum_{k=1}^{n} \left( \sum_{i=1}^{k} \hat{\eta}_{i,\ell} - \frac{k}{n} \sum_{i=1}^{n} \hat{\eta}_{i,\ell} \right)^2
\]
and established its limit distribution under the null hypothesis, as well as its consistency under the alternative. For the convenience of the reader, these results are stated as a theorem.

**Theorem 5.1.** Let $\mathbb{E}[\|Y_1\|^4] < \infty$. Then, it holds under $H_0$ that
\[
S_{n,d} \xrightarrow{D} \sum_{\ell=1}^{d} \int_0^1 B_{\ell}^2(x) dx \quad (n \to \infty),
\]
where $\xrightarrow{D}$ indicates convergence in distribution and $(B_\ell(x) : x \in [0, 1]), 1 \leq \ell \leq d,$ denotes independent standard Brownian bridges. If $\Delta$ is not orthogonal to the subspace spanned by the eigenfunctions $\varphi_1, \ldots, \varphi_d$, then it holds under $H_A$ that $S_{n,d} \xrightarrow{P} \infty$ as $n \to \infty$.

While the theorem guarantees in its second part that $S_{n,d}$ will eventually detect a change given that there are sufficiently many observations, it does not contain information on how to locate the change-point, and what the distributional properties of an appropriate estimator are. The main aim of the present paper is therefore to introduce an estimator $\hat{k}_n^*$ for $k^*$ and to derive its limit distribution under different assumptions on the function $\Delta$ which determines the type of change. This will be done in Section 5.2. In Section 5.3, we evaluate the finite sample behavior via a small simulation study. All proofs are relegated to Section 5.4.
5.2 Change-point estimator and its limit distribution

It is assumed throughout this section that the alternative hypothesis $H_A$ holds true. Letting $\mathbf{x}^T$ denote the transpose of a vector $\mathbf{x}$, define $\hat{\mathbf{\eta}}_i = (\hat{\eta}_{i,1}, \ldots, \hat{\eta}_{i,d})^T$ and the diagonal matrix $\hat{\Sigma} = \text{diag}(\hat{\lambda}_\ell : \ell = 1, \ldots, d)$. Introducing the quantities

$$\hat{\kappa}_n(k) = \sum_{i=1}^k \hat{\eta}_i - \frac{k}{n} \sum_{i=1}^n \hat{\eta}_i,$$

and the quadratic forms

$$\hat{Q}_n(k) = \frac{1}{n} \hat{\kappa}_n(k)^T \hat{\Sigma}^{-1} \hat{\kappa}_n(k),$$

a suitable estimator for $k^*$ is given by

$$\hat{k}^*_n = \min \left\{ k : \hat{Q}_n(k) = \max_{1 \leq j \leq n} \hat{Q}_n(j) \right\}.$$  

With this procedure, we select as change-point the time $k$ that maximizes the random quadratic form $\hat{Q}_n(k)$ which is directly linked to the test statistic $S_{n,d}$ from the previous section via the equality $S_{n,d} = \int_0^1 \hat{Q}_n([nx]) \, dx$. Because $\hat{Q}_n(k)$ lives on the subspace spanned by the first $d$ estimated eigenfunctions $\hat{\phi}_1, \ldots, \hat{\phi}_d$ of the covariance operator $\hat{K}(s, t)$, we need to determine the behavior of $\hat{K}(s, t)$ under $H_A$. Due to the additional $\Delta$ appearing after the change-point $k^*$, it cannot be expected that $\hat{K}(s, t)$ provides an estimator for $K(s, t)$ anymore. Indeed, the following holds true instead. If we let

$$K_A(s, t) = K(s, t) + \theta (1 - \theta) \Delta(t) \Delta(s), \quad s, t \in \mathcal{T},$$

then $K_A(s, t)$ is symmetric, square integrable and positive-definite, so it admits a representation

$$K_A(s, t) = \sum_{j=1}^\infty \gamma_{\ell \ell} \psi_\ell(s) \psi_\ell(t).$$
with eigenfunctions $\psi_\ell$ and eigenvalues $\gamma_\ell$ obtained from solving the system $\int_T K_A(s,t)\psi_\ell(t)dt = \gamma_\ell \psi_\ell(s)$. The relation between the pairs $(\gamma_\ell, \psi_\ell)$ and $(\hat{\lambda}_\ell, \hat{\varphi}_\ell)$ is established in lemma 5.2 whose proof is given in Berkes et al. (2009a).

**Lemma 5.2.** Under $H_A$ it holds that, for all $1 \leq \ell \leq d$,

(i) $|\hat{\lambda}_\ell - \gamma_\ell| = o_P(1)$ as $n \to \infty$ and

(ii) $\|\hat{\varphi}_\ell - \hat{c}_\ell \psi_\ell\| = o_P(1)$ as $n \to \infty$,

where $\hat{c}_\ell = \text{sign} \int_T \psi_\ell(t) \hat{\varphi}_\ell(t) dt$.

The lemma identifies $\gamma_\ell$ and $\psi_\ell$ (up to a sign) as the stochastic limits of their estimated versions $\hat{\lambda}_\ell$ and $\hat{\varphi}_\ell$. As a consequence, it implies that the limit distribution of $\hat{k}^*_n$ depends on the behavior of the projection of $\Delta$ on the subspace spanned by the eigenfunctions $\psi_1, \ldots, \psi_d$. For $1 \leq \ell \leq d$, denote by

$$\zeta_{i,\ell} = \sqrt{\gamma_\ell} \xi_{i,\ell} = \int_T Y_i(t) \psi_\ell(t) dt \quad \text{and} \quad \beta_\ell = \sqrt{\gamma_\ell} \delta_\ell = \int_T \Delta(t) \psi_\ell(t) dt$$

the principal component scores and set

$$\zeta_i = (\zeta_{i,1}, \ldots, \zeta_{i,d})^T, \quad \xi_i = (\xi_{i,1}, \ldots, \xi_{i,d})^T, \quad \delta = (\delta_1, \ldots, \delta_d)^T.$$

We distinguish two cases

(5.2.2) $\delta \neq 0$ is constant

and

(5.2.3) $\delta = \delta_n \neq 0$ such that $\|\delta_n\|_2 \to 0$ ($n \to \infty$),

where $\| \cdot \|_2$ denotes Euclidean norm on $\mathbb{R}^d$. Assumptions (5.2.2) and (5.2.3) reflect two common approaches to deriving an asymptotic distribution of change point estimators, see for example, Csörgő and Horváth (1997) and references therein.
We first state the result for the case (5.2.2).

**Theorem 5.3.** Let $E[\|Y_1\|^4] < \infty$. If $\delta \neq 0$ is constant, then it holds under $H_A$ that

$$\hat{k}_n^* - k^* \xrightarrow{p} \min \{ k : P(k) = \sup_j P(j) \} \quad (n \to \infty),$$

where

$$P(k) = \begin{cases} 
(1 - \theta)\|\delta\|^2_k + \delta^T S_k & \text{if } k < 0, \\
0 & \text{if } k = 0, \\
-\theta\|\delta\|^2_k + \delta^T S_k & \text{if } k > 0,
\end{cases}$$

with $S_k$ defined by

$$S_k = \sum_{i=1}^{k} \xi_i + \sum_{i=-k}^{-1} \xi_i, \quad -\infty < k < \infty.$$

Here $(\xi_{-i})$ denotes an independent copy of $(\xi_i)$ for all $i \geq 1$ and, as usual, an empty sum is set to equal zero.

Since $\delta$ does not vary with the number of observations, it appears naturally also in the limit variable, which is given as the argument of the maximum of a two-sided sequence of random variables with drift.

A corresponding result holds true for the case (5.2.3). It is stated next.

**Theorem 5.4.** Let $E[\|Y_1\|^4] < \infty$. If $\delta = \delta_n \neq 0$ is such that

$$\|\delta_n\|_2 \to 0, \quad \text{but} \quad \frac{n\|\delta_n\|^2_2}{\log \log n} \to \infty \quad (n \to \infty),$$

then it holds under $H_A$ that

$$\|\delta_n\|^2_2(\hat{k}_n^* - k^*) \xrightarrow{p} \min \{ t : V(t) = \sup_s V(s) \} \quad (n \to \infty),$$
where

\[
V(t) = \begin{cases} 
(1 - \theta)t + W(t) & \text{if } t \leq 0, \\
0 & \text{if } t = 0, \\
-\theta t + W(t) & \text{if } t > 0,
\end{cases}
\]

with \((W(t) : -\infty < t < \infty)\) denoting a two-sided standard Brownian motion.

Note that the limit processes \(P(k)\) and \(V(t)\) contain drift terms which attain their maximum at 0, and whose slope on the negative and positive half line is determined by the location \(\theta\) of the change-point. If \(\theta = 1/2\), then the drift parts are symmetric, while the change-point detection becomes significantly harder if \(\theta\) is close to 0 (or 1). In these cases, the slope of the drift for positive (or negative) arguments is close to zero. In the case of Theorem 5.3, the constant order of magnitude of \(\|\delta\|_2\) also plays a role, with larger changes naturally being more easily identifiable. Theorems 5.3 and 5.4 thus provide clear theoretical justification of the empirical properties discussed in Section 5.3.

