Hilbert Space Valued Signal Plus Noise Models: Analysis of Structural Breaks under High Dimensionality and Temporal Dependence

Inaugural-Dissertation

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

vorgelegt von
Leonid Torgovitski
aus Sankt Petersburg, Russland

Köln 2016

Berichterstatter: Prof. i.R. Dr. Josef G. Steinebach, Universität zu Köln ....
Prof. i.R. Dr. Wolfgang Wefelmeyer, Universität zu Köln ....
Prof. Dr. Piotr S. Kokoszka, Colorado State University, Fort Collins, USA

Tag der mündlichen Prüfung: 31. Oktober 2016

> “Now, here, you see, it takes all the running you can do, to keep in the same place. If you want to get somewhere else, you must run at least twice as fast as that!
>
> *Through the Looking-Glass*
> — Lewis Carroll

ABSTRACT

This thesis is concerned with change point analysis for time series, i.e. with detection of structural breaks in time-ordered, random data. This long-standing research field regained popularity over the last few years and is still undergoing, as statistical analysis in general, a transformation to high-dimensional problems. We focus on the fundamental »change in the mean« problem and provide extensions of the classical non-parametric Darling-Erdős-type cumulative sum (CUSUM) testing and estimation theory within high-dimensional Hilbert space settings.

In the first part we contribute to (long run) principal component based testing methods for Hilbert space valued time series under a rather broad (abrupt, epidemic, gradual, multiple) change setting and under dependence. For the dependence structure we consider either traditional m-dependence assumptions or more recently developed m-approximability conditions which cover, e.g., MA, AR and ARCH models. We derive Gumbel and Brownian bridge type approximations of the distribution of the test statistic under the null hypothesis of no change and consistency conditions under the alternative. A new formulation of the test statistic using projections on subspaces allows us to simplify the standard proof techniques and to weaken common assumptions on the covariance structure. Furthermore, we propose to adjust the principal components by an implicit estimation of a (possible) change direction. This approach adds flexibility to projection based methods, weakens typical technical conditions and provides better consistency properties under the alternative.

In the second part we contribute to estimation methods for common changes in the means of panels of Hilbert space valued time series. We analyze weighted CUSUM estimates within a recently proposed »high-dimensional low sample size (HDLSS)« framework, where the sample size is fixed but the number of panels increases. We derive sharp conditions on »pointwise asymptotic accuracy« or »uniform asymptotic accuracy« of those estimates in terms of the weighting function. Particularly, we prove that a covariance-based correction of Darling-Erdős-type CUSUM estimates is required to guarantee uniform asymptotic accuracy under moderate dependence conditions within panels and that these conditions are fulfilled, e.g., by any MA(1) time series. As a counterexample we show that for AR(1) time series, close to the non-stationary case, the dependence is too strong and uniform asymptotic accuracy cannot be ensured.

Finally, we conduct simulations to demonstrate that our results are practically applicable and that our methodological suggestions are advantageous.

ZUSAMMENFASSUNG

Diese Arbeit beschäftigt sich mit der Changepoint-Analyse von Zeitreihen, d.h. mit der Aufdeckung von Strukturbrüchen in zeitlich angeordneten Zufallsdaten. Dies ist ein seit langem bestehender Forschungsbereich, welcher in den letzten Jahren an Popularität wiedererlangt hat, und, wie die statistische Analyse im Allgemeinen, eine Transformation hin zu hochdimensionalen Problemstellungen vollzieht. Wir betrachten das grundlegende Problem einer »Änderung des Erwartungswertes« und erweitern die klassische Theorie der nichtparameterischen Cumulative Sum (CUSUM) Test- und Schätzverfahren vom Darling-Erdős-Typ auf hochdimensionale Hilbertraum-Modelle.

Im ersten Teil liefern wir Beiträge zu Testverfahren für Hilbertraum-wertige Zeitreihen basierend auf (long run) Hauptkomponenten unter einem breiten Spektrum an Änderungsmodellen (abrupt, epidemisch, graduell, mehrfach) sowie unter Abhängigkeiten. Als Abhängigkeitsstruktur betrachten wir die traditionellen Annahmen der m-Abhängigkeit sowie die neueren Bedingungen der m-Approximierbarkeit, welche insgesamt z.B. die MA-, AR- und ARCH-Modelle abdecken. Unter der Nullhypothese leiten wir Approximationen der Testverteilung sowohl vom Gumbel-Typ als auch mittels Brownscher Brücken her und unter der Alternativhypothese weisen wir Konsistenzbedingungen nach. Eine neue Formulierung der Teststatistik unter Verwendung von Projektionen auf Teilräume ermöglicht uns, die gängige Beweistechnik zu vereinfachen und die üblichen Bedingungen an die Kovarianzstruktur abzuschwächen. Des Weiteren schlagen wir vor, die Hauptkomponenten durch implizite Schätzung der (eventuellen) Änderungsrichtung der Erwartungswerte zu korrigieren. Dieser Ansatz bringt mehr Flexibilität für projektionsbasierte Methoden, ermöglicht es die üblichen technischen Bedingungen abzuschwächen und liefert bessere Konsistenzeigenschaften unter der Alternative.

Im zweiten Teil liefern wir Beiträge zu Schätzverfahren für gleichzeitige Änderungen in Erwartungswerten von Paneldaten Hilbertraum-wertiger Zeitreihen. Wir analysieren gewichtete CUSUM Schätzer im einem erst kürzlich vorgeschlagenem »high-dimensional low sample size (HDLSS)« Setting. In diesem bleibt der Stichprobenumfang fix und lediglich die Anzahl der Panels wächst an. Wir leiten scharfe Bedingungen an die »punktweise asymptotische Exaktheit« sowie an die »gleichmäßige asymptotische Exaktheit« der Schätzer in Abhängigkeit von der Gewichtsfunktion her. Insbesondere weisen wir nach, dass eine Kovarianz-basierte Korrektur der Darling-Erdős-Typ CUSUM Schätzer benötigt wird, um gleichmäßige asymptotische Exaktheit unter moderaten Bedingungen an Abhängigkeiten innerhalb einzelner Panels sicherzustellen, und zeigen, dass diese Bedingungen von beliebigen MA(1) Zeitreihen erfüllt werden. Als Gegenbeispiel zeigen wir, dass AR(1) Zeitreihen nahe der Nichtstationarität schon eine zu hohe Abhängigkeit aufweisen und somit eine gleichmäßige asymptotische Exaktheit der Schätzung nicht vorliegen kann.

In Simulationen zeigen wir, dass unsere Resultate sowohl praktisch anwendbar sind als auch, dass unsere methodischen Vorschläge vorteilhaft sind.

ACKNOWLEDGEMENT

First of all, I would like to sincerely thank Prof. J. G. Steinebach for giving me the opportunity to change from the field of numerical analysis to statistics and to work on change point analysis for high-dimensional data. At the very beginning he gave me a draft version of the book of Horváth & Kokoszka (2012) which determined the functional data direction of my research. I am thankful for his numerous reviews of my articles, for his precise comments and for a lot of good advice that always led to the desired results and especially helped to create this thesis. Moreover, I am grateful that he supported all my (academic) traveling plans which allowed me to benefit from his wide network of colleagues.

I would also like to thank Prof. H. Schmidli, Prof. W. Wefelmeyer and Prof. A. Drewitz together with all colleagues at the working-group of Prof. J. G. Steinebach for the enjoyable years and for the pleasant atmosphere. Particularly, I want to thank Christoph Heuser for the motivating and encouraging time in our office.

Several parts of this work were conducted during my research stays at the University of California in Davis, the Colorado State University in Fort Collins, the University of Utah in Salt Lake City and at the University of Cambridge. I am grateful to Prof. A. Aue, Prof. D. Paul, Prof. P. Kokoszka, Prof. L. Horváth and Prof. J. Aston for their kind hospitality as well as to the Friedrich-Ebert-Foundation for the generous financial support. All these visits became important puzzle pieces of this thesis.

To be slightly more specific I am indebted to Prof. A. Aue and Prof. D. Paul, on the one hand, for their attention, which made my stays at the UC Davis a remarkable experience, and, on the other hand, for giving me the opportunity to dive into the fascinating theory of random matrices. I am grateful to Prof. J. Aston for the fruitful discussions on projection based functional data methods and on modern fMRI applications. I am thankful to Prof. L. Horváth, Prof. P. Kokoszka and Prof. G. Rice for generously sharing preprints of their articles which was especially important at the beginning of my research work. During my visit of Prof. P. Kokoszka I experienced much hospitality from the students Daniel Hernandez-Stumpfhauser, Ben Goldman, Charlie Vollmer and from Łukasz Kidziński. Discussions with Łukasz were one motivation to (re)state the results of this thesis in an abstract Hilbert space setting which led to substantial improvements of the final version.

I also want to thank my friends Sascha Baumanns, Michael Matthes, Lennart Jansen, Frank Kirchhoff and Carsten Kobe. Their support and motivation was important to me during these years.

A special separate thank goes to Prof. J. G. Steinebach, Prof. W. Wefelmeyer and to Prof. P. Kokoszka for refereeing this dissertation.

Great big thanks go to my parents and my grandparents for all the encouragement, trust and patience. Finally, my greatest thanks go to Hella and Maxim for the very best possible support during this challenge!

CONTENTS

Abstract **vii**

Acknowledgement **xi**

Notation **xvii**

1 General introduction **1**
1.1 A standard change point testing and estimation problem 3
1.2 Extensions of standard methodology 5
1.3 Preliminaries on Hilbert space data and on dependence concepts 13
1.4 Structure of the thesis 19

2 Change point testing for Hilbert space valued data **21**
2.1 Introduction 21
2.2 A change point problem 25
2.2.1 Signal plus noise model 25
2.2.2 The testing problem 26
2.3 Testing for change points with weighted CUSUM procedures 29
2.3.1 Preliminaries on dimension reduction 29
Static and long run principal components 31
Fully empirical CUSUM based on long run principal components 35
Perturbed long run covariance operator 36
2.3.2 Limit theorems for long run principal components 37
Estimation under the null hypothesis 38
Estimation under the alternative hypothesis 43
2.3.3 Limit theorems for weighted CUSUM tests 45
Asymptotics under the null hypothesis 49
Asymptotics under the alternative hypothesis 52
Testing via change-aligned principal components 54
2.4 A small simulation study 57
2.5 Proofs 65
2.5.1 Proofs for Subsection 2.3.2 65
2.5.2 Proofs for Subsection 2.3.3 74
2.5.3 Auxiliary results for the proofs 86
2.6 Notes 91

3 Change point estimation for Hilbert space valued data **93**
3.1 Introduction . . . 93
3.2 A change point problem . . . 97
3.2.1 Signal plus noise model . . . 97
3.2.2 The estimation problem . . . 99
3.3 Estimating change points with weighted CUSUM procedures . . . 101
3.3.1 Preliminaries on the variance of cumulated noises . . . 102
3.3.2 Limit theorems for weighted CUSUM estimates . . . 106
Pointwise accurate estimation . . . 106
Uniformly accurate estimation . . . 110
3.3.3 Relation between weighted CUSUM and LASSO estimates . . . 114
3.4 A small simulation study . . . 119
3.5 Proofs . . . 123
3.6 Notes . . . 135

Bibliography **137**

NOTATION

For convenience of the reader, we provide a list of some frequently used notations, symbols and abbreviations in this thesis.

General notation

$\mathbb{N}, \mathbb{N}_0, \mathbb{Z}$	$\mathbb{N} = \{1,2,3,\ldots\}$ are the natural numbers and $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$. $\mathbb{Z} = \{\ldots,-2,-1,0,1,2,\ldots\}$ are the integers ....
$\mathbb{R}, \mathbb{R}_+, \bar{\mathbb{R}}$	$\mathbb{R}$ are real numbers, $\mathbb{R}_+ = [0,\infty)$ are non-negative real numbers and $\bar{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\}$ ....
$\mathcal{B}$	Borel sigma-algebra ....
$\delta_{i,j}$	Kronecker's delta ....
x'	Transpose of a vector $x \in \mathbb{R}^d$ ....
$\lfloor x \rfloor, \lceil x \rceil$	Floor function $\lfloor x \rfloor = \max\{m \in \mathbb{Z} \,\vert\, m \leq x\}$ and ceiling function $\lfloor x \rfloor = \min\{m \in \mathbb{Z} \,\vert\, m \geq x\}$ for any $x \in \mathbb{R}$ ....
Γ	Gamma function ....
$\mathbb{1}, \mathbb{1}_M(x)$	Vector $\mathbb{1} = [1,\ldots,1]' \in \mathbb{R}^d$ and the indicator function $\mathbb{1}_M$ defined for a subset $M \subseteq \mathbb{R}$ by $\mathbb{1}_M(x) = 1$ if $x \in M$ and by $\mathbb{1}_M(x) = 0$ otherwise ....
$x_n \downarrow a$	The sequence x_n converges monotonically decreasing towards $a \in \bar{\mathbb{R}}$. The case $x_n \uparrow a$ is defined analogously ....
$f^a(x)$	$f^a(x) := [f(x)]^a$ ....
$\int$	$\int := \int_{[0,1]}$ ....
$z_\alpha(X)$	α-quantile of a generic random variable X ....
$\vert x\vert_2$	Euclidean norm $\vert x\vert_2 = (x'x)^{1/2}$ of a vector $x \in \mathbb{R}^d$. We use the abbreviation $\vert x\vert = \vert x\vert_2$, whenever no confusion is possible ....
$\vert x\vert_\Sigma$	Energy norm $\vert x\vert_\Sigma = (x'\Sigma^{-1}x)^{1/2} = \vert\Sigma^{-1/2}x\vert$ for a vector $x \in \mathbb{R}^d$ where the matrix $\Sigma \in \mathbb{R}^{d\times d}$ is supposed to be symmetric and positive definite ....
$\vert A\vert_F$	Frobenius norm $\vert A\vert_F = (\sum_{i=1}^m \sum_{j=1}^n a_{i,j}^2)^{1/2}$ of a matrix $A \in \mathbb{R}^{m\times n}$ ....

Notation for Hilbert spaces for Chapter 2 and Chapter 3

$H, L^2[0,1]$	A generic real Hilbert space H and the specific Hilbert space $L^2[0,1] = L^2([0,1], \mathcal{B}_{[0,1]}, \lambda_{[0,1]})$ of real-valued functions with domain $[0,1]$ that are measurable with respect to the Borel sigma-algebra $\mathcal{B}_{[0,1]}$ and are square integrable with respect to the Lebesgue-measure $\lambda_{[0,1]}$ ....
$d_H, \mathbb{N}_H$	$d_H \in \mathbb{N} \cup \{\infty\}$ is the dimension of the Hilbert space H. We set $\mathbb{N}_H = \{1, \ldots, d_H - 1\}$ if $d_H \in \mathbb{N}$ and $\mathbb{N}_H = \mathbb{N}$ if $d_H = \infty$ ..
$\langle v, w \rangle_H, \|v\|_H$	Hilbert space inner product $\langle v, w \rangle_H$ of two elements $v, w \in H$ and the corresponding norm $\|v\|_H$ ....
$\langle \mathscr{H}, \mathscr{K} \rangle_\mathcal{S}, \|\mathscr{H}\|_\mathcal{S}$	$\langle \mathscr{H}, \mathscr{K} \rangle_\mathcal{S} = \sum_{i=1}^{d_H} \langle \mathscr{H}(e_i), \mathscr{K}(e_i) \rangle_H$ is the Hilbert-Schmidt inner product of two linear operators $\mathscr{H}, \mathscr{K} \colon H \to H$ where $\{e_1, e_2, \ldots\}$ is a basis of a separable Hilbert space H. $\|\mathscr{H}\|_\mathcal{S}$ denotes the corresponding norm ....
$\|\mathscr{H}\|_\mathcal{L}$	Operator norm of a linear operator $\mathscr{H} \colon H \to H$ ....
$f \otimes g$	Tensor product of $f, g \in H$ defined by $(f \otimes g)h := f \langle g, h \rangle_H$, $h \in H$, which is a rank-one Hilbert-Schmidt operator on H. The mapping is linear in all arguments $f, g, h \in H$. We use the notation $E[f \otimes g]h = E[(f \otimes g)h]$ for H-valued random elements f, g and h ....

Notation for CUSUM testing via long run principal components in Chapter 2

The notation below will be used in Subsection 2.3.2 and Subsection 2.3.3. In the introductory Chapter 1 and in the preliminary discussions of Subsection 2.3.1 we use some of this notation slightly different.

Y_i, ε_i, m_i	$Y_i = \varepsilon_i + m_i$, $1 \le i \le n$, is a Hilbert space valued signal plus noise model. ε_i's are the random noise terms and m_i's are the deterministic means ....
$\mathscr{C}, \mathscr{C}_r, \hat{\mathscr{C}}, \hat{\mathscr{C}}_r$	Long run covariance operator $\mathscr{C} = \sum_{r \in \mathbb{Z}} \mathscr{C}_r$, where $\mathscr{C}_r = E[\varepsilon_0 \otimes \varepsilon_r]$, $r \in \mathbb{Z}$, are the lagged covariance operators. $\hat{\mathscr{C}} = \sum_{r=-n}^{n} \mathscr{K}(r/h) \hat{\mathscr{C}}_r$ is the corresponding Bartlett-type estimate with cross-covariance operator estimates $\hat{\mathscr{C}}_r = \sum_{i=1}^{n-r} [\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i+r}]/n$ for $r \ge 0$ and $\hat{\mathscr{C}}_r = \sum_{i=1}^{n+r} [\hat{\varepsilon}_{i-r} \otimes \hat{\varepsilon}_i]/n$ for $r < 0$, where $\hat{\varepsilon}_i = Y_i - \bar{Y}_n$. $\mathscr{K}$ is the kernel and h is the bandwidth ....

$\mathscr{G}(g_i,g_j), \mathscr{G}(g_j), \mathcal{G}_{g_j}$	We set $\mathscr{G}(g_i,g_j) = \int_0^1 g_i(x)g_j(x)dx - \int_0^1 g_i(x)dx \int_0^1 g_j(x)dx$ and $\mathscr{G}(g_j) = \mathscr{G}(g_j,g_j)$ for piecewise Lipschitz continuous trend functions $g_j(x)$, $x \in [0,1], 1 \le j \le \varrho$ (cf. p. 26 and 43). Moreover, we define $\mathcal{G}_{g_j}(x) = \int_0^x g_j(y)dy - x\int_0^1 g_j(y)dy$, $x \in [0,1]$, for $1 \le j \le \varrho$ ....
$\mathscr{C}_{\alpha,\beta}$	We set $\mathscr{C}_{\alpha,\beta} = \alpha\mathscr{C} + \sum_{i,j=1}^{\varrho} \beta_{i,j}[\Delta_i \otimes \Delta_j]$ under H_A and formally $\mathscr{C}_{1,0} := \mathscr{C}$ under H_0. $\mathscr{C}_{\alpha,\beta}$ is a perturbed version of the long run covariance operator $\mathscr{C}$ in the former case. The Δ_j's are the ϱ change directions, $\alpha \in \mathbb{R}_+$ is a scalar and $\beta = \beta_h\beta_{\mathcal{K}}\beta_{\mathcal{G}}$ is a matrix-valued parameter. The factors $\beta_h = (2h+1)$ and $\beta_{\mathcal{K}} = \int_0^\infty \mathscr{K}(x)dx$ are scalars that depend on the bandwidth h and on the kernel $\mathscr{K}$ of the Bartlett-type estimate $\hat{\mathscr{C}}$. Furthermore, the matrix $\beta_{\mathcal{G}} = (\mathscr{G}(g_i,g_j))_{i,j=1,\dots,\varrho}$ depends on the $\varrho \in \mathbb{N}_H$ piecewise Lipschitz continuous trend functions g_j. (See p. 26 for the definition of the latter and for the formulation of the testing hypotheses H_0 and H_A.) ....
$(\lambda_j, v_j), (\hat{\lambda}_j, \hat{v}_j)$	$(\lambda_j,\ v_j)$ are the eigenelements of the operator $\mathscr{C}_{\alpha,\beta}$ given by the spectral decomposition. λ_j are the eigenvalues and v_j are the corresponding eigenvectors. In case of $\mathscr{C}_{1,0} = \mathscr{C}$ these are the population long run principal components. $(\hat{\lambda}_j,\ \hat{v}_j)$ are eigenelements of the Bartlett-type estimate $\hat{\mathscr{C}}$ given by the spectral decomposition. These are the empirical long run principal components ....
$\boldsymbol{y}, \hat{\boldsymbol{y}}, \hat{\boldsymbol{y}}^{\Delta}$	$\boldsymbol{y}$ represents the sequence $\boldsymbol{y}_i = [\langle Y_i, v_1\rangle, \dots, \langle Y_i, v_d\rangle]'$ of the population principal component scores, $\hat{\boldsymbol{y}}$ represents the sequence $\hat{\boldsymbol{y}}_i = [\langle Y_i, \hat{v}_1\rangle, \dots, \langle Y_i, \hat{v}_d\rangle]'$ of their empirical counterparts and finally $\hat{\boldsymbol{y}}^{\Delta}$ represents their change-aligned empirical version $\hat{\boldsymbol{y}}_i^{\Delta} = [\langle Y_i, \hat{v}_1^{\Delta}\rangle, \langle Y_i, \hat{v}_2\rangle, \dots, \langle Y_i, \hat{v}_d\rangle]'$. (See p. 55 for the construction of $\hat{v}_1^{\Delta}$.) ....
$\Sigma, \hat{\Sigma}$	$\Sigma = \sum_{r\in\mathbb{Z}} \mathrm{Cov}(\boldsymbol{y}_0, \boldsymbol{y}_r)$ is the (formal) long run covariance matrix of the time series $\{\boldsymbol{y}_i\}_{i\in\mathbb{Z}}$ of long run principal component scores, where $\mathrm{Cov}(\boldsymbol{y}_0, \boldsymbol{y}_r) = E[(\boldsymbol{y}_0 - E\boldsymbol{y}_0)(\boldsymbol{y}_r - E\boldsymbol{y}_r)']$. A corresponding estimate is given by $\hat{\Sigma} = \mathrm{diag}(\lvert\hat{\lambda}_1\rvert, \dots, \lvert\hat{\lambda}_d\rvert)$, where $\hat{\lambda}_j$'s are eigenelements of the Bartlett-type estimate $\hat{\mathscr{C}}$ ......
$w(x)$	Darling-Erdős-type weighting function $w(x) = [x(1-x)]^{-1/2}$ for $x \in (0,1)$ ....
$S_n(x;\varepsilon), S_n'(x;\varepsilon)$	Partial sum $S_n(x;\varepsilon) = \sum_{j=1}^{\lfloor nx\rfloor}(\varepsilon_j - \bar{\varepsilon}_n)/n^{1/2}$ and a modification $S_n'(x;\varepsilon) = (\sum_{j=1}^{\lfloor nx\rfloor}\varepsilon_j - x\sum_{i=1}^{n}\varepsilon_i)/n^{1/2}$, where ε represents any Hilbert space valued time series $\{\varepsilon_i\}$ ....

$\mathscr{T}(x), \hat{\mathscr{T}}(x), \hat{\mathscr{T}}^{\Delta}(x)$	$\mathscr{T}(x) = \mathscr{T}(x; \boldsymbol{y}) = \lvert S_n(x; \boldsymbol{y})\rvert_{\Sigma}$ is a CUSUM detector based on population long run principal components $\boldsymbol{y}$. $\hat{\mathscr{T}}(x) = \hat{\mathscr{T}}(x; \hat{\boldsymbol{y}}) = \lvert S_n(x; \hat{\boldsymbol{y}})\rvert_{\hat{\Sigma}}$ is a CUSUM detector based on empirical, i.e. estimated, long run principal component scores $\hat{\boldsymbol{y}}$. Finally, $\hat{\mathscr{T}}^{\Delta}(x) = \hat{\mathscr{T}}^{\Delta}(x; \hat{\boldsymbol{y}}^{\Delta}) = \lvert S_n(x; \hat{\boldsymbol{y}}^{\Delta})\rvert_{\hat{\Sigma}}$ is a CUSUM detector that is based on empirical change-aligned principal component scores $\hat{\boldsymbol{y}}^{\Delta}$ ....
$\mathscr{M}_n, \hat{\mathscr{M}}_n, \hat{\mathscr{M}}_n^{\Delta}$	$\mathscr{M}_n = \mathscr{M}_n(\boldsymbol{y}) = \max_{1\leq k<n} w(k/n)\mathscr{T}(k/n)$ is the Darling-Erdős-type CUSUM statistic based on population long run principal components $\boldsymbol{y}$. $\hat{\mathscr{M}}_n = \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) = \max_{1\leq k<n} w(k/n)\hat{\mathscr{T}}(k/n)$ is the Darling-Erdős-type CUSUM statistic based on empirical, i.e. estimated, long run principal component scores $\hat{\boldsymbol{y}}$. $\hat{\mathscr{M}}_n^{\Delta} = \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) = \max_{1\leq k<n} w(k/n)\hat{\mathscr{T}}^{\Delta}(k/n)$ is the Darling-Erdős-type CUSUM statistic based on empirical change-aligned principal component scores $\hat{\boldsymbol{y}}^{\Delta}$ ....
$\boldsymbol{W}(t)$	Standard d-dimensional Wiener process, i.e. a centered Gaussian process $\boldsymbol{W}(t) = [W_1(t), \ldots, W_d(t)]'$ with covariance function $\operatorname{Cov}(W_i(t), W_j(s)) = \delta_{i,j}\min(t,s)$ for every $t, s \in [0, \infty)$. The W_i are standard univariate Wiener processes ....
$\boldsymbol{B}(t)$	Standard d-dimensional Brownian bridge, i.e. a centered Gaussian process $\boldsymbol{B}(t) = [B_1(t), \ldots, B_d(t)]'$ with covariance function $\operatorname{Cov}(B_i(t), B_j(s)) = \delta_{i,j}[\min(t,s) - ts]$ for every $t, s \in [0, 1]$. The B_i are standard univariate Brownian bridges .....
$\boldsymbol{U}(t)$	Standard d-dimensional Ornstein-Uhlenbeck process, i.e. a centered Gaussian process $\boldsymbol{U}(t) = [U_1(t), \ldots, U_d(t)]'$ with covariance function $\operatorname{Cov}(U_i(t), U_j(s)) = \delta_{i,j}\exp(-\lvert t-s\rvert/2)$ for all $t, s \in (-\infty, \infty)$. The components U_i are standard univariate Ornstein-Uhlenbeck processes ....

Notation for CUSUM based estimation in panel data in Chapter 3

$Y_{i,k}, \varepsilon_{i,k}, m_{i,k}, \zeta_i, \gamma_k$	$Y_{i,k} = \varepsilon_{i,k} + m_{i,k} + \gamma_k\zeta_i$, $1 \leq i \leq n$, $1 \leq k \leq d$, is a Hilbert space valued signal plus noise model. $\varepsilon_{i,k}$'s are the noise terms, $m_{i,k}$'s are the means, ζ_i's are the common factors and γ_k's are the corresponding factor loadings. k represents the panels and i the time points ....
I_n, x	$I_n = \{1/n, 2/n, \ldots, (n-1)/n\}$, $n \in \mathbb{N}$, is the discrete, time-rescaled domain. $\mathsf{x} = \mathsf{x}_n$ is a discrete variable with values restricted to the grid I_n. We use this notation to distinguish between discrete-time and continuous-time arguments. For the latter we use x which can take any value in $[0, 1]$ ....

$S_{n,k}(x;\varepsilon)$	$S_{n,k}(x;\varepsilon) = \sum_{j=1}^{\lfloor nx \rfloor}(\varepsilon_{j,k} - \bar{\varepsilon}_{n,k})/n^{1/2}$ with $\bar{\varepsilon}_{n,k} = \sum_{i=1}^{n} \varepsilon_{i,k}/n$ and $x \in I_n$ are the centered partial sums of the noise sequences $\{\varepsilon_{i,k}\}_{i=1,\dots,n}$ for $1 \le k \le d$. Note that ε represents the noise array $\{\varepsilon_{i,k}\}_{1\le i\le n, 1\le k\le d}$ ....
$\rho, \sigma, \Delta, \Delta_k, \bar{\Delta}_d$	Rescaled noise-to-change ratio $\rho(\Delta,\sigma,n) = \sigma^2/(n\Delta)$, where $\Delta = \lim_{d\to\infty} \bar{\Delta}_d \in (0,\infty)$ is the total average change, $\bar{\Delta}_d = \sum_{k=1}^{d} \lVert\Delta_k\rVert_H^2/d$ is the average change and $\sigma^2 = E\lVert\varepsilon_{i,k}\rVert^2$ are the second moments. Parameter $\sigma \in (0,\infty)$ is the same for all $1 \le i \le n, 1 \le k \le d$. (See p. 99 for the formulation of the common change point setting and p. 100 for the definition of changes Δ_k.) ....
$V(x)$	$V^2(x) = (E\lVert S_{n,1}(x;\varepsilon)\rVert_H^2)/\sigma^2, x \in I_n$, is the variance of the cumulated noises ....
$w(x), w_\gamma(x), w_\star(x)$	$w(x)$, $x \in I_n$, is an arbitrary positive weighting function. $w_\gamma(x) = [x(1-x)]^{-\gamma}$, $x \in I_n$, are the classical CUSUM weighting functions with $\gamma \in [0, 1/2]$. Finally, $w_\star(x) = w_{1/2}(x)/h(x)$, $x \in I_n$, is a covariance-based weighting function, where $h(x) = 1/V(x)$, $x \in I_n$ ....
$C(x;y,\rho), H(x,y)$	Critical function $C(x;y,\rho) = w^2(x)[V^2(x)\rho + H^2(x,y)]$, $x, y \in I_n$, where $H(x,y) = \min\{x,y\}(1-\max\{x,y\})$ ....
$\mathscr{T}(x)$	CUSUM detector $\mathscr{T}(x) = (\sum_{k=1}^{d} \lVert S_{n,k}(x;Y)\rVert_H^2)^{1/2}$, $x \in I_n$ ..
$u, \hat{u}, s, \varsigma$	Any $\hat{u} \in \arg\max_{1\le i<n} w(i/n)\mathscr{T}(i/n)$ is a CUSUM estimate of a common change point u. (See p. 99 for the formulation of the corresponding change point setting.) Furthermore, we set $s = u/n$ and $\varsigma(s) = \max\{s, 1-s\}$ ....
$\Sigma, \gamma(k,l)$	$\Sigma = (\gamma(k,r))_{k,l=1,\dots,n}$ with $\gamma(k,r) = E\langle\varepsilon_{k,1}, \varepsilon_{r,1}\rangle_H$. It is a covariance matrix if all $\varepsilon_{k,1}$'s are univariate and real-valued ...
$\hat{\Sigma}, \hat{\Sigma}'$	$\hat{\Sigma}$ and $\hat{\Sigma}'$ are estimates of Σ. (See p. 113.) ....
$\mathcal{F}(x), \Gamma_x$	$\mathcal{F}(x) = x(1-x)[\Gamma_x + \Gamma_{1-x} - \Gamma_1]/\sigma^2$, $x \in I_n$. Furthermore, $\Gamma_x = \sum_{k,r=1}^{\lfloor nx \rfloor} \gamma(k,r)/\lfloor nx \rfloor$ for $x \in I_n$ ....
$R(x,s), F(x), G(x,s)$	$R(x,s) = [G(x,s) - G(s,s)]/[F(s) - F(x)]$, where $F(x) = [w(x)V(x)]^2$, $G(x,s) = [w(x)H(x,s)]^2$, $x, s \in I_n$ ....
λ	Shrinkage parameter λ in the penalty term of the penalized least squares and of the corresponding group fused LASSO estimates ....
$\hat{U}(\lambda), \hat{\beta}(\lambda), \hat{\mathscr{E}}(\lambda)$	$\hat{U}(\lambda)$ and $\hat{\beta}(\lambda)$ are solutions of the penalized least squares and of the corresponding group fused LASSO approaches. $\hat{\mathscr{E}}(\lambda)$ is the resulting set of common change point estimates ......

Some abbreviations

CUSUM	Cumulative sums ....
HDLSS	High-dimension low sample size (framework) ....
KKT	Karush-Kuhn-Tucker (conditions) ....
LASSO	Least absolute shrinkage and selection operator ....

Abbreviations for time series models:

WN	White noise ....
AR, FAR	Autoregressive and functional autoregressive (time series) ...
MA	Moving average (time series) ....
ARMA	Autoregressive moving average (time series) ....
ARCH	Autoregressive conditional heteroscedastic (time series) ....

CHAPTER 1

GENERAL INTRODUCTION

> "Change point analysis (...) originated in the 1940s and initially focused on data-driven quality control techniques. Over time, methods in change point analysis have been developed to address data analytic questions in fields ranging from biology to finance, and in many cases such methodology has become standard.*
>
> — HORVÁTH & RICE (2014, P. 220)

Data can be collected nowadays with ease all the time and everywhere. As a consequence a lot of data in practice comes in modern high-dimensional data types. Two prominent examples are infinite dimensional functional time series and high-dimensional panel data. This thesis will contribute to change point analysis techniques for both data types.

Change point analysis is a broad statistical testing and estimation methodology that provides answers to questions of parameter stability in random data. According to Brodsky & Darkhovsky (1993, p. 11): *»The change point problem can be considered to be one of the central problems of statistical inference, linking together statistical control theory, the theory of estimation and testing hypotheses, classical and Bayesian approaches, and fixed sample and sequential procedures.«* Among the early milestones are, e.g., the famous articles by Page (1954, 1955) that are concerned with quality control problems. Since then, change point analysis has evolved into a dynamic research field that covers many different disciplines with a huge amount of different approaches and well-developed theory. For a systematic introduction to (some) theoretical foundations we refer, e.g., to two well-known books by Brodsky & Darkhovsky (1993) and Csörgő & Horváth (1997). Especially the latter is of importance for this thesis and contains Darling-Erdős-type limit theorems that set the stage for our own theoretical results.

In this thesis we focus on Darling-Erdős-type cumulative sum (CUSUM) procedures and we contribute to change point testing and estimation theory for functional time series and for panel data. Modern developments in change point analysis within these environments are reviewed, e.g., by Aue & Horváth (2013), by Jandhyala *et al.* (2013) and in a discussion paper by Horváth & Rice (2014). The latter address functional data, panel data, as well as Darling-Erdős-type laws, separately, and provide references to relevant literature. Aue & Horváth (2013) focus on CUSUM-type procedures for time series. Jandhyala *et al.* (2013) provide an overview of estimation methods for multiple change points including the total variation denoising approach that we will also contribute to. A survey of change point literature within different fields that range from econometrics to machine learning is collected in Frick *et al.* (2014).

Note that our research is to a large extent inspired by the book of Horváth & Kokoszka (2012). It contains a variety of results on change point methodology for functional data and is an excellent starting point to get familiar with the Hilbert space change point setting that we are working with. (For overviews of general functional data analysis we refer, e.g., to Ramsay & Silverman, 2005, Ferraty & Vieu, 2006 and to Ramsay *et al.*, 2009.)

A standard change point testing and estimation problem 1.1

For a discussion of recent research developments in change point analysis and of our contributions to this research we will introduce some notation and be slightly more formal in the following. We begin with a setting that is a central building block of our research and that will be important throughout the whole thesis. Consider a multivariate $\mathbb{R}^d$-valued time series

$$\boldsymbol{Y}_i = \boldsymbol{m}_i + \boldsymbol{e}_i, \quad i \in \mathbb{Z}, \tag{1.1.1}$$

in a signal plus noise model where $\boldsymbol{e}_i$, $i \in \mathbb{Z}$, are some random error terms, i.e. the noises, and $\boldsymbol{m}_i = E\boldsymbol{Y}_i$, $i \in \mathbb{Z}$, are the expectations, i.e. the signals. A basic question of parameter stability, which we will work on, is to decide whether an observable sample $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ has constantly same means $\boldsymbol{m}_1, \ldots, \boldsymbol{m}_n$ or whether these means change over time. In terms of hypothesis testing the aim is to decide whether we are under the null hypothesis

$$H_0 : \quad \boldsymbol{m}_1 = \ldots = \boldsymbol{m}_n$$

or under the alternative H_A that H_0 does not hold. If the null hypothesis is rejected, then one typically is interested in estimating the time point of the change, or more generally speaking, in estimating the change pattern. A lot of testing and estimation theory is concerned with, or is motivated by, the idealized abrupt change setting

$$H_A^{\text{abrupt}} : \quad \boldsymbol{m}_1 = \ldots = \boldsymbol{m}_u \neq \boldsymbol{m}_{u+1} = \ldots = \boldsymbol{m}_n \tag{1.1.2}$$

that is also the focus of this thesis. One standard approach to test H_0 against H_A^{abrupt}, e.g., in stationary weakly dependent settings, is to work with maximum-type CUSUM test statistics

$$\mathscr{M}_n(\boldsymbol{Y}) = \max_{1 \le k < n} w(k/n)\mathscr{T}(k/n), \tag{1.1.3}$$

where $\mathscr{T}(x)$ is a detector given by

$$\mathscr{T}(x) = \mathscr{T}(x; \boldsymbol{Y}) = |\Sigma^{-1/2} S_n(x; \boldsymbol{Y})| = |S_n(x; \boldsymbol{Y})|_\Sigma \tag{1.1.4}$$

and where $w(x)$, $x \in (0, 1)$, is a weighting function that will be specified further below. Moreover, $\boldsymbol{Y}$ is an abbreviation for $\{\boldsymbol{Y}_i\}_{i \in \mathbb{Z}}$, Σ is the long run covariance matrix of $\{\boldsymbol{Y}_i\}_{i \in \mathbb{Z}}$, $S_n(x; \boldsymbol{Y}) = \sum_{i=1}^{\lfloor nx \rfloor} (\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}_n)/n^{1/2}$, $|\cdot|$ is the Euclidean norm and $|\cdot|_\Sigma = |\Sigma^{-1/2} \cdot |$.

A standard way to estimate changes under H_A^{abrupt} is to rely on

$$\hat{u} \in \underset{1\le k<n}{\arg\max}\, w(k/n)\mathscr{T}(k/n) \tag{1.1.5}$$

using the same detector and same weighting function as for the test statistic before.[1] The detector $\mathscr{T}(x)$ relies essentially on a weighted comparison of means

$$\left[\frac{1}{(k/n)(1-k/n)}\right] S_n(k/n;\boldsymbol{Y}) = \left[\frac{1}{k}\sum_{i=1}^{k}\mathbf{Y}_i - \frac{1}{n-k}\sum_{i=k+1}^{n}\mathbf{Y}_i\right] n^{1/2} \tag{1.1.6}$$

over all possible change locations $1 \le k < n$. This relation indicates why the statistic $\mathscr{M}_n(\boldsymbol{Y})$ is heuristically reasonable for testing of H_0 against H_A^{abrupt} for large sample sizes n. Under H_0 the difference of means in (1.1.6) is likely to be small for all k whereas under H_A^{abrupt} it is expected to be comparably large for $k \approx u$ where u is the actual change point. Moreover, in the latter situation it is likely to have a peak (i.e. the largest difference) around u which indicates why the estimate $\hat{u}$ is a reasonable choice.

The weighting function w, in (1.1.3) and in (1.1.5), controls the fluctuations of mean differences (1.1.6) at borders, i.e. for $k \approx 1$ and $k \approx n$. The behavior of (1.1.6) in these regions is more erratic which may lead by pure chance, on the one hand, to false rejections with $\mathscr{M}_n(\boldsymbol{Y})$ under the null hypothesis and, on the other hand, to incorrect estimation with $\hat{u}$ under the alternative (if, e.g., the true abrupt change lies rather centered). Generally, the weighting function w can be chosen quite flexible but among all the possibilities the weighting function

$$w(x) = [x(1-x)]^{-1/2}, \tag{1.1.7}$$

$x \in (0,1)$, has an outstanding position: the statistic (1.1.3) and the estimate (1.1.5) with this particular weighting arise, e.g., via the quasi-maximum likelihood approach(es) and they are commonly referred to as being of Darling-Erdős-type.[2] This is due to their large sample asymptotics which rely on the already mentioned class of Darling-Erdős-type limit theorems that can be traced back to Darling & Erdős (1956) (cf., e.g., Csörgő & Horváth, 1997 and our discussion on p. 48). Note that, for simplicity, we will also call (1.1.7) a Darling-Erdős-type weighting function. This specific weighting function will be of a particular importance throughout this thesis for testing and for estimation.

We are now in a position to explain the concept of this thesis and to highlight our contributions in the next paragraphs. Additional details on our contributions are given in the introductions of the main chapters.

[1] Note that $\arg\max$ is defined here as a set. Above notation means that any element $\hat{u}$ within this set may be chosen as an estimate.

[2] Note that (1.1.5) with weighting function (1.1.7) is also the least squares estimate.

Extensions of standard methodology 1.2

A lot of research on change point analysis focuses on extensions of standard methodology that was developed for the above multivariate abrupt change setting. In the following we discuss three directions that are important within this thesis.

I. Extensions to high-dimensional data

The first direction of extensions is the adaptation of methods to signal plus noise models (1.1.1) for modern high-dimensional data types such as functional and panel data which are the focus of change point analysis over the last five to fifteen years (cf., e.g., Bai, 2010, Aue & Horváth, 2013 and Horváth & Rice, 2014). Some early works on change point problems in functional data settings are by Berkes *et al.* (2009), Hörmann & Kokoszka (2010) and Aston & Kirch (2012a), which all serve as a basis for our considerations.[1,2] According to Bai (2010) and to Horváth & Hušková (2012), change point analysis for panel data gained some attention in the early 2000s but goes back at least to Joseph & Wolfson (1992, 1993). In this thesis we are interested in a more recent high-dimensional low sample size (HDLSS) panel data setting where the number of observations in the time domain $n \in \mathbb{N}$ is fixed but the amount of panels $d \in \mathbb{N}$, i.e. the dimension, tends to infinity. This new HDLSS asymptotic framework is interesting because it demonstrates effects for high dimensions d that fade out if we let $n \to \infty$. For change point problems this setting was considered first by Bai (2010) and Bleakley & Vert (2010, 2011a) relying on least squares approaches. (The latter adapted a modern penalized least squares estimation procedure from the machine learning community.) Even more recently, this setting was picked up for CUSUM-type procedures, e.g., by Peštová & Pešta (2015) and by Horváth *et al.* (2016).

We will contribute to change point methods for infinite dimensional functional data in Chapter 2 (cf. Figure 1.1, below) and for HDLSS panel data in Chapter 3 (cf. Figure 1.2, below). In both chapters we work on abstract Hilbert space valued time series, i.e. particularly covering the popular $L^2[0,1]$ scenario.

1. In the functional data setting of Chapter 2 we are concerned with testing. We base our research upon the principal component dimension reduction approach suggested by Berkes *et al.* (2009) in a non-parametric setting for CUSUM procedures. We extend their non-parametric methodology to the Darling-Erdős-type case of $w(x) = [x(1-x)]^{-1/2}$, $x \in (0,1)$. (Note that Darling-Erdős-type procedures are also considered by Zhou (2011) in a Gaussian special case.) Furthermore, we suggest to use a new tensor based formulation of CUSUM statistics using projections on whole principal component subspaces rather than separate projections on principal component directions. This formulation has advantages: for one thing, it simplifies and clarifies the proofs, and, for another, it allows to weaken the usual

[1] Note that they are all formulated for $L^2[0,1]$-valued data.

[2] The functional two-sample testing problem has been considered even earlier. (Cf. Berkes *et al.*, 2009.)

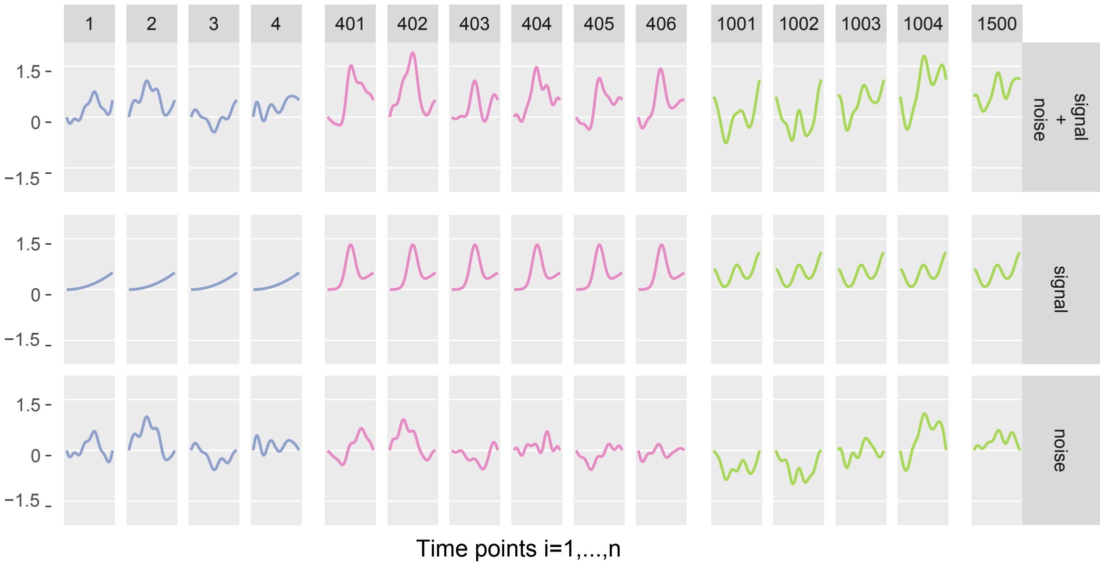

Figure 1.1: This figure indicates the functional signal plus noise model of Chapter 2. The top panel shows a subset of observations from a functional time series of length $n = 1500$. (We will get back to the whole time series in Remark 1.3 and in Figure 1.3, below.) The middle panel shows the corresponding signals (or means) that have abrupt changes at time points $i = 400$ and at $i = 1000$. The lower panel shows the corresponding random functional noise.

assumptions on eigenvalue separation using the results of Reimherr (2015). Moreover, it clarifies the relation between projection based procedures and traditional multivariate approaches.

2. In the panel data setting of Chapter 3 we work on estimation of common change points and base our research upon Bleakley & Vert (2010, 2011a) but our results complement the findings of Bai (2010), Peštová & Pešta (2015) and Horváth *et al.* (2016), as well. We investigate the relation of the weighted penalized least squares approach of Bleakley & Vert (2010, 2011a) to the traditional weighted CUSUM estimates in a single change point framework and show that they coincide under mild assumptions. For instance, this connects the work of Bai (2010) with the works of Bleakley & Vert (2010, 2011a). Relying on this observation we propose to study a general class of weighted CUSUM-type estimates in the HDLSS setting with a focus on their accuracy for $d \to \infty$. Moreover, we propose to study panels of Hilbert space valued time series which, to the best of our knowledge, is new in the change point panel data context. Intuitively, one would expect accuracy to increase if we add more panels that share a common change point, i.e. if we have more data with more information about the change position. However, as observed by Bleakley & Vert (2010) estimation might be (surprisingly) less accurate. Bleakley & Vert (2010, 2011a) study a penalized least squares approach using weighting functions $w(x) = 1$ or $w(x) = [x(1-x)]^{-1/2}$, $x \in (0,1)$, and derive sharp bounds on the change-to-noise ratio that leads to accurate estimation under Gaussianity and under i.i.d. assumptions. (These bounds are functions of change magnitudes and of change locations.) We continue their study for CUSUM

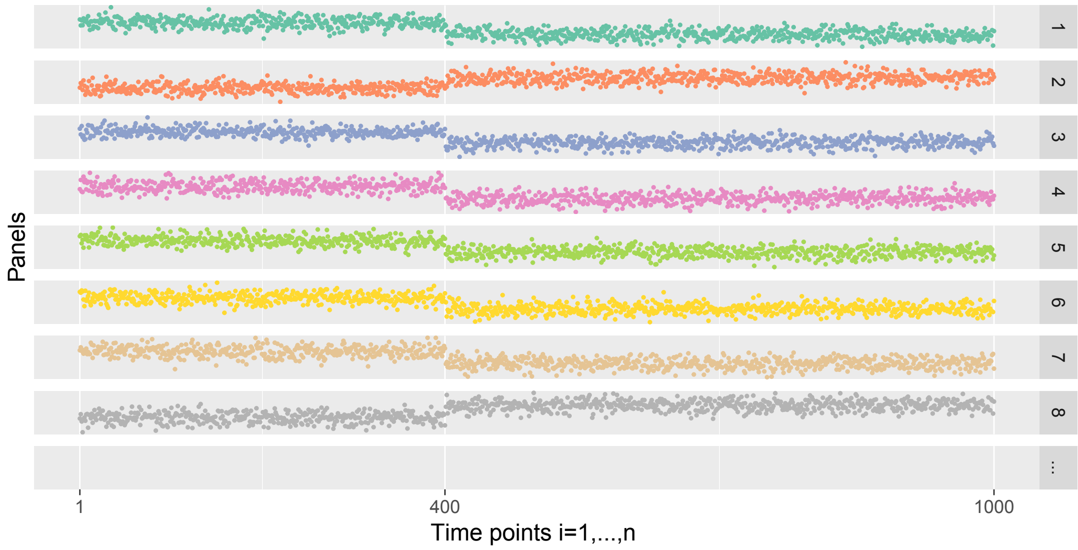

Figure 1.2: This figure indicates the signal plus noise model in the panel data setting of Chapter 3. It shows a subset of eight panels of univariate real-valued time series of length $n = 1000$ with a common change point at time point $i = 400$.

estimates and consider a class of common CUSUM weighting functions

$$w_\gamma(x) = [x(1-x)]^{-\gamma}, \tag{1.2.1}$$

$x \in (0,1)$, $\gamma \in (0,1/2)$, that contains the weighting functions of Bleakley & Vert (2010, 2011a) as limiting cases for $\gamma = 0$ and $\gamma = 1/2$. Weights (1.2.1) are a well-known choice in classical CUSUM theory but were not considered in the HDLSS panel data setting before.

II. Extensions to dependent data

The second aim of this thesis is the adaptation of existing theory to more general distributional assumptions and, particularly, the extension to more (modern) dependence concepts.[1] We are working on test statistics and estimates, (1.1.3) and (1.1.5), that have been well-known for a long time. Their asymptotic properties are thoroughly studied for univariate and multivariate settings with linear and nonlinear time series and a variety of classical dependence conditions (cf., e.g., Csörgő & Horváth, 1997). However, the situation is different for corresponding dimension reduction based tests in functional data and for estimation theory in high-dimensional panel data that we are looking at. For functional data Berkes *et al.* (2009) considered dimension reduction via static principal components in an i.i.d. setting. This was then extended under stationarity within the weak dependence concept of m-approximability by Hörmann & Kokoszka (2010), Aston & Kirch (2012a) and, using long run principal components, by Horváth

[1] Note that new methods are often developed under i.i.d. and normality assumptions which are later on typically replaced by stationarity together with some weak dependence concept.

et al. (2014).[1] For panel data Bai (2010) and Horváth & Hušková (2012) considered linear time series under dependence conditions formulated in terms of the coefficients decay. More recently, on the one hand, rather flexible moment type conditions were considered by Horváth *et al.* (2016) and, on the other hand, Peštová & Pešta (2015) contributed to testing and estimation within a HDLSS framework under certain monotonicity conditions on the autocovariances.[2] Finally, note that Bleakley & Vert (2010, 2011a) considered an i.i.d. HDLSS framework under Gaussianity.

We will extend the Darling-Erdős-type functional and panel data procedures to weak dependence concepts and avoid the normality assumptions of Bleakley & Vert (2010, 2011a) in the latter panel data setting.

1. For functional data in Chapter 2 we will consider Darling-Erdős-type procedures under m-dependence or m-approximability and work, as Berkes *et al.* (2013), with long run principal components in both cases. We will introduce these dependence concepts in detail in Section 1.3.[3]

2. For panel data in Chapter 3 we will adapt the theory of Bleakley & Vert (2010, 2011a) to the time series context and, additionally, incorporate common factors into the signal plus noise model. (The latter are highly popular in the econometric literature.) Bleakley & Vert (2011a) showed that the Darling-Erdős-type weighting is »optimal« in the following sense: accuracy of change estimation always increases if we consider more panels with same change position and (on average) same change magnitudes across those panels. However, as we will see, this result is limited to the i.i.d. case. Under dependence we may lose precision instead of benefiting from additional information if the noise is too dominant or the change is too small. We show that under dependence »optimality« may be (re-)obtained by taking the autocovariance structure of panels into account. More precisely, we introduce a new weighting function $w_\star$ that combines the traditional Darling-Erdős-type weighting function $w_{1/2}$ with a non-trivial covariance-based correction term h as follows:

$$w_\star(i/n) = w_{1/2}(i/n)/h(i/n),$$

$i = 1, \ldots, n$. Note that the correction term h may have a strong influence and thus the weights $w_\star$ may be surprisingly different from the traditional Darling-Erdős-type weights $w_{1/2}$ in finite samples (cf. (1.1.7) or (1.2.1)).

III. Extensions to general change models

A third direction of extensions that is important for our research are more general alternatives than H_A^{abrupt}. Popular frameworks that go beyond the abrupt change point setting are, e.g., multiple changes or gradual changes. These are well-studied in traditional settings. (Cf., e.g., Brodsky & Darkhovsky, 1993, Csörgő & Horváth, 1997 and

[1] Note that Aston & Kirch (2012a) considered also some mixing type conditions.
[2] Their setting as well as their findings are related to our results.
[3] The consideration of m-dependence is largely motivated by Horváth *et al.* (1999).

Jandhyala *et al.*, 2013.) Again, the situation is different for new high-dimensional settings where less literature is available.

1. In Chapter 2 we consider a multi-directional change point alternative in functional data context (cf. Figure 1.1) which extends the setting of Horváth *et al.* (2014) in a straightforward way and allows for gradual changes.[1] On the one hand, we derive conditions under which our method is consistent. On the other hand, we propose a new methodology of change-aligned principal components which ensures consistency in finite and infinite dimensional Hilbert spaces for all possible changes that are under consideration. This is achieved by incorporating a fully-functional estimate of an abrupt change.[2] Additionally, a remarkable property of this approach is that we may use a one-directional test for multi-directional changes. (Note that one-directional tests are also considered, e.g., by Aston & Kirch (2014).)

2. The research in Chapter 3 is inspired by a method that was developed by Harchaoui & Lévy-Leduc (2008) for multiple changes of means in univariate time series and then extended by Bleakley & Vert (2010, 2011a) to high-dimensional frameworks (cf. Figure 1.2). We will develop our theory as Bleakley & Vert (2010, 2011a) under a single abrupt change setting but it is worth noting that extensions to multiple change point scenarios are also possible (cf., e.g., Torgovitski, 2015b) and are part of ongoing research.

Overall, we consider generalizations of the signal plus noise setting (1.1.1) and of the change in the mean problems for (weakly) dependent finite or infinite dimensional Hilbert space data where we essentially rely on modifications of the test statistic (1.1.3) and of the estimate (1.1.5). An important goal of this thesis is to extend existing change point methodology such that more changes can be tested and accurately estimated in modern settings where the dimension is either infinite or finite (but high).

We finish this paragraph by a few remarks which indicate some similarities and some differences between the high-dimensional settings of Chapter 2 and Chapter 3.

Remark 1.1. Chapter 2 is motivated by the infinite dimensional functional setting of $L^2[0,1]$ and by the (functional) principal component analysis, yet we formulated our theory in a general Hilbert space including also classical multivariate data. One reason for this is that our theory does not use any specific functional structure and thus directly covers the multivariate case - using dimension reduction - as well.[3] Moreover, the abstract Hilbert space notation appears to be more concise and elegant in the sense that parallels between the traditional multivariate and the infinite dimensional settings become evident. In Chapter 3 we have the opposite situation. We are rather interested in panels of univariate real-valued time series but the theory (mostly) extends to panels of Hilbert space time series and thus is presented in this generality.

[1] The consideration of gradual changes is, again, motivated by Horváth *et al.* (1999).
[2] We work with this estimate even if the actual change is more complex.
[3] Note that we tacitly assume that the dimension of finite dimensional data is too high for traditional multivariate approaches (in which case dimension reduction, e.g., via principal components becomes appealing).

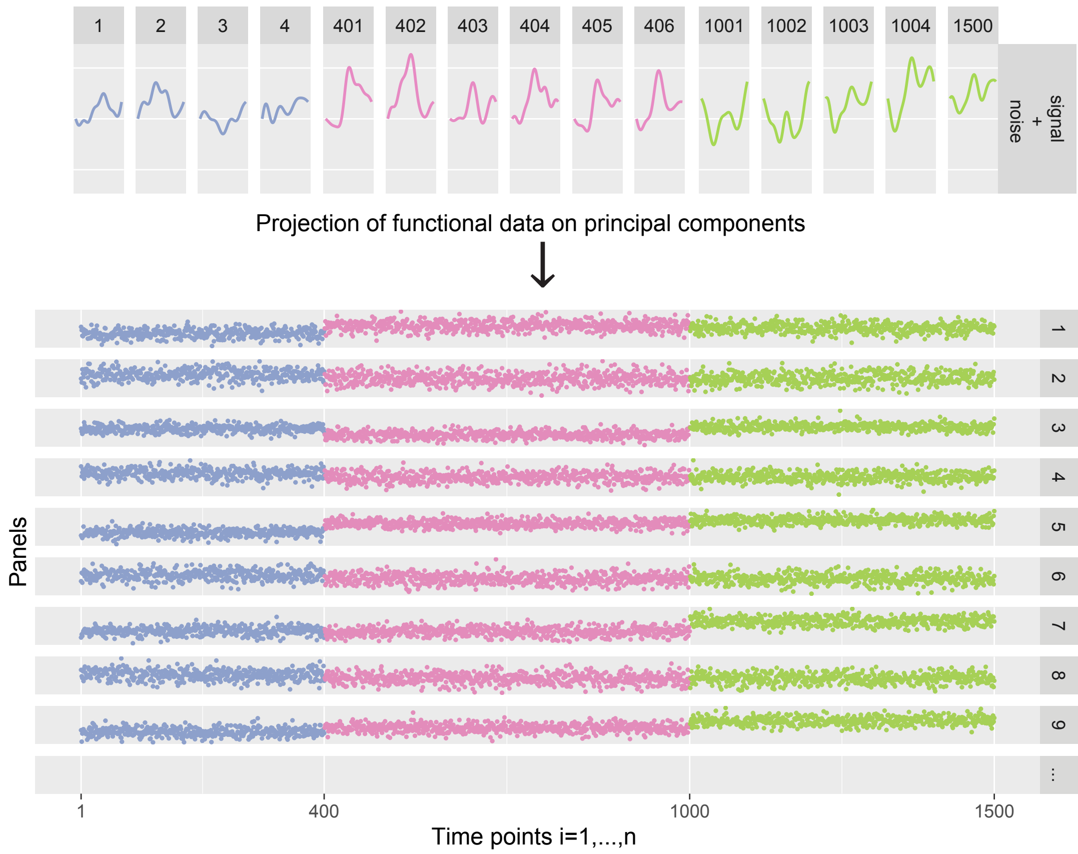

Figure 1.3: We represent the functional time series from Figure 1.1 as panel data via whitened projections on principal components and show the first nine panels. The two changes in the functional signals in Figure 1.1 at time points $i = 400$ and $i = 1000$ correspond here to common changes in the means across the panels at the same time points.

