

Advancements in structural break testing

Citation for published version (APA):

Blatt, D. P. (2018). *Advancements in structural break testing*. [Doctoral Thesis, Maastricht University]. Datawyse / Universitaire Pers Maastricht. <https://doi.org/10.26481/dis.20180621db>

Document status and date:

Published: 01/01/2018

DOI:

[10.26481/dis.20180621db](https://doi.org/10.26481/dis.20180621db)

Document Version:

Publisher's PDF, also known as Version of record

Please check the document version of this publication:

- A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
- The final author version and the galley proof are versions of the publication after peer review.
- The final published version features the final layout of the paper including the volume, issue and page numbers.

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license above, please follow below link for the End User Agreement:

www.umlib.nl/taverne-license

Take down policy

If you believe that this document breaches copyright please contact us at:

repository@maastrichtuniversity.nl

providing details and we will investigate your claim.

Advancements in structural break testing

Dominik Blatt

© Dominik Blatt, Maastricht 2018

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted in any form, or by any means, electronic, mechanical, photocopying, recording or otherwise, without the prior permission in writing from the author.

The book was typeset by the author using L^AT_EX.

Published by Datawyse - Universitaire Pers Maastricht



Advancements in structural break testing

DISSERTATION

to obtain the degree of Doctor at Maastricht University,
on the authority of the Rector Magnificus,
Prof. dr. Rianne M. Letschert
in accordance with the decision of the Board of Deans,
to be defended in public
on Thursday 21 June 2018, at 12:00 hours

by

Dominik Peter Blatt

First Promoter

Prof. dr. Bertrand Candelon

Second Promoter

Prof. dr. Hans Manner

Assessment Committee

Prof. dr. Clemens Kool (chair)

Dr. Kausik Chaudhuri, Leeds University

Prof. dr. Alain Hecq

Prof. dr. Dominik Wied, Universität zu Köln

Why do you dislike it so when you can't say why?

—I don't know.

This mystery of life is a holy thing. It is our freedom. We have shot out of physical reality, we exist now in a kind of godlike freedom, and the mystery is integral to it.

—No. We are still physical reality. Atoms in their rounds. Determined on most scales, random on some others.

Ah well. We disagree. But either way, the scientist's job is to explore everything. No matter the difficulties! To stay open, to accept ambiguity. To attempt to fuse with the object of knowledge. To admit that there are values shot through the whole enterprise. To love it. To work toward discovering the values by which we should live. To work to enact those values in the world. To explore—and more than that—to create!

—I'll have to think about that.

Kim Stanley Robinson, *Green Mars*

Acknowledgements

Thank you to everyone who has helped me write this dissertation.

Bertrand Candelon, thank you for your guidance, patience and insight, and for making this amazing experience possible. Hans Manner, thank you for your support and the effort you have put into my academic progress. You had warned me that writing a Doctor thesis is a test of endurance. Thank you to the assessment committee, Clemens Kool, Kausik Chaudhuri, Alain Hecq and Dominik Wied, for reading and evaluating the thesis.

My family, friends and colleagues, I am very grateful for your inspiration, for your support and your criticism, and the joy we shared during this time. You have contributed a lot to the making of this thesis.

A special thank you to Frederik Daemen, Marlis Blatt-Homrich, Norbert Blatt, Tom Stolp, Jan Lohmeyer, Sylvia Beenen, David Behler, Sabine Blatt, Verena Blatt, Helge Braun, Christoph Deters, Seher Fazlioglu, Lennart Freitag, Peter Funk, Liudmila Galiullina, Mieke Jansen, Andrey Kateshov, Julian Keuthen, Hae-Ryun Kim, Felix Kortmann, Adam Kuszczinski, Hub Lentjes, Lenard Lieb, Svenja Mann, Maddy Meijers, Eric de Regt, Fleur Roemers, Marko Roos, Silvana de Sanctis, Benjamin Schmidt, Sebastian Schmidtko, Philipp Seegers, Christian Weyand, Carola Wießmann, Alexander Willms, Kika Wolff-Metternich, Wladimir Wolstein, and Patricia Zöller.

Contents

Acknowledgements	i
Contents	iii
Figures	viii
Tables	ix
1 Introduction	1
1.1 Objective	1
1.2 Preview of the results	4
2 Literature review	7
2.1 Structural break testing literature	7
2.2 Maximum likelihood ratio testing	9
2.2.1 Assumptions	9
2.2.2 Detecting multiple structural breaks	10
2.3 Bootstraps	13
3 Detecting contagion in a multivariate time series system	19
3.1 Introduction	19
3.2 Methodology	22
3.2.1 The multivariate model	22
3.2.2 Shift contagion in the multivariate model	22
3.2.3 Testing for common break dates	25
3.2.4 The spillover index	26
3.3 Empirical Study	27
3.3.1 Contagion during the European sovereign debt crisis	27
3.3.2 Spillovers vs shift contagion	32
3.4 Conclusion	34
3.A Appendix	35
4 A rolling window algorithm for detecting multiple structural breaks	43
4.1 Introduction	43
4.2 Methodology	46
4.2.1 The model	46
4.2.2 The rolling window algorithm	46

CONTENTS

4.2.3	Illustration	48
4.2.4	Testing power and method combination	49
4.3	Monte Carlo simulations	51
4.3.1	Introduction	51
4.3.2	General results	52
4.3.3	Execution time	57
4.3.4	Structural break estimation	59
4.4	Empirical Application	62
4.5	Conclusion	72
4.A	Appendix	73
5	Testing parameter change abruptness with a sum of likelihood ratios	77
5.1	Introduction	77
5.2	Methodology	79
5.2.1	The model	79
5.2.2	Structural break testing	80
5.2.3	Proposed non-abruptness test	81
5.2.4	Illustration of the test statistic	83
5.3	Monte Carlo simulations	86
5.3.1	Introduction	86
5.3.2	General results	87
5.3.3	Testing power	89
5.3.4	Impact of the summation range parameter	91
5.4	Empirical Application	93
5.5	Conclusion	98
5.A	Appendix	100
6	Detecting structural breaks by cross-validated error estimation	113
6.1	Introduction	113
6.2	Methodology	117
6.2.1	The model	117
6.2.2	Cross-validated error estimation	118
6.2.3	Structural breaks	119
6.2.4	Test statistic	120
6.3	Monte Carlo simulations	121
6.3.1	Introduction	121
6.3.2	Testing power	122
6.3.3	Break estimation	124
6.4	Empirical Application	127
6.5	Conclusion	135
6.A	Appendix	137
7	Summary	143
	Bibliography	145
	Valorisation addendum	159

Curriculum Vitae

163

CONTENTS

List of Figures

3.1	Spillover in the euro area	32
3.2	Directional spillover indexes for Greece	33
3.3	Directional spillover indexes for Austria	35
3.4	Directional spillover indexes for Belgium	38
3.5	Directional spillover indexes for Finland	38
3.6	Directional spillover indexes for France	39
3.7	Directional spillover indexes for Ireland	39
3.8	Directional spillover indexes for Italy	40
3.9	Directional spillover indexes for the Netherlands	40
3.10	Directional spillover indexes for Portugal	41
3.11	Directional spillover indexes for Spain	41
4.1	Illustration of the rolling window algorithm	49
4.2	Differences in break date estimation with shift reversal 1/2	56
4.3	Relative execution time of the rolling window algorithm	58
4.4	Differences in break date estimation for three upward shifts in one series	62
4.5	Residual plots and significant changes in volatility 1/2	65
4.6	Residual plots and significant changes in volatility 2/2	66
4.7	Differences in break date estimation with shift reversal 2/2	74
5.1	Visualization of two parameter transitions	83
5.2	Visualization of test results for two parameter transitions	85
5.3	Error term standard deviation over time 1/9	96
5.4	Error term standard deviation over time 2/9	104
5.5	Error term standard deviation over time 3/9	105
5.6	Error term standard deviation over time 4/9	106
5.7	Error term standard deviation over time 5/9	107
5.8	Error term standard deviation over time 6/9	108
5.9	Error term standard deviation over time 7/9	109
5.10	Error term standard deviation over time 8/9	110
5.11	Error term standard deviation over time 9/9	111
6.1	Absolute deviation in break date estimation	126
6.2	Break date estimates in prediction error repetitions, the Netherlands	130
6.3	Linear regressions for pre-, no-, and post-break data, part 1/6	133
6.4	Linear regressions for pre-, no-, and post-break data, part 2/6	134
6.5	Linear regressions for pre-, no-, and post-break data, part 3/6	139

LIST OF FIGURES

6.6 Linear regressions for pre-, no-, and post-break data, part 4/6 140
6.7 Linear regressions for pre-, no-, and post-break data, part 5/6 141
6.8 Linear regressions for pre-, no-, and post-break data, part 6/6 142

List of Tables

3.1	Example of a sequence of crisis events	23
3.2	Structural breaks for mean and standard deviation	28
3.3	Timeline of estimated structural breaks	30
3.4	Structural breaks for correlations part 1/2	36
3.5	Structural breaks for correlations part 2/2	37
4.1	General testing results	53
4.2	Mean execution time comparison	57
4.3	Structural break estimation comparison	60
4.4	RW test results for volatility changes 1/3	67
4.5	RW test results for volatility changes 2/3	68
4.6	RW test results for volatility changes 3/3	69
4.7	Full sample regression models	75
5.1	General testing results	88
5.2	Test results for different specifications of the DGP	89
5.3	The impact of ν on SLR power	91
5.4	The impact of ν on SLR power for additional DGP specifications	92
5.5	SLR test results for a moderate summation parameter value	94
5.6	Full SLR test results for Germany and Spain	98
5.7	Full SLR test results 1/3	101
5.8	Full SLR test results 2/3	102
5.9	Full SLR test results 3/3	103
6.1	Testing power for various DGP specifications	123
6.2	Break date estimation for various DGP specifications	124
6.3	Structural break estimation for European countries, lagged U.S. covariate	128
6.4	Estimation results for possible break dates, the Netherlands	132
6.5	The impact of CVE settings on testing	138
6.6	Parameter shift estimation	138

LIST OF TABLES

1

Introduction

1.1 Objective

The aim of this book is to introduce novel structural break test ideas and methods. In time series modeling, it is important to recognize that relations between variables may change over time. The adequacy of an estimated model may be increased, if its parameters are allowed to change values. A structural break test evaluates this possibility. It will be associated with a break date, which is sometimes also referred to as a change point in the literature. Parameters that may otherwise be invariant throughout the data will change their value at that date, if a structural break exists. Regardless of the simplicity or the complexity of a model's functional form, considering the existence of parameter changes increases the adaptability of a model to the given data. This makes structural break testing a convenient and powerful tool in time series analysis. The objective of this thesis is to contribute to this important field of research by presenting new ways by which structural break testing can be improved further.

First, Chapter 2 will give a compact overview of the current state of structural break testing. Chapter 3 and Chapter 4, each, describe the development of a structural break testing procedure that is recommended in a specific context. Chapter 3 describes an original, sequential testing procedure that is designed for the detection of financial contagion in time series data. The procedure anticipates various potential scenarios of shock transmission and incorporates appropriate structural break tests and test settings. The tests find structural breaks associated with contagion endogenously and determine whether the detected events can be assumed to have occurred simultaneously. The use of multivariate data systems, and joined break hypotheses, if appropriate, increases the efficiency of financial contagion detection.

The computational burden of analyzing structural breaks can be large, especially when multiple breaks are estimated, multiple times in an iterative sequence, in mul-

tivariate data sets. A time-efficient program for finding multiple structural breaks is presented in Chapter 4. Instead of locating all structural breaks at once, they are individually located in rolling data windows. The proposed algorithm is a heuristic for finding structural breaks that optimize the model fit, and so the time-efficiency comes at the expense of maximum testing power. Thus, the added value of the program design depends on the goals of the applicant, and whether the computational burden of multiple break detection tests is a concern. The reduction in estimation time can be drastic, if the data set is large, or if the break test must be performed frequently.

The procedures described in Chapters 3 and 4 apply likelihood ratio statistics in an a-posteriori test design. That is, for a definite data set, the statistic evaluates the fit of a model that includes abrupt changes in parameter values. Dates associated with maximum likelihood ratio values are break date estimates, if the test statistic is sufficiently large. Chapter 2 on the existing structural break testing literature provides an overview of all aspects of maximum likelihood ratio testing that are relevant for the remainder of this book. The motivation for another chapter of this book, Chapter 5, is derived from a discussion in Andrews (1993), which contains an original description of the asymptotic distribution of the maximum likelihood ratio statistic. The author acknowledges that the associated test functions as a test of model adequacy in a broader sense. It can detect abrupt parameter shifts, as well as more general forms of parameter change.

In Chapter 5, a new statistic is described that tests the abruptness of parameter shifts to facilitate (and strengthen) the interpretation of structural break estimates. Conveniently, it uses data that is already available during maximum likelihood ratio testing, but that is usually neglected: For maximum likelihood ratio testing, likelihood ratio values associated with dates other than the estimated break date are calculated, but usually not further analyzed. The chapter describes a test statistic that adds the values in a sum of likelihood ratios within a range of dates surrounding the break date estimate. The new statistic can uncover parameter transitions that are significantly more gradual than abrupt changes.

The final advancement in structural break testing is a novel break detection test statistic described in Chapter 6. The statistic is based on the estimation of prediction errors, sometimes referred to as the generalization errors of models, see for example Nadeau and Bengio (2000). The strategy is adopted from the more general topic of model selection and finding the optimal capacity of a model. Like other a-posteriori test statistics, the statistic evaluates the change in the fit of the model, when a change in parameters is permitted at a break date. Whereas the maximum likelihood ratio statistic measures the model fit to the available observations, the new statistic estimates the model fit to unavailable observations. A structural break is detected, when the resulting model is estimated to lead to significantly smaller errors with respect to the prediction of data. This way, the inference of structural breaks is directly based on a measure of the out-of-sample fit of a model.