It is possible to develop a feel for the size of the function \(\Delta = \Delta_n\) which implies the assumptions of Theorem 2.2. If \(\|\Delta_n\| \to 0\), then \(\|\hat{K} - K\| \to 0\), so by inequalities (4.38) and (4.44) of Bosq (2000), \(\|\hat{\varphi}_\ell - \hat{c}_\ell \varphi_\ell\| \to 0\) and \(\hat{\lambda}_\ell \to \lambda_\ell\) in probability. In view of lemma 5.2, we have that eigenvalues and eigenfunctions under \(H_0\) and \(H_A\) coincide in the limit. This means that \(\delta_{n,\ell} \approx c_{n,\ell} \lambda_\ell^{-1} \int_T \Delta_n(t)\varphi_\ell(t)dt\) and so \(\|\delta_n\|^2 \approx \sum_{\ell=1}^d \lambda_\ell^{-2} \left(\int_T \Delta_n(t)\varphi_\ell(t)dt\right)^2\). Thus, by the Cauchy-Schwartz inequality, \(\|\Delta_n\| \to 0\) implies \(\|\delta_n\| \to 0\). A sufficient condition for \(n\|\delta_n\|^2/(\log \log n) \to \infty\) cannot be stated as easily, but it is roughly \(n\|\Delta_n\|^2/(\log \log n) \to \infty\) because by Parseval’s inequality, for sufficiently large \(d\), \(\int T \Delta_n^2(t)dt \approx \sum_{\ell=1}^d \left(\int_T \Delta_n(t)\varphi_\ell(t)dt\right)^2\). These approximate calculations could be formalized, but our goal is to merely indicate that Theorem 2.2 holds if \(\|\Delta_n\|\) tends to zero at the rate slower than \(n^{-1/2}\).

Finally, we discuss the consistency of the estimator. Observe that we have assumed in (5.2.2) and (5.2.3) that \(\delta \neq 0\). This means that there exists \(1 \leq \ell \leq d\) such
that $\int_T \Delta(t) \psi_\ell(t) dt \neq 0$. If instead the change function $\Delta$ is orthogonal to $\psi_1, \ldots, \psi_d$, that is if
$$
\int_0^1 \Delta(t) \psi_\ell(t) dt = 0 \quad \text{for all} \quad \ell = 1, \ldots, d,
$$
then $\hat{k}_n^*$ cannot be a consistent estimator of $k^*$, since the principal components analysis has been performed in an eigenspace with a too small dimension to capture the change. On the other hand, see e.g. Chapter 8 of Ramsay and Silverman (2005), using large $d$ is not practical because it bears the difficulty of interpreting a multitude of principal components. Moreover, since for large $\ell$ the eigenvalues $\lambda_\ell$ are generally very small (the $\lambda_\ell$ are arranged in the decreasing order), such $\psi_\ell$ explain only a very small part of the variability of the data. Therefore the impact of a change occurring in a subspace spanned by the $\psi_\ell$ with large $\ell$ is small, and its detection less crucial.

5.3 Finite sample behavior

We carried out simulations to illustrate our theoretical results in finite samples. We simulated change-point processes under conditions of Theorems 5.3 and 5.4 for different sample sizes, and always used 1,000 replications. For each replication we estimated the location of a change-point $k^*$. We generated functional observations according to (5.1.1). Without loss of generality, $\mu$ was chosen to be equal to zero. Two different cases of $Y_i$ were considered, namely the trajectories of the standard Brownian motion (BM), and the Brownian bridge (BB). The number $d$ of the principal components was chosen to be equal to 2 and 3 in order to explain at least 75% of variability. The properties of the sampling distributions of the change-point estimator $\hat{k}_n^*$ are now briefly discussed.

To illustrate the simulation results based on Theorem 2.1 we introduced the quantity $\tau_n^* = k_n^*/n$ and the corresponding estimator $\hat{\tau}_n^* = \hat{k}_n^*/n$. We concentrated on $\hat{\tau}_n^* - \tau^*$ rather than on $\hat{k}_n^* - k^*$ to show the effect of the increase in sample size more clearly. Various functions $\Delta$ were analyzed: $\Delta = t, t^2, \sqrt{t}, \exp(t), \sin(t)$, and $\cos(t)$. To assess the accuracy of the estimator, bias, root mean square error (RMSE),
and mean absolute error (MAE) of $\hat{\tau}_n^*$ were computed. To conserve space, we do not display the whole set of tables we obtained, but rather display representative results in Table 5.1, and discuss general findings. From Table 5.1 we see that by increasing the sample size we attain a smaller bias, RMSE, and MAE. A similar pattern is observed for the increase in the number of principal components. In all cases we considered, the summary statistics indicate that estimation is more accurate if BB was used, even though the same number of principal components explains more variability for BM. This is easy to understand because the BB is a "smaller" process in the sense that $E[\|BB\|^2] = 1/6$ and $E[\|BM\|^2] = 1/4$, so the same change function $\Delta$ is more pronounced if the $Y_i$ are the BB. As expected from the discussion following Theorem 5.4, the closer the change point is to the middle of the sample, the better the estimator is. For $\tau^*$ equal to 0.25 and 0.75 an increased bias is observed.

Next we illustrate Theorem 5.4 which deals with nonconstant $\Delta$. We chose $\Delta = \Delta_n$ satisfying conditions of Theorem 2.2 and carried out the change-point estimation. Several different forms of $\Delta_n$ were considered, namely $\sin(t)\frac{n^\alpha}{\sqrt{n}}, t\frac{n^\alpha}{\sqrt{n}}, \sqrt{t}\frac{n^\alpha}{\sqrt{n}}, \cos(t)\frac{n^\alpha}{\sqrt{n}}, e^t\frac{n^\alpha}{\sqrt{n}}$, where $\alpha \in (0, 0.5)$. To illustrate Theorem 5.4, we concentrated on the distribution of $\|\delta_n\|^2_2(\hat{k}_n^* - k^*)$. We computed $\delta_\ell$ from $\sqrt{\gamma_\ell}\delta_\ell = \int T \Delta(t)\psi_\ell(t) dt$, where for $\ell = 1, \ldots, d$

$$\psi_\ell(t) = \sqrt{2} \sin \left(\frac{2\ell + 1}{2}\pi t\right), \quad t \in [0, 1], \quad \text{and} \quad \gamma_\ell = \frac{4}{\pi(2\ell + 1)^2}$$

are the eigenfunctions and eigenvalues of the BM and

$$\psi_\ell(t) = \sqrt{2} \sin (\ell\pi t), \quad t \in [0, 1], \quad \text{and} \quad \gamma_\ell = \frac{1}{[\ell\pi]^2}$$

are the corresponding eigenfunctions and eigenvalues of the BB.

As before, we chose $k_n^*$ to be the lower, middle and upper quartile of the sample size. The graphs of the estimated density of $\|\delta_n\|^2_2(\hat{k}_n^* - k^*)$ are shown in Figures 5.1 and 5.2. The densities are close to each other, as Theorem 5.4 implies that they
Table 5.1: Summary statistics for the change-point estimator. The change-point processes were generated by combining BB and $t + BB$ for three different locations of the change-point $\tau^*$. We used $d = 2$ and $d = 3$ (in parenthesis).