Remark 1.2. The settings of Chapter 2 and Chapter 3 are both high-dimensional (cf. previous remark) but the asymptotic frameworks are different. On the one hand, in Chapter 2 we consider a projection based approach to a modern infinite dimensional time series model in a traditional asymptotic setting, where the dimension d of the projection subspace is fixed and only the sample size n tends to infinity (i.e. $d \in \mathbb{N}$, $n \to \infty$). On the other hand, in Chapter 3 we consider a more traditional multivariate panel data model in the modern HDLSS framework, where the sample size n is fixed and the number of panels d, which is the dimension, tends to infinity (i.e. $n \in \mathbb{N}$, $d \to \infty$).

Remark 1.3 (Functional data interpreted as panel data). Our functional framework of Chapter 2 is interconnected with our panel data framework of Chapter 3 and we indicate this connection informally below. In Chapter 2 we use principal components as a dimension reduction technique. The projections of the original functional data on the principal components may be interpreted as panels of time series. (Each row of the projected data in Table 2.1, below, corresponds to a panel.) Assuming independence of the observed functional data and assuming Gaussianity we obtain independent projected panels that fit well (after a suitable normalization) into the HDLSS panel data framework of Chapter 3 for the following reason: in the infinite dimensional functional setting we may pick (theoretically) an arbitrarily large number of projections d. Hence, we may actually consider $d \to \infty$ asymptotics for the projected panels. Moreover, multiple abrupt changes in functional data necessarily correspond to common changes in panels (see Figure 1.3) for which the method of Bleakley & Vert (2011a) is designed. This interpretation was our first motivation to work on estimation in panel data that is presented in Chapter 3. (Obviously, this interpretation has limitations. For instance, we lose independence between panels if we do not assume Gaussianity. Moreover, the conditions on change magnitudes, that we will impose in Chapter 3, will be generally too restrictive for this panel data approach within a functional setting.)

Remark 1.4 (Simulations). We demonstrate the performance of the change point tests for functional data and of the estimates for change points in panel data by conducting simulations in R.[1] The utilized packages are indicated in the corresponding sections later on. (All R scripts can be obtained on request from the author.[2] Additionally, a small demonstration application for panel data is implemented in MATLAB which is also available from the author or can be downloaded from `www.mi.uni-koeln.de/~ltorgovi`.)

[1] R version 3.2.3 (Wooden Christmas-Tree, 2015-12-10).
[2] Email: ltorgovi@math.uni-koeln.de

1.3 Preliminaries on Hilbert space data and on dependence concepts

We start with a brief formal description of Hilbert space valued time series (including functional data) and then turn to the concepts of short range dependence that will be under consideration in this thesis.

Hilbert space data and functional time series

Let H be a real finite or infinite dimensional Hilbert space with an inner product $\langle x, y\rangle_H$ for $x, y \in H$ and with a corresponding norm $\|x\|_H = \langle x, x\rangle_H^{1/2}$.[1] Furthermore, let $\mathcal{B}_H$ be the associated Borel sigma-algebra. A Hilbert space time series $\{X_i\}_{i\in\mathbb{Z}}$ is formally a collection of random elements (i.e. measurable mappings)

$$X_i : (\Omega, \mathcal{A}, P) \to (H, \mathcal{B}_H) \tag{1.3.1}$$

defined on some common probability space $(\Omega, \mathcal{A}, P)$. (Thus, the X_i's are random variables but with a more abstract state space.) In this thesis we study finite and infinite dimensional Hilbert space valued time series simultaneously which cover the two following important situations: 1.) Multivariate $\mathbb{R}^d$-valued time series. 2.) Functional $L^2[0,1]$-valued time series. Since the latter functional time series is an abstract object, we provide a somewhat more intuitive interpretation in the next remark.

Remark 1.5 (Real-valued processes interpreted as functional data). Assume a sequence of real-valued random processes $\{\varepsilon_i(t),\ t \in [0,1]\}_{i\in\mathbb{Z}}$, where for each $i \in \mathbb{Z}$ we have a collection of random variables $\varepsilon_i(t),\ t \in [0,1]$ that are defined on some common probability space $(\Omega, \mathcal{A}, P)$. A condition which allows us to interpret these processes as an $L^2[0,1]$-valued time series is given, e.g., in Hsing & Eubank (2015, Theorem 7.4.1): for all $i \in \mathbb{Z}$ the sample paths $\varepsilon_i(\cdot)$ are in $L^2[0,1]$ and the variables $\varepsilon_i(t) = \varepsilon_i(t,\omega)$ are jointly measurable in (t,ω).

Remark 1.6 (A historic remark). The functional point-of-view gained popularity within the statistical community over the last two decades and a rich statistical theory has been developed. (Cf., e.g., the timeline in Section 1.1 of Cuevas, 2014.) A sometimes overlooked fact is that the foundations of functional data analysis lie rather far back in time. According to Wang (2015, p. 2): »*(...) the term "functional data analysis" was coined by Ramsay (1982) and Ramsay & Dalzell (1991), [but] the history of this area is much older and dates back to Grenander (1950) and Rao (1958)*«. Note also the pioneering works of the French school on probability theory in function spaces which are dating back to (roughly) the same time-period (e.g. Fortet & Mourier, 1955 and Mourier, 1956). Taking a more abstract position we can even go further back in time. As formulated by Mas (2008, p. 136): »*It turns out that probabilists have studied such random elements for*

[1] For the sake of readability, we will suppress the subscript H if the context allows us to do so.

a much longer time than statisticians (first works on the Brownian motion date back to the XIXth century), the first monograph dedicated to functional data was published in 1991«.

We proceed by describing two closely related frameworks of short range dependence for Hilbert space valued data which we will study subsequently in Chapter 2. The first concept is that of m-dependent time series which is then used to define the broad class of so-called L^κ-m-approximable time series.

m-dependent time series

As usual, m-dependence is defined via the pairwise independence of sufficiently distant sigma-algebras. We worked with this concept and with the following definition already in Torgovitski (2015a).

Definition 1.7 (m-dependence). Let H be a Hilbert space. An H-valued time series $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is called m-dependent if all sigma-algebras $\sigma\{\ldots,\varepsilon_{j-2},\varepsilon_{j-1},\varepsilon_j\}$ and $\sigma\{\varepsilon_k,\varepsilon_{k+1},\varepsilon_{k+2},\ldots\}$ that are separated by $k-j>m$ are independent.

The concept of m-dependence may be interpreted as a (first) straightforward deviation from the independence assumption where the latter, obviously, corresponds to $m=0$. As should be expected, there exists a rich literature dedicated to m-dependent time series that ranges from rather classical limit theorems of probability theory up to more specific statistical applications. (Cf., e.g., Hoeffding & Robbins (1948) for the former and, e.g., Horváth *et al.* (1999) or Choudhury *et al.* (1999) for the latter.) Hence, studying procedures under m-dependence we may rely on a generally well-developed machinery. Prominent statistical models for this dependence concept are linear MA(q) time series $(m=q)$ and examples of real-life data that can be reasonably approximated by MA(q), $q=1,2,3,$ series are given, e.g., in the textbook of Shumway & Stoffer (2011, cf. Examples 3.32, 3.38 and 3.40). The following quote by Choudhury *et al.* (1999, p. 347, partly cited in Gombay, 2010), stated in a regression context, underpins that studying MA(q) time series and thus the framework of m-dependence is of importance: *»Historically, the inability to use OLS to obtain parameter estimates for MA(q) and ARMA(p,q) models discouraged practitioners from using such models. (...) In fact, the computational ease with which AR error processes are handled is arguably the main reason why AR error models have become the paradigm for economic modeling (...), even though the MA error model frequently is more plausible (...).«*

Remark 1.8 (Estimating the order of MA(q) time series). Even though the dependence parameter q is rarely known exactly in practice some reasonable estimates for q can be identified, e.g., either ad-hoc by visual inspection of the *sample autocorrelation* together with the *partial autocorrelation* functions or more scientifically by fitting different MA($\hat{q}$) time series and choosing their order $\hat{q}$ via some common model-selection criteria such as, e.g., the AIC or the BIC. Examples of real-life applications where q is known exactly also exist as, e.g., time series that have been preprocessed by some averaging filter. (Cf. the related discussion in Torgovitski, 2015a.)

Remark 1.9 (Infinite dependence). Many popular time series models that are used in practice and extensively studied in theory (e.g., the already mentioned autoregressive models) depend on innovations from the infinite past or possibly also from the (infinite) future and thus do not fit into an m-dependent framework. Such processes are covered in Chapter 2 by the study of m-approximable time series which are introduced further below.

m-approximable time series

The concept of L^κ-m-approximability (or m-approximability for short) gained a lot of attention in recent literature. It was popularized - amongst many others - by Aue *et al.* (2009) and Hörmann & Kokoszka (2010) where the former considered a multivariate and the latter a functional setting. (Cf. also Chochola *et al.*, 2013, Jirak, 2013 and Hörmann & Kidziński, 2015.) An appealing aspect of this approach is that it covers many relevant strictly stationary models including, e.g., ARMA and ARCH time series. Before we state the precise definition of this concept, we recall that for an H-valued random element X (cf. (1.3.1)), where H is a Hilbert space, the $L^\kappa(\Omega, P)$-norms are defined via

$$\nu_\kappa(X) = [E\|X\|_H^\kappa]^{1/\kappa} \tag{1.3.2}$$

for any $\kappa \geq 1$. Loosely speaking, L^κ-m-approximable time series are L^κ-limits of suitable m-dependent approximations. We already worked with this particular formulation in Torgovitski (2016, Definition 2.1) and also used it in Torgovitski (2015c).

Definition 1.10 (m-approximability). Let H be a Hilbert space. An H-valued time series $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is L^κ-m-approximable with an approximation rate function $\delta\colon \mathbb{N}_0 \to \mathbb{R}_+$, if $E\|\varepsilon_i\|_H^\kappa < \infty$, for some $\kappa \geq 1$, and, if the following conditions hold true:

1. *Bernoulli shift representation:* It holds that $\varepsilon_i = f(\ldots, \eta_{i+2}, \eta_{i+1}, \eta_i, \eta_{i-1}, \eta_{i-2}, \ldots)$, where $f\colon S^{\mathbb{Z}} \to H$ is a measurable mapping from some measurable space S and $\{\eta_i\}_{i\in\mathbb{Z}}$ is a sequence of i.i.d. S-valued random elements.

2. *L^κ-approximability via an m-dependent coupling:* Let m-dependent copies of ε_i be defined for all $m \in \mathbb{N}_0$, $m' = \lfloor m/2 \rfloor$ by

$$\begin{aligned}\varepsilon_i^{(m)} = f(\ldots, \eta_{i+m'+1}^{(m,i)}, \eta_{i+m'}^{(m,i)}, \eta_{i+m'-1}, \ldots, \eta_i,\\ \ldots, \eta_{i-m'+1}, \eta_{i-m'}^{(m,i)}, \eta_{i-m'-1}^{(m,i)}, \ldots),\end{aligned} \tag{1.3.3}$$

using a family $\{\eta_r, \eta_i^{(k,j)},\ i,j,r,k \in \mathbb{Z}\}$ of i.i.d. random variables.[1] Using definition (1.3.2) it holds that

$$\nu_\kappa\big(\varepsilon_0 - \varepsilon_0^{(m)}\big) \leq c\delta(m) \tag{1.3.4}$$

for all $m \in \mathbb{N}_0$, $\delta(0) = 1$, some $c > 0$ and with $\sum_{m=0}^\infty \delta(m) < \infty$.

[1] We formally set $\varepsilon_i^{(m)} = f(\ldots, \eta_{i+1}^{(m,i)}, \eta_i^{(m,i)}, \eta_{i-1}^{(m,i)}, \ldots)$ for $m = 0, 1$, i.e. for $m' = 0$.

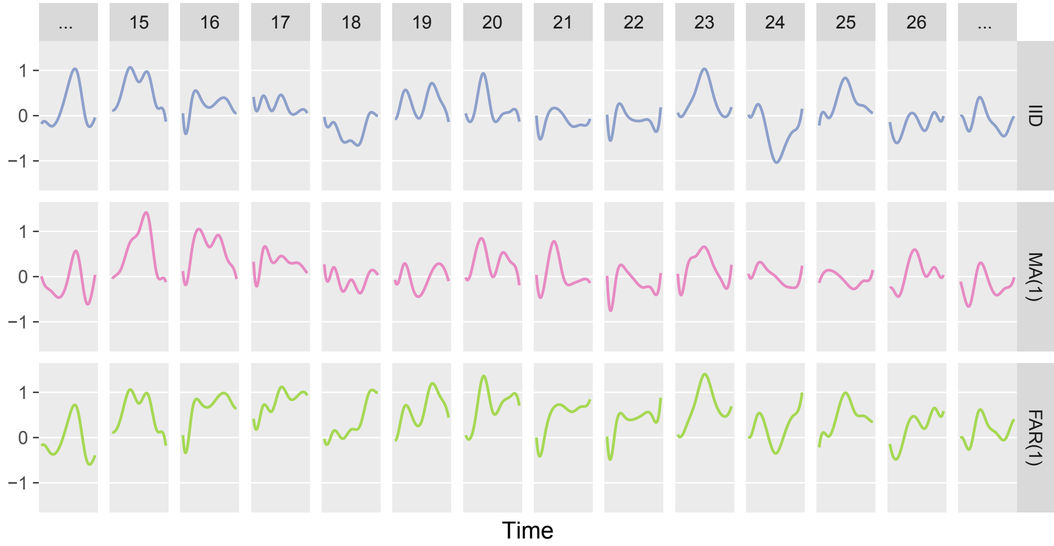

Figure 1.4: This figure gives an impression of $L^2[0,1]$-valued time series that are m-dependent or m-approximable. The top row shows a functional i.i.d. innovation sequence. The panel in the middle shows (1-dependent) observations of an MA(1) time series which are generated as averages of the i.i.d. sequence from the top panel. Finally, the bottom panel shows an m-approximable functional AR(1) time series generated from the same innovation sequence, using a Wiener kernel (cf., e.g., Horváth & Kokoszka, 2012).

Remark 1.11 (Typical rates of decay). The decay rate $\delta(m)$, $m \in \mathbb{N}$, quantifies the intensity of weak dependence via (1.3.4) and is directly related to the decay of lagged correlations. The latter is the underlying reason why many widely used functional and multivariate time series models do have an exponential rate of decay, i.e. $\delta(m) = \exp(-\nu m)$ with some $\nu > 0$ (cf., e.g., Aue *et al.*, 2009, Horváth & Kokoszka, 2012 and Hörmann *et al.*, 2013). In this thesis we will work (mostly) with a *polynomial* rate $\delta(m) = m^{-\nu}$ for some $\nu > 1$.

We finish this chapter by two remarks, which, on the one hand, explain the relation of m-approximability to other dependence concepts that are commonly found in literature and, on the other hand, indicate why we consider a non-causal formulation in Definition 1.7 and in Definition 1.10.

Remark 1.12 (Related weak dependence concepts). It is important to note that the ideas behind the concept of L^κ-m-approximability may be found in a variety of related approaches like the famous concept of *near-epoch dependent* (NED) time series that goes back to the works of Billingsley and McLeish in the 1960's and 1970's (cf., e.g., Ling, 2007), or to that of L_p-*approximability* of Pötscher & Prucha (1997) (given that the underlying and defining so-termed *basis series* are i.i.d. in both cases).[1] It is also closely related to the concept of *»physical dependence (measure)«* of Wu (2011). Contrary to the general near-epoch dependent approach, one key idea of L^κ-m-approximability and of

[1] The motivation of the L^κ-m-approximability concept, its relation to NED and (some of) its advantages over mixing type concepts are discussed in detail in Hörmann (2009), Hörmann & Kokoszka (2010) and in Berkes *et al.* (2011). For further information on Bernoulli shifts and on weak dependence concepts we refer to Doukhan *et al.* (2003) and to Dedecker *et al.* (2007).

the approach of Wu (2011) is to restrict the framework to i.i.d. basis series, i.e. to shifts of an i.i.d. sequence $\{\eta_i\}_{i\in\mathbb{Z}}$, which allows for an explicit and convenient construction of m-dependent approximating sequences (1.3.3). Note that some authors treat L^κ-m-approximability and the physical dependence concept of Wu (2011) simultaneously as *»filters«* (cf., e.g., Jirak, 2013). (But even though they share the same idea it is important to point out that within the concept of Wu (2011) the approximating sequences are constructed differently replacing only one innovation at a time instead of replacing an infinite number of innovations simultaneously.)

Remark 1.13 (Causal and non-causal time series). The consideration of m-approximable time series that are based on two-sided shifts (as in our situation of Definition 1.7 and Definition 1.10) is generally by no means just of purely theoretical interest. Davis & Wu (2010, p. 98) highlight that *»[non-causal time series] arise frequently in the modeling of real data«* and they also provide a real-life example. However, the motivation to study non-causal series in this thesis is different: Some of the presented proof techniques in Chapter 2 involve *time-reversion* of time series and thus non-causal shifts emerge anyway even if we worked (only) under causality, i.e. have

$$\varepsilon_i = f(\eta_i, \eta_{i-1}, \eta_{i-2}, \ldots) \tag{1.3.5}$$

instead of the two-sided shift in Definition 1.10.

Note that within the m-approximable framework we will rely on literature where often only a causal representation is assumed for analogues of Definition 1.10. To this end Hörmann & Kokoszka (2010, p. 1851) pointed out that *»(...) only a straightforward modification is necessary in order to generalize the theory (...) to non-causal processes (...)«*. Hence, wherever we could apply some (straightforward) modifications we state our results under the general non-causal setting (cf., e.g., Remarks 2.37, 2.64 and Remark 2.66). In a few more complicated cases we restrict ourselves to the often assumed one-sided shifts (cf., e.g., Theorem 2.45 and Theorem 2.51).

Structure of the thesis 1.4

This thesis is subdivided into this introductory Chapter 1 and two subsequent main Chapters 2 and 3. Chapter 2 is on testing in Hilbert space data and Chapter 3 is on estimation in Hilbert space panel data. The proofs are always postponed to the end of each chapter. (Note that Section 1.3 introduces Hilbert space data and some dependence concepts for time series. Furthermore, for convenience of the reader, we summarize the most important notation at the beginning of this thesis on page xvii.)

Relation of this thesis to previous publications (and preprints)

In this thesis we present the results that were obtained during the years as a postgraduate student at the University of Cologne in the working group of Prof. J. G. Steinebach. We combine and generalize the results of our »*Teilpublikationen*«[1] Torgovitski (2015a,b,c,d, 2016) and embed them in an overall context in a Hilbert space time series framework. The relations of the theoretical results of Chapter 2 to Torgovitski (2015a,c, 2016) and of Chapter 3 to Torgovitski (2015b,d) are explained at the beginning of both chapters and complemented by detailed references at the end of both chapters in Section 2.6 and in Section 3.6, respectively. (Note that additional references are provided at the beginning of the proofs.)

[1] Preprints, submitted articles and published articles.

CHAPTER 2

CHANGE POINT TESTING FOR HILBERT SPACE VALUED DATA

> "*An increasing number of applications from biology to image sequences in medical imaging involve data that can be well represented as functional time series. This has led to a rapid progression of theory associated with functional data.*
>
> — Aston & Kirch (2012b, p. 1906)

Introduction 2.1

In this chapter we focus on testing for changes in means in a non-parametric, large sample size framework for high-dimensional Hilbert space valued data.[1] Change point analysis for high-dimensional settings developed rapidly over the last decade and a lot of literature, especially in the infinite dimensional functional setting, focuses on projection based CUSUM-type procedures using principal components (see Remark 2.3, below). We aim to extend, refine and combine our own results on projection based CUSUM procedures from Torgovitski (2015a,c, 2016) towards a more general, unifying framework.[2] The theory is developed for Darling-Erdős-type procedures within a multi-directional change framework in a general Hilbert space setting including the infinite dimensional and the finite dimensional cases. We derive asymptotic distributions of Darling-Erdős-type test statistics under the null hypothesis of no change in the mean and conditions for consistency under the alternative. Moreover, we use the tensor-based CUSUM formulation of the recent article Torgovitski (2015c) and verify that change alignment is advantageous for Darling-Erdős-type procedures as well.[3] The simulations in Section 2.4 complement those in Torgovitski (2015a,c, 2016). Finally, note that the

[1] In the sense of Remark 1.1.

[2] Note that the combined and extended results are shown by the same proof techniques which therefore partly differ from either of the proof techniques used in Torgovitski (2015a,c, 2016).

[3] Note that we looked at a differently weighted CUSUM procedure in Torgovitski (2015c).

main theoretical part of this chapter is contained in Section 2.3 and subdivided into a Subsection 2.3.2 on long run covariance and principal component estimation and a Subsection 2.3.3 on testing based on those estimates.

For convenience of the reader, we explain some main differences and similarities of Torgovitski (2015a,c, 2016). The results of Torgovitski (2015a, 2016) are originally both stated for Darling-Erdős-type CUSUM procedures in the $L^2[0,1]$-valued functional framework under a one-directional change setting, working with long run principal components. Despite these similarities there are the following differences: the former article considers a gradual (piecewise linear) change scenario under m-dependence whereas the latter article considers an abrupt change scenario under the more sophisticated dependence concept of m-approximability. In Torgovitski (2015c) we look at a differently weighted CUSUM procedure within an abstract, infinite dimensional Hilbert space framework and (essentially) in a one-directional gradual change setting which is more general than that of Torgovitski (2015a, 2016). As in Torgovitski (2016) we consider m-approximability and indicate extensions to the multi-directional change alternative but without rigorous proofs. Overall, the setting of Torgovitski (2015c) is closest to this thesis but, as already mentioned, it considers a different test statistic and does not include the multivariate case.

It is also worth mentioning that in Torgovitski (2015a, 2016) we focus on adaptations of techniques that were already available in the change point respectively two-sample testing context for related (but different) settings and test statistics.[1] In Torgovitski (2015c) we suggest new methodology. For instance, we propose to work with change-aligned projections which simplify conditions on detectability of changes and that allow us to detect changes which are less likely to be captured by the leading principal components. (Detectability particularly increases when working with static principal components. The latter are appealing due to optimality properties, as will be discussed in Remark 2.19, below.) These findings are to the best of our knowledge not only new in the infinite dimensional setting but are also new in the multivariate context. Torgovitski (2015c) additionally contains a new, already mentioned, tensor-based CUSUM formulation which allows for a more direct proof technique than in the previous articles Torgovitski (2015a, 2016) and enables us to avoid conditions on eigenvalue separation as well as some technicalities in the proofs.[2]

More details on the relation between the results of this chapter and the results in Torgovitski (2015a,c, 2016) will be provided at the end of this chapter in Section 2.6.

Remark 2.1 (Dependence assumptions). In this chapter we will distinguish between the assumptions of m-dependence and m-approximability. The reasons for doing so are manifold: we already mentioned that the results of this chapter under m-dependence (see Assumption M1, below) are derived for more restricted linear alternatives in Torgovitski (2015a) whereas the results under m-approximability (see Assumption M2, below) are partly contained in Torgovitski (2015c, 2016). As we will see in Subsection 2.3.2 and Subsection 2.3.3, our procedures under m-dependence allow us the use of (slightly)

[1] See our Remark 2.3, below, and cf., e.g., Berkes *et al.* (2009), Hörmann & Kokoszka (2010), Aston & Kirch (2012a), Horváth *et al.* (2013) and Horváth *et al.* (2014).

[2] This contribution is essentially based on a recent result by Reimherr (2015).

different long run covariance estimates and long run principal component estimates than the corresponding procedures under m-approximability. Also, the asymptotics for change point testing rely on different versions of strong approximations for both situations. Hence, the theoretical results in both subsections will be formulated side-by-side for both concepts to illustrate similarities and differences. The results on m-approximability do have a broader applicability. However, the assumption of m-dependence yields sharper results for long run covariance estimation under the null hypothesis (as should be expected) and also a different asymptotic stabilization behavior of these estimates under the alternative. Moreover, it provides *»Brownian bridge type«* approximations for the test statistic which are beyond the scope of this thesis under the assumption of m-approximability.

Remark 2.2. To the best of our knowledge Torgovitski (2015a) is the first contribution to the functional setting where non-parametric Darling-Erdős-type CUSUM-procedures are studied, where piecewise linear change alternatives are considered and where less than fourth moments are assumed.[1] It is also the first contribution under temporal dependence where long run functional principal components, as suggested by Horváth *et al.* (2013) in a two-sample testing setting, are incorporated into functional CUSUM-type procedures. These studies are continued by Torgovitski (2016, arXiv:1407.3625v1), which, again to the best of our knowledge, is the first contribution which considers Darling-Erdős-type CUSUM-procedures under the concept of m-approximability. Finally, note that in Torgovitski (2015c) we introduce change-aligned principal components and a subspace based formulation of dimension reduction based CUSUM statistics.

Remark 2.3. For change point problems in the specific Hilbert space $L^2[0,1]$ the basis given by the (empirical) functional principal components was considered by several authors for related CUSUM-type statistics. Under temporal independence it was used by Berkes *et al.* (2009) for the weighting w_0 and under Gaussianity by Zhou (2011) for our Darling-Erdős-type weighting $w_{1/2}$ (cf. (1.2.1) for the definitions of the weights). Later on, this basis was used by Hörmann & Kokoszka (2010) and by Aston & Kirch (2012a), again for the weighting w_0, under temporal dependence within the concept of m-approximability, which is the concept that we are also working with. As already mentioned, Horváth *et al.* (2013) considered long run principal components for two-sample testing under m-approximability and Horváth *et al.* (2014) extended this approach to change point testing (for CUSUM statistics with the weighting w_0) under similar dependence assumptions.

Consistency of Bartlett-type long run covariance estimates and of the corresponding long run principal components has been proven for functional m-approximable time series by Horváth *et al.* (2013) in the $L^2[0,1]$ setting. Note that convergence rates in m-approximable functional and Hilbert space settings were meanwhile also obtained by, e.g., Horváth *et al.* (2014), Hörmann *et al.* (2015) and Berkes *et al.* (2016). In subsequent sections we will contribute to the existing theory and obtain polynomial convergence rates under rather mild assumptions. Our technique for the proof may be (retrospectively) interpreted as a combination of the aforementioned literature. Finally, note that logarithmic convergence rates of long run covariance estimation are mandatory for asymptotics of Darling-Erdős-type CUSUM procedures.

[1] For an abrupt-change setting under Gaussianity we refer, again, to Zhou (2011).

Structure of the chapter

The structure of this chapter is as follows. We begin by introducing our signal plus noise model and the change point testing problem in Section 2.2. In Section 2.3 we explain the testing approach in general Hilbert spaces and provide some necessary preliminaries on dimension reduction and on (long run or static) principal components in Subsection 2.3.1. We present our theoretical results in Subsection 2.3.2 and in Subsection 2.3.3: the former one addresses estimation of the long run principal components and the latter one focuses on asymptotics for CUSUM tests based on those estimates and additionally on change-aligned extensions. All proofs are postponed to Section 2.5 which is subdivided into three parts: Subsection 2.5.1 treats, again, the estimation of the principal components, Subsection 2.5.2 provides the proofs for CUSUM tests based on those estimated principal components and Subsection 2.5.3 contains some auxiliary results. (Following a mathematical tradition we close this chapter with some notes in Section 2.6 that show the relation of this chapter to our own previous publications and preprints.[1])

[1] Cf., e.g., Csörgő & Horváth (1997), Bosq (2000) or Tavakoli (2014).

A change point problem 2.2

Assumption 2.4. Unless stated otherwise, throughout Chapter 2 we will assume $\{Y_i\}_{i\in\mathbb{Z}}$ to be an observable time series of H-valued random elements where H is a separable, real Hilbert space with dimension $d_H \in \mathbb{N} \cup \{\infty\}$, $d_H > 1$. We use the notation $\mathbb{N}_H = \{1, \dots, d_H - 1\}$ for the finite dimensional case $d_H < \infty$ and stick to the convention $\mathbb{N}_H = \mathbb{N}$ for the infinite dimensional case $d_H = \infty$.[1]

Signal plus noise model 2.2.1

We assume that the time series $\{Y_i\}_{i\in\mathbb{Z}}$ follows a Hilbert space signal plus noise model

$$Y_i = m_i + \varepsilon_i, \tag{2.2.1}$$

for $i \in \mathbb{Z}$, where $m_i \in H$ are the deterministic Hilbert space signals and $\varepsilon_i \in H$ are random Hilbert space noises that shall fulfill Assumption S1 and either one of the Assumptions M1 or M2, given below. Throughout this chapter we assume stationarity of $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ and at least $E\|\varepsilon_1\| < \infty$ in which case the expectation $E\varepsilon_1$ is well-defined and characterized as the unique solution of $E\langle\varepsilon_1, v\rangle = \langle E\varepsilon_1, v\rangle$ for all $v \in H$.[2]

Assumption S1. The sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ is H-valued centered, strictly stationary and $E\|\varepsilon_1\|^\kappa < \infty$ holds true for some $\kappa > 2$.

Assumption M1. The sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ fulfills Assumption S1 and is m-dependent.

Assumption M2. The sequence $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ fulfills Assumption S1 and is m-approximable.

We continue with a remark on the relation between Assumptions S1 and M2 as well as by clarifying the interaction between Assumptions M1 and M2.

Remark 2.5 (Parallel structure). For the sake of a parallel structure in this chapter, we accept some redundancy in above Assumptions S1 and M2 which is that an L^κ-m-approximable series is always strictly stationary and always fulfills a moment condition.

It is of some interest to discuss briefly whether Assumption M1 is an implication of Assumption M2 or only intersects with the latter. The relation of these two concepts is somewhat surprisingly not entirely transparent and some facts that can be found in literature are explained in the remark below.

[1] We consider $\mathbb{N}_H = \{1, \dots, d_H - 1\}$ instead of $\mathbb{N}_H = \{1, \dots, d_H\}$, for $1 < d_H < \infty$, only for notational reasons: we need to ensure that λ_{d+1} is well-defined in Assumption E3 (below) within a multivariate setting. All considerations are also restricted to $d_H \geq 2$ to exclude the case $\mathbb{N}_H = \emptyset$ for $d_H = 1$.

[2] See also the discussion in the Section 3.1 of Cuevas (2014) for more general definitions of expectations in Banach spaces and for a convenient pointwise definition for expectations of random functions. Moreover, cf. Section 4.1 of Cuevas (2014) for a further intuitive distance-based interpretation.

Remark 2.6 (Relation of m-dependence and m-approximability).

1. Strictly stationary m-dependent series form evidently the basis of m-approximability. Thus, constructing a series that fulfills both Assumptions M1 and M2 is straightforward. A series that satisfies the first assumption of Definition 1.10 with representation

$$\varepsilon_i = f(\eta_{i+k}, \ldots, \eta_i, \ldots, \eta_{i-k}), \tag{2.2.2}$$

$k \in \mathbb{N}$, is $2k$-dependent and thus given that $E\|\varepsilon_0\|^\kappa < \infty$ also fulfills (1.3.4) with the trivial rate function $\delta(m) = 0$ for $m \geq 2(k+1)$.

2. Constructing a series that only fulfills Assumption M1 is far less trivial. In Berkes *et al.* (2011) it is claimed that stationary m-dependent processes without a shift representation exist and thus in this case Assumption M1 is fulfilled whereas Assumption M2 is not. It is likely a correct statement but (nevertheless) we were not able to find an explicit and rigorous proof for this claim (at least) within the referenced literature.[1]

3. It is certainly true that not all strictly stationary m-dependent series do have a truncated shift representation (2.2.2) with an i.i.d. sequence $\{\eta_i\}_{i\in\mathbb{Z}}$. For instance, Burton *et al.* (1993) provide a counterexample that does not have a finite shift representation as given in (2.2.2) for *any* finite $k \in \mathbb{N}$. Such series can only be (if at all possible) m-approximable with an infinite shift representation (in which case the determination of a decay rate $\delta(m)$ would remain an additional issue).

The testing problem 2.2.2

Our aim is to test the null hypothesis of no change in the Hilbert space means

$$H_0 : \; m_1 = \ldots = m_n, \tag{2.2.3}$$

$m_i \in H$, against the multi-directional change alternative of

$$H_A : \; m_i = m + \sum_{j=1}^{\varrho} g_j(i/n)\varDelta_j, \tag{2.2.4}$$

for $1 \leq i \leq n$, $n \in \mathbb{N}$, $\varrho \in \mathbb{N}$ and $m \in H$.[2] The $\varDelta_j \in H$ are orthonormal *change-directions* and the g_j are piecewise Lipschitz continuous *trend-functions* on $[0,1]$ for which we assume $g_j(0) = 0$ for all $1 \leq j \leq \varrho$.[3] Furthermore, we assume $g_j(t_j) \neq 0$ for some $t_j \in (0,1)$, again, for all $1 \leq j \leq \varrho$. Our conditions on g_j's imply that $g_j(1/n) = g_j(2/n) = \ldots = g_j((n-1)/n) = g_j(n/n)$ cannot hold true for each $1 \leq j \leq \varrho$

[1] Note that in Berkes *et al.* (2011) the stationarity of an m-dependent series is not explicitly required. It seems to be implicitly assumed since otherwise counterexamples are trivial.
[2] cf. Section 6 in Torgovitski (2015c).
[3] The domain of the g_j's can be partitioned into a finite number of non-empty intervals on which the functions are Lipschitz continuous.

and all $n > n_0$ with some $n_0 \in \mathbb{N}$. This ensures that H_A does not coincide with H_0 for $\varrho > 0$ if $n > n_0$. (Note that the case $\varrho = 0$ formally corresponds to H_0.) Overall, the multi-directional change term $\sum_{j=1}^{\varrho} g_j(t)\Delta_j$ quantifies the deviation from m and thus from the null-hypothesis in time via g_j and in space via Δ_j (cf. Figure 1.1). Finally, note that we assume m, Δ_j, g_j and ϱ to be unknown.

Remark 2.7. The alternative (2.2.4) may be interpreted in the multivariate setting as a multivariate regression with fixed design and thus as a special case of Gombay (2010). (For related high-dimensional settings we refer to Aston & Kirch, 2014.)

Remark 2.8 (Popular change settings). Horváth *et al.* (2014) considered a one-directional piecewise Lipschitz continuous alternative (2.2.4) in the functional setting for CUSUM tests which are related to the tests considered in Section 2.3, below. Furthermore, the following popular special cases $g_1 \in \{a \cdot g_A, a \cdot g_E, a \cdot g_L\}$, $a \in \mathbb{R}\backslash\{0\}$, $\varrho = 1$ were studied for analogous tests in one-directional functional frameworks previously:

1. *abrupt change* $\quad g_A(x;\theta_1) = \mathbb{1}_{(\theta_1,1]}(x)$,
2. *epidemic change* $\quad g_E(x;\theta_1,\theta_2) = \mathbb{1}_{(\theta_1,\theta_2]}(x)$,
3. *piecewise linear change* $\quad g_L(x;\theta_1,\theta_2) = \mathbb{1}_{(\theta_1,\theta_2]}(x)h(x) + \mathbb{1}_{(\theta_2,1]}(x)$,

with $h(x) = (x-\theta_1)/(\theta_2-\theta_1)$, $x \in [0,1]$, and where $0 < \theta_1 < \theta_2 \le 1$. The case of an abrupt alternative $g_A(x;\theta_1)$ was considered first in the functional setting by Berkes *et al.* (2009), the epidemic change case $g_E(x;\theta_1,\theta_2)$ by Aston & Kirch (2012a) and the piecewise linear alternative case $g_L(x;\theta_1,\theta_2)$ by Torgovitski (2015a)[1].

Remark 2.9 (Abrupt change setting). In the subsequent Section 2.3 we will consider a statistic that is tailored to test H_0 against the single abrupt change in the mean alternative

$$H_A^{\text{abrupt}}: \; m_1 = \ldots = m_u \neq m_{u+1} = \ldots = m_n, \tag{2.2.5}$$

where the position u is unknown. For asymptotic theory u is typically assumed to be proportional to the sample size, i.e. $u = \lfloor n\theta \rfloor$ for some (again unknown) parameter $\theta \in (0,1)$.[2] Under this assumption the latter case of $u = \lfloor n\theta \rfloor$ corresponds to (2.2.4) with $\varrho = 1$ and with the trend function $g_1(x) = g_A(x;\theta)$ from Remark 2.8. Consideration of a broader class of changes in (2.2.4) is important to reassure that our CUSUM test in Section 2.3, below, is sensitive to more realistic deviations from H_0 than the single abrupt change (2.2.5), e.g. to multiple and possibly smooth changes in various directions.[3]

[1] The piecewise linear setting results of Torgovitski (2015a) with $g = a \cdot g_L$, $a \in \mathbb{R}\backslash\{0\}$ were submitted in March 2013 and published in March 2014 whereas Horváth *et al.* (2014) was published online later in June 2014.

[2] As indicated in the introductory Section 1.1 this is a widely accepted and probably most studied type of alternatives in the change point context.

[3] For related results on such robustness for the one-directional case of $\varrho = 1$ we refer to Horváth *et al.* (1999, 2014), Aston & Kirch (2012a) and to Torgovitski (2015a).

Remark 2.10 (Further extensions of our change setting). Our results on change-aligned testing (cf. p. 56, below) can be theoretically extended to a more flexible ∞-directional setting, where $\varrho = \infty$ in (2.2.4) and where g_j fulfill some additional assumptions. However, this case is beyond the scope of this thesis and will be studied elsewhere.

Testing for change points with weighted CUSUM procedures (using principal components) 2.3

We begin this section by introducing the test statistic together with a preliminary discussion of a *dimension reduction* technique that is based on (long run) principal components, i.e. on the Karhunen-Loève expansion. Note that according to Panaretos & Tavakoli (2013, p. 2780): »*The Karhunen-Loève expansion representation has become both the object of and the means for much of the statistical methodology developed for functional data. It has defined what today is accepted as the canonical framework for functional data analysis and has provided a bridge allowing for a technology transfer of tools from multivariate statistics to problems of functional statistics.*«

Preliminaries on dimension reduction (and on long run principal components) 2.3.1

Assumption 2.11 (on weak dependence). Throughout Subsection 2.3.1 we will tacitly assume the signal plus noise model (2.2.1) and that either Assumption M1 or Assumption M2 on weak dependence holds true. Both assumptions imply that $E\|\varepsilon_0 \otimes \varepsilon_0\|_{\mathcal{S}} = E\|\varepsilon_0\|^2 < \infty$ and that $\sum_{r\in\mathbb{Z}} \|E[\varepsilon_0 \otimes \varepsilon_r]\|_{\mathcal{S}} < \infty$ hold true.[1]

Given any orthonormal basis $\{v_j\}$ of H we can always represent each Y_i in a *generalized Fourier expansion*

$$Y_i = \sum_{j=1}^{d_H} \langle Y_i, v_j \rangle v_j, \tag{2.3.1}$$

$d_H \in \mathbb{N}_H$ (cf. Assumption 2.4), where in the infinite dimensional case the convergence of these series is meant with respect to the underlying norm $\|\cdot\|_H$. To obtain a lower finite dimensional approximation of Y_i the usual practice is to truncate the expansion (2.3.1) by $Y_i \approx \sum_{j=1}^{d} \langle Y_i, v_j \rangle v_j$ with some finite $d \in \mathbb{N}_H$, which is an orthogonal projection on the subspace spanned by $\{v_1, \dots, v_d\}$. Note that all information on the Y_i's, with respect to the chosen subspace, is now captured by the multivariate *projections* $\boldsymbol{y}_i = [\langle Y_i, v_1 \rangle, \dots, \langle Y_i, v_d \rangle]'$, which consist of the so-called *scores* $\langle Y_i, v_j \rangle$ (cf. the scheme in Table 2.1, below).

A usual practice for high-dimensional data (either it is finite or infinite dimensional), is to apply classical multivariate analysis techniques directly to the lower dimensional projected time series $\{\boldsymbol{y}_i\}_{i\in\mathbb{Z}}$. For our testing problem, which we introduced in the previous Section 2.2, we will base the analysis on the multivariate weighted CUSUM

[1] Cf. Hörmann *et al.* (2015) (and also Horváth *et al.*, 2013).

statistic introduced in (1.1.3) and given by

$$\mathscr{M}_n(\boldsymbol{y}) = \max_{1\le k<n} w(k/n)\mathscr{T}(k/n) \tag{2.3.2}$$

with a detector

$$\mathscr{T}(x) = \mathscr{T}(x;\boldsymbol{y}) = |\Sigma^{-1/2}S_n(x;\boldsymbol{y})| = |S_n(x;\boldsymbol{y})|_\Sigma. \tag{2.3.3}$$

Throughout this chapter $w(x) = w_{1/2}(x) = [x(1-x)]^{-1/2}$, $x \in (0,1)$, is the specific Darling-Erdős-type weighting function which we introduced in (1.1.7),

$$\Sigma = \sum_{r\in\mathbb{Z}} \mathrm{Cov}(\boldsymbol{y}_0, \boldsymbol{y}_r) \tag{2.3.4}$$

is the long run covariance matrix of $\{\boldsymbol{y}_i\}_{i\in\mathbb{Z}}$ with $\mathrm{Cov}(\boldsymbol{y}_0,\boldsymbol{y}_r) = E[(\boldsymbol{y}_0 - E\boldsymbol{y}_0)(\boldsymbol{y}_r - E\boldsymbol{y}_r)']$ and as before $S_n(x;\boldsymbol{y}) = \sum_{i=1}^{\lfloor nx\rfloor}(\boldsymbol{y}_i - \bar{\boldsymbol{y}}_n)/n^{1/2}$. Finally, $|\cdot|$ is the Euclidean norm and $|\cdot|_\Sigma = |\Sigma^{-1/2}\cdot|$. This statistic is designed to test for an abrupt change in the mean but is known to detect some broader classes of alternatives with still reasonable power (cf., e.g., Horváth *et al.* (1999) in a finite dimensional setting). Our aim is to study the asymptotics for (2.3.2) under H_0 and under H_A and simultaneously in infinite and finite dimensional Hilbert spaces with respect to the alternative of multi-directional changes (2.2.4). Obviously, the choice of the appropriate basis functions v_j and of the dimension d is crucial and needs to be done carefully to provide accurate low-dimensional approximations which capture sufficient *»relevant«* information of the observations Y_i. In our case it is certainly desirable if the basis functions are approximately aligned - and at least not all orthogonal - with the change in the mean directions Δ_j. (We will get back to this issue in Remarks 2.47, 2.48 and 2.49, below.)

Remark 2.12 (on principal components). A recommendation for the choice of v_j's that is found in a vast amount of literature on general high-dimensional (particularly functional) data analysis is to rely on the basis given by the data-driven principal components. Those are appealing due to the well-known optimality properties which will be discussed briefly in Remark 2.19 further below. (Note that we will follow the terminology of Hörmann *et al.* (2015) and call this basis the *»static principal components«* to be able to distinguish between other related bases.)

Subsequently, we will define and discuss the basis of static principal components and the basis of *»long run principal components«* that takes additionally autocorrelation into account. The former choice is well-established in change point literature on functional data whereas the latter approach turns out to be mathematically more convenient for our purposes and becomes quite popular in change point literature as well (cf. Remark 2.18, below).[1]

[1] We will also briefly mention the *»dynamic principal components«* in Remark 2.20, below. However, they are not in the scope of this thesis.

Static and long run principal components

Assumption 2.13 (on the spectral decomposition). Let $\mathcal{H}$ be any self-adjoint Hilbert-Schmidt operator on a separable Hilbert space H. Then, $\mathcal{H}$ possesses a *»spectral decomposition«*

$$\mathcal{H} = \sum_{j=1}^{d_H} \gamma_j (e_j \otimes e_j)$$

into real-valued eigenvalues γ_j and rank one operators $(e_j \otimes e_j)$. The eigenvectors e_j form an orthonormal basis of H. We call (γ_j, e_j) the *»eigenelements«*[1] and always assume $|\gamma_i| \geq |\gamma_j|$ for all $i > j$.

Definition 2.14 (Static principal components). The static principal components of an H-valued random element Y_0 are given by the eigenelements (λ_j, v_j) of the covariance operator[2]

$$\mathscr{C}_0 = E[\varepsilon_0 \otimes \varepsilon_0]. \tag{2.3.5}$$

It is a positive, self-adjoint Hilbert-Schmidt operator with a spectral decomposition $\mathscr{C}_0 = \sum_{j=1}^{d_H} \lambda_j (v_j \otimes v_j)$. (Cf. Remark 2.16, below.)

Definition 2.15 (Long run principal components). The long run principal components of an H-valued time series $\{Y_i\}_{i \in \mathbb{Z}}$ are given by the eigenelements (λ_j, v_j) of the long run covariance operator

$$\mathscr{C} = \sum_{r \in \mathbb{Z}} \mathscr{C}_r, \qquad \mathscr{C}_r = E[\varepsilon_0 \otimes \varepsilon_r], \tag{2.3.6}$$

which is also a positive, self-adjoint Hilbert-Schmidt operator and thus has a spectral decomposition $\mathscr{C} = \sum_{j=1}^{d_H} \lambda_j (v_j \otimes v_j)$. (Again, see Remark 2.16, below.)

Remark 2.16 (Properties of the covariance operators). We consider the covariance operator (2.3.5) first. Due to $E\|\varepsilon_0 \otimes \varepsilon_0\|_{\mathcal{S}} < \infty$ it is well defined and via

$$\langle \mathscr{C}_0(x), y \rangle = E\langle \varepsilon_0, y \rangle \langle \varepsilon_0, x \rangle, \tag{2.3.7}$$

for any x, $y \in H$, it is straightforward to see that it is self-adjoint and positive. Furthermore, (2.3.7) yields $E\langle \varepsilon_0, v_i \rangle \langle \varepsilon_0, v_j \rangle = \delta_{i,j} \lambda_i$. Now, we turn to the long run covariance operator (2.3.6). In view of $\sum_{r \in \mathbb{Z}} \|E[\varepsilon_0 \otimes \varepsilon_r]\|_{\mathcal{S}} < \infty$ it is well defined, too. To see that (2.3.6) is self-adjoint with non-negative eigenvalues λ_j it is sufficient to observe the identity

$$\langle \mathscr{C}(x), y \rangle = \sum_{r \in \mathbb{Z}} E\langle \varepsilon_0, y \rangle \langle \varepsilon_r, x \rangle, \tag{2.3.8}$$

[1] This notation and terminology for eigenvalues and eigenvectors is used, e.g., in Bosq (2000).
[2] Note that in the multivariate case of $H = \mathbb{R}^d$ (and using the Euclidean inner product) this operator corresponds to the mapping $\mathscr{C}_0(x) = E[\langle \varepsilon_0, x \rangle \varepsilon_0] = E[\varepsilon_0 \varepsilon_0'] x = \mathrm{Cov}(\varepsilon_0) x$.

which can be easily verified for any $x, y \in H$.[1] Using (2.3.8) we see that $\langle \mathscr{C}(x), x\rangle \geq 0$ holds true for all $x \in H$ since it is the long run variance of the projected series $\{\langle \varepsilon_r, x\rangle\}_{r\in\mathbb{Z}}$. Moreover, by stationarity, we obtain $\langle \mathscr{C}(x), y\rangle = \langle \mathscr{C}(y), x\rangle$ for any x, $y \in H$ and this verifies that $\mathscr{C}$ is self-adjoint. Finally, $\langle \mathscr{C}(v_i), v_j\rangle = \delta_{i,j}\lambda_i$ implies $\sum_{r\in\mathbb{Z}} E\langle \varepsilon_0, v_i\rangle\langle \varepsilon_r, v_j\rangle = \delta_{i,j}\lambda_i$.

Y_1	Y_2	$\cdots$	Y_n	1. Start with an H-valued time series.
$\downarrow$	$\downarrow$		$\downarrow$	2. Project on a basis $v_1, v_2, \ldots$
$\boldsymbol{y}_1$	$\boldsymbol{y}_2$	$\cdots$	$\boldsymbol{y}_n$	3. Work with a d-dimensional time series.
$=$	$=$		$=$	
$\langle Y_1, v_1\rangle$	$\langle Y_2, v_1\rangle$	$\cdots$	$\langle Y_n, v_1\rangle$	
$\langle Y_1, v_2\rangle$	$\langle Y_2, v_2\rangle$	$\cdots$	$\langle Y_n, v_2\rangle$	
$\langle Y_1, v_3\rangle$	$\langle Y_2, v_3\rangle$	$\cdots$	$\langle Y_n, v_3\rangle$	
$\vdots$	$\vdots$		$\vdots$	
$\langle Y_1, v_d\rangle$	$\langle Y_2, v_d\rangle$	$\cdots$	$\langle Y_n, v_d\rangle$	
$\langle Y_1, v_{d+1}\rangle$	$\langle Y_2, v_{d+1}\rangle$	$\cdots$	$\langle Y_n, v_{d+1}\rangle$	This part of the expansion (2.3.1) is considered to be less informative and is neglected.
$\vdots$	$\vdots$	$\cdots$	$\vdots$	

Table 2.1: A scheme of the projection based approach with $d \in \mathbb{N}_H$.

Since our aim is to develop non-parametric procedures, we need estimates of the static and of the long run principal components. The general approach to obtain them is to rely on eigenelements $(\hat{\lambda}_j, \hat{v}_j)$ of empirical counterparts $\hat{\mathscr{C}_0}$ and $\hat{\mathscr{C}}$ of the operators $\mathscr{C}_0$ and $\mathscr{C}$. They will be defined subsequently and convergence properties will be obtained later on in Subsection 2.3.2.

Definition 2.17 (Empirical static and long run principal components).

1. The covariance operator $\mathscr{C}_0$ is estimated via

$$\hat{\mathscr{C}_0} = \sum_{i=1}^{n} [\hat{\varepsilon}_i \otimes \hat{\varepsilon}_i]/n$$

   based on the residuals $\hat{\varepsilon}_i = Y_i - \bar{Y}_n$. $\hat{\mathscr{C}_0}$ is a self-adjoint, positive, Hilbert-Schmidt operator with a spectral decomposition $\hat{\mathscr{C}_0} = \sum_{j=1}^{d_H} \hat{\lambda}_j (\hat{v}_j \otimes \hat{v}_j)$. The eigenelements $(\hat{\lambda}_j, \hat{v}_j)$ are called the »*empirical static principal components*« and are estimates of the population counterparts.

[1] We may interchange the summation, the inner product and the expectation due to the definition of expectation in Hilbert spaces and by the continuity of inner products.

2. To estimate the long run covariance operator $\mathscr{C}$ we will work with *»Bartlett-type«* estimates $\hat{\mathscr{C}}$ defined in a straightforward manner by

$$\hat{\mathscr{C}} = \sum_{r=-n}^{n} \mathscr{K}(r/h)\hat{\mathscr{C}}_r \tag{2.3.9}$$

and based on weights $\mathscr{K}(r/h)$, where $\mathscr{K}(x)$ is a symmetric *»window«* function with bounded support and $h \in \mathbb{N}$ is the *»bandwidth«*. Both will be specified later on in Subsection 2.3.2 and will differ under Assumption M1 and Assumption M2. The $\hat{\mathscr{C}}_r$ are plug-in estimates for the *»lagged cross-covariance operators«* $\mathscr{C}_r$ and are naturally defined by

$$\hat{\mathscr{C}}_r = \begin{cases} \sum_{i=1}^{n-r} [\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i+r}]/n, & r \geq 0, \\ \sum_{i=1}^{n+r} [\hat{\varepsilon}_{i-r} \otimes \hat{\varepsilon}_i]/n, & r < 0, \end{cases}$$

with $\hat{\varepsilon}_i = Y_i - \bar{Y}_n$. It is, again, easy to verify that $\hat{\mathscr{C}}$ is a self-adjoint (however, not necessarily positive) Hilbert-Schmidt operator which therefore has a spectral decomposition

$$\hat{\mathscr{C}} = \sum_{j=1}^{d_H} \hat{\lambda}_j (\hat{v}_j \otimes \hat{v}_j). \tag{2.3.10}$$

The eigenelements $(\hat{\lambda}_j, \hat{v}_j)$ are called the *»empirical long run principal components«*.

Remark 2.18 (Properties of the principal components I). Recall that our CUSUM statistic involves the long run covariance matrix of the projected time series for a suitable standardization in (2.3.3).

1. In view of Remark 2.16 it holds that $\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d)$, if we rely on static principal components and if we are in the i.i.d. case. A convenient estimate is then given by $\hat{\Sigma} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_d)$ using the empirical static principal components. However, this does not hold true in the general time dependent situation since we have correlations within the projected scores.[1] To estimate the long run covariance of the projected data we may use, e.g., Bartlett-type estimates as suggested by Hörmann & Kokoszka (2010, Proposition 4.1). The authors point out that the proof is »delicate« which is due to the data-driven projection setting and the corresponding sign issues.

2. If we rely on long run principal components, as suggested by Horváth *et al.* (2013) in a two-sample context, then, in view of Remark 2.16, the long run covariance matrix (2.3.4) equals

$$\Sigma = \text{diag}(\lambda_1, \dots, \lambda_d). \tag{2.3.11}$$

A convenient estimate for Σ is now given, again, by $\hat{\Sigma} = \text{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_d)$ using the empirical long run principal components. Note that we are working with $\Sigma^{-1/2}$ in (2.3.2) and (2.3.3). A corresponding regularity condition will be imposed in Assumption E2, below.

[1] Cf., e.g., the discussion in Hörmann *et al.* (2015, p. 320).

Remark 2.19 (Properties of the principal components II). According to Remark 2.18 we may work with both static and long run principal components. However, under the considered dependence structure working with CUSUM statistics based on (empirical) long run principal components is mathematically more convenient whereas the use of static principal components becomes more technical. For that reason we will stick to the former approach throughout this thesis. (Note that under independence both static and long run principal components coincide.)

Nevertheless, we want to point out two features that make the static principal components generally appealing and which are the fundamental reasons for functional static principal components being one of the most important tools in functional data analysis: the scores $\langle \varepsilon_0, v_k \rangle$ are uncorrelated in k and the truncated series represents the data optimally in the mean-square sense, i.e.

$$E\|\varepsilon_0 - \sum_{j=1}^{d} \langle \varepsilon_0, v_j \rangle v_j\|^2 = E\| \sum_{j=d+1}^{d_H} \langle \varepsilon_0, v_j \rangle v_j\|^2 = \sum_{j=d+1}^{d_H} \mathrm{Var}\left[\langle \varepsilon_0, v_j \rangle\right] = \sum_{j=d+1}^{d_H} \lambda_j \tag{2.3.12}$$

is minimal amongst all possible projections on arbitrary orthonormal systems for any fixed finite $d \in \mathbb{N}_H$ $(d < d_H)$. Note that (2.3.12) follows directly by *Parsevals's identity* and by

$$\lambda_j = \langle \mathscr{C}_0(v_j), v_j \rangle = \mathrm{Var}\left[\langle \varepsilon_0, v_j \rangle\right].$$

In other words (2.3.12) states that the leading static principal components $v_1, \ldots, v_d$ capture as most of the variation of ε_0 as is possible by projections in a d-dimensional subspace. This follows, e.g., analogously to the explanations in Section 3 of Horváth & Kokoszka (2012).

Finally, we would like to mention another appealing dimension reduction possibility that is not the focus of this thesis.

Remark 2.20 (Dimension reduction via dynamic principal components). Hörmann *et al.* (2015) study Hilbert space spectral density operators $\mathcal{F}_\theta$, $\theta \in [-\pi, \pi]$, which obviously coincide with (2.3.6) at frequency zero, i.e. $\mathcal{F}_0 = \mathscr{C}$. They use the eigenelements of these operators to construct *»functional dynamic principal component scores«* via the so-called *»filters«*. These dynamic principal component scores generalize the long run principal component scores. On the one hand, they also ensure the diagonal structure of the long run covariance matrix of the d-dimensional time series. On the other hand, they provide *»optimal«* finite dimensional representation under dependence - in a sense similar to (2.3.12) (cf. Theorem 2 and Proposition 3 of Hörmann *et al.*, 2015).

In an introductory discussion the authors mention an extension of CUSUM-type statistics within their dynamic setup (but without being rigorous). Furthermore, they point out the general importance of a diagonal long run covariance matrix for this purpose (cf. also Remark 2.18).