This book seeks to enrich the existing structural break literature and stimulate future research by presenting novel testing ideas. The testing concepts have been developed with specific goals in mind. The sum of likelihood ratios of Chapter 5, for example, focuses the attention towards additional information contained in available likelihood ratio values, and so augments the analysis of parameter change. The test of Chapter 6 introduces a model selection measure to an established structural break test design, because the

statistic is more relevant to the external validity of the resulting model than conventional data fit statistics. The emphasis of the book lies on the introduction and illustration of the concepts and on proof of validity as shown by Monte Carlo simulation studies.

The Monte Carlo simulations are used to evaluate and compare the testing properties of the structural break testing methods. As a possibility for future work, descriptions of theoretical properties of the methods can be developed, but so far, no general distributions have been found. In the current structural break testing literature, it is not common to describe exact test statistic distributions, instead, asymptotic theory is used to describe limit distributions. In practice, for finite data applications, testing power is lost when asymptotically derived, approximate critical values are used, see Davidson and MacKinnon (2004). Therefore, Chapters 5 and 6 contain appropriate bootstrap methods that enable empirical, simulation-based estimations of test distributions in practice. The literature review Chapter 2 also contains an overview of relevant bootstrapping techniques used throughout the book.

Each of the Chapters 3–6 contains an empirical application for the purpose of illustration of the presented methods. The applications seek to uncover and interpret structural changes during the European sovereign debt crisis. At a relatively young age, the euro area has experienced diverse episodes of turmoil in its financial markets. The European sovereign debt crisis involved fierce, supranational debates on reported and unreported government debt levels, on existing and non-existing institutions of the European Central Bank, and on conventional and unconventional monetary and fiscal policy tools. The debt obligations of euro area governments, each seen as a safe investment after the introduction of the euro, quickly became a focus point of the debates. Frieden and Walter (2017) argue that other than in previous cases of debt and balance-of-payments crisis, countries with and without financial problems are joined in a monetary union that strongly limits the available policy options, and the burden of resolving the crisis has been largely put on debtor countries and tax payers of the euro area.

Reasons for the crisis have been thoroughly studied, such as in Copelovitch et al. (2016); in the euro area, macroeconomic asymmetry is paired with limited degrees of integration, which hinders a quick resolution of adverse shocks. In this context and given the rapidly increased government debt at the end of the Global Financial Crisis, a poor outcome for the euro area can be caused by lack of fiscal policy coordination, fragmented financial regulation and a non-credible no-bailout commitment. Jones et al. (2016) argue that the European Union is perceived to be in economic crisis constantly. They argue that the diversity in the union and the necessity to find common denominators cause a cycle of incomplete solutions to policy failure. While it is important to analyze problems that occur in Europe with theoretical reasoning, the development of econometric methods is equally important to detect concrete evidence for crisis.

Structural break tests can help detect market stress and market interdependence and are therefore important guides for policy making. To this end, the chapters of this book will illustrate how they can be applied to time series of daily government bond prices. Chapter 3 holistically studies the propagation of shocks between a set of euro area government bond yield spreads, looking for common spillover patterns between 2009 and 2014. Chapters 4–6 extend the analysis to the government bond yields of a larger set of 18 European countries. Structural break tests are set to detect possible changes to the extent that the variation in individual government bond yields cannot be explained

by other financial variables. Initially, in Chapters 4 and 5, the error term of each series model is investigated for multiple significant changes in the volatility that have either occurred abruptly or non-abruptly. For each country, a unique sequence of more and less volatile data segments is described. Using the final structural break test of the book, the linear relation of the individual series to U.S. government bond yields is studied in Chapter 6 to see how it may have changed in recent years.

1.2 Preview of the results

Chapter 2 begins with an overview of past and current contributions to the structural break testing literature. This provides some context for the structural break testing methods presented in Chapters 3–6. As the method of maximum likelihood ratio testing is used throughout this book, either in application or as a reference for comparison purposes, it is formally described in the second section of Chapter 2. The section contains technical assumptions, descriptions of the components of the test statistic, and a description of limit distributions. Further, Chapters 3, 5 and 6 make use of bootstrap methods, either to find confidence intervals of break date estimates, or for an empirical derivation of the distribution of a break test statistic. Therefore, the final section of Chapter 2 provides an overview of different bootstrap methods found in the literature that are relevant to the remainder of the book.

The sequential testing procedure of Chapter 3 is tailored to the detection of contagion, using a multivariate time series model and likelihood ratio statistics. The procedure includes separate tests for structural breaks in the conditional mean parameters of a model, in the volatility of the error term, and finally in the correlation coefficients of the error term. The iterative method re-estimates the three types of breaks, conditional on the previous results, until convergence. The procedure includes tests for the distinctiveness of the detected break dates. When applied to sovereign debt yield spreads of euro area countries, the detected volatility breaks are interpreted as separating periods of crisis from periods of relatively calm markets. An episode of particularly high volatility throughout the euro area starts in July 2011 and ends in October 2012. Shift contagion is identified by a significant increase of correlation coefficients between markets during a volatile data segment. Some incidents of shift contagion are found, but none between Greece and other countries. This is surprising, arguably, as the Greek government debt burden is typically a prominent feature in debates on the European sovereign debt crisis. A spillover analysis is performed to further study the transition of shocks over time. The analysis is enriched from the structural break results of the sequential testing procedure, as spillover indexes can be calculated for each data segment of the model. The total spillover index is high for the euro area as a whole, and spillover from the Greek debt market is particularly high during the crisis period between July 2011 and October 2012.

Chapter 4 presents a rolling window algorithm that can be used to detect multiple structural breaks. The algorithm is a heuristic for optimally positioning multiple breaks in time series data. The algorithm is only time-efficient, if at least two structural breaks are estimated. It does not locate structural breaks at the combination of dates that jointly maximize the likelihood ratio statistic. Instead, it locates structural breaks at the dates that individually produce the largest likelihood ratio values within given data

windows. Following a description and illustration, the method is evaluated using Monte Carlo simulations. The algorithm is shown to have slightly lower detection power than a global optimization program. Occasionally, the rolling window algorithm estimates break dates more efficiently, due to its local optimization approach. This happens, for example, when parameters shift in value multiple times in the same direction. Since the direction of parameter shifts is not known prior to testing, neither method is strictly better than the other based on the estimation results. A clear result of the simulation study is that the rolling window algorithm drastically reduces the execution time, when applied instead of the time-efficient global optimization program of Bai and Perron (2003).

Chapter 5 describes a sum of likelihood ratios statistic that uncovers non-abrupt parameter change. The statistic can be used in a follow-up test, after a structural break has been discovered. The test statistic produces large values, when the parameter change associated with a structural break happens gradually instead of abruptly. This property is observed in a Monte Carlo simulation study, in which the covariate coefficients of a linear regression model transition smoothly between values. The Monte Carlo study also confirms that the maximum likelihood ratio test typically finds a structural break in such data of gradual parameter change. This result supports the set-up of the sum of likelihood ratios as a test that follows maximum likelihood ratio testing. The application is convenient in practice, as all required data (likelihood ratio values of dates surrounding the break date estimate) has already been obtained during the original structural break testing. Although the test is developed using likelihood ratios, the summation concept can be applied to other break tests statistics.

An alternative structural break test statistic to the maximum likelihood ratio is presented in Chapter 6, using cross-validated error estimates. In the cross-validation method, mean squared error estimates are minimized to minimize the estimated prediction error that is associated with a break date candidate and the resulting model. To minimize the impact of each random assignment of validation sets, the estimation is repeated a large number of times with a random validation set composition in each. Each repetition is associated with a date that minimizes the estimated error. The median date among the dates can be taken as a point estimate of the break date, if the test statistic is sufficiently large. The overall distribution of preliminary break date estimates, and additional distribution measures, can be investigated further. A Monte Carlo study compares this cross-validated error estimation test to the maximum likelihood ratio test in simulated data with structural breaks in the mean regression parameters. Either method frequently detects break dates closer to the true location in the simulated data than the other method. Neither method is significantly better, but more often the cross-validated error test has more accurate estimation results than the maximum likelihood ratio test.

In Chapters 4, 5 and 6, structural breaks are detected and interpreted in models for long-term, European government bond yields between 2003-1-6 and 2016-08-03. Differences in the yields are regressed on other financial series, including lagged differences in U.S. government bond yields, to control for global risk factors. The rolling window method finds multiple structural breaks in the error term volatility, as described in Chapter 4. The structural breaks separate various episodes of the Global Financial Crisis and the European sovereign debt crisis, which can be studied for each country separately. The abruptness of the volatility changes is investigated in Chapter 5 by the application of the sum of likelihood ratios statistic. For some of the detected structural breaks, the

test rejects the hypothesis that change has happened abruptly, especially for the breaks leading up to the various stages of crisis in Europe. In contrast, abruptness of change is less frequently rejected in the later part of the sample.

The years between 2010 and 2016 have been characterized by several adjustments of unconventional monetary policy by the ECB. In Chapter 6 the quantitative easing program of 2015 is found to have left a significant mark on the relation between European and U.S. government bond prices, according to the cross-validated error test. For several central euro area countries, a structural break is detected in the coefficient of the U.S. government yield variable and the error term volatility. Breaks are estimated to have occurred in April 2015, shortly after first data had been published on the success of the program. The quantitative easing program, which directly involves government bond purchases, increased the alignment between the monetary policy measures of the euro area and the United States. Correspondingly, the linear association between several European bond yields and lagged differences in U.S. government bond yields is significantly larger after April 2015 than before.

2

Literature review

2.1 Structural break testing literature

This section provides a compact look at the field of structural break testing in general. The development of structural break tests goes back until the 1950's, when Hartley (1950) introduced a test that can identify variance differences between data groups, and when Page (1954) and Page (1955) offered a method to find changes that affect model parameters over time. In the decades of statistical research that followed, structural break tests have been increasingly generalized and extended, so that current tests can uncover a variety of structural instability in many settings. Whereas the test of Hartley (1950) is designed to uncover variance heterogeneity between equally sized, normally distributed groups of data, modern structural break tests can detect differences and changes in various parameter types, and in less restrictive model formulations.

Structural break tests can be designed as a-posteriori tests, which retrospectively find breaks in a complete data set, or they can be monitoring tests, where the data is updated sequentially. Some tests use estimated model parameters directly, others use the values of optimized objective functions. In contrast, nonparametric structural break tests make use of metrics that can be calculated in more general settings. Because the properties of the (unknown) underlying data distribution are unused in non-parametric tests, the increase in applicability comes at the cost of lower testing power for well-defined models, see Andrews (1993). The remainder of the book will focus on a-posteriori tests, so monitoring tests will only be introduced shortly in the following.

One monitoring test is presented in Page (1954) as a sequential process inspection scheme. The idea has been developed further into the standardized CUSUM test described in Brown et al. (1975). For this method, differences between observations and their respective forecasts are calculated, where the forecasts are derived from preceding data only. These recursive residuals are added up to form the cumulative sum (CUSUM)

statistic. A structural break that concerns the conditional mean model is found when the calculated cumulative sum deviates sufficiently from its expected value. The CUSUM approach is adopted for instability in more general model parameters in Lee et al. (2003). Gombay and Serban (2009) describe critical values for a generalization of the CUSUM test that is applicable for change in any collection of parameters of an autoregressive time series model. It is shown to exhibit good size even for long monitoring periods, but its power is negatively affected by nuisance parameters, see Lai (1995) for an extensive study of sequential break detection methods. Groen et al. (2011) find that the testing power of an adjusted CUSUM test is increased by using multivariate data systems, made possible by considering the maximum absolute value of individual CUSUM statistics at each date. The parametric equivalent to the CUSUM test is the fluctuation monitoring procedure, in which parameters are recursively estimated, as described in Chu et al. (1996) and analyzed by a simulation study in Carsoule and Franses (2003).

A pioneer in the strand of a-posteriori methods has been the Chow test. With this test the significance of a structural break at a given date can be determined, as described in Chow (1960). At the time, the author studied the demand for automobile ownership and the demand for new cars. He wanted to see whether two groups of economic units can be assumed to share the same regression parameters, and so developed a test for the equality of linear regression coefficients. Chow (1960) proposes the use of sums of squared residuals and sums of squared differences in fitted values, using full sample and subsample coefficient estimates. The resulting statistic can be used to infer a structural break in a time series, which separates a sample in two parts with significantly different model parameter values. For this test, the break date candidate that may separate the sample is exogenously given.

In the following years and decades, limitations of the Chow test were gradually addressed. Goldfeld and Quandt (1965) is another seminal publication in the field of break testing and contains a test for the equality of error term variances. The test statistic compares sums of squared residuals between two equal-sized samples. Both the test statistics of Chow (1960) and Goldfeld and Quandt (1965) assume normal linear regression models, while the structural break test of Ploberger et al. (1989) can be applied to dynamic regression models with stationary and ergodic errors. An important advancement in a-posteriori structural break testing has been the ability to detect structural breaks endogenously, such as in Hansen (1992), instead of testing for a structural break at a preselected date.