<table>
<thead>
<tr>
<th>$\tau^*$</th>
<th>Average($\hat{\tau}$)</th>
<th>Bias($\hat{\tau}$)</th>
<th>Median($\hat{\tau}$)</th>
<th>RMSE($\hat{\tau}$)</th>
<th>MAE($\hat{\tau}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 60$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.27 (0.26)</td>
<td>0.0152 (0.0107)</td>
<td>0.25 (0.25)</td>
<td>0.0336 (0.0252)</td>
<td>0.0158 (0.0108)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>0.0002 (-0.0003)</td>
<td>0.50 (0.50)</td>
<td>0.0108 (0.0058)</td>
<td>0.0038 (0.0018)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.73 (0.74)</td>
<td>-0.0152 (-0.0087)</td>
<td>0.75 (0.75)</td>
<td>0.0356 (0.0205)</td>
<td>0.0157 (0.0088)</td>
</tr>
<tr>
<td></td>
<td>$n = 100$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.26 (0.26)</td>
<td>0.0096 (0.0052)</td>
<td>0.25 (0.25)</td>
<td>0.0220 (0.0122)</td>
<td>0.0101 (0.0053)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>0.0002 (0.0000)</td>
<td>0.50 (0.50)</td>
<td>0.0063 (0.0039)</td>
<td>0.0024 (0.0011)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.74 (0.74)</td>
<td>-0.0096 (-0.0052)</td>
<td>0.75 (0.75)</td>
<td>0.0215 (0.0155)</td>
<td>0.0100 (0.0063)</td>
</tr>
<tr>
<td></td>
<td>$n = 140$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.26 (0.25)</td>
<td>0.0062 (0.0039)</td>
<td>0.25 (0.25)</td>
<td>0.0141 (0.0096)</td>
<td>0.0064 (0.0040)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>-0.0001 (-0.0001)</td>
<td>0.50 (0.50)</td>
<td>0.0043 (0.0027)</td>
<td>0.0017 (0.0007)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.74 (0.75)</td>
<td>-0.0071 (-0.0039)</td>
<td>0.75 (0.75)</td>
<td>0.0147 (0.0093)</td>
<td>0.0068 (0.0040)</td>
</tr>
<tr>
<td></td>
<td>$n = 200$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.25 (0.25)</td>
<td>0.0046 (0.0030)</td>
<td>0.25 (0.25)</td>
<td>0.0107 (0.0070)</td>
<td>0.0050 (0.0031)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>0.0001 (0.0000)</td>
<td>0.50 (0.50)</td>
<td>0.0033 (0.0016)</td>
<td>0.0013 (0.0005)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.75 (0.75)</td>
<td>-0.0050 (-0.0023)</td>
<td>0.75 (0.75)</td>
<td>0.0110 (0.0062)</td>
<td>0.0052 (0.0024)</td>
</tr>
<tr>
<td></td>
<td>$n = 300$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.25 (0.25)</td>
<td>0.0030 (0.0018)</td>
<td>0.25 (0.25)</td>
<td>0.0066 (0.0047)</td>
<td>0.0032 (0.0019)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>0.0000 (0.0001)</td>
<td>0.50 (0.50)</td>
<td>0.0021 (0.0012)</td>
<td>0.0008 (0.0004)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.75 (0.75)</td>
<td>-0.0032 (-0.0018)</td>
<td>0.75 (0.75)</td>
<td>0.0079 (0.0048)</td>
<td>0.0034 (0.0019)</td>
</tr>
<tr>
<td></td>
<td>$n = 600$</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.25</td>
<td>0.25 (0.25)</td>
<td>0.0015 (0.0007)</td>
<td>0.25 (0.25)</td>
<td>0.0036 (0.0019)</td>
<td>0.0016 (0.0008)</td>
</tr>
<tr>
<td>0.50</td>
<td>0.50 (0.50)</td>
<td>0.0000 (0.0000)</td>
<td>0.50 (0.50)</td>
<td>0.0010 (0.0006)</td>
<td>0.0004 (0.0002)</td>
</tr>
<tr>
<td>0.75</td>
<td>0.75 (0.75)</td>
<td>-0.0015 (-0.0009)</td>
<td>0.75 (0.75)</td>
<td>0.0037 (0.0022)</td>
<td>0.0016 (0.0009)</td>
</tr>
</tbody>
</table>
Fig. 5.1: Estimated density of $\|\delta_n\|_2^2(\hat{k}_n^* - k^*)$ for the process obtained combining $BM$ and $t^{0.05}_n \sqrt{n} + BM$. 
Fig. 5.2: Estimated density of $\|\delta_n\|_2^2(\hat{k}_n^* - k^*)$ for the process obtained combining $BB$ and $\sin(t)\frac{n^{0.45}}{\sqrt{n}} + BB$. 
must be close to the limit distribution. In most cases, a convergence with increasing $n$ is also clearly visible. For example, in the top and middle panels of Figure 5.2, the densities for $n = 600$ and $n = 900$ almost coincide. These properties hold for all choices of $\alpha \in (0, 0.5)$, Figures 5.1 and 5.2 show the extreme cases of $\alpha = 0.05$ and $\alpha = 0.45$.

5.4 Proofs

The proof section is divided into three parts. In the first subsection, we derive a decomposition that will be used to derive Theorems 5.3 and 5.4, whose proofs will be pursued in Subsections 5.4.2 and 5.4.3, respectively.

5.4.1 Preliminary calculations

Let $\hat{R}_n(k) = \hat{Q}_n(k) - \hat{Q}_n(k^*)$. Since $\hat{R}_n(k)$ and the original $\hat{Q}_n(k)$ differ only by the value $\hat{Q}_n(k^*)$ which is independent of $k$, it holds that they attain their maximum for the same value of $k$. Consequently, we have

$$\hat{k}_n^* = \min \left\{ k : \hat{R}_n(k) = \max_{1 \leq j \leq n} \hat{R}_n(j) \right\}.$$

Denote by $\hat{\zeta}_{i,\ell} = \sqrt{\lambda_{\ell}} \hat{\xi}_{i,\ell} = \int_T Y_i(t) \hat{\phi}_{\ell}(t) dt$ and $\hat{\beta}_\ell = \sqrt{\lambda_{\ell}} \hat{\delta}_\ell = \int_T \Delta(t) \hat{\phi}_{\ell}(t) dt$ the counterparts of $\zeta_{i,\ell}$ and $\beta_{\ell}$ which are obtained by replacing the true eigenvalues and eigenfunctions with the estimated versions. Note that the quantities $\hat{\zeta}_{i,\ell}$, $\hat{\xi}_{i,\ell}$, $\hat{\beta}_\ell$ and $\hat{\delta}_\ell$ are unobservable. The proofs to come will fall back on the following decomposition
of $\hat{R}_n(k)$. First, we have for $1 \leq k < k^*$ that

$$\hat{R}_n(k) = \frac{1}{n} \sum_{\ell=1}^{d} \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} - \frac{k}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - k \frac{n - k^*}{n} \hat{\delta}_\ell \right)^2$$

$$- \frac{1}{n} \sum_{\ell=1}^{d} \left( \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - k^* \frac{n - k^*}{n} \hat{\delta}_\ell \right)^2$$

$$= \frac{1}{n} \sum_{\ell=1}^{d} \left( - \sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k - k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (k - k^*) \frac{n - k^*}{n} \hat{\delta}_\ell \right)$$

$$\times \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (k + k^*) \frac{n - k^*}{n} \hat{\delta}_\ell \right)$$

(5.4.1)

$$= \frac{1}{n} \sum_{\ell=1}^{d} \left( \hat{E}^{(1)}_{k,\ell} + \hat{D}^{(1)}_{k,\ell} \right) \left( \hat{E}^{(2)}_{k,\ell} + \hat{D}^{(2)}_{k,\ell} \right),$$

where $\hat{E}^{(1)}_{k,\ell}$ and $\hat{E}^{(2)}_{k,\ell}$ ($\hat{D}^{(1)}_{k,\ell}$ and $\hat{D}^{(2)}_{k,\ell}$) denote the estimated random part (estimated deterministic part) in the first and second bracket of (5.4.1), respectively. We will refer to the expressions $\hat{D}^{(1)}_{k,\ell}$ and $\hat{D}^{(2)}_{k,\ell}$ as estimated deterministic in the following even though it is understood that they depend on $\hat{\varphi}_\ell$. A similar expression can be obtained
if \( k^* < k \leq n \). Here it holds,

\[
\hat{R}_n(k) = \frac{1}{n} \sum_{\ell=1}^{d} \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} - \frac{k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (n-k^*) \frac{k^*}{n} \hat{\delta}_{\ell} \right)^2

- \frac{1}{n} \sum_{\ell=1}^{d} \left( \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (n-k^*) \frac{k^*}{n} \hat{\delta}_{\ell} \right)^2

= \frac{1}{n} \sum_{\ell=1}^{d} \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} - \frac{k}{k^*+1} \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \frac{k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (2n-k-k^*) \frac{k^*}{n} \hat{\delta}_{\ell} \right)

\times \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k+k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} - (2n-k-k^*) \frac{k^*}{n} \hat{\delta}_{\ell} \right)

(5.4.2) = \frac{1}{n} \sum_{\ell=1}^{d} \left( \hat{E}_{k,\ell}^{(3)} + \hat{D}_{k,\ell}^{(3)} \right) \left( \hat{E}_{k,\ell}^{(4)} + \hat{D}_{k,\ell}^{(4)} \right),

\]

where \( \hat{E}_{k,\ell}^{(3)} \), \( \hat{E}_{k,\ell}^{(4)} \), \( \hat{D}_{k,\ell}^{(3)} \), and \( \hat{D}_{k,\ell}^{(4)} \) are the corresponding estimated random and drift parts. Using (5.4.1) and (5.4.2), we proceed with the proof of Theorem 5.3 in the next subsection. Since the arguments to be employed are symmetric for time lags before and after the change-point, detailed expositions will only be given for \( 1 \leq k < k^* \).

### 5.4.2 Proof of Theorem 5.3

The proof is divided into two parts. At first, we show that the estimator \( \hat{k}_n^* \) will be close to \( k^* \) by showing that \( \hat{R}_n(k) \) will attain its maximum not too far from the change-point. In the second step, we will derive the limit distribution.