Fully empirical CUSUM based on long run principal components

The CUSUM statistic that we will work with is a fully empirical version of (2.3.2) and is defined as follows

$$\hat{\mathscr{M}}_n = \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}(k/n), \tag{2.3.13}$$

with the same Darling-Erdős-type weighting function $w(x) = [x(1-x)]^{-1/2}, x \in (0,1)$, and the detector

$$\hat{\mathscr{T}}(x) = \hat{\mathscr{T}}(x;\hat{\boldsymbol{y}}) = |\hat{\Sigma}^{-1/2} S_n(x;\hat{\boldsymbol{y}})| = |S_n(x;\hat{\boldsymbol{y}})|_{\hat{\Sigma}}.$$

It is based on the empirical long run principal components $(\hat{\lambda}_j, \hat{v}_j)$ via *empirical projections* $\hat{\boldsymbol{y}}_i = [\langle Y_i, \hat{v}_1\rangle, \dots, \langle Y_i, \hat{v}_d\rangle]'$, for some finite $d \in \mathbb{N}_H$, and with an estimate $\hat{\Sigma} = \mathrm{diag}(|\hat{\lambda}_1|, \dots, |\hat{\lambda}_d|)$. An equivalent and mathematically sometimes more convenient *»tensor-based«* formulation of (2.3.13) is given by

$$\hat{\mathscr{M}}_n = \hat{\mathscr{M}}'_n(Y) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}'(k/n) \tag{2.3.14}$$

with

$$\hat{\mathscr{T}}'(x) = \hat{\mathscr{T}}'(x;Y) = \|\hat{\mathscr{C}}^{-1/2,d} S_n(x;Y)\| \tag{2.3.15}$$

and using

$$\hat{\mathscr{C}}^{-1/2,d} = \sum_{j=1}^{d} |\hat{\lambda}_j|^{-1/2} (\hat{v}_j \otimes \hat{v}_j). \tag{2.3.16}$$

The operator $\hat{\mathscr{C}}^{-1/2,d}$, $d \in \mathbb{N}_H$, is a truncated inverse square root of $\hat{\mathscr{C}}$. This formulation of the CUSUM statistic has been suggested by Torgovitski (2015c) and therein studied for the case $w(x) = 1, x \in (0,1)$, whereas in this thesis we consider the Darling-Erdős-type weighting. The identity between (2.3.13) and (2.3.14) can be verified by *Parseval's identity*.

Remark 2.21 (Regularization). For the long run covariance estimation we use $\hat{\Sigma} = \mathrm{diag}(|\hat{\lambda}_1|, \dots, |\hat{\lambda}_d|)$ as a simple ad-hoc regularization of the naive choice $\hat{\Sigma} = \mathrm{diag}(\hat{\lambda}_1, \dots, \hat{\lambda}_d)$. Note that the eigenvalues $\hat{\lambda}_j$ may be negative since the positivity of $\hat{\mathscr{C}}$ is generally not guaranteed.[1] By using the absolute values the matrix $\hat{\Sigma}^{-1/2}$, the mapping $|\cdot|_{\hat{\Sigma}}$ and the operator $\hat{\mathscr{C}}^{-1/2,d}$ remain well-defined as long as $\hat{\Sigma}$ has full rank. To cope with the singular case, i.e. with zero eigenvalues, we can formally set $\hat{\mathscr{M}}_n := \infty$ in that situation.

[1] For instance, estimates $\hat{\mathscr{C}}$ with flat-top kernels $\mathscr{K}$ (cf. Example 2.23, below) are known to violate this condition.

Perturbed long run covariance operator

We finish this subsection by introducing the *»perturbed long run covariance operator«* $\mathscr{C}_{\alpha,\beta}$ as a finite rank perturbation of a rescaled $\mathscr{C}$ (if $\beta \neq 0$). It will allow us a concise discussion of the asymptotic behavior of $\hat{\mathscr{C}}$ and correspondingly of its eigenelements simultaneously under the null and under the alternative.[1] Under the null hypothesis we set formally

$$\mathscr{C}_{1,0} := \mathscr{C} \tag{2.3.17}$$

whereas under the alternative we define the operator $\mathscr{C}_{\alpha,\beta}$ by the following linear combination

$$\mathscr{C}_{\alpha,\beta} = \alpha\mathscr{C} + \sum_{i,j=1}^{\varrho} \beta_{i,j}(\varDelta_i \otimes \varDelta_j) \tag{2.3.18}$$

of the long run covariance operator $\mathscr{C}$ and of the tensor products of all ϱ change directions $\varDelta_i \otimes \varDelta_j$.[2] The parameter $\beta \in \mathbb{R}^{\varrho\times\varrho}$ stands for a symmetric matrix and $\alpha \in \mathbb{R}_+$ for a non-negative scalar. (Both parameters will be specified in Subsection 2.3.2, below.) The finite rank perturbation $\mathscr{C}_{\alpha,\beta} - \alpha\mathscr{C}$ in (2.3.18) is determined by the change directions $\varDelta_j$ and is a self-adjoint Hilbert-Schmidt operator. Moreover, as we already discussed, $\mathscr{C}$ is a self-adjoint Hilbert-Schmidt operator, too. Thus, $\mathscr{C}_{\alpha,\beta}$ must be also self-adjoint Hilbert-Schmidt and therefore $\mathscr{C}_{\alpha,\beta}$ has a spectral decomposition

$$\mathscr{C}_{\alpha,\beta} = \sum_{j=1}^{d_H} \lambda_j(v_j \otimes v_j) \tag{2.3.19}$$

in both cases (2.3.17) and (2.3.18). (Under the null hypothesis this coincides with the spectral decomposition in Definition 2.15.)

The following truncated operator $\mathscr{C}_{\alpha,\beta}^{-1/2,d}$ will turn out to be useful in the next subsection for analyzing the asymptotic behavior of $\hat{\mathscr{C}}^{-1/2,d}$, cf. (2.3.16), and of our test statistic. It is constructed using the eigenelements of $\mathscr{C}_{\alpha,\beta}$ by formally setting

$$\mathscr{C}_{\alpha,\beta}^{-1/2,d} = \sum_{j=1}^{d} \lambda_j^{-1/2}(v_j \otimes v_j), \tag{2.3.20}$$

$d \in \mathbb{N}_H$. Finally, we state and briefly discuss various conditions on eigenvalues of $\mathscr{C}_{\alpha,\beta}$ that will be of importance in this chapter.

Assumption E1 (Positive definiteness). It holds that $\lambda_1 \geq \lambda_2 \geq \ldots \geq 0$.

Assumption E2 (Data dimensionality). It holds that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_d > 0,\ d \in \mathbb{N}_H$.

[1] The limiting operator of $\hat{\mathscr{C}}$ will correspond to $\mathscr{C}_{\alpha,\beta}$ with $\alpha = 1$ and $\beta = 0$ under the null hypothesis and with $\beta \neq 0$ (and some $\alpha > 0$) under the alternative.

[2] Note that we distinguish between the null hypothesis and the alternative in the definition of $\mathscr{C}_{\alpha,\beta}$ because the change directions $\varDelta_j$'s are only defined in the latter situation.

Assumption E3 (Complete separation of eigenvalues). It holds that $\lambda_1 > \lambda_2 > \ldots > \lambda_{d'} > \lambda_{d'+1} \geq 0$ for some $d' \leq d$, where $d \in \mathbb{N}_H$.

Assumption E3' (Spectral gap). It holds that $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_{d'} > \lambda_{d'+1} \geq 0$ for some $d' \leq d$, where $d \in \mathbb{N}_H$.

Assumption E1 is a technical condition that imposes positive definiteness on $\mathscr{C}_{\alpha,\beta}$. Assumption E2 will be used under H_0, i.e. for $\mathscr{C}_{1,0}$, to ensure that the data is at least d-dimensional and that the long run covariance matrix Σ has full rank. (It cannot be avoided.) The complete separation Assumption E3 will be used under H_0, as usual with $d' = d$, and under H_A with $d' \leq d$. Under the alternative this assumption is slightly weaker than the usual assumptions in the literature. (For instance for static principal components $d' = d$ is typically required as under the null in Assumption E2.) Broadly speaking the purpose of this assumption is to ensure that the directions of each $v_1, \ldots, v_{d'}$ can be estimated separately. The substantial weakening of Assumption E3' (if compared to Assumption E3) relies on the recent work of Reimherr (2015) and is introduced for CUSUM-type procedures by Torgovitski (2015c). It will be used under H_0 and under H_A in Remarks 2.28 and 2.35, below.

Finally, to fix the notation and to avoid ambiguities throughout this chapter we state the next assumption on the notation.

Notation for dimension parameters. Throughout this Chapter 2 the number d_H denotes the dimension of the separable Hilbert space H, the number $\varrho \in \mathbb{N}_H$ denotes the number of change-directions Δ_j and the number $d \in \mathbb{N}_H$ denotes the dimension of the finite dimensional subspace of H which is used to construct the CUSUM statistics (2.3.2) or (2.3.13), respectively, in order to detect the structural changes. In particular the case $\varrho > d$ is possible.

Limit theorems for long run principal components (and estimation of the long run covariance operator) 2.3.2

To understand the behavior of our statistic $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ under the null and under the alternative hypotheses we begin with a study of the asymptotic behavior of $\hat{\mathscr{C}}$ and of its eigenstructure for both scenarios. We will derive consistency with convergence rates under the null (which is necessary for the Darling-Erdős-type weighting) and show that this estimate is not consistent under the alternative. As already pointed out in the literature, e.g., by Aston & Kirch (2012a,b) and by Horváth *et al.* (2014), the estimates are affected under the alternative by the change in such a way that their asymptotic behavior (especially of the eigenvectors $\hat{v}_j$) becomes advantageous for a principal component based dimension reduction approach: the aforementioned authors show for the one-directional change setting that more changes become detectable (if compared to a consistent estimate) due to implicit alignment of $\hat{v}_1$ with the particular change direction Δ_1. This comes as a surprise because usually one would rather try to construct estimates that are not only consistent under the null but also under the alternative. We will discuss this issue in Subsection 2.3.3, below.

Estimation under the null hypothesis

As already mentioned we will distinguish between Assumption M1 and Assumption M2. In the former case we may use a more tailored estimate $\hat{\mathscr{C}}$. Hence, the results are different and sharper than under the latter assumptions. Moreover, notice that Remark 2.6 indicates that Assumption M1 covers cases which are not included by Assumption M2.

In the m-dependent setting of Assumption M1 it holds that $\mathscr{C}_r = 0$ for $r > m$ which follows immediately due to $\langle \mathscr{C}_r(x), y\rangle = E\langle \varepsilon_0, y\rangle\langle \varepsilon_r, y\rangle$ for all $x, y \in H$. Hence, knowing m, it seems natural to only estimate the »truncated« long run covariance operator $\mathscr{C} = \sum_{|r|\leq m} \mathscr{C}_r$, i.e. to avoid estimation of zeros. This is reasonably done by the following plug-in estimate which is a Hilbert space analogue of the estimate considered, e.g., by Horváth *et al.* (1999, p. 100).

Assumption K1 (Kernel and bandwidth under Assumption M1). In the m-dependent setting we use the estimate $\hat{\mathscr{C}}$, defined in (2.3.9), with $\mathscr{K}(x) = \mathbb{1}_{[-1,1]}(x)$, $x \in \mathbb{R}$ and with a fixed *»bandwidth«* $h \geq m$, $h \in \mathbb{N}$.

Note that we allow for a »misspecified« $h > m$ which, as it turns out, has some theoretical advantages for our CUSUM testing procedure that will be discussed later on. In fact some »moderate overfitting« can increase the power. In the next Proposition 2.22 we obtain convergence rates for the estimate $\hat{\mathscr{C}}$ under m-dependence based on a *Marcinkiewicz-Zygmund* type law of large numbers which allows us to assume less than four moments.

Proposition 2.22. *Let Assumption M1 and Assumption K1 hold true and $\mathscr{C}_{1,0}$ be defined as in (2.3.17). Under H_0 it holds that, as $n \to \infty$,*

$$\|\hat{\mathscr{C}} - \mathscr{C}_{1,0}\|_{\mathcal{S}} = \mathcal{O}_P(r_n)$$

with the rate $r_n = n^{-1+2/\min\{\kappa,4\}}$.

In the general case of m-approximable time series of Assumption M2 we need (necessarily) to estimate the complete long run covariance operator $\mathscr{C} = \sum_{r\in\mathbb{Z}} \mathscr{C}_r$ if we aim to work with the long run principal components rather than with the static ones. A typical class of candidates for such estimates is given next.

Assumption K2 (Kernel and bandwidth under Assumption M2). In the m-approximable setting we use the estimate $\hat{\mathscr{C}}$, defined in (2.3.9), with a bandwidth $h \to \infty$, $h = o(n)$ as $n \to \infty$, $h \in \mathbb{N}$, and with a window $\mathscr{K}$ for which we assume:

1. *boundedness,* $\mathscr{K}(0) = 1$ and $|\mathscr{K}(x)| \leq 1$ for any x,
2. *symmetry,* $\mathscr{K}(x) = \mathscr{K}(1 - x)$ for all $x \in \mathbb{R}$,
3. *a bounded support,* $\mathscr{K}(x) = 0$ for all $|x| > a$ for some $a > 0$,
4. *continuity,* $\mathscr{K}(x)$ is piecewise continuous,

5. *and a convergence rate,* $\lim_{x\downarrow 0}|\mathscr{K}(x)-1|/x^{\varsigma}$ exists and is finite for some $\varsigma \geq 1$.

Above assumptions are common (cf., e.g., Andrews (1991) in a finite-dimensional setup and, e.g., Horváth *et al.* (2013, 2014) and Berkes *et al.* (2016) in related functional data frameworks). The fifth condition implies continuity of $\mathscr{K}(x)$ at $x=0$ and the largest number ς that provides a non-zero limit is the so-termed *»characteristic exponent«* of $\mathscr{K}$ (cf., e.g., Parzen, 1957 and Andrews, 1991).

Example 2.23 (Popular windows). *Some typically used window functions are*

1. *Plain-window:* $\mathscr{K}(x) = \mathbb{1}_{[-a,a]}(x)$ *for some* $a>0$,

2. *Flat-top-window:* $\mathscr{K}(x) = \min\{\max\{a-|x|,0\},1\}$ *for some* $a>1$,

3. *Bartlett-window:* $\mathscr{K}(x) = \max\{1-|x|,0\}$.

The first two windows provide the finiteness of $\lim_{x\downarrow 0}|\mathscr{K}(x)-1|/x^{\varsigma}$ *for any* $\varsigma \geq 1$ *and the third for* $\varsigma = 1$. *For further common examples we refer to Brockwell & Davis (1991) and to Andrews (1991). Notice that some of their examples are not included in our considerations. As pointed out in the latter article some popular kernels, such as the »Quadratic-spectral« window, have an unbounded support and thus do not belong to the class defined by Assumption K2.*

To prove convergence rates under m-approximability our proof techniques require higher moment assumptions compared to Proposition 2.22 (since we work with the variance of the estimates in this case) and, as should be expected, the obtained rates in the more general setting are also weaker. The next theorem contributes, e.g., to the results of Horváth *et al.* (2013, Theorem 2), Aston & Kirch (2012a, Lemma 2.3), Hörmann & Kokoszka (2010, Theorem 3.1) and Hörmann *et al.* (2015, Proposition 4).

Theorem 2.24. *Let Assumption M2 be fulfilled with* $\kappa = 4$. *Furthermore, let Assumption K2 hold true and* $\mathscr{C}_{1,0}$ *be defined as in* (2.3.17). *Then it holds under* H_0 *that, as* $n\to\infty$,

$$\|\hat{\mathscr{C}} - \mathscr{C}_{1,0}\|_{\mathcal{S}} = \mathcal{O}_P(r_n) \tag{2.3.21}$$

with the rate

$$r_n = (h/n)^{1/2}h + (1/n)\sum_{r=1}^{m} r\|\mathscr{C}_r\|_{\mathcal{S}} + (1/h^{\varsigma})\sum_{r=1}^{m} r^{\varsigma}\|\mathscr{C}_r\|_{\mathcal{S}} \\ + \sum_{r=m+1}^{\infty}\delta(r) + m\delta(m) \tag{2.3.22}$$

for any sequence $m = m_n \to \infty$, $m\in\mathbb{N}$, $m\leq h$ *and with* $h = o(n)$ *as* $n\to\infty$. *(The kernel-parameter* ς *is specified in Assumption K2.)*

Above rate (2.3.22) tends to zero if $h = o(n^{1/3})$ given that $\sum_{m=1}^{\infty} m^{\varsigma}\delta(m) < \infty$. (The latter condition ensures convergence of the second, third, fourth and fifth terms in (2.3.22) since the bound $\|\mathscr{C}_r\|_{\mathcal{S}} \leq c\,\delta(r)$ holds true with some $c > 0$ that is independent of r.)

Note that this rate reflects, as expected, the typical trade-off between the variance and the bias in kernel estimation: a larger bandwidth yields higher variance on the price of a smaller bias and vice versa. Moreover, a higher smoothness parameter ς lowers the bias contribution, too.

Remark 2.25 (Comparison to Hörmann *et al.*, 2015). Let us briefly compare our results to the findings of Hörmann *et al.* (2015) for the Bartlett-window $\mathscr{K}(x) = \max\{1 - |x|, 0\}$ and therefore consider $\varsigma = 1$ and $m = h$ which simplifies (2.3.22) to

$$r_n = \mathcal{O}\Big((h/n)^{1/2}h + (1/h)\sum_{r=1}^{h} r\delta(r) + \sum_{r=h+1}^{\infty}\delta(r) + m\delta(m)\Big). \tag{2.3.23}$$

In Hörmann *et al.* (2015) the authors showed a slightly better (but surprisingly essentially the same) rate

$$\begin{aligned} r_n &= (h/n)^{1/2}h + (1/h)\sum_{r=1}^{h} r\|\mathscr{C}_r\|_{\mathcal{S}} + \sum_{r=h+1}^{\infty}\|\mathscr{C}_r\|_{\mathcal{S}} \\ &= \mathcal{O}\Big((h/n)^{1/2}h + (1/h)\sum_{r=1}^{h} r\delta(r) + \sum_{r=h+1}^{\infty}\delta(r)\Big), \end{aligned} \tag{2.3.24}$$

where (2.3.24) corresponds to the order of (2.3.23) but without the (usually negligible) term $m\delta(m)$.[1] They use a different but related approach where they rely on uniform bounds for lagged estimates $\hat{\mathscr{C}}_r$.

For the clarity of our exposition it will be convenient to restrict ourselves subsequently to polynomial bandwidths.

Assumption PB (Polynomial bandwidths). We use the estimate $\hat{\mathscr{C}}$, defined in (2.3.9), with a kernel $\mathscr{K}$ that fulfills the Assumption K2 and we assume a bandwidth $h = \lfloor n^{1/\mu} \rfloor$ for any $\mu > 3$.

Corollary 2.26. *Let Assumption M2 be fulfilled with $\delta(m) = m^{-\nu}$, $\nu > \varsigma + 1$ and with $\kappa = 4$. Furthermore, let Assumption PB be fulfilled, too. Then (2.3.21) holds true with a polynomial rate of convergence $r_n = \mathcal{O}(n^{-\varepsilon})$ for some $\varepsilon > 0$.*

[1] Hörmann *et al.* (2015) was published online on 18th of July 2014 but Hörmann *et al.* (2015, arXiv:1210.7192v1), a preprint of the latter article, was previously available on arXiv.org since 26th of October 2012 and, as it turns out, already contained these rates (implicitly) in the proof of Proposition 11 of Appendix B. In the final and in the updated preprint version Hörmann *et al.* (2015, arXiv:1210.7192v3) this result was shifted to the main part of the articles and became Proposition 4. Our proof technique is different but it is also worth emphasizing that we were not aware of the results in Hörmann *et al.* (2015) and of the latter restructured presentations when Torgovitski (2016, arXiv:1407.3625v1), i.e. the first version of Torgovitski (2016), was published on arXiv.org on 14th of July 2014.

Corollary 2.27. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 hold true. Furthermore, let* (λ_j, v_j) *be the eigenelements of* $\mathscr{C}_{1,0}$*, which is defined in* (2.3.17). *Under* H_0 *and the Assumption E2 it holds that, as* $n \to \infty$,

$$\max_{1\le k\le d} |\hat{\lambda}_k - \lambda_k| = o_P\big(n^{-\varepsilon_1}\big) \tag{2.3.25}$$

and under the additional Assumption E3 it holds with $\hat{s}_k = \operatorname{sign}\langle \hat{v}_k, v_k\rangle$ *that*

$$\max_{1\le k\le d'} \|\hat{s}_k \hat{v}_k - v_k\| = o_P\big(n^{-\varepsilon_1}\big) \tag{2.3.26}$$

for some $\varepsilon_1 > 0$. *In the latter case we have that*

$$\|\hat{\mathscr{C}}^{-1/2,d'} - \mathscr{C}_{1,0}^{-1/2,d'}\|_{\mathcal{S}} = o_P\big(n^{-\varepsilon_2}\big) \tag{2.3.27}$$

for some $\varepsilon_2 > 0$, *where the operators* $\hat{\mathscr{C}}^{-1/2,d'}$ *and* $\mathscr{C}_{1,0}^{-1/2,d'}$ *are defined in* (2.3.16) *and in* (2.3.20).

Note that related results on convergence rates (2.3.25) and (2.3.26) for principal components of functional data go back at least to the famous work of Dauxois *et al.* (1982). For further related results under different dependence assumptions and with different rates we refer, e.g., to Hörmann & Kokoszka (2010, Theorem 3.2), to Aston & Kirch (2012a, Lemma 2.3), to Horváth & Kokoszka (2012, Chapter 2) and to Bosq (2000, Chapter 4).

Remark 2.28 (Eigenvalue separation I). If $\hat{\mathscr{C}}$ is positive definite (or if d_H is finite), then relation (2.3.27) holds also true under the more general Assumption E3' instead of Assumption E3. We will give a proof in Subsection 2.5.1 on p. 70.

Note that we have positive definiteness of $\hat{\mathscr{C}}$ for instance under the assumptions of Proposition 2.22 if we are in the 0-dependent (i.i.d.) setting and if we work with the sample covariance estimate $\hat{\mathscr{C}}$ with $h = 0$. (Recall that the long run principal components coincide with the static ones in this scenario.) Furthermore, positivity may be ensured under the assumptions of Corollary 2.26 by an appropriate kernel function $\mathscr{K}$: if we consider $\langle \hat{\mathscr{C}} x, x\rangle$ for any fixed $x \in H$ and evaluate it similarly to (2.3.8), then we see that the same class of kernels provides positive definite estimates as in the univariate setting.[1] Finally, note that our testing approach is not restricted to Bartlett-type estimates at all. We could also work with any other estimates or positive definite modifications instead.[2] (However, we would need to provide convergence rates as in (2.3.21).)

The literature on long run covariance estimation has a long tradition. A dominating amount of literature that deals particularly with convergence rates assumes the finiteness of fourth moments which then allows to study the mean squared error of the estimates $\hat{\mathscr{C}}$. It would be desirable to have convergence rates for these estimates in our theoretical settings with less than finite fourth moments, since our Darling-Erdős-type CUSUM (such as various other statistical procedures), do not intrinsically rely on such

[1] Examples for the latter are given in Andrews (1991). For instance, the Bartlett-kernel is included in this class.

[2] Research in this direction is announced in Berkes *et al.* (2016).

a moment assumption. In finite dimensional cases it is known that the long run covariance matrix may be estimated with suitable rates assuming only finite $2+\delta$-moments with any $\delta > 0.$ Notable results in this direction are, e.g., Steinebach (1995) or Antoch *et al.* (1997). The latter have shown the rate $r_n' = (h/n)^{1/2}(h\log h)^{1/2} + 1/h$ in a univariate linear time series setup using a Bartlett-window and relying on a Marcinkiewicz-Zygmund type law of large numbers. Steinebach (1995) showed polynomial rates for another class of the so-called *»batch-mean«* estimates. The theory in the latter article relies mainly on suitable strong invariance principles - which are available under a variety of weak dependence assumptions - and the convergence rates depend on the approximation rates of those invariance principles.[1]

Steinebach (1995) and Antoch *et al.* (1997) can be restated to a large extent for weakly dependent multivariate time series in a straightforward manner and similar results should also hold true for some restricted classes of functional time series as well. The extension of the results of Steinebach (1995) or of Antoch *et al.* (1997) to general linear and nonlinear time series in our Hilbert space m-approximable framework is beyond the scope of this thesis and could be part of future research.

Meanwhile, amongst others, Horváth *et al.* (2014) and Berkes *et al.* (2016) have studied the estimate $\hat{\mathscr{C}}$ in a similar framework.[2] The results of Horváth *et al.* (2014) and of Berkes *et al.* (2016) appear stronger since they study asymptotic normality, derive mean squared error rates of order $(h/n)^{1/2} + 1/h^{\varsigma}$ (with a parameter ς as in Assumption K2) and provide guidance on the selection of the bandwidth $h,$ too. First of all, it is important to point out that we use a different technique compared to Horváth *et al.* (2014) and Berkes *et al.* (2016) which is an interesting aspect of our results. Moreover, our results are not directly comparable due to differences in the assumptions. Our result does not require $\nu > 4$ for polynomial rates of approximation $\delta(m) = m^{-\nu}$ as is required in Horváth *et al.* (2014). We also do not require the existence of more than the fourth moment which is required in Berkes *et al.* (2016). Therein, either $E\|\varepsilon_1\|^8 < \infty$ is assumed, in which case $h = \lfloor n^{1/\mu} \rfloor$ with any $\mu > 1$ is admissible, or $E\|\varepsilon_1\|^{4+\delta_1} < \infty$ for some $\delta_1 > 0$ is required, in which case the bandwidth is constrained by $\mu > (4+2\delta_1)/\delta_1$ and thus $\mu \to \infty$ as $\delta_1 \to 0.$

Remark 2.29 (Cumulant-based results). For the sake of completeness, we would like to mention that results on consistency of long run covariance estimates are often derived based on *»cumulant-type«* conditions (cf., e.g., the classical results in Andrews (1991) or the newer results of Hörmann & Kokoszka (2010) in the m-approximable setting). It is claimed in Horváth *et al.* (2014) that cumulant-type conditions are typically rather difficult to verify which is the motivation of the authors to provide, on the one hand, a *»simplified cumulant«* condition, and on the other hand, to show how the latter can be verified indirectly but more conveniently via m-approximability.

[1] In this context, it is worth to highlight the recent and striking result of Berkes *et al.* (2015). The authors showed optimal *Komlós-Major-Tusnády* type invariance principles in a dependent univariate framework of Wu (2011) - a dependence setting that is closely related to m-approximability.

[2] Note that Horváth *et al.* (2014) was published online on 12th of June 2014 whereas Torgovitski (2016, arXiv:1407.3625v1) was published shortly after on 14th of July 2014.

Remark 2.30 (Bartlett-type estimates under m-dependence). So far, we have formally used different estimates under m-dependence and under m-approximability. Clearly, it is also reasonable to use the general estimates $\hat{\mathscr{C}}$ from Assumption K2 in the m-dependent setting of Assumption M1, as well. Many results of this chapter that are stated under Assumptions M2 and K2 could then be restated under Assumptions M1 and K2 but our proofs and most theorems that we rely on would require some minor modifications (cf. the discussion on p. 26). To avoid confusion we do not consider this case explicitly.

Estimation under the alternative hypothesis

We show how the limiting behavior of $\hat{\mathscr{C}}$ and of its eigenelements $(\hat{\lambda}_j, \hat{v}_j)$ is characterized by the contaminated operator $\mathscr{C}_{\alpha,\beta}$, defined via (2.3.18), and by its eigenelements (λ_j, v_j). We will see that the parameters α and β are affected by the directions of the changes Δ_j, the shape of the trend-functions g_j, the shape of the window $\mathscr{K}$ and also by the chosen bandwidth h. The results of this subsection are extensions of Torgovitski (2015a,c, 2016). They coincide for $d_H = \infty$ with the counterparts in Berkes *et al.* (2009) and Aston & Kirch (2012a) for a one-directional abrupt change (or also an epidemic change in the latter article) if we are in the special setting of Proposition 2.22 with $h = m = 0$ and $\kappa = 4$, i.e. if we are in the independent case and work in $L^2[0,1]$ with the functional static principal components under finite fourth moment assumptions. The following mapping is important not only for this but also for the subsequent sections. We set

$$\mathscr{G}(g_i, g_j) = \Big[\int_0^1 g_i(x) g_j(x) dx\Big] - \Big[\int_0^1 g_i(x) dx \int_0^1 g_j(x) dx\Big], \tag{2.3.28}$$

where g_i are the piecewise Lipschitz continuous »trend-change« functions that describe the deviations under the alternative in (2.2.4). Note that $\mathscr{G}(g_i) := \mathscr{G}(g_i, g_i)$ is positive under H_A since g_i is assumed to be non-constant.

Remark 2.31 (Evaluation of $\mathscr{G}$ for popular change settings). For the abrupt, epidemic and piecewise linear change functions g_A, g_E and g_L from Remark 2.8 we have

$$\begin{aligned}
&\mathscr{G}(g_A) = f(\theta_1), \\
&\mathscr{G}(g_E) = f(\theta_2 - \theta_1), \\
&\mathscr{G}(g_L) = f((\theta_2 + \theta_1)/2) - (\theta_2 - \theta_1)/6,
\end{aligned}$$

with $f(x) = x(1-x)$ and $0 < \theta_1 < \theta_2 \le 1$.

Theorem 2.32. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 hold true and $\mathscr{C}_{1,\beta}$ be defined as in (2.3.18). Under H_A it holds that, as $n \to \infty$,*

$$\|\hat{\mathscr{C}} - \mathscr{C}_{1,\beta}\|_{\mathcal{S}} = o_P(\beta_h), \tag{2.3.29}$$

where we use $\beta = \beta_h \beta_{\mathcal{K}} \beta_{\mathcal{G}}$ *with*

$$\beta_{\mathcal{K}} = \int_0^\infty \mathscr{K}(x)dx, \quad \beta_h = 2h+1, \quad \beta_{\mathcal{G}} \in \mathbb{R}^{\varrho \times \varrho}, \tag{2.3.30}$$

with matrix entries $(\beta_{\mathcal{G}})_{i,j} = \mathscr{G}(g_i, g_j)$ *as defined in* (2.3.28).

This theorem contributes to and extends, e.g., the results of Horváth *et al.* (2013, Theorem 7), Horváth *et al.* (2014, Lemma B.2) and Aston & Kirch (2012a, Lemma 2.4).

Example 2.33 (The parameter matrix $\beta_{\mathcal{G}}$ and the operator $\mathscr{C}_{1,\beta}$). *Consider a multiple two-directional change setting* (2.2.4) *with abrupt changes* $g_1(x) = g_A(x, \theta_1)$*,* $g_2(x) = g_A(x, \theta_2)$*,* $0 < \theta_1 < \theta_2 < 1$ *in some orthonormal directions* Δ_1 *and* Δ_2 *(cf. Remark 2.8). In this case we obtain*

$$\beta_{\mathcal{G}} = \begin{bmatrix} \theta_1(1-\theta_1) & \theta_1(1-\theta_2) \\ \theta_1(1-\theta_2) & \theta_2(1-\theta_2) \end{bmatrix}$$

for (2.3.30) *which yields*

$$\mathscr{C}_{1,\beta} = \mathscr{C} + (2h+1) \int_0^\infty \mathscr{K}(x)dx \Big[\theta_1(1-\theta_1)\Delta_1 \otimes \Delta_1 + \theta_2(1-\theta_2)\Delta_2 \otimes \Delta_2 + \\ \theta_1(1-\theta_2)\big(\Delta_1 \otimes \Delta_2 + \Delta_2 \otimes \Delta_1\big)\Big]$$

as the operator in (2.3.29).

Note that under the assumptions of Proposition 2.22 the bandwidth h is fixed and that $\int_0^\infty \mathscr{K}(x)dx = 1$ since we work with the plain window $\mathscr{K}(x) = \mathbb{1}_{[-1,1]}(x)$ (cf. Assumption K1). Whereas, under the assumptions of Theorem 2.24 the bandwidth necessarily must increase with the sample size to capture the correlation structure (yet the choice of the windows is more flexible).

We see a structurally different behavior of estimates if the bandwidth tends to infinity: the estimate $\hat{\mathscr{C}}$ does not stabilize without an appropriate rescaling since β_h in (2.3.30) tends to infinity. This might appear as a drawback at first sight but in fact this turns out to be an advantage if we look at the behavior of the eigenstructure (cf., e.g., Aston & Kirch, 2012a,b, Horváth *et al.*, 2014 and the related discussions in Remark 2.47, below).

The next corollary contributes to and extends, e.g., Lemma 1 of Berkes *et al.* (2009) and Theorem 2.1 of Aston & Kirch (2012a).[1]

Corollary 2.34. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 hold true and let* (λ_j, v_j) *be the eigenelements of* $\mathscr{C}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}$*, which is defined in* (2.3.18)*. Moreover, let* $\alpha = \lim_{n\to\infty} 1/\beta_h$ *and* β_h*,* $\beta_{\mathcal{K}}$*,* $\beta_{\mathcal{G}}$ *be set as in* (2.3.30)*. Under* H_A *and the Assumptions E1 and E3 it holds that, as* $n \to \infty$*,*

$$\max_{1\le k\le d'} |\hat{\lambda}_k/\beta_h - \lambda_k| = o_P(1) \tag{2.3.31}$$

[1] Cf. also Horváth *et al.* (2014).

and with $\hat{s}_k = \operatorname{sign}\langle \hat{v}_k, v_k \rangle$ *that*

$$\max_{1 \le k \le d'} \|\hat{s}_k \hat{v}_k - v_k\| = o_P(1). \tag{2.3.32}$$

Furthermore, it holds that

$$\|\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}_{\alpha,\beta_\mathcal{K}\beta_\mathcal{G}}^{-1/2,d'}\|_\mathcal{S} = o_P(1), \tag{2.3.33}$$

where the operators $\hat{\mathscr{C}}^{-1/2,d'}$ *and* $\mathscr{C}_{\alpha,\beta_\mathcal{K}\beta_\mathcal{G}}^{-1/2,d'}$ *are defined in* (2.3.16) *and in* (2.3.20).

We would like to point out the notation within the preceding corollary. We have

$$\alpha = \lim_{n\to\infty} 1/\beta_h = \begin{cases} 1/\beta_h, & \text{under the Assumptions of Proposition 2.22,} \\ 0, & \text{under the Assumptions of Corollary 2.26.} \end{cases}$$

(Recall that we set $\beta_h = 2h + 1$. On the one hand, h is fixed in the first case, and, on the other hand, h tends to infinity in the second one.)

Remark 2.35 (Eigenvalue separation II). Assume that $\hat{\mathscr{C}}$ is positive definite and that Assumption E1 holds true. Then (2.3.33) holds also true under Assumption E3' instead of the stronger Assumption E3. The verification is similar as for Remark 2.28 and relies on Proposition 2.54. (Regarding positivity of $\hat{\mathscr{C}}$ we refer to Remark 2.28, too.)

2.3.3 *Limit theorems for weighted CUSUM tests (based on long run principal components)*

Before we present our theoretical results on principal component based Darling-Erdős-type tests (starting on p. 49) we would like to begin with a preliminary discussion that explains and summarizes some theoretical foundations of this section. This should provide a basic intuition for our results. For the sake of a clearer presentation, let us introduce the multivariate special cases of Assumptions S1, M1 and M2.

Assumption S1'. The sequence $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ is $\mathbb{R}^d$-valued, centered and strictly stationary with $E|\boldsymbol{e}_1|^\kappa < \infty$ for some $\kappa > 2$.

For the moment, we will assume the sequence $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ to have a full rank long run covariance matrix Σ in which case we may (without loss of generality) rescale the time series and thus assume that the long run covariance matrix is the identity matrix.

Assumption M1'. The sequence $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ fulfills Assumption S1' and is m-dependent with the identity matrix as the long run covariance matrix.

Assumption M2'. The sequence $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ fulfills Assumption S1' and is m-approximable with the identity matrix as the long run covariance matrix.

The asymptotic theory for Darling-Erdős-type CUSUM procedures relies essentially on (strong) invariance principles that provide the connection between weighted tied-down partial sums

$$\mathscr{M}_n(\boldsymbol{e}) = \max_{1\le k<n} w(k/n)\mathscr{T}(k/n),$$

with $\mathscr{T}(k/n) = |S_n(k/n;\boldsymbol{e})|$, and functionals of continuous-time multivariate Gaussian processes (cf., e.g., Proposition 2.57, below). For the latter the asymptotic behavior (for $n \to \infty$) as well as convenient tail approximations are well-known. These processes and asymptotics will be discussed subsequently followed by a paragraph on appropriate invariance principles in our situations of Assumption M1' or of Assumption M2'.

I. Preliminaries on multivariate Gaussian processes (Extreme value asymptotics and tail approximations)

Let

$$\boldsymbol{W} = \boldsymbol{W}^{(d)} = [W_1, \dots, W_d]'$$

denote a standard d-dimensional *»Wiener process«* (cf. p. xx) where all coordinates consist of independent standard Wiener processes $\{W_i(t),\ t\in[0,\infty)\}$, $1\le i\le d$. Define

$$\begin{aligned}\boldsymbol{B} &= \boldsymbol{B}^{(d)} = [B_1,\dots,B_d]',\\ \boldsymbol{U} &= \boldsymbol{U}^{(d)} = [U_1,\dots,U_d]',\end{aligned}$$

where the $\{B_i(t),\ t\in[0,1]\}$ are independent standard *»Brownian bridges«* and where the latter $\{U_i(t),\ t\in(-\infty,\infty)\}$ are independent *»Ornstein-Uhlenbeck processes«* (cf. p. xx, too). Above processes are known to be transformations of each other. For instance, the relation

$$\begin{aligned}&\{\boldsymbol{W}(e^t)/e^{t/2},\ t\in\mathbb{R}\}\\ &\stackrel{\mathcal{D}}{=} \{\boldsymbol{U}(t),\ t\in\mathbb{R}\}\\ &\stackrel{\mathcal{D}}{=} \{\boldsymbol{B}(e^t/(1+e^t))(1+e^t)/e^{t/2},\ \ t\in\mathbb{R}\}\end{aligned} \tag{2.3.34}$$

follows by a direct comparison of the covariance structure (cf., e.g., Section 1.9 of Csörgő & Révész, 1981). A well-known fundamental observation, on which our theory on CUSUM testing essentially relies, is that the supremum of the Euclidean norm of a multivariate Ornstein-Uhlenbeck process has a *»Gumbel type«* limit distribution

$$\lim_{T\to\infty} P\Big(a(T)\sup_{t\in[0,T]}|\boldsymbol{U}^{(d)}(t)| - b_d^*(T) \le x\Big) = \exp(-2\exp(-x)) \tag{2.3.35}$$

utilizing suitable correction functions

$$\begin{aligned}a(t) &= (2\log t)^{1/2},\\ b_d(t) &= (2\log t) + (d/2)\log\log t - \log\Gamma(d/2),\\ b_d^*(t) &= (2\log t) + (d/2)\log\log t - \log\Gamma(d/2) - \log(2) = b_d(t) - \log(2)\end{aligned} \tag{2.3.36}$$

and where Γ stands for the usual *Gamma function*.[1] As a consequence, the supremum of the Euclidean norms of a multivariate Wiener process and of a Brownian bridge fulfill

$$\lim_{n\to\infty} P\Big(a_n \sup_{t\in[1,n]} |\boldsymbol{W}^{(d)}(t)|/t^{1/2} - b^*_{n,d} \leq x\Big) = \exp(-2\exp(-x)), \tag{2.3.37}$$

$$\lim_{n\to\infty} P\Big(a_n \sup_{t\in[h,1-h]} |\boldsymbol{B}^{(d)}(t)|/(t(1-t))^{1/2} - b_{n,d} \leq x\Big) = \exp(-2\exp(-x)), \tag{2.3.38}$$

for all $x \in \mathbb{R}$, using (2.3.36) as follows

$$\begin{aligned} a_n &= a(\log n), \\ b_{n,d} &= b_d(\log n), \\ b^*_{n,d} &= b^*_d(\log n), \end{aligned} \tag{2.3.39}$$

with sequences $h = h_n \geq 1/n$ that fulfill

$$h = \mathcal{O}(\exp((\log n)^{1-\varepsilon})/n), \tag{2.3.40}$$

for some $\varepsilon \in (0,1)$ and as $n \to \infty$. Finally, let us mention a well-known tail approximation by Vostrikova (1981) (recall (2.3.34)) as a complement to (2.3.38). It states for $0 < h < 1/2$, that

$$P\Big(\sup_{t\in[h,1-h]} |\boldsymbol{B}^{(d)}(t)|/(t(1-t))^{1/2} \geq x\Big) = \frac{x^d \exp(-x^2/2)}{2^{d/2}\Gamma(d/2)} R(x;h,d) \tag{2.3.41}$$

holds true with $R(x;h,d) = (1-d/x^2)\log[(1/h-1)^2] + 4/x^2 + \mathcal{O}(1/x^4)$ and as $x \to \infty$.

Remark 2.36 (References for above results). The asymptotics in (2.3.35) are discussed in Theorem A.3.2 in Csörgő & Horváth (1997). The asymptotics in (2.3.37) and (2.3.38) follow essentially from (2.3.34) together with (2.3.35). Note that for the latter (2.3.38) one has to repeat the steps of Corollary A.3.1 of Csörgő & Horváth (1997) and therein take Theorem A.3.4 of Csörgő & Horváth (1997) into account.

II. Preliminaries on invariance principles (Gumbel type and Brownian bridge type approximations)

Let $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ fulfill either Assumption M1' or Assumption M2' with at least a polynomial rate of decay of order $\nu > 2$. In both cases a d-dimensional standard Wiener process exists such that

$$|\sum_{i=1}^{n} \boldsymbol{e}_i - \boldsymbol{W}(n)| = \mathcal{O}(n^{1/2-\eta}) \qquad \text{a.s.} \tag{2.3.42}$$

holds true as $n \to \infty$ for some $\eta > 0$ (cf. Remark 2.37, below).[2] The existence of such a strong approximation (2.3.42) is under both assumptions sufficient to derive the

[1] Cf., e.g., Schmitz (2011).

[2] In the construction of such an approximation it is assumed that the sequence $\{\boldsymbol{e}_i\}_{i\in\mathbb{Z}}$ is redefined on a richer new probability space leaving the distribution unchanged. As common, the probability space and the sequences are denoted as before.

following limiting distribution of the multivariate Darling-Erdős-type CUSUM statistic as

$$\lim_{n\to\infty} P\big(a_n \mathscr{M}_n(\boldsymbol{e}) - b_{n,d} \le x\big) = \exp(-2\exp(-x)), \tag{2.3.43}$$

for all $x \in \mathbb{R}$. (Cf. Proposition 2.57, below.) As will be discussed later on, this is based on the fundamental limit theorems for multivariate Gaussian processes of the last paragraph and forms the basis of all our tests. A reason for $\mathscr{M}_n(\boldsymbol{e})$ to be commonly referred to as the »Darling-Erdős-type CUSUM« detector is that the univariate version of (2.3.43) can be traced back essentially to the work of Darling & Erdős (1956). This is demonstrated in Csörgő & Horváth (1993, p. 256). Some additional insight into the behavior of $\mathscr{M}_n(\boldsymbol{e})$ and an another proof of (2.3.43) (cf. Remark 2.38, below) may be gained by subsequent Brownian bridge type approximations. Let

$$V(\chi; h) := \sup_{x\in[h,1-h]} \chi(x,d), \tag{2.3.44}$$

with

$$\chi(x,d) := w(x)|\boldsymbol{B}^{(d)}(x)| = |\boldsymbol{B}^{(d)}(x)|/(x(1-x))^{1/2}. \tag{2.3.45}$$

Under Assumption M1' it is possible to show that

$$\mathscr{M}_n(\boldsymbol{e}) = V(\chi_n; h) + o_P\big(\exp(-(\log n)^{1-\varepsilon})\big), \tag{2.3.46}$$

holds true, as $n \to \infty$, for a sequence $h \ge 1/n$ chosen according to (2.3.40) and where

$$\{\chi_n(t,d),\, t \in (0,1)\} \stackrel{\mathcal{D}}{=} \{\chi(t,d),\, t \in (0,1)\}, \tag{2.3.47}$$

for $n \in \mathbb{N}$.[1] The notation (2.3.44) will be convenient and contribute to the clarity of our statements and of the corresponding proofs. In particular it emphasizes that $\chi^2(x,d)$ has for each $x \in (0,1)$ a chi-square distribution with d degrees of freedom.

Remark 2.37 (References for the above strong approximation). Under Assumption M1' the approximation (2.3.42) is shown in Horváth *et al.* (1999, Lemma 4.1), via an application of general invariance principles of Eberlein (1986, Theorem 1) which in turn are based on the fundamental results of Berkes & Philipp (1979). Note that the work of Berkes & Philipp (1979) is also essential for Aue *et al.* (2014, Theorem S2.1) which states (2.3.42) under Assumption M2' in the causal case. We checked that (2.3.42) may be restated in the non-causal setting via Aue *et al.* (2014, Theorem S2.1), as well, by taking decomposition (2.5.66), below, into account.

Remark 2.38 (References for the Gumbel type approximation). The general techniques to obtain (2.3.43) are, e.g., sketched for univariate mixing linear processes in Theorem 4.1.3 of Csörgő & Horváth (1997) based on (2.3.42). Techniques to obtain this limit relying on somewhat more sophisticated Brownian bridge type approximations

[1] This should also be possible under Assumption M2' but is beyond the scope of this thesis.

(similar to (2.3.46)) are presented in detail in Schmitz (2011). Brownian bridge type approximations (2.3.46), themselves, are derived

1. in the multivariate i.i.d. case, e.g., in Theorems 1.1.2, 1.3.1 and 1.3.2 of Csörgő & Horváth (1997),

2. for multivariate m-dependent stationary processes, e.g., in Horváth *et al.* (1999).[1]

Extension and adaptation of the above results to our situation will be an important part of our Theorem 2.40 and Theorem 2.43, below.[2] Therein, the limit of the test statistic follows essentially on combining (2.3.42) with (2.3.37).

Remark 2.39. The common distinction between »Brownian bridge type approximations« and »Gumbel type approximations« is somewhat misleading since the derivation of the latter (typically) involves some Brownian bridge approximation(s), as well.[3]

Asymptotics under the null hypothesis

Recall that we are interested in the CUSUM test statistic $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ based on projections on empirical long run principal components in a non-parametric setting. (Cf. (2.3.13) for the traditional and (2.3.14) for a more recent tensor based formulation of this statistic.) The aim of this paragraph is to derive asymptotically correct critical values for the corresponding test for $n \to \infty$. The results are based, on the one hand, on asymptotics for the statistic $\mathscr{M}_n(\boldsymbol{y})$ and, on the other hand, on replacement of the unobservable scores $\boldsymbol{y}$ by the empirical scores $\hat{\boldsymbol{y}}$. The asymptotics for $\mathscr{M}_n(\boldsymbol{y})$ follow by verifying and utilizing the Gumbel type limit theorems and the Brownian bridge type approximations indicated for $\mathscr{M}_n(\boldsymbol{e})$ in (2.3.43) and in (2.3.46). (The structure of this paragraph follows Torgovitski, 2015a.)

Theorem 2.40. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 hold true. Let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{1,0}$, which is defined in* (2.3.17)*, and assume that Assumption E3 holds true with $d' = d$, i.e. particularly with $\lambda_d > 0$. Then under H_0 it holds that*

$$\lim_{n\to\infty} P\big(a_n \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) - b_{n,d} \le x\big) = \exp(-2\exp(-x)) \tag{2.3.48}$$

for all $x \in \mathbb{R}$. The sequences a_n and $b_{n,d}$ are given in (2.3.39).

Remark 2.41 (Eigenvalue separation III). The proof of Theorem 2.40 (p. 74) relies on the asymptotics for $\hat{\mathscr{C}}^{-1/2,d} - \mathscr{C}_{1,0}^{-1/2,d}$ in (2.3.27). Hence, if $\hat{\mathscr{C}}$ is positive definite,

[1] For linear univariate processes such approximations may be derived, e.g., via the Beveridge-Nelson decomposition in Berkes *et al.* (2009). The approximation (2.3.42) follows then, e.g., via Lemma 3.1 of Berkes *et al.* (2009). However, a linear time series setting is not the focus of our work.

[2] Note that Theorems 1.3.1, 1.3.2 and Theorem 4.1.3 of Csörgő & Horváth (1997) use slightly different »truncation arguments«. For the sake of consistency, we will use the same truncation arguments for Theorem 2.40 and Theorem 2.43, below.

[3] Cf., e.g., Csörgő & Horváth (1997) and Schmitz (2011).

we may follow Remark 2.28 and replace Assumption E3 in Theorem 2.40 by the weaker Assumption E3'. (The same applies to Corollary 2.42, to Theorem 2.43 and to Corollary 2.50, below.)

Based on (2.3.48) we may use quantiles of the limiting Gumbel distribution as critical values

$$c_n(\alpha) = [-\log(-\log(1-\alpha)/2) + b_{d,n}]/a_n \tag{2.3.49}$$

to test H_0 versus H_A via $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ at a significance level of α. Unfortunately, such critical values provide a reasonable approximation to the actual ones only for large sample sizes and small dimensions whereas this approximation is poor in moderate and higher dimensions.[1] This is indicated in the top panel of Figure 2.1: the approximations $c_n(\alpha)$ are (heavily) decreasing for increasing d due to the $\Gamma(d/2)$ term in the normalizing $b_{d,n}$ sequence. The approximations for larger d are far away from the exact ones, since we know that the exact critical values for $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ must be monotonically increasing in d.

At first sight these facts seem to prohibit to use the principal component based Darling-Erdős-type statistic in practical applications. However, there is another (more direct) approach to derive critical values for $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ which we will discuss before turning to the results under the alternative. This approach relies on Corollary 2.42, below, and yields substantially better critical values $c'_n(\alpha)$ in our simulations than $c_n(\alpha)$ based on (2.3.48). (Cf. the bottom panel of Figure 2.1 and also Gombay & Horváth (1996) for results in a multivariate setup.)

Notice that we use $z_\alpha(X)$ to denote α-quantiles of a random variable X in the next corollary.

Corollary 2.42. *Under the assumptions of Theorem 2.40 and for sequences $h = h_n \geq 1/n$ that fulfill (2.3.40) it holds under H_0 that*

$$\lim_{n\to\infty} P\Big(\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) \leq z_\alpha\big(V(\chi;h)\big)\Big) = \alpha, \tag{2.3.50}$$

where $V(\chi;h)$ and $\chi = \chi(x,d)$ are defined (via Brownian bridges) in (2.3.44) and in (2.3.45). Moreover, it holds that

$$z_\alpha(\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})) = z_\alpha(V(\chi;h)) + o\big((\log\log n)^{-1/2}\big),$$

for all $\alpha \in (0,1)$ and as $n \to \infty$.

This corollary implies that reasonable critical values for testing at the significance level of α are provided directly by $c'_n(\alpha) = z_{1-\alpha}(V(\chi;h))$ which in turn may be computed via the already mentioned approximation (2.3.41) of Vostrikova. Even though the latter corollary ensures asymptotic correctness and a rate for the approximating quantiles of $V(\chi;h)$ we would like, analogously to Gombay & Horváth (1996) and Gombay (2010), to underpin and motivate the good performance of the approximation (2.3.50), that we observe in our simulations, by the next Theorem 2.43 that corresponds

[1] Cf. also the related comments in Horváth *et al.* (1999, p. 106) and in Gombay & Horváth (1996, p. 126).

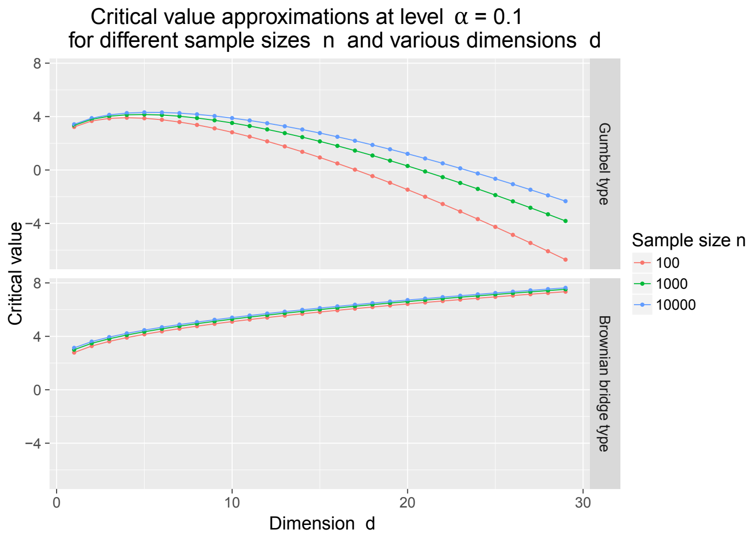

Figure 2.1: The top panel shows the critical values $c_n(\alpha)$ based on the Gumbel type approximations (2.3.48) and (2.3.49) for different sample sizes n and different dimensions d. The bottom panel shows critical values $c'_n(\alpha)$ obtained via the Brownian bridge type approximation (2.3.50) together with Vostrikova's expansion (2.3.41). We use the sequence $h = (\log n)^{3/2}/n$ which is suggested by Gombay & Horváth (1996). Note that $c_n(\alpha)$ and $c'_n(\alpha)$ must converge to each other (and to the actual critical value), as $n \to \infty$, in view of (2.3.38) and of (2.3.48).

to Theorem 4.2 of Torgovitski (2015a). It shows a quite strong rate and thus indicates the closeness of the approximating random variables $V(\chi_n; h_n)$ to $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$.

In our setting we need an extension of results of Gombay & Horváth (1996) to the multivariate m-dependent case and additionally we have to take the estimation of the Hilbert space principal components into account. Note that under the assumptions of Proposition 2.22 this yields also an alternative proof of (2.3.48) via (2.3.38).

Theorem 2.43. *Let the assumptions of Proposition 2.22 hold true and let $h = h_n \geq 1/n$ be a sequence that fulfills (2.3.40). Moreover, let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{1,0}$, which is defined in (2.3.17), and assume that Assumption E3 holds true with $d' = d$. Then it holds under H_0 that, as $n \to \infty$,*

$$\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) = V(\chi_n; h) + o_P\big(\exp(-(\log n)^{1-\varepsilon})\big) \tag{2.3.51}$$

for all $\varepsilon \in (0,1)$ and for some $\chi_n = \chi_n(x,d)$ that fulfill (2.3.47) via (2.3.45). $V(\chi_n; h)$ is defined in (2.3.44).

Remark 2.44. Theorem 3.1 of Gombay & Horváth (1996) corresponds to our Corollary 2.42 where the former relies on Theorem 2.2 of Gombay & Horváth (1996) and which in turn is the multivariate counterpart of our Theorem 2.43. Note, that as a by-product we showed that Theorem 2.43 is not necessarily needed to prove Corollary 2.42 and thus also Theorem 3.1 of Gombay & Horváth (1996) may be shown without invoking Theorem 2.2 of Gombay & Horváth (1996), too.

Asymptotics under the alternative hypothesis

We proceed with the study of »consistency« under the alternative. In this context consistency means that the probability to reject the null hypothesis tends asymptotically to 1 at any level of significance. Towards this end we need to define the functions

$$\mathcal{G}_{g_j}(x) = \int_0^x g_j(y)dy - x\int_0^1 g_j(y)dy \tag{2.3.52}$$

for $1 \leq j \leq \varrho$ (cf., e.g., Section 3.3 of Horváth *et al.*, 2014). The g_j's are the trend-functions from the alternative hypothesis (2.2.4). It holds that all $\mathcal{G}_{g_j}$ are non-constant since the trend functions g_j are assumed to be non-constant. Following a projection based dimension-reduction approach it is evident that detection of (possibly multiple and gradual) changes under H_A is only feasible if those changes are visible in the subspace spanned by the estimates $\{\hat{v}_1, \ldots, \hat{v}_d\}$. Taking Corollary 2.34 into account this motivates the following *»asymptotic visibility and tractability«* assumption.

Assumption G (The interplay between changes and principal components). (λ_j, v_j) are the eigenelements of $\mathscr{C}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}$ with $\alpha = \lim_{n\to\infty} 1/\beta_h$, where $\beta_h, \beta_{\mathcal{K}}, \beta_{\mathcal{G}}$ are defined as in (2.3.30).

i. We assume that Assumption E1 and Assumption E3 hold true with some $d' \leq d$.

ii. We assume that

$$\sum_{k=1}^{\varrho} \mathcal{G}_{g_k}(x)\langle \Delta_k, v_j\rangle \neq 0 \tag{2.3.53}$$

holds true for some $1 \leq j \leq d'$ and some $x \in [0,1]$. (d' is the same as under i.)

A general bottleneck of condition (2.3.53) is that it requires us to know the trend-functions, the change-directions and the eigenfunctions of the contaminated limiting operator. We will get back to this assumption in Remark 2.47 and Remark 2.48, below, but later on, we will consider change-aligned principal components which will allow us to avoid (2.3.53) in all situations under consideration and in this sense be mathematically more convenient.

The next theorem shows that our test with the critical values obtained via Theorem 2.40 or Corollary 2.42 is consistent, i.e. it rejects asymptotically under H_A at any significance level.[1]

Theorem 2.45. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 together with* (1.3.5) *hold true and Assumption G be fulfilled. Then under H_A it holds that, as $n \to \infty$,*

$$(\log\log n)^{-1/2}\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) \xrightarrow{P} \infty.$$

[1] This theorem is a Darling-Erdős-type analogue, e.g., of Berkes *et al.* (2009, Corollary 1) and Aston & Kirch (2012a, Theorem 3.2). See also Hörmann & Kokoszka (2010, Theorem 5.2).

Remark 2.46 (Eigenvalue separation IV). For positive definite $\hat{\mathscr{C}}$ we may follow Remark 2.35 and replace Assumption E3 of Theorem 2.45, which is implicitly assumed in Assumption G, by the weaker Assumption E3'. (This is analogous to Remark 2.41.)