In endogenous structural break testing, the most likely breaks are located at those dates that cause the maximum data fit improvement among all candidate break dates. Determining the significance of these most likely breaks is different, because the maximum of a break test statistic has a larger expected value than the break test statistic itself, even if no structural break exists. To enable the detection of structural break dates endogenously, a likelihood ratio (LR) statistic is presented in Hinkley (1970) for linear regression models with normally distributed errors and mean parameter change. Similarly, a Wald statistic is presented in Hawkins (1987) for models with independent and identically distributed errors and a change in all model parameters at the break date. Andrews (1993) expands the applicability of endogenous break detection to models that can be non-linear, non-stationary and may contain temporally dependent data. Andrews (1993) describes the limit distribution of a group of asymptotically equivalent

tests, using either a maximum Wald statistic, a maximum Lagrange multiplier statistic, or a maximum LR statistic.

A large fraction of the subsequent structural break testing literature has further generalized the results of Andrews (1993). Bai and Perron (2003) describe the asymptotic distribution of the maximum LR statistic when used to find multiple structural breaks in a sample at once. More recently, Qu and Perron (2007) describe the asymptotic distribution of the maximum LR statistic, when the data system contains more than one dependent variable, and multiple structural breaks are tested in any subset of model parameters, including covariance terms of error distributions. The required assumptions are weak, and reproduced in Section 2.2, alongside the test statistics and limit distributions.

Other break test contributions concern methods and properties of testing that are important for specific situations. For example, Bataa et al. (2013) deals with conditional testing, when various types of parameters must be tested that are nuisance parameters to each other. Structural breaks that are not asymptotically distinct but locally ordered are covered in Perron and Oka (2011). Lin and Teräsvirta (1994) consider more general parameter change than the abrupt parameter change usually assumed in structural break testing. The authors describe a test within the smooth transition model framework, with which parameter stability can be rejected in favor of a smooth transition alternative hypothesis. Finally, it should be noted that structural break tests are a subcategory of the more general field of model diagnostics. For further reading on methods to evaluate the adequacy of fitted models, refer to Engle (1982) and Brockwell and Davis (2016).

2.2 Maximum likelihood ratio testing

Maximum likelihood ratios are used throughout this book as a basis for a-posteriori structural break testing. When calculating values of the maximum likelihood ratio statistic, usually the parameters of the multivariate Gaussian distribution function are estimated by maximum likelihood estimation. When data is not normal, both the corresponding limit distribution function described by Qu and Perron (2007), and empirical distributions derived from bootstrap methods can account for the non-normality. In the following, the maximum likelihood ratio statistic is formally described, as well as the limit distribution function of Qu and Perron (2007). Finally, the dynamic program by Bai and Perron (2003) is presented, which efficiently finds an optimal distribution of multiple break dates. At the beginning of this section, the required assumptions for the application of the maximum likelihood ratio test are given.

2.2.1 Assumptions

When using the maximum likelihood ratio statistic for multiple structural break testing, the following assumptions are made about the error term, as adopted from Qu and Perron (2007). The assumptions are formulated with respect to time series segments, where $\varepsilon_t \sim (0, \Sigma_m)$, for $[\lambda_{m-1}T] + 1 \leq t \leq [\lambda_m T]$, with data segments $\lambda_m = k_m/T$, $\lambda_0 = 0$, $\lambda_{M+1} = 1$, where k_m is the m -th break date in a time series of T observations. The assumptions cover the typical features observed in financial returns. In particular, the presence of conditional heteroscedasticity typically encountered in financial data is permitted, as long as the generating process is of short memory with bounded fourth

moments. Assumption A3 guarantees asymptotic distinctiveness of the segments between structural breaks.

- Assumption A1: Let $\mathcal{F}_t = \sigma - \text{field}\{\dots, \varepsilon_{t-2}, \varepsilon_{t-1}\}$. ε_t is weakly stationary within each segment and (a) $\{\varepsilon_t, \mathcal{F}_t\}$ forms a strongly mixing (α -mixing) sequence with size $-4r/(r-2)$ for some $8 > r > 2$, (b) $E(\varepsilon_t) = 0$ and $\sup_t \|\varepsilon_t\|_{2r+\delta} < N < \infty$ for some $\delta > 0$ and $N > 0$, where $\|X\|_r = (\sum_i \sum_m E|X_{i,m}|^r)^{1/r}$, for $r \geq 1$, is the L_r -norm of a random matrix X , (c) let $S_{c,m}(l) = \sum_{T_{m-1}^0+l+1}^{T_m^0+l+c} (\varepsilon_t \varepsilon_t')$, $m = 1, \dots, M$, for each $e \in \mathbb{R}^n$ of length 1, $\text{var}(\langle e, S_{c,l}(0) \rangle) \geq v(c)$ for some function $v(c) \rightarrow \infty$ as $c \rightarrow \infty$ (with $\langle \cdot, \cdot \rangle$, the usual inner product).
- Assumption A2: $\{\varepsilon_t \varepsilon_t' - \Sigma_m^0\}$ satisfies the conditions stated for ε_t in Assumption A1.
- Assumption A3: $0 < \lambda_1^0 < \dots < \lambda_M^0 < 1$, where $k_m^0 = [T\lambda_m^0]$, $m = 1, \dots, M+1$.

2.2.2 Detecting multiple structural breaks

Apart from an irrelevant constant term, the multivariate Gaussian log-likelihood function for n dependent variables $y_t = [y_{1,t}, \dots, y_{n,t}]'$ is given below, with a vector of $q+1$ covariates $x_t = [1, x_{1,t}, \dots, x_{q,1}]'$ and dates $t = 1, \dots, T$. For each date t , the set of model parameters θ_t contains a $(q+1) \times n$ -matrix of covariate coefficients, β_t , and an $n \times n$ -covariance matrix of the error term, Σ_t , so that

$$\begin{aligned} l(\theta_t) &= l(\beta_t, \Sigma_t) \\ &= -\frac{1}{2} \sum_{t=1}^T \log(|\Sigma_t|) - \frac{1}{2} \sum_{t=1}^T (y_t - x_t' \beta_t)' \Sigma_t^{-1} (y_t - x_t' \beta_t). \end{aligned}$$

For structural break testing, the likelihood ratio compares the maximized value of a restricted log-likelihood function to the maximized value of a less restricted log-likelihood function. Under the null hypothesis of parameter stability, the model is delimited by a set of only two ending dates, $\mathcal{T}_0 = \{0, T\}$. Accordingly, when maximizing the restricted log-likelihood function, the model parameters are constant throughout the sample, $\theta_t = \theta$, for $t = 1, \dots, T$. Let β denote the constant covariate coefficients and Σ the constant covariance matrix. Under the alternative hypothesis, M structural breaks exist that affect a non-empty subset of the model parameters θ_t^b , while other parameters may be unaffected. The structural break dates are included in the set of ending dates, $\mathcal{T}_M = \{0, k_1, \dots, k_M, T\}$. Accordingly, when maximizing the less restricted function, some parameters may still be constant throughout the sample, $\theta_t^c = \theta^c$, for $t = 1, \dots, T$. The other parameters will only be constant within each regime of the model. That is, $\theta_t^b = \theta_m$, for $t = k_{m-1} + 1, \dots, k_m$ and for $m = 1, \dots, M+1$, where $k_0 = 0$ and $k_{M+1} = T$. Let β_m denote the covariate coefficients and Σ_m the covariance matrix of segment m .

For a given set of ending dates \mathcal{T}_M , the likelihood ratio $LR(\mathcal{T}_M)$ for M structural breaks is computed from the following equations, where $*$ denotes maximized values, as in

$$l_0^* = \max_{\{\beta, \Sigma\}} \left(-\frac{T}{2} \log(|\Sigma|) - \frac{1}{2} \sum_{t=1}^T (y_t - x_t' \beta)' \Sigma^{-1} (y_t - x_t' \beta) \right),$$

$$\begin{aligned}
 l_A^*(\mathcal{J}_M) &= \max_{\left\{ \begin{array}{l} \beta_1, \dots, \beta_{M+1}, \\ \Sigma_1, \dots, \Sigma_{M+1} \end{array} \right\}} l(\mathcal{J}_M, \beta_1, \dots, \beta_{M+1}, \Sigma_1, \dots, \Sigma_{M+1}) \\
 &= \max_{\left\{ \begin{array}{l} \beta_1, \dots, \beta_{M+1}, \\ \Sigma_1, \dots, \Sigma_{M+1} \end{array} \right\}} \sum_{m=1}^{M+1} \left(\begin{array}{l} -\frac{k_m - k_{m-1}}{2} \log(|\Sigma_m|) \\ -\frac{1}{2} \sum_{t=k_{m-1}+1}^{k_m} (y_t - x_t' \beta_m)' (\Sigma_m)^{-1} (y_t - x_t' \beta_m) \end{array} \right), \\
 LR(\mathcal{J}_M) &= 2(l_A^*(\mathcal{J}_M) - l_0^*). \tag{2.1}
 \end{aligned}$$

For the detection of break dates that are unknown, the set \mathcal{J}_M must be estimated except for the known dates $k_0 = 0$ and $k_{M+1} = T$. For the break detection test to be well-behaved, the difference between each pair of ending dates must be at least $k_m - k_{m-1} = \omega = \lceil \kappa T \rceil$, where $\kappa \in (0, 0.5]$, and where $\lceil \cdot \rceil$ rounds a number to the next largest integer. Let \mathcal{J}_κ be the set that contains all permitted combinations of ending dates. The set of break date estimates $\hat{\mathcal{J}}_M \in \mathcal{J}_\kappa$ is identified as the set that maximizes the likelihood ratio of equation (2.1). Formally, the test statistic is given as the supremum value, $\sup LR(\mathcal{J}_M)$, such that

$$\hat{\mathcal{J}}_M = \arg \sup_{\mathcal{J}_M \in \mathcal{J}_\kappa} LR(\mathcal{J}_M). \tag{2.2}$$

The limit distribution of the maximum likelihood ratio statistic $\sup LR(\mathcal{J}_M)$ has been described in Qu and Perron (2007) for data generating processes that satisfy the assumptions in the previous Section 2.2.1, as in

$$\begin{aligned}
 \sup LR(\mathcal{J}_M) &\xrightarrow{d} QP(M, bq, bs, \kappa), \\
 QP(M, bq, bs, \kappa) &= \sup_{\lambda \in \Lambda_\kappa} \sum_{m=1}^M LR_m(\lambda, bq, bs), \\
 LR_m(\lambda, bq, bs) &= \frac{\|\lambda_m W_{bq}(\lambda_{m+1}) - \lambda_{m+1} W_{bq}(\lambda_m)\|^2}{(\lambda_{m+1} - \lambda_m) \lambda_m \lambda_{m+1}} \\
 &\quad + \frac{1}{2} (\lambda_m W_{bs}(\lambda_{m+1}) - \lambda_{m+1} W_{bs}(\lambda_m))' H \Omega H' \\
 &\quad \times \frac{(\lambda_m W_{bs}(\lambda_{m+1}) - \lambda_{m+1} W_{bs}(\lambda_m))}{(\lambda_{m+1} - \lambda_m) \lambda_m \lambda_{m+1}},
 \end{aligned}$$

where $\lambda = \{\lambda_1, \dots, \lambda_M\}$, $\lambda_{M+1} = 1$, and Λ_κ is the set of all permitted combinations of break fractions according to the trimming parameter κ . $W_b(\cdot)$ is a b -dimensional vector of independent Wiener processes.

Critical values of the test statistic $\sup LR(\mathcal{J}_M)$ are obtained by simulating the limit distribution $QP(M, bq, bs, \kappa)$ repeatedly. For this, the independent Wiener processes in $W(\cdot)$ are simulated by partial sums of i.i.d. normal random variables. A large number of observations must be simulated a large number of times for the derivation of critical values. For example, Andrews (1993) chooses to simulate 3600 observations 10000 times. Alternatively, a bootstrap method can be applied, so that the distribution of $\sup LR(\mathcal{J}_M)$ under the null hypothesis of no structural break is derived from empirical resampling. This will be further discussed in the next section.

The limit distribution $QP(M, bq, bs, \kappa)$ depends on the number of breaks M , the number of covariate coefficients allowed to change bq , the number of covariance coefficients allowed to change bs , where each off-diagonal element counts twice, and on the trimming parameter κ . The limit distribution further depends on H and Ω . The matrix H identifies the coefficients in the covariance matrix Σ allowed to change. Specifically, H must be a full row rank matrix of dimension $bs \times n^2$, such that $Hvec(\Sigma)$ is the bs -dimensional vector of covariance elements allowed to change. For example, consider the case of $n = 4$ time series, and a test for structural breaks in the three covariance coefficients between the first three series. Since covariance coefficients are off-diagonal elements of Σ , it follows that $bs = 2 \cdot 3 = 6$. Further,

$$vec(\Sigma) = (\sigma_1^2, \sigma_{12}, \sigma_{13}, \sigma_{14}, \sigma_{21}, \sigma_2^2, \sigma_{23}, \dots, \sigma_4^2)',$$

$$H = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$Hvec(\Sigma) = (\sigma_{12}, \sigma_{13}, \sigma_{21}, \sigma_{23}, \sigma_{31}, \sigma_{32})'.$$

Ω is a covariance matrix of scaled error term products and cross-products over time. With standardized residuals $\tilde{\varepsilon}$, let $\xi_t = \tilde{\varepsilon}_t \tilde{\varepsilon}_t' - I_n$ be an $n \times n$ matrix for all $t = 1, \dots, T$. Further, $vec(\xi)$ is a $T \times n^2$ matrix, where the t -th row contains the n^2 scaled products and cross-products of standardized residuals at date t . Then Ω is the $n^2 \times n^2$ matrix $\Omega = Var(vec(\xi))$. Ω prevents size distortions of the test, when data is non-normal and structural breaks occur in the covariance matrix. Ω is larger when the error distribution is leptokurtic. This is relevant for financial data, which is often conditionally heteroscedastic. Ω can be estimated using a heteroscedasticity and autocorrelation robust estimator, see also Bai and Perron (2003). In this book, the covariance estimator presented in Newey and West (1987) is applied.