**Lemma 5.5.** Under the assumptions of Theorem 5.3, it holds that

\[
\left| \hat{k}_n^* - k^* \right| = O_P(1) \quad (n \to \infty).
\]

**Proof.** To show the assertion of the lemma, we determine the behavior of those \( k \) satisfying \( 1 \leq k \leq k^* - N \) or \( k^* + N \leq k \leq n \) for some \( N \geq 1 \). Let \( 1 \leq \ell \leq d \). At
first, we derive the order of magnitude of the estimated deterministic term in (5.4.1)
that is, of \( \frac{1}{n} \hat{D}^{(1)}_{k,\ell} \hat{D}^{(2)}_{k,\ell} \). To this end, note that
\[
\max_{1 \leq k \leq k^* - N} \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 \rightarrow 2\theta(1 - \theta)^2 \quad (n \rightarrow \infty).
\]

In the next step, we shall replace \( \hat{\delta}_\ell \) by \( \delta_\ell \). To do so, observe that \( \hat{\eta}_\ell = \sqrt{\hat{\lambda}_\ell} \hat{\delta}_\ell \) and \( \beta_\ell = \sqrt{\gamma_\ell} \delta_\ell \) by definition. Moreover, part (ii) of Proposition 5.2 states that
\[
\| \varphi_\ell - \hat{c}_\ell \psi_\ell \| \rightarrow 0 \text{ in probability.}
\]
Therefore,
\[
\hat{\eta}_\ell = \int_\mathcal{T} \Delta(t) \hat{\varphi}_\ell(t) dt = \hat{c}_\ell \int_\mathcal{T} \Delta(t) \psi_\ell(t) dt + o_P(1) = \hat{c}_\ell \beta_\ell + o_P(1) \quad (n \rightarrow \infty),
\]
using that \( \Delta(t) \in \mathcal{L}^2(\mathcal{T}) \). Consequently, \( \hat{\eta}^2_\ell = \beta^2_\ell + o_P(1) \). Since the estimated eigenvalues \( \hat{\lambda}_\ell \) converge in probability to \( \gamma_\ell \) (see part (i) of Proposition 5.2), we arrive at
\[
\hat{\delta}^2_\ell = \delta^2_\ell + o_P(1) \quad (n \rightarrow \infty).
\]
Combining the above arguments yields
\[
\max_{1 \leq k \leq k^* - N} \frac{1}{n} \hat{D}^{(1)}_{k,\ell} \hat{D}^{(2)}_{k,\ell} = \max_{1 \leq k \leq k^* - N} (k - k^*) \hat{\sigma}^2_\ell \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 + o_P(1) = -2\theta^2(1 - \theta)^2 N + o_P(1).
\]

It is shown in the next sections that this deterministic part is the dominating term in (5.4.1). It follows thus that, for all \( K > 0 \),
\[
\lim_{N \rightarrow \infty} \lim_{n \rightarrow \infty} P \left( \max_{1 \leq k \leq k^* - N} \hat{R}_n(k) > -K \right) = 0.
\]
On the other hand, using (5.4.2), it can be proved in a similar fashion that
\[
\max_{k^* + N \leq k \leq n} \frac{1}{n} \hat{D}^{(3)}_{k,\ell} \hat{D}^{(4)}_{k,\ell} = -2\theta^2(1 - \theta) N + o_P(1),
\]
which implies that, for all \( K > 0 \),

\[
(5.4.5) \quad \lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{k^*+N \leq k \leq n} \hat{R}_n(k) > -K \right) = 0.
\]

Equations (5.4.4) and (5.4.5) now yield that

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \{ \hat{k}_n^* < k^* - N \} \cup \{ \hat{k}_n^* > k^* + N \} \right) = 0,
\]

which consequently finishes the proof of the lemma.

To derive the limit distribution, it suffices to investigate the asymptotic behavior of \( \hat{R}_n(k) \) for the range \( k^* - N \leq k \leq k^* + N \) of those time lags close to the change-point. The result is presented as a lemma.

**Lemma 5.6.** Under the assumptions of Theorem 5.3, it holds that, for any \( N \geq 1 \),

\[
\{ \hat{R}_n(k + k^*) : -N \leq k \leq N \} \overset{D}{\rightarrow} \{ 2\theta(1 - \theta)P(k) : -N \leq k \leq N \} \quad (n \to \infty).
\]

**Proof.** Let \( 1 \leq \ell \leq d \). Using (5.4.3), it is easy to see that, for any fixed \( N \geq 1 \) and as \( n \to \infty \),

\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)} - 2\theta(1 - \theta)^2 \delta^2_\ell (k - k^*) \right|
\]

\[
= \delta^2_\ell N \max_{k^*-N \leq k \leq k^*} \left| \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 - 2\theta(1 - \theta)^2 \right| + o_P(1)
\]

\[
= o_P(1).
\]

By a similar argument,

\[
\max_{k^* \leq k \leq k^* + N} \left| \frac{1}{n} \hat{D}_{k,\ell}^{(3)} \hat{D}_{k,\ell}^{(4)} + 2\theta^2(1 - \theta) \delta^2_\ell (k - k^*) \right| = o_P(1) \quad (n \to \infty).
\]
In the following, we are dealing with the estimated random parts. The functional central limit theorem implies that, for all \( x \in [0, 1] \),

\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} \zeta_i \xrightarrow{d} \Gamma(x) \quad (n \to \infty),
\]

where \( \xrightarrow{d} \) indicates weak convergence in the Skorohod space \( D^d[0, 1] \) and \( \Gamma(x) : x \in [0, 1] \) is an \( \mathbb{R}^d \)-valued, zero mean stochastic process with covariance matrix \( \Sigma \). Then,

\[
\sup_{x \in (0, 1)} \frac{1}{\sqrt{n}} \left| \sum_{i=1}^{\lfloor nx \rfloor} \zeta_{i,t} - \sum_{i=1}^{\lfloor nx \rfloor} \hat{\zeta}_{i,t} \right| = \sup_{x \in (0, 1)} \left| \int_{T} \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} Y_i(t) [\hat{c}_\ell \psi_\ell(t) - \hat{\phi}_\ell(t)] dt \right|
\]

\[
\leq \sup_{x \in (0, 1)} \left( \int_{T} \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nx \rfloor} Y_i(t) \right]^2 dt \right)^{1/2} \left( \int_{T} [\hat{c}_\ell \psi_\ell(t) - \hat{\phi}_\ell(t)]^2 dt \right)^{1/2}
\]

(5.4.6) \( = o_P(1) \)

by an application of Proposition 5.2. The same statement holds true also if \( \xi_{i,t} \) and \( \hat{\xi}_{i,t} \) are used in place of \( \zeta_{i,t} \) and \( \hat{\zeta}_{i,t} \).
Equations (5.4.3) and (5.4.6) imply now that
\[
\max_{k^*-N \leq k \leq k^*} \left| \left( \frac{k - k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \left( \frac{k + k^*}{n} - \frac{k^*}{n} \hat{\delta}_\ell \right) \right| \\
= \max_{k^*-N \leq k \leq k^*} \left| \left( \frac{k - k^*}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right) \left( \frac{k + k^*}{n} - \frac{k^*}{n} \delta_\ell \right) \right| + o_P(1) \\
= O(1) \frac{N}{n} \left| \sum_{i=1}^{n} Y_i(t) \psi_\ell(t) dt \right| + o_P(1) \\
= o_P(1).
\]

Hence,
\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \hat{E}^{(1)}_{k,\ell} \hat{E}^{(2)}_{k,\ell} + 2\theta(1 - \theta)\delta_\ell \sum_{i=k+1}^{k^*} \xi_{i,\ell} \right| = o_P(1)
\]
as \(n \to \infty\) for any \(N \geq 1\) which follows from (5.4.3) and (5.4.6) as well. Similar arguments apply also to \(\frac{1}{n} \hat{E}^{(3)}_{k,\ell} \hat{D}^{(4)}_{k,\ell}\) for which \(k^* \leq k \leq k^* + N\) holds. In view of the definition of the limit process \(P(k)\) in Theorem 5.3, it suffices to verify that the remaining terms in (5.4.1) and (5.4.2) do not contribute asymptotically. To this end, write
\[
\max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \hat{E}^{(1)}_{k,\ell} \hat{E}^{(2)}_{k,\ell} \right| \\
= \max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \left( \sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell} + \frac{k - k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right) \right| \\
\leq \max_{k^*-N \leq k \leq k^*} \left| \sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell} + \frac{k - k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right| \max_{k^*-N \leq k \leq k^*} \left| \frac{1}{n} \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right| \\
= o_P(1).
\]

Here, the first maximum is \(O_P(1)\), since the first sum \(\sum_{i=k+1}^{k^*} \hat{\xi}_{i,\ell}\) contains at most \(N\) terms, while the second sum is \(o_P(1)\) because of (5.4.6). Another application of
(5.4.6) gives that the second maximum is $o_P(1)$. Moreover,

$$
\max_{k^*-N \leq k \leq k^*} \frac{1}{n} \left| \hat{\delta}_{k,\ell}^{(1)} \hat{\delta}_{k,\ell}^{(2)} \right|
$$

$$
= \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \left( k - k^* \right) \frac{n - k^*}{n} \hat{\delta}_\ell \left( \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right)
$$

$$
\leq \max_{k^*-N \leq k \leq k^*} \left| \left( k - k^* \right) \frac{n - k^*}{n} \hat{\delta}_\ell \right| \max_{k^*-N \leq k \leq k^*} \frac{1}{n} \left| \sum_{i=1}^{k} \hat{\xi}_{i,\ell} + \sum_{i=1}^{k^*} \hat{\xi}_{i,\ell} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{\xi}_{i,\ell} \right|
$$

$$
= o_P(1).
$$

The same arguments apply also to the remaining terms in (5.4.2) and the proof of the lemma is therefore complete. ■

**Proof of Theorem 5.3.** The assertion follows immediately from Lemmas 5.5 and 5.6. ■

### 5.4.3 Proof of Theorem 5.4

We follow the proof steps developed in the previous subsection.