It is important to gain some insight into Assumption G and to verify that it is fulfilled under some reasonable constellations. In the spirit of the observations in Aston & Kirch (2012a, Theorem 4.1) and in Horváth *et al.* (2014) we formulate the following remark which states that »*dominant changes must be eventually detectable*«.[1]

Remark 2.47 (Improved visibility due to dominant trends). Consider the situation of $\varrho \in \mathbb{N}$ trend directions in (2.2.4) and set

$$g_1(t) = ag(t), \tag{2.3.54}$$

$a \in \mathbb{R}_+$, with some piecewise Lipschitz continuous function g that fulfills $g(0) = 0$ but is non-constant. (The other trend-functions g_j, $2 \leq j \leq \varrho$, do not depend on a.) Let (λ_j, v_j) be the eigenelements of the operator $\mathscr{C}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}$ with $\alpha = \lim_{n\to\infty} 1/\beta_h$, where $\beta_h, \beta_{\mathcal{K}}, \beta_{\mathcal{G}}$ are set as in (2.3.30). Furthermore, let Assumption E1 be fulfilled. Our claim is that Assumption G is always satisfied with $d' = 1$ given that a is sufficiently large, i.e. if the trend in direction of the first change Δ_1 is dominant. Hence, we need to show that $\lambda_1 > \lambda_2$ and that

$$\sup_{x\in[0,1]} |\sum_{k=1}^{\varrho} \mathcal{G}_{g_k}(x)\langle \Delta_k, v_1\rangle| > 0 \tag{2.3.55}$$

hold true for a being large.[2] To this end we observe that

$$\lim_{a\to\infty} \mathscr{G}(g_k, g_j)/\mathscr{G}(g_1) = \begin{cases} 1, & k = j = 1, \\ 0, & \text{otherwise}, \end{cases}$$

holds true for $\mathscr{G}$ defined in (2.3.28) and this immediately implies

$$\lim_{a\to\infty} \| \sum_{k,j=1}^{\varrho} \frac{\mathscr{G}(g_k, g_j)}{\mathscr{G}(g_1)} [\Delta_k \otimes \Delta_j] - [\Delta_1 \otimes \Delta_1]\|_{\mathcal{S}} = 0. \tag{2.3.56}$$

Clearly, we have also $\lim_{a\to\infty} \|\alpha\mathscr{C}\|_{\mathcal{S}}/\mathscr{G}(g_1) = 0$ and recalling (2.3.18) together with the normalization assumption on Δ_1 we arrive at $\lim_{a\to\infty} \lambda_1/\mathscr{G}(g_1) = 1$ and at $\lim_{a\to\infty} \lambda_2/\mathscr{G}(g_1) = 0$, i.e. $\lambda_1 > \lambda_2$ holds true for large a. This ensures the separation of the first one-dimensional subspace associated with v_1 and yields $\lim_{a\to\infty} \|s_1 v_1 - \Delta_1\| = 0$ with $s_1 = \text{sign}\langle v_1, \Delta_1\rangle$, e.g., via Lemma 2.3 of Horváth & Kokoszka (2012). Now, due to the orthonormality of Δ_j's, we observe that, as $a \to \infty$,

$$\sup_{x\in[0,1]} |\sum_{k=1}^{\varrho} \mathcal{G}_{g_k}(x)\langle \Delta_k, v_1\rangle|/a = \sup_{x\in[0,1]} |\sum_{k=1}^{\varrho} \mathcal{G}_{g_k}(x)\langle \Delta_k, \Delta_1\rangle|/a + o(1)$$

[1] According to Horváth *et al.* (2014) our Assumption G (i & ii) is always satisfied with $d' = 1$ in the m-approximable setting of Theorem 2.45 if $\varrho = 1$. However, this does not hold true in the m-dependent setting of Theorem 2.45.

[2] The situation is less clear if we have more than one dominant trend-direction. In the latter situation the limiting operator in (2.3.56) would not be of rank-one anymore.

$$= \sup_{x\in[0,1]} |\mathcal{G}_{g_1}(x)|/a + o(1) = \sup_{x\in[0,1]} |\mathcal{G}_g(x)| + o(1).$$

Hence, (2.3.55) follows for large a, i.e. for a dominant trend g_1 (see (2.3.54)).

We continue with two (rather counterintuitive) remarks related to Remark 4.5 of Torgovitski (2015a) regarding the one-directional m-dependent setting. Loosely speaking the message of the first remark is that estimating principal components is better than knowing them. The message of the second is that »overfitting« the bandwidth has also some advantage.

Remark 2.48 (Improved visibility due to perturbation). Theorem 4.1 of Aston & Kirch (2012a) may be restated in our one-directional $\varrho = 1$ situation under the m-dependent setting of Theorem 2.45. To verify this we need only to observe that the *»rank one«* perturbation term $\mathscr{G}(g_1)[\varDelta_1 \otimes \varDelta_1]$ is a positive, self-adjoint Hilbert-Schmidt operator.

This minor adaptation of Theorem 4.1 of Aston & Kirch (2012a) implies an important observation: all one-directional changes that fulfill (2.3.53) of Assumption G with respect to the eigenfunctions of $\mathscr{C}$, will fulfill the same assumption with respect to the eigenfunctions of $\mathscr{C}_{\alpha,\beta}$ with $\alpha = 1/\beta_h$ and $\beta = \mathscr{G}(g_1)$ (cf. Corollary 2.34), i.e. the visibility of a change with respect to the subspace generated by $v_1, \ldots, v_d$ improves under the alternative due to the perturbed estimation of these principal components via $\hat{\mathscr{C}}$.[1]

Remark 2.49 (Improved visibility due to misspecification). Let us consider again the one-directional $\varrho = 1$ situation within the m-dependent setting of Theorem 2.45. It turns out that $\hat{\mathscr{C}}$ based on a misspecified parameter $h \gg m$ is beneficial asymptotically under H_A, as $n \to \infty$, despite being a worse estimate of $\mathscr{C}$ under H_0. The reasoning is similar to Remark 2.47. A larger h yields a smaller value $\alpha = 1/\beta_h$ (cf. Corollary 2.34 and the proof of Theorem 2.45, below) and the rank-one perturbation, i.e. the change contribution, becomes the dominating part in the sum (2.3.18). This implies that v_1 gradually aligns with the subspace that is associated with $\varDelta_1$, as $h \to \infty$, and thus (2.3.53) in Assumption G is asymptotically fulfilled.

Testing via change-aligned principal components

In this paragraph we will show how Assumption G (ii) of Theorem 2.45 may be avoided under H_A while preserving the convergence in Theorem 2.40 and also in Corollary 2.42 under H_0. As introduced in Torgovitski (2015c) we propose to adjust the principal components such that the first direction captures enough information on any multi-directional change. As a consequence we may test with only one direction, i.e. using $d = 1$. This enables us to draw conclusions about infinite dimensional random

[1] For multi-directional changes the verification of Theorem 4.1 of Aston & Kirch (2012a) is more intricate since even the verification of positivity of the perturbation term $\mathscr{C}_{1/\beta_h, \beta_{\mathcal{K}}\beta_{\mathcal{G}}} - \mathscr{C}/\beta_h$ in Corollary 2.34 (with a rank being higher than 1) is generally not obvious anymore. The case of multi-directional changes under m-approximability is also unclear for the same reason.

objects by means of proper chosen one-dimensional subspaces. This is heuristically reasonable because we do not focus on the whole distribution but solely on level-shifts which are less complex objects.

Recall our statistic $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}(k/n)$ and the detector

$$\hat{\mathscr{T}}(x) = \hat{\mathscr{T}}(x;\hat{\boldsymbol{y}}) = |\hat{\Sigma}^{-1/2} S_n(x;\hat{\boldsymbol{y}})|,$$

that is based on projected data $\hat{\boldsymbol{y}}_i = [\langle Y_i, \hat{v}_1\rangle, \ldots, \langle Y_i, \hat{v}_d\rangle]'$. We will consider a slightly different *»change-aligned«* statistic

$$\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}^{\Delta}(k/n) \tag{2.3.57}$$

with a corrected detector

$$\hat{\mathscr{T}}^{\Delta}(x) = \hat{\mathscr{T}}(x;\hat{\boldsymbol{y}}^{\Delta}).$$

The first score in the projected data $\hat{\boldsymbol{y}}_i^{\Delta} = [\langle Y_i, \hat{v}_1^{\Delta}\rangle, \langle Y_i, \hat{v}_2\rangle, \ldots, \langle Y_i, \hat{v}_d\rangle]'$ is modified by a change-aligned direction

$$\hat{v}_1^{\Delta} = z/\|z\| = \tilde{z}/\|\tilde{z}\|,$$

with

$$z = z(\hat{v}_1, \gamma, n) = \hat{v}_1/n^{\gamma} + \hat{s}\hat{u} \tag{2.3.58}$$

or with

$$\tilde{z} = \tilde{z}(\hat{v}_1, \gamma, n) = \hat{v}_1 + \hat{s}(n^{\gamma}\hat{u}).$$

The direction z (or $\tilde{z}$) is constructed as follows:

1. First, we choose any $\hat{k}$ such that $\|S_n(\hat{k}/n;Y)\| = \max_{1\le k<n}\|S_n(k/n;Y)\|$ holds true.
2. This »estimate« is then used to define $\hat{u} = S_n(\hat{k}/n;Y)/n^{1/2}$.
3. The sign $\hat{s} = \operatorname{sign}\langle\hat{v}_1, \hat{u}\rangle$ prevents the cancellation of $\hat{v}_1$ and $\hat{u}$ in (2.3.58) which would occur in cases where both align in the same direction (e.g., in the direction of a one-directional change and given that n^{γ} is not too large).
4. The parameter $\gamma \in (0, 1/2)$ controls the influence of the change-direction estimate $\hat{u}$ under H_0 and under H_A.

If, for the sake of argument, we restrict the considerations to a single abrupt change point setting, then our approach combines a fully-functional estimate of a (possible) change point with a principal component based testing procedure. Consistency and further distributional properties of the fully-functional estimate are derived recently by Aue *et al.* (2015). Interestingly, our proofs do not rely on the consistency of this estimate explicitly which allows us to extend the theory to multi-directional trends where there is no single change point and no single change direction, i.e. where consistency (in the usual sense) becomes meaningless.

We formulate the direct analogues of Theorem 2.40 and of Theorem 2.45. They adapt the results of Corollary 5.1 and of Theorem 5.2 of Torgovitski (2015c) to the Darling-Erdős-type situation. (The derivation of Corollary 2.42 is then straightforward and thus skipped.)

Corollary 2.50. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 together with (1.3.5) hold true. Let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{1,0}$, which is defined in (2.3.17), and assume that Assumption E3 holds true with $d' = d$. Then under H_0 it holds that*

$$\lim_{n\to\infty} P\big(a_n \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) - b_{n,d} \leq x\big) = \exp(-2\exp(-x))$$

for all $x \in \mathbb{R}$. The sequences a_n and $b_{n,d}$ are given in (2.3.39).

Theorem 2.51. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 together with (1.3.5) hold true and Assumption G (i) be fulfilled.*[1] *Then under H_A it holds that, as $n \to \infty$,*

$$(\log\log n)^{-1/2} \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) \xrightarrow{P} \infty.$$

Remark 2.52 (Change-alignment as regularization). Typical principal component based CUSUM approaches have all in common that eigenelements (λ_j, v_j)'s have to be estimated and that inverses $\hat{\lambda}_j^{-1/2}$'s are then used for standardization. (Note that fully-functional CUSUM approaches usually still involve estimation of λ_j's to simulate the limiting distribution.) Estimation of eigenelements is known to be difficult in high(er) dimensions and estimation of eigenvalues close to zero may be troublesome. Change-alignment allow us to rely on a single direction, i.e. to work with $d = 1$, and thus to avoid some of the estimation issues.

[1] Note that we avoid condition (2.3.53) of Assumption G (ii).

A small simulation study 2.4

In this section we provide a simulation study based on synthetically generated data to gain some empirical insight on the performance of the Darling-Erdős-type CUSUM statistics $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}(k/n)$ and on the effect of change-alignment for the corresponding corrected test statistics $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) = \max_{1\le k<n} w(k/n)\hat{\mathscr{T}}^{\Delta}(k/n)$. (Cf. (2.3.13) and (2.3.57).) We consider $L^2[0,1]$-valued functional time series which serve as the standard example for infinite dimensional Hilbert space valued data.

Remark 2.53 (Parameter selection). The principal component based Darling-Erdős-type CUSUM statistics $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ and $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta})$ offer a class of tests where we have several degrees of freedom that will be specified later on:

1. We have to choose the subspace dimension $d \in \mathbb{N}_H$ that is used for testing, i.e. we need to select the number of long run principal components (λ_j, v_j).

2. We need to specify the long run covariance operator estimate $\hat{\mathscr{C}}$ which defines the estimates of the long run principal components $(\hat{\lambda}_j, \hat{v}_j)$. This requires the selection of a kernel $\mathscr{K}$ and of a bandwidth h.

3. We have to decide whether critical values are computed relying on a Gumbel type approximation, i.e. using (2.3.48) and (2.3.49), or via Brownian bridge type approximations, i.e. relying on (2.3.50) in combination with (2.3.41). (The latter approach involves a bandwidth selection as well.)

4. Finally, we have to specify the correction parameter γ (cf. (2.3.58)).

Performance for synthetic data

We consider the $L^2[0,1]$-version of the Hilbert space signal plus noise model (2.2.1):

$$Y_i(t) = m_i(t) + \varepsilon_i(t),$$

$t \in [0,1]$, $1 \le i \le n$, with the $L^2[0,1]$-valued means $\{m_i(\cdot)\}$ and the $L^2[0,1]$-valued noise $\{\varepsilon_i(\cdot)\}$. We show simulation results for the following two settings: in the first setting we assume i.i.d. ε_i's which are modeled as paths of independent standard Brownian motions (cf. Remark 1.5). In the second setting we assume the ε_i's to be a functional AR(1) time series which is generated by an $L^2[0,1]$-valued i.i.d. sequence and where the latter innovations are also modeled by paths of independent standard Brownian motions. The implementation details are similar to Torgovitski (2015a). We simulate the paths of a Brownian motion using the {e1071}-R-package. For our further implementation we use the approach of Berkes *et al.* (2009) and work with the convenient {fda}-R-package to represent the discrete data as functional objects using 25 B-spline basis functions (of order 4). A detailed description of the generation process for the functional AR(1) time series is provided in Torgovitski (2016, Section 5) and involves a usual burn-in approach.[1] We consider a functional autoregressive time series that has a *»parabolic kernel«*

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

$t, s \in [0,1]$, where $\psi > 0$ is chosen such that the $L^2([0,1]\times[0,1])$ norm of $\Psi(\cdot,\cdot)$ is $1/4$. (Cf., e.g., Horváth & Kokoszka, 2012, p. 201 and Torgovitski, 2016.) In the first i.i.d. setting we may rely on the theoretical framework of Chapter 2 under Assumption M1 and work with static principal components, i.e. with eigenelements of the covariance operator. This corresponds to $h = 0$ in Assumption K1. The second autoregressive setting fits into the framework of Assumption M2 and in this case we work with long run principal components, i.e. with eigenelements of the long run covariance operator. In that case, we choose the kernel $\mathscr{K}(x) = \mathbb{1}_{[-1,1]}(x)$ and set the bandwidth for demonstration purposes to $h = 2$.

We consider the following simulation scenarios where we assume (2.2.3) with $m_1 = \ldots = m_n = 0$ under the null hypothesis and (2.2.4) with $m = 0$ under the alternative. Note that our scenarios contain the one-directional, i.e. $\varrho = 1$, and the two-directional, i.e. $\varrho = 2$, change models involving abrupt, epidemic or linear changes (see also Remark 2.8):

H0 ($\varrho = 0$): no change,

SIN ($\varrho = 1$): $\Delta_1(t) = c\sin(t)$ with an abrupt model $g_1(x) = g_A(x;1/2)/\xi$,

T ($\varrho = 1$): $\Delta_1(t) = \tilde{c}\,t$ with an abrupt model $g_1(x) = g_A(x;1/2)/\xi$,

BM3 ($\varrho = 1$): $\Delta_1(t) = u_3(t)$ with an abrupt model $g_1(x) = g_A(x;1/2)/\xi$,

BM10 ($\varrho = 1$): $\Delta_1(t) = u_{10}(t)$ with an abrupt model $g_1(x) = g_A(x;1/2)/\xi$,

BM15 ($\varrho = 1$): $\Delta_1(t) = u_{15}(t)$ with an abrupt model $g_1(x) = g_A(x;1/2)/\xi$,

MULT-E ($\varrho = 2$): $\Delta_1(t) = u_{10}(t)$ and $\Delta_2(t) = u_{15}(t)$ with two epidemic change functions $g_1(x) = g_E(x;1/4,2/3)/(4\xi)$, $g_2(x) = g_E(x;1/5,1/3)/(4\xi)$,

MULT-G ($\varrho = 2$): $\Delta_1(t) = u_{10}(t)$ and $\Delta_2(t) = u_{15}(t)$ with linear and epidemic change functions $g_1(x) = g_L(x;1/5,1/2)/(4\xi)$, $g_2(x) = g_E(x;1/3,2/3)/(4\xi)$.

The orthonormal functions $u_j(t) = 2^{1/2}\sin((j-1/2)\pi t)$, $t \in [0,1]$, stem from the Karhunen-Loève expansion of a Brownian motion, i.e. are exactly the static principal components in our i.i.d. scenario. The constants $c, \tilde{c} > 0$ in the SIN and T settings are implicitly determined by our normalization assumption $\|\Delta_1\| = 1$. Furthermore, we use a scaling constant $\xi = 5$ for the i.i.d. setting and $\xi = 7.5$ for the functional autoregressive case. (We use different values to demonstrate the effect of change-alignment. The power is too high for this purpose using $\xi = 5$ in the autoregressive setting because the Bartlett-type estimate enhances the detectability.)

[1] In our simulations we use a burn-in length of 50.

Figures 2.2 - 2.5 show the performance of the Darling-Erdős-type tests for different dimensions $d = 1, 2, 3, 4, 5$ of principal component subspaces, for different sample sizes $n = 100, 500, 1000$ and for various types of change models that were specified above. The empirical rejection rates are reported based on 1000 repetitions and the theoretical significance level is set to $\alpha = 10\%$. Also, note that the critical values are obtained, as in Figure 2.1, via the Brownian bridge type approximation (2.3.50) together with Vostrikova's expansion (2.3.41) using the sequence $h = (\log n)^{3/2}/n$ as is suggested by Gombay & Horváth (1996).

Figure 2.2 and Figure 2.3 show the results for the i.i.d. setting where the results for the latter were obtained using the change-alignment method. Figure 2.4 and Figure 2.5 contain the corresponding results for the functional AR(1) case where, again, the alignment approach is shown in the latter figure. The first main observation is that the Darling-Erdős-type testing approaches work well for reasonable sample sizes: we have approximately 10% of rejections under the null hypothesis and we have asymptotic consistency under a variety of alternatives, namely abrupt, epidemic, gradual and multiple changes. We observe that the change-alignment has a positive effect under the alternative hypothesis which is stronger than the negative impact under the null hypothesis. Particularly, we see the most benefit if we work in the i.i.d. setting with static principal components and if we use only the first dimension. Furthermore, Figure 2.2 and Figure 2.3 show that the effect on higher frequency changes (e.g. for BM10 and BM15) is stronger than on lower frequency changes (e.g. for SIN, T and BM3). Note that the effect of change-alignment is less visible in the functional AR(1) setting of Figure 2.2 and Figure 2.3 than in the i.i.d. setting of Figure 2.2 and Figure 2.3. The reason is that the long run principal components already tend to align in the direction of the change (cf., e.g., Horváth *et al.*, 2014). Nevertheless, we see a substantial increase in the power, e.g., for $d = 1$ and $n = 1000$.

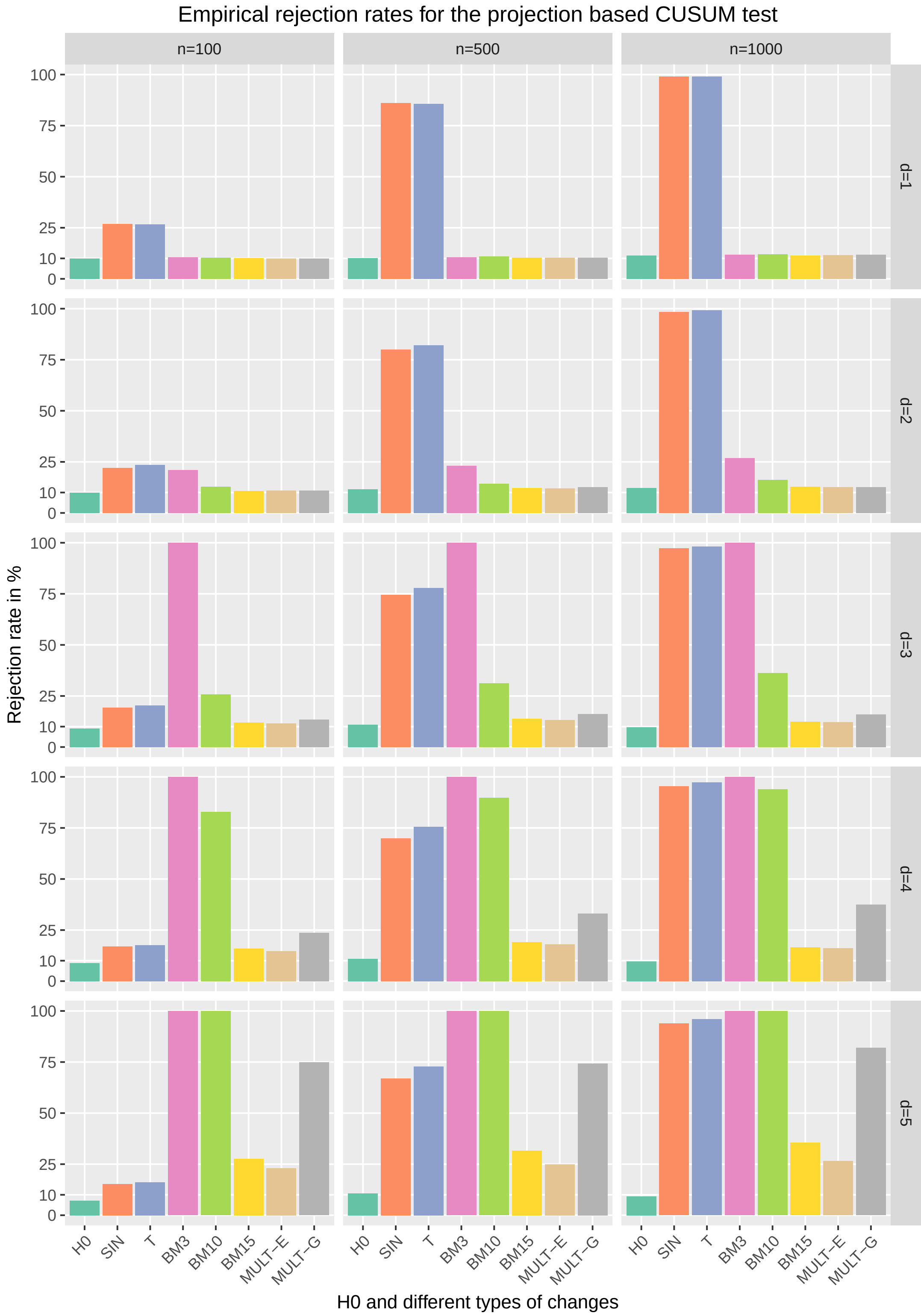

Figure 2.2: Empirical rejection rates for the Darling-Erdős-type CUSUM test based on $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ and in the setting of a functional i.i.d. error sequence $\{\varepsilon_i\}$, working with (empirical) static principal components. As a critical value we use $c_n'(10\%)$.

Figure 2.3: Empirical rejection rates for the change-aligned Darling-Erdős-type CUSUM test based on $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta})$ and in the setting of a functional i.i.d. error sequence $\{\varepsilon_i\}$, working with (empirical) static principal components. We use $\gamma = 0.49$ for alignment and as a critical value we use $c_n'(10\%)$.

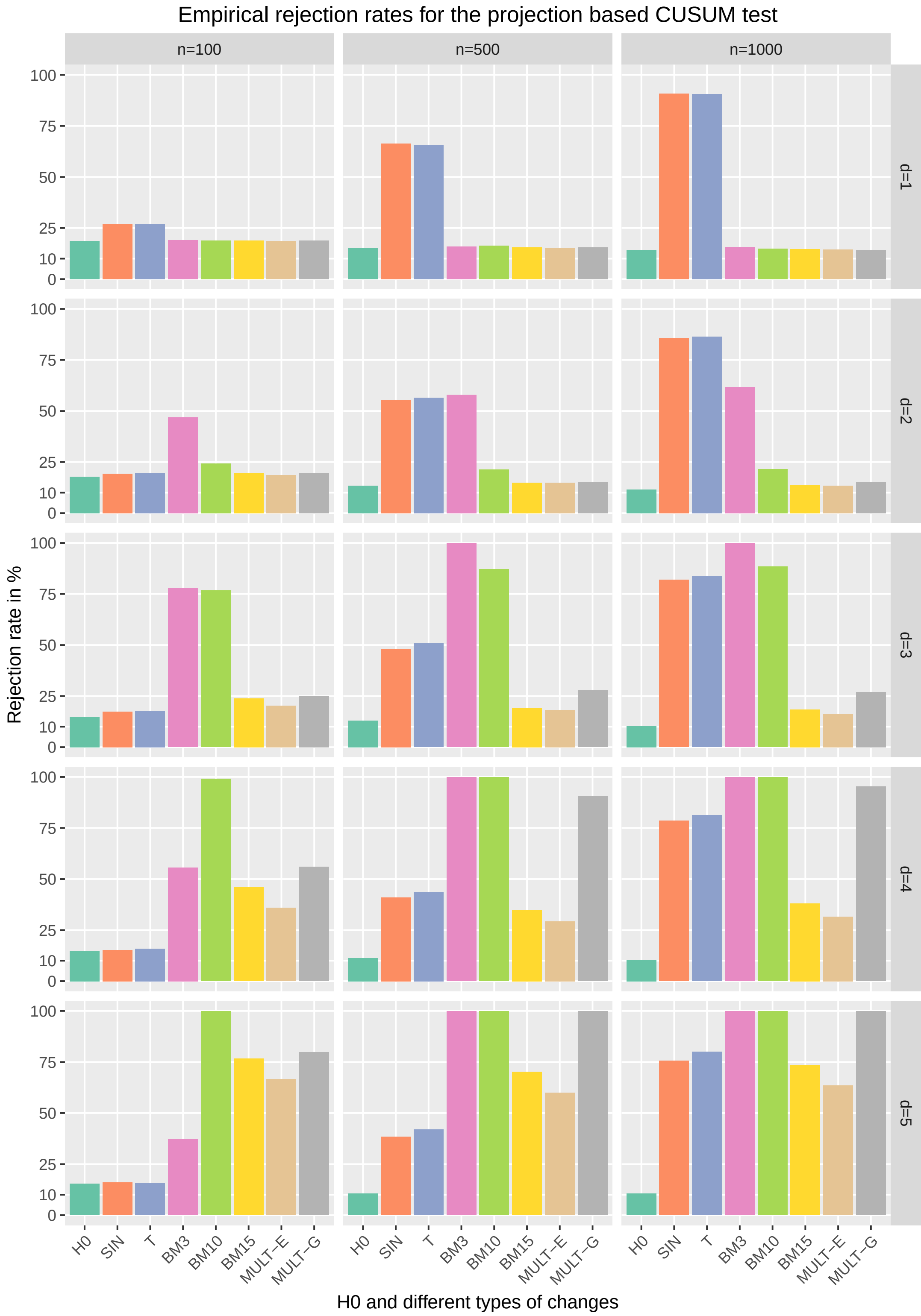

Figure 2.4: Empirical rejection rates for the Darling-Erdős-type CUSUM test based on $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ and in the setting of a functional AR(1) error sequence $\{\varepsilon_i\}$, working with (empirical) long run principal components. The principal components are computed using the kernel $\mathscr{K}(x) = \mathbb{1}_{[-1,1]}(x)$ with bandwidth $h = 2$. As a critical value we use $c_n'(10\%)$.

Figure 2.5: Empirical rejection rates for the change-aligned Darling-Erdős-type CUSUM test based on $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta})$ and in the setting of a functional AR(1) error sequence $\{\varepsilon_i\}$, working with (empirical) long run principal components. The principal components are computed using the kernel $\mathscr{K}(x) = \mathbb{1}_{[-1,1]}(x)$ with bandwidth $h = 2$. We use $\gamma = 0.49$ for alignment and as a critical value we use $c_n'(10\%)$.

Proofs 2.5

Proofs for Subsection 2.3.2 (on covariance estimation) 2.5.1

For convenience of the reader, let us recall the definition (2.3.9) of $\hat{\mathscr{C}}$ and also define a corresponding analogue $\hat{\mathscr{E}}$ as follows:

$$\hat{\mathscr{C}} = \sum_{r=-n}^{n} \mathscr{K}(r/h)\hat{\mathscr{C}}_r, \qquad \hat{\mathscr{C}}_r = \begin{cases} \sum_{i=1}^{n-r}[\hat{\varepsilon}_i \otimes \hat{\varepsilon}_{i+r}]/n, & r \geq 0, \\ \sum_{i=1}^{n+r}[\hat{\varepsilon}_{i-r} \otimes \hat{\varepsilon}_i]/n, & r < 0, \end{cases}$$
$$\hat{\mathscr{E}} = \sum_{r=-n}^{n} \mathscr{K}(r/h)\hat{\mathscr{E}}_r, \qquad \hat{\mathscr{E}}_r = \begin{cases} \sum_{i=1}^{n-r}[\varepsilon_i \otimes \varepsilon_{i+r}]/n, & r \geq 0, \\ \sum_{i=1}^{n+r}[\varepsilon_{i-r} \otimes \varepsilon_i]/n, & r < 0. \end{cases} \tag{2.5.1}$$

Note that the estimate $\hat{\mathscr{E}}$ is based on the unobservable (mean zero) errors ε_i whereas $\hat{\mathscr{C}}$ is based on the observable (empirically centered) residuals $\hat{\varepsilon}_i = Y_i - \bar{Y}_n$.

We begin with estimation of the long run covariance structure under the null-hypothesis and continue with the proofs for the test statistic later on. The next proof of Proposition 2.22 is an application of rather standard techniques.

Proof of Proposition 2.22. We give a detailed Hilbert space version of the proof of Torgovitski (2015a, Proposition 3.1) where we distinguish between the cases $\kappa \in (2,4)$ and $\kappa \geq 4$. First, let us assume that $\kappa \in (2,4)$. Using a Hilbert space *Marcinkiewicz-Zygmund* type strong law of large numbers (cf., e.g., Remark 1 and Theorem 9 of Dedecker & Merlevède, 2008) we observe that

$$\|\hat{\mathscr{E}} - \mathscr{C}_{1,0}\|_{\mathcal{S}} = \mathcal{O}\big(n^{-1+2/\kappa}\big) \quad \text{a.s.} \tag{2.5.2}$$

holds true as $n \to \infty$. Next, assume that $\kappa \geq 4$. Using stationarity, m-dependence and Proposition 2.65 (below) with $\nu = 0$ we see that $E\|\sum_{i=1}^{n-r}\varepsilon_i \otimes \varepsilon_{i+r}\|_{\mathcal{S}}^2 = \mathcal{O}(n)$ for any fixed $r \in \{1,\dots,h\}$. Hence, proceeding straightforward as for (2.5.2) we obtain $\|\hat{\mathscr{E}} - \mathscr{C}_{1,0}\|_{\mathcal{S}} = \mathcal{O}_P\big(n^{-1/2}\big)$. Finally, we observe $\|\hat{\mathscr{C}} - \hat{\mathscr{E}}\|_{\mathcal{S}} = \mathcal{O}_P(n^{-1})$, e.g. by mimicking the corresponding steps in the proof of Horváth *et al.* (2013, Theorem 2) under the Assumption M1 and Assumption K1, which completes our proof.[1] □

Proof of Theorem 2.24. This is a generalized and extended Hilbert space version of the proof of Torgovitski (2016, Theorem 4.1). First, we observe $\|\hat{\mathscr{E}}_0 - \mathscr{C}_0\| = \mathcal{O}_P(n^{-1/2})$ on using Proposition 2.65 or by invoking Hörmann & Kokoszka (2010, Theorem 3.1). The latter can be applied in our case of non-causality in a straightforward manner via modifications similar to (2.5.66), below. Hence, it remains to investigate only the (lagged) long run part of the estimate $\hat{\mathscr{C}}$. Moreover, due to symmetry it suffices to consider only the term $\hat{\mathscr{C}}^f = \sum_{i=1}^{n} \mathscr{K}(i/h)\hat{\mathscr{C}}_i$ of (2.5.1).

[1] Recall that we do not assume a shift representation in this case and that we assume the bandwidth $h \geq m$ to be fixed.

In the following we will work with the decomposition:

$$\begin{aligned}\hat{\mathscr{C}}^f - \mathscr{C}^f &= \hat{\mathscr{C}}^f - \mathscr{E}^f \\ &= \left[\hat{\mathscr{E}}^{f,(m)} - E\hat{\mathscr{E}}^{f,(m)}\right] + \left[E\hat{\mathscr{E}}^{f,(m)} - \mathscr{E}^{f,(m)}\right] \\ &\quad + \left[\hat{\mathscr{E}}^f - \hat{\mathscr{E}}^{f,(m)}\right] + \left[\mathscr{E}^{f,(m)} - \mathscr{E}^f\right] + \left[\hat{\mathscr{C}}^f - \hat{\mathscr{E}}^f\right] \\ &=: A_1 + A_2 + A_3 + A_4 + A_5,\end{aligned}$$

where we set

$$\begin{aligned}&\mathscr{C}^f = \mathscr{E}^f = \sum_{i=1}^{\infty} \mathscr{E}_i, && \mathscr{C}^{f,(m)} = \mathscr{E}^{f,(m)} = \sum_{i=1}^{m} \mathscr{E}_i^{(m)}, \\ &\hat{\mathscr{C}}^f = \sum_{i=1}^{n} \mathscr{K}(i/h)\hat{\mathscr{C}}_i, && \hat{\mathscr{E}}^f = \sum_{i=1}^{n} \mathscr{K}(i/h)\hat{\mathscr{E}}_i, && \hat{\mathscr{E}}^{f,(m)} = \sum_{i=1}^{n} \mathscr{K}(i/h)\hat{\mathscr{E}}_i^{(m)}, \\ &\hat{\mathscr{C}}_i = \sum_{j=1}^{n-i}[\hat{\varepsilon}_j \otimes \hat{\varepsilon}_{j+i}]/n, && \hat{\mathscr{E}}_i = \sum_{j=1}^{n-i}[\varepsilon_j \otimes \varepsilon_{j+i}]/n, && \hat{\mathscr{E}}_i^{(m)} = \sum_{j=1}^{n-i}[\varepsilon_j^{(m)} \otimes \varepsilon_{j+i}^{(m)}]/n,\end{aligned}$$

using $\mathscr{E}_i = E[\varepsilon_0 \otimes \varepsilon_i]$ and $\mathscr{E}_i^{(m)} = E[\varepsilon_0^{(m)} \otimes \varepsilon_i^{(m)}]$. Furthermore, we use $\hat{\varepsilon}_i = Y_i - \bar{Y}_n$ and the superscript indicates the m-dependent quantities (cf. (1.3.3)). Let us briefly explain the terms A_1 - A_5 beginning with the last term: A_5 captures the bias of the transition from the observable Y_j's to the unobservable ε_j's. The term A_4 is then the bias due to the neglected covariances for lags larger than m, whereas A_3 captures the bias due to estimation of the covariance structure of the m-dependent approximations $\{\varepsilon_i^{(m)}\}_{i\in\mathbb{Z}}$. To derive bounds for these three terms we borrow results of Horváth *et al.* (2013, proof of Theorem 2) with the necessary straightforward modifications in view of non-causality. The crucial calculations are for the first terms A_1 and A_2 via $E\|A_1\|_{\mathcal{S}}^2$ and $E\|A_2\|_{\mathcal{S}}$ which are the *»variance«* and the bias of the estimate $\hat{\mathscr{E}}^{f,(m)}$. Note that our main extension of the proof of Horváth *et al.* (2013) is the introduction of the additional term $\hat{\mathscr{E}}^{f,(m)}$ and that we use a dynamic approximation where the dependency of $\hat{\mathscr{E}}^{f,(m)}$ increases with the sample size, since we require $m = m_n \to \infty$ as $n \to \infty$. (Recall also that the sequences m and $h = h_n$ fulfill $m = o(h)$ as $n \to \infty$.)

Rates for A_1: Evaluating $E\|A_1\|_{\mathcal{S}}^2$ we have at most $(nh)^2$ non-zero terms (which all contain the factor n^{-2}) and a naive bound is thus $E\|A_1\|_{\mathcal{S}}^2 = \mathcal{O}(h^2)$. The point of subsequent considerations is that it may be improved to $\mathcal{O}(h^3/n) = \mathcal{O}(h^2(h/n))$ by taking m-dependence into account. First, we observe via the Cauchy-Schwarz inequality that

$$\begin{aligned}\left[E\langle \varepsilon_k^{(m)} \otimes \varepsilon_{k+i}^{(m)}, \varepsilon_r^{(m)} \otimes \varepsilon_{r+j}^{(m)}\rangle_{\mathcal{S}}\right]^2 &\le E\|\varepsilon_k^{(m)} \otimes \varepsilon_{k+i}^{(m)}\|_{\mathcal{S}}^2 E\|\varepsilon_r^{(m)} \otimes \varepsilon_{r+j}^{(m)}\|_{\mathcal{S}}^2 \\ &= \left[E\|\varepsilon_k^{(m)}\|_H^2\|\varepsilon_{k+i}^{(m)}\|_H^2\right]\left[E\|\varepsilon_r^{(m)}\|_H^2\|\varepsilon_{r+j}^{(m)}\|_H^2\right] \\ &\le \left[E\|\varepsilon_0^{(m)}\|_H^4\right]^2 \\ &= \left[E\|\varepsilon_0\|_H^4\right]^2.\end{aligned}$$

(Cf. analogous arguments in the proof of Lemma 4 in the arXiv:1210.7192v3 version of Hörmann *et al.*, 2015.) Hence, in view of $|\mathscr{K}(x)| \leq 1$ we obtain

$$
\begin{aligned}
E\|\hat{\mathscr{E}}^{f,(m)}\|_{\mathcal{S}}^2 &\leq \sum_{i,j=1}^{\lfloor ch \rfloor} |E\langle \hat{\mathscr{E}}_i^{(m)}, \hat{\mathscr{E}}_j^{(m)} \rangle_{\mathcal{S}}| \\
&\leq \sum_{i,j=1}^{\lfloor ch \rfloor} \sum_{k,r=1}^{n} |E\langle \varepsilon_k^{(m)} \otimes \varepsilon_{k+i}^{(m)}, \varepsilon_r^{(m)} \otimes \varepsilon_{r+j}^{(m)} \rangle_{\mathcal{S}}|/n^2 \\
&\leq E\|\varepsilon_0\|^4 \sum_{i,j=1}^{\lfloor ch \rfloor} \sum_{k,r=1}^{n} \delta_{k,r}^{i,j}/n^2
\end{aligned}
$$

for some $c > 0$. Here, we take into account that $\mathscr{K}(x) = 0$ for $x > \tilde{c}$ for some $\tilde{c} > 0$ and we define $\delta_{k,r}^{i,j}$ as

$$
\delta_{k,r}^{i,j} := \begin{cases} 0, & r-(k+i) > m,\ r \geq k, \\ 0, & k-(r+j) > m,\ r \leq k, \\ 1, & r-(k+i) \leq m,\ r \geq k, \\ 1, & k-(r+j) \leq m,\ r \leq k, \end{cases}
$$

which count the zero terms that appear due to m-dependence and which simplify to

$$
\delta_{k,r}^{i,j} = \begin{cases} 1, & 0 \leq r-k \leq m+i, \\ 1, & 0 \leq k-r \leq m+j, \\ 0, & \text{otherwise.} \end{cases}
$$

Since the values $\delta_{k,r}^{i,j}$ depend only on i, j and on the difference of $|k-r|$, we have

$$
\begin{aligned}
\sum_{i,j=1}^{\lfloor ch \rfloor} \sum_{k,r=1}^{n} \delta_{k,r}^{i,j} &\leq \sum_{i,j=1}^{\lfloor ch \rfloor} \sum_{z=1}^{n} \left[\sum_{q=0}^{n} \delta_{z,z+q}^{i,j} + \sum_{q=0}^{n} \delta_{z+q,z}^{i,j} \right] \\
&= \sum_{i,j=1}^{\lfloor ch \rfloor} \sum_{z=1}^{n} \left[\sum_{q=0}^{n} \delta_{0,q}^{i,j} + \sum_{q=0}^{n} \delta_{q,0}^{i,j} \right] \\
&\leq n \sum_{i,j=1}^{\lfloor ch \rfloor} (2m+i+j) \leq c' n h^3
\end{aligned}
$$

for some $c' > 0$. (Recall that $m \leq h$.) From above considerations we obtain

$$
E\|\hat{\mathscr{E}}^{f,(m)} - E\hat{\mathscr{E}}^{f,(m)}\|_{\mathcal{S}}^2 \leq E\|\hat{\mathscr{E}}^{f,(m)}\|_{\mathcal{S}}^2 = \mathcal{O}(h^3/n).
$$

Rates for A_2: Using $\lim_{n\to\infty} |\mathscr{K}(m/h) - 1|(m/h)^{-s} = \mathcal{O}(1)$ and stationarity, we see that, as $n \to \infty$,

$$
\begin{aligned}
&\|E\hat{\mathscr{E}}^{f,(m)} - \mathscr{E}^{f,(m)}\|_{\mathcal{S}} \\
&= \|\sum_{i=1}^{m} [\mathscr{K}(i/h)(1-i/n) - 1] \mathscr{E}_i^{(m)}\|_{\mathcal{S}}
\end{aligned}
$$

$$\leq c\Big[\sum_{i=1}^{m} i\|\mathscr{E}_i^{(m)}\|_{\mathcal{S}}/n + \Big(\max_{1\leq r\leq m}|\mathscr{K}(r/h)-1|(h/r)^{\varsigma}\Big)(1/h)^{\varsigma}\sum_{i=1}^{m} i^{\varsigma}\|\mathscr{E}_i^{(m)}\|_{\mathcal{S}}\Big]$$
$$= \mathcal{O}\Big((1/n)\sum_{i=1}^{m} i\|\mathscr{E}_i^{(m)}\|_{\mathcal{S}} + (1/h)^{\varsigma}\sum_{i=1}^{m} i^{\varsigma}\|\mathscr{E}_i^{(m)}\|_{\mathcal{S}}\Big) \tag{2.5.3}$$

for some $c > 0$.

Rates for A_3: By Horváth *et al.* (2013, proof of Theorem 2) we observe that

$$E\|\hat{\mathscr{E}}^f - \hat{\mathscr{E}}^{f,(m)}\|_{\mathcal{S}} = \mathcal{O}\Big(m\Big[E\|\varepsilon_0-\varepsilon_0^{(m)}\|^2\Big]^{1/2} + \sum_{i=m+1}^{\infty}\Big[E\|\varepsilon_0-\varepsilon_0^{(i)}\|^2\Big]^{1/2}\Big)$$
$$= \mathcal{O}\Big(m\delta(m) + \sum_{i=m+1}^{\infty}\delta(i)\Big).$$

Rates for A_4: Similarly to (2.5.66), below, we have

$$\mathscr{E}_i^{(m)} - \mathscr{E}_i = E\Big(\varepsilon_0^{(m)}\otimes\Big[\varepsilon_i^{(m)}-\varepsilon_i\Big] + \Big[\varepsilon_0^{(m)}-\varepsilon_0\Big]\otimes\varepsilon_i\Big). \tag{2.5.4}$$

Using this decomposition (2.5.4) twice and taking the independence of $\varepsilon_0^{(i)}$ and $\varepsilon_i^{(i)}$ into account, we get via stationarity that

$$\|\mathscr{E}_f^{(m)} - \mathscr{E}_f\|_{\mathcal{S}}$$
$$\leq \|\sum_{i=1}^{m}(\mathscr{E}_i^{(m)}-\mathscr{E}_i)\|_{\mathcal{S}} + \|\sum_{i=m+1}^{\infty}\mathscr{E}_i\|_{\mathcal{S}}$$
$$= \mathcal{O}\Big(m\Big[E\|\varepsilon_0-\varepsilon_0^{(m)}\|^2\Big]^{1/2} + \|\sum_{i=m+1}^{\infty}E\Big[(\varepsilon_0-\varepsilon_0^{(i)})\otimes\varepsilon_i + \varepsilon_0^{(i)}\otimes(\varepsilon_i-\varepsilon_i^{(i)})\Big]\|_{\mathcal{S}}\Big)$$
$$= \mathcal{O}\Big(m\Big[E\|\varepsilon_0-\varepsilon_0^{(m)}\|^2\Big]^{1/2} + \sum_{i=m+1}^{\infty}\big[E\|\varepsilon_0-\varepsilon_0^{(i)}\|^2\big]^{1/2}\Big)$$
$$= \mathcal{O}\Big(m\delta(m) + \sum_{i=m+1}^{\infty}\delta(i)\Big),$$

as $n\to\infty$.

Rates for A_5: Finally, we recall that

$$\|\hat{\mathscr{C}}^f - \hat{\mathscr{E}}^f\|_{\mathcal{S}} = \mathcal{O}_P(h/n) = \mathcal{O}_P((h/n)^{1/2})$$

which can be shown (after obvious modifications) as in the proof of Theorem 2 of Horváth *et al.* (2013).

To finish our proof we have to replace the m-dependent $\mathscr{E}_i^{(m)}$ in (2.5.3) by their original counterparts $\mathscr{E}_i$. This is straightforward and we observe via (2.5.4) that by stationarity and by the Cauchy-Schwarz inequality it holds that

$$\sum_{i=1}^{m} i^{\varsigma}\|(\mathscr{E}_i^{(m)}-\mathscr{E}_i)\|_{\mathcal{S}} \leq 2(E\|\varepsilon_0\|^2)^{1/2}\sum_{i=1}^{m} i^{\varsigma}(E\|\varepsilon_0^{(m)}-\varepsilon_0\|^2)^{1/2}$$
$$= \mathcal{O}\big(m^{\varsigma+1}\delta(m)\big)$$

for any $\varsigma\in[1,\infty)$. (Note that $h^{-\varsigma}m^{\varsigma+1}\delta(m) = \mathcal{O}(m\delta(m))$.) □

As already mentioned in Remark 2.25 the above proof is related to Hörmann *et al.* (2015). The convergence rate is comparable but the incorporation of m-dependence is different. Horváth *et al.* (2014) and, more recently, Berkes *et al.* (2016) showed $\mathcal{O}_P(h/n)$ rates under the same dependence concept for $H = L^2[0,1]$. They worked with slightly different assumptions using m-dependent approximations with some fixed but arbitrary large $m \in \mathbb{N}$ (and by far more subtle arguments).

Proof of Corollary 2.26. The assertion follows from Theorem 2.24 by setting $m = \lfloor h^{\varepsilon'} \rfloor$ for some $0 < \varepsilon' < 1$. □

Proof of Corollary 2.27. Due to Assumption E3, we may formally assume that $\hat{\lambda}_j > 0$ holds true for $1 \leq j \leq d' \leq d$ since the probability of this situation is asymptotically 1 as $n \to \infty$. To prove (2.3.27) it is sufficient to show the following two bounds

$$\|\sum_{j=1}^{d'} \lambda_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j) - \sum_{j=1}^{d'} \lambda_j^{-1/2}(v_j \otimes v_j)\|_{\mathcal{S}} = \mathcal{O}_P(n^{-\varepsilon_2}) \tag{2.5.5}$$

and

$$\|\sum_{j=1}^{d'} \hat{\lambda}_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j) - \sum_{j=1}^{d'} \lambda_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j)\|_{\mathcal{S}} = \mathcal{O}_P(n^{-\varepsilon_2}) \tag{2.5.6}$$

for some $\varepsilon_2 > 0$. (Notice that we use $\varepsilon_1 > 0$ in (2.3.26).) We begin with the first bound (2.5.5) and observe that

$$\begin{aligned}
&\|(\hat{v}_j \otimes \hat{v}_j) - (v_j \otimes v_j)\|_{\mathcal{S}} \\
&= \|(\hat{s}_j\hat{v}_j \otimes \hat{s}_j\hat{v}_j) - (v_j \otimes v_j)\|_{\mathcal{S}} \\
&\leq \|(\hat{s}_j\hat{v}_j \otimes \hat{s}_j\hat{v}_j) - (\hat{s}_j\hat{v}_j \otimes v_j)\|_{\mathcal{S}} + \|(\hat{s}_j\hat{v}_j \otimes v_j) - (v_j \otimes v_j)\|_{\mathcal{S}} \\
&= \|\hat{s}_j\hat{v}_j\|_H \|\hat{s}_j\hat{v}_j - v_j\|_H + \|\hat{s}_j\hat{v}_j - v_j\|_H \|v_j\|_H,
\end{aligned} \tag{2.5.7}$$

where $\hat{s}_j = \operatorname{sign}\langle \hat{v}_j, v_j\rangle$. Hence, relation (2.5.5) follows by combining (2.5.7) with (2.3.26). For the second bound (2.5.6) we get

$$\|\sum_{j=1}^{d'} \hat{\lambda}_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j) - \sum_{j=1}^{d'} \lambda_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j)\|_{\mathcal{S}}^2 = \sum_{j=1}^{d'} (\hat{\lambda}_j^{-1/2} - \lambda_j^{-1/2})^2 \tag{2.5.8}$$

via Parseval's identity (cf., e.g., Lemma 3.2 of Reimherr, 2015). Now, since we are under Assumption E3, relation (2.5.6) follows from (2.3.25) together with (2.5.8). (Recall that Proposition 2.22 provides a polynomial rate of convergence for the estimates $\hat{\mathscr{C}}$ and thus also for their eigenelements in Corollary 2.27 under the Assumption M1. Analogously, Corollary 2.26 ensures a polynomial rate of convergence of the operators and of their eigenelements under the Assumption M2.) □

The following proposition is implicitly contained in the proof of Proposition 4.2 in Torgovitski (2015c) and is based on the findings of Reimherr (2015, Lemmas 3.1 and 3.2). It provides a mathematically convenient relation between the convergence of $\hat{\mathscr{C}}^{-1/2,d'} - \mathscr{C}_{\alpha,\beta}^{-1/2,d'}$ and the convergence of $\hat{\mathscr{C}} - \mathscr{C}_{\alpha,\beta}$.

Proposition 2.54. *Consider $\hat{\mathscr{C}}$ and $\mathscr{C}_{\alpha,\beta}$ as defined in (2.3.9) and in (2.3.18) with spectral decompositions in (2.3.10) and in (2.3.19). Let $\hat{\mathscr{C}}^{-1/2,d}$ and $\mathscr{C}_{\alpha,\beta}^{-1/2,d}$ be their truncated (inverse square root) counterparts as defined in (2.3.16) and in (2.3.20). Let Assumption E1 and Assumption E3' hold true. Moreover, assume that $\hat{\mathscr{C}}$ is positive definite with $\hat{\lambda}_{d'} > \hat{\lambda}_{d'+1}$. Then it holds that*

$$\|\sum_{j=1}^{d'} \lambda_j^{-1/2}(v_j \otimes v_j) - \sum_{j=1}^{d'} \lambda_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j)\|_{\mathcal{S}}^2 \leq f_1(d')\|\hat{\mathscr{C}} - \mathscr{C}_{\alpha,\beta}\|_{\mathcal{S}}^2$$

and

$$\|\sum_{j=1}^{d'} \lambda_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j) - \sum_{j=1}^{d'} \hat{\lambda}_j^{-1/2}(\hat{v}_j \otimes \hat{v}_j)\|_{\mathcal{S}}^2 \leq f_2(d')\|\hat{\mathscr{C}} - \mathscr{C}_{\alpha,\beta}\|_{\mathcal{S}}^2,$$

where

$$f_1(r) = 4r\Big((\lambda_r - \lambda_{r+1})^{-2} + (\hat{\lambda}_r - \hat{\lambda}_{r+1})^{-2} + \lambda_r^{-2}\Big)/\lambda_r, \quad f_2(r) = r/\min\{\hat{\lambda}_r, \lambda_r\}^3.$$

Proof of Proposition 2.54. This follows by carefully repeating Lemmas 3.1 and 3.2 of Reimherr (2015). (Note that we consider the general case of parameters α and β. Under H_0 we have $\alpha = 1$ and $\beta = 0$ and Assumption E1 is then automatically fulfilled for the long run covariance operator.) □

Using this proposition we are able to provide a proof for Remark 2.28.

Proof of Remark 2.28. Under Assumption E3' and under positive definiteness of $\hat{\mathscr{C}}$ we know that $\hat{\lambda}_{d'} > \hat{\lambda}_{d'+1} \geq 0$ is asymptotically valid with probability tending to 1. Hence, we obtain the same bounds as in (2.5.5) and (2.5.6) by using Proposition 2.54 together with Proposition 2.22, Corollary 2.26 and (2.3.25).

Note that our need to impose positive definiteness of $\hat{\mathscr{C}}$ in Remark 2.28 stems from Proposition 2.54: If $d_H = \infty$ and Assumption E3 or Assumption E3' hold true then, on the one hand, we know that the probability of a fixed finite number of eigenvalues fulfilling $\hat{\lambda}_j \geq 0$ for $1 \leq j \leq d'$ is asymptotically 1, but, on the other hand, we cannot guarantee that the probability of infinitely many eigenvalues fulfilling $\hat{\lambda}_j \geq 0$ for $j \in \mathbb{N}$ is asymptotically 1, as well, i.e. that the covariance estimate becomes eventually positive definite. (However, in the multivariate case, where d_H is finite, we do not encounter asymptotically any such problems under the null hypothesis given that $\lambda_{d_H} > 0$ since the probability of $\hat{\mathscr{C}}$ being positive definite tends to 1.) □

We proceed with the estimation of the long run covariance structure under the alternative and particularly with the corresponding proofs of Theorem 2.32 and Corollary 2.34. They are both applications of Theorem 2.56 which shows under H_A how the behavior of $\hat{\mathscr{C}}$ deviates from consistent estimation of $\mathscr{C}_{1,0} = \mathscr{C}$ and is itself based on the following Lemma 2.55. Note that Lemma 2.55 extends Lemma 8.2 from Torgovitski (2015c), that Theorem 2.56 extends Theorem 8.3 from Torgovitski (2015c) and that these results are only indicated therein in this generality.

Lemma 2.55. *Set* $\beta_{n,i,j} = g_j(i/n) - \int_0^1 g_j(x)dx$ *and* $\mathscr{G}(g_i, g_j)$ *according to* (2.3.28) *and assume that all* g_j *are Lipschitz continuous. It holds that, as* $n \to \infty$,

$$\max_{0 \le r \le h} \Big| \sum_{i=1}^{n-r} \beta_{n,i,k}\beta_{n,i+r,j}/n - \mathscr{G}(g_k, g_j)\Big| = o(1),$$

for $1 \le k, j \le \varrho$, *where* $h \to \infty$ *with* $h = o(n)$ *as* $n \to \infty$.

Theorem 2.56. *Let the assumptions of Proposition 2.22 or the assumptions of Corollary 2.26 hold true. It holds under* H_A *that, as* $n \to \infty$,

$$\Big\|(\hat{\mathscr{C}} - \mathscr{C}_{1,0}) - \sum_{r=-n}^{n} \mathscr{K}(r/h) \sum_{k,j=1}^{\varrho} \mathscr{G}(g_k, g_j)[\Delta_k \otimes \Delta_j]\Big\|_{\mathcal{S}} = o_P(h), \tag{2.5.9}$$

where $\mathscr{G}(g_k, g_j)$ *is defined in* (2.3.28).