When determining the significance of M structural breaks, an optimal distribution of the set of break date \mathcal{J}_M will be used, according to equation (2.2). To find the globally optimal distribution of break dates in the sample, all possible combinations of dates must be considered. To reduce the computational burden of this partitioning task, Bai and Perron (2003) present a dynamic program, whose origin is attributed earlier cluster analysis literature such as Guthery (1974). While running, the dynamic programming algorithm dismisses certain break date pairings that cannot be part of the final optimal solution. Bai and Perron (2003) formulate the program in the context of OLS estimation and sum of squared residuals, while Qu and Perron (2007) translate the concept to maximum likelihood ratio testing as follows,

$$l_A^*(\mathcal{J}_M) = \max_{\omega \cdot M \leq t \leq T - \omega} (l_A^*(\mathcal{J}_{M-1,t}) + l^*(t+1, T)),$$

where $\mathcal{J}_{M-1,t}$ is the optimal partitioning of $M - 1$ breaks using only the first t observations, $l^*(t+1, T)$ is the maximized log-likelihood function using the observations between

$t + 1$ and T , and where ω is still the minimum difference between break dates. Note that a maximum value $l_A^*(\mathcal{J}_M)$ also maximizes the likelihood ratio $LR(\mathcal{J}_M) = 2(l_A^*(\mathcal{J}_M) - l_0^*)$ defined in equation (2.1), as the value for l_0^* is independent from the partitioning.

To provide some intuition, consider the case that $M = 3$, $T = 100$ and $\omega = 10$. If the set of break dates candidates $(10, 21, 31)$ yields a larger likelihood ratio than another set $(11, 21, 31)$, then the set $(10, 21, 32)$ must also yield a larger likelihood ratio than the set $(11, 21, 32)$. Differently stated, for a given middle date k_2 , the optimal positioning of date k_1 is independent of the optimal positioning of date k_3 . The presented dynamic program is therefore only effective when $M \geq 3$. The larger the number of tested breaks, the larger the reduction in computation time compared to a naive comparison of all partitioning options. Chapter 4 will introduce a heuristic that can reduce the computation time even more drastically.

In this final part of the section, the influence of partial breaks on maximum likelihood ratio testing is explained. Constant parameters θ^c can be implemented by adding corresponding equality restrictions on the parameters. This, however, will affect the dynamic program of finding multiple breaks. An iterative scheme for this scenario is described in Qu and Perron (2007), as well as in Bai and Perron (2003). In a first step, the equality restrictions are neglected when estimating break dates and coefficients. In a second step, the coefficients are estimated under the restrictions, conditional on the given break date estimates from the earlier step. In a third step, the break dates and changing coefficients are estimated, using the constant coefficient estimates from the earlier step. Finally, the second and third step are repeated until convergence. Similar iterative schemes are developed and employed in Bataa et al. (2013), and also in Chapter 3, when distinct mean, variance and correlation changes are investigated for the topic of financial contagion.

2.3 Bootstraps

Throughout the book, bootstraps will be helpful for the application of the described statistics in practice. First, concept, history and overall purpose of the bootstrap will be described, before specified bootstrap techniques are presented later in this section. The techniques are initially described in their general form and then in the form that can be used in structural break testing.

The bootstrap is a resampling method that is used to estimate statistics based on random resampling of the given data with replacement. The method was introduced by Efron (1979) and derives its name from the phrase “to put oneself up by one’s own bootstrap”. Its concept is to derive additional information about the behavior of a measure in a given data sample by randomly drawing new samples from that same data. Its concept is therefore closely related to the concept of the jackknife method, earlier developed by Quenouille (1949) and Tukey (1958). However, the jackknife systematically leaves out each observation of a sample. Efron and Tibshirani (1994) refers to the jackknife as a simple closed-form approximation to the bootstrap in the context of standard errors and bias estimations. The introduction of the bootstrap with its use of random draws has improved the performance of the resampling method, and current research on different bootstrap approaches and applications is extensive. A range of appropriate application

possibilities are discussed in Efron (1979) and Bickel and Freedman (1981), or more recently in Politis (2003),

Bootstraps are frequently used to estimate standard errors of parameter estimates, such as regression model parameters. They are useful in structural break testing to analyze the uncertainty associated with break date estimates, when the distribution of break date estimators is non-standard. Confidence intervals can be derived from a bootstrap and overlapping confidence intervals of two structural break date estimates may be used to formulate a hypothesis about the true break date being the same for both breaks.

Bootstraps can also be used for inference of a structural break test result, when the distribution of the respective test statistic under the null hypothesis model is not known or non-standard. Bootstrap resampling and repeated testing will produce an empirical distribution of the test statistic. Such empirically derived critical values may be preferred to critical values that rely on an asymptotic distribution of the structural break test as an approximation, especially if the number of available observations is limited. As noted by Kirch (2008), frequently convergence to the limit distribution of structural break tests is slow. Bootstrapped tests generally perform better than tests based on approximate asymptotic distributions, according to Davidson and MacKinnon (2004). Before describing in more detail, how bootstraps can be used in structural break testing, an overview of relevant bootstrapping techniques will be given in the following.

As discussed in Efron and Tibshirani (1994), it is important to point out that the concept of bootstrapping is compatible with a large variety of methods of data resampling. None of the presented structural break tests of this book rely on a specific, uniquely defined bootstrap method. Bootstraps range from fully parametric bootstraps, in which estimated models are used to make simulations of the dependent variable, to non-parametric resampling bootstraps, in which original observations are randomly reused. In practical applications of the presented tests, the choice of a bootstrap method will depend on the confidence with which the underlying probability model of a data sample is interpreted. The choice can be supported by arguments found in the bootstrapping literature.

One method used in the context of time series data and structural break testing is residual resampling. It is a parametric bootstrap method that can also be referred to as regression bootstrap, see Hušková and Kirch (2012). Consider a newly created bootstrap sample where a synthetic observation y_t^* is given as

$$y_t^* = \hat{f}(x_t; \hat{\theta}) + \hat{\varepsilon}_j, \text{ for } t = 1, \dots, T, \text{ with } j \in \{1, \dots, T\}, \quad (2.3)$$

and where \hat{f} is the estimated functional form of the model with estimated parameters $\hat{\theta}$. In each newly created sample, all $\hat{\varepsilon}_j$ are randomly drawn from the set of regression residuals with replacement. As in all bootstrap methods, a large number of data samples are simulated according to the specified design.

With this residual resampling method, all T original covariate vectors x_t are used; they are treated as fixed quantities and not as observations of random variables. Efron and Tibshirani (1994) explain this practice in the context of estimating standard errors of model parameter estimates. By treating x_t as fixed, the bootstrapped standard error estimate reflects the precision associated with the given sample of x_t . If x_j were randomly

drawn in each bootstrap repetition, such that $x_j \in \{x_1, \dots, x_T\}$, the standard deviation between all x_i values in the resulting set would increase or decrease (depending on the draw), and that would decrease or increase the standard error of the parameter estimate. Therefore, a static set of x_t is preferable to a variable set. It can be shown that bootstrap standard errors derived from this residual resampling method are the same as arithmetic ones.

Case resampling, or pair bootstrapping, is another type of bootstrap method. In this non-parametric method, covariate observations are randomly drawn (with replacement) along with the corresponding observation of the dependent variable. A bootstrap sample consists of T observations, as given by

$$(y_t, x_t)^* = (y_j, x_j), \text{ for } t = 1, \dots, T, \text{ with } j \in \{1, \dots, T\}. \quad (2.4)$$

Here, the dependence between observation pairs is directly preserved. No model estimates are used in the simulations and no artificial observations are produced. Instead, existing observations are re-shuffled to create artificial samples of the data. Compared to residual resampling, this random drawing of pairs is less sensitive to assumptions made about the functional form of the model. However, it can be shown, for large data sets and a correctly specified model, that the answer given by case resampling approaches the answer given by residual resampling, according to Efron and Tibshirani (1994).

In practice, time series processes such as financial time series may exhibit autocorrelation or volatility clustering. Such forms of temporal dependence can be preserved in either bootstrap method by using blocks of data. The data blocks can have equal size b , which can be any number such as $b = 15$, depending on the degree of temporal dependence that is estimated or assumed about the data. For residual resampling, successive residuals are joined in distinct blocks, such that the j -th block contains residuals $\hat{\varepsilon}_{(j-1)b+1}, \hat{\varepsilon}_{(j-1)b+2}, \dots, \hat{\varepsilon}_{jb}$. Analogously, for case resampling, the distinct blocks contain successive pairs, such that the j -th block contains $(x_{(j-1)b+1}, y_{(j-1)b+1}), (x_{(j-1)b+2}, y_{(j-1)b+2}), \dots, (x_{jb}, y_{jb})$. Lahiri et al. (1999) provide an analysis of this and other types of block bootstraps, with random block length or overlapping blocks. When appropriate, a sieve bootstraps can be implemented to preserve error dynamics as an alternative to the block bootstrap. In this parametric method, residuals are first fit to an autoregressive model before the resulting residuals are resampled. The more general method of block bootstraps is applied for the empirical samples in the remaining chapters of this book.

As mentioned before, a classic application for bootstraps is the derivation of confidence intervals, see for example DiCiccio and Efron (1996). Structural breaks in time series models are associated with an estimate of the break date \hat{k} . A bootstrap can be applied to find a lower limit k_l and an upper limit k_u , such that they correspond to the $\alpha/2$ -th and the $(1 - \alpha/2)$ -th empirical percentile of ordered bootstrap break date estimates, respectively. According to Efron and Tibshirani (1994), when percentile-based confidence intervals are equal-tailed, the probability that a true break date k_0 with $k_0 < k_l$ yields a break data estimate $\tilde{k} > \hat{k}$ is smaller than $\alpha/2$. Since the analogous conclusion can be drawn for true break dates larger than the upper limit, the break date estimate \hat{k} corresponds to a true break date between the limits, with $1 - \alpha$ confidence.

In the empirical study of Chapter 3, percentile-based confidence intervals of the break date estimates are derived using a block bootstrap.

When a bootstrap method is used for hypothesis testing, the original data must be resampled in such a way that new data is known to satisfy the null hypothesis, even if the original data fails to satisfy the null hypothesis (see Hall and Wilson, 1991). For structural break testing specifically, the data is resampled in a way that the potential structural break is eliminated, but other properties of the data are preserved. When the test statistic is applied to a large number of such samples, the results are stored and ordered by value. If the test result of the original data is unusual, such that only a fraction α of the bootstrap simulations produce a test result at least equally extreme, then the null hypothesis will be rejected in favor of the alternative hypothesis with $1 - \alpha$ confidence. Note that α is interpreted as the p -value of the null hypothesis, as in other types of hypothesis tests. Accordingly, when interpreting the p -value, base probabilities for either hypothesis should be considered. For a discussion on typical misconceptions about the p -value, refer to Greenland et al. (2016).

For the derivation of an empirical null hypothesis distribution in structural break testing, it is possible to use case resampling or residual resampling, and in either case with or without blocks of data. A straightforward application is described in Chapter 6 for a test that finds a structural break in a regression model against the null hypothesis of no structural break. For the residual resampling bootstrap method, model parameter estimates that correspond to a linear regression model without structural break are used to calculate the synthetic values y_t^* . Thus, the no-break parameter estimates are used for $\hat{\theta}$ in equation (2.3). That way, samples can be simulated that are known to satisfy the null hypothesis of no break. Alternatively, the case resampling bootstrap method can be applied as described above. Here, the redistribution of actual data is expected to erase the temporal ordering of the data that is fundamental to any hypothesized structural break.

Other chapters of this book contain tests, whereby regression models under the null hypothesis contain a structural break. In Chapter 3, a common break date test is applied, where multiple parameters change their value at a common break date under the null hypothesis and at distinct times under the alternative hypothesis. To derive the test statistic distribution under the null hypothesis, data must be simulated where the parameter shift is known to happen at the common break date. Similarly, Chapter 5 introduces a test statistics that tests the null hypothesis of an abrupt parameter change at the estimated break date against the alternative hypothesis of non-abrupt parameter change close to the estimated break date. Again, data must be simulated where the tested parameters abruptly shift in value at the estimated break date.

When data is simulated that contains a (common) structural break, as usual, residuals or cases can be resampled, with or without data blocks. The corresponding methods shown in equations (2.3) and (2.4) are only adjusted with respect to the data source sets. If the (common) break date estimate is \hat{k} , then the first \hat{k} simulated observations must sample from the first \hat{k} original observations, and the remaining $T - \hat{k}$ simulated observations must sample from the remaining $T - \hat{k}$ original observations. Accordingly, if the bootstrap method is parametric, pre-break parameter estimates are used for the first \hat{k} simulated observations and post-break parameter estimates are used for the remaining $T - \hat{k}$ simulated observations.

Naturally, many other bootstrap methods can be found in the literature that relate to various aspects of structural break testing. The topic of non-stationary time series data is covered in publications such as Palm et al. (2008), in which the properties of bootstrap unit-root tests are studied. Some publications cover the topic of bootstrapping for tests that monitor structural breaks. In this case, due to the continual increase of the historical data set, a choice must be made whether and in which way the bootstrapped critical values must be updated as well. Steland (2006) proposes a type of residual resampling bootstrap that has infrequent updates and uses convex combinations of earlier bootstrap distributions. This drastically reduces the computational burden compared to a continually updated bootstrap, while testing power is minimally reduced, as shown in simulation studies of Kirch (2008), and more recently Hlávka et al. (2016).