**Lemma 5.7.** Under the assumptions of Theorem 5.4, it holds that

$$
\|\hat{\delta}_n\|_2 k^*_n - k^* = O_P(1) \quad (n \to \infty).
$$

**Proof.** At first, we derive the order of magnitude of $\frac{1}{n} \hat{\delta}_{k,\ell}^{(1)} \hat{\delta}_{k,\ell}^{(2)}$ in (5.4.1). Let $N \geq 1$ and define $N_\delta = N\|\delta_n\|_2^{-2}$. Recognizing that $n^{-1}N_\delta \to 0$, since by assumption $n\|\delta_n\|_2^2 \to \infty$, it follows that

$$
\max_{1 \leq k \leq k^* - N_\delta} \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 \frac{2k^* - N_\delta}{n} \left( \frac{n - k^*}{n} \right)^2 \to 2\theta(1 - \theta)^2 \quad (n \to \infty).
$$
Consequently, (5.4.3) yields

\[
\max_{1 \leq k \leq k^* - N_\delta} \frac{1}{n} \sum_{\ell=1}^{d} \hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)} = \max_{1 \leq k \leq k^* - N_\delta} (k - k^*) \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 \sum_{\ell=1}^{d} \delta_{\ell}^2
\]

\[
= \max_{1 \leq k \leq k^* - N_\delta} (k - k^*) \frac{k + k^*}{n} \left( \frac{n - k^*}{n} \right)^2 \sum_{\ell=1}^{d} \delta_{\ell}^2 + o_P(1)
\]

\[
= -2\theta(1 - \theta)^2 N + o_P(1).
\]

It is shown in Section 5.4.5 that, under the assumptions of Theorem 5.4, this deterministic part is the dominating contributor in (5.4.1). It follows thus that, for all \( K > 0 \),

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N_\delta} \hat{R}_n(k) > -K \right) = 0
\]

Moreover, utilizing the decomposition in display (5.4.2), it can be proved similarly that

\[
\max_{k^* + N_\delta \leq k \leq n} \frac{1}{n} \sum_{\ell=1}^{d} \hat{D}_{k,\ell}^{(3)} \hat{D}_{k,\ell}^{(4)} = -2\theta^2(1 - \theta)N + o_P(1),
\]

which implies that, for all \( K > 0 \),

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{k^* + N_\delta \leq k \leq n} \hat{R}_n(k) > -K \right) = 0.
\]

Equations (5.4.7) and (5.4.8) now yield that

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \{ \hat{k}_n^* < k^* - N_\delta \} \cup \{ \hat{k}_n^* > k^* + N_\delta \} \right) = 0,
\]

which, noticing the definition of \( N_\delta \), completes the proof of the lemma. \( \blacksquare \)
Lemma 5.8. Under the assumptions of Theorem 5.4, it holds that, for any \( N \geq 1 \),

\[
\{ \hat{R}_n(k^* + \lfloor t\|\delta_n\|_2^{-2} \rfloor) : t \in [-N, N] \} \xrightarrow{d} \{ 2\theta(1-\theta)V(t) : t \in [-N, N] \} \quad (n \to \infty),
\]

where \( \xrightarrow{d} \) indicates weak convergence in the Skorohod space \( D[-N, N] \).

Proof. Denote by \( k \) the integer part of \( t\|\delta_n\|_2^{-2} \). Then, as \( n \to \infty \),

\[
\sup_{t \in [-N, 0]} \left| \frac{1}{n} \sum_{\ell=1}^{d} \hat{D}_{k^* + k, \ell}^{(1)} - 2\theta(1-\theta)t \right| = o_P(1) \quad \sup_{t \in [-N, 0]} \left( t - \|\delta_n\|_2 \|t\|_2^{-2} \right) = o_P(1).
\]

Similarly,

\[
\sup_{t \in [0, N]} \left| \frac{1}{n} \sum_{\ell=1}^{d} \hat{D}_{k^* + k, \ell}^{(3)} - 2\theta^2(1-\theta)t \right| = o_P(1) \quad (n \to \infty).
\]

Note next that, after an application of (5.4.6) and the law of the iterated logarithm [see, for example, Section 9 in Billingsley (1995)]

\[
\sup_{t \in [-N, 0]} \frac{|t|}{n\|\delta_n\|_2^2} \left| \sum_{\ell=1}^{d} \delta_{\ell}^\perp \sum_{i=1}^{n} \xi_{i,\ell} \right| = o_P(1) \quad \frac{1}{n\|\delta_n\|_2^2} \left| \sum_{\ell=1}^{d} \sum_{i=1}^{n} \xi_{i,\ell} \right| = o_P(1) \sqrt{\frac{\log \log n}{n\|\delta_n\|_2^2}} = o_P(1)
\]

by assumption on \( \delta_n \). It follows from the weak convergence of partial sum processes that there exist independent standard Brownian motions \( (W_{\ell}(t) : t \geq 0), 1 \leq \ell \leq d \) [see, for example, Billingsley (1968)], such that

\[
\|\delta_n\|_2^{-2} \sum_{i=k^*+k+1}^{k^*} \xi_{i,\ell} \xrightarrow{d} \|\delta_n\|_2^{-2} \sum_{i=1}^{k} \xi_{i,\ell} \xrightarrow{d[-N,0]} W_{\ell}(-t),
\]
where $k$ is the integer part of $t \| \delta_n \|^{-2}$. Checking the finite-dimensional distributions, it follows that the process
\[
\left( \frac{1}{\| \delta \|_2^2} \sum_{\ell=1}^{d} \delta_\ell W_\ell(t) : t \geq 0 \right)
\]
is a standard Brownian motion. Therefore, there exists a standard Brownian motion $(W^{(1)}(t) : t \geq 0)$ such that
\[
\sup_{t \in [-N,0]} \left| \frac{1}{n} \sum_{\ell=1}^{d} \hat{E}^{(1)}_{k+k^*,\ell} \hat{D}^{(2)}_{k+k^*,\ell} - 2\theta (1 - \theta) W^{(1)}(t) \right| = \Omega(1)
\]
\[
= \sup_{t \in [-N,0]} \left| \frac{2 k^* + k n - k^*}{n} \sum_{\ell=1}^{d} \delta_\ell \left( \sum_{i=k^*+k+1}^{k^*} \xi_{i,\ell} + \frac{t}{n \| \delta_n \|_2^2} \sum_{i=1}^{n} \xi_{i,\ell} \right) - 2\theta (1 - \theta) W^{(1)}(t) \right| + o_P(1)
\]
\[
= \Omega(1)
\]
\[
= o_P(1).
\]
A similar string of arguments yields that there is a standard Brownian motion $(W^{(2)}(t) : t \geq 0)$ such that
\[
\sup_{t \in [0,N]} \left| \frac{1}{n} \sum_{\ell=1}^{d} \hat{E}^{(3)}_{k+k^*,\ell} \hat{D}^{(4)}_{k+k^*,\ell} - 2\theta (1 - \theta) W^{(2)}(t) \right| = o_P(1).
\]
It remains to verify that the remaining parts in displays (5.4.1) and (5.4.2) do not contribute to the limit distribution. So, consider first

$$\max_{k^*-N \delta \leq k \leq k^*} \frac{1}{n} |\hat{D}^{(1)}_{k,\ell} \hat{E}^{(2)}_{k,\ell}|$$

$$= \max_{k^*-N \delta \leq k \leq k^*} \left| k^* - k \frac{n - k^*}{n} \delta_{\ell} \left( \sum_{i=1}^{k^*} \xi_{i,\ell} + \sum_{i=1}^{k^*} \xi_{i,\ell} - k + k^* \frac{n}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right) \right| + o_P(1)$$