Proof of Lemma 2.55. We present a version of the proof of Torgovitski (2015c, Lemma 8.2) for the multi-directional change case. It holds that, as $n \to \infty$,

$$\begin{aligned}
&\max_{0 \le r \le h} \Big| \sum_{i=1}^{n-r} \beta_{n,i,k}\beta_{n,i+r,j}/n - \mathscr{G}(g_k, g_j)\Big| \\
&\le \max_{0 \le r \le h} \Big| \sum_{i=1}^{n-r} \beta_{n,i,k}\beta_{n,i,j}/n - \mathscr{G}(g_k, g_j)\Big| + \max_{0 \le r \le h} \Big(\sum_{i=1}^{n-r} |\beta_{n,i,k}||\beta_{n,i,j} - \beta_{n,i+r,j}|/n \Big) \\
&\le \Big| \sum_{i=1}^{n} \beta_{n,i,k}\beta_{n,i,j}/n - \mathscr{G}(g_k, g_j)\Big| + \max_{0 \le r \le h} \Big| \sum_{i=n-r+1}^{n} \beta_{n,i,k}\beta_{n,i,j}\Big|/n \\
&\quad + c\Big(\max_{0 \le r \le h} \max_{1 \le i \le n-r} |\beta_{n,i,j} - \beta_{n,i+r,j}|\Big)\Big(\sum_{i=1}^{n} |\beta_{n,i,k}|/n \Big) \\
&= o(1) + \mathcal{O}(h)\mathcal{O}(1/n) + \mathcal{O}(h/n)\mathcal{O}(1) = o(1),
\end{aligned}$$

with some $c > 0$, which finishes the proof. (The last relation follows from the Lipschitz continuity of $g(x)$ for $x \in [0,1]$. It implies also the boundedness $|\beta_{n,i+r,k}| \le c'$ for all $n \in \mathbb{N}$, $1 \le i + r \le n$ and $1 \le k \le \varrho$ with some $c' > 0$.) □

Proof of Theorem 2.56. This is a more detailed multi-directional version of the proof of Torgovitski (2015c, Theorem 8.3).[1] We assume that g is Lipschitz continuous since the piecewise Lipschitz continuous case can be handled in an analogous manner. Set $\beta_{n,i,k} = g_k(i/n) - \int_0^1 g_k(x)dx$ and observe that

$$\begin{aligned}
Y_{i+r} - \bar{Y}_n &= \varepsilon_{i+r} - \bar{\varepsilon}_n + \sum_{j=1}^{\varrho} \Big[g_j((i+r)/n) - \sum_{l=1}^{n} g_j(l/n)/n\Big]\Delta_j \\
&= \varepsilon_{i+r} - \bar{\varepsilon}_n + \sum_{j=1}^{\varrho} \beta_{n,i,j}\Delta_j + r_n,
\end{aligned} \tag{2.5.10}$$

[1] For the special case of a gradual piecewise linear change in an m-dependent framework a simpler version is given in Torgovitski (2015a, Proof of Lemma 3.3).

where $r_n = o(1)$, as $n \to \infty$, and where the convergence depends only on n, i.e. is uniform in i and r. We consider the decomposition

$$
\begin{aligned}
&(Y_i - \bar{Y}_n) \otimes (Y_{i+r} - \bar{Y}_n) \\
&= \Big[(\varepsilon_i - \bar{\varepsilon}_n) \otimes (\varepsilon_{i+r} - \bar{\varepsilon}_n) + \sum_{k,j=1}^{\varrho} \beta_{n,i,k}\beta_{n,i+r,j}(\Delta_k \otimes \Delta_j)\Big] && (2.5.11)\\
&+ \sum_{j=1}^{\varrho} \Big[\beta_{n,i+r,j}(\varepsilon_i \otimes \Delta_j) + \beta_{n,i,j}(\Delta_j \otimes \varepsilon_{i+r}) \\
&\qquad - \beta_{n,i+r,j}(\bar{\varepsilon}_n \otimes \Delta_j) - \beta_{n,i,j}(\Delta_j \otimes \bar{\varepsilon}_n) && (2.5.12)\\
&\qquad + \beta_{n,i+r,j}(r_n \otimes \Delta_j) + \beta_{n,i,j}(\Delta_j \otimes r_n)\Big] \\
&+ \varepsilon_i \otimes r_n + r_n \otimes \varepsilon_{i+r} - \bar{\varepsilon}_n \otimes r_n - r_n \otimes \bar{\varepsilon}_n + r_n \otimes r_n. && (2.5.13)
\end{aligned}
$$

The contribution of the first term in (2.5.11) is covered by Proposition 2.22 or Corollary 2.26, respectively, and yields the term $\mathscr{C}_{1,0}$ in (2.5.9). Next, we will see that only the second term in (2.5.11) does additionally contribute to the overall asymptotics of $\hat{\mathscr{C}}$ by showing that the remaining terms vanish asymptotically. Therefore, it is sufficient to restrict our considerations to the contribution of the third term (and to consider any $1 \le j \le \varrho$) since all subsequent terms can be treated in a similar fashion. Due to the boundedness of $|\mathscr{K}(x)| \le 1$, $x \in \mathbb{R}$, and due to $\|\Delta_j\| = 1$ we obtain

$$
\begin{aligned}
&\Big\| \sum_{r=1}^{n} \mathscr{K}(r/h) \sum_{i=1}^{n-r} \beta_{n,i+r,j}(\varepsilon_i \otimes \Delta_j)/n \Big\|_{\mathcal{S}} \\
&\le \sum_{r=1}^{\lfloor ch \rfloor} \Big\| \sum_{i=1}^{n-r} \beta_{n,i+r,j}\varepsilon_i \Big\|/n \\
&\le \lfloor ch \rfloor \max_{1 \le r \le \lfloor ch \rfloor} \Big\| \sum_{i=1}^{n-r} \beta_{n,i,j}\varepsilon_i \Big\|/n + \sum_{r=1}^{\lfloor ch \rfloor} \Big(\sum_{i=1}^{n-r} |\beta_{n,i,j} - \beta_{n,i+r,j}| \|\varepsilon_i\|/n \Big) && (2.5.14)\\
&\le \lfloor ch \rfloor \max_{1 \le r \le \lfloor ch \rfloor} \Big\| \sum_{i=1}^{n-r} \beta_{n,i,j}\varepsilon_i \Big\|/n \\
&\qquad + \lfloor ch \rfloor \max_{1 \le r \le \lfloor ch \rfloor} \Big(\max_{1 \le i \le n-r} |\beta_{n,i,j} - \beta_{n,i+r,j}| \Big) \sum_{i=1}^{n} \|\varepsilon_i\|/n
\end{aligned}
$$

for some $c > 0$. Using Proposition 2.65 (which is postponed to Subsection 2.5.3) and $E(\sum_{i=n-\lfloor ch \rfloor+1}^{n} |\beta_{n,i,j}| \|\varepsilon_i\|) = \mathcal{O}(h)$ we observe that

$$
\begin{aligned}
&\max_{1 \le r \le \lfloor ch \rfloor} \Big\| \sum_{i=1}^{n-r} \beta_{n,i,j}\varepsilon_i \Big\|/n \\
&\le \Big\| \sum_{i=1}^{n} \beta_{n,i,j}\varepsilon_i \Big\|/n + \max_{1 \le r \le \lfloor ch \rfloor} \Big\| \sum_{i=n-r+1}^{n} \beta_{n,i,j}\varepsilon_i \Big\|/n \\
&\le \Big\| \sum_{i=1}^{n} \beta_{n,i,j}\varepsilon_i \Big\|/n + \sum_{i=n-\lfloor ch \rfloor+1}^{n} |\beta_{n,i,j}| \|\varepsilon_i\|/n = o_P(1) + \mathcal{O}_P(h/n) = o_P(1).
\end{aligned}
$$

By the Lipschitz continuity and the law of large numbers we get

$$\max_{1\le r\le \lfloor ch\rfloor}\Big[\max_{1\le i\le n-r}|\beta_{n,i,j}-\beta_{n,i+r,j}|\Big]\Big[\sum_{i=1}^{n}\|\varepsilon_i\|/n\Big]=\mathcal{O}(h/n)\mathcal{O}_P(1)=o_P(1),$$

as $n\to\infty$, and altogether, we obtain

$$\Big\|\sum_{r=1}^{n}\mathscr{K}(r/h)\sum_{i=1}^{n-r}\beta_{n,i+r,j}\,[\varepsilon_i\otimes\Delta_j]/n\Big\|_{\mathcal{S}}=o_P(h).$$

The remaining terms in (2.5.12) and the term (2.5.13) are also all of order $o_P(h)$ since $r_n=o(1)$ converges uniformly (cf. (2.5.10)) and since $\sum_{i=1}^{n}\beta_{n,i,j}\varepsilon_i/n=o_P(1)$ holds true in view of Proposition 2.65. Finally, we use Lemma 2.55 to obtain

$$\sum_{r=0}^{n}|\mathscr{K}(r/h)|\Big\|\sum_{i=1}^{n-r}\sum_{k,j=1}^{\varrho}\beta_{n,i,k}\beta_{n,i+r,j}[\Delta_k\otimes\Delta_j]/n \\ -\sum_{k,j=1}^{\varrho}\mathscr{G}(g_k,g_j)[\Delta_k\otimes\Delta_j]\Big\|_{\mathcal{S}}=o(h).$$

Note, that if g_j are *piecewise* Lipschitz continuous then the arguments are analogous but require a technical modification: for instance, we would have to exclude some terms in the last sum of (2.5.14) for all of discontinuities of g_j (the number of which is finite). However, all skipped terms are uniformly bounded by $\tilde{c}/n$ for some $\tilde{c}>0$ and thus the last sum still vanishes asymptotically at the rate $o_P(h)$. Also we have to adapt the proof of Lemma 2.55 accordingly. We avoid those straightforward modifications to keep the presentation clear and compact. □

Proof of Theorem 2.32. The assertion follows from Theorem 2.56 by taking following two considerations into account: under the Assumptions of Proposition 2.22 the bandwidth $h\ge m$ is fixed and we get $\sum_{r=-n}^{n}\mathscr{K}(r/h)/(2h+1)=1=\int_0^\infty\mathscr{K}(x)dx$. Whereas, under the assumptions of Corollary 2.26 we use that $\mathscr{K}$ is piecewise continuous, symmetric and that it has a bounded support to obtain

$$\lim_{n\to\infty}\sum_{r=-n}^{n}\mathscr{K}(r/h)/(2h+1)=\int_{-\infty}^{\infty}\mathscr{K}(x)dx/2=\int_0^{\infty}\mathscr{K}(x)dx$$

with a well-defined integral. □

Proof of Corollary 2.34. The first and the second statements (2.3.31) and (2.3.32) are immediate consequences of Theorem 2.32. The last statement (2.3.33) follows then as in the proof of Corollary 2.27 on p. 69. □

Proofs for Subsection 2.3.3 (on testing with CUSUM) 2.5.2

We proceed to study the CUSUM procedure and the replacement of the population Hilbert space principal components by their empirical counterparts. Let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{1,0}$ and $(\hat{\lambda}_j, \hat{v}_j)$ be the eigenelements of $\hat{\mathscr{C}}$. We recall the definition of the projection scores

$$\begin{aligned}\boldsymbol{y}_i &= \left[Y_{i,1}, \ldots, Y_{i,d}\right]', & Y_{i,r} &= \langle Y_i, v_r\rangle,\\ \hat{\boldsymbol{y}}_i &= \left[\hat{Y}_{i,1}, \ldots, \hat{Y}_{i,d}\right]', & \hat{Y}_{i,r} &= \langle Y_i, \hat{v}_r\rangle\end{aligned}$$

and additionally introduce

$$\begin{aligned}\boldsymbol{\epsilon}_i &= \left[\varepsilon_{i,1}, \ldots, \varepsilon_{i,d}\right]', & \varepsilon_{i,r} &= \langle \varepsilon_i, v_r\rangle,\\ \hat{\boldsymbol{\epsilon}}_i &= \left[\hat{\varepsilon}_{i,1}, \ldots, \hat{\varepsilon}_{i,d}\right]', & \hat{\varepsilon}_{i,r} &= \langle \varepsilon_i, \hat{v}_r\rangle.\end{aligned} \tag{2.5.15}$$

The corresponding test statistics are

$$\begin{aligned}\mathscr{M}_n(\boldsymbol{y}) &= \max_{1\le k<n} w(k/n)|S_n(k/n;\boldsymbol{y})|_{\Sigma},\\ \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) &= \max_{1\le k<n} w(k/n)|S_n(k/n;\hat{\boldsymbol{y}})|_{\hat{\Sigma}},\\ \mathscr{M}_n(\boldsymbol{\epsilon}) &= \max_{1\le k<n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})|_{\Sigma},\\ \hat{\mathscr{M}}_n(\hat{\boldsymbol{\epsilon}}) &= \max_{1\le k<n} w(k/n)|S_n(k/n;\hat{\boldsymbol{\epsilon}})|_{\hat{\Sigma}},\end{aligned} \tag{2.5.16}$$

where $\hat{\Sigma} = \operatorname{diag}(|\hat{\lambda}_1|, \ldots, |\hat{\lambda}_d|)$ and $\Sigma = \operatorname{diag}(\lambda_1, \ldots, \lambda_d)$.[1]

Proposition 2.57. *Let Assumption M1 or Assumption M2 hold true with $\delta(m) = m^{-\nu}$ for some $\nu > 2$. Furthermore, let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{1,0} = \mathscr{C}$, which is defined in (2.3.17), and assume that Assumption E2 holds true. Then under H_0 it holds that*

$$\lim_{n\to\infty} P\left(a_n \mathscr{M}_n(\boldsymbol{\epsilon}) - b_{n,d} \le x\right) = \exp(-2\exp(-x)), \tag{2.5.17}$$

for all $x \in \mathbb{R}$, where $\mathscr{M}_n(\boldsymbol{\epsilon})$ is defined via (2.5.15) and (2.5.16).

Showing that (2.5.17) is still true for $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ is straightforward if we work with simultaneous projections on subspaces. Notice that without having this approach we would need a series of (interconnected) lemmas as in Torgovitski (2015a).

Proof of Theorem 2.40. Let $\mathscr{M}_n(\boldsymbol{\epsilon})$ be defined via (2.5.15) and (2.5.16). Closely following the proof of Torgovitski (2015c, Theorem 4.3) we observe, using the tensor based CUSUM representation (2.3.15), that

$$\begin{aligned}&|\mathscr{M}_n(\boldsymbol{\epsilon}) - \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})|\\ &= \Big|\max_{1\le k<n} w(k/n)|\Sigma^{-1/2}S_n(k/n;\boldsymbol{\epsilon})| - \max_{1\le k<n} w(k/n)|\hat{\Sigma}^{-1/2}S_n(k/n;\hat{\boldsymbol{\epsilon}})|\Big|\end{aligned}$$

[1] Note a subtle difference between $\mathscr{M}_n(\boldsymbol{e})$, defined in the preliminary discussion of Subsection 2.3.3, and the principal component based $\mathscr{M}_n(\boldsymbol{\epsilon})$, defined in (2.5.16). (This should not lead to any confusion.)

$$
\begin{aligned}
&= | \max_{1\le k<n} w(k/n) \|\mathscr{C}_{1,0}^{-1/2,d} S_n(k/n;\varepsilon)\| - \max_{1\le k<n} w(k/n)\|\hat{\mathscr{C}}^{-1/2,d} S_n(k/n;\varepsilon)\|| \\
&\le \max_{1\le k<n} w(k/n)\|[\mathscr{C}_{1,0}^{-1/2,d} - \hat{\mathscr{C}}^{-1/2,d}] S_n(k/n;\varepsilon)\| \qquad (2.5.18)\\
&\le \|\mathscr{C}_{1,0}^{-1/2,d} - \hat{\mathscr{C}}^{-1/2,d}\|_{\mathcal{L}} \max_{1\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\|.
\end{aligned}
$$

It holds for the operator and the Hilbert-Schmidt norms that $\|\mathscr{C}_{1,0}^{-1/2,d} - \hat{\mathscr{C}}^{-1/2,d}\|_{\mathcal{L}} \le \|\mathscr{C}_{1,0}^{-1/2,d} - \hat{\mathscr{C}}^{-1/2,d}\|_{\mathcal{S}}$ and the proof is finished if, on the one hand, we combine Proposition 2.57 with (2.5.18) and if, on the other hand, we use the rates for weighted partial sums in Proposition 2.63 (see Subsection 2.5.3) together with the rates of estimation for truncated (inverse square root) operators of Corollary 2.27. □

Proof of Proposition 2.57. This is a combined, modified and more detailed version of the proofs for Lemma 6.3 in Torgovitski (2015a) and for Theorem 3.10 in Torgovitski (2016).[1] First, we observe that the *orthogonal projection* $\varepsilon_i \mapsto \boldsymbol{\epsilon}_i$ is a measurable, bounded linear mapping. This implies that the $\boldsymbol{\epsilon}_i$'s are centered and that they inherit the strict stationarity and the overall dependence structure from the Hilbert space valued ε_i's. Hence, they remain centered and strictly stationary m-dependent in the case of Assumption M1 and L^κ-m-approximable - with the same rate $\delta(m)$ and the same κ - in the case of Assumption M2 (but now as a sequence in $\mathbb{R}^d$). As already mentioned in (2.3.11) the long run covariance matrix of $\{\boldsymbol{\epsilon}_i\}_{i\in\mathbb{Z}}$ is given by $\operatorname{diag}(\lambda_1,\dots,\lambda_d)$, which has full rank due to Assumption E2. Hence, for convenience, we will tacitly assume that this long run covariance matrix is the identity matrix, i.e. that the error sequence is rescaled.[2]

In the case of Assumption M1 the assertion follows immediately by (Horváth *et al.*, 1999, Theorem 1.3). In the case of Assumption M2 this requires more effort and may be obtained following Csörgő & Horváth (1997, Theorem 4.1.3) and combining the latter with Schmitz (2011). Note that a related proof is also indicated in Kamgaing & Kirch (2016). We provide a detailed proof and introduce the *»truncation and shifting techniques«* that will be (re)used, e.g., in the proof of Theorem 2.43 later on but in a continuous-time setup. Set

$$
\begin{aligned}
s_n &= \lfloor \exp((\log n)^{1-\varepsilon}) \rfloor,\\
u_n &= \lfloor n/(\log n) \rfloor
\end{aligned}
$$

for any $\varepsilon \in (0,1)$ and observe that

$$
\begin{aligned}
&[(\log\log u_n)/(\log\log n)]^{1/2} = 1 + o(1),\\
&[(\log\log s_n)/(\log\log n)]^{1/2} = (1-\varepsilon)^{1/2} + o(1),
\end{aligned}
\qquad (2.5.19)
$$

as $n\to\infty$. Subsequently, we tacitly assume that $n \ge n_0 \in \mathbb{N}$, which is chosen such that $s_n < u_n$ for all $n \ge n_0$. Recall also that we assume $\delta(m) = m^{-\nu}$ for some $\nu > 2$. Hence, by the invariance principle (2.3.42) we obtain that, as $n\to\infty$,

$$
\max_{1\le k<n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i - \boldsymbol{W}(k)|/k^{1/2} = \mathcal{O}_P(1), \qquad (2.5.20)
$$

[1] In Torgovitski (2016, Theorem 3.10) we considered formally only the causal case. Here, we allow for two-sided shifts as well.

[2] Cf., e.g., the discussion on p. 45.

$$\max_{s_n \le k \le n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i - \boldsymbol{W}(k)|/k^{1/2} = o_P((\log\log n)^{-1/2}), \tag{2.5.21}$$

$$\max_{u_n \le k \le n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i - \boldsymbol{W}(k)|/k^{1/2-\eta_1} = o_P(1) \tag{2.5.22}$$

for some $\eta_1 > 0$. Following the proof of Schmitz (2011, Theorem 2.1.4) we obtain by using (2.5.22) that

$$\lim_{n\to\infty} P\Big[a(w_n) \max_{u_n < k < n-u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_d^*(w_n) \le x\Big] = \exp(-2\exp(-x)) \tag{2.5.23}$$

for all $x \in \mathbb{R}$ and with $w_n = 2\log((\log n) - 1)$. Note that the polynomial rate of approximation in the invariance principle is crucial in the proof of Schmitz (2011, Theorem 2.1.4). Subsequently, we will explain why the following Darling-Erdős-type results

$$\lim_{n\to\infty} P\Big[a_n \max_{1\le k \le u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d}^* \le x\Big] = \exp(-2\exp(-x)), \tag{2.5.24}$$

$$\lim_{n\to\infty} P\Big[a_n \max_{n-u_n\le k < n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d}^* \le x\Big] = \exp(-2\exp(-x)), \tag{2.5.25}$$

for all $x \in \mathbb{R}$, hold true under our dependence assumptions. Before we turn to the proofs, we note that above asymptotics (2.5.23), (2.5.24) and (2.5.25) imply that

$$\begin{aligned}
&(2\log\log n)^{-1/2} \max_{u_n<k<n-u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| = o_P(1),\\
&(2\log\log n)^{-1/2} \max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| = 1 + o_P(1),\\
&(2\log\log n)^{-1/2} \max_{n-u_n\le k<n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| = 1 + o_P(1).
\end{aligned} \tag{2.5.26}$$

Hence, the overall assertion would follow from (2.5.24), (2.5.25) and (2.5.26) after showing the following asymptotic independence

$$\begin{aligned}
&\lim_{n\to\infty} P\Big[a_n \max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d} \le t,\\
&\qquad a_n \max_{n-u_n\le k<n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d} \le s\Big]\\
&= \lim_{n\to\infty} P\Big[a_n \max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d} \le t\Big]\\
&\qquad \times \lim_{n\to\infty} P\Big[a_n \max_{n-u_n\le k<n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| - b_{n,d} \le s\Big].
\end{aligned} \tag{2.5.27}$$

We proceed by showing (2.5.24) and (2.5.25) first and verifying (2.5.27) afterwards. From Lemma 2.2 of Horváth (1993) (see also (2.3.37)) we observe via $\boldsymbol{W}(k) = \sum_{i=1}^{k}[\boldsymbol{W}(i) - \boldsymbol{W}(i-1)]$ (a.s.) that

$$\lim_{n\to\infty} P\Big[a(\log w_n) \max_{1\le k\le w_n} |\boldsymbol{W}(k)|/k^{1/2} - b^*(\log w_n) \le x\Big] = \exp(-2\exp(-x)), \tag{2.5.28}$$

$$(2\log\log n)^{-1/2} \max_{1\le k\le u_n} |\boldsymbol{W}(k)|/k^{1/2} = 1 + o_P(1), \tag{2.5.29}$$

$$(2\log\log n)^{-1/2} \max_{1\le k\le s_n} |\boldsymbol{W}(k)|/k^{1/2} = (1-\varepsilon)^{1/2} + o_P(1) \tag{2.5.30}$$

for any $w_n \to \infty$, $w_n \in \mathbb{N}$ (cf. (2.5.19)). Via the first invariance principle (2.5.20) relations (2.5.29) and (2.5.30) yield that

$$\begin{aligned}(2\log\log n)^{-1/2} \max_{1\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} &= 1 + o_P(1),\\ (2\log\log n)^{-1/2} \max_{1\le k\le s_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} &= (1-\varepsilon)^{1/2} + o_P(1).\end{aligned} \tag{2.5.31}$$

Choose any $d_n, g_n \in [1, u_n] \cap \mathbb{N}$ such that

$$\max_{1\le k\le u_n} |\boldsymbol{W}(k)|/k^{1/2} = |\boldsymbol{W}(d_n)|/d_n^{1/2},$$

$$\max_{1\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} = |\sum_{i=1}^{g_n} \boldsymbol{\epsilon}_i|/g_n^{1/2}$$

and observe from (2.5.29), (2.5.30) and (2.5.31) that

$$\lim_{n\to\infty} P(d_n, g_n \in [s_n, u_n]) = 1. \tag{2.5.32}$$

A direct application of (2.5.21) yields

$$\Big| \max_{s_n\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} - \max_{s_n\le k\le u_n} |\boldsymbol{W}(k)|/k^{1/2}\Big| = o_P((\log\log n)^{-1/2})$$

and by taking (2.5.32) into account this implies

$$\Big| \max_{1\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} - \max_{1\le k\le u_n} |\boldsymbol{W}(k)|/k^{1/2}\Big| = o_P((\log\log n)^{-1/2}). \tag{2.5.33}$$

Using (2.5.33) together with (2.5.29), we arrive at

$$\max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| = \max_{1\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} + o_P((\log\log n)^{-1/2}), \tag{2.5.34}$$

which then combined with (2.5.28) and (2.5.33) yields

$$\begin{aligned}\lim_{n\to\infty} P\Big[a(\log u_n) \max_{1\le k\le u_n} |\sum_{i=1}^{k} \boldsymbol{\epsilon}_i|/k^{1/2} - b_d^*(\log u_n) \le x\Big]\\ = \exp(-2\exp(-x)),\end{aligned} \tag{2.5.35}$$

for all $x \in \mathbb{R}$, and thus shows (2.5.24). Note that the precise calculations for (2.5.34) and the transition from (2.5.35) to (2.5.24) are routine and skipped since a detailed explanation in the univariate case can be found in Lemma 2.1.5 and under (2.1.67) and (2.1.68) of Schmitz (2011). Note that, therein, the backward invariance principle may be replaced by the law of the iterated logarithm (which is implied by the invariance principle) after invoking stationarity using similar arguments as for the subsequent considerations.

To cope with (2.5.25) we set $\boldsymbol{z}_i^{(n)} := \boldsymbol{\epsilon}_{n-i+1}$ and observe that for all fixed n the sequences $\{\boldsymbol{z}_i^{(n)}\}_{i\in\mathbb{N}}$ are L^κ-m-approximable with same κ and same rate $\delta(m)$. We may repeat all considerations with respect to the time-inversed series $\{\boldsymbol{z}_i^{(1)}\}_{i\in\mathbb{N}}$ which also satisfies the invariance principles (2.5.20), (2.5.21) and (2.5.22) but with a different Wiener process. Thus, via stationarity and symmetry we obtain from (2.5.24) that

$$\begin{aligned}
&\max_{n-u_n\le k<n} w(k/n)|S_n(k/n;\boldsymbol{\epsilon})| \\
&\quad= \max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{z}^{(n)})| \\
&\quad\overset{\mathcal{D}}{=} \max_{1\le k\le u_n} w(k/n)|S_n(k/n;\boldsymbol{z}^{(1)})| \\
&\quad= \max_{1\le k\le u_n} \Big|\sum_{i=1}^{k}\boldsymbol{z}_i^{(1)}\Big|/k^{1/2} + o_P((\log\log n)^{-1/2}) && (2.5.36)\\
&\quad\overset{\mathcal{D}}{=} \max_{n-u_n\le k<n} \Big|\sum_{i=k+1}^{n}\boldsymbol{\epsilon}_i\Big|/(n-k)^{1/2} + o_P((\log\log n)^{-1/2}) && (2.5.37)
\end{aligned}$$

with a slight abuse of notation in the last step where the equality in distribution is only meant for the maximum of the partial sums (and not for the o_P terms). Note that (2.5.36) follows by repeating the verification of (2.5.34). Via (2.5.35) and (2.5.36) we obtain (2.5.25). Now, we use (2.5.34) and (2.5.37) to show (2.5.27). Therefore, we will adapt the approach of Csörgő & Horváth (1997, Theorem 4.1.3) by considering m-dependent copies $\boldsymbol{\epsilon}_i^{(m)}$ as in (1.3.3) and by setting $m = m_n = \lfloor n-3u_n\rfloor/2$. This sequence is chosen such that $[\boldsymbol{\epsilon}_1^{(m)},\ldots,\boldsymbol{\epsilon}_{u_n}^{(m)}]$ and $[\boldsymbol{\epsilon}_{n-u_n}^{(m)},\ldots,\boldsymbol{\epsilon}_n^{(m)}]$ are independent for all sufficiently large n. The representation of $\boldsymbol{\epsilon}_i$'s as a shift of i.i.d. random variables and the construction of the $\boldsymbol{\epsilon}_i^{(m)}$'s ensures that $Z_{k,r} := \boldsymbol{\epsilon}_{k,r} - \boldsymbol{\epsilon}_{k,r}^{(m)}$ are equally distributed for all k. Hence, it holds that

$$E|Z_{k,r}|^2 = E|Z_{0,r}|^2 = \mathcal{O}(\delta^2(m)) = \mathcal{O}(m^{-2\gamma})$$

as $n\to\infty$ for all $k\in\mathbb{Z}$ and all $1\le r\le d$. Furthermore, we observe

$$\begin{aligned}
&\Big|\max_{1\le k\le u_n}\Big|\sum_{i=1}^{k}\boldsymbol{\epsilon}_i\Big|/k^{1/2} - \max_{1\le k\le u_n}\Big|\sum_{i=1}^{k}\boldsymbol{\epsilon}_i^{(m)}\Big|/k^{1/2}\Big| \\
&\le \max_{1\le k\le u_n}\Big|\sum_{i=1}^{k}\big(\boldsymbol{\epsilon}_i-\boldsymbol{\epsilon}_i^{(m)}\big)\Big|/k^{1/2} \le \sum_{r=1}^{d}\max_{1\le k\le u_n}\Big|\sum_{i=1}^{k}Z_{i,r}\Big|/k^{1/2}.
\end{aligned}$$

An application of the Hájek-Rényi type inequality of Kounias & Weng (1969, Theorem 2) yields that

$$\begin{aligned}
&P\Big(\max_{1\le k\le u_n}\Big|\sum_{i=1}^{k}Z_{i,r}\Big|/k^{1/2} > (\log n)^{-1/2}\Big) \\
&\le \Big[(\log n)^{1/2}\sum_{k=1}^{u_n}\big(E|Z_{k,r}|^2\big)^{1/2}/k^{1/2}\Big]^2 \\
&= (\log n)E|Z_{0,r}|^2\Big[\sum_{k=1}^{u_n}k^{-1/2}\Big]^2 = \mathcal{O}((\log n)m^{-2\nu}u_n) = \mathcal{O}(n^{1-2\nu}),
\end{aligned}$$

which implies

$$\max_{1\le k\le u_n}\Big|\sum_{i=k+1}^{n}\boldsymbol{\epsilon}_i\Big|/k^{1/2} = \max_{1\le k\le u_n}\Big|\sum_{i=k+1}^{n}\boldsymbol{\epsilon}_i^{(m)}\Big|/k^{1/2} + o_P((\log\log n)^{-1/2}),$$

$$\max_{n-u_n\le k<n}|\sum_{i=k+1}^{n}\boldsymbol{\epsilon}_i|/(n-k)^{1/2}=\max_{n-u_n\le k<n}|\sum_{i=k+1}^{n}\boldsymbol{\epsilon}_i^{(m)}|/(n-k)^{1/2}+o_P((\log\log n)^{-1/2}),$$

where the second relation follows exactly as the first. The proof of (2.5.27) follows by taking (2.5.24), (2.5.25), (2.5.34), (2.5.37) and Davidson (1994, Lemma 29.5) into account. This completes the proof of the whole proposition. □

Proof of Corollary 2.42. This is a slightly more detailed version of the proof of Torgovitski (2015a, Corollary 4.3). First, we want to clarify that $V(\chi;h)$ has a continuous distribution function. This is often tacitly assumed in the literature (cf., e.g., Gombay & Horváth, 1996). One way to verify this continuity is to start with the fact that the dual norm of the Euclidean norm is the Euclidean norm itself. Thus, following Piterbarg (2012, p. 115) we may write

$$V(\chi;h)=\sup_{t\in[h,1-h]}|\boldsymbol{B}(t)|/\big(t(1-t)\big)^{1/2}=\sup_{t\in[h,1-h]}\Big[\sup_{\boldsymbol{u}\in S}|\boldsymbol{G}(t,\boldsymbol{u})|\Big]=\sup_{(t,\boldsymbol{u})\in[h,1-h]\times S}\boldsymbol{G}(t,\boldsymbol{u}),$$

where $\boldsymbol{G}(t,\boldsymbol{u})=\langle\boldsymbol{B}(t),\boldsymbol{u}\rangle/\big(t(1-t)\big)^{1/2}$ is now a Gaussian random field and S is a d-dimensional unit sphere in $\mathbb{R}^d$. Now, it is also easy to check that $\operatorname{Var}(\boldsymbol{G}(t,\boldsymbol{u}))=1$ holds true for all $(t,\boldsymbol{u})\in[h,1-h]\times S$. Therefore, all conditions of Tsirelson (1976, Theorem 3) are fulfilled (cf. Proposition 2 in Lifshits, 1984), which implies a continuous distribution function of $\sup_{(t,\boldsymbol{u})\in[h,1-h]\times S}\boldsymbol{G}(t,\boldsymbol{u})$. Hence, we may always find a sequence ξ_n that satisfies $P(V(\chi;h)\le\xi_n)=\alpha$. From (2.3.38) and Remark 2.36 we know that

$$\lim_{n\to\infty}P\left(a_nV(\chi;h)-b_{n,d}\le x\right)=\exp(-2\exp(-x)) \tag{2.5.38}$$

holds true, where $h=h_n$. Altogether, a combination of Theorem 2.40 with (2.5.38) states that, as $n\to\infty$,

$$\begin{aligned}
&\big|P(\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})\le\xi_n)-\alpha\big|\\
&=\big|P(\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})\le\xi_n)-P(V(\chi;h)\le\xi_n)\big|\\
&\le\sup_{x\in\mathbb{R}}\big|P\big(a_n\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})-b_{n,d}\le x\big)-P\big(a_nV(\chi;h)-b_{n,d}\le x\big)\big|\\
&\le\sup_{x\in\mathbb{R}}\big|P\big(a_n\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})-b_{n,d}\le x\big)-\exp(-2\exp(-x))\big|\\
&\quad+\sup_{x\in\mathbb{R}}\big|P\big(a_nV(\chi;h)-b_{n,d}\le x\big)-\exp(-2\exp(-x))\big|=o(1).
\end{aligned}$$

The uniform convergence in the last step is justified by the continuity of the limiting Gumbel distribution. Again, on using Theorem 2.40 and (2.5.38) we obtain

$$\begin{aligned}
\lim_{n\to\infty}(a_nc_n'(\alpha)-b_{n,d})&=\lim_{n\to\infty}z_\alpha(a_nV(\chi;h)-b_{n,d})\\
&=\lim_{n\to\infty}z_\alpha(a_n\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})-b_{n,d})=\lim_{n\to\infty}(a_nc_n(\alpha)-b_{n,d}).
\end{aligned} \tag{2.5.39}$$

Here, we used that $z_\alpha(aX+b)=az_\alpha(X)+b,\ a>0$ for any random variable X and the fact that convergence in distribution implies convergence of quantiles at continuity points of the quantile function, as well. (Recall, that the limiting distribution is continuous in our case.) Clearly, (2.5.39) implies $c_n'(\alpha)-c_n(\alpha)=o(b_{n,d}/a_n)$ as $n\to\infty$, which finishes the proof. □

Proof of Theorem 2.43. We present a modified and more detailed proof of Torgovitski (2015a, Theorem 4.2). The strategy of the proof follows Csörgő & Horváth (1997, Theorem 1.3.2) but is based on the invariance principle in Horváth *et al.* (1999, Lemma 4.3) for m-dependent sequences. Before we continue, let us assume that $h = h_n$ fulfills (2.3.40) for some $\varepsilon^* \in (0,1)$ and define the intervals

$$\begin{aligned} J &= J_n = [1/n, 1-1/n], \\ I &= I_n = [h, 1-h], \\ J^1 &= J_n^1 = [s, 1-s], \\ J^2 &= J_n^2 = J_n^l \cup J_n^r, \end{aligned}$$

where $J_n^l = [1/n, s)$ and $J_n^r = (1-s, 1-1/n]$ with the sequence

$$s = s_n = \lfloor \exp((\log n)^{1-\varepsilon_1}) \rfloor / n$$

for arbitrary $\varepsilon_1 \in (0, \varepsilon^*)$. Note that the following inclusions

$$J^1 \subset I \subset J \tag{2.5.40}$$

hold true by construction and by the choice of ε_1. Furthermore, we set

$$\begin{aligned} T_n(t) &= w(t)|S_n'(t;\boldsymbol{\epsilon})|_\Sigma, \\ T_n'(t) &= w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n;\boldsymbol{\epsilon})|_\Sigma = w(\lfloor nt \rfloor/n)|S_n(t;\boldsymbol{\epsilon})|_\Sigma, \end{aligned}$$

with $\{\boldsymbol{\epsilon}_i\}$ defined via (2.5.15), with $|\cdot|_\Sigma = |\Sigma^{-1/2}\cdot|$ and where $S_n'(t;\boldsymbol{\epsilon}) = (\sum_{i=1}^{\lfloor nt \rfloor}\boldsymbol{\epsilon}_i - t\sum_{i=1}^{n}\boldsymbol{\epsilon}_i)/n^{1/2}$ is a minor modification of $S_n(t;\boldsymbol{\epsilon})$. Note the subtle difference that $S_n(t;\boldsymbol{\epsilon})$ is piece-wise constant in t whereas $S_n'(t;\boldsymbol{\epsilon})$ is not. As in the proof of Proposition 2.57 we will tacitly assume that the long run covariance matrix Σ is the identity matrix, i.e. that the error sequence is rescaled.

Horváth *et al.* (1999, Lemma 4.3) states that for strictly stationary and m-dependent sequences, $\{\boldsymbol{\epsilon}_i\}$, there exist equivalent processes $\{\chi_n(t,d),\ t \in (0,1)\} \overset{\mathcal{D}}{=} \{\chi(t,d),\ t \in (0,1)\}$ (with $\chi(t,d)$ being defined via Brownian bridges as in (2.3.44)) such that

$$n^\alpha \sup_{t\in J}[t(1-t)]^\alpha |T_n^2(t) - \chi_n^2(t,d)| = \mathcal{O}_P(1) \tag{2.5.41}$$

holds true for all $\alpha \in [0,\delta]$ and some $\delta > 0$. This yields

$$\begin{aligned} \sup_{t\in J^1}|T_n^2(t) - \chi_n^2(t,d)| &\le \sup_{t\in J^1}[t(1-t)]^{-\alpha}\Big[\sup_{t\in J^1}[t(1-t)]^\alpha|T_n^2(t) - \chi_n^2(t,d)|\Big] \\ &= \mathcal{O}_P((ns)^{-\alpha}) \\ &= \mathcal{O}_P(\exp(-\alpha(\log n)^{1-\varepsilon_1})) \end{aligned} \tag{2.5.42}$$

for some $\alpha > 0$. From Csörgő & Horváth (1997, Theorem A.3.4) we know that

$$\begin{aligned} (2\log\log n)^{-1/2}\sup_{t\in J^1}\chi_n(t,d) &= 1 + o_P(1), \\ (2\log\log n)^{-1/2}\sup_{t\in J^2}\chi_n(t,d) &= (1-\varepsilon_1)^{1/2} + o_P(1) \end{aligned} \tag{2.5.43}$$

hold true. On account of (2.5.41) with $\alpha = 0,$ we get from (2.5.43), that

$$\begin{aligned} &(2\log\log n)^{-1/2} \sup_{t\in J^1} T_n(t) = 1 + o_P(1), \\ &(2\log\log n)^{-1/2} \sup_{t\in J^2} T_n(t) = (1-\varepsilon_1)^{1/2} + o_P(1) \end{aligned} \tag{2.5.44}$$

and, as we will show further below, also that

$$(2\log\log n)^{-1/2} \sup_{t\in J^1} T_n'(t) = 1 + o_P(1), \tag{2.5.45}$$

$$(2\log\log n)^{-1/2} \sup_{t\in J^2} T_n'(t) = (1-\varepsilon_1)^{1/2} + o_P(1). \tag{2.5.46}$$

Now, we choose any δ_n, γ_n and γ_n' that fulfill

$$\begin{aligned} &\sup_{t\in J} \chi_n(t,d) = \chi_n(\delta_n, d), \\ &\sup_{t\in J} T_n(t) = T_n(\gamma_n), \\ &\sup_{t\in J} T_n'(t) = T_n'(\gamma_n') \end{aligned}$$

and observe from (2.5.43) - (2.5.46) that

$$\lim_{n\to\infty} P(\delta_n, \gamma_n, \gamma_n' \in J^1) = 1. \tag{2.5.47}$$

Furthermore, (2.5.42) yields, in view of the first relation in (2.5.43), that

$$\Big|\sup_{t\in J^1} T_n(t) - \sup_{t\in J^1} \chi_n(t,d)\Big| = \mathcal{O}_P(\exp(-(\log n)^{1-\varepsilon_2})), \tag{2.5.48}$$

with any $\varepsilon_2 \in (\varepsilon_1, \varepsilon^*)$. This rate is slightly weaker since we have squared terms in (2.5.42) and since the logarithmic sequence in (2.5.43) needs to be compensated to get the rates for the non-squared expressions. The latter (2.5.48) implies

$$\Big|\sup_{t\in J^1} T_n'(t) - \sup_{t\in J^1} \chi_n(t,d)\Big| = \mathcal{O}_P(\exp(-(\log n)^{1-\varepsilon_3})),$$

for any $\varepsilon_3 \in (\varepsilon_2, \varepsilon^*)$ due to

$$\begin{aligned} \Big|\sup_{t\in J^1} T_n'(t) - \sup_{t\in J^1} T_n(t)\Big| &= \mathcal{O}_P((\log\log n)^{1/2} \exp(-(\log n)^{1-\varepsilon_1})) \\ &= \mathcal{O}_P(\exp(-(\log n)^{1-\varepsilon_3})), \end{aligned} \tag{2.5.49}$$

which will be shown further below. In view of (2.5.47) and (2.5.40) we can replace J^1 with I or J in (2.5.49), respectively, and end up with

$$\Big|\mathscr{M}_n(\boldsymbol{\epsilon}) - \sup_{t\in I} \chi_n(t,d)\Big| = \Big|\sup_{t\in J} T_n'(t) - \sup_{t\in I} \chi_n(t,d)\Big| = \mathcal{O}_P(\exp(-(\log n)^{1-\varepsilon_3})).$$

The assertions follow now on replacing $\mathscr{M}_n(\boldsymbol{\epsilon})$ by $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ using (2.5.18), Corollary 2.27 and Proposition 2.63 (see Subsection 2.5.3, below). Furthermore, we have to take the subsequent discussion into account. (Note that $\varepsilon_1 < \varepsilon_2 < \varepsilon_3$ can be chosen arbitrarily small and thus (2.3.51) holds true with any $\varepsilon \in (0,1)$.)

To complete the proof we need to justify relations (2.5.45), (2.5.46) and (2.5.49). We observe that

$$\begin{aligned}
&\sup_{t\in J} w(\lfloor nt\rfloor/n)\big|S_n'(\lfloor nt\rfloor/n;\boldsymbol{\epsilon}) - S_n'(t;\boldsymbol{\epsilon})\big| \\
&\le \Big|\sum_{i=1}^n \boldsymbol{\epsilon}_i\Big|/n^{1/2}\Big(\sup_{t\in J}|\lfloor nt\rfloor - t|/n\Big)\Big(\sup_{t\in J} w(\lfloor nt\rfloor/n)\Big) \qquad (2.5.50)\\
&= \mathcal{O}_P(1)\mathcal{O}(1/n)\mathcal{O}(n^{1/2}) = \mathcal{O}_P(n^{-1/2})
\end{aligned}$$

holds true by the central limit theorem. It remains to study

$$\begin{aligned}
&\sup_{t\in J^1}\big|w(\lfloor nt\rfloor/n)S_n'(t;\boldsymbol{\epsilon}) - w(t)S_n'(t;\boldsymbol{\epsilon})\big| \\
&\le \big(\sup_{t\in J^1}|w(\lfloor nt\rfloor/n)/w(t) - 1|\big)\big(\sup_{t\in J} w(t)|S_n'(t;\boldsymbol{\epsilon})|\big),
\end{aligned} \qquad (2.5.51)$$

where $\sup_{t\in J} w(t)|S_n'(t;\boldsymbol{\epsilon})| = \mathcal{O}_P((\log\log n)^{1/2})$ which follows from Horváth *et al.* (1999, disp. (4.28)). Note that it holds that

$$\begin{aligned}
&\sup_{t\in J^1}|w(\lfloor nt\rfloor/n)/w(t) - 1| \\
&\le |w(\lfloor ns\rfloor/n)/w(s) - 1| = \mathcal{O}(|[w(\lfloor ns\rfloor/n)/w(s)]^2 - 1|).
\end{aligned} \qquad (2.5.52)$$

Now, we set $z_n = \lfloor ns\rfloor$ and $\xi = ns - \lfloor ns\rfloor$ and get

$$\begin{aligned}
&[w(z_n/n)/w((z_n+\xi)/n)]^2 - 1 \\
&= [(z_n+\xi)(n-z_n-\xi) - z_n(n-z_n)]/[z_n(n-z_n)] \\
&= [-2z_n\xi + \xi(n-\xi)]/[z_n(n-z_n)] \\
&= -2\xi/(n-z_n) + [\xi/z_n][(n-\xi)/(n-z_n)] = \mathcal{O}(1/z_n) = \mathcal{O}(1/(ns))
\end{aligned} \qquad (2.5.53)$$

since $|\xi| \le 1$ and $z_n = o(n)$. Furthermore, it holds that

$$(\log\log n)^{1/2}\exp(-(\log n)^{1-\varepsilon_1}) \le \exp(-(\log n)^{1-\varepsilon_3})$$

is equivalent to $\log((\log\log n)^{1/2}) \le (\log n)^{1-\varepsilon_1}(1 - (\log n)^{\varepsilon_1-\varepsilon_3})$ and the latter is fulfilled for sufficiently large n since $\varepsilon_3 > \varepsilon_1$. Hence, above relations (2.5.50) and (2.5.51) combined with (2.5.52) and (2.5.53) yield (2.5.49) which then together with (2.5.44) implies (2.5.45). It remains to show (2.5.46). Since $|S_n(\lfloor nt\rfloor/n;\boldsymbol{\epsilon})|$ is a random non-negative step function in t, and due to convexity of $w(t)$, we have the relation

$$\begin{aligned}
&\sup_{t\in J^l} w(\lfloor nt\rfloor/n)|S_n'(\lfloor nt\rfloor/n;\boldsymbol{\epsilon})| \\
&= \sup_{t\in J^l} w(t)|S_n'(\lfloor nt\rfloor/n;\boldsymbol{\epsilon})| \\
&= \sup_{t\in J^l} w(t)|S_n'(t;\boldsymbol{\epsilon})| + \mathcal{O}_P(n^{-1/2}),
\end{aligned} \qquad (2.5.54)$$

where, in the last step, we tacitly repeated the arguments of (2.5.50). The second relation of (2.5.44) holds also true if we replace the two-sided set J^2 by the one-sided set J^l. Hence, (2.5.54) together with the one-sided modification of (2.5.44) yield, in probability and as $n \to \infty$,

$$(2 \log\log n)^{-1/2} \sup_{t \in J^l} w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n; \boldsymbol{\epsilon})| = (1-\varepsilon_1)^{1/2} + o_P(1). \tag{2.5.55}$$

Finally, to show (2.5.46), it remains to conclude that

$$(2 \log\log n)^{-1/2} \sup_{t \in J^r} w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n; \boldsymbol{\epsilon})| = (1-\varepsilon_1)^{1/2} + o_P(1) \tag{2.5.56}$$

holds true. This follows on account of stationarity and symmetry similar to Proposition 2.63 (cf. Subsection 2.5.3, below). Therefore, we define $\boldsymbol{z}_i^{(n)} := \boldsymbol{\epsilon}_{n-i+1}$ and observe that for all fixed n the sequences $\{\boldsymbol{z}_i^{(n)}\}_{i\in\mathbb{N}}$ are strictly stationary and m-dependent. Hence,

$$\begin{aligned}
&\sup_{t \in J^r} w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n; \boldsymbol{\epsilon})| \\
&= \sup_{t \in J^l} w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n; \boldsymbol{z}^{(n)})| \\
&\overset{\mathcal{D}}{=} \sup_{t \in J^l} w(\lfloor nt \rfloor/n)|S_n'(\lfloor nt \rfloor/n; \boldsymbol{z}^{(1)})|
\end{aligned}$$

and (2.5.56) follows immediately via (2.5.55). □

We turn to the behavior of $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$ under H_A. The next proposition and corollary form extensions of Berkes *et al.* (2009, Lemma 2), Aston & Kirch (2012a, Lemma 3.2), Horváth *et al.* (2014) and Torgovitski (2015c, (8.12)). Proposition 2.58 will be used to show consistency of our test under m-dependence and m-approximability.

Proposition 2.58. *Let Assumption M1 or Assumption M2 together with the causal representation* (1.3.5) *(in the latter case) hold true. It holds under* H_A *that, as* $n \to \infty$,

$$\sup_{x\in[0,1]} \big\| S_n(x;Y)/n^{1/2} - \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j \big\| = \mathcal{O}_P(n^{-1/2}), \tag{2.5.57}$$

where $\mathcal{G}_{g_j}$ *are defined in* (2.3.52). *Furthermore, it holds that*

$$\sup_{x\in[0,1]} \big\| \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j \big\|^2 = \sup_{x\in[0,1]} \big[\sum_{j=1}^{\varrho} \mathcal{G}_{g_j}^2(x) \big] > 0. \tag{2.5.58}$$

Proof of Proposition 2.58. Using, e.g., Jirak (2013) we observe under Assumption M2 that

$$\sup_{x\in[0,1]} \|S_n(x;\varepsilon)\| = \mathcal{O}_P(1) \tag{2.5.59}$$

holds true, as $n \to \infty$. Under Assumption M1 we can deduce (2.5.59) from the i.i.d. case, e.g., by repeating the steps in the proof of Proposition 2.59 on p. 89, below. In both cases (2.5.59) implies

$$\sup_{x\in[0,1]} \big\| \big(\sum_{i=1}^{\lfloor nx \rfloor} Y_i - (\lfloor nx \rfloor/n) \sum_{i=1}^{n} Y_i \big)/n - \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j \big\|$$

$$\leq \sup_{x\in[0,1]} \Big\| \sum_{i=1}^{\lfloor nx\rfloor} \varepsilon_i - (\lfloor nx\rfloor/n) \sum_{i=1}^{n} \varepsilon_i \Big\|/n$$
$$+ \sum_{j=1}^{\varrho} \sup_{x\in[0,1]} \Big|\Big(\sum_{i=1}^{\lfloor nx\rfloor} g_j(i/n) - (\lfloor nx\rfloor/n) \sum_{i=1}^{n} g_j(i/n)\Big)/n - \mathcal{G}_{g_j}(x)\Big|$$
$$= \mathcal{O}_P(n^{-1/2})$$

by the piecewise Lipschitz continuity of the g_j's. Parseval's identity yields (2.5.58). □

Proof of Theorem 2.45. We proceed similarly as in the proof of Torgovitski (2016, Theorem 3.7) but in a more general setting, using operators $\hat{\mathscr{C}}^{-1/2,d}$. (Cf. also Torgovitski, 2015c, Theorem 4.7.) First, we observe

$$\begin{aligned}
\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) &= \max_{1\leq k<n} w(k/n) \|\hat{\mathscr{C}}^{-1/2,d} S_n(k/n;Y)\| \\
&\geq \max_{1\leq k<n} w(k/n) \|\hat{\mathscr{C}}^{-1/2,d'} S_n(k/n;Y)\| \\
&\geq w(1/2) \max_{1\leq k<n} \Big| \Big\| \Big[\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}\Big] S_n(k/n;Y)/n^{1/2} \Big\| \\
&\qquad - \|\mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}} S_n(k/n;Y)/n^{1/2}\| \Big| \Big[\frac{n^{1/2}(\log\log n)^{1/2}}{\beta_h^{1/2}(\log\log n)^{1/2}}\Big] \\
&\geq w(1/2) \Big| \max_{1\leq k<n} \Big\| \Big[\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}\Big] S_n(k/n;Y)/n^{1/2} \Big\| \\
&\qquad - \max_{1\leq k<n} \|\mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}} S_n(k/n;Y)/n^{1/2}\| \Big| \Big[\frac{n^{1/2}}{\beta_h^{1/2}(\log\log n)^{1/2}}\Big] (\log\log n)^{1/2}.
\end{aligned} \tag{2.5.60}$$

Now, on the one hand, we get in view of Proposition 2.58

$$\sup_{x\in[0,1]} \|\mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}} S_n(x;Y)/n^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}} \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j\| = o_P(1),$$

and due to Assumption G (ii) we obtain

$$\sup_{x\in[0,1]} \|\mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}} \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j\|^2 = \sup_{x\in[0,1]} \sum_{r=1}^{d'} \lambda_r^{-1} \langle \sum_{j=1}^{\varrho} \mathcal{G}_{g_j}(x)\Delta_j, v_r\rangle^2 > 0.$$

We get, on the other hand, that

$$\begin{aligned}
&\max_{1\leq k<n} \Big\| \Big[\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}\Big] S_n(k/n;Y)\|/n^{1/2} \\
&\leq \|\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}\|_{\mathcal{L}} \max_{1\leq k<n} \|S_n(k/n;Y)\|/n^{1/2} \\
&\leq \|\hat{\mathscr{C}}^{-1/2,d'} \beta_h^{1/2} - \mathscr{C}^{-1/2,d'}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}\|_{\mathcal{S}} \max_{1\leq k<n} \|S_n(k/n;Y)\|/n^{1/2} = o_P(1).
\end{aligned}$$

The last relation holds true in view of Corollary 2.34 and, again, due to Proposition 2.58. The proof is finished since we have $\beta_h = o_P(n/(\log\log n))$ in (2.5.60). □

Proof of Corollary 2.50 (as given in Corollary 5.1 of Torgovitski, 2015c). The change corrected statistic $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta})$ only differs in the first component of the scores from the original statistic $\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}})$. Hence, we obtain via the Cauchy-Schwarz inequality that

$$
\begin{aligned}
&|\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}) - \hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta})| \\
&\le \max_{1\le k<n} w(k/n)|S_n(k/n;\hat{\boldsymbol{\epsilon}}) - S_n(k/n;\hat{\boldsymbol{\epsilon}}^{\Delta})|_{\hat{\Sigma}} \\
&= \max_{1\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\|\|\hat{v}_1 - \hat{v}_1^{\Delta}\|/|\hat{\lambda}_1|^{1/2} \\
&\le \max_{1\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\|\frac{\|\hat{v}_1\||1-\|\hat{v}_1+n^{\gamma}\hat{s}\hat{u}\||+\|n^{\gamma}\hat{u}\|}{\|\hat{v}_1+n^{\gamma}\hat{s}\hat{u}\|}\frac{1}{|\hat{\lambda}_1|^{1/2}} \\
&\le 2\max_{1\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\|\frac{\|n^{\gamma}\hat{u}\|}{\|\hat{v}_1+n^{\gamma}\hat{s}\hat{u}\|}\frac{1}{|\hat{\lambda}_1|^{1/2}} &\quad(2.5.61)\\
&= \mathcal{O}_P((\log n)^{1/2}n^{-1/2+\gamma}) = o_P((\log\log n)^{-1/2}), &\quad(2.5.62)
\end{aligned}
$$

where we rely on Proposition 2.63 (cf. Subsection 2.5.3, below). In (2.5.61) we used that

$$|1-\|\hat{v}_1+n^{\gamma}\hat{s}\hat{u}\|| = |\|\hat{v}_1\| - \|\hat{v}_1+n^{\gamma}\hat{s}\hat{u}\|| \le \|n^{\gamma}\hat{u}\|$$

and that the eigenfunctions are normalized, i.e. $\|\hat{v}_1\|=1$. Furthermore, we have

$$\|n^{\gamma}\hat{u}\| = \max_{1\le k<n}\|S_n(k/n;\varepsilon)\|n^{-1/2+\gamma} = \mathcal{O}_P(n^{-1/2+\gamma}) = o_P(1)$$

in view of (2.5.59) and since $\gamma<1/2$. Finally, note that the first step in (2.5.62) holds true due to $\hat{\lambda}_1 = \lambda_1 + o_P(1)$ and due to $\|\hat{v}_1\| = \|v_1\| + o_P(1) = 1 + o_P(1)$, which are both consequences of $\|\mathscr{C}-\hat{\mathscr{C}}\|_{\mathcal{S}} = o_P(1)$. □

Proof of Theorem 2.51 (as given in Theorem 5.2 of Torgovitski, 2015c). The relations in (2.5.57) and (2.5.58) yield that

$$\|\hat{u}\| = \max_{1\le k<n}\|S_n(k/n;Y)/n^{1/2}\| = \sup_{x\in[0,1]}\Big[\sum_{j=1}^{\varrho}\mathcal{G}_{g_j}^2(x)\Big]^{1/2} + o_P(1) =: \mathscr{S} + o_P(1)$$

and, similarly, that $\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\| = \mathscr{S}+o_P(1)$ with $\mathscr{S}>0$ since $\gamma>0$ and $\|\hat{v}_1\|=1$. It is sufficient to consider the behavior of the statistic in the first direction which corresponds to

$$
\begin{aligned}
&\hat{\mathscr{M}}_n(\hat{\boldsymbol{y}}^{\Delta}) \\
&\ge \max_{1\le k<n} w(k/n)|\langle S_n(k/n;Y)/n^{1/2}, \hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\rangle| \\
&\qquad\qquad\qquad\times\frac{|(n/\log\log n)/\hat{\lambda}_1|^{1/2}}{\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\|}(\log\log n)^{1/2} \\
&\ge w(1/2)\Big|\max_{1\le k<n}|\langle S_n(k/n;Y)/n^{1/2}, \hat{v}_1/n^{\gamma}\rangle| \\
&\quad -\max_{1\le k<n}|\langle S_n(k/n;Y)/n^{1/2},\hat{u}\rangle|\Big|\frac{|(n/\log\log n)/\beta_h|^{1/2}|\beta_h/\hat{\lambda}_1|^{1/2}}{\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\|}(\log\log n)^{1/2}.
\end{aligned}
$$

Via (2.5.57) we observe, on the one hand, that

$$\max_{1\le k<n} \frac{|\langle S_n(k/n;Y)/n^{1/2}, \hat{v}_1/n^{\gamma}\rangle|}{\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\|} \le \frac{\|\hat{u}\|}{\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\|}\frac{\|\hat{v}_1\|}{n^{\gamma}} = \mathcal{O}_P(1)\mathcal{O}_P(n^{-\gamma})$$

and, on the other hand, by evaluating at $k=\hat{k}$, that

$$\max_{1\le k<n} \frac{|\langle S_n(k/n;Y)/n^{1/2}, \hat{u}\rangle|}{\|\hat{v}_1/n^{\gamma}+\hat{s}\hat{u}\|} \ge \|\hat{u}\|\frac{\|\hat{u}\|}{\|\hat{v}_1/n^{\gamma}+\hat{u}\|} = \mathscr{S}+o_P(1).$$

Finally, let (λ_j, v_j) be the eigenelements of $\mathscr{C}_{\alpha,\beta_{\mathcal{K}}\beta_{\mathcal{G}}}$ with $\alpha=\lim_{n\to\infty}1/\beta_h$, where $\beta_h,\ \beta_{\mathcal{K}},\ \beta_{\mathcal{G}}$ are set as in (2.3.30) and recall the convergence of the first eigenvalue $\hat{\lambda}_1$ due to (2.3.31) in Corollary 2.34. The assertion follows by $(n/\log\log n)/\beta_h\to\infty$ and by $P(|\hat{\lambda}_1/\beta_h|<2\lambda_1)\to 1$, as $n\to\infty$, where $\lambda_1>0$ in view of Assumption G (i). □

Auxiliary results for proofs 2.5.3 (Convergence rates for partial sums of Hilbert space time series)

This section contains limit theorems for weighted partial sums that were used in the previous subsection. All proofs are given at the end of this section.