3

Detecting contagion in a multivariate time series system¹

3.1 Introduction

The recent European sovereign debt crisis has rapidly diffused across borders in the euro area. Yet, the diffusion has been different for each euro area country. For some countries, particularly Southern European countries such as Italy, Portugal and Spain, the impact has been quick and negative, whereas the direct impact on Northern and central European countries such as Germany and Austria appears less severe. For policy makers, such observations are crucial for the establishment of efficient firewalls to stop the diffusion of the turmoil, and they stress the problem of finding a universal, European response to the crisis. This issue has been the starting point of a revival of empirical studies on the transmission of shocks and financial contagion (Metiu, 2012, De Santis, 2012, Caporin et al., 2013, Beirne and Fratzscher, 2013 or Claeys and Vašíček, 2014, to name a few). These studies rely on distinct definitions of contagion and methodologies and lead to different conclusions. The current chapter seeks to refine the empirical methodology to better detect the transmission of crises in time series data.

Since the seminal papers of King and Wadhvani (1990), Calvo and Mendoza (2000) and Baig and Goldfajn (1999), contagion has often been identified by a significant increase in the correlation between financial series of two countries, for example stock market indexes. A key methodological contribution is Forbes and Rigobon (2002), in which it is shown that contagion is over-accepted, if one ignores changes that occur in the variance when testing for correlation changes. Following Diebold and Yilmaz (2009, 2012), another strand of literature focuses on the crisis transmission channels and inves-

¹The chapter is based on Blatt, Candelon, and Manner (2015)

tigates the stability of spillovers between time series. The two approaches adopt different timing perspectives, either studying abrupt changes in the contemporaneous dependence, or changes in dynamic spillovers. Both approaches will be included in the study of this chapter and it will be important to have clear definitions of relevant terms.

The following terminology will be adopted in this chapter. Interdependence refers to existing linkages between markets. Crisis is turmoil in financial markets that occurs as a result of shocks, and it materializes in the form of increased volatility. Contagion in general refers to the situation in which the degree of interdependence increases beyond its usual level during crisis. Shift contagion refers to an increase in contemporaneous correlation, i.e., an increase in the immediate shock transmission. Spillover contagion is associated with breaks in the dynamics of transmission channels. This form of contagion is directional, and typically requires a certain time lag to materialize. Finally, by flight-to-quality we mean the situation in which the level of interdependence decreases during crisis, indicating a decoupling of safe haven countries that are believed to be unaffected by the crisis. See Baur and Lucey (2009) for a comparison between contagion and flight-to-quality in the case of stocks and bonds markets.

Three important methodological issues in contagion testing must be addressed. First, many studies exogenously determine crisis dates. In other words, structural break dates are not inferred from the data but specified before. Several procedures for endogenous break date determination have been proposed, e.g., in Eichengreen et al. (1995, 1996), Favero and Giavazzi (2002), Candelon and Manner (2010) or Metiu (2012).

Second, volatility changes, representing increased turmoil and the outbreak of the crisis, and correlation changes, representing shift contagion, are typically assumed to occur simultaneously. Candelon and Manner (2010) challenge this assumption and observe that during the 1997 Asian Financial Crisis, variance changes preceded correlation shifts in most cases. The economic motivation behind this finding is that the intensification of interdependence is not immediate and takes place only when markets are already stressed. In such situations, assuming simultaneity of volatility and correlation shifts can obscure the individual effects and the presence of contagion may be underestimated.

Third, many existing papers such as Forbes and Rigobon (2002), Metiu (2012), or Candelon and Manner (2010) exclusively analyze pairwise correlations. Dungey et al. (2004) note that a bivariate analysis can lead to biased conclusions. A multivariate approach is recommended, so that the diffusion process can be analyzed fully. A shock that originates in country/market X does not necessarily impact country/market Y directly, but may indirectly transit via country/market Z . Furthermore, from a purely econometric perspective, Bai et al. (1998), Groen et al. (2011) and Qu and Perron (2007) show that the date of a structural break is estimated more precisely in multivariate data sets.

The contribution of this chapter is to propose and apply a novel sequential testing procedure (STP) for contagion detection that addresses the three described methodological issues. Relying on the theory developed in Qu and Perron (2007), we specify a vector autoregressive model that is subject to multiple structural breaks. The key methodological innovation is to separate the structural breaks in the parameters for the conditional mean, the error variance and the correlation between series. Furthermore, the procedure tests whether the inferred breaks are distinct from one another, as proposed in Perron and Oka (2011).

The sequential procedure is performed in a multivariate dynamic set-up (of dimension 10) to benefit from high testing power and more precise estimates of the break dates. This will improve contagion detection. The idea of decomposing structural breaks that involve mean, variance and correlation parameters is similar to the one proposed by Bataa et al. (2013), who study structural breaks in cross-country inflation relations. An advantage of the proposed STP is that the covariance matrix is decomposed before testing and not after finding a break in the covariance matrix. This means that the STP involves structural break testing in fewer parameters, which lowers the degrees of freedom and increases the power of the applied break tests.

Similarly to Missio and Watzka (2011), Metiu (2012), De Santis (2012), Caporin et al. (2013), Beirne and Fratzscher (2013) and Claeys and Vašíček (2014), we analyze 10-year bond yield spreads over Germany. Our analysis confirms previous conclusions regarding the absence of shift contagion between Greece and other countries. We can identify and date the sequence of regimes in the European sovereign bond markets. We observe that volatility changes occur simultaneously in the whole system but are distinct from the structural breaks of the conditional mean parameters, which are again common across all equations. This finding stresses the coincidence of unrest hitting the euro area economies during the European sovereign debt crisis.

In the application, changes in correlation occur at various dates and are not common to all country pairs. Still, many of the corresponding structural breaks cluster around a small number of dates, and the distinctiveness tests enable us to identify a few 'contagion clubs', i.e., countries that have common breaks that relate to correlation changes. Furthermore, we find strong evidence of a flight-to-quality mechanism at various stages of the crisis that is associated with a decrease in correlation.

The study also contains a spillover analysis, which benefits from the structural breaks in mean, variance and correlation parameters detected by the STP. Spillover indexes can be calculated segmentwise to incorporate the structural break test results Diebold and Yilmaz (2009, 2012). An increase in the spillover index between data segments indicates the presence of spillover contagion. For the empirical application, changes in the spillover indexes over time are plotted to visualize the dynamics of shock transmission, similarly to Claeys and Vašíček (2014). We demonstrate significant time variation in the spillovers between euro area countries. There is evidence of spillover contagion from Greece at various times during the crisis, so in this analysis the role of Greece is different to its role in the shift contagion analysis.

The chapter is structured as follows. Section 3.2 motivates and explains the model and the STP methodology. The section includes a description of the applied structural break detection test and the break distinctiveness test. The last part of the section describes spillover indexes. Section 3.3 presents the empirical application of our method to the case of the European sovereign debt crisis, using STP and spillover analysis. Concluding remarks are made in Section 3.4.

3.2 Methodology

3.2.1 The multivariate model

Our analysis of financial contagion builds on the following vector autoregressive model (VAR) for a vector of financial time series $y_t = [y_{1,t}, \dots, y_{n,t}]'$, such that

$$y_t = \beta_{0,t} + \sum_{i=1}^p B_{i,t} y_{t-i} + B_{x,t} x_t + \varepsilon_t, \quad (3.1)$$

for $t = 1, \dots, T$. x_t is a vector of exogenous variables, and the coefficient matrices $B_{i,t}$ and $B_{x,t}$ as well as the intercept vector $\beta_{0,t}$ are potentially time-varying. The n -dimensional vector of error terms $\varepsilon_t = [\varepsilon_{1,t}, \varepsilon_{2,t}, \dots, \varepsilon_{n,t}]'$ follows some (unknown) distribution with covariance matrix Σ_t . Assumptions about the innovations are mild. They allow for the typical features observed in financial returns, in particular, conditional heteroscedasticity and autocorrelation. The assumptions are found in the first section of Chapter 2.

The parameters in equation (3.1) are allowed to be time-varying by being subject to structural breaks at unknown points in time. Below we explain how to test for and date structural breaks in the conditional mean, in the conditional variance and in conditional correlations in a manner similar to that described in Bataa et al. (2013). The estimated break dates and the corresponding changes in parameter estimates in turn allow us to infer whether and to what extent contagion has occurred.

3.2.2 Shift contagion in the multivariate model

For the moment, let us ignore any time variation in parameters of the conditional mean equation. Shift contagion is detected when the correlation between markets increases beyond its pre-crisis level. Because contemporaneous dependence is not part of the conditional mean model, it is captured by the covariance matrix Σ_t of the errors ε_t . Thus, within this model framework, testing for shift contagion means testing for an increase in the dependence between the residuals $\hat{\varepsilon}_t$. However, as noted by Forbes and Rigobon (2002), a change in the covariance matrix Σ_t does not allow for the identification of contagion. The origin of an increase in a covariance term $\sigma_{ij} = \sigma_i \rho_{ij} \sigma_j$ can result from an increase in the correlation or from a rise in the variance, where the latter is typically a sign of crisis outbreak or intensification. Therefore, we decompose the covariance matrix as follows,

$$\Sigma_t = S_t R_t S_t.$$

R_t is the matrix of $n(n-1)/2$ correlation coefficients $\rho_{ij,t}$ and S_t is a diagonal matrix containing n standard deviations $\sigma_{i,t}$, for $i, j = 1, \dots, n$. A test for shift contagion looks for an increase in the elements of the correlation matrix R_t , which measure contemporaneous interdependence only. In our approach, breaks that affect volatility and correlation are not assumed to be simultaneous, but it can be tested whether the break dates are the same.

Table 3.1 illustrates a hypothetical crisis scenario that can be identified using our procedure. The crisis breaks out in market 1 first, which increases the standard deviation σ_1 . Market 3 then enters a high-volatility crisis state, followed by market 2. The last

Table 3.1: Example of a sequence of crisis events

$$\begin{aligned}
& \begin{bmatrix} \sigma_1^{(1)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(1)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12}^{(1)} & \rho_{13}^{(1)} & \rho_{14} \\ \rho_{12}^{(1)} & 1 & \rho_{23}^{(1)} & \rho_{24} \\ \rho_{13}^{(1)} & \rho_{23}^{(1)} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{(1)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(1)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} = \begin{bmatrix} \sigma_1^{2(1)} & \sigma_{12}^{(1)} & \sigma_{13}^{(1)} & \sigma_{14}^{(1)} \\ \sigma_{12}^{(1)} & \sigma_2^{2(1)} & \sigma_{23}^{(1)} & \sigma_{24}^{(1)} \\ \sigma_{13}^{(1)} & \sigma_{23}^{(1)} & \sigma_3^{2(1)} & \sigma_{34}^{(1)} \\ \sigma_{14}^{(1)} & \sigma_{24}^{(1)} & \sigma_{34}^{(1)} & \sigma_4^2 \end{bmatrix} = \Sigma^{(1)} \\
& \text{Variance change in series 1} \\
& \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(1)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12}^{(1)} & \rho_{13}^{(1)} & \rho_{14} \\ \rho_{12}^{(1)} & 1 & \rho_{23}^{(1)} & \rho_{24} \\ \rho_{13}^{(1)} & \rho_{23}^{(1)} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(1)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} = \begin{bmatrix} \sigma_1^{2(2)} & \sigma_{12}^{(2)} & \sigma_{13}^{(2)} & \sigma_{14}^{(2)} \\ \sigma_{12}^{(2)} & \sigma_2^{2(1)} & \sigma_{23}^{(1)} & \sigma_{24}^{(1)} \\ \sigma_{13}^{(2)} & \sigma_{23}^{(1)} & \sigma_3^{2(1)} & \sigma_{34}^{(1)} \\ \sigma_{14}^{(2)} & \sigma_{24}^{(1)} & \sigma_{34}^{(1)} & \sigma_4^2 \end{bmatrix} = \Sigma^{(2)} \\
& \text{Variance change in series 3} \\
& \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12}^{(1)} & \rho_{13}^{(1)} & \rho_{14} \\ \rho_{12}^{(1)} & 1 & \rho_{23}^{(1)} & \rho_{24} \\ \rho_{13}^{(1)} & \rho_{23}^{(1)} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(1)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} = \begin{bmatrix} \sigma_1^{2(2)} & \sigma_{12}^{(2)} & \sigma_{13}^{(3)} & \sigma_{14}^{(2)} \\ \sigma_{12}^{(2)} & \sigma_2^{2(1)} & \sigma_{23}^{(2)} & \sigma_{24}^{(1)} \\ \sigma_{13}^{(3)} & \sigma_{23}^{(2)} & \sigma_3^{2(2)} & \sigma_{34}^{(1)} \\ \sigma_{14}^{(2)} & \sigma_{24}^{(1)} & \sigma_{34}^{(1)} & \sigma_4^2 \end{bmatrix} = \Sigma^{(3)} \\
& \text{Variance change in series 2 and correlation changes between series 1, 2 and 3} \\
& \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(2)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} \begin{bmatrix} 1 & \rho_{12}^{(2)} & \rho_{13}^{(2)} & \rho_{14} \\ \rho_{12}^{(2)} & 1 & \rho_{23}^{(2)} & \rho_{24} \\ \rho_{13}^{(2)} & \rho_{23}^{(2)} & 1 & \rho_{34} \\ \rho_{14} & \rho_{24} & \rho_{34} & 1 \end{bmatrix} \begin{bmatrix} \sigma_1^{(2)} & 0 & 0 & 0 \\ 0 & \sigma_2^{(2)} & 0 & 0 \\ 0 & 0 & \sigma_3^{(2)} & 0 \\ 0 & 0 & 0 & \sigma_4 \end{bmatrix} = \begin{bmatrix} \sigma_1^{2(2)} & \sigma_{12}^{(3)} & \sigma_{13}^{(4)} & \sigma_{14}^{(2)} \\ \sigma_{12}^{(3)} & \sigma_2^{2(2)} & \sigma_{23}^{(3)} & \sigma_{24}^{(2)} \\ \sigma_{13}^{(4)} & \sigma_{23}^{(3)} & \sigma_3^{2(2)} & \sigma_{34}^{(2)} \\ \sigma_{14}^{(2)} & \sigma_{24}^{(2)} & \sigma_{34}^{(2)} & \sigma_4^2 \end{bmatrix} = \Sigma^{(4)}
\end{aligned}$$

Note: Example of crisis and contagion events resulting in four covariance matrix segments $\Sigma^{(m)}$, $m = 1, \dots, 4$. The left side of the term indicates the decomposed covariance matrix, $SRS = \Sigma$. All parameter changes are highlighted in bold. Crisis breaks out in market 1 first, then in market 3 and reaches market 2 last, whereas contagion occurs at the same time as the crisis outbreak in market 2. Market 4 remains unaffected.

standard deviation change occurs simultaneously with shift contagion between the first three markets, which means that the three correlation coefficients ρ_{12} , ρ_{13} and ρ_{23} shift to higher values. The fourth market is not affected by the crisis other than via the usual market connections, but nevertheless, its covariance with the other markets changes. A test seeking instability in, for example, the covariance $\sigma_{13} = \sigma_1 \rho_{13} \sigma_3$ would produce a biased break date estimate in between the three distinct breaks, as all components change separately. A direct test for a structural break that involves covariances could not identify the source of instability. Furthermore, the number of covariance components changing is smaller than the number of affected covariances. Thus, a test on covariances would require more degrees of freedom, which would negatively affect the power of the test.