$$= \mathcal{O}(1) \max_{k^*-N \delta \leq k \leq k^*} \frac{1}{n \|\delta_n\|_2} \left| \sum_{i=1}^{k^*} \xi_{i,\ell} + \sum_{i=1}^{k^*} \xi_{i,\ell} - k + k^* \frac{n}{n} \sum_{i=1}^{n} \xi_{i,\ell} \right| + o_P(1)$$

$$= o_P(1),$$

since, for example,

$$\max_{k^*-N \delta \leq k \leq k^*} \frac{1}{n \|\delta_n\|_2} \left| \sum_{i=1}^{k^*} \xi_{i,\ell} \right| = \mathcal{O}_{P}(1) \sqrt{\frac{\log \log n}{n \|\delta_n\|_2^2}} + o_P(1) = o_P(1)$$

by (5.4.6), the law of the iterated logarithm and assumption on $\delta_n$. All other terms can be handled in the same way. Next, note that by (5.4.6) and the law of the iterated logarithm, it holds that

$$\max_{k^*-N \delta \leq k \leq k^*} \sum_{i=k+1}^{k^*} \xi_{i,\ell} = \mathcal{O}_{P} \left( \sqrt{N \delta} \right) \quad \text{and} \quad \max_{k^*-N \delta \leq k \leq k^*} \frac{k^* - k}{n} \left| \sum_{i=1}^{n} \xi_{i,\ell} \right| = o_P(1),$$
respectively. Hence,

$$\max_{k^* - N \leq k \leq k^*} \frac{1}{n} |\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|$$

$$= \max_{k^* - N \leq k \leq k^*} \frac{1}{n} \left| \left( \sum_{i=k+1}^{k^*} \hat{x}_{i,\ell} + \frac{k - k^*}{n} \sum_{i=1}^{n} \hat{x}_{i,\ell} \right) \left( \sum_{i=1}^{k} \hat{x}_{i,\ell} + \sum_{i=1}^{k^*} \sum_{i=1}^{n} \hat{x}_{i,\ell} \right) \right|$$

$$\Theta_P(1) \max_{k^* - N \leq k \leq k^*} \frac{\sqrt{N}}{n} \left| \sum_{i=1}^{k} \hat{x}_{i,\ell} + \sum_{i=1}^{k^*} - \frac{k + k^*}{n} \sum_{i=1}^{n} \hat{x}_{i,\ell} \right|$$

$$= o_P(1),$$

since, by (5.4.6) and the law of the iterated logarithm,

$$\max_{k^* - N \leq k \leq k^*} \frac{\sqrt{N}}{n \|\delta_n\|_2} \left| \sum_{i=1}^{k} \hat{x}_{i,\ell} \right| = \Theta_P(1) \sqrt{\log \log n} \frac{\|\delta_n\|_2}{\sqrt{n}} + o_P(1) = o_P(1).$$

Similar for the other two terms and also for the terms coming from (5.4.2). The proof is complete. 

### 5.4.4 Verification of equation (5.4.4)

**Lemma 5.9.** Under the assumptions of Theorem 5.3 it holds that, for all $1 \leq \ell \leq d$ and $\varepsilon > 0$,

$$\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \geq \varepsilon \right) = 0.$$

**Proof.** Let $1 \leq \ell \leq d$ and $1 \leq k \leq k^* - N$ for some $N \geq 1$. From the definition in (5.4.1) and the argument leading to display (5.4.3) it follows that the absolute value of the estimated deterministic term $|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|$ has precise stochastic order $n(k^* - k)$. Hence,

$$\max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = o(1) \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k,\ell}^{(1)}|}{k^* - k} \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k,\ell}^{(2)}|}{n} = o_P(1) M_1(N, n) M_2(N, n).$$
We start by examining $M_1(N, n)$. The law of the iterated logarithm in combination with (5.4.6) imply that

$$\max_{1 \leq k \leq k^* - N} \frac{1}{k^* - k} \left| \sum_{i=k+1}^{k^*} \hat{\epsilon}_{i, \ell} \right| = \mathcal{O}_P \left( \max_{1 \leq k \leq k^* - N} \frac{1}{(k^* - k)^{1-\alpha}} \right) = \mathcal{O}_P \left( \frac{1}{N^{1-\alpha}} \right),$$

for any $1/2 < \alpha < 1$. Moreover, on account of (5.4.6),

$$\max_{1 \leq k \leq k^* - N} \frac{1}{n} \left| \sum_{i=1}^{n} \hat{\epsilon}_{i, \ell} \right| = \frac{1}{n} \sum_{i=1}^{n} \hat{\epsilon}_{i, \ell} = o_P(1) \quad (n \to \infty).$$

Three further applications of (5.4.6) to $M_2(N, n)$ yield that

$$\lim_{N \to \infty} \limsup_{n \to \infty} P \left( M_1(N, n) M_2(N, n) \geq \varepsilon \right) = 0$$

and the lemma is proved. ■

**Lemma 5.10.** Under the assumptions of Theorem 5.3 it holds that, for all $1 \leq \ell \leq d$ and $\varepsilon > 0$,

$$\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k, \ell}^{(1)} \hat{D}_{k, \ell}^{(2)}|}{|\hat{D}_{k, \ell}^{(1)} \hat{D}_{k, \ell}^{(2)}|} \geq \varepsilon \right) = 0.$$

**Proof.** Write

$$\max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k, \ell}^{(1)} \hat{D}_{k, \ell}^{(2)}|}{|\hat{D}_{k, \ell}^{(1)} \hat{D}_{k, \ell}^{(2)}|} = \mathcal{O}_P(1) \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k, \ell}^{(1)}|}{k^* - k} \max_{1 \leq k \leq k^* - N} \frac{|\hat{D}_{k, \ell}^{(2)}|}{n} = \mathcal{O}_P(1) M_1(N, n) M_3(N, n),$$

where $M_1(N, n)$ has already been dealt with in Lemma 5.9. Noticing that

$$M_3(N, n) = \max_{1 \leq k \leq k^* - N} \frac{k + k^* n - k^*}{n} \delta_{\ell} + o_P(1) = \mathcal{O}_P(1)$$

hence yields the assertion. ■
Lemma 5.11. Under the assumptions of Theorem 5.3 it holds that, for all $1 \leq \ell \leq d$ and $\varepsilon > 0$,

$$
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \geq \varepsilon \right) = 0.
$$

Proof. In an analogous fashion, we obtain

$$
\max_{1 \leq k \leq k^* - N} \frac{|\hat{D}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = O(1) \max_{1 \leq k \leq k^* - N} \frac{|\hat{D}_{k,\ell}^{(1)}|}{k^* - k} \max_{1 \leq k \leq k^* - N} \frac{|\hat{E}_{k,\ell}^{(2)}|}{n} = O_P(1) M_4(N, n) M_2(N, n)
$$

with $M_2(N, n)$ from Lemma 5.9. Therefore

$$
M_4(N, n) = \max_{1 \leq k \leq k^* - N} \frac{n - k^*}{n} \frac{k^* - k}{k^* - k} \delta_\ell + o_P(1) = O_P(1)
$$

gives the result. ■

Similar calculations can be be performed for the terms appearing in display (5.4.2). Details are omitted.

5.4.5 Verification of equation (5.4.7)

Lemma 5.12. Under the assumptions of Theorem 4.4 it holds that, for all $\varepsilon > 0$,

$$
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N} \frac{\sum_{\ell=1}^{d} |\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^{d} |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \geq \varepsilon \right) = 0.
$$

Proof. Observe that, uniformly in $k$,

$$
\sum_{\ell=1}^{d} |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}| \sim_P n(k^* - k) \|\delta_n\|_2^2.
$$
Therefore, for any \( 1 \leq \ell \leq d \),

\[
\max_{1 \leq k \leq k^* - N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = \mathcal{O}_P(1) \max_{1 \leq k \leq k^* - N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)}|}{(k^* - k)\|\delta_n\|_2} \max_{1 \leq k \leq k^* - N_\delta} \frac{|\hat{E}_{k,\ell}^{(2)}|}{n\|\delta_n\|_2} = \mathcal{O}_P(1) M_1^\delta(N, n) M_2^\delta(N, n).
\]

We first study the asymptotics of \( M_1^\delta(N, n) \). To this end note that

\[
\max_{1 \leq k \leq k^* - N_\delta} \frac{|\sum_{i=k+1}^{k^*} \xi_{i,\ell}|}{(k^* - k)\|\delta_n\|_2} = \max_{1 \leq k \leq k^* - N_\delta} \frac{|\sum_{i=1}^{k^* - k} \xi_{i,\ell}|}{(k^* - k)\|\delta_n\|_2} = \mathcal{O}_P(1) \max_{1 \leq k \leq k^* - N_\delta} \frac{1}{(k^* - k)\|\delta_n\|_2} = \mathcal{O}_P\left(\frac{1}{\sqrt{N}}\right).
\]

Furthermore, from the law of the iterated logarithm,

\[
\max_{1 \leq k \leq k^* - N_\delta} \frac{|\sum_{i=1}^n \xi_{i,\ell}|}{n\|\delta_n\|_2} = o_P(1) \quad (n \to \infty).
\]