Proposition 2.59. *Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ fulfill Assumption M1 with $\kappa=2$.*[1] *Then it holds that, as $n\to\infty$,*

$$\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}\varepsilon_i\Big\|/k^{1/2} = \mathcal{O}_P\big((\log\log n)^{1/2}\big)$$

and, similarly, the time-inversed partial sums fulfill

$$\max_{1\le k<n}\Big\|\sum_{i=1}^{k}\varepsilon_{-i}\Big\|/k^{1/2} = \mathcal{O}_P\big((\log\log n)^{1/2}\big)$$

since $\{\varepsilon_{-i}\}_{i\in\mathbb{Z}}$ fulfills Assumption M1 with $\kappa=2$, too.

Proposition 2.59 is proven under Assumption M1 using the law of the iterated logarithm.[2] We need also an analogue of this result under Assumption M2. However, it seems that no law of iterated logarithm is proven in the infinite dimensional setting, i.e. when $d_H=\infty$, under the generality of Assumption M2. To this end we follow another rather general and well known approach to state a result in Proposition 2.62 which is similar to Proposition 2.59 by using a combination of theorems of Móricz (1976),

[1] Note that Assumption M1 relies on Assumption S1 which was formally restricted to $\kappa>2$ for the sake of a clearer presentation in Subsection 2.3.3. In this section we may tacitly assume $\kappa=2$ to be also admissible.

[2] Note that in finite dimensional frameworks the $(\log\log n)^{1/2}$-rate is implied by Darling-Erdős-type results. We are not aware of any suitable extensions of finite-dimensional Darling-Erdős-type results to our infinite dimensional situation of $d_H=\infty$.

Fazekas & Klesov (2001) and of Berkes *et al.* (2011). The rates are slightly weaker but the result is still interesting on its own. Proposition 2.62 will be a consequence of the next Proposition 2.60 together with Proposition 2.61 and is essentially based on rates for moments of unweighted partial sums. To be more specific, we will assume that

$$E\Big\|\sum_{i=1}^{n}\varepsilon_i\Big\|^{\kappa}=\mathcal{O}(n^{\kappa/2}) \tag{2.5.63}$$

holds true for some $\kappa>2$ and as $n\to\infty$. We begin with a proposition under broad conditions to provide a connection of (2.5.63) to maxima of partial sums.

Proposition 2.60. *Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ be an H-valued time series such that (2.5.63) holds true for some $\kappa>2$. Then it holds that, as $n\to\infty$,*

$$\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}\varepsilon_i\Big\|/k^{1/2}=\mathcal{O}_P\big((\log n)^{1/\kappa}\big). \tag{2.5.64}$$

Proposition 2.61. *Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ fulfill Assumption M2 with some $\kappa\ge 2$. Then (2.5.63) holds true.*

Proposition 2.62. *Let $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ fulfill Assumption M2 with some $\kappa>2$. Then it holds that, as $n\to\infty$,*

$$\max_{1\le k<n}\Big\|\sum_{i=1}^{k}\varepsilon_i\Big\|/k^{1/2}=\mathcal{O}_P((\log n)^{1/2})$$

and similarly the time-inversed partial sums fulfill

$$\max_{1\le k<n}\Big\|\sum_{i=1}^{k}\varepsilon_{-i}\Big\|/k^{1/2}=\mathcal{O}_P((\log n)^{1/2})$$

since $\{\varepsilon_{-i}\}_{i\in\mathbb{Z}}$ fulfills Assumption M2 with same parameter κ, too.

The next proposition is a consequence of the previous Proposition 2.59 and Proposition 2.62.

Proposition 2.63. *It holds under Assumption M1 or Assumption M2 that, as $n\to\infty$,*

$$\max_{1\le k<n}w(k/n)\|S_n(k/n;\varepsilon)\|=\mathcal{O}_P(g(n)),$$

with $g(n)=(\log\log n)^{1/2}$ in the former and $g(n)=(\log n)^{1/2}$ in the latter situation.

Remark 2.64. Rates such as (2.5.63) are shown in the functional framework by Berkes *et al.* (2013, cf. Theorems 3.1 and 3.3) for causal m-approximable time series with any $\kappa\in(2,3)$ in which case (2.5.64) holds true. Since we need (2.5.64) also for time-inversed series $\{\varepsilon_{-i}\}_{i\in\mathbb{Z}}$ in Subsection 2.5.2, we state a corresponding extension of Berkes *et al.* (2013, Theorem 3.3) for the non-causal case in Proposition 2.61.

Furthermore, note that Berkes *et al.* (2013, Theorem 3.3) is stated without an explicit proof. The authors refer instead to the related proof of Theorem 3.1 which deals with m-dependent approximations and is not adequate for our purposes. It could be possible to adapt Theorem 3.3 of Berkes *et al.* (2013) to the non-causal case via Theorem 3.1 but, on the one hand, it would be an unnecessary detour, and on the other hand, it is easy to see that substantial modifications would be required. Thus, it appears to be more transparent and simpler to rely directly on Berkes *et al.* (2011, Proposition 4) for our purposes. The latter is a univariate two-sided version of Berkes *et al.* (2013, Theorem 3.3).

The next proposition is not related to the previous Propositions 2.59-2.63 but is used in Subsection 2.5.2 as well.

Proposition 2.65. *Let* $\{\varepsilon_i\}_{i\in\mathbb{Z}}$ *be an* H*-valued, centered time series such that*

$$\sum_{i,j=1}^{n} |E\langle\varepsilon_i,\varepsilon_j\rangle| = \mathcal{O}(n) \tag{2.5.65}$$

holds true as $n\to\infty$*. Furthermore, let* $\{a_{n,i}\}_{n,i\in\mathbb{N}}$*,* $a_{n,i}\in\mathbb{R}$*, be an array which fulfills*

$$n^{\nu/2}\max_{1\le i\le n}|a_{n,i}| = \mathcal{O}(1),$$

$\nu\ge 0$*, as* $n\to\infty$*. Then it holds that* $E\|\sum_{i=1}^{n} a_{n,i}\varepsilon_i\|^2 = \mathcal{O}(n^{1-\nu})$ *as* $n\to\infty$*.*

Remark 2.66 (on condition (2.5.65) of Proposition 2.65). Condition (2.5.65) holds true in both cases of Assumption M1 and Assumption M2 for the following reasons. Under stationarity we have

$$\sum_{i,j=1}^{n} |E\langle\varepsilon_i,\varepsilon_j\rangle| \le n\sum_{r=1}^{n}|E\langle\varepsilon_0,\varepsilon_r\rangle|$$

and $|E\langle\varepsilon_0,\varepsilon_r\rangle|\le E\|\varepsilon_0\|^2$. Now, on the one hand, $|E\langle\varepsilon_0,\varepsilon_r\rangle| = 0$ holds true for $r>m$ under Assumption M1 due to m-dependence. On the other hand, $|E\langle\varepsilon_0,\varepsilon_r\rangle|\le c\delta(r)$, $c>0$, holds true in the setting of Assumption M2. The latter follows via the basic decomposition

$$\langle\varepsilon_0,\varepsilon_j\rangle = \langle\varepsilon_0-\varepsilon_0^{(j)},\varepsilon_j\rangle + \langle\varepsilon_0^{(j)},\varepsilon_j-\varepsilon_j^{(j)}\rangle + \langle\varepsilon_0^{(j)},\varepsilon_j^{(j)}\rangle \tag{2.5.66}$$

which yields

$$\begin{aligned}|E\langle\varepsilon_0,\varepsilon_j\rangle| &\le E\|\varepsilon_0-\varepsilon_0^{(j)}\|\|\varepsilon_j\| + E\|\varepsilon_j-\varepsilon_j^{(j)}\|\|\varepsilon_0^{(j)}\| + E\langle\varepsilon_0^{(j)},\varepsilon_j^{(j)}\rangle\\ &\le 2\Big[E\|\varepsilon_0-\varepsilon_0^{(j)}\|^2E\|\varepsilon_0\|^2\Big]^{1/2}\le c\delta(j)\end{aligned}$$

for all $j\in\mathbb{N}$ and some $c>0$.

We turn to the proofs of Propositions 2.59 - 2.65.

Proof of Proposition 2.59. For any $k \in \mathbb{N}$ we decompose

$$\sum_{i=1}^{k} \varepsilon_i = \sum_{j=1}^{N_{k,1}} \tilde{\varepsilon}_{j,1} + \ldots + \sum_{j=1}^{N_{k,m+1}} \tilde{\varepsilon}_{j,m+1}$$

into $(m+1)$ partial sums of i.i.d. random variables $\tilde{\varepsilon}_{j,r} := \varepsilon_{(j-1)(m+1)+r}$ and where $N_{k,1}, \ldots, N_{k,m+1} \in \mathbb{N}$ are set in an obvious manner. We observe

$$\begin{aligned}\max_{1\le k\le n} \|\sum_{i=1}^{k} \varepsilon_i\|/k^{1/2} &\le \max_{1\le i\le (m+1)} \Big[\max_{1\le k\le n} \mathbb{1}_{\{N_{k,i}\neq 0\}} \|\sum_{j=1}^{N_{k,i}} \tilde{\varepsilon}_{j,i}\|/N_{k,i}^{1/2}\Big] \\ &\qquad \times (m+1) \max_{1\le i\le (m+1)} \Big[\max_{1\le k\le n} N_{k,i}/k\Big]^{1/2} \\ &\le (m+1) \max_{1\le i\le (m+1)} \max_{1\le k\le n} \|\sum_{j=1}^{k} \tilde{\varepsilon}_{j,i}\|/k^{1/2}.\end{aligned}$$

Hence, an application of the law of iterated logarithm for i.i.d. Hilbert space random elements (cf., e.g., Theorem 8.6 of Ledoux & Talagrand, 1991) completes the proof. □

Proof of Proposition 2.60. This is a simplified »in expectation« version of the corresponding proof of Torgovitski (2016, Proposition 3.11). The well-known results of Móricz (1976) state that moment inequalities for partial sums yield analogous moment inequalities for maxima of partial sums. As shown, e.g., in Section B.1 of Kirch (2006) those results may be combined with Fazekas & Klesov (2001, Theorem 1.1) to obtain Hájek-Rényi type inequalities for weighted maxima of partial sums. The mentioned results are stated for univariate real-valued random variables. However, carefully inspecting the proofs of Móricz (1976, Theorem 1) and of Fazekas & Klesov (2001, Theorem 1.1) we observe that they can be restated in our Hilbert space setting with $\kappa > 2$, as well. Therefore, Móricz (1976, Theorem 1) together with assumption (2.5.63) yield

$$E\Big[\max_{1\le k\le n} \|\sum_{i=1}^{k} \varepsilon_i\|^{\kappa}\Big] \le c_1 n^{\kappa/2}$$

for all $n \in \mathbb{N}$ and some $c_1 > 0$. Next, we use $n^{\kappa/2} = \mathcal{O}(\sum_{k=1}^{n} k^{\kappa/2-1})$ and apply Fazekas & Klesov (2001, Theorem 1.1), taking (2.5.63) into account, to obtain,

$$E\Big[\max_{1\le k\le n} \|\sum_{i=1}^{k} \varepsilon_i\|^{\kappa}/k^{1/2}\Big] \le c_2 \sum_{k=1}^{n} k^{-1} \le c_3 \log n,$$

for all $n \in \mathbb{N}$ and some $c_2, c_3 > 0$ which finishes the proof. □

Proof of Proposition 2.61. The assertion follows directly by a straightforward extension of Berkes *et al.* (2011, Proposition 4). More details on this extension were given previously in Proposition 3.9 of Torgovitski (2016, arXiv:1407.3625v1) under the restriction of $\kappa \in [2, 3)$, which relies on Berkes *et al.* (2013, Lemma 3.1). One of the authors (G. Rice, in a private communication) pointed out that the application of the latter lemma can be easily avoided in our situation and thus the result holds true for all $\kappa \ge 2$ (cf. also Section A of Aue *et al.*, 2015). □

Proof of Proposition 2.63. We adapt the proof of Torgovitski (2015a, Lemma 6.2) to both m-dependent and m-approximable scenarios. In the following we use (again) the shifting technique that helped us several times within this chapter. Set $z_i^{(n)} := \varepsilon_{n-i+1}$ and observe that for any fixed $n \in \mathbb{N}$ the sequences $\{z_i^{(n)}\}_{i\in\mathbb{N}}$ are again strictly stationary and fulfill Assumption M1 or Assumption M2, respectively. Now, observe that due to symmetry and stationarity we obtain

$$\begin{aligned}
&\max_{1\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\| \\
&\le \max_{1\le k\le n/2} w(k/n)\|S_n(k/n;\varepsilon)\| + \max_{n/2\le k<n} w(k/n)\|S_n(k/n;\varepsilon)\| \\
&\le \max_{1\le k\le n/2} w(k/n)\|S_n(k/n;\varepsilon)\| + \max_{1\le k\le n/2} w(k/n)\|S_n(k/n;z^{(n)})\|,
\end{aligned}$$

where

$$\max_{1\le k\le n/2} w(k/n)\|S_n(k/n;z^{(n)})\| \overset{\mathcal{D}}{=} \max_{1\le k\le n/2} w(k/n)\|S_n(k/n;z^{(1)})\|,$$

and also that

$$\begin{aligned}
&\max_{1\le k\le n/2} w(k/n)\|S_n(k/n;\varepsilon)\| \\
&\le \max_{1\le k\le n/2} (n/(n-k))^{1/2}\Big\|\sum_{i=1}^{k}\varepsilon_i\Big\|/k^{1/2} \\
&\quad + \max_{1\le k\le n/2} (n/(n-k))^{1/2}(k/n)^{1/2}\Big\|\sum_{i=1}^{n}\varepsilon_i\Big\|/n^{1/2} \\
&\le 4\max_{1\le k\le n}\Big\|\sum_{i=1}^{k}\varepsilon_i\Big\|/k^{1/2}
\end{aligned} \tag{2.5.67}$$

holds true for all $n \in \mathbb{N}$. In the situation of Assumption M2 we obtain the $\mathcal{O}_P(g(n))$ bound for (2.5.67) via Proposition 2.62 and in the situation of Assumption M1 we may rely on Proposition 2.59. Finally, the same rate applies to

$$\max_{1\le k\le n/2} w(k/n)\|S_n(k/n;z^{(1)})\|$$

and thus the proof is complete. □

Proof of Proposition 2.65. It holds that $E\|\sum_{i=1}^n a_{n,i}\varepsilon_i\|^2 \le \sum_{i,j=1}^n |a_{n,i}||a_{n,j}||E\langle\varepsilon_i,\varepsilon_j\rangle|$, which directly implies the assertion. □

Notes 2.6

(Relation of this chapter to previous publications and preprints)

As explained in the general introduction of Chapter 1 (and in the introduction in Section 2.1 of Chapter 2) within this chapter we combined and extended the findings of Torgovitski (2015a, 2016, 2015c) to establish a unifying theory and to embed all our results of this chapter in the same Hilbert space change point setting. For the sake of clarity and for convenience we provide the details of the relations between this chapter and Torgovitski (2015a, 2016, 2015c), below. Many results are new derivations from the aforementioned works and also their presentation is different. Furthermore, note that some proofs in this chapter are substantial modifications and extensions (with notational adaptations) of the proofs presented in the latter articles, whereas others are only notational or technical adaptations. For this reason several proofs in Section 2.5 contain separate references to the previous publications, as well.

Assumptions: Assumption 2.4 and Assumption G are new. Assumptions S1, M1 and M2 combine our general assumptions in Torgovitski (2015c) with Assumption M in Torgovitski (2016) and Assumption 1 in Torgovitski (2015a). Assumption S1', Assumption M1' and Assumption M2' are implicitly contained in Torgovitski (2015a, 2016, 2015c). Assumption 2.11 is implicitly contained in (3.1) in Torgovitski (2015c). Assumption 2.13 and Assumption E2 are implicitly contained in Torgovitski (2015a, 2016, 2015c). Assumption E1 is implicitly assumed in Torgovitski (2015c). Assumption E3 corresponds to (3.5) of Torgovitski (2015a). Assumption E3' extends Assumption 4.1 from Torgovitski (2015c). Assumption K1 originates from (3.2) of Torgovitski (2015a) and its counterpart Assumption K2 is closely related to assumptions of Torgovitski (2016, Section 4) and Torgovitski (2015c, p. 6). Assumption PB is partly contained in Theorem 4.1 of Torgovitski (2016).

Definitions: The Definitions 2.14, 2.15 and 2.17 are implicitly used in Torgovitski (2015a, 2016) and in Torgovitski (2015c).

Theorems: Theorem 2.24 extends Theorem 4.1 of Torgovitski (2016). Theorem 2.32 extends both Theorem 8.3 of Torgovitski (2015c) and Lemma 3.3 of Torgovitski (2015a) and is indicated in this generality in the former article. Theorem 2.40 combines Theorem 4.1 of Torgovitski (2015a) with Theorem 3.4 of Torgovitski (2016). Theorem 2.43 corresponds to Theorem 4.2 of Torgovitski (2015a). Theorem 2.45 combines and extends Theorem 4.4 of Torgovitski (2015a), Theorem 3.7 of Torgovitski (2016) as well as Propositions 4.6 and 4.7 of Torgovitski (2015c). Theorem 2.51 adapts Theorem 5.2 of Torgovitski (2015c). Theorem 2.56 extends Theorem 8.3 from Torgovitski (2015c) and is indicated therein in this generality.

Propositions: Proposition 2.22 corresponds to Proposition 3.1 of Torgovitski (2015a). Proposition 2.54 is implicitly contained in Proposition 4.2 of Torgovitski (2015c). Proposition 2.57 combines Lemma 6.3 of Torgovitski (2015a) and (3.7) with Theorem 3.10 of Torgovitski (2015c). Proposition 2.58 extends Torgovitski (2015c, (8.12)) and is indicated therein in this generality. (Cf. also Torgovitski, 2015a, Lemma 6.8.) Proposition 2.59 is contained in Lemma 6.2 of Torgovitski (2015a). Propositions 2.60, 2.61, 2.62 and 2.63 extend Propositions 3.11 and 3.12 as well as Corollary 3.13 of Torgovitski (2016) and Lemma 6.2 of Torgovitski (2015a). Proposition 2.65 slightly extends Proposition 8.1 from Torgovitski (2015c). Note that a similar argument was used in the proof of Torgovitski (2015a, Lemma 3.3).

Lemmas: Lemma 2.55 extends Lemma 8.2 from Torgovitski (2015c) and is indicated therein in this generality.

Corollaries: Corollary 2.26 extends Theorem 4.1 (and the subsequent discussion) of Torgovitski (2016). Corollary 2.27 is new and merges Theorem 4.1 of Torgovitski (2016) with Corollary 3.2 of Torgovitski (2015a) and with Proposition 4.2 of Torgovitski (2015c).

Corollary 2.34 extends Lemma 3.3 of Torgovitski (2015a). The one-directional case is used implicitly in Proposition 4.7 of Torgovitski (2015c). Corollary 2.42 corresponds to Corollary 4.3 of Torgovitski (2015a) and is indicated in Torgovitski (2016). Corollary 2.50 adapts Corollary 5.1 of Torgovitski (2015c).

Examples: Example 2.23 is new but is implicitly contained in Section 4 of Torgovitski (2016). Example 2.33 is new.

Remarks: Remarks 2.1, 2.2, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, 2.12, 2.19, 2.20, 2.25, 2.28, 2.29, 2.30, 2.31, 2.35, 2.39, 2.41, 2.44, 2.46, 2.47 and 2.52 are new. Remark 2.3 stays close to the introduction in Torgovitski (2015a). Remarks 2.53, 2.16 and 2.18 are implicitly contained in Torgovitski (2015a, 2016) and in Torgovitski (2015c). Remark 2.21 is implicitly used in Torgovitski (2015c) and in Torgovitski (2016). Remarks 2.36, 2.37 and 2.38 are partly (implicitly) contained in Torgovitski (2015a, 2016). Remark 2.48 and Remark 2.49 are new but related to Remark 4.5 of Torgovitski (2015a). Remark 2.64 is close to Torgovitski (2016, p. 9). Remark 2.66 was implicitly used in Theorem 8.3 of Torgovitski (2015c).

Figures and Tables: Figure 2.1 is new. Figures 2.2, 2.3, 2.4 and 2.5 are also new but related to the Tables in Torgovitski (2015a, 2016). Finally, Table 2.1 is new.

CHAPTER 3

CHANGE POINT ESTIMATION FOR HILBERT SPACE VALUED DATA

> *"Common breaks in panel data are wide spread phenomena. For example, a credit crunch or debt crisis may affect every company's stock returns, and an oil price shock may impact every country's output. A tax policy change may alter each firm's investment incentive. A fad or fashion can influence a large section of the society.*
>
> — Bai (2010, p. 78)

Introduction 3.1

In this chapter we assume a model of $d \in \mathbb{N}$ panels

$$\{Y_{1,k}, \dots, Y_{n,k}\}_{k=1,\dots,d} \tag{3.1.1}$$

of similarly structured time series of length $n \in \mathbb{N}$ and we focus on a *»common change in the means«* model, where we assume that a change occurs simultaneously within some fraction of the panels (3.1.1) at the same time point $u \in \{1, \dots, n-1\}$. Moreover, we assume that all affected panels have on average a comparable change magnitude (see (3.2.6) and (3.2.9), below). Bleakley & Vert (2011a, p. 1) emphasize that such settings, and especially multiple change point extensions thereof, are important in many applications (e.g. in image processing and for the analysis of genomic profiles). Our interest is, as in the latter article, to study *»accuracy«* of a change point estimate $\hat{u}$ for a common change point u, where accuracy denotes $P(\hat{u} = u)$. We consider a HDLSS asymptotic framework, where the sample size n is fixed and the dimension d, i.e. the number of panels, increases. Asymptotic accuracy of change point estimates $\hat{u} = \hat{u}_d$ in such an asymptotic framework, i.e. $\lim_{d\to\infty} P(\hat{u} = u)$, was only recently addressed in the literature and to the best of our knowledge first results are by Bai (2010) and by the already mentioned Bleakley & Vert (2010, 2011a).[1] Motivated by studies of segmentation methods for genomic profile data, Bleakley & Vert (2010) sought for multiple change point estimates that provide a trade-off between, on the one hand, the faster but less accurate local methods (such as binary segmentation) and, on the other hand, the slower but more accurate global approaches (such as least squares methods). As a solution Bleakley & Vert (2010, 2011a) came up with a global »total variation denoising« method[2] which is a multivariate extension of the LASSO-type suggestion of Harchaoui & Lévy-Leduc (2008). Empirical studies in Bleakley & Vert (2011a) demonstrate promising computational properties and a rather high accuracy of this estimate. (Note that the contributions of this chapter rely largely on the latter work.)

Analogously to Chapter 2 this chapter combines and extends the results of Torgovitski (2015d) and of the corresponding preprints Torgovitski (2015d, arXiv:1501.00177v1, 1501.00177v2). The theory of the latter articles is furthermore simplified and presented in more detail. The various directions of generalizations will be explained later on. First, note that we focus on a class of weighted CUSUM estimates whereas Bleakley & Vert (2010, 2011a) consider weighted, penalized least squares estimates. We will discuss that both types of estimates are equivalent for panels of real-valued univariate time series under a single change point scenario (see p. 114, below), which is a known result in the traditional real-valued univariate setting. Due to this equivalence most of our results for CUSUM estimates - restricted to the real-valued univariate setting - apply to the total variation denoising approach and vice versa, i.e. the results of Bleakley & Vert (2010, 2011a) may be interpreted for the CUSUM estimates as well. Note that Bleakley & Vert (2010, 2011a) consider an i.i.d. single change point setting under Gaussianity and derive sharp conditions on change point locations and change sizes that ensure $\lim_{d\to\infty} P(\hat{u} = u) = 1$ for the weighting w_0, which is defined in (1.2.1). Furthermore, Bleakley & Vert (2011a) show that the Darling-Erdős-type weighting function $w_{1/2}$, which is also defined in (1.2.1), leads asymptotically to a higher accuracy than w_0 and ensures

$$\lim_{d\to\infty} P(\hat{u} = u) = 1 \tag{3.1.2}$$

under broad assumptions on change point locations and change sizes. In this thesis we will extend the findings of Bleakley & Vert (2010, 2011a) to a non-parametric setting of panels of weakly dependent Hilbert space valued time series that share the same autocovariance structure. First, we will show for weighted CUSUM estimates that the asymptotic accuracy properties (3.1.2) are determined analytically by the shape of the so-called *»critical function«* C (cf. Definition 3.5, below, and see also Torgovitski, 2015d, arXiv:1501.00177v1). Based on this observation we derive sharp bounds for asymptotic accuracy of estimates with traditional CUSUM weights w_γ, $\gamma \in (0, 1/2)$ (as set in (1.2.1)), which extend the bounds for w_0 and $w_{1/2}$ of Bleakley & Vert (2011a). Moreover, we propose a new covariance-based weighting scheme $w_\star$ that ensures (3.1.2) within a rather broad range of dependence models under the same assumptions on the change locations and magnitudes as required for the weighting $w_{1/2}$ in the i.i.d. setting. (In this sense these weights $w_\star$ are optimal.) The admissible dependence models are

[1] See also Hadri *et al.* (2012) and Kim (2014).

[2] A penalized least squares method with a »total variation penalty«.

characterized by a monotonicity property of the so-called *»variance of cumulated noises«*. This property holds true for panels of MA(1) time series and for further (small) deviations from the i.i.d. assumption. Furthermore, we will see that weights $w_\star$ coincide with the Darling-Erdős-type weights $w_{1/2}$ in a white noise setting but outperform the latter weights in terms of asymptotic accuracy for dependent data under the mentioned monotonicity condition. Note that $w_{1/2}$ necessarily leads to spurious estimation under dependence if the change is sufficiently small. Finally, notice that we indicate in Torgovitski (2015b,d) that our findings are applicable to random changes, to multiple changes and to spatial data settings, as well. Particularly, our findings may serve as a motivation for further studies of covariance-based weightings for the group fused LASSO approach in multiple change point settings.

As in Section 2.1 and for convenience of the reader, we explain some main similarities and differences between this chapter, Torgovitski (2015d) and Torgovitski (2015d, arXiv:1501.00177v1, 1501.00177v2, 1501.00177v3). First, note that the final publication Torgovitski (2015d) includes a common factor model which is not contained in the corresponding preprints. Also note that the signal plus noise model in the latter article is only stated for panels of univariate time series but that a Hilbert space panel data model is indicated in the preprints Torgovitski (2015d, arXiv:1501.00177v1, 1501.00177v2). In all these publications, monotonicity conditions for uniform accuracy are only evaluated for MA(1) time series. Finally, note that the estimation of the covariance-based weighting scheme is, for one thing, only discussed informally in all previous publications and, for another, restricted to a univariate setting. In this chapter most of the results of Torgovitski (2015d) and of the corresponding preprints are extended to panels of Hilbert space valued time series including common factors. Our theoretical results are stated in a more convenient time-rescaled setting which simplifies the presentation substantially. Furthermore, we provide a convenient expression for the covariance-based weighting function $w_\star$ which clarifies its relation to the Darling-Erdős-type weighting function $w_{1/2}$ and also allows for an elegant verification of the monotonicity conditions. The latter conditions are evaluated for MA(q) and AR(1) time series. Particularly the AR(1) case provides new counterexamples that show the limits of the presented theory.

We will provide additional details on the relation between the results of this chapter and the results in Torgovitski (2015d) at the end of this chapter in Section 3.6.

Structure of the chapter

We begin by introducing our common change point panel model in Section 3.2 where we start with a Hilbert space signal plus noise setting, present our general assumptions in Subsection 3.2.1 and proceed with CUSUM estimates in Subsection 3.2.2. The corresponding theory is presented in Section 3.3. Here, we start with our key result on the *»critical function«* in Theorem 3.6 which involves the *»variance of cumulated noises«* and forms the basis for most of the subsequent considerations. Properties of the variance of cumulated noises are discussed in Subsection 3.3.1 and then consistency properties of estimates are derived in Subsection 3.3.2. The first part of the latter subsection is on pointwise asymptotic accuracy of classical weightings whereas the second part is on uniform asymptotic accuracy of new covariance-based weightings as well as on consistency of their estimates. This section is complemented by a discussion of the relation between the weighted CUSUM and the weighted *»group fused LASSO«* approaches in Subsection 3.3.3. Some simulation results are shown in Section 3.4. All proofs are given at the end of this chapter in Section 3.5. Finally, some notes on the relation to the previous publications and preprints complement this chapter in Section 3.6.

A change point problem 3.2

Signal plus noise model (with common factors) 3.2.1

We assume a signal plus noise panel design with Hilbert space valued time series and with common factors given by

$$\begin{bmatrix} Y_{1,1} & \cdots & Y_{1,d} \\ \vdots & \ddots & \vdots \\ Y_{n,1} & \cdots & Y_{n,d} \end{bmatrix} = \begin{bmatrix} m_{1,1} & \cdots & m_{1,d} \\ \vdots & \ddots & \vdots \\ m_{n,1} & \cdots & m_{n,d} \end{bmatrix} + \begin{bmatrix} \varepsilon_{1,1} & \cdots & \varepsilon_{1,d} \\ \vdots & \ddots & \vdots \\ \varepsilon_{n,1} & \cdots & \varepsilon_{n,d} \end{bmatrix} + \begin{bmatrix} \gamma_1\zeta_1 & \cdots & \gamma_d\zeta_1 \\ \vdots & \ddots & \vdots \\ \gamma_1\zeta_n & \cdots & \gamma_d\zeta_n \end{bmatrix}, \tag{3.2.1}$$

where n is the time parameter and d is the number of panels.[1] The $\varepsilon_{i,k}$'s are centered noise terms, the $m_{i,k}$'s are means (or signals), the ζ_i's are *»common factors«* and the γ_k's are *»factor loadings«*. Beside the scalars $\gamma_k \in \mathbb{R}$, we assume that all random elements are Hilbert space valued and that the underlying Hilbert space H is the same within all panels. (As before, the inner product is denoted by $\langle v, w\rangle_H$ for $v, w \in H$ and the corresponding norm by $\|v\|_H$. The subscripts will be suppressed if the context allows us to do so.) We may also allow for a *»mixed panel«* setting, i.e. the state spaces to be different in different panels. However, this would lead to more notation and distract from our main ideas. We will get back to this more flexible setting briefly in Remark 3.8, below.[2]

The presented theory is motivated by the penalized least squares method of Bleakley & Vert (2010, 2011a) that is designed for a multiple (common) change point model

$$m_{i,k} = \begin{cases} \mu_{1,k}, & i = 1, \ldots, u_1, \\ \mu_{2,k}, & i = u_1 + 1, \ldots, u_2, \\ \ldots, & \ldots, \\ \mu_{p+1,k}, & i = u_p + 1, \ldots, n, \end{cases} \tag{3.2.2}$$

where $1 \le p < n$, $1 \le k \le d$ and $n \ge 3$.[3] The u_r (for $1 \le r \le p$) are *change points* for which level-shifts $\mu_{r,l} \ne \mu_{r+1,l}$ hold true for some panels $1 \le l \le d$ and where l may depend on r. Bleakley & Vert (2010) study their method theoretically in a (restricted) single common change point scenario, i.e. for $p = 1$, and within this setting it turns out to correspond to a traditional CUSUM approach. (Cf. Subsection 3.3.3, below.) Our theory extends Bleakley & Vert (2011a) and is restricted to the single common change point framework as well. Since we are focusing on CUSUM procedures in this thesis,

[1] This is a generalization of Torgovitski (2015d) where panels are univariate. The Hilbert space setting is briefly sketched in Torgovitski (2015d, arXiv:1501.00177v2) but without common factors.

[2] Cf. Section 2.4 of Torgovitski (2015d, arXiv:1501.00177v2).

[3] The penalized least squares method of Bleakley & Vert (2010, 2011a) extends the class of total variation denoising approaches and is reformulated as a group fused LASSO method, i.e. as a regression problem. (Cf. Subsection 3.3.3, below.)

we state our theory with respect to CUSUM-type estimates and justify the relation to the penalized least squares approach of Bleakley & Vert (2011a) afterwards. (Note that the latter approach is restricted to a univariate real-valued panel data setting.)

Notation 3.1 (Signal plus noise model in the matrix notation). Our primary interest is to study estimates $\hat{u}_r$ for change points u_r for fixed n and as $d \to \infty$. Particularly, we will focus on traditional weighted CUSUM estimates that will be introduced in (3.2.7) and (3.2.8), below. Later on in Subsection 3.3.3, we will take a closer look at the total variation denoising approach of Bleakley & Vert (2010, 2011a) where a (re-)formulation of the signal plus noise model (3.2.1) in the matrix notation

$$Y = \mathcal{M} + \mathcal{E} \tag{3.2.3}$$

with entries $\mathcal{M}_{i,k} = m_{i,k}$ and $\mathcal{E}_{i,k} = \varepsilon_{i,k} + \gamma_k \zeta_i$ will be useful. Furthermore, we will use $Y_{\bullet,k} = [Y_{1,k}, \ldots, Y_{n,k}]'$, $\mathcal{E}_{\bullet,k} = [\mathcal{E}_{1,k}, \ldots, \mathcal{E}_{n,k}]'$ and $Y_{i,\bullet} = [Y_{i,1}, \ldots, Y_{i,d}]$, $\mathcal{E}_{i,\bullet} = [\mathcal{E}_{i,1}, \ldots, \mathcal{E}_{i,d}]$ to denote the columns and the rows of the matrices Y and $\mathcal{E}$. We already worked with this matrix notation in Torgovitski (2015d) (cf. also Bleakley & Vert, 2010).

To state our basic conditions on the panels more precisely, we follow Torgovitski (2015d) and introduce the *»cumulated noises«*

$$S_n(x;\mathcal{E}) = \sum_{j=1}^{\lfloor nx \rfloor} (\mathcal{E}_{j,\bullet} - \bar{\mathcal{E}}_{n,\bullet})/n^{1/2} = [S_{n,1}(x;\varepsilon), \ldots, S_{n,d}(x;\varepsilon)]'$$

with $\bar{\mathcal{E}}_{n,\bullet} = \sum_{i=1}^{n} \mathcal{E}_{i,\bullet}/n$. The contribution of the k-th panel is here defined by

$$S_{n,k}(x;\varepsilon) = \sum_{j=1}^{\lfloor nx \rfloor} (\varepsilon_{j,k} - \bar{\varepsilon}_{n,k})/n^{1/2}$$

with $\bar{\varepsilon}_{n,k} = \sum_{i=1}^{n} \varepsilon_{i,k}/n$ for all $k \in \mathbb{N}$. Furthermore, the variable $x \in I_n$ denotes values on the grid I_n, where

$$I_n = \{1/n, 2/n, \ldots, (n-1)/n\} \subset (0,1) \tag{3.2.4}$$

is a discrete time domain. (We will explain why we work with $S_{n,k}(x;\varepsilon)$, $x \in I_n$, rather than with $S_{n,k}(x;\varepsilon)$, $x \in (0,1)$, in Remark 3.11, below.[1])

Assumption N1 (on the noise).

1. The noise terms $\{\varepsilon_{i,k}\}_{i,k\in\mathbb{N}}$ are centered with same second moments $E\|\varepsilon_{i,k}\|^2 = \sigma^2$, $\sigma \in (0,\infty)$, for all $i,k \in \mathbb{N}$, and have uniformly bounded finite fourth moments

$$\sup_{i,k\in\mathbb{N}} E\|\varepsilon_{i,k}\|^4 < \infty.$$

[1] One should be aware that the domain of x always depends on n but that this dependence is suppressed in the notation of x.

2. The second moments of cumulated noises are independent of the panel k and of the time point x, i.e.

$$E\|S_{n,1}(x;\varepsilon)\|^2 = E\|S_{n,k}(x;\varepsilon)\|^2$$

for all $k \in \mathbb{N}$ and all $x \in I_n$. We define

$$V(x) = (E\|S_{n,1}(x;\varepsilon)\|^2)^{1/2}/\sigma \tag{3.2.5}$$

for all $x \in I_n$ and call V^2 the standardized *»variance of cumulated noises«*[1].

Assumption CF (on common factors). The array of noises $\{\varepsilon_{i,k}\}_{i,k\in\mathbb{N}}$ and the sequence of common factors $\{\zeta_i\}_{i\in\mathbb{N}}$ are independent of each other. The latter sequence is mean zero and fulfills $E\|\zeta_i\|^4 < \infty$ for all $i \in \mathbb{N}$. Furthermore, the factor loadings γ_k satisfy, as $d \to \infty$,

$$\sum_{k=1}^{d} \gamma_k^2/d = o(1).$$

The estimation problem 3.2.2

As indicated previously, we work in a single common change point scenario where we assume the means to follow the model

$$m_{i,k} = \begin{cases} \mu_{1,k}, & i = 1,\dots,u, \\ \mu_{2,k}, & i = u+1,\dots,n \end{cases} \tag{3.2.6}$$

in all panels $1 \le k \le d$ and with $n \ge 3$. The time point u is called a *»common change point«* since it is the only possible location for a jump in the mean $\mu_{1,k} \neq \mu_{2,k}$. Note that for real-valued and univariate panels this corresponds to the multivariate abrupt change alternative introduced in (1.1.2).

Notation 3.2 (Change point location). Instead of working with $u \in \{1,\dots,n-1\}$, we will often use the rescaled time domain (3.2.4) and thus work with the *rescaled common change point* that will be denoted by $s = u/n \in I_n$. (Notice also the symmetric version $\varsigma(s) = \max\{s, 1-s\}$ that we will use, e.g., in Proposition 3.20, below.)

As an estimate $\hat{u}$ for the change point u, we consider a Hilbert space version of a classical weighted CUSUM estimate which is defined to be any element

$$\hat{u} \in \operatorname*{arg\,max}_{1\le i<n} w(i/n)\mathscr{T}(i/n) \tag{3.2.7}$$

and where the detector is defined as

$$\mathscr{T}(x) = \mathscr{T}(x;Y) = \|S_n(x;Y)\|_{2,H} = \left[\sum_{k=1}^{d} \left\|S_{n,k}(x;Y)\right\|^2\right]^{1/2}, \tag{3.2.8}$$

[1] This terminology is motivated by the univariate real-valued setting where $V(x) = \mathrm{Var}(S_{n,1}(x;\varepsilon))^{1/2}/\sigma$.

using the product norm $\|x\|_{2,H} = [\sum_{k=1}^{d} \|x_k\|_H^2]^{1/2}$ for $x = (x_1, \ldots, x_d)' \in H^d$. The general weighting function $w(x) > 0$ is defined formally only on a discrete domain, i.e. only for all $x \in I_n$. Notice that our detector (3.2.8) is an analogue of the detectors in (1.1.4) and in (2.3.3) but now for Hilbert space valued panel data. (Furthermore, notice that estimates (3.2.7) and (1.1.5) coincide if we neglect the common factors and if the $\mathbb{R}^d$-valued error sequence $\{\mathcal{E}_{\bullet,i}\}_{i=1,\ldots,n}$ is i.i.d. with $\mathrm{Cov}(\mathcal{E}_{\bullet,i}) = \mathrm{diag}(1, \ldots, 1)$, i.e. with $\sigma = 1$ in Assumption N1.)

Remark 3.3 (Definition of argmax). The definition of $\arg\max$ as a set is a minor but necessary technicality (cf. also (1.1.5)): if we define it as usual to be the smallest argument for which the maximum is attained, then (3.3.5), below, is generally not valid anymore. Note that our results for $\hat{u}$ will not depend on the actual choice in (3.2.7).

It is evident that not all changes can be detected with (3.2.7). For instance, a change in only one panel in (3.2.6) will be not detectable as its contribution fades out asymptotically for $d \to \infty$. For this reason we define the *»total average change«* as

$$\Delta = \lim_{d\to\infty} \bar{\Delta}_d \in (0, \infty), \qquad \bar{\Delta}_d = \sum_{k=1}^{d} \|\Delta_k\|^2 / d \tag{3.2.9}$$

with $\Delta_k = \mu_{2,k} - \mu_{1,k}$ (cf. (3.2.6)). On the one hand, this definition quantifies the magnitude of common changes via $\bar{\Delta}_d$ and, on the other hand, imposes an asymptotic condition on the magnitude and on the size of changes $\Delta \in (0, \infty)$ that we will work with in the following.

Remark 3.4 (Conditions on the change for different asymptotics). In the fixed d and $n \to \infty$ setting, it is usually assumed that the change location $s = s_n$ (or equivalently $u = u_n$) is asymptotically proportional to the sample size n (cf. Remark 2.9), i.e.

$$\lim_{n\to\infty} s_n = \lim_{n\to\infty} u_n/n \in (0, 1),$$

and that the magnitude of the change may be arbitrary, as long it is non-zero. Opposed to this, in our fixed n and $d \to \infty$ setting, we require $\Delta \in (0, \infty)$, i.e. that the squared change magnitudes have to be proportional to the number of panels d (cf. (3.2.9)) but the change point position $s \in I_n$ may be arbitrary instead.

3.3 Estimating change points with weighted CUSUM procedures

In the following we will study the influence of different weighting functions w, several distributions of the panels and, in particular, various parameters Δ, σ, u and n on the asymptotic accuracy (or consistency) of the change point estimate $\hat{u}$, i.e. we investigate whether $P(\hat{u} = u) \to 1$ holds true as $d \to \infty$ or not. The results depend on the behavior of the following critical function which is related to Bleakley & Vert (2011a, disp. (11)).

Definition 3.5 (of the critical function). We define a *»critical function«* as

$$C(x; y, \rho) = w^2(x)\big[V^2(x)\rho + H^2(x, y)\big] \tag{3.3.1}$$

on the discrete domain $x, y \in I_n$ with $V(x)$ as in (3.2.5), with the *»noise to change«* ratio

$$\rho = \rho(\Delta, \sigma, n) = \frac{1}{n}\frac{\sigma^2}{\Delta}$$

and with $H(x, y) = \min\{x, y\}[1 - \max\{x, y\}]$.

Note that in Definition 3.5 we implicitly restrict ourselves to $\rho \in (0, \infty)$ since we already assumed $\sigma \in (0, \infty)$ in Assumption N1 and $\Delta \in (0, \infty)$ in (3.2.9) before. The next theorem extends Lemma 1 and Theorem 3 of Bleakley & Vert (2011a) as well as Theorem 2.9 of Torgovitski (2015d) to a dependent Hilbert space setting (with common factors) and is our main tool for the analysis of accuracy in the following sections. Particularly, conditions (3.3.3) and (3.3.4) are different compared to (2.13) from Torgovitski (2015d) which is a consequence of the Hilbert space setting.

Theorem 3.6. *Let Assumption N1 and Assumption CF be fulfilled. Assume that it holds that*

$$\lim_{d\to\infty} \sum_{k,r=1}^{d} \big|\operatorname{Cov}\big(\langle\varepsilon_{j,k}, \varepsilon_{h,k}\rangle, \langle\varepsilon_{l,r}, \varepsilon_{p,r}\rangle\big)\big|/d^2 = 0, \tag{3.3.2}$$

$$\lim_{d\to\infty} \sum_{k,r=1}^{d} \big|E\big(\langle\Delta_k, \varepsilon_{j,k}\rangle\langle\Delta_r, \varepsilon_{l,r}\rangle\big)\big|/d^2 = 0, \tag{3.3.3}$$

$$\lim_{d\to\infty} \sum_{k,r=1}^{d} |\gamma_k\gamma_r|\big|E\big(\langle\varepsilon_{j,k}, \varepsilon_{l,r}\rangle\big)\big|/d^2 = 0 \tag{3.3.4}$$

for all $1 \le j,\ h,\ l,\ p \le n$. *Then, for any change point* $s \in I_n$ *and any ratio* $\rho \in (0, \infty)$ *it holds that*

$$\lim_{d\to\infty} P\Big(\hat{u} \in \operatorname*{arg\,max}_{x\in I_n} C(x; s, \rho)\Big) = 1. \tag{3.3.5}$$

Remark 3.7 (on conditions of Theorem 3.6). Assume that the panels $\{\varepsilon_{i,k};\ i \in \mathbb{N}\}_{k\in\mathbb{N}}$ are independent. A repeated application of the Cauchy-Schwarz inequality shows that Assumption N1 and Assumption CF together with $\Delta \in (0,\infty)$ imply that conditions (3.3.2)-(3.3.4) are fulfilled. Evidently, some minor deviations beyond this independence assumption are possible.

Theorem 3.6 states, under rather mild assumptions, that consistent estimation is determined by the shape of the critical function C, i.e. by the interplay between the functions V, H and the weights w. We proceed with a detailed study of C and begin by analyzing V with respect to different underlying distributions.

Remark 3.8 (Mixed panels). Notice that if we assume our model to be without common factors, then, as already mentioned in the introduction, all our results can be extended to panels in different Hilbert spaces. Towards this end we would have to adapt the norms in (3.2.8) appropriately for each k. Having this »generality« in mind we prefer to keep the notation simple and the presentation clear and thus we continue to restrict ourselves to some common Hilbert space within all panels.

Preliminaries on the variance of cumulated noises 3.3.1

In this subsection we analyze the function V, defined in Assumption N1, more detailed. First, we derive different representations for V and establish conditions for positivity and symmetry which will be important for our investigations in the following sections. Afterwards, we present further results on the shape of V for different distributions of $\{\varepsilon_{i,k}\}$ covering the three basic models: panels of uncorrelated, moving average and autoregressive noises. Each of these examples will play a major role in our subsequent analysis. Before we study the shape of V we introduce some notation and define the matrix

$$\Sigma = \begin{bmatrix} \gamma(1,1) & \cdots & \gamma(1,n) \\ \vdots & \ddots & \vdots \\ \gamma(n,1) & \cdots & \gamma(n,n) \end{bmatrix} \tag{3.3.6}$$

with $\gamma(k,r) = E\langle \varepsilon_{k,1}, \varepsilon_{r,1}\rangle$. Furthermore, we define the function

$$\mathcal{F}(x) = x(1-x)\big[\Gamma_x + \Gamma_{1-x} - \Gamma_1\big]/\sigma^2 \tag{3.3.7}$$

with

$$\Gamma_x = \Gamma_x^{(n)} = \sum_{k,r=1}^{\lfloor nx \rfloor} \gamma(k,r)/\lfloor nx \rfloor, \tag{3.3.8}$$

where $x \in I_n$. Finally, we need to introduce the quadratic forms

$$f_{i,n}(\Sigma) = a^{(i)}\Sigma a^{(i)\prime}/\sigma^2, \tag{3.3.9}$$

Model	Variance of cumulated noises	Correction term
WN	$V^2(x) = x(1-x)$	
MA(1)	$c_1 V^2(x) = x(1-x)\alpha(\phi) - \mathcal{R}(\phi)$	$\alpha(\phi) = (n/2)[1+\phi^2] + \phi + \phi n$
AR(1)	$c_2 V^2(x) = x(1-x)\alpha(\phi) - \mathcal{R}(\phi, x)$	$\alpha(\phi) = (n/2)[1-\phi^2] + \phi - \phi^{n+1}$

Table 3.1: Representation of the variance of cumulated noises V^2 for Examples 3.12, 3.13 and 3.14, i.e. under different dependence assumptions. The terms for $\mathcal{R}(\phi)$ and $\mathcal{R}(\phi, x)$ are provided within the respective examples. Furthermore, c_1 and c_2 are some positive constants that are independent of x. (WN is the usual abbreviation for a white noise model.)

which are based on vectors $a^{(i)}$ given by

$$a^{(i)} = (a_1^{(i)}, \dots, a_i^{(i)} \,|\, a_{i+1}^{(i)}, \dots, a_n^{(i)}) = (n-i, \dots, n-i \,|\, -i, \dots, -i)/n^{3/2}. \quad (3.3.10)$$

We suppress the dependence of $a^{(i)}$'s on n which shall not lead to any confusion.

Remark 3.9. Note that within the real-valued univariate case the matrix Σ is the usual covariance matrix of $(\varepsilon_{1,1}, \dots, \varepsilon_{n,1})'$ and that Γ_x corresponds to the variance of $\sum_{r=1}^{\lfloor nx \rfloor} \varepsilon_{r,1}/\lfloor nx \rfloor^{1/2}$.

The next lemma states the connection between the functions $V, \mathcal{F}$ and $f_{i,n}$. We will call V *»symmetric«* whenever $V(x) = V(1-x)$ holds true for all $x \in I_n$ and *»positive«* whenever $V(x) > 0$ holds true for all $x \in I_n$. (Note that V is always non-negative.)

Lemma 3.10. *Under Assumption N1 it holds that*

$$V^2(i/n) = f_{i,n}(\Sigma) \quad (3.3.11)$$

for $1 \le i < n$ and where $f_{i,n}(\Sigma)$ is defined in (3.3.9) by using (3.3.6). Hence, a positive definite matrix Σ implies a positive function V. Furthermore, if the time series $\{\varepsilon_{r,1}\}_{r=1,\dots,n}$ is additionally weakly stationary, then we obtain

$$V^2(i/n) = \mathcal{F}(i/n) \quad (3.3.12)$$

for $1 \le i < n$ and where $\mathcal{F}$ is defined in (3.3.7). Note that V is symmetric in this case.

Remark 3.11 (The necessity of a discrete domain). If $V(x) = (E\|S_{n,1}(x;\varepsilon)\|^2)^{1/2}/\sigma$ is symmetric on the discrete time domain, i.e. for $x \in I_n$, it does not imply $V(x) = (E\|S_{n,1}(x;\varepsilon)\|^2)^{1/2}/\sigma$ to be symmetric on the time continuous domain, i.e. for all $x \in [0,1]$. This subtle difference is one reason for the restriction of our considerations to I_n.

We proceed with a few examples that illustrate the function V for different dependence structures by making use of Lemma 3.10 and in particular of the expression (3.3.7). We will specify V only up to some scalar factor $c > 0$ that is negligible since our results will be invariant under such rescaling. (Note that the verification of all expressions of V given in these examples is postponed to Section 3.5 and that some expressions are additionally summarized in Table 3.1.)

Example 3.12 (for white noise). *Let* $\{\varepsilon_{i,k}\}$ *be an array of Hilbert space valued random variables with* $\gamma(l,r)=0$ *for* $l \neq r$ *and such that part 1 of Assumption N1 holds true. We observe that*

$$V^2(x) = x(1-x) \tag{3.3.13}$$

for all $x \in I_n$ *since* $\Gamma_x \equiv \Gamma_{1-x} \equiv \Gamma_1 = \sigma^2$ *holds true for any* $x \in I_n$ *(cf. (3.3.12)). The function* V *is obviously symmetric and positive on the whole domain* I_n.

For the next three (more general) Examples 3.13, 3.14 and 3.18 we assume $\{\eta_{i,k}\}$ to consist of i.i.d. centered Hilbert space valued random variables with $E\|\eta_{1,1}\|^4 < \infty$ and $E\|\eta_{1,1}\|^2 = \tilde{\sigma}^2$ for $\tilde{\sigma}^2 \in (0,\infty)$.

Example 3.13 (for moving average noise I). *Let* $\{\varepsilon_{i,k}\}_{i\in\mathbb{Z}}$*,* $k \in \mathbb{N}$*, be identically distributed MA(1) time series, defined via*

$$\varepsilon_{i,k} = \eta_{i,k} + \phi\eta_{i-1,k},$$

for some $\phi \in \mathbb{R}$. *It holds that*

$$cV^2(x) = x(1-x)\alpha(\phi) - \mathcal{R}(\phi), \tag{3.3.14}$$

for all $x \in I_n$*, with some* $c > 0$ *(that depends on* n*), with*

$$\alpha(\phi) = (n/2)[1+\phi^2] + [\phi + \phi n]$$

and with a remainder term $\mathcal{R}(\phi) = \phi$*. Based on Lemma 3.10 we conclude that* V *is positive which follows due to the positive definiteness of* Σ *(cf. (3.3.6)). The latter is easily verified, since the matrix is irreducibly diagonally dominant and thus invertible for any parameter* ϕ *(cf., e.g., Theorem 1.21 of Varga, 2000). Notice that the term* $\alpha(\phi)$ *may be zero (for some negative* ϕ*) or even become negative for some* ϕ *close to* -1*. Hence,* V *may be a convex, a concave or even a constant function.*

Example 3.14 (for autoregressive noise). *Let* $\{\varepsilon_{i,k}\}_{i\in\mathbb{Z}}$*,* $k \in \mathbb{N}$*, be identically distributed AR(1) time series, defined via*

$$\varepsilon_{i,k} = \sum_{j=0}^{\infty} \phi^j \eta_{i-j,k}$$

for some $\phi \in (-1,1)$*. It holds that*

$$cV^2(x) = x(1-x)\alpha(\phi) - \mathcal{R}(\phi, x), \tag{3.3.15}$$

for all $x \in I_n$*, with some* $c > 0$ *(that depends on* ϕ *and* n*), with*

$$\alpha(\phi) = (n/2)[1-\phi^2] + [\phi - \phi^{n+1}]$$

and with a remainder term

$$\mathcal{R}(\phi, x) = \phi - (1-x)\phi^{nx+1} - x\phi^{n(1-x)+1}.$$

Again, using Lemma 3.10, we conclude that V *is positive. This follows by the matrix* Σ *in (3.3.6) being positive definite for any* $\phi \in (-1,1)$ *(cf., e.g., disp. (1.1) of Sutradhar & Kumar, 2003). We will observe later on that* V *is generally oscillating and thus neither convex nor concave.*

In Subsection 3.3.2 it will turn out to be important whether $V(x)/x$ is (strictly) decreasing in the domain I_n or not. The following lemma slightly simplifies this verification and will help us with Example 3.13 in Lemma 3.16 below.

Lemma 3.15. *Let V be positive, strictly concave and fulfill $V(z)/z > V(y)/y$ for $z = 1/n$ and all $y \in I_n$ with $y > z$. Then $V(x)/x$ is strictly decreasing for $x \in I_n$.*

The next lemma shows that $V(x)/x$ remains strictly decreasing for larger deviations from (3.3.13).

Lemma 3.16. *In the MA(1) situation of Example 3.13 it holds that $V(x)/x$ is strictly decreasing for $x \in I_n$ and any parameter $\phi \in \mathbb{R}$.*

$V(x)/x$ depends continuously on Σ and is strictly decreasing in the i.i.d. case (cf. (3.3.13)). Hence, by continuity, it must necessarily be strictly decreasing for a variety of minor deviations from this scenario (beyond moving averages). In the AR(1) situation of Example 3.14 this is, e.g., the case for any $\phi \in (-\delta_n, \delta_n)$ for some appropriate $\delta_n \in (0,1)$. However, the same AR(1) model provides interesting counterexamples where $V(x)/x$ does not decrease strictly monotonically.

Due to the finite panel length n it is difficult to derive sharp conditions on the violation of our monotonicity assumption. Hence, we restrict ourselves to some specific asymptotic considerations of $\phi \to -1$ or of a sufficiently large n.

Counterexample 3.17 (for autoregressive noise). *Let $\{\varepsilon_{i,k}\}_{i\in\mathbb{Z}}$, $k \in \mathbb{N}$, be identically distributed AR(1) time series as in Example 3.14.*

1. *The first counterexample is given by observing that*

$$\lim_{n\to\infty} \left[\frac{V((n-2)/n)}{(n-2)/n}\right]^2 - \left[\frac{V((n-1)/n)}{(n-1)/n}\right]^2 = c\,(1 - 3\phi^2 + 2\phi^3)$$

   holds true for some $c > 0$. The right-hand side is negative for any $\phi \in (-1, -1/2)$ which implies that $V(x)/x$ is non decreasing for $x \in \{1-2/n, 1-1/n\}$ and such ϕ's, given a sufficiently large sample size n.

2. *The second counterexample is given by assuming that n is even and observing that*

$$\lim_{\phi\to -1} cV^2(x) = 1 + (-1)^{nx+1} \tag{3.3.16}$$

   holds true for all $x \in I_n$ and with some $c > 0$. The right-hand side is oscillating between 0 and 1 and, as a consequence, $V(x)/x$ cannot be monotonically decreasing on I_n for ϕ sufficiently close to -1 (and for n being even). Notice that the limit in (3.3.16) would be different for n being odd but would still provide a counterexample (cf., Figure 3.2, below).

For the sake of generality, we close this subsection by mentioning a concise expression for V in case of general MA(q) time series (which extends (3.3.14)).

Example 3.18 (for moving average noise II). *Let* $\{\varepsilon_{i,k}\}_{i\in\mathbb{Z}}$, $k \in \mathbb{N}$, *be identically distributed MA(q) time series with* $1 < q < n/2$ *defined via*

$$\varepsilon_{i,k} = \sum_{j=0}^{\infty} \phi_j \eta_{i-j,k}$$

for some $\phi_j \in \mathbb{R}$ *with* $\phi_0 = 1$ *and* $\phi_j = 0$ *for* $j > q$.[1] *For discrete valued*

$$x \in (q/n, 1-q/n) \cap I_n \tag{3.3.17}$$

it holds that

$$cV^2(x) = x(1-x)\alpha(\phi_1, \phi_2, \ldots) - \mathcal{R}(\phi_1, \phi_2, \ldots) \tag{3.3.18}$$

for some $c > 0$ *(that depends on* n*) and with*

$$\alpha(\phi_1, \phi_2, \ldots) = (n/2) \sum_{j=0}^{q} \phi_j^2 + n \sum_{k=1}^{q} \sum_{j=0}^{q} \phi_j \phi_{j+k} + \mathcal{R}(\phi_1, \phi_2, \ldots),$$

$$\mathcal{R}(\phi_1, \phi_2, \ldots) = \sum_{k=1}^{q} \sum_{j=0}^{q} k(\phi_j \phi_{j+k}). \tag{3.3.19}$$

The compact expressions (3.3.18) - (3.3.19) *do not hold true on the whole domain* I_n. *(However, the general expression in* (3.3.12) *with* (3.3.7) *is still valid and holds true for arbitrary* $q \in \mathbb{N}$.*)*

Limit theorems for weighted CUSUM estimates 3.3.2

Pointwise accurate estimation

The aim of this section is to study the pointwise accuracy (or consistency) of $\hat{u}$. Given that (3.3.5) holds true, the estimate $\hat{u}$ of a change point u with a change to noise ratio $\rho > 0$ is accurate, i.e. $\lim_{d\to\infty} P(\hat{u} = u) = 1$, if the following Assumption U1 holds true and is not accurate if Assumption U2 holds true instead. The former is the analogue of Assumption A1 of Torgovitski (2015d).