To summarize, our test for shift contagion is a test for a structural break that affects the correlation matrix at an unknown date, conditional on structural breaks at (possibly distinct) unknown dates that affect the volatility of the series. However, the VAR coefficients in equation (3.1) are likely to also be subject to structural breaks; see Bataa et al. (2013) and Claeys and Vařicek (2014). Therefore, prior to testing for and estimating structural breaks in correlations and variances, we test for breaks in the set of VAR parameters $B_t = \{\beta_{0,t}, B_{i,t}, B_{x,t}\}$. A test for multiple structural breaks in the parameters of the conditional mean and covariance of a multivariate system is proposed in Qu and Perron (2007), and we adopt this approach. Bataa et al. (2013) propose an iterative procedure for separating breaks in the coefficients and the covariance matrix. We adapt

this procedure to our setting, but further allow for distinct breaks in correlations and volatilities. Our testing procedure operates as follows.

1. Determine M_B , the number of breaks in the VAR coefficients, and estimate the break dates $k_1^B, \dots, k_{M_B}^B$. Estimate the coefficients B_t for the corresponding data segments.
2. Compute the residuals $\hat{\varepsilon}_t$, and conditional on the breaks from Step 1, determine M_S , the number of breaks in the standard deviations in S_t and estimate the corresponding break dates $k_1^S, \dots, k_{M_S}^S$. Estimate the segment-specific standard deviations σ_{it} .
3. Compute the standardized residuals $\tilde{\varepsilon}_{it} = \hat{\varepsilon}_{it}/\hat{\sigma}_{it}$. Conditional on the breaks from Steps 1 and 2, determine M_R , the number of breaks in the correlation matrix R_t and the estimate the corresponding break dates $k_1^R, \dots, k_{M_R}^R$.
4. Conditional on the breaks in R_t , re-do Step 2.
5. Conditional on the breaks in S_t and R_t re-do Step 1.
6. Iterate between Steps 1 to 5 until the number of breaks and the estimated break dates do not change.

All tests are based on (pseudo) likelihood ratio (LR) statistics derived from a multivariate normal distribution function. Break date estimates are identified as the dates that jointly maximize the LR statistic. Chapter 2 contains a general description of the maximum LR statistic and the limiting distribution that has been described by Qu and Perron (2007). The location of the multiple break dates can be estimated using the dynamic optimization program proposed in Bai and Perron (2003). Alternatively, Chapter 4 proposes a rolling window detection algorithm, a heuristic that reduces the calculation time of iterative, multiple break tests substantially. Critical values are obtained by simulating the limiting distribution, given the number of break dates and the number of changing parameters. Critical values will also depend on the data trimming, that means, the minimum required amount of observations between break dates. Although the LR statistic uses a normal distribution function, non-normality of the data is permitted in the assumptions. The error term covariance is estimated using an estimator that is robust to heteroscedasticity and autocorrelation. The outcome affects the critical values of the maximum LR test, as well.

A few additional remarks must be made concerning the details of the proposed procedure. First, the maximum number of breaks M of each type must be set prior to the analysis. Also, the minimum segment length between two breaks must be chosen. In the empirical study of this chapter, for example, we allow for a maximum $M = 3$ breaks per model parameter and restrict each segment to contain at least 10% of all observations. Second, a feasible generalized least squares estimator is applied to estimate the VAR coefficients in Step 5, conditional on the breaks in the covariance matrix that have been detected in earlier steps. For details of this procedure, see Bataa et al. (2013). Third, a bootstrap procedure can be used to estimate confidence intervals of the break dates. In the empirical study of this chapter, blocks of data are resampled between the estimated break dates.

A final remark concerns the assumption of common break dates, which will be treated in the next part of this methodology section. Although the break dates in B_t , S_t and R_t are allowed to occur at distinct dates, all parameters within each type are assumed

to share common break dates, for now. This strong assumption results in favorable properties for the break detection test. However, when this assumption is violated, the test may not detect any structural change due to low power, or the break date estimates may be biased.

3.2.3 Testing for common break dates

Structural break tests have more power and produce more reliable estimates of the break dates when applied to multivariate data systems according to Bai et al. (1998), Groen et al. (2011) and Qu and Perron (2007). Unreported Monte Carlo simulation results confirm this effect for small samples. This is one motivation for looking for joint break dates of a large system of series, instead of testing only bivariate data for contagion. Before advantages of common break date testing can be exploited, it is recommended that the common break date assumption is tested.

For the parameters in B_t and S_t , separately, we suggest to test whether the n equations share common break dates or not. For R_t , we suggest to test whether the $\frac{n(n-1)}{2}$ pairs of correlations share common break dates or not. A test for the coincidence of break dates in different model parameters is proposed by Perron and Oka (2011). For each type of parameter, we test

$$H_0 : \text{The parameters share common break dates}$$

against the alternative that the break dates are different across the n equations (or the $\frac{n(n-1)}{2}$ pairs of correlations). This null hypothesis can be tested using a likelihood ratio test. The restricted likelihood assumes common break locations in the data, whereas under the alternative, the number of breaks and their locations are determined for each equation/pair separately. Critical values are obtained using the following bootstrap algorithm.

1. Assuming common break dates, determine the number of structural breaks M and their dates $\hat{k}_1, \dots, \hat{k}_M$. Compute the log-likelihood under H_0 , LL_0 . Furthermore, determine the number of breaks, the break dates and the resulting log-likelihood LL_1 under the alternative. Compute the likelihood ratio statistic $LR = -2(LL_1 - LL_0)$.
2. Using the common break dates from Step 1, resample the multivariate observations using a block bootstrap scheme for each segment separately, i.e., k_1 observations from $t = 1, \dots, k_1$, then $k_2 - k_1$ observations from $t = k_1 + 1, \dots, k_2$, etc.
3. Using the resampled data, determine the number of breaks, their locations and the log-likelihood under H_0 and H_1 , and compute the bootstrap test statistic LR^* .
4. Repeat Steps 2 and 3 a large number of times to obtain the bootstrap distribution of the likelihood ratio statistic.

Note that in Step 2 the data is resampled either from the raw data in a case resampling bootstrap, or using either $\hat{\varepsilon}_t$ or $\tilde{\varepsilon}_t$ in a residual resampling bootstrap.

If the null hypothesis of common breaks is rejected, we suggest studying the estimated break dates and their confidence intervals to determine subsets of the data that may still share common breaks. For this, common breaks can be considered when confidence

intervals overlap. The presence of subsets with common break dates is beneficial in terms of break detection power and break estimation efficiency. Furthermore, the information received from the detected parameter groups is economically relevant. Conclusions must be drawn carefully, especially if common break hypotheses are not rejected for some combinations of parameters, but rejected for other combinations of the same parameters. Finally, overall, our procedure requires a large number of hypothesis tests, implying potential problems of multiple testing. Therefore, it is important to use conservative test sizes to control for the overall size of the procedure. Because the number of hypothesis tests to be performed is not known prior to the analysis, we use a size of 1% in our application.

3.2.4 The spillover index

Recently, Diebold and Yilmaz (2009, 2012) have presented a tool for analyzing the relationships between financial markets via spillovers. They construct spillover indexes that build on the structural, dynamic model linkages between countries. Spillover values can provide information about the source and destination of dynamic shock transmission during the sample segment. A VAR model framework such as ours enables the computation of the indexes, see Claeys and Vašíček (2014) for a similar application. An index increase indicates spillover contagion.

Let λ_{ij}^h be the h -step-ahead forecast error variance decomposition, i.e., the fraction of the forecast error variance for forecasting variable i that is due to shocks to variable j . Variance decomposition requires orthogonal innovations, and Diebold and Yilmaz (2012) suggest using the generalized impulse response framework of Koop et al. (1996) and Pesaran and Shin (1998), which is invariant to the ordering of the variables. Furthermore, λ_{ij}^h are normalized to satisfy $\sum_{j=1}^n \lambda_{ij}^h = 1$ and $\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij}^h = n$. The total spillover index is defined as the fraction of overall forecast error variance that is due to shocks to other markets, such that

$$TS^h = 100 \cdot \frac{\sum_{i=1}^n \sum_{j=1, j \neq i}^n \lambda_{ij}^h}{\sum_{i=1}^n \sum_{j=1}^n \lambda_{ij}^h}.$$

A directional spillover index is defined as either the spillover transmitted from i to all other markets, $DS_{\leftarrow i}^h$, or as the spillover received by i from the other markets, $DS_{\rightarrow i}^h$. Further, NS_i^h , a net spillover from market i can be calculated, so that

$$\begin{aligned} DS_{\leftarrow i}^h &= 100 \cdot \frac{\sum_{j=1, j \neq i}^n \lambda_{ji}^h}{n}, \\ DS_{\rightarrow i}^h &= 100 \cdot \frac{\sum_{j=1, j \neq i}^n \lambda_{ij}^h}{n}, \\ NS_i^h &= DS_{\leftarrow i}^h - DS_{\rightarrow i}^h. \end{aligned}$$

The spillover index depends on the VAR coefficients and on the covariance matrix of the residuals. We compute (and plot) these indexes by taking into account the structural breaks of each parameter type identified using our procedure described above. This yields an easily interpretable summary of the effects that the various parameter changes have

and allows the researcher to assess whether spillover contagion has occurred. Additionally, the interpretation of shift contagion is enhanced, as each previously detected data segment delimited by structural breaks can be associated with a spillover index value.

3.3 Empirical Study

3.3.1 Contagion during the European sovereign debt crisis

We consider daily 10-year sovereign bond yield spreads of $n = 10$ euro area countries, Austria (AUS), Belgium (BEL), Finland (FIN), France (FRA), Greece (GRE), Ireland (IRE), Italy (ITA), the Netherlands (NET), Portugal (POR) and Spain (SPA), over the yield of Germany (GER). Data is extracted from Thomson Reuters datastream, and the sample covers the period from 2009-01-02 until 2014-08-01. A preliminary analysis reveals that the unit root hypothesis cannot be rejected, and thus, the first differences in bond yield spreads are considered. According to the Bayesian Information Criterion (BIC), $p = 1$ lag is chosen for the VAR model. Because our sample covers the European sovereign debt crisis, 2 exogenous variables are included to control for systemic risks. As in Arghyrou and Kontonikas (2012), Beirne and Fratzscher (2013) and other studies, the Chicago Board Options Exchange Index (VIX) is included with one lag as a measure of global risk. Similarly to Claeys and Vašíček (2014), Caporin et al. (2013) and Tonzer and Buchholz (2014), the lagged spread between the Euribor 3-month lending rate and the overnight reference rate EONIA is additionally included as a measure of European financial market stress.

Given the large sample length of $T = 1456$, a moderate trimming $\kappa = 0.1$ has been considered, allowing for a minimal data segment length of $\omega = \lceil \kappa T \rceil = 146$ days. As noted above, we allow for a maximum of $M = 3$ breaks in each type of parameter, which corresponds to the number of (common) breaks found in Claeys and Vašíček (2014). As previously described, we sequentially search for structural breaks assumed to be common within each class of parameters. Thus, in Step 1, we search for a maximum of three simultaneous co-breaks involving $n(n + 1 + 2) = 130$ regression coefficients, Step 2 looks for a maximum of three simultaneous co-breaks involving $n = 10$ standard deviations (conditional on the breaks found in Step 1), and Step 3 looks for a maximum of three simultaneous co-breaks involving $\frac{n(n-1)}{2} = 45$ correlation coefficients (conditional on the breaks found in Steps 1 and 2).