Since the same arguments apply also to the term \( M_2^\delta(N, n) \), it follows from (5.4.6) that

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( M_1^\delta(N, n) M_2^\delta(N, n) \geq \varepsilon \right) = 0.
\]

This proves the assertion. \( \blacksquare \)

**Lemma 5.13.** Under the assumptions of Theorem 5.4 it holds that, for all \( \varepsilon > 0 \),

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^* - N_\delta} \frac{\sum_{\ell=1}^d |\hat{E}_{k,\ell}^{(1)} \hat{E}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \geq \varepsilon \right) = 0.
\]
Proof. Along the lines of the previous proof, we may write

\[
\max_{1 \leq k \leq k^*-N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = \mathcal{O}_P(1) \max_{1 \leq k \leq k^*-N_\delta} \frac{|\hat{E}_{k,\ell}^{(1)}|}{(k^* - k) \|\delta_n\|_2} \max_{1 \leq k \leq k^*-N_\delta} \frac{|\hat{D}_{k,\ell}^{(2)}|}{n \|\delta_n\|_2} = \mathcal{O}_P(1) M_3^\delta(N, n) M_\delta^4(N, n),
\]

where

\[
M_3^\delta(N, n) = \max_{1 \leq k \leq k^*-N_\delta} \frac{k + k^* n - k^* \delta_\ell}{n \|\delta_n\|_2} + o_P(1) = \mathcal{O}_P(1).
\]

Since \(M_1^\delta(N, n)\) has already been estimated in Lemma 5.12, the proof is complete.

Lemma 5.14. Under the assumptions of Theorem 5.4 it holds that, for all \(\varepsilon > 0\),

\[
\lim_{N \to \infty} \limsup_{n \to \infty} P \left( \max_{1 \leq k \leq k^*-N_\delta} \frac{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} \geq \varepsilon \right) = 0.
\]

Proof. Write

\[
\max_{1 \leq k \leq k^*-N_\delta} \frac{|\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|}{\sum_{\ell=1}^d |\hat{D}_{k,\ell}^{(1)} \hat{D}_{k,\ell}^{(2)}|} = \mathcal{O}_P(1) M_4^\delta(N, n) M_\delta^6(N, n)
\]

with

\[
M_4^\delta(N, n) = \max_{1 \leq k \leq k^*-N_\delta} \frac{k - k^* n - k^* \delta_\ell}{n \|\delta_n\|_2} + o_P(1) = \mathcal{O}_P(1)
\]

and the lemma is proved.

Again, the same arguments give the corresponding results for the terms in (5.4.2).
CHAPTER 6
SUMMARY AND CONCLUSIONS

This thesis introduced some novel methods in functional data analysis, namely goodness-of-fit and change point tests. These tests were motivated by data derived from nearly continuous times series records.

We developed a test for independence and identical distribution of functional observations. To reduce dimension, curves were projected on the most important functional principal components. Then a test statistic based on lagged cross–covariances of the resulting vectors was constructed. We showed that this dimension reduction step introduces asymptotically negligible terms, i.e. the projections behave asymptotically as iid vector–valued observations. A complete asymptotic theory based on correlations of random matrices, functional principal component expansions and Hilbert space techniques was developed. The test statistic has $\chi^2$ asymptotic null distribution. It can be readily computed using the \texttt{R} package \texttt{fda}. The test has good empirical size and power which, in our simulations and examples, is not affected by the choice of the functional basis. Its application is illustrated on two data sets: credit card sales activity and geomagnetic records.

We proposed two inferential tests for error correlation in the functional linear model, which complement the available graphical goodness of fit checks. To construct them, finite dimensional residuals were computed in two different ways, and then their autocorrelations were suitably defined. From these autocorrelation matrices, two quadratic forms were constructed whose limiting distribution are chi–squared with known numbers of degrees of freedom (different for the two forms). The asymptotic approximations are suitable for moderate sample sizes. The test statistics can be relatively easily computed using the \texttt{R} package \texttt{fda}, or similar MATLAB software. Application of the tests was illustrated by weather, magnetometer and financial data.
The asymptotic theory emphasizes the differences between the standard vector linear regression and the functional linear regression. To understand the behavior of the residuals obtained from the functional linear model, the interplay of three types of approximation errors must be considered, whose sources are: projection on a finite dimensional subspace, estimation of the optimal subspace and estimation of the regression kernel.

We developed a test, based on functional principal component scores, for detecting a change point in the mean of functional observations. The test can be readily computed in the R package fda. The null distribution of the test statistic is asymptotically pivotal with a well-known asymptotic distribution. The asymptotic test has excellent finite sample performance. Its application is illustrated by temperature data from England.

We also developed a comprehensive asymptotic theory for the estimation of a change–point in the mean function of functional observations. We considered both: the case of a constant change size, and the case of a change whose size approaches zero, as the sample size tends to infinity. We showed how the limit distribution of a suitably defined change–point estimator depends on the size and location of the change. The theoretical insights were confirmed by a simulation study which illustrated the behavior of the estimator in finite samples.

Functional data analysis still has many interesting open questions and the work in this dissertation shows many appealing techniques and properties that have great potential in applied statistics.
REFERENCES


June 2nd 2010

Robertas Gabrys
Department of Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, UT 84322-3900
E-mail: gabrys.robertas@gmail.com
Phone: (435) 512-9192
Fax: (435) 797-1822

Dear Mr. Sampson:
I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of the American Statistical Association articles, Portmanteau Test of Independence for Functional Observations (Vol. 102, No. 480, pages 1338–1348, 2007), and Tests for Error Correlation in the Functional Linear Model (forthcoming), of which I am one of the authors, report an essential part of my dissertation research. I would like permission to reprint them as two chapters in my dissertation. (Reprinting them as chapters may require making some revision.) Please note that USU sends dissertations to Bell & Howell Dissertation Services to be made available for reproduction.
I will include an acknowledgment to the articles on the first page of the chapters, as shown below. Copyright and permission information will be included in a special
appendix. If you would like a different acknowledgment, please so indicate. Please indicate your approval of this request by signing in the space provided below, and attach any other form necessary to confirm permission. If you charge a reprint fee for use of an article by the author, please indicate that as well. If you have any questions, please call me at the number above or send me an e-mail message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested articles in his dissertation, with the following acknowledgment:


Signed

Date
Hi Robertas:

Please allow me to grant this permission. Because you are an author, you have the right to reprint these articles in your dissertation. All fees are waived.

If you require anything further, please do not hesitate to let me know.

All the best and good luck!

Eric Sampson
Journals Manager ASA
June 2nd 2010

Robertas Gabrys
Department of Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, UT 84322-3900
E-mail: gabrys.robertas@gmail.com
Phone: (435) 512-9192
Fax: (435) 797-1822

Journal of the Royal Statistical Society Series B
Phone: (44) 020-7614-3914
Fax: (44) 020-7614-3905
E-mail: m.owen@rss.org.uk

Dear Dr. Owen:

I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of the Royal Statistical Society Series B article, Detecting Changes in the Mean of Functional Observations (Vol. 71, Issue 5, pages 927–946, 2009), of which I am one of the authors, reports an essential part of my dissertation research. I would like permission to reprint it as a chapter in my dissertation. (Reprinting it as a chapter may require making some revision.) Please note that USU sends dissertations to Bell & Howell Dissertation Services to be made available for reproduction. I will include an acknowledgment to the article on the first page of the chapters, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgment, please so indicate.

Please indicate your approval of this request by signing in the space provided below,
and attach any other form necessary to confirm permission. If you charge a reprint fee for use of an article by the author, please indicate that as well.

If you have any questions, please call me at the number above or send me an e-mail message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested article in his dissertation, with the following acknowledgment:


Signed

Date
from Owen, Martin <M.Owen@rss.org.uk>

to Robertas Gabrys <gabrys.robertas@gmail.com>

date       Fri, Jun 11, 2010 at 1:55 AM subject

RE: Request for permission to reprint my article published in JRSS
(B)

Dear Robertas

As I said, the form indicates that you do not need to apply for
permission for use in this way, and the Society has no objection.

Yours sincerely

Martin Owen

Executive Editor

Royal Statistical Society

12 Errol Street

London, EC1Y 8LX, UK
Direct: (44) 020 7614 3914

Switchboard: (44) 020 7638 8998

Fax: (44) 020 7614 3905

E-mail: m.owen@rss.org.uk
Internet: http://www.rss.org.uk
June 2nd 2010

Robertas Gabrys
Department of Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, UT 84322-3900
E-mail: gabrys.robertas@gmail.com
Phone: (435) 512-9192
Fax: (435) 797-1822

Journal of Multivariate Analysis
Phone: (+44) 1865-843-830
Fax: (+44) 1865-853-333
E-mail: permissions@elsevier.com

To whom it may concern:
I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of Multivariate Analysis article, Estimation of a Change–Point in the Mean Function of Functional Data (Vol. 100, No. 10, pages 2254–2269, 2009), of which I am one of the authors, reports an essential part of my dissertation research. I would like permission to reprint it as a chapter in my dissertation. (Reprinting it as a chapter may require making some revision.) Please note that USU sends dissertations to Bell & Howell Dissertation Services to be made available for reproduction. I will include an acknowledgment to the article on the first page of the chapters, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgment, please so indicate.