Assumption U1 (on the interplay between the change point and C**).** Let the critical function C be defined as in (3.3.1). For a change at $s = u/n$ it holds that $C(s; s, \rho) > C(x; s, \rho)$ for all $x \neq s$, $x \in I_n$.

Assumption U2 (as a counterpart to Assumption U1). Let the critical function C be defined as in (3.3.1). For a change at $s = u/n$ there is some $x \in I_n$ such that $C(s; s, \rho) < C(x; s, \rho)$ holds true.

[1] Recall the definition of $\{\eta_{i,k}\}$ on p. 104.

Neither of the Assumptions U1 or U2 is fulfilled if $C(x; s, \rho)$ has a non-unique maximum as a function of $x \in I_n$. This situation is not of interest in this thesis since in that case asymptotic accuracy may neither be proven nor disproven based on (3.3.5) without additional information.

Let us define

$$R(x, s) := [G(x, s) - G(s, s)]/[F(s) - F(x)] \tag{3.3.20}$$

with $F(x) = [w(x)V(x)]^2$, $G(x, s) = [w(x)H(x, s)]^2$ and set

$$\rho_{\min}(s) := \sup_{\substack{x \in I_n, \\ F(x) < F(s)}} R(x, s), \qquad \rho_{\max}(s) := \inf_{\substack{x \in I_n, \\ F(x) > F(s)}} R(x, s) \tag{3.3.21}$$

with $\sup \emptyset = -\infty$ and $\inf \emptyset = \infty$, $x, s \in I_n$. (Cf. Definition 3.5.) The next two propositions are straightforward to verify and are generalized versions of Theorem 2.16 of Torgovitski (2015d).

Proposition 3.19. *Assume that for all $x \neq s$, $x \in I_n$, either $F(x) \neq F(s)$ or $F(x) = F(s)$ with $G(x, s) < G(s, s)$ holds true, where F and G are defined in (3.3.20). For a change at $s = u/n$ it holds that*

1. *Assumption U1 is fulfilled for all $\rho \in (0, \infty) \cap \big(\rho_{\min}(s), \rho_{\max}(s)\big)$,*
2. *Assumption U2 is fulfilled for all $\rho \in (0, \infty) \setminus \big(\rho_{\min}(s), \rho_{\max}(s)\big)$,*

where we set $(\rho_{\min}, \rho_{\max}) = \emptyset$ for $\rho_{\min} \geq \rho_{\max}$ (cf. (3.3.21)).

Note that if $F(s) = F(x)$ and $G(x, s) < G(s, s)$ hold true for all $x \neq s$, $x \in I_n$, then we have $(\rho_{\min}(s), \rho_{\max}(s)) = \mathbb{R}$ and thus Assumption U1 is fulfilled for all $\rho \in (0, \infty) \cap \mathbb{R} = (0, \infty)$. (On the other hand, Assumption U2 is fulfilled for all $\rho \in (0, \infty)$, if $F(s) = F(x)$ and $G(x, s) > G(s, s)$ hold true for some $x \neq s$, $x \in I_n$.[1])

Proposition 3.20. *Assume that the variance of cumulated noises V^2, defined in (3.2.5), and the weighting w are symmetric and that F, defined in (3.3.20), is strictly concave. Furthermore, assume that $w(x)x$ is strictly increasing and $w(x)(1-x)$ is strictly decreasing for $x \in I_n$. Then it holds that*

$$\rho_{\min}(s) < 0 < \rho_{\max}(s) = \inf_{\substack{1/2 \leq x < \varsigma, \\ x \in I_n}} R(x, \varsigma),$$

where $\varsigma(s) = \max\{s, 1-s\}$. Furthermore, we have $\{1/2 \leq x < \varsigma,\ x \in I_n\} = \emptyset$ if and only if $\varsigma = \lceil n/2 \rceil / n$.

[1] Notice that this case is not included in Proposition 3.19.

Pointwise accurate estimation with classical weighting schemes

It is straightforward to verify that all assumptions of Proposition 3.20 are satisfied for the variance of cumulated noises $V^2(x) = x(1-x)$ (cf. Example 3.12) together with the weights w given by the well-known class of weighting functions

$$w_\gamma(x) = \left[x(1-x)\right]^{-\gamma},$$

$x \in I_n$, where $\gamma \in (0, 1/2)$. We introduced this class already in (1.2.1). Notice that the excluded limiting cases of weightings $w_{1/2}$ and w_0 were studied in related settings by Bleakley & Vert (2010, 2011a) and that $w_{1/2}$ was studied independently by Bai (2010). Altogether, Bleakley & Vert (2011a, Theorem 3) and Bai (2010, Theorem 3.1) showed, on the one hand, that $\rho_{\max}(s) = \infty$ holds true for the weighting $w_{1/2}$ for any change point s and, on the other hand, Bleakley & Vert (2011a, Theorem 2) derived a closed-form expression of $\rho_{\max}(s)$ for the weighting w_0.

The next theorem of Torgovitski (2015d) is restated in a more convenient and insightful manner. It extends the latter results of Bleakley & Vert (2011a) and Bai (2010) and provides a closed-form expression for $\rho_{\max}(s)$ for the weightings w_γ within the whole range $\gamma \in (0, 1/2)$.

Theorem 3.21. *Let $V^2(x) = x(1-x)$ and assume that we use w_γ with $\gamma \in (0, 1/2)$. It holds that*

$$\rho_{\max}(s) = \inf_{\substack{1/2 \le x < \varsigma, \\ x \in I_n}} R(x, \varsigma) \begin{cases} = R(\varsigma - 1/n, \varsigma), & 1/2 + 1/n < \varsigma \le \mathscr{B}(\gamma), \\ < R(\varsigma - 1/n, \varsigma), & \mathscr{B}(\gamma) + 2/n < \varsigma < 1, \end{cases}$$

where $R(x, y)$ is defined in (3.3.20), $\rho_{\max}$ is defined in (3.3.21), $\varsigma(s) = \max\{s, 1-s\}$ and where the bound $\mathscr{B}(\gamma)$ is defined as

$$\mathscr{B}(\gamma) = \frac{(4\gamma^2 + 6\gamma^{3/2} - 3\gamma^{1/2} - 1)}{(8\gamma^2 + 8\gamma^{3/2} - 4\gamma^{1/2} - 1) - 2\gamma} \tag{3.3.22}$$

for $\gamma \in (0, 1/2)$. $\mathscr{B}$ is continuous and $\mathscr{B}(\gamma) \downarrow \mathscr{B}(1/2) = 2^{-1/2}$ as $\gamma \uparrow 1/2$.

To get a feeling for the magnitude of $\rho_{\max}(s)$ it is proposed by Torgovitski (2015d) to evaluate $R(\varsigma - 1/n, \varsigma)$ for $n \to \infty$ and a change point that is proportional to the sample size, i.e. that satisfies $\varsigma = \lfloor n\theta \rfloor / n$ with some $\theta \in (1/2, 1)$. We observe that

$$\lim_{n\to\infty} R(\varsigma - 1/n, \varsigma) = 2\theta[\theta - f(\gamma)]f(\theta) \tag{3.3.23}$$

with $f(x) = (1-x)/(1-2x)$ which is according to Theorem 3.21 an asymptotic expression for $\rho_{\max}(s)$ if $\theta \in (1/2, \mathscr{B}(\gamma)]$ yet generally not the correct asymptotic expression for $\rho_{\max}(s)$ if $\theta \in (\mathscr{B}(\gamma), 1)$ (cf. Proposition 3.22 and Figure 3.1, below). In the following proposition, that corresponds to Torgovitski (2015d, Proposition 2.18), we show an explicit representation for $\theta \in (\mathscr{B}(\gamma), 1/2)$ for $\gamma = 1/4$ based on a different approach that does not rely on the evaluation of $R(\varsigma - 1/n, \varsigma)$. (It is, however, not clear how to extend this approach to other values of γ.)

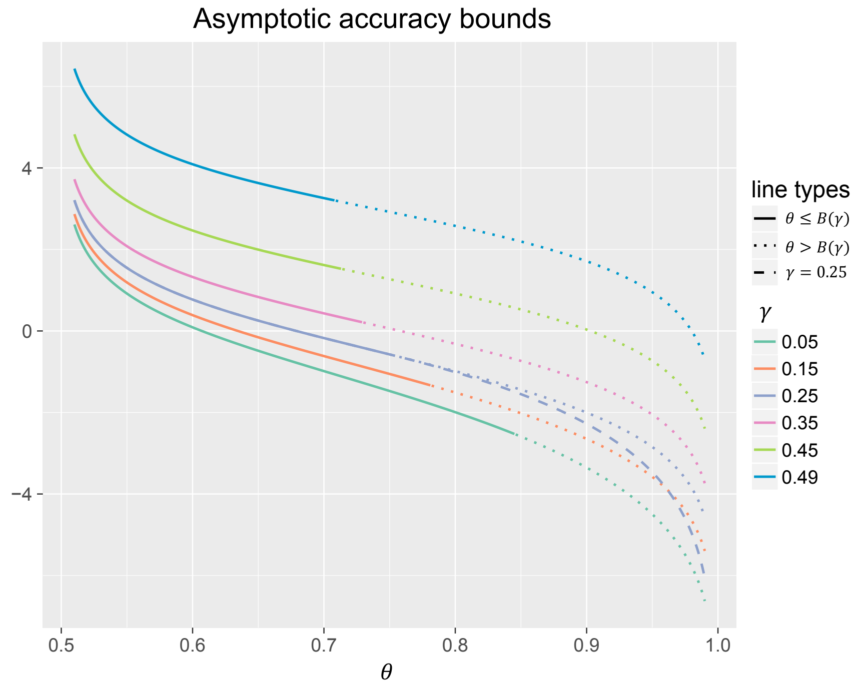

Figure 3.1: This figure shows $\log(2\theta[\theta - f(\gamma)]f(\theta))$, i.e. the logarithm of the expression (3.3.23). We use solid lines for $\theta \leq \mathscr{B}(\gamma)$ and dotted lines for $\theta > \mathscr{B}(\gamma)$. The solid part corresponds to parameters for which equality holds true in (3.3.22) in the limit, i.e. as $n \to \infty$. For the special case of $\gamma = 0.25$ we use a dashed line to show the logarithm of the expression (3.3.24) for all parameters $\theta \in (1/2, 1)$. We see that (3.3.22) is generally not a valid expression for $\theta > \mathscr{B}(\gamma)$ since the dashed and the dotted lines for $\gamma = 0.25$ are different.

Proposition 3.22. *Let $V^2(x) = x(1-x)$, $x \in I_n$, and assume that we use w_γ with $\gamma = 1/4$. Let $\rho_{\max}$ be defined as in (3.3.21), $\mathscr{B}(\gamma)$ as in (3.3.22) and set $\varsigma(s) = \max\{s, 1-s\}$. Further, assume that $\varsigma = \lfloor n\theta \rfloor / n$ with $\theta \in (1/2, 1)$. Then it holds that*

$$\lim_{n\to\infty} \rho_{\max}(s) = (1-\theta)^2 \begin{cases} \frac{(3-2\theta)\theta}{(1-\theta)(2\theta-1)}, & \theta \in (1/2, \mathscr{B}(1/4)], \\ \frac{2\theta(\theta+1)+\theta(1-\theta)^{1/2}-2(1-\theta)^{1/2}-2}{2\theta(\theta-1)+\theta(1-\theta)^{1/2}}, & \theta \in (\mathscr{B}(1/4), 1). \end{cases} \tag{3.3.24}$$

(The right-hand side is a continuously differentiable function for all $\theta \in (1/2, 1)$.)

Remark 3.23 (Spurious estimation under H_0). Assume that there is no change in all panels, i.e. that we have $\mu_{1,k} = \mu_{2,k}$ for all $k \in \mathbb{N}$ in (3.2.6), and that $V^2(x) = x(1-x)$, $x \in I_n$, holds true. It is pointed out in Torgovitski (2015d, Remark 2.19) that our change point estimate $\hat{u}$, with weights w_γ and $\gamma \in [0, 1/2)$, lies asymptotically in the set $S' = \{\lfloor n/2 \rfloor, \lceil n/2 \rceil\}$, i.e. $P(\hat{u} \in S') \to 1$ as $d \to \infty$. In Torgovitski (2015b) we indicated how such a spurious estimation may be used as a building block for accurate estimation of multiple common changes via a scanning-type approach yet these considerations are not the focus of this thesis and will be studied further elsewhere.

Uniformly accurate estimation (Estimation with covariance-corrected weighting schemes)

A striking feature of the weighting $w_{1/2}$, in case of panels with white noise, is that pointwise asymptotic accuracy, i.e.

$$\lim_{d\to\infty} P(\hat{u} = u) = 1, \tag{3.3.25}$$

holds true uniformly in the sense that (3.3.25) is independent of the change position u and independent of the change to noise ratio $\rho \in (0, \infty)$. We call this property *»uniform asymptotic accuracy«* and in this section we are concerned with conditions on the dependence structure of panel data under which this property holds true for estimates $\hat{u}$. We will see that in a wide dependence framework uniform asymptotic accuracy requires a weighting function $w(x) = c\, w_\star(x)$ with

$$w_\star(x) = 1/V(x) = w_{1/2}(x)/h(x), \tag{3.3.26}$$

where $c > 0$ is an arbitrary constant, where $x \in I_n$ and where $h(x)$ is specified in (3.3.27), below. (In other words the inference on change-locations with $w_{1/2}$ is misleading without an appropriate correction term.) The second equality in (3.3.26) holds true due to (3.3.12) since $V(x)$ may be represented via the standard Darling-Erdős-type weighting scheme $w_{1/2}$ corrected by a covariance-based term

$$h(x) = [\Gamma_x + \Gamma_{1-x} - \Gamma_1]^{1/2}/\sigma. \tag{3.3.27}$$

Furthermore, note that the latter term collapses to $h(x) = 1$, $x \in I_n$, in the white noise case and thus we obtain $w_\star(x) = w_{1/2}(x)$ in (3.3.26) for this specific situation.

To study uniform asymptotic accuracy, we need to check whether Assumption U1 holds true for arbitrary $s \in I_n$ and all $\rho \in (0, \infty)$ or whether Assumption U2 is satisfied instead.

Theorem 3.24. *Let the matrix Σ from* (3.3.6) *be positive definite, i.e. (according to Lemma 3.10) let the function V from* (3.2.5) *be positive. If Assumption U1 holds true for all change points $s = u/n \in I_n$ (i.e. for all $u = 1, \ldots, n-1$) and all change to noise ratios $\rho \in (0, \infty)$, then $w = cw_\star$, where c is an arbitrary, positive constant. If $w \neq cw_\star$ for all $c > 0$, then there is some change point $s \in I_n$ and some ratio $\rho \in (0, \infty)$ such that Assumption U2 holds true.*

To state a sufficient condition we need a stronger assumption on the noise sequence and on the function V.

Theorem 3.25. *Let the matrix Σ from* (3.3.6) *be positive definite and the noise sequence $\{\varepsilon_{i,1}\}$ be weakly stationary, i.e. (according to Lemma 3.10) let the function V from* (3.2.5) *be positive and symmetric. Assumption U1 is fulfilled for $w = cw_\star$ with any constant $c > 0$ for all change points $s \in I_n$ and all ratios $\rho \in (0, \infty)$ if and only if $V(x)/x$ is strictly decreasing for $x \in I_n$.*

Peštová & Pešta (2015) study a related (but different) specific change point estimate for real-valued univariate panels. In a second updated version Peštová & Pešta (2016) consider a general class of weighted estimates for real-valued univariate panels as in this thesis and as in Torgovitski (2015d). They have a somewhat different focus on the consistency problem and thus obtain different but closely related monotonicity conditions on the variance of cumulated noises in their framework, too.

Remark 3.26 (on admissible oscillations of V). A strictly decreasing $V(x)/x, x \in I_n$, restricts the possible oscillations of V in Theorem 3.25 and thus may be interpreted as a smoothness condition. Moreover, due to the symmetry of V, Theorem 3.25 implies the following additional restrictions on the function V:

$$\alpha_1(x)V(x) < V(x + 1/n) < \alpha_2(x)V(x)$$

with $\alpha_1(x) = 1 - [n(1-x)]^{-1} < 1$, $\alpha_2(x) = 1 + (nx)^{-1} > 1$ for all $x \in I_n$, $x < 1 - 1/n$.

Panels based on MA(1) noise, specified in Example 3.13, and on AR(1) noise, specified in Example 3.14, fulfill Assumption N1. As discussed in Lemma 3.16, the respective functions $V(x)/x$, $x \in I_n$, satisfy the monotonicity criterion of Theorem 3.25 for all parameters ϕ in the MA(1) case and for some parameters ϕ in the AR(1) case. Thus, e.g., for MA(1) panels that satisfy Assumption CF on the common factors, relation (3.3.5) holds true and we have uniform asymptotic accuracy for arbitrary parameters ϕ if we use $w = cw_\star$ with $c > 0$. On the contrary, as follows from Counterexample 3.17, AR(1) panels that satisfy Assumption CF immediately yield counterexamples to uniform asymptotic accuracy for parameters ϕ close to -1 or for large sample sizes n. (Cf. Figure 3.2.)

Remark 3.27. As a final remark, we would like to point out that the difficulties in the evaluation of $V(x)$ via Lemma 3.10, using Γ_x, stem from the facts that n is finite and that $x \in I_n$ depends on n. (Note that for weakly stationary time series $\{\varepsilon_{r,1}\}_{r\in\mathbb{Z}}$ with $\sum_{r=0}^{\infty} |\gamma(1, 1+r)| < \infty$ the limit $\lim_{n\to\infty} \Gamma_{\lfloor nx\rfloor/n}$ does not depend on $x \in [0, 1]$.)

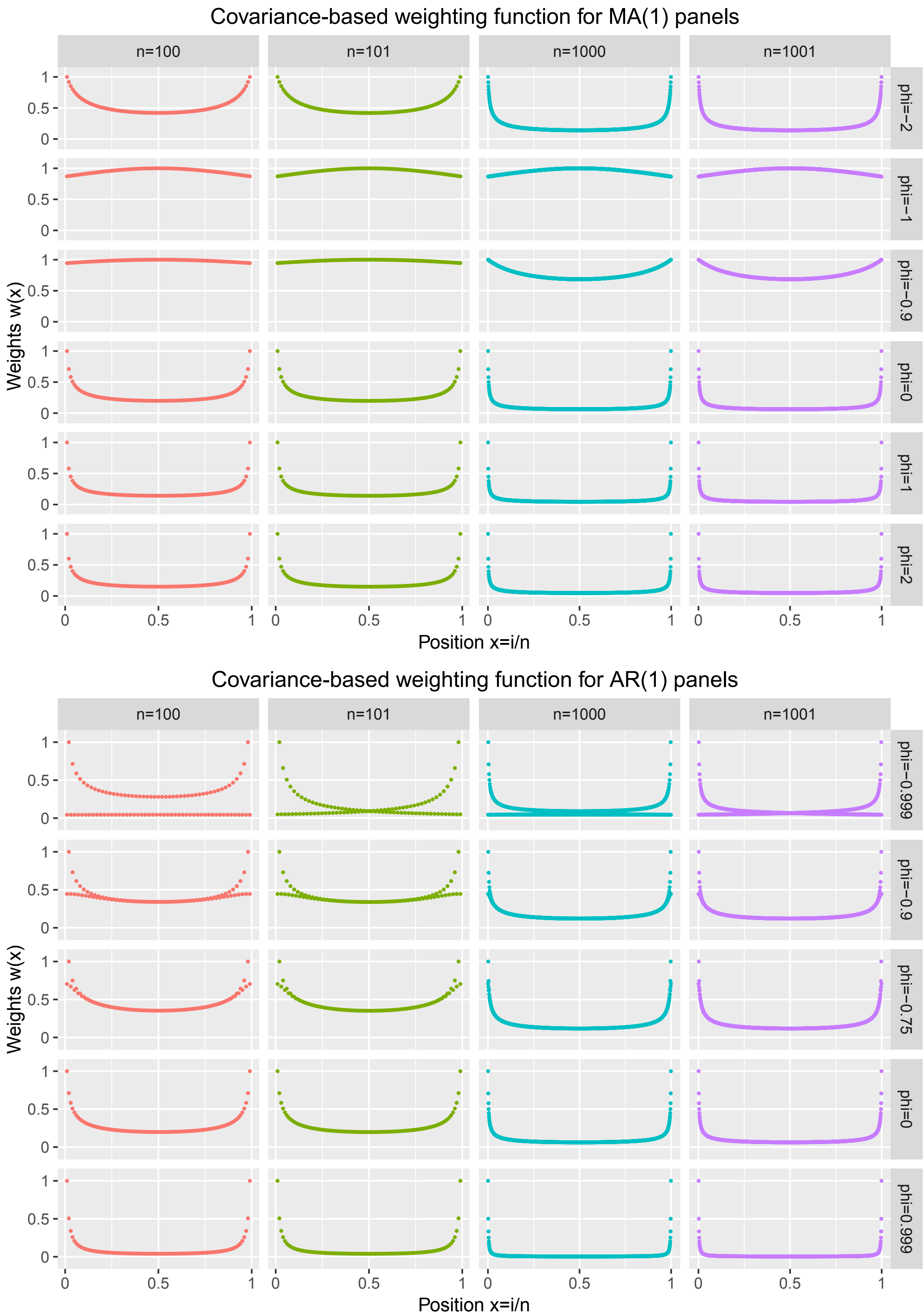

Figure 3.2: The upper panels show the covariance-based weights $w = cw_\star$ from Theorem 3.25 for MA(1) noise and based on (3.3.14) and the lower panels show the same type of weights for AR(1) noise based on (3.3.15). In both cases $c = (\max_{1\leq i<n} w_\star(i/n))^{-1}$ is used as a scaling for the sake of comparison and in both cases $\phi = 0$ corresponds to the Darling-Erdős-type weights.

Estimation of the covariance-based correction term

So far, we know $w_\star$ only theoretically and the computation requires the knowledge of the underlying time series model. In practice this information is often not available and we need to estimate $w_\star$ which will be briefly dealt with in this subsection closely following Torgovitski (2015d). First, note that the problem of estimating $w_\star$ is essentially the problem of estimating $\varSigma$, defined in (3.3.6), which follows by Lemma 3.10.

The aim of this subsection is to summarize conditions under which the entries $\varSigma_{j,k}$ can be estimated using the basic plug-in estimate

$$\hat{\varSigma}_{j,k} = \sum_{p=1}^{d} \langle Y_{j,p} - \bar{Y}_{j,d}, Y_{k,p} - \bar{Y}_{k,d} \rangle / d, \tag{3.3.28}$$

for $1 \le j, k \le n$ with means $\bar{Y}_{j,d} = \sum_{p=1}^{d} Y_{j,p}/d$.

Assumption N2 (on the noise). The weak law of large numbers holds true for the noise sequences $\{\varepsilon_{j,k}\}_{k\in\mathbb{N}}$ and $\{\langle \varepsilon_{j,k}, \varepsilon_{l,k} \rangle\}_{k\in\mathbb{N}}$ for any j and any l. The preceding Assumption N1 and Assumption CF hold true and the matrix $\varSigma$ is positive definite.

The next proposition describes further conditions on the means, shows that estimation is theoretically possible and that the influence of the common factors to such estimation is negligible. (Cf. also Remark 3.30, below.)

Proposition 3.28. *Let Assumption N2 hold true and assume that $m_{i,k} = c_i$ holds for each time point $1 \le i \le n$, for some $c_i \in \mathbb{R}$ and for all $k \in \mathbb{N}$, i.e. that the means, at any time point, are equal in all panels. Then it holds that $\hat{\varSigma} \xrightarrow{P} \varSigma$, as $d \to \infty$.*

The basic estimate (3.3.28) provides only rough approximations for small d. Two standard techniques can be used to improve this estimate, e.g., for MA(q) panels. On the one hand, we may average estimates across different time points. On the other hand, we may incorporate banding of $\hat{\varSigma}$ to reduce the variability by neglecting estimates of zero covariances. Let us assume a so-called training period $\{n_1, \dots, n_2\} \subset \{1, \dots, n\}$ such that $m_{i,k} = c_i$ holds true for any $n_1 \le i \le n_2$ and for all $k \in \mathbb{N}$. We follow Torgovitski (2015d) and define a banded and averaged estimate via

$$\hat{\varSigma}'_{j,j+r} = \hat{\varSigma}'_{j+r,j} = \frac{1}{n_2 - n_1 - r + 1} \begin{cases} \sum_{i=n_1}^{n_2-r} \hat{\varSigma}_{i,i+r}, & r \in \{0, \dots, \tilde{r}_j\}, \\ 0, & r \in \{\tilde{r}_j + 1, \dots, n - j\} \end{cases}$$

for $1 \le j \le n$ and with $\tilde{r}_j = \min\{h, n - j\}$. The parameter $h \in \{0, \dots, n_2 - n_1\}$ is the so-called *bandwidth*.

Corollary 3.29. *Assume MA(q) panels defined in Example 3.18. Let Assumption N1 and Assumption CF hold true and assume that $m_{i,k} = c_i$ holds for each time point $1 \le i \le n$ for some $c_i \in \mathbb{R}$ and independent of $k \in \mathbb{N}$. Then it holds that $\hat{\varSigma}' \xrightarrow{P} \varSigma$, as $d \to \infty$ and for $0 \le q \le h$.*

Remark 3.30. Proposition 3.28 and Corollary 3.29 require rather restrictive assumptions on the means $m_{i,k}$ for $1 \le i \le n$ and $k \in \mathbb{N}$. We discussed in Torgovitski (2015d) how these conditions may be ensured (approximatively) by an appropriate panel-wise centering and demonstrated in the corresponding simulations that this approach is beneficial in practice.

Remark 3.31 (Differencing of time series). Consider MA(1) panels defined in Example 3.18. The estimation of $w_\star$ or equivalently of Σ can be handled differently by using a differencing approach which we describe informally: after taking the differences within all panels of time series, the resulting panels are all mean zero (up to the one location of the common change point which needs to be excluded from further calculations). Hence, this allows to proceed similarly as under Proposition 3.28 or under Corollary 3.29 and to estimate the parameter ϕ which can then be used to estimate the tridiagonal matrix Σ. (This approach can be applied to MA(q) panels in a straightforward way.)

Remark 3.32 (Positive weights). Only strictly positive weighting schemes are reasonable for the estimate (3.2.7). Hence, estimates of weighting schemes shall be strictly positive, too, which is asymptotically ensured in Proposition 3.28 with probability tending to 1 by Σ being positive definite and the corresponding estimate $\hat{\Sigma}$ being consistent.

3.3.3 *Relation between weighted CUSUM and LASSO estimates*

We already explained in Section 3.2 that this chapter is motivated by Bleakley & Vert (2010, 2011a) and that our results are built upon their findings. They propose extensions of the weighted total variation denoising (or weighted LASSO) estimates in a multiple change point setting and present their theory under the single change point assumption in a high-dimensional low sample size scenario (i.e. for fixed n and $d \to \infty$). We describe their procedure for a real-valued univariate panel data setting and show for the single change point case (cf. Proposition 3.36, below) that the multivariate weighted LASSO approach coincides with the weighted CUSUM estimates under minimal restrictions. Hence, in such a single change point scenario, all results on uniformly asymptotic accurate estimation regarding CUSUM hold for weighted LASSO as well. Note that Bleakley & Vert (2011a) were (slightly) more concerned with algorithmic and computational aspects (and limited their framework to the i.i.d. setting) while our focus is more on the probabilistic aspects of a time series setting.

In the following we use the matrix terminology from Notation 3.1 and additionally write $\|\cdot\| = \|\cdot\|_2$ for the Euclidean norm to distinguish it from the Frobenius norm $\|\cdot\|_F$.

Weighted penalized least squares estimates

Consider the multiple common change point problem (3.2.2) for panels of univariate real-valued time series (3.2.1) in the matrix notation (3.2.3). Bleakley & Vert (2011a) propose to use *global* change point estimates that are obtained via the following penalized minimization problem

$$\underset{U\in\mathbb{R}^{n\times d}}{\text{Minimize}}\ \frac{1}{2}\|Y-U\|_F^2+\lambda\sum_{i=1}^{n-1}\|U_{i+1,\bullet}-U_{i,\bullet}\|_2/w(i/n) \tag{3.3.29}$$

with weights $w(x)>0,\ x\in I_n$, and a *»shrinkage«* parameter $\lambda\geq 0$ as the *»degrees of freedom«*. As in Torgovitski (2015d) we use

$$\hat{U}(\lambda)=(\hat{U}(\lambda)_{l,r})_{l=1,\dots,n,\,r=1,\dots,d}$$

to denote a matrix-valued solution of (3.3.29) for a chosen λ. Then, each time point v with $\hat{U}_{v,j}(\lambda)\neq\hat{U}_{v+1,j}(\lambda)$ in the solution of (3.3.29) is called a *»jump«* (in the j-th panel) and a jump corresponds, independently of j, to an estimate of a change point at v. More precisely, the set of change point estimates is defined via

$$\hat{\mathscr{E}}(\lambda)=\{v\mid \hat{U}_{v,\bullet}(\lambda)\neq\hat{U}_{v+1,\bullet}(\lambda)\}, \tag{3.3.30}$$

where the number of estimates is controlled by λ (cf. Notation 3.1). The minimization problem (3.3.29) aims to fit a *»piecewise constant signal«* to each panel such that all signals tend to have (common) jumps at the same positions. Therefore, (3.3.29) balances via λ between the first term that ensures a close fit of $\hat{U}$ and Y and the second term which controls the complexity (sparseness) of the model, i.e. controls the number of jumps or equivalently of estimates.

Remark 3.33 (Uniqueness of the solution). The solution of (3.3.29) is unique which follows due to the strict convexity of the objective function (the first term is strictly convex and the second term is convex) and by taking into account that the unbounded minimization domain may be formally redefined to be a compact set.

Some remarks on the shrinkage parameter

Via the shrinkage parameter λ one may control the number of estimated jumps. While λ increases less changes are identified - or in other words more estimates are *fused*. Seeking for p change points we want to select a λ such that (3.3.30) yields exactly p estimates. However, one has to be aware of following problematic and (at least) ambiguous situations:

1. A λ that yields exactly p change points might not exist. This may be observed already in the trivial case where $Y_{i,k}\equiv Y_{j,r}$ holds true for all $1\leq i,j\leq n$, $1\leq k,r\leq d$ since $\hat{\mathscr{E}}(\lambda)=\emptyset$ holds in this case for all λ (cf., e.g., Remark 2.2 of Torgovitski, 2015d).

2. The set $\hat{\mathscr{E}}(\lambda)$ is generally non-monotonic in λ for $d > 1$ (cf., e.g., Figure 1 of Torgovitski, 2015d). Hence, different λ's that lead to p estimated jumps might lead to different estimated jump locations, i.e. the choice of λ might be not »well-defined«.

To avoid such problems - at least in the single change point scenario - we will define an unambiguous estimate under mild assumptions. These assumptions ensure, on the one hand, that a λ' which yields exactly one change point estimate exists and is identifiable, and, on the other hand, that the set $\hat{\mathscr{E}}(\lambda)$ behaves monotonically in λ around that λ'.

Weighted group fused LASSO estimates

We begin by summarizing how Bleakley & Vert (2011a) restate (3.3.29) as a group fused LASSO. Note that this idea was applied previously by Harchaoui & Lévy-Leduc (2008) to the univariate case and substitutes the minimization problem (3.3.29) with the equivalent LASSO version

$$\underset{\beta\in\mathbb{R}^{(n-1)\times d}}{\text{Minimize}} \frac{1}{2}\|\bar{Y}-\bar{D}\beta\|_F^2+\lambda\sum_{i=1}^{n-1}\|\beta_{i,\bullet}\|_2, \tag{3.3.31}$$

where λ is chosen as in (3.3.29). Here, $D \in \mathbb{R}^{n\times(n-1)}$ is a *fixed design matrix* defined through the weighting scheme w via

$$D = \begin{bmatrix} 0 & 0 & 0 & \dots & & 0 \\ w(1/n) & 0 & 0 & & & \\ w(1/n) & w(2/n) & 0 & \ddots & & \vdots \\ w(1/n) & w(2/n) & w(3/n) & \ddots & & \vdots \\ \vdots, & \vdots & \vdots & \ddots & 0 & 0 \\ w(1/n) & w(2/n) & w(3/n) & \dots & w(1-2/n) & 0 \\ w(1/n) & w(2/n) & w(3/n) & \dots & w(1-2/n) & w(1-1/n) \end{bmatrix} \tag{3.3.32}$$

and $\bar{D}$ is the corresponding column-wise centered version of D, i.e. $\bar{D}_{i,\bullet} = D_{i,\bullet} - \sum_{j=1}^{n} D_{j,\bullet}/n$ for $1 \le i \le n$. Similarly, $\bar{Y}$ denotes the column-wise centered version of Y. The two minimization problems of (3.3.29) and (3.3.31) are equivalent in the sense that any solution $\hat{U} = \hat{U}(\lambda)$ of (3.3.29) yields a solution of (3.3.31) via

$$\hat{\beta}_{i,\bullet} = (\hat{U}_{i+1,\bullet} - \hat{U}_{i,\bullet})/w(i/n)$$

and, vice versa, any solution $\hat{\beta} = \hat{\beta}(\lambda)$ of (3.3.31) yields a solution of (3.3.29) via

$$\hat{U} = \mathbb{1}\hat{\gamma} + D\hat{\beta},$$

where $\hat{\gamma} = \mathbb{1}'(Y - D\hat{\beta})/n$. Hence, we obtain the correspondence

$$\hat{\mathscr{E}}(\lambda) = \{u \mid \hat{U}_{u,\bullet}(\lambda) \neq 0\} = \{u \mid \hat{\beta}_{u,\bullet}(\lambda) \neq 0\}$$

for the set of estimated change points. Furthermore, a β solves (3.3.31) if and only if the following *Karush-Kuhn-Tucker (KKT)* conditions are jointly satisfied:

$$\bar{D}'_{\bullet,i}(\bar{Y} - \bar{D}\beta) = \lambda\beta_{i,\bullet}/\|\beta_{i,\bullet}\| \text{ holds for all } i \text{ with rows } \beta_{i,\bullet} \neq 0. \tag{KKT1}$$

$$\|\bar{D}'_{\bullet,i}(\bar{Y} - \bar{D}\beta)\| \leq \lambda \text{ holds for all } i \text{ with rows } \beta_{i,\bullet} = 0. \tag{KKT2}$$

Weighted group fused LASSO in the single change point scenario

The next proposition extends Proposition 5.1 of Torgovitski (2015d). Recall that we only consider univariate real-valued panels in this subsection and that we write $\|\cdot\| = \|\cdot\|_2$.

Proposition 3.34. *Consider the random matrix $\hat{c} = \bar{D}'\bar{Y}$, where the fixed design matrix D is defined in (3.3.32) and the observable random matrix Y is defined in (3.2.1). Set $t_i = \|\hat{c}_{i,\bullet}\|$ for $i = 1, \ldots, n-1$ and assume that $t_{i_1} \leq t_{i_2} \leq \ldots \leq t_{i_{n-1}}$ for $i_k \neq i_r$ with $k \neq r$. Then we obtain for $M = i_{n-1}$ and $m = i_{n-2}$ that:*

1. *For any $\lambda \in [t_M, \infty)$ the matrix $\beta = 0$ fulfills conditions (KKT1) and (KKT2). Thus, it is a solution of (3.3.31) and we obtain $\hat{\mathscr{E}}(\lambda) = \{\emptyset\}$.*

2. *For any $\lambda \in (\lambda_{\min}, t_M)$, where $\lambda_{\min} \in (t_m, t_M)$ is random and $t_m < t_M$ (i.e. $\{t_i\}_{i=1,\ldots,n-1}$ has a unique maximum), the matrix β with rows*

$$\beta_{i,\bullet} = \alpha_M \begin{cases} \hat{c}_{M,\bullet}, & i = M, \\ 0, & i \neq M \end{cases} \tag{3.3.33}$$

   and $\alpha_M = (t_M - \lambda)/((\bar{D}_{\bullet,M})'\bar{D}_{\bullet,M}t_M)$ fulfills conditions (KKT1) and (KKT2). Thus it is a solution of (3.3.31) and we obtain $\hat{\mathscr{E}}(\lambda) = \{M\}$.

3. *For any $\lambda \in [\lambda_{\min}, t_M)$, where $\lambda_{\min} \in [0, t_M)$ is random and given that $t_m = t_M$ with $\hat{c}_{m,\bullet} = \hat{c}_{M,\bullet}$, there is no matrix β with rows*

$$\beta_{i,\bullet} \begin{cases} \neq 0, & i = m \text{ or } i = M, \\ = 0, & \text{otherwise}, \end{cases}$$

   which fulfills both conditions (KKT1) and (KKT2) at the same time, given that

$$\xi := 1 - ((\bar{D}_{\bullet,\tilde{m}})'\bar{D}_{\bullet,\tilde{M}})/((\bar{D}_{\bullet,\tilde{M}})'\bar{D}_{\bullet,\tilde{M}}) > 0 \tag{3.3.34}$$

   holds true for $\tilde{m} = \min\{m, M\}$ and $\tilde{M} = \max\{m, M\}$.

Remark 3.35 (Selection of λ in the single change point scenario). The first and the second statement of Proposition 3.34 show monotonicity of $\hat{\mathscr{E}}(\lambda)$ with respect to large λ. In the case of $t_M > t_m$ we see that any $\lambda \in (\lambda_{\min}, t_M)$ results in the same unambiguous estimate $\hat{\mathscr{E}}(\lambda) = \{M\}$ of a single change point at position M.

Finally, we are in the position to show the relation of the weighted LASSO to the weighted CUSUM estimates. The following proposition is given in Torgovitski (2015d, Proposition 2.4).

Proposition 3.36. *Assume that* $\{t_i\}_{i=1,\ldots,n-1}$ *has a unique maximum*[1] *and that* λ *is selected according to Remark 3.35. Then it holds that*

$$\underset{1\le i<n}{\arg\max}\, w(i/n)\mathscr{T}(i/n) = \hat{\mathscr{E}}(\lambda). \tag{3.3.35}$$

The CUSUM estimate on the left-hand side is specified in (3.2.7) and the group fused LASSO estimate on the right-hand side is specified in Remark 3.35. (We assume that we use the same weighting function w *for both estimates.)*

Remark 3.37. Proposition 3.36 states the identity (3.3.35) assuming that $\{t_i\}_{i=1,\ldots,n-1}$ has a unique maximum. Proposition 3.34 adds a minor clarification of the behavior of the group fused LASSO in the opposite case. Let $\{t_i\}_{i=1,\ldots,n-1}$ have a non-unique maximum and consider the following situation $t_{i_{n-1}} = t_M = t_m = t_{i_{n-2}} > t_{i_{n-3}}$ (see the notation in Proposition 3.34). If (3.3.34) is fulfilled, then the third statement of Proposition 3.34 implies that there is some random $\lambda_{\min}$ such that no $\lambda \in (\lambda_{\min}, t_M)$ provides a single change point estimate $\hat{\mathscr{E}}(\lambda) = \{m\}$ or $\hat{\mathscr{E}}(\lambda) = \{M\}$. (The latter condition is fulfilled, e.g., for the weighting $w_0 \equiv 1$ which follows immediately by $((\bar{D}_{\bullet,m})'\bar{D}_{\bullet,M})/((\bar{D}_{\bullet,M})'\bar{D}_{\bullet,M}) = [m(n-M)]/[M(n-M)] = m/M \in (0,1)$ in case of $m < M$.)

[1] This means that $t_M > t_m$ holds true in Proposition 3.34.

A small simulation study 3.4

The aim of this section is to demonstrate that the CUSUM estimates $\hat{u}$ from (3.2.7) with a covariance-based weighting function $w_\star$ outperform the traditional estimates with a Darling-Erdős-type weighting function $w_{1/2}$ in the HDLSS panel data framework under some moderate temporal dependence.[1] For simplicity we restrict ourselves to univariate real-valued panels

$$Y_{i,k} = m_{i,k} + \varepsilon_{i,k}, \tag{3.4.1}$$

with $1 \leq i \leq n,\ 1 \leq k \leq d$. Moreover, we do not take any common factors into account by formally setting $\gamma_k = 0$ for all $1 \leq k \leq d$ in (3.2.1). For the sake of a clearer presentation, we assume a simple common change point model

$$m_{i,k} = \begin{cases} 0, & i = 1, \dots, u, \\ m, & i = u+1, \dots, n, \end{cases}$$

with the same change magnitude $m > 0$ across all panels, i.e. with $\Delta = m^2$ (cf. (3.2.6) and (3.2.9)). For the noise terms $\{\varepsilon_{i,k}\}$ in (3.4.1) we consider the MA(1) and AR(1) settings from Examples 3.13 and 3.14, i.e. the noise sequence is either given by

$$\varepsilon_{i,k} = \eta_{i,k} + \phi\eta_{i-1,k} \tag{3.4.2}$$

or by

$$\varepsilon_{i,k} = \sum_{j=0}^{\infty} \phi^j \eta_{i-j,k}, \tag{3.4.3}$$

for all $1 \leq i \leq n,\ 1 \leq k \leq d$. For the innovation sequence $\{\eta_{i,k}\}$ we use in both cases independent and standard normally distributed random variables.

In this section we complement the thorough simulation study on accuracy for MA(1) panels in Torgovitski (2015d) and add the AR(1) case. Our simulations are implemented in R where we simulate both (3.4.2) and (3.4.3) using the arima.sim function from the basic {stats}-R-package. Figure 3.3 and Figure 3.4 show the accuracy of the CUSUM estimates for various dimensions d, for a range of parameters ϕ and for different weighting functions w_γ, where all reported results base on 1000 repetitions. Both figures show a matrix of plots where the columns correspond to different dimensions and the rows to different weighting functions. Each plot in this matrix shows the accuracy for a range of parameters ϕ. For example, the leftmost plot in the top panel shows the accuracy for the combination of w_0 with $d = 1$. Note that for the sake of a more concise notation, we denote $w_\star$ by w_γ with $\gamma := \star$. (Recall that the weighting $w_\star$ is shown in Figure 3.2 for different sample sizes and different parameters ϕ and also that convenient formulas for this weighting are summarized in Table 3.1.) Our simulation results show that estimates weighted by $w_\star$ are asymptotically more accurate than those weighted by w_0 or by $w_{1/2}$ and that the accuracy of estimates weighted by $w_\star$ increases, as $d \to \infty$. These observations are in accordance with the theoretical results of this section.

[1] We choose $\hat{u}$ as the smallest element in $\arg\max_{1 \leq i < n} w(i/n)\mathscr{T}(i/n)$.

Note that w_0 and $w_{1/2}$ show a complementary behavior for the chosen parameters: CUSUM estimates with w_0 have high accuracy whenever $w_{1/2}$ yields low accuracy and vice versa. For instance, we see in Figure 3.3 that using w_0 is better for parameters $\phi \leq 0$ whereas using $w_{1/2}$ is advantageous for $\phi > 0$. We observe a similar pattern in Figure 3.3. In the demonstrated simulations the covariance-based weighting $w_\star$ seems to combine the good properties of w_0 for negative ϕ and of $w_{1/2}$ for positive ϕ. The reason, at least in the MA(1) case, is that $w_\star$ indeed approximates w_0 for parameters ϕ close to -1 and approximates $w_{1/2}$ for positive parameters ϕ (cf. Figure 3.2). We finish our discussion by two observations. In Figure 3.3 and in Figure 3.4 we see that a weaker dependence structure yields higher accuracy for estimates weighted by $w_\star$ and that w_0 leads to spurious estimation in the i.i.d situation of $\phi = 0$. (The latter was already observed by Bleakley & Vert, 2011a.)

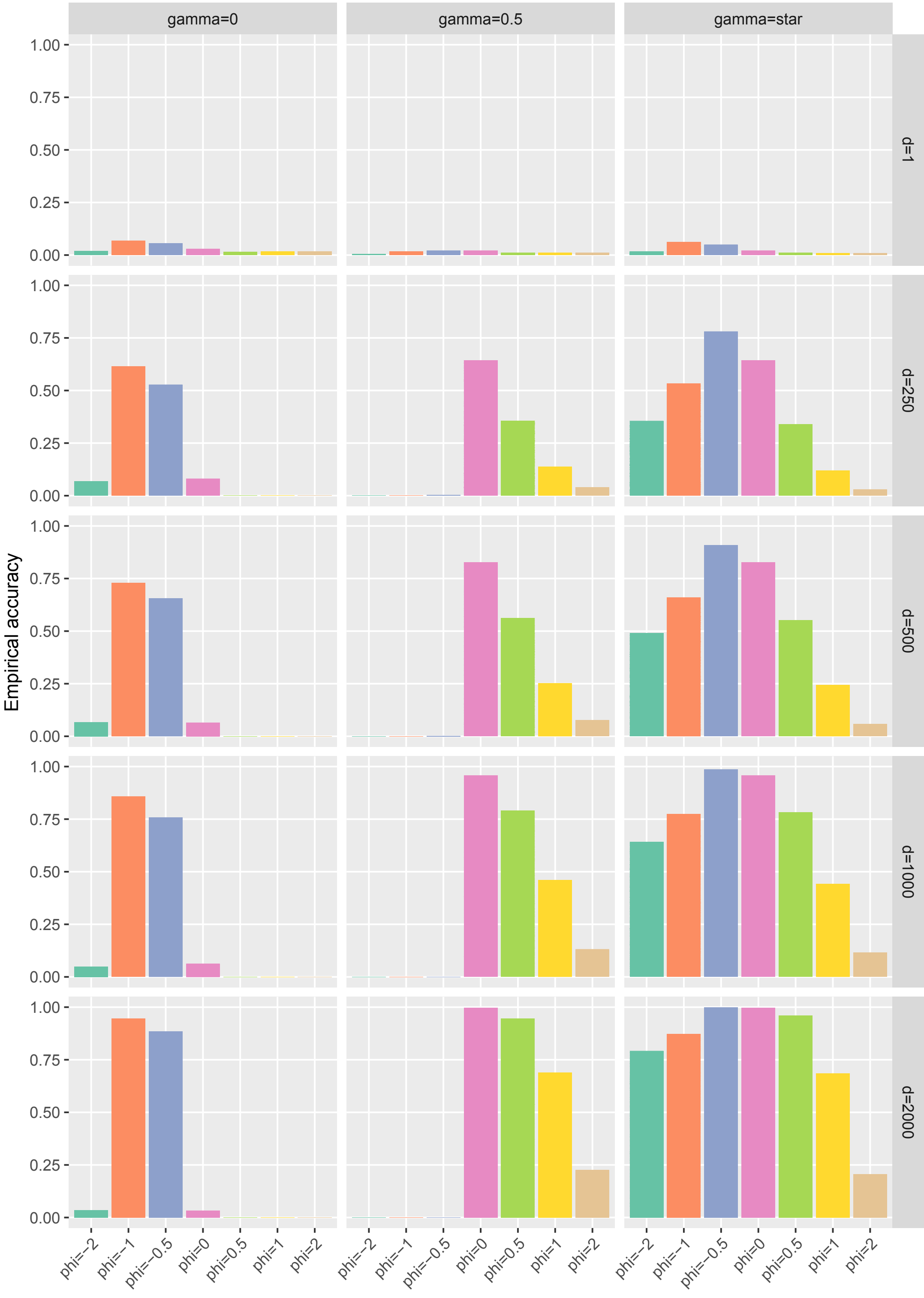

Figure 3.3: CUSUM estimates $\hat{u}$ weighted by w_γ in the MA(1) setting. The sample size is $n = 100$, the common change point is at $u = 75$ and the change magnitude is $m = 1/5$.

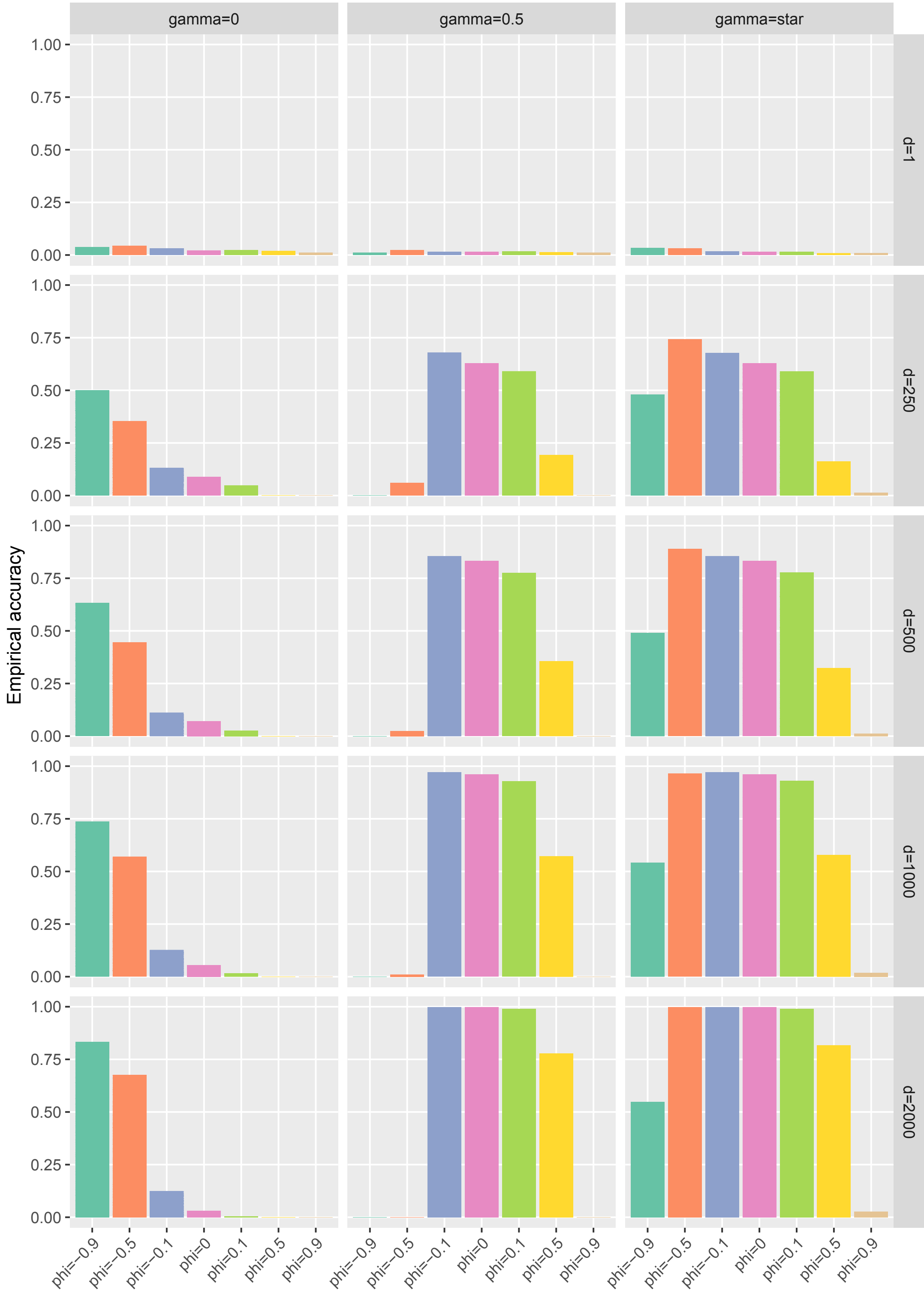

Figure 3.4: CUSUM estimates $\hat{u}$ weighted by w_γ in the AR(1) setting. The sample size is $n = 100$, the common change point is at $u = 75$ and the change magnitude is $m = 1/5$.

Proofs 3.5

Proof of Theorem 3.6. We provide a detailed version of the proof of Torgovitski (2015d, Theorem 2.9) that is also extended to the Hilbert space valued setting (cf. also Torgovitski, 2015d, arXiv:1501.00177v2, Theorem 2.18). Let $x = i/n$ be the rescaled time point for any fixed $1 \le i \le n$. First, we decompose the partial sums into the contributions of the noise, the common factors and the changes, i.e.

$$S_{n,k}(x; Y) = S_{n,k}(x; \varepsilon) + \gamma_k S_{n,k}(x; \zeta) - n^{1/2} \Delta_k H(x, s). \tag{3.5.1}$$

In the following we will write $S_n(x;\zeta) = S_{n,k}(x;\zeta)$ since the partial sum of common factors is independent of k and use the notation from (3.2.9) for Δ. For the squared norm of (3.5.1) we have

$$\begin{aligned}
&\|S_{n,k}(x;Y)\|^2 \\
&= \|S_{n,k}(x;\varepsilon) + \gamma_k S_n(x;\zeta)\|^2 \\
&\quad - 2\langle S_{n,k}(x;\varepsilon) + \gamma_k S_n(x;\zeta), \Delta_k\rangle n^{1/2} H(x,s) + n\|\Delta_k\|^2 H^2(x,s) \\
&= \|S_{n,k}(x;\varepsilon)\|^2 + 2\langle \gamma_k S_{n,k}(x;\varepsilon), S_n(x;\zeta)\rangle + \gamma_k^2 \|S_n(x;\zeta)\|^2 \\
&\quad - 2\langle S_{n,k}(x;\varepsilon), \Delta_k\rangle n^{1/2} H(x,s) - 2\langle S_n(x;\zeta), \gamma_k \Delta_k\rangle n^{1/2} H(x,s) \\
&\quad + n\|\Delta_k\|^2 H^2(x,s).
\end{aligned} \tag{3.5.2}$$

Applying the Cauchy-Schwarz inequality twice, we obtain

$$E\|S_n(x;\zeta)\|^2 = \sum_{j,q=1}^{n} a_j^{(i)} a_q^{(i)} E\langle \zeta_j, \zeta_q\rangle \le \sum_{j,q=1}^{n} a_j^{(i)} a_q^{(i)} \left[E\|\zeta_j\|^2 E\|\zeta_q\|^2\right]^{1/2} < \infty, \tag{3.5.3}$$

where the coefficients $a_j^{(i)}$ are defined via (3.3.10) and $x = i/n$. Similarly, we get that $E\|S_{n,k}(x;\varepsilon)\|^2 < \infty$, where in both cases we use that $\{\zeta_j\}$ and $\{\varepsilon_{j,k}\}$ have finite second moments. Since $S_{n,k}(x;\varepsilon)$ and $S_n(x;\zeta)$ are independent and centered, we observe, using (3.5.2) and (3.5.3), that

$$\begin{aligned}
&E\|S_{n,k}(x;Y)\|^2 \\
&= E\|S_{n,k}(x;\varepsilon) + \gamma_k S_n(x;\zeta)\|^2 \\
&\quad - 2\langle E[S_{n,k}(x;\varepsilon) + \gamma_k S_n(x;\zeta)], \Delta_k\rangle n^{1/2} H(x,s) + n\|\Delta_k\|^2 H^2(x,s) \\
&= E\|S_{n,k}(x;\varepsilon)\|^2 + \gamma_k^2 E\|S_n(x;\zeta)\|^2 + n\|\Delta_k\|^2 H^2(x,s).
\end{aligned}$$

This yields, as $d \to \infty$,

$$\begin{aligned}
&E\sum_{k=1}^{d} \|S_{n,k}(x;Y)\|^2/d \\
&= \sum_{k=1}^{d} E\|S_{n,k}(x;\varepsilon)\|^2/d \\
&\quad + \left[\sum_{k=1}^{d} \gamma_k^2/d\right] E\|S_n(x;\zeta)\|^2 + \left[\sum_{k=1}^{d} \|\Delta_k\|^2/d\right] nH^2(x,s) \\
&= V^2(x)\sigma^2 + H^2(x,s)(n\Delta) + o(1),
\end{aligned} \tag{3.5.4}$$

for any $x \in I_n$, where in the last line we used Assumption N1, Assumption CF and (3.2.9). In the next step we show that the convergence (3.5.4) also holds in probability. Therefore, by Chebyshev's inequality, it is sufficient to show that

$$\operatorname{Var}\left(\sum_{k=1}^{d} \|S_{n,k}(x;Y)\|^2/d\right) = o(1) \tag{3.5.5}$$

holds true for any $x \in I_n$, as $d \to \infty$. Below, we will show that

$$\operatorname{Var}\left(\sum_{k=1}^{d} \|S_{n,k}(x;\varepsilon)\|^2\right) = o(d^2), \tag{V1}$$

$$\operatorname{Var}\left(\langle S_n(x;\zeta), \sum_{k=1}^{d} \gamma_k S_{n,k}(x;\varepsilon)\rangle\right) = o(d^2), \tag{V2}$$

$$\operatorname{Var}\left(\sum_{k=1}^{d} \gamma_k^2 \|S_n(x;\zeta)\|^2\right) = o(d^2), \tag{V3}$$

$$\operatorname{Var}\left(\sum_{k=1}^{d} \langle \Delta_k, S_{n,k}(x;\varepsilon)\rangle\right) = o(d^2), \tag{V4}$$

$$\operatorname{Var}\left(\sum_{k=1}^{d} \langle \Delta_k \gamma_k, S_n(x;\zeta)\rangle\right) = o(d^2) \tag{V5}$$

hold true for any $x \in I_n$ and therefore (3.5.5) follows via (3.5.2) by the Cauchy-Schwarz inequality.