Table 3.2 reports the results of testing for breaks involving mean coefficients and involving the variance parameters of the model error terms. Note that the results presented are the final results obtained after iterating the break detection until convergence, as explained in Section 3.2. The STP detects $M = 3$ significant, common breaks in both cases. The three breaks in the conditional mean equations are common across all equations, at p -value = 0.18. The breaks in the variances can also be assumed to be common across all equations, with p -value = 0.32. The null hypothesis that the breaks in variances and means are located at common dates is rejected, with p -value < 0.01. This finding supports our intuition that the breaks in different types of parameters occur at distinct times. Therefore, assuming their synchronicity can lead to a bias in the estimated break dates and the segment-specific parameter estimates. The three breaks in conditional means (August 2011, March 2012 and October 2012) represent times of

Table 3.2: Structural breaks for mean and standard deviation

Mean regression	$M_B = 3$ breaks		
max LR	1010.63		
99% critical value	470.71		
p -value	< 0.001		
Break date estimates	2011-08-15	2012-03-12	2012-10-02
95% confidence intervals	[2011-04-20,2011-08-19]	[2012-02-27,2012-03-12]	[2012-10-02,2014-01-13]
Covariates	coefficient change	coefficient change	coefficient change
<i>Const.</i>	-3.635	+3.709	-0.161
<i>AUS</i> _{$t-1$}	+1.757	-4.764	+5.308
<i>BEL</i> _{$t-1$}	+0.447	+5.577	-4.478
<i>FIN</i> _{$t-1$}	+8.131	-7.895	-0.838
<i>FRA</i> _{$t-1$}	-0.001	-1.139	+1.357
<i>GRE</i> _{$t-1$}	+0.646	-0.453	0.297
<i>IRE</i> _{$t-1$}	-0.740	-1.776	+1.032
<i>ITA</i> _{$t-1$}	-2.741	+1.557	-0.493
<i>NET</i> _{$t-1$}	-13.218	+10.529	+2.320
<i>POR</i> _{$t-1$}	-0.063	+0.016	-0.279
<i>SPA</i> _{$t-1$}	+3.189	-2.425	-0.354
<i>VIX</i> _{$t-1$}	+0.066	-0.099	+0.027
<i>(Euribor - EONIA)</i> _{$t-1$}	+2.538	-0.869	-1.100
Standard deviations	$M_S = 3$ breaks		
max LR	6877.00		
99% critical value	1222.76		
p -value	< 0.001		
Break date estimates	2010-01-22	2011-07-05	2012-10-01
95% confidence intervals	[2009-10-30,2010-03-11]	[2011-04-22,2011-08-01]	[2012-09-03,2012-10-25]
Series	s.d. change	s.d. change	s.d. change
<i>AUS</i>	-0.001	+0.028	-0.034
<i>BEL</i>	+0.020	+0.036	-0.057
<i>FIN</i>	-0.002	+0.009	-0.011
<i>FRA</i>	+0.004	+0.040	-0.039
<i>GRE</i>	+0.242	+1.303	-1.421
<i>IRE</i>	+0.084	+0.012	-0.099
<i>ITA</i>	+0.022	+0.095	-0.088
<i>NET</i>	-0.002	+0.010	-0.013
<i>POR</i>	+0.116	+0.092	-0.140
<i>SPA</i>	+0.045	+0.068	-0.077

Note: Test results, estimated break dates and corresponding changes in coefficients for the conditional mean parameters in the upper part, and the standard deviations in the lower part. 95% confidence intervals we obtained via a case resampling block bootstrap. For breaks in the mean regression, changes in coefficients are summarized for each regressor listed in the first column: The shifts in the estimated regression coefficients are sums across all equations.

important systemic changes in Europe: The August 2011 break can be associated with the issuance of 4.6 billion euros to assist Ireland and Romania, as well as several meetings of the Ecofin and the first meeting of the European Systemic Board. The March 2012 and October 2012 breaks can be associated with encouraging speeches of M. Draghi and O. Rehn and first IMF/ECB/EC reports about the improvement of financial stability in the euro area.

The first break that involves volatility changes in January 2010 is common to all tested countries, and may relate to a Eurostat report questioning the Greek figures on public debt and deficit. At the time, financial markets respond to the potential default of Greece, asking for a higher risk premium for holding Greek public bonds. Whereas this shock is associated with an increase in volatility for most countries, volatility actually decreases (by a small margin) for Austria, Finland and the Netherlands, which highlights a perceived heterogeneity between euro area countries at the time. The next volatility break is found in July 2011, when a new financial plan to support Greece is stated at a Eurogroup meeting. Volatility in the following segment is increased, this time for all sampled series. This can be interpreted as a sign that by then, it had become apparent that the crisis affects the financial health of the euro area as a whole, including countries whose debt policy is typically seen as stable. Finally, in October 2012 we notice a return to a quieter sample segment and lower volatility following the IMF/EC reports signaling that Ireland and Portugal will satisfy the objectives conditioning the safety plans. Again, the reduction in volatility is observed for all sample series.

The third step of the STP consists of testing for significant changes in the correlation coefficients, conditional on those previously found in the mean and the variance. Ultimately, no common break is detected in the correlation matrix of all euro area countries. The finding shows that the diffusion of shocks is not the same for all euro area countries. In a next step, we analyze pairwise correlations and study multiple breaks in each pair. Having identified suitable candidate groups, we seek to identify subsets of countries that are characterized by common breaks that involve correlations.

The common correlation change hypothesis cannot be rejected for AUS–POR–SPA at a p -value of 0.35. The correlations shift in the same direction for the first two breaks between these countries, while the last break on 2013-12-30 indicates shift contagion between Portugal and Spain but of flight-to-quality towards Austria. The second contagion club is BEL–FRA–NET at a p -value of 0.85. This second cluster stresses the potential contagion between this core of European countries, which share strong cross-border banking activities (Dexia, ING, etc.). Finally, we observe a third cluster composed of BEL–FIN–ITA at a p -value of 0.40.² In Table 3.3, we summarize the timeline of the estimated structural breaks in all parameters, including common break date estimates for contagion clubs, but without quantitative parameter estimates. Appendix 3.A contains two additional tables that report the detailed correlation changes, including break date confidence intervals and correlation estimates.

We observe correlation changes for 13 country pairs in and around August 2009, already, before the first estimated change in volatility in January 2010. All changes in correlation are positive, providing preliminary evidence of the danger of contagion. The correlation between Portugal and Spain rises by 0.318 to a high 0.841 on 2009-07-29,

²The common break date hypothesis is rejected for other cluster candidates, BEL–FIN–FRA–ITA–NET, BEL–ITA–NET, FIN–FRA–NET, BEL–FIN–FRA, and BEL–FIN–NET.

Table 3.3: Timeline of estimated structural breaks

Date	Break	Countries	Date	Break	Countries
2009-07-27	↑ <i>R</i>	FIN-SPA	2012-01-09	↑ <i>R</i>	AUS-BEL
28	↑ <i>R</i>	FIN-NET	03-01	↑ <i>R</i>	FRA-POR
08-03	↑ <i>R</i>	BEL-FIN-ITA	02	↑ <i>R</i>	BEL-FRA, BEL-NET
05	↑ <i>R</i>	ITA-NET		↓ <i>R</i>	FRA-NET
11	↑ <i>R</i>	FIN-FRA, FIN-POR,	12	<i>B</i>	euro area
		NET-POR	05-31	↓ <i>R</i>	BEL-SPA
13	↑ <i>R</i>	AUS-POR-SPA	08-07	↓ <i>R</i>	AUS-BEL
10-01	↑ <i>R</i>	ITA-SPA	10-01	↓ <i>S</i>	euro area
2010-01-22	↑ <i>S</i>	BEL, FRA, GRE, IRE,	02	<i>B</i>	euro area
		ITA, POR, SPA	08	↓ <i>R</i>	BEL-FRA-NET
	↓ <i>S</i>	AUS, FIN, NET	11-05	↓ <i>R</i>	NET-SPA
02-25	↓ <i>R</i>	BEL-FIN-ITA,	07	↓ <i>R</i>	BEL-ITA, FIN-ITA
		ITA-NET		↑ <i>R</i>	BEL-FIN
26	↓ <i>R</i>	BEL-FRA-NET	2013-02-27	↓ <i>R</i>	AUS-BEL
03-03	↓ <i>R</i>	FIN-NET, FRA-POR,	04-01	↓ <i>R</i>	FIN-NET
		NET-POR, NET-SPA	10	↑ <i>R</i>	FRA-IRE, IRE-NET
31	↓ <i>R</i>	AUS-IRE,	05-21	↓ <i>R</i>	AUS-FRA
04-02	↓ <i>R</i>	AUS-POR-SPA	06-10	↑ <i>R</i>	NET-SPA
06	↑ <i>R</i>	FIN-POR, FIN-SPA	13	↑ <i>R</i>	AUS-IRE
05-28	↓ <i>R</i>	FIN-FRA	10-28	↑ <i>R</i>	ITA-SPA
10-15	↓ <i>R</i>	FRA-IRE	11-01	↓ <i>R</i>	FIN-FRA
11-04	↑ <i>R</i>	IRE-NET	29	↓ <i>R</i>	FRA-IRE
25	↓ <i>R</i>	AUS-FRA	12-11	↑ <i>R</i>	AUS-FRA
		ITA-POR	30	↑ <i>R</i>	ITA-POR, POR-SPA
2011-01-12	↑ <i>R</i>			↓ <i>R</i>	AUS-POR, AUS-SPA
02-28	↓ <i>R</i>	BEL-SPA	2014-01-08	↓ <i>R</i>	AUS-IRE
03-15	↓ <i>R</i>	FIN-SPA	09	↓ <i>R</i>	IRE-NET
31	↓ <i>R</i>	NET-POR			
04-29	↓ <i>R</i>	ITA-NET			
07-05	↑ <i>S</i>	FRA-POR			
08-04	↓ <i>R</i>	euro area			
15	<i>B</i>	BEL-SPA			
11-03	↓ <i>R</i>	euro area			
		ITA-SPA			

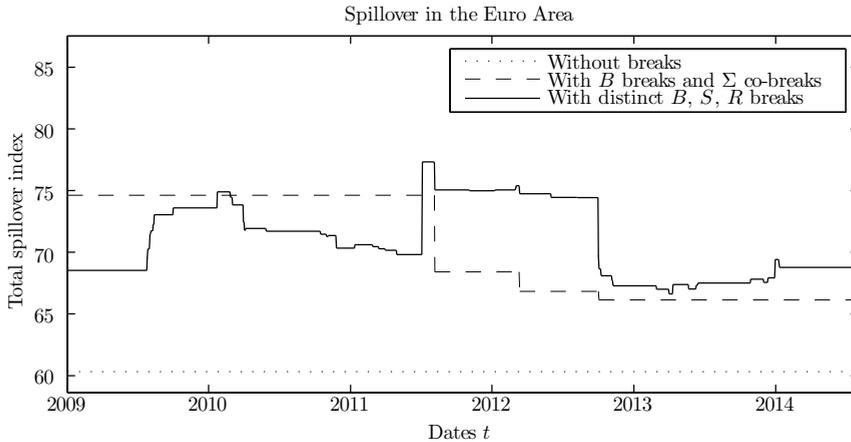
Note: Summary of all break detection results. *B*, *S* and *R* represent breaks in parameters for the conditional mean, the variance and the correlation, respectively. The arrows indicate whether the parameters have increased or decreased at the corresponding break dates.

and the correlation between Italy and Spain rises by 0.359 to a remarkable level of 0.874 on 2009-10-01. While the volatility has not yet increased at the time, European debt markets already exhibit an intensification of their interdependence. This may be a consequence of government bailouts during the Global Financial Crisis and growing debt levels throughout the euro area, and following the financial distress of countries such as Latvia or Hungary. This interpretation is supported by several reports, such as the one published by the European Commission on January 2009, giving rise to concerns about debt sustainability (EUobserver, 2009 and The New York Times, 2009).

Following the first volatility change in January 2010, in early 2010 we observe correlation changes in a total of 19 country pairs. Almost all of the breaks are associated with a decrease in correlations. The breaks clearly indicate the presence of a flight-to-quality mechanism, in favor of Austria, Finland and the Netherlands. They appear to be relatively detached from the developments in other countries, and benefit from a decrease in their yield spread and thus more affordable refinancing conditions. As previously mentioned, this result highlights a high level of heterogeneity among euro area countries, supporting the conclusions of De Santis (2012) and Metiu (2012), who find that the EMU periphery crisis mainly spread to other periphery countries as well as to France and Belgium, but not to Austria, Finland or the Netherlands. In the second half of 2010 and the first half of 2011, several further breaks in correlations occur, and so the flight-to-quality mechanisms appear to persist.

After the second volatility change in July 2011, which signals the entry into a higher volatility period, the evolution of the correlation becomes ambiguous. The STP detects only a few breaks in bilateral correlations, both positive and negative. It can be envisioned that market participants realize that the euro area is not particularly homogeneous and begin to make arbitrage between countries, leading to various effects on correlations. Similar conclusions can be drawn after the third variance shock in October 2012, after which relatively few breaks in correlation are observed. In late 2013 and early 2014, again, several breaks are observed over a short period, involving 10 country pairs. Correlations among periphery countries are rising (POR-SPA or POR-ITA), but there is also evidence of an amplification of the flight-to-quality mechanism involving mostly Austria and the Netherlands. Between 2012 and 2013, most correlation changes involving France are negative, but some shifts are also positive. Interestingly, France has two breaks with the Netherlands and Belgium, each, in 2012, and two breaks with Austria and Ireland, each, in 2013. These findings highlight the uncertainty concerning the role of France in the crisis and specific problems associated with the French economy.