Please indicate your approval of this request by signing in the space provided below,
and attach any other form necessary to confirm permission. If you charge a reprint fee for use of an article by the author, please indicate that as well. If you have any questions, please call me at the number above or send me an e-mail message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested article in his dissertation, with the following acknowledgment:


Signed

Date
Dear Dr Gabrys,

Thank you for your e-mail.

As a journal author, you retain article rights for large number of author uses, including use by your employing institute or company. These rights are retained and permitted without the need to obtain specific permission from Elsevier. This includes:

- the right to include the journal article, in full or in part, in a thesis or dissertation;

Or contact our Global Rights Department:

Elsevier Global Rights Department phone (+44) 1865 843 830 fax (+44) 1865 853 333 email: permissions@elsevier.com

Yours sincerely,

Nobilie F. Niz

Elsevier Customer Support
June 2nd 2010

Robertas Gabrys
Department of Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, UT 84322-3900
E-mail: gabrys.robertas@gmail.com
Phone: (435) 512-9192
Fax: (435) 797-1822

Dear Prof. Lajos Horváth:
I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of the American Statistical Association article, Tests for Error Correlation in the Functional Linear Model (forthcoming), The Journal of the Royal Statistical Society Series B article, Detecting Changes in the Mean of Functional Observations (Vol. 71, Issue 5, pages 927–946, 2009), and The Journal of Multivariate Analysis article, Estimation of a Change–Point in the Mean Function of Functional Data (Vol. 100, No. 10, pages 2254–2269, 2009), of which I am one of the authors,
report an essential part of my dissertation research. I would like permission to reprint
them as three chapters in my dissertation. (Reprinting them as chapters may require
making some revision.) Please note that USU sends dissertations to Bell & Howell
Dissertation Services to be made available for reproduction.

I will include an acknowledgment to the articles on the first page of the chapters,
as shown below. Copyright and permission information will be included in a special
appendix. If you would like a different acknowledgment, please so indicate.

Please indicate your approval of this request by signing in the space provided below,
and attach any other form necessary to confirm permission.

If you have any questions, please call me at the number above or send me an e-mail
message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested articles in his
dissertation, with the following acknowledgment:

Gabrys, R., Horváth, L., Kokoszka, P. (2010) Tests for error correlation in the func-

Berkes, I., Gabrys, R. Horváth, L., Kokoszka, P. (2009). Detecting changes in the
mean of functional observations, *Journal of the Royal Statistical Society, Series B*,
71, 927–946.

point in the mean function of functional data, *Journal of Multivariate Analysis*, 100,
2254–2269.
Signed

Date
from horvath@math.utah.edu

to Robertas Gabrys <gabrys.robertas@gmail.com>

date Mon, Jun 14, 2010 at 8:01 PM subject

Re: [Fwd: Permission to reprint the articles we coauthored in my dissertation] mailed-by math.utah.edu

Please accept this as confirmation that I gladly agree to you including our joint journal articles, in full or in part, in your thesis or dissertation.

Lajos Horvath
June 2\textsuperscript{nd} 2010

Robertas Gabrys  
Department of Mathematics and Statistics  
Utah State University  
3900 Old Main Hill  
Logan, UT 84322-3900  
E-mail: gabrys.robertas@gmail.com  
Phone: (435) 512-9192  
Fax: (435) 797-1822

Prof. Alexander Aue  
Department of Statistics  
University of California  
One Shields Avenue  
Davis, CA 95616  
E-mail: alexaue@wald.ucdavis.edu  
Phone: (530) 554-1555  
Fax: (530) 752-7099

Dear Prof. Aue:

I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of Multivariate Analysis article, Estimation of a Change–Point in the Mean Function of Functional Data (Vol. 100, No. 10, pages 2254–2269, 2009), of which I am one of the authors, report an essential part of my dissertation research. I would like permission to reprint them as three chapters in my dissertation. (Reprinting it as a chapter may require making some revision.) Please note that USU sends
dissertations to Bell & Howell Dissertation Services to be made available for reproduction.

I will include an acknowledgment to the article on the first page of the chapters, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgment, please so indicate.

Please indicate your approval of this request by signing in the space provided below. If you have any questions, please call me at the number above or send me an e-mail message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested article in his dissertation, with the following acknowledgment:


Signed

Date
from Alexander Aue <alexaue@wald.ucdavis.edu>

to Robertas Gabrys <gabrys.robertas@gmail.com>

date Thu, Jun 10, 2010 at 4:40 PM subject

Re: Permission to reprint the article we coauthored in my dissertation mailed-by wald.ucdavis.edu

Dear Robertas,

Please accept this as confirmation that I gladly agree to you including our joint journal article, in full or in part, in your thesis or dissertation.

Best,
Alex
June 2nd 2010
Robertas Gabrys
Department of Mathematics and Statistics
Utah State University
3900 Old Main Hill
Logan, UT 84322-3900
E-mail: gabrys.robertas@gmail.com
Phone: (435) 512-9192
Fax: (435) 797-1822

Prof. István Berkes
Institute of Statistics
Graz University of Technology
Mnzgrabenstrae 11/III
A-8010 Graz, Austria
E-mail: berkes@tugraz.at
Phone: (+43) 316-873-6976
Fax: (+43) 316-873-6977

Dear Prof. Berkes:
I am preparing my dissertation in the Department of Mathematics and Statistics at Utah State University and I plan to complete my degree in the Summer of 2010. The Journal of the Royal Statistical Society Series B article, Detecting Changes in the Mean of Functional Observations (Vol. 71, Issue 5, pages 927–946, 2009), of which I am one of the authors, report an essential part of my dissertation research. I would like permission to reprint them as three chapters in my dissertation. (Reprinting it as a chapter may require making some revision.) Please note that USU sends dissertations to Bell & Howell Dissertation Services to be made available for reproduction.
I will include an acknowledgment to the article on the first page of the chapters, as shown below. Copyright and permission information will be included in a special appendix. If you would like a different acknowledgment, please so indicate.

Please indicate your approval of this request by signing in the space provided below. If you have any questions, please call me at the number above or send me an e-mail message at the above address.

Thank you for your assistance.

Robertas Gabrys

I hereby give permission to Robertas Gabrys to reprint the requested article in his dissertation, with the following acknowledgment:


Signed

Date
Dear Robertas,

Thank you for your e-mail.

I herewith give you permission to reprint the article: "Detecting changes in the mean of functional observations", J. Royal Stat. Soc. B 71 (2009), 927--946, coauthored by you, Lajos Horvath and myself, in your PhD dissertation or use it in any way you find appropriate. Do not hesitate to contact me if you have any questions.

Best regards,

Istvan Berkes
EDUCATION

**Ph.D.** Statistics (GPA 4.0)  \(2006 - 2010\)
Utah State University, Logan, UT, USA

*Dissertation title:* Goodness-of-fit and change-point tests for functional data
*Advisor:* Piotr S. Kokoszka

**M.S.** Applied Mathematics (Finance and Insurance emphasis)  \(2003 - 2005\)
Vilnius University, Vilnius, Lithuania

*Thesis title:* Market index investigation using functional data analysis tools
*Advisor:* Alfredas Račkauskas

**B.A.** (CUM LAUDE) Applied Mathematics (Finance and Insurance emphasis)  \(1999 - 2003\)
Vilnius University, Vilnius, Lithuania

*Thesis title:* Value at risk
*Advisor:* Alfredas Račkauskas

RESEARCH ACTIVITY

**Graduate Research Assistant**  \(2006 - 2010\)

*Accepted/Published Articles*

forthcoming.


RESEARCH INTERESTS

Functional data analysis and applications

Statistical computing

Wavelet methods in time series analysis

Applications in environmental sciences, financial econometrics, actuarial mathematics, space physics

TEACHING EXPERIENCE

Utah State University, Logan UT

Instructor 2006 – 2010

Courses Taught
Course: Business Statistics
Text: Bowerman, O’Connell, Orris, and Murphree (2010)

Essentials of Business Statistics
Semester: Summer 2010

Course: Statistical Methods

Introduction to the Practice of Statistics
Semesters: Spring 2008, Fall 2008, Fall 2009, Spring 2009

Course: Calculus II
Text: Stewart (2005)

Calculus
Semester: Summer 2007

Course: Introduction to statistics

Statistics
Semesters: Fall 2006, Spring 2007

INDUSTRIAL WORK EXPERIENCE

UAB SEB VB Life Insurance, Vilnius Lithuania

Actuary 2002 – 2005