Bound for (V1)*:* From $\|S_{n,k}(x;\varepsilon)\|^2 = \sum_{j,q=1}^{n} a_j^{(i)} a_q^{(i)} \langle \varepsilon_{j,k}, \varepsilon_{q,k}\rangle$, $x = i/n$, we obtain

$$\begin{aligned}
\operatorname{Var}\left(\sum_{k=1}^{d} \|S_{n,k}(x;\varepsilon)\|^2\right) &= \sum_{k,r=1}^{d} \sum_{j,q,l,m=1}^{n} a_j^{(i)} a_q^{(i)} a_l^{(i)} a_m^{(i)} \operatorname{Cov}\Big(\langle \varepsilon_{j,k}, \varepsilon_{q,k}\rangle, \langle \varepsilon_{l,r}, \varepsilon_{m,r}\rangle\Big) \\
&= \sum_{j,q,l,m=1}^{n} a_j^{(i)} a_q^{(i)} a_l^{(i)} a_m^{(i)} \sum_{k,r=1}^{d} \operatorname{Cov}\Big(\langle \varepsilon_{j,k}, \varepsilon_{q,k}\rangle, \langle \varepsilon_{l,r}, \varepsilon_{m,r}\rangle\Big)
\end{aligned}$$

and hence (3.3.2) implies (V1).

Bound for (V2)*:* Since $S_n(x;\zeta)$ and $S_{n,k}(x;\varepsilon)$ are independent, centered and due to (3.5.3), we obtain

$$\begin{aligned}
&\operatorname{Var}\left(\langle S_n(x;\zeta), \sum_{k=1}^{d} \gamma_k S_{n,k}(x;\varepsilon)\rangle\right) \\
&= E\langle S_n(x;\zeta), \sum_{k=1}^{d} \gamma_k S_{n,k}(x;\varepsilon)\rangle^2 \\
&\le E\|S_n(x;\zeta)\|^2 E\|\sum_{k=1}^{d} \gamma_k S_{n,k}(x;\varepsilon)\|^2
\end{aligned}$$

$$
\begin{aligned}
&= E\|S_n(x;\zeta)\|^2 \sum_{k,r=1}^{d} |\gamma_k\gamma_r| E\langle S_{n,k}(x;\varepsilon), S_{n,k}(x;\varepsilon)\rangle \\
&= E\|S_n(x;\zeta)\|^2 \sum_{j,q=1}^{n} a_j a_q \sum_{k,r=1}^{d} |\gamma_k\gamma_r| E\langle \varepsilon_{j,k}, \varepsilon_{q,r}\rangle
\end{aligned}
$$

and thus (V2) follows in view of (3.3.4).

Bound for (V3)*:* Analogously to (3.5.3), we obtain by the finiteness of $E(\|\zeta_i\|^4)$ that

$$
\begin{aligned}
&\operatorname{Var}\left(\|S_n(x;\zeta)\|^2\right) \\
&= \sum_{j,q,l,m=1}^{n} a_j^{(i)} a_q^{(i)} a_l^{(i)} a_m^{(i)} \operatorname{Cov}\Big(\langle\zeta_j,\zeta_q\rangle, \langle\zeta_l,\zeta_m\rangle\Big) \\
&\leq \sum_{j,q,l,m=1}^{n} a_j^{(i)} a_q^{(i)} a_l^{(i)} a_m^{(i)} \left[\operatorname{Var}\Big(\|\zeta_j\|\|\zeta_q\|\Big) \operatorname{Var}\Big(\|\zeta_l\|\|\zeta_m\|\Big)\right]^{1/2} < \infty
\end{aligned}
\tag{3.5.6}
$$

and (V3) follows from

$$
\operatorname{Var}\left(\sum_{k=1}^{d} \gamma_k^2 \|S_n(x;\zeta)\|^2\right) = \left[\sum_{k=1}^{d} \gamma_k^2\right]^2 \operatorname{Var}\left(\|S_n(x;\zeta)\|^2\right)
$$

combined with Assumption CF.

Bound for (V4)*:* It holds that

$$
\operatorname{Var}\left(\sum_{k=1}^{d} \langle\Delta_k, S_{n,k}(x;\varepsilon)\rangle\right) = \sum_{j,q=1}^{n} a_j^{(i)} a_l^{(i)} \sum_{k,r=1}^{d} E\Big(\langle\Delta_k, \varepsilon_{j,k}\rangle\langle\Delta_r, \varepsilon_{l,r}\rangle\Big)
$$

and therefore (3.3.3) implies (V4).

Bound for (V5)*:* Our assumptions on Δ_k (cf. (3.2.9)) and Assumption CF yield, via the Cauchy-Schwarz inequality, that, as $d \to \infty$,

$$
\begin{aligned}
\operatorname{Var}\left(\sum_{k=1}^{d} \langle\Delta_k\gamma_k, S_n(x;\zeta)\rangle\right) &\leq \left[\sum_{k=1}^{d} |\gamma_k|\|\Delta_k\|\right]^2 \operatorname{Var}\Big(\|S_n(x;\zeta)\|\Big) \\
&\leq \left[\sum_{k=1}^{d} \gamma_k^2\right]\left[\sum_{k=1}^{d} \|\Delta_k\|^2\right] \operatorname{Var}\Big(\|S_n(x;\zeta)\|\Big) \\
&= o(d)\mathcal{O}(d) \operatorname{Var}\Big(\|S_n(x;\zeta)\|\Big).
\end{aligned}
$$

Hence, (V5) follows in view of (3.5.6).

Altogether, the relations (V1)-(V5) imply (3.5.5) and making use of (3.5.4) we obtain

$$
[n\Delta d]^{-1}\left[w^2(x) \sum_{k=1}^{d} \|S_n(x;Y)\|^2\right] \xrightarrow{P} C(x;s,\rho),
\tag{3.5.7}
$$

as $d \to \infty$. To finish our proof we set $S = \arg\max_{1\le i<n} \mathcal{C}(i/n;\mathsf{s},\rho)$ and tacitly assume that $S \subsetneq \{1,\ldots,n-1\}$ because the case of $S = \{1,\ldots,n-1\}$ is trivial. Relation (3.5.7) now yields that, as $d \to \infty$,

$$P\big(\arg\max_{1\le i<n} w(i/n)\mathscr{T}(i/n) \subseteq S\big) \ge P\big(\max_{i\in S} w(i/n)\mathscr{T}(i/n) > \max_{j\notin S} w(j/n)\mathscr{T}(j/n)\big)$$
$$\to P\big(\max_{i\in S} \mathcal{C}(i/n;\mathsf{s},\rho) > \max_{j\notin S} \mathcal{C}(j/n;\mathsf{s},\rho)\big) = 1$$

and the assertion follows. □

Proof of Lemma 3.10. Straightforward calculations yield

$$\begin{aligned}
c\,\mathcal{V}^2(i/n) &= E\Big\|(n-i)\sum_{j=1}^{i}\varepsilon_{j,1} - i\sum_{j=i+1}^{n}\varepsilon_{j,1}\Big\|^2 \\
&= (n-i)^2\sum_{k,r=1}^{i} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle + i^2\sum_{k,r=i+1}^{n} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle \\
&\quad - 2(n-i)i\sum_{k=1}^{i}\sum_{r=i+1}^{n} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle \\
&= [n(n-i)]\sum_{k,r=1}^{i} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle + [ni]\sum_{k,r=i+1}^{n} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle \\
&\quad - [i(n-i)]\sum_{k,r=1}^{n} E\langle\varepsilon_{k,1},\varepsilon_{r,1}\rangle,
\end{aligned}$$

which shows (3.3.11) for $c = \sigma^2 n^3$. Under weak stationarity of $\{\varepsilon_{r,1}\}_{r=1,\ldots,n}$ the last expression shows that

$$c\,\mathcal{V}^2(i/n) = [n(n-i)]\Big(i\Gamma_{i/n}\Big) + [ni]\Big((n-i)\Gamma_{1-i/n}\Big) - [i(n-i)]\Big(n\Gamma_1\Big)$$

holds true with Γ_x defined in (3.3.8) and this in turn verifies (3.3.12). □

Based on Lemma 3.10 we are able to show a different and simpler proof for the MA(1)-Example 3.13 than in Torgovitski (2015d). Similar considerations then extend to the autoregressive and the MA(q) cases.

Proof of Example 3.13. First, notice that

$$c\,\Gamma_{i/n} = \Big[(1+\phi^2)i + 2(i-1)\phi\Big]/i = (1+\phi^2+2\phi) - 2\phi/i$$

holds true for some $c > 0$. We use formula (3.3.7) to obtain

$$c\,\mathcal{F}(i/n) = (i/n)(1-i/n)\Big([1+\phi^2+2\phi] - 2[\phi/i + \phi/(n-i)] + 2\phi/n\Big)$$

and the assertion follows via Lemma 3.10 since $1/i + 1/(n-i) = [n(i/n)(1-i/n)]^{-1}$. □

We continue with the variance of cumulated noises in the AR(1) case.

Proof of Example 3.14. Standard calculations yield

$$\begin{aligned}
c\,\Gamma_{i/n} &= \sum_{r=1}^{i}\sum_{j=1}^{i}\gamma(r,j)/i \\
&= \sum_{r=1}^{i}\Bigg(\sum_{j=0}^{i-r}\phi^j+\sum_{j=0}^{r-1}\phi^j-\phi^0\Bigg)/i \\
&= 2\sum_{r=1}^{i}\sum_{j=0}^{r-1}\phi^j/i-1 \\
&= 2(1-\phi)^{-1}-2(\phi-\phi^{i+1})(1-\phi)^{-2}/i-1 \\
&= [(1-\phi^2)-2(\phi-\phi^{i+1})/i](1-\phi)^{-2}
\end{aligned}$$

for some $c>0$. Thus, $\tilde{c}\,\Gamma_{i/n}=(1-\phi^2)-2(\phi-\phi^{i+1})/i$ for some $\tilde{c}>0$ and we finish the proof by using formula (3.3.7) together with Lemma 3.10. □

Proof of Lemma 3.15 (as given in Lemma 2.13 in Torgovitski, 2015d). We may define for any s a strictly concave function $f_s(x)=V(x)/V(s)$ and a linearly increasing function $g_s(x)=x/s$. It holds that $f_s(s)=g_s(s)$ and that $f_s(1/n)>g_s(1/n)$. Hence, strict concavity yields $f_s(x)>g_s(x)$ for all $x<s$ which finishes the proof. □

Proof of Lemma 3.16. We begin with the MA(1) case and show a slightly modified proof of Remark 2.14 of Torgovitski (2015d). We treat the cases $\alpha(\phi)>0$ and $\alpha(\phi)\le 0$ separately to illustrate how Lemma 3.15 may be applied in the former »concave« situation.

$\alpha(\phi)>0$: According to Lemma 3.15 it suffices to verify the positivity of

$$\left[V^2(1/n)/(1/n)^2-V^2(y)/y^2\right]=c\,(ny-1)\Big[\phi^2 ny-2\phi(1-y)+ny\Big]/y^2 \tag{3.5.8}$$

$$=c\,(ny-1)\Big[(ny+ny\phi^2-2\phi)+2\phi y\Big]/y^2 \tag{3.5.9}$$

for $y=2/n,\ldots,(n-1)/n$, where $c>0$ is some positive constant. The term on the right-hand side of (3.5.8) is positive for all $\phi\le 0$ and the term on the right-hand side of (3.5.9) is positive for all $\phi\ge 0$ since $(ny+ny\phi^2-2\phi)>1+\phi^2-2\phi\ge 0$.

$\alpha(\phi)\le 0$: We consider the function

$$h(x):=[x(1-x)\alpha(\phi)-\mathcal{R}(\phi)]/x^2,$$

where x takes values in the whole domain $x\in(0,1)$, and show that the derivative $\partial_x h(x)$ is negative which then implies the assertion. It holds that

$$cx^3\partial_x h(x)=-\alpha(\phi)x+2\phi<-\alpha(\phi)+2\phi$$

for some $c>0$ and all $x\in(0,1)$. Now, the assertion follows since the right-hand side is negative if $1+\phi^2>(2-2/n)\phi$, which is fulfilled because $\alpha(\phi)\le 0$ implies $\phi<0$. (Note that we may apply a similar argument in the first case of $\alpha(\phi)>0$ without making use of Lemma 3.15.) □

The next proof is for the variance of cumulated noises within the MA(q) case.

Proof of Example 3.18. After some basic calculations we arrive at

$$c\Gamma_{i/n} = \sum_{j=0}^{q} \phi_j^2 + 2\sum_{k=1}^{i-1}(1-k/i)\sum_{j=0}^{q}\phi_j\phi_{j+k},$$

for $1 \leq i \leq n$, which simplifies to

$$\begin{aligned} c\Gamma_{i/n} &= \sum_{j=0}^{q} \phi_j^2 + 2\sum_{k=1}^{q}(1-k/i)\sum_{j=0}^{q}\phi_j\phi_{j+k} \\ &= \Big[\sum_{j=0}^{q}\phi_j^2 + 2\sum_{k=1}^{q}\sum_{j=0}^{q}\phi_j\phi_{j+k}\Big] - \Big[2\sum_{k=1}^{q}\sum_{j=0}^{q}k(\phi_j\phi_{j+k})\Big]/i \end{aligned}$$

for $q < i$ (cf. (3.3.17)) and for some $c > 0$. As before, we use formula (3.3.7) and Lemma 3.10 to finish the proof. □

Remark 3.38. Recall the Definition 3.5 of $C(x;\varsigma,\rho)$ and, furthermore, that we consider $w_\gamma(x) = [x(1-x)]^{-\gamma}$, $\gamma \in (0,1/2)$ and $V^2(x) = x(1-x)$ in the subsequent proofs of Theorem 3.21 and of Proposition 3.22.

For the following proofs it is convenient to define the differentiable versions of the discrete critical function $C(x;\varsigma,\rho)$ and of (3.3.20), namely

$$\mathscr{C}(x;\rho) := \mathscr{F}(x)\rho + \mathscr{G}(x,\varsigma)$$

and

$$\mathscr{R}(x,s) := [\mathscr{G}(x,s) - \mathscr{G}(s,s)]/[\mathscr{F}(s) - \mathscr{F}(x)]$$

with $\mathscr{F}(x) = [x(1-x)]^{1-2\gamma}$ and $\mathscr{G}(x,s) = [x(1-x)]^{-2\gamma}[x(1-s)]^2$, where $x \in [0,1)$. Notice that the identity $\mathscr{C}(x;\rho) \equiv C(x;\varsigma,\rho)$ holds only for $x \in [0,\varsigma] \cap I_n$, where $\varsigma(s) = \max\{s, 1-s\}$ is defined as before, and that by definition we have

$$\inf_{1/2 \leq x < \varsigma} R(x,\varsigma) = \inf_{1/2 \leq x < \varsigma} \mathscr{R}(x,\varsigma) = \inf\{\rho \,|\, \mathscr{C}(x;\rho) = \mathscr{C}(\varsigma;\rho), 1/2 \leq x < \varsigma\}.$$

Proof of Theorem 3.21. This is a detailed version of the proof of Theorem 2.17 in Torgovitski (2015d) but developed in a (as it turns out more convenient) time-rescaled setting to clarify the ideas. We consider the partial derivative

$$\partial_x \mathscr{C}(x;\rho+\varepsilon) = \partial_x \mathscr{C}(x;\rho) + \varepsilon \partial_x \mathscr{F}(x), \tag{3.5.10}$$

for $\rho > 0$, $\varepsilon > 0$, and the function

$$\mathscr{P}(x,\rho) := h(x)\partial_x \mathscr{C}(x;\rho),$$

where $h(x) = (x(1-x))^{2\gamma+1}/x$ and $\mathscr{P}(x,\rho)$ are restricted to $x \in (0,1)$. Notice that the derivative $\partial_x \mathscr{C}(x;\rho)$ and the rescaled derivative $\mathscr{P}(x,\rho)$ have the same zeros as functions of x since $h(x) > 0$.

A combination of the following properties P1) - P3) (essentially) yields the assertion:

P1) It holds that $\mathscr{C}(0;\rho)=0$ and that $\mathscr{C}(x;\rho)\uparrow\infty$ for $x\uparrow 1$. Hence, by the continuity of $\mathscr{C}$, a (strict) local maximum of $\mathscr{C}(x;\rho)$ at some $x_{\max}(\rho)\in(0,1)$ implies the existence of a a (strict) local minimum at some $x_{\min}(\rho)\in(x_{\max}(\rho),1)$ and vice versa. Since $\mathscr{P}(x,\rho)$ is a polynomial in x with degree 2, $\mathscr{C}(x;\rho)$ may have on $x\in(0,1)$ either one isolated critical point as a saddle point or two isolated critical points, i.e. exactly one strict local maximum and one strict local minimum.

P2) Since $\rho_{\max}(s)<\infty$, we know by P1 that $\partial_x\mathscr{C}(x^*;\rho^*)=0$ holds true for some isolated saddle point $x^*\in(0,1)$ given some $\rho^*>0$. Hence, a strict local maximum and a strict local minimum of $\mathscr{C}(x;\rho^*+\varepsilon)$, restricted to $x\in(0,1)$, are always located for $\rho=\rho^*+\varepsilon$, $\varepsilon>0$ at some

$$0<x_{\max}(\rho)<x^*(\rho^*)<x_{\min}(\rho)<1.$$

This follows again from property P1 in combination with (3.5.10) and due to

$$\partial_x\mathscr{C}(x^*;\rho^*+\varepsilon)<0$$

for any $\varepsilon>0$.

P3) The discriminant of $\mathscr{P}(x,\rho)$, denoted by $D(\rho)$, is independent of x. It is itself a polynomial in ρ with degree 2 and its roots are given by

$$\rho_1=\frac{-(2\gamma+2)+4\gamma^{1/2}}{2\gamma-1}(1-\varsigma)^2,\quad \rho_2=\frac{-(2\gamma+2)-4\gamma^{1/2}}{2\gamma-1}(1-\varsigma)^2,$$

where ρ_2 denotes the larger root. We observe that $x_1=\mathscr{B}(\gamma)\in(1/2,1)$ is the unique solution of $\mathscr{P}(x,\rho_2)=0$ and therefore also of $\partial_x\mathscr{C}(x;\rho_2)=0$. (Note that the properties of $\mathscr{B}(\gamma)$, stated in this Theorem 3.21, follow by an application of l'Hôpital's rule.) Furthermore, notice that there are no critical points of $\mathscr{C}(x;\rho)$ for $\rho\in[0,\rho_2)$: P2 implies that $D(\rho)$ must be positive for $\rho\in(\rho_2,\infty)$, negative for $\rho\in(\rho_1,\rho_2)$ and again positive for $\rho\in[0,\rho_1)$. Hence, $\partial_x\mathscr{C}(x;\rho)=0$ does not have a solution for $\rho\in(\rho_1,\rho_2)$. By P2, $\partial_x\mathscr{C}(x;\rho)=0$ for some $\rho\in[0,\rho_1]$ implies the existence of a local maximum for some $\rho\in(\rho_1,\rho_2)$ which is a contradiction. Hence, $\partial_x\mathscr{C}(x;\rho)=0$ does not have a solution for $\rho\in[0,\rho_1)$ either.

In conclusion, combining P1-P3, we have no critical points of $\mathscr{C}(x;\rho)$ (restricted to $x\in(0,1)$) for $\rho<\rho_1$ and exactly two critical points - a strict local maximum and a strict local minimum - at some

$$0<x_{\max}(\rho)<\mathscr{B}(\gamma)<x_{\min}(\rho)<1$$

for $\rho\in(\rho_2,\infty)$. Based on this, we observe the following two cases that finish our proof:

Case 1: Let $\lceil n/2\rceil/n<\varsigma\leq\mathscr{B}(\gamma)$ and $\rho_a>0$. If $\mathscr{C}(x;\rho_a)=(>)\mathscr{C}(\varsigma;\rho_a)$ for some $1/2\leq x<\varsigma-1/n$, $x\in I_n$, then $\mathscr{C}(\varsigma-1/n;\rho_a)\geq(>)\mathscr{C}(\varsigma;\rho_a)$ holds true since we have an isolated saddle point at $\mathscr{B}(\gamma)$ and since $x_{\min}>\mathscr{B}(\gamma)$. If $\mathscr{C}(\varsigma-1/n;\rho_a)<\mathscr{C}(\varsigma;\rho_a)$, then we have $\mathscr{C}(x;\rho_a)<\mathscr{C}(\varsigma;\rho_a)$ for all $1/2\leq x<\varsigma-1/n$, $x\in I_n$.

Case 2: Let $\mathscr{B}(\gamma)+2/n<\varsigma<1$ and $\rho_a>0$. If we have $\mathscr{C}(\varsigma-1/n;\rho_a)=\mathscr{C}(\varsigma;\rho_a)$ then we necessarily obtain $\mathscr{C}(\varsigma-2/n;\rho_a)>\mathscr{C}(\varsigma-1/n;\rho_a)$ since we have an isolated saddle point at $\mathscr{B}(\gamma)$ and $x_{\min}>\mathscr{B}(\gamma)$. □

Proof of Proposition 3.22. We present an extended version of the proof of Proposition 2.18 of Torgovitski (2015d) in a time-rescaled setting. The case of $\theta \in (1/2, 3/4]$ is already shown in (3.3.23) in Theorem 3.21 and we continue with $\theta \in (3/4, 1)$. First, we define

$$\rho_\star(s) := \inf_{x \in (1-\varsigma, \varsigma)} \mathscr{R}(x, \varsigma) = \inf_{x \in [1/2, \varsigma)} \mathscr{R}(x, \varsigma) \leq \inf_{\substack{1/2 < x < \varsigma, \\ x \in I_n}} R(x, \varsigma) = \rho_{\max}(s).$$

Notice that the function $\mathscr{R}(x, \varsigma)$ is well-defined and differentiable on $x \in (1-\varsigma, \varsigma)$ and that $\mathscr{R}(x, \varsigma) < \mathscr{R}(1-x, \varsigma)$ for $x \in (1/2, \varsigma)$ which yields the second equality. In the following we assume that n is sufficiently large such that $\mathscr{B}(\gamma) + 2/n < \varsigma$ holds true in which case we know from previous considerations that $\mathscr{R}(x, \varsigma)$ must have a (strict) local minimum within $[1/2, \varsigma)$. It holds that

$$\partial_x \mathscr{R}(x, \varsigma) = -\frac{x(1-\varsigma)^2}{2} \left(\frac{\varsigma(1-\varsigma)}{x(1-x)} \right)^{1/2} \frac{f_1(x, \varsigma) - g_1(x, \varsigma)}{[f_2(x, \varsigma) - g_2(x, \varsigma)]^2} \tag{3.5.11}$$

with

$$\begin{aligned}
f_1(x, \varsigma) &= 2x[x(1-x)]^{1/2}[\varsigma(1-\varsigma)]^{1/2}, \\
g_1(x, \varsigma) &= 2\varsigma x^2 + \varsigma^2 - 3\varsigma x, \\
f_2(x, \varsigma) &= x(1-x)[\varsigma(1-\varsigma)]^{1/2}, \\
g_2(x, \varsigma) &= \varsigma(1-\varsigma)[x(1-x)]^{1/2},
\end{aligned}$$

defined on $x \in (1-\varsigma, \varsigma)$. Notice that $f_2(x, \varsigma) = g_2(x, \varsigma)$ is equivalent to

$$x(1-x) = \varsigma(1-\varsigma)$$

which is impossible for $x \in (1-\varsigma, \varsigma)$, i.e. the above expression (3.5.11) is well-defined for all $x \in (1-\varsigma, \varsigma)$.

We have $\partial_x \mathscr{R}(x, \varsigma) = 0$ for $x \in (1-\varsigma, \varsigma)$ if and only if $f_1(x, \varsigma) = g_1(x, \varsigma)$ holds true. Further, we observe that

$$-\varsigma(4x^2 - 4x + \varsigma)(\varsigma - x)^2 = f_1^2(x, \varsigma) - g_1^2(x, \varsigma) = 0$$

has the solutions

$$x_{1,2} = [1 \pm (1-\varsigma)^{1/2}]/2.$$

Notice that $x_\star = [1 + (1-\varsigma)^{1/2}]/2$ is the only critical point that fulfills $x_\star \in [1/2, \varsigma)$. Hence, we conclude that this is necessarily the sought argument of the local minimum, i.e. $\rho_\star(s) = \mathscr{R}(x_\star, \varsigma)$, and see that $\rho_\star(s)$ converges to the right-hand side of (3.3.24) due to $\lim_{n\to\infty} \varsigma = \theta$ and due to the continuity of $\mathscr{R}(x, y)$. It holds that

$$\rho_{\max}(s) = \inf_{1/2 \leq x < \varsigma} R(x, \varsigma) = \inf_{1/2 \leq x < \varsigma} \mathscr{R}(x, \varsigma) = \min\{\mathscr{R}(\lfloor nx_\star \rfloor/n, \varsigma), \mathscr{R}(\lceil nx_\star \rceil/n, \varsigma)\}$$

since $x_\star$ is the only critical point of $\mathscr{R}(\lfloor nx_\star \rfloor/n, \varsigma)$ in $[1/2, \varsigma)$. By the mean value theorem, we obtain, as $n \to \infty$,

$$|\rho_{\max}(s) - \rho_\star(s)| = |\min\{\mathscr{R}(\lfloor nx_\star \rfloor/n, \varsigma), \mathscr{R}(\lceil nx_\star \rceil/n, \varsigma)\} - \mathscr{R}(x_\star, \varsigma)|$$

$$\leq \sup_{\substack{\delta\in[-1/n,1/n],\\ x_\star+\delta\in[1/2,\varsigma)}} \left|\mathscr{R}(x_\star+\delta,\varsigma)-\mathscr{R}(x_\star,\varsigma)\right|$$
$$\leq c \sup_{\substack{\delta\in[-1/n,1/n],\\ x_\star+\delta\in[1/2,\varsigma)}} \left|\partial_x\mathscr{R}(x_\star+\delta,\varsigma)\right|/n = o(1)$$

for some $c > 0$, which proves (3.3.24) for $\rho_{\max}(s)$, too. Finally, the smoothness properties follow on applying l'Hôpital's rule. (Notice that if we could solve $\partial_x\mathscr{R}(x,\varsigma) = 0$ for values of γ other than $1/4$, then this Proposition 3.22 may be extended to these cases, too.) □

Proof of Theorem 3.24 (as in Theorem 2.11 of Torgovitski, 2015d). The critical function $C(\cdot; s, \rho)$ is positive. It has a maximum at $x = s$ for all ratios $\rho > 0$ if and only if

$$1 \leq C(s; s, \rho)/C(x; s, \rho) \tag{3.5.12}$$

holds true for all $\rho > 0$ and all $x \in I_n$. This can only be fulfilled if

$$1 \leq \lim_{\rho\to\infty} C(s; s, \rho)/C(x; s, \rho) = \left[(w(s)V(s))/(w(x)V(x))\right]^2 \tag{3.5.13}$$

holds true for every x. Notice that the weights $w = cw_\star$ fulfill this constraint for any $c > 0$. Now, recall, that we require that w is positive. Hence, if $w = cw_\star$ does not hold with some $c > 0$, then

$$1 < (w(z)V(z))/(w(y)V(y))$$

must hold true for some $z \neq y$ which contradicts condition (3.5.13) for $s = y$ and $x = z$ and therefore also contradicts (3.5.12) for some $\rho > 0$. □

Proof of Theorem 3.25 (as given in Theorem 2.12 of Torgovitski, 2015d). The weights $w = cw_\star$ imply that the difference

$$C(s; s, \rho) - C(x; s, \rho) = \left[w(s)s(1-s)\right]^2 - \left[w(x)\min\{x, s\}(1-\max\{x, s\})\right]^2 \tag{3.5.14}$$

does not depend on ρ for all $x \neq s$. A unique maximum of $C(x; s, \rho)$ is obviously at $x = s$ if and only if the right-hand side of (3.5.14) is positive for all $x \neq s$ which in turn is the case if and only if

$$w(s)/w(x) > \begin{cases} x/s, & x < s, \\ (1-x)/(1-s), & x > s \end{cases}$$

holds true for all $x \neq s$, $x \in I_n$. For $x < s$ this is equivalent to

$$V(x)/x > V(s)/s \tag{3.5.15}$$

and, due to symmetry, for $x > s$ it is equivalent to

$$V(1-x)/(1-x) > V(1-s)/(1-s). \tag{3.5.16}$$

On the one hand, if $V(x)/x$ is strictly decreasing, then (3.5.15) and (3.5.16) follow for any combination of x and s. On the other hand, if (3.5.15) holds true for any s and for all $x < s$, then $V(x)/x$ is necessarily strictly decreasing which completes the proof. □

Proof of Proposition 3.28. It holds that

$$\begin{aligned}
\hat{\Sigma}_{j,k} &= \sum_{p=1}^{d} \big\langle (Y_{j,p} - \bar{Y}_{j,d}), (Y_{k,p} - \bar{Y}_{k,d}) \big\rangle / d, \\
&= \sum_{p=1}^{d} \big\langle (\varepsilon_{j,p} - \bar{\varepsilon}_{j,d}) + (\gamma_p - \bar{\gamma}_d)\zeta_j, (\varepsilon_{k,p} - \bar{\varepsilon}_{k,d}) + (\gamma_p - \bar{\gamma}_d)\zeta_k \big\rangle / d \\
&= \sum_{p=1}^{d} \big\langle (\varepsilon_{j,p} - \bar{\varepsilon}_{j,d}), (\varepsilon_{k,p} - \bar{\varepsilon}_{k,d}) \big\rangle / d + \langle \zeta_j, \zeta_k \rangle \Big[\sum_{p=1}^{d} (\gamma_p - \bar{\gamma}_d)^2 / d \Big] \\
&\quad + \big\langle \zeta_j, \sum_{p=1}^{d} (\gamma_p - \bar{\gamma}_d)(\varepsilon_{k,p} - \bar{\varepsilon}_{k,d}) / d \big\rangle + \big\langle \sum_{p=1}^{d} (\gamma_p - \bar{\gamma}_d)(\varepsilon_{j,p} - \bar{\varepsilon}_{j,d}) / d, \zeta_k \big\rangle, \\
&=: A_1 + A_2 + A_3 + A_4,
\end{aligned}$$

where $\bar{\gamma}_d = \sum_{p=1}^{d} \gamma_p / d$. Assumption N2 ensures that $\Sigma_{j,k} = A_1 + o_P(1)$ as $d \to \infty$. Hence, it remains to verify that the terms A_2, A_3 and A_4, which involve the common factors, are asymptotically negligible. It holds that

$$\sum_{p=1}^{d} (\gamma_p - \bar{\gamma}_d)^2 / d = \Big[\sum_{p=1}^{d} \gamma_p^2 / d \Big] - \Big[\sum_{p=1}^{d} \gamma_p / d \Big]^2 \tag{3.5.17}$$

and

$$\begin{aligned}
&\Big\| \sum_{p=1}^{d} (\gamma_p - \bar{\gamma}_d)(\varepsilon_{k,p} - \bar{\varepsilon}_{k,d}) / d \Big\| \\
&\le \sum_{p=1}^{d} \gamma_p \|\varepsilon_{k,p}\| / d + \Big[\sum_{p=1}^{d} \gamma_p / d \Big] \|\bar{\varepsilon}_{k,d}\| \\
&\le \Big[\sum_{p=1}^{d} \gamma_p^2 / d \Big]^{1/2} \Big[\sum_{p=1}^{d} \|\varepsilon_{k,p}\|^2 / d \Big]^{1/2} + \Big[\sum_{p=1}^{d} \gamma_p / d \Big] \|\bar{\varepsilon}_{k,d}\|.
\end{aligned} \tag{3.5.18}$$

A combination of Assumption CF together with Assumption N2 and Jensen's inequality yields that (3.5.17) and (3.5.18) are of order $o(1)$ and $o_P(1)$, as $d \to \infty$. An application of the Cauchy-Schwarz inequality, using the finiteness of the second moments of ζ_j, yields $A_2 + A_3 + A_4 = o_P(1)$, which finishes the proof. □

Proof of Proposition 3.34. We treat the three cases separately. The proofs for the first two cases were already shown in Torgovitski (2015d) whereas the proof of the third case has not been published previously.

Case 1: We assume that $\lambda \in [t_M, \infty)$ and that $\beta = 0$. Hence, condition (KKT1) is fulfilled. Condition (KKT2) directly translates to $t_M = \max_{1 \le i < n} \|\hat{c}_{i,\bullet}\| = \max_{1 \le i < n} \|(\bar{D}_{\bullet,i})' \bar{Y}\| \le \lambda$ and is also satisfied.

Case 2: We assume $\lambda \in (t_m, t_M)$ and that β is set according to (3.3.33). We rearrange (KKT1) and observe that it is equivalent to

$$\begin{aligned}
\hat{c}_{M,\bullet} &= \lambda \beta_{M,\bullet} / \|\beta_{M,\bullet}\| + (\bar{D}_{\bullet,M})' \bar{D} \beta_{M,\bullet} \\
&= \big(\lambda / \|\beta_{M,\bullet}\| + (\bar{D}_{\bullet,M})' \bar{D}_{\bullet,M} \big) \beta_{M,\bullet},
\end{aligned} \tag{3.5.19}$$

which is satisfied since $\|\beta_{M,\bullet}\| = (t_M - \lambda)/((\bar{D}_{\bullet,M})'\bar{D}_{\bullet,M})$ holds true. Next, we observe that (KKT2) is fulfilled whenever it holds that

$$t_m = \max_{i \neq M} \|(\bar{D}_{\bullet,i})'\bar{Y}\| \leq \lambda - \|\bar{D}\beta\|,$$

which is the case for all $\lambda \in (\lambda_{\min}, t_M)$ with some $\lambda_{\min} \in (t_m, t_M)$ since $\beta \to 0$ as $\lambda \uparrow t_M$.

Case 3: We assume that $\lambda \in (0, t_m)$. Furthermore, we may assume without loss of generality that $m < M$. Proceeding as under (3.5.19) we observe that β has to be defined according to (3.3.33) to fulfill (KKT1). This contradicts condition (KKT2) which may be seen as follows: assumption $\hat{c}_{m,\bullet} = \hat{c}_{M,\bullet}$ implies that

$$\begin{aligned}
\|(\bar{D}_{\bullet,m})'(\bar{Y} - \bar{D}\beta)\| &= \|(\bar{D}_{\bullet,m})'\bar{Y} - (\bar{D}_{\bullet,m})'(\bar{D}_{\bullet,M}\beta_{M,\bullet})\| \\
&= \|\hat{c}_{m,\bullet} - (\bar{D}_{\bullet,m})'\bar{D}_{\bullet,M}(\alpha_M \hat{c}_{M,\bullet})\| \\
&= |1 - \alpha_M(\bar{D}_{\bullet,m})'\bar{D}_{\bullet,M}|\|\hat{c}_{M,\bullet}\| \\
&= |1 - \alpha_M(\bar{D}_{\bullet,m})'\bar{D}_{\bullet,M}|t_M
\end{aligned}$$

and (KKT2) requires that $\|(\bar{D}_{\bullet,m})'(\bar{Y} - \bar{D}\beta)\| \leq \lambda$. If λ is sufficiently close to t_M, this is equivalent to $\xi t_M \leq \xi\lambda$ with

$$\xi = 1 - (\bar{D}_{\bullet,m})'\bar{D}_{\bullet,M})/((\bar{D}_{\bullet,M})'\bar{D}_{\bullet,M}).$$

Thus, if $\xi > 0$, we end up with $t_M \leq \lambda$ which contradicts our previously made assumption of $\lambda < t_M$. $\square$

Now, Proposition 3.36 is a direct implication of Proposition 3.34.

Proof of Proposition 3.36 (as given in Proposition 2.4 of Torgovitski, 2015d). An evaluation of $\hat{c} = \bar{D}'\bar{Y}$ yields

$$t_i = \|\hat{c}_{i,\bullet}\| = w(i/n)\|\sum_{j=1}^{i}(Y_{j,\bullet} - \bar{Y}_{n,\bullet})\| = n^{1/2}w(i/n)\mathscr{T}(i/n)$$

for all $1 \leq i < n$ and the second statement of Proposition 3.34 yields the assertion. $\square$

Notes 3.6
(Relation of this chapter to previous publications and preprints)

The theory presented and developed within this chapter is based essentially on Torgovitski (2015d) and on the corresponding preprints Torgovitski (2015d, arXiv:1501.00177). As explained in the general introduction in Chapter 1 (and in Section 3.1 of this Chapter 3) the theory is refined and extended in manifold ways. Moreover, the presentation of the results is structured differently. For the sake of clarity, we summarize the relations between this chapter and the results in Torgovitski (2015d), below. Note that the proofs in this chapter are modifications and extensions (with notational adaptations) of the proofs presented in Torgovitski (2015d) and also in the corresponding previous arXiv.org versions Torgovitski (2015d, arXiv:1501.00177). (For this reason the proofs in Section 3.5 contain additional references, as well.)

Assumptions: Assumption N1, Assumption CF, Assumption N2 correspond to Assumptions 2.5, 2.6 and also to Assumptions of Section 2.2.4 in Torgovitski (2015d). Assumption U1 and Assumption U2 both correspond to Assumption A1 of Torgovitski (2015d).

Definitions: Definition 3.5 is a time-rescaled Hilbert space version of (2.11) of Torgovitski (2015d). (Cf. also Section 2.4 of Torgovitski, 2015d, arXiv:1501.00177v2.)

Theorems: Theorems 3.6, 3.21, 3.24 and 3.25 correspond to Theorems 2.9, 2.17, 2.11 and 2.12 of Torgovitski (2015d). Particularly, (3.3.23) is a new more compact representation of the limit (2.23) in Torgovitski (2015d).

Propositions: Propositions 3.22 and 3.28 correspond to Proposition 2.18 and the informally stated results of Section 2.2.4 in Torgovitski (2015d). Proposition 3.19 is new whereas Proposition 3.20 is related to Theorem 2.16 of Torgovitski (2015d). Proposition 3.34 extends Proposition 5.1 of Torgovitski (2015d) and Proposition 3.36 corresponds to Proposition 2.4 of Torgovitski (2015d).

Lemmas: Lemma 3.10 is a substantial extension of (2.27) of Torgovitski (2015d). Lemma 3.15 and Lemma 3.16 correspond to Lemma 2.13 and Remark 2.14 of Torgovitski (2015d).

Corollaries: Corollary 3.29 is new but implicitly contained as a discussion in Section 2.2.4 of Torgovitski (2015d).

Examples: Examples 3.12 and 3.13, for WN and MA(1) panels, correspond to Examples 2.7 and 2.8 of Torgovitski (2015d). Examples 3.14 and 3.18, for AR(1) and MA(q) panels, are entirely new. Furthermore, Counterexample 3.17 is new, too.

Remarks: Remarks 3.4, 3.9, 3.11, 3.26, 3.27, 3.31 and 3.38 are new. Remarks 3.3, 3.7, 3.23, 3.32 and 3.33 correspond to Footnote 7, to Remarks 2.10, 2.19, 2.20 and to Footnote 4 of Torgovitski (2015d). Remark 3.8 is related to Section 2.4 of Torgovitski (2015d, arXiv:1501.00177v2). Remarks 3.30 and 3.35 are implicitly contained in Section 2 of Torgovitski (2015d). Remark 3.37 refines the explanation in Section 2.2.1 of Torgovitski (2015d).

Notations: Notation 3.1 is borrowed from Torgovitski (2015d). Notation 3.2 is new.

Figures and Tables: Figure 3.1 and Figure 3.2 are new. Figure 3.3 and Figure 3.4 are new but related to Torgovitski (2015d, Figures 3 and 4). Finally, Table 3.1 is new.

BIBLIOGRAPHY

Andrews D.W.K. (1991)
Heteroskedasticity and autocorrelation consistent covariance matrix estimation.
Econometrica, 59(3):817–858

Antoch J., Hušková M., Prášková Z. (1997)
Effect of dependence on statistics for determination of change.
Journal of Statistical Planning and Inference, 60(2):291–310

Aston J.A.D., Kirch C. (2012a)
Detecting and estimating changes in dependent functional data.
Journal of Multivariate Analysis, 109:204–220

Aston J.A.D., Kirch C. (2012b)
Evaluating stationarity via change-point alternatives with applications to fMRI data.
Annals of Applied Statistics, 6(4):1906–1948

Aston J.A.D., Kirch C. (2014)
Change points in high dimensional settings.
arXiv:1409.1771v1, 1–34. Preprint, first version.

Aue A., Hörmann S., Horváth L., Hušková M. (2014)
Dependent functional linear models with applications to monitoring structural change.
Statistica Sinica: Supplement, 24(3):S1–S13

Aue A., Hörmann S., Horváth L., Reimherr M. (2009)
Break detection in the covariance structure of multivariate time series models.
Annals of Statistics, 37(6B):4046–4087

Aue A., Horváth L. (2013)
Structural breaks in time series.
Journal of Time Series Analysis, 34(1):1–16

Aue A., Rice G., Sönmez O. (2015)
Dating structural breaks in functional data without dimension reduction.
arXiv:1511.04020v1, 1–30. Preprint, first version.

Bai J. (2010)
Common breaks in means and variances for panel data.
Journal of Econometrics, 157(1):78–92

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(5):927–946

Berkes I., Gombay E., Horváth L. (2009)
Testing for changes in the covariance structure of linear processes.
Journal of Statistical Planning and Inference, 139(6):2044–2063

Berkes I., Hörmann S., Schauer J. (2011)
Split invariance principles for stationary processes.
Annals of Probability, 39(6):2441–2473

Berkes I., Horváth L., Rice G. (2013)
Weak invariance principles for sums of dependent random functions.
Stochastic Processes and their Applications, 123(2):385–403

Berkes I., Horváth L., Rice G. (2016)
On the asymptotic normality of kernel estimators of the long run covariance of functional time series.
Journal of Multivariate Analysis, 144:150–175.

Berkes I., Liu W., Wu W.B. (2014)
Komlós-Major-Tusnády approximation under dependence.
Annals of Probability, 42(2):794–817

Berkes I., Philipp W. (1979)
Approximation theorems for independent and weakly dependent random vectors.
Annals of Probability, 7(1):29–54

Bleakley K., Vert J.-P. (2010)
Fast detection of multiple change-points shared by many signals using group LARS.
Advances in Neural Information Processing Systems, 23:2343–2351

Bleakley K., Vert J.-P. (2011a)
The group fused Lasso for multiple change-point detection.
arXiv:1106.4199v1, 1–25. Preprint and an extension of Bleakley & Vert (2010).
(Available as a technical report: HAL-00602121, 1–25)

Bosq D. (2000)
Linear Processes in Function Spaces: Theory and Applications.
Lecture Notes in Statistics. Springer, New York.

Brockwell P.J., Davis R.A. (1991)
Time Series: Theory and Methods.
Springer Series in Statistics. Springer, New York.

Brodsky B., Darkhovsky B. (1993)
Nonparametric Methods in Change Point Problems.
Dordrecht, Kluwer.

Burton R.M., Goulet M., Meester R. (1993)
On 1-dependent processes and k-block factors.
Annals of Probability, 21(4):2157–2168

Chochola O., Hušková M., Prášková Z., Steinebach J.G. (2013)
Robust monitoring of CAPM portfolio betas.
Journal of Multivariate Analysis, 115:374–395

Choudhury A.H., Hubata R., St. Louis R.D. (1999).
Understanding time-series regression estimators.
American Statistician, 53(4):342–348

Csörgő M., Horváth L. (1993)
Weighted Approximations in Probability and Statistics.
Wiley, Chichester.

Csörgő M., Horváth L. (1997)
Limit Theorems in Change-Point Analysis.
Wiley, Chichester.

Csörgő M., Révész P. (1981)
Strong Approximations in Probability and Statistics.
Academic Press, New York.

Cuevas A. (2014)
A partial overview of the theory of statistics with functional data.
Journal of Statistical Planning and Inference, 147:1–23

Darling D.A., Erdős P. (1956)
A limit theorem for the maximum of normalized sums of independent random variables.
Duke Mathematical Journal, 23(1):143–155

Dauxois J., Pousse A., Romain Y. (1982)
Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference.
Journal of Multivariate Analysis, 12(1):136–154

Davidson J. (1994)
Stochastic Limit Theory: An Introduction for Econometricians.
Oxford university press, New York.

Davis R.A., Wu R. (2010)
Least absolute deviation estimation for general autoregressive moving average time-series models.
Journal of Time Series Analysis, 31(2):98–112

Dedecker J., Doukhan P., Lang G., León J.R., Louhichi S., Prieur C. (2007)
Weak Dependence: With Examples and Applications.
Lecture Notes in Statistics. Springer, New York.

Dedecker J., Merlevède F. (2008)
Convergence rates in the law of large numbers for Banach-valued dependent variables.
Theory of Probability and Its Applications, 52(3):416–438

Doukhan P., Oppenheim G., Taqqu M.S. (2003)
Theory and Applications of Long-Range Dependence.
Birkhäuser, Basel.

Eberlein E. (1986)
On strong invariance principles under dependence assumptions.
Annals of Probability, 14(1):260–270

Fazekas I., Klesov O. (2001)
A general approach to the strong law of large numbers.
Theory of Probability and Its Applications, 45(3):436–449

Ferraty F., Vieu P. (2006)
Nonparametric Functional Data Analysis: Theory and Practice.
Springer, New York.

Fortet R., Mourier E. (1955)
Les fonctions aléatoires comme éléments aléatoires dans les espace de Banach.
Studia Mathematica, 15(1):62–79

Frick K., Munk A., Sieling H. (2014)
Multiscale change point inference.
Journal of the Royal Statistical Society: Series B, 76(3):495–580

Gombay E. (2010)
Change detection in linear regression with time series errors.
Canadian Journal of Statistics, 38(1):65–79

Gombay E., Horváth L. (1996)
On the rate of approximations for maximum likelihood tests in change-point models.
Journal of Multivariate Analysis, 56(1):120–152

Hadri K., Larsson R., Rao Y. (2012)
Testing for stationarity with break in panels where the time dimension is finite.
Bulletin of Economic Research, 64:s123–s148

Harchaoui Z., Lévy-Leduc C. (2008)
Catching change-points with Lasso.
Advances in Neural Information Processing Systems, 20:617–624

Hoeffding W., Robbins H. (1948)
The central limit theorem for dependent random variables.
Duke Mathematical Journal, 15(3):773–780

Hörmann S. (2009)
Berry-Esseen bounds for econometric time series.
Latin American Journal of Probability and Mathematical Statistics, 6:37–397

Hörmann S., Horváth L., Reeder R. (2013)
A functional version of the ARCH model.
Econometric Theory, 29(2):267–288

Hörmann S., Kidziński Ł. (2015)
A note on estimation in Hilbertian linear models.
Scandinavian Journal of Statistics, 42(1):43–62

Hörmann S., Kidziński Ł., Hallin M. (2015)
Dynamic functional principal components.
Journal of the Royal Statistical Society: Series B, 77(2):319–348
(All preprint versions are arXiv:1210.7192v1, arXiv:1210.7192v2, arXiv:1210.7192v3, arXiv:1210.7192v4 and arXiv:1210.7192v5)

Hörmann S., Kokoszka P. (2010)
Weakly dependent functional data.
Annals of Statistics, 38(3):1845–1884

Horváth L. (1993)
The maximum likelihood method for testing changes in the parameters of normal observations.
Annals of Statistics, 21(2):671–680

Horváth L., Hušková M. (2012)
Change-point detection in panel data.
Journal of Time Series Analysis, 33(4):631–648

Horváth L., Hušková M., Rice G., Wang J. (2016)
Asymptotic properties of the CUSUM estimator for the time of change in linear panel data models.
Econometric Theory, 1–47. Forthcoming.

Horváth L., Kokoszka P. (2012)
Inference for Functional Data with Applications.
Springer Series in Statistics. Springer, New York.

Horváth L., Kokoszka P., Reeder R. (2013)
Estimation of the mean of functional time series and a two-sample problem.
Journal of the Royal Statistical Society: Series B, 75(1):103–122
(A preprint is available on arXiv:1105.0019v1.)

Horváth L., Kokoszka P., Rice G. (2014)
Testing stationarity of functional time series.
Journal of Econometrics, 179(1):66–82

Horváth L., Kokoszka P., Steinebach J.G. (1999)
Testing for changes in multivariate dependent observations with an application to temperature changes.
Journal of Multivariate Analysis, 68(1):96–119

Horváth L., Rice G. (2014)
Extensions of some classical methods in change point analysis.
TEST, 23(2):219–255

Horváth L., Rice G. (2015)
Testing equality of means when the observations are from functional time series.
Journal of Time Series Analysis, 36(1):84–108

Horváth L., Rice G., Whipple S. (2014)
Adaptive bandwidth selection in the long run covariance estimator of functional time series.
Computational Statistics and Data Analysis, 1–18. Forthcoming.

Hsing T., Eubank R. (2015)
Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators.
Wiley, Chichester.

Jandhyala V., Fotopoulos S., MacNeill I., Liu P. (2013)
Inference for single and multiple change-points in time series.
Journal of Time Series Analysis, 1–24. Forthcoming.

Jirak M. (2012)
Change-point analysis in increasing dimension.
Journal of Multivariate Analysis, 111:136–159

Jirak M. (2013)
On weak invariance principles for sums of dependent random functionals.
Statistics and Probability Letters, 83(10):2291–2296

Johnstone I.M., Titterington D.M. (2009)
Statistical challenges of high-dimensional data.
Philosophical Transactions of the Royal Society A, 367(1906):4237–4253

Joseph L., Wolfson D.B. (1992)
Estimation in multi-path change-point problems.
Communications in Statistics - Theory and Methods, 21(4):897–913

Joseph L., Wolfson D.B. (1993)
Maximum likelihood estimation in the multi-path change-point problem.
Annals of the Institute of Statistical Mathematics, 45(3):511–530

Kamgaing J.T., Kirch C. (2016)
Detection of change points in discrete valued time series.
In: Handbook of Discrete Valued Time series. Handbooks of Modern Statistical Methods. Chapman & Hall/CRC.

Kim D. (2014)
Common breaks in time trends for large panel data with a factor structure.
Econometrics Journal, 17(3):301–337

Kirch C. (2006)
Resampling methods for the change analysis of dependent data.
PhD Thesis, University of Cologne.

Kounias E.G., Weng T.-S. (1969)
An inequality and almost sure convergence.
Annals of Mathematical Statistics, 40(3):1091–1093

Ledoux M., Talagrand M. (1991)
Probability in Banach Spaces: Isoperimetry and Processes.
Springer, Berlin.

Ling S. (2007)
Testing for change points in time series models and limiting theorems for NED sequences.
Annals of Statistics, 35(3):1213–1237

Lifshits M.A. (1984)
Absolute continuity of functionals of supremum type for Gaussian processes.
Journal of Soviet Mathematics, 27(5):3103–3112

Mas A. (2008)
Local functional principal component analysis.
Complex Analysis and Operator Theory, 2(1):135–167

Móricz F. (1976)
Moment inequalities and the strong laws of large numbers.
Probability Theory and Related Fields, 35(4):299–314

Mourier E. (1956)
L-random Elements and L^*-random Elements in Banach Spaces.
Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, University of California Press, 231–242

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

Page E.S. (1955)
A test for a change in a parameter occurring at an unknown point.
Biometrika, 42(3-4):523–527

Panaretos V.M., Tavakoli S. (2013)
Cramér–Karhunen-Loève representation and harmonic principal component analysis of functional time series.
Stochastic Processes and their Applications, 123(7):2779–2807

Parzen E. (1957)
On consistent estimates of the spectrum of a stationary time series.
Annals of Mathematical Statistics, 28(2):329–348

Peštová B., Pešta M. (2015)
Testing structural changes in panel data with small fixed panel size and bootstrap.
Metrika, 78(6):665–689.
(Online version published in 2014.)

Peštová B., Pešta M. (2016)
Erratum to: Testing structural changes in panel data with small fixed panel size and bootstrap.
Metrika, 79(2):237–238.
Corrected version of Peštová & Pešta (2015) is on arXiv:1509.01291v1, 1–23

Piterbarg V.I. (2012)
Asymptotic Methods in the Theory of Gaussian Processes and Fields.
American Mathematical Society, Providence.

Pötscher B.M., Prucha I.R. (1997)
Dynamic Nonlinear Econometric Models: Asymptotic Theory.
Springer, Berlin.

Politis D.N., Romano J.P. (1994)
Bias-corrected nonparametric spectral estimation II.
Department of Statistics, Purdue University, 94(5):1–55. Technical report.

Ramsay J.O., Hooker G., Graves S. (2009)
Functional Data Analysis with R and MATLAB.
Springer, New York.

Ramsay J.O., Silverman B.W. (2005)
Functional Data Analysis.
Springer, New York.

Reimherr M. (2015)
Functional regression with repeated eigenvalues.
Statistics and Probability Letters, 107:62–70

Schmitz A. (2011)
Limit theorems in change-point analysis for dependent data.
PhD Thesis, University of Cologne.

Shumway R.H., Stoffer D.S. (2011)
Time Series Analysis and Its Applications. With R Examples. Third edition.
Springer Texts in Statistics. Springer, New York.

Steinebach J.G. (1995)
Variance estimation based on invariance principles.
Statistics, 27(1-2):15–25

Sutradhar B.C., Kumar P. (2003)
The inversion of correlation matrix for MA(1) process.
Applied Mathematics Letters, 16(3):317–321

Tavakoli S. (2014)
Fourier analysis of functional time series, with applications to DNA dynamics.
Thèse EPFL, n°6320.

Torgovitski L. (2015a)
A Darling-Erdős-type CUSUM-procedure for functional data.
Metrika, 78(1):1–27.

Torgovitski L. (2015b)
Algorithm for overlapping estimation of common change-sets in spatial data of fixed size.
arXiv:1502.04890v1, 1–17, Preprint, first version.
(All preprint versions are arXiv:1502.04890 and arXiv:1502.04890v1.)

Torgovitski L. (2015c)
Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components.
arXiv:1509.07409v2, 1–16, Preprint, second version.
(All preprint versions are arXiv:1509.07409, arXiv:1509.07409v1, arXiv:1509.07409v2.)

Torgovitski L. (2015d)
Panel data segmentation under finite time horizon.
Journal of Statistical Planning and Inference, 167:69–89.
Online version published in June 2015. Printed version appeared in December 2015.
(All preprint versions are arXiv:1501.00177, arXiv:1501.00177v1, arXiv:1501.00177v2 and arXiv:1501.00177v3.)

Torgovitski L. (2016)
A Darling-Erdős-type CUSUM-procedure for functional data II.
arXiv:1407.3625v2, 1–35, Preprint, second version.
Submitted in March 2016 to the Electronic Journal of Statistics, (EJS).
(All preprint versions are arXiv:1407.3625, arXiv:1407.3625v1 and arXiv:1407.3625v2.)

Tsirelson B.S. (1976)
The density of the distribution of the maximum of a Gaussian process.
Theory of Probability and Its Applications, 20(4):847–856

Varga R.S. (2000)
Matrix Iterative Analysis.
Springer Series in Computational Mathematics. Springer, Berlin.

Vostrikova L. (1981)
Detection of a "disorder" in a Wiener process.
Theory of Probability and Its Applications, 26(2):356–362

Wang J.-L., Chiou J.-M., Müller H.-G. (2015)
Review of functional data analysis.
arXiv:1507.05135v1, 1–47. Preprint, first version.

Wu W.B. (2011)
Asymptotic theory for stationary processes.
Statistics and Its Interface, 4(2):207–226

Zhou J. (2011)
Maximum likelihood ratio test for the stability of sequence of Gaussian random processes.
Computational Statistics and Data Analysis, 55(6):2114–2127

“Всё переплетено ...

ERKLÄRUNG

Ich versichere, dass ich die von mir vorgelegte Dissertation selbstständig angefertigt, die benutzten Quellen und Hilfsmittel vollständig angegeben und die Stellen der Arbeit - einschließlich Tabellen, Karten und Abbildungen -, die anderen Werken im Wortlaut oder dem Sinn nach entnommen sind, in jedem Einzelfall als Entlehnung kenntlich gemacht habe; dass diese Dissertation noch keiner anderen Fakultät oder Universität zur Prüfung vorgelegen hat; dass sie - abgesehen von unten angegebenen Teilpublikationen - noch nicht veröffentlicht worden ist sowie, dass ich eine solche Veröffentlichung vor Abschluss des Promotionsverfahrens nicht vornehmen werde. Die Bestimmungen der Promotionsordnung sind mir bekannt. Die von mir vorgelegte Dissertation ist von Prof. i.R. Dr. Josef G. Steinebach betreut worden.

Köln, im Dezember 2016

(Leonid Torgovitski)

Teilpublikationen

Torgovitski L. (2015a) . . . .
A Darling-Erdős-type CUSUM-procedure for functional data. . . . .
Metrika, 78(1):1–27. . . . .

Torgovitski L. (2015b) . . . .
Algorithm for overlapping estimation of common change-sets in spatial data of fixed size.
arXiv:1502.04890v1, 1–17, Preprint, first version. . . . .
(All preprint versions are arXiv:1502.04890 and arXiv:1502.04890v1.) . . . .

Torgovitski L. (2015c) . . . .
Detecting changes in Hilbert space data based on "repeated" and change-aligned principal components. . . . .
arXiv:1509.07409v2, 1–16, Preprint, second version. . . . .
(All preprint versions are arXiv:1509.07409, arXiv:1509.07409v1, arXiv:1509.07409v2.)

Torgovitski L. (2015d) . . . .
Panel data segmentation under finite time horizon. . . . .
Journal of Statistical Planning and Inference, 167:69–89. . . . .
Online version published in June 2015. Printed version appeared in December 2015. . . .
(All preprint versions are arXiv:1501.00177, arXiv:1501.00177v1, arXiv:1501.00177v2 and arXiv:1501.00177v3.) . . . .

Torgovitski L. (2016) . . . .
A Darling-Erdős-type CUSUM-procedure for functional data II. . . . .
arXiv:1407.3625v2, 1–35, Preprint, second version. . . . .
Submitted in March 2016 to the Electronic Journal of Statistics (EJS) and is currently under a major revision. . . . .
(All preprint versions are arXiv:1407.3625, arXiv:1407.3625v1 and arXiv:1407.3625v2.)