A striking result is that STP does not detect any significant correlation change involving Greece and only few changes involving Ireland (only with Austria, France and the Netherlands). This means we find no evidence of shift contagion from Greece/Ireland to the rest of the European countries. It appears that the contemporaneous shock transmission to the rest of Europe is not intensified. Formally, this gives support to our idea of separating the dates of breaks in conditional means, variances and correlations. In terms of the interpretation, the relatively low interdependence of the unexplained variation of Greek/Irish yield spreads with other series may have several explanations, for example the de-leveraging of German and Dutch banks of their Greek assets (see Candelon and Bicu, 2013), and the regulation measures taken by the Irish government. The results indicate that a Greek debt default may not be characterized by contagion to the rest of

Figure 3.1: Spillover in the euro area

Note: Total spillover index, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01. Either the model contains no break, or it may contain breaks in the mean parameters B and covariance terms at the same dates, or it may contain distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP.

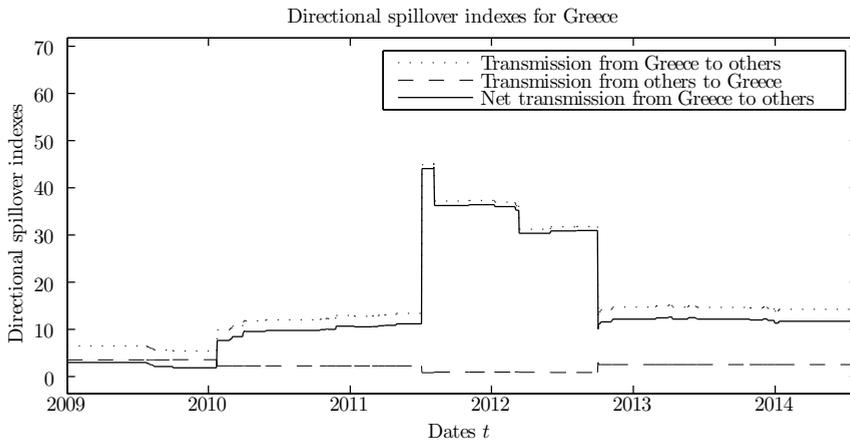
Europe. Before this result is accepted as final, it should be confirmed by complementary contagion studies.

3.3.2 Spillovers vs shift contagion

To better understand the situation of Greece, but also to gain additional insight concerning other euro area countries, we adopt the spillover approach. We calculate indexes based on the analysis of the forecast error variance decomposition. Spillover contagion is identified, when abrupt changes in spillovers are observed. We consider a 20-step-ahead forecast error variance decomposition, corresponding to approximately one trading month, which is calculated for each country (transmitting or receiving) and the euro area as a whole.³ Three VAR models are considered to compute spillover indexes. The first is the VAR model without any structural breaks. The second follows the approach of Claeys and Vařicek (2014), as the model assumes common breaks in the conditional mean and the covariance matrix of the errors. The third model is obtained by applying the STP and assumes distinct breaks in the conditional means, variances and correlations. Most of the breaks within the correlation matrix are also distinct. As a consequence, a large number of regimes are observed, which leads to notable time variation in the spillover indexes. This larger number of regimes enables a refined interpretation of the results.

Figure 3.1 shows the total spillover index TS , calculated as the sum of either all transmitted spillover indexes or all received spillover indexes, excluding each own variation share. The figure reveals that spillovers from other countries explain a large fraction of yield dynamics. When considering the VAR model without breaks, we obtain a spillover index of approximately 60, whereas the complete model with break leads to spillovers

³The results for other forecast horizons such as 10 steps are virtually identical to the ones reported here.

Figure 3.2: Directional spillover indexes for Greece

Note: Directional spillovers of Greece, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Greece to other countries, spillover transmitted from other countries to Greece, and net spillover transmitted from Greece to other countries.

explaining more than 70% of the forecast error variance. The figure illustrates how additional information can be gained from permitting distinct breaks in variance and correlation parameters, as the plot of the solid line is noticeably different than the plot of the dashed line, which represents the model with common break date assumption for all parameters. This result clearly suggests that a failure to consider structural breaks can lead to an underestimation of the importance of spillovers, and that setting unjustified restrictions on break dates affects the resulting dynamics in the spillover index.

An analysis of the plot with distinct breaks allows us to assess the importance of the individual breaks for the shock transmission over time. The index is particularly high between mid-2011 and mid-2012, with a value of around 75, before it stabilizes at a value of close to 70. During this period, the crisis was particularly intense, and the increase in mid-2011 means that spillover contagion occurred. On the other hand, no evidence of shift contagion has been found in the earlier part of the study at this time in 2011. According to the results, there has been no increase in the contemporary interdependence between series at the time, but the overall degree of spillover over time has increased. The result underlines the importance of differentiating between contagion concepts and appropriate methods for a refined analysis of the data.

Focusing on the specific case of Greece, Figure 3.2 plots the directional spillover indexes that are relevant to Greece. Corresponding figures for the other countries can be found in Appendix 3.A. We observe a positive net transmission from Greece to the other countries, indicating that the linkages to Greece are structural sources of spillovers among European countries. Beginning in 2010, several structural breaks show that the transmission from Greece has increased severely, before decreasing sharply in late 2012. This result confirms the previous findings of Metiu (2012) and Claey's and Vašíček (2014). After the decrease, the spillover transmitted from Greece is still relatively high, at $DS_{\leftarrow GRE} > 10$. The results show that over time, a large fraction of Greek yield spread fluctuations spill over to the rest of the euro area. From a political perspective, the message is thus more

pessimistic than the one gathered from the previous subsection. A Greek debt default may not immediately affect the other countries beyond the usual correlation, when the level of interdependence remains stable. Still, shocks will be transmitted over time, and the shock transmission could even intensify. On a positive note, a negative wave could give authorities some time to react.

We shortly remark on the situation of Ireland, for which the net transmission has declined sharply in mid-2011, confirming that the financial markets have recognized the successful crisis management in this country. A positive net transmission is observed for three other periphery countries (Italy, Spain and Portugal). The directional transmission indexes of these countries are lower than the Greek index (around 5 for Italy and Portugal and 2.5 for Spain). The net transmissions each decrease in the middle of the sample, and then increase in late 2012. According to the contagion analysis, the countries belong to clusters involving European core countries. This result suggests that higher tension with respect to the repayment of sovereign debt could immediately impact European core countries, and would also transmit dynamically.

3.4 Conclusion

This chapter proposes a new approach to testing for contagion that has three distinct features. First, the approach distinguishes breaks in conditional mean, volatility, and correlation using a sequence of tests, permitting distinct break dates. The tests apply the techniques developed in Qu and Perron (2007). Second, the approach is implemented within a multivariate system, instead of using bivariate systems as in many other studies. Third, although structural breaks are generally allowed to occur at distinct dates, we test whether some parameters share a common break date, using an approach similar to Perron and Oka (2011). This approach can lead to a simpler model and to more reliable testing results. It further allows the researcher to challenge restrictive assumptions concerning the equality of break dates.

The application of our approach to the recent European sovereign debt crisis offers new insight into the way the crisis has spread over the region. The implementation of the sequential testing procedure indicates that mean breaks are common between euro area countries. The same is true for the volatility breaks. These two types of breaks are observed to occur at distinct dates. Conditional on these breaks, a large number of distinct changes in correlations occur. Many of these changes cluster around a few dates, thus common breaks can be assumed to occur within a few small groups of countries. We observe some contagion clubs but notice that Greece and Ireland belong to none of them. This finding is particularly interesting for Greece, which constitutes a case study for European authorities. Such findings support the idea that it is not recommendable to assume all structural breaks are simultaneous, as it is usually done in the literature. Crisis periods begin with mean and volatility breaks, and shocks are transmitted to other markets with some delay. However, there is also evidence of sudden increases in correlation, indicating the presence of immediate shift contagion.

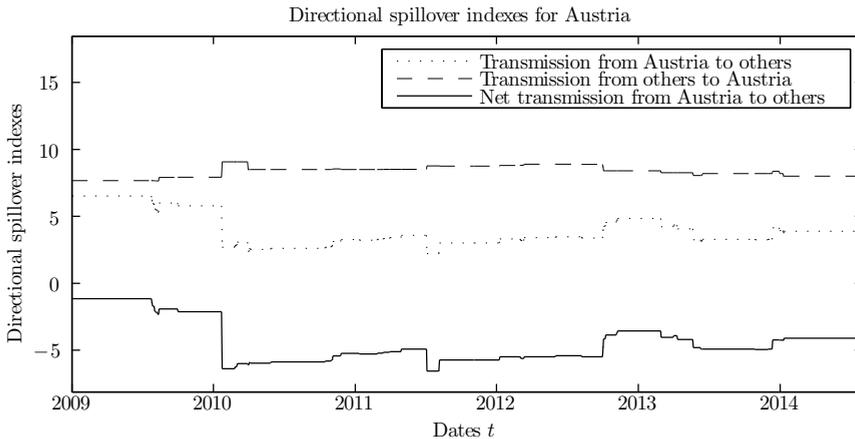
To refine our conclusions, we consider the spillover index proposed by Diebold and Yilmaz (2009, 2012). We notice that global spillovers vary significantly over time, following the break dates estimated by our sequential testing procedure. A high spillover

degree is observed between mid-2011 and mid-2012 during the peak of the turmoil. More specifically, periphery countries transmit an enormous part of their shocks to the rest of Europe, the most important country being Greece. This finding demonstrates that even if Greek distress may not necessarily be immediately contagious, it does transmit to the rest of Europe, hence confirming the findings of Claeys and Vašíček (2014).

Some guidelines for European policy can be derived from the results of this chapter. First, a default or a higher risk premium on the Greek debt may not immediately affect the yield spreads of other euro area countries above the usual interdependence. Such shocks will, however, spill over eventually, which leaves authorities little time to set up firewalls or a bail-out plan. In the case of Italy, Spain and Portugal, the spillover index is lower than that for Greece, suggesting a lower transmission. Nevertheless, these countries are potentially contagious for Europe, as correlations between these countries have abruptly increased during the crisis and stayed on high levels. The study has shown a substantial degree of heterogeneity between euro area countries, but has also detected credible threads. Appropriate regulations and specific policy actions are required to reduce the systemic risk.

3.A Appendix

Figure 3.3: Directional spillover indexes for Austria



Note: Directional spillovers of Austria, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Austria to other countries, spillover transmitted from other countries to Austria, and net spillover transmitted from Austria to other countries.

Table 3.4: Structural breaks for correlations part 1/2

	AUS	BEL	FIN	FRA	GRE
AUS		(0.646) <u>+0.231</u> <u>-0.052</u> <u>-0.320</u> (0.505)	0.497	0.644 <u>+0.182</u> <u>-0.475</u> <u>+0.461</u> 0.811	0.278
BEL	2012-01-09 [2012-05-26,2012-01-13] 2012-08-07 [2012-07-31,2012-08-07] 2013-02-27 [2013-02-27,2014-01-09]		0.287 <u>+0.471</u> <u>-0.365</u> <u>+0.039</u> 0.432	0.725 <u>-0.020</u> <u>+0.222</u> <u>-0.332</u> 0.595	0.335
FIN	-	2009-08-03 [2009-07-24,2009-08-05] 2010-02-25 [2010-02-23,2012-03-28] 2012-11-07 [2010-12-02,2013-10-18]		0.356 <u>+0.163</u> <u>+0.340</u> <u>-0.413</u> 0.446	0.175
FRA	2010-11-04 [2009-10-20,2012-05-05] 2013-05-21 [2011-05-27,2013-05-21] 2013-12-11 [2013-12-11,2014-01-09]	2010-02-26 [2009-08-06,2011-02-22] 2012-03-02 [2010-09-20,2012-03-16] 2012-10-08 [2012-09-24,2014-01-03]	2009-08-11 [2009-07-24,2009-09-14] 2010-04-06 [2010-03-03,2010-04-09] 2013-11-01 [2010-10-27,2013-11-13]		0.293
GRE	-	-	-	-	-
IRE	2010-03-31 [2009-08-31,2011-07-16] 2013-06-13 [2010-11-04,2013-05-29] 2014-01-08 [2014-01-03,2014-01-09]	-	-	2010-05-28 [2009-10-26,2011-09-16] 2013-04-10 [2010-12-21,2013-05-06] 2013-11-29 [2013-10-31,2014-01-09]	-
ITA	-	2009-08-03 [2009-07-24,2009-08-05] 2010-02-25 [2010-02-23,2012-03-28] 2012-11-07 [2010-12-02,2013-10-18]	2009-08-03 [2009-07-24,2009-08-05] 2010-02-25 [2010-02-23,2012-03-28] 2012-11-07 [2010-12-02,2013-10-18]	-	-
NET	-	2010-02-26 [2009-11-06,2011-02-22] 2012-03-02 [2010-09-20,2012-03-16] 2012-10-08 [2012-09-24,2014-01-03]	2009-07-28 [2009-07-24,2009-08-04] 2010-03-03 [2010-02-17,2012-08-16] 2013-04-01 [2010-09-28,2013-12-26]	2010-02-26 [2009-08-06,2011-02-22] 2012-03-02 [2010-09-20,2012-03-16] 2012-10-08 [2012-09-24,2014-01-03]	-
POR	2009-08-13 [2009-07-24,2009-09-03] 2010-03-31 [2010-03-05,2013-04-09] 2013-12-30 [2010-11-02,2014-01-09]	-	2009-08-11 [2009-07-24,2009-08-21] 2010-04-02 [2010-03-04,2013-12-04]	2010-03-03 [2009-08-04,2010-08-31] 2011-04-29 [2010-09-28,2011-08-10] 2012-03-01 [2011-12-08,2013-07-30]	-
SPA	2009-08-13 [2009-07-24,2009-09-03] 2010-03-31 [2010-03-05,2013-04-09] 2013-12-30 [2010-11-02,2014-01-09]	2011-01-12 [2009-07-24,2011-01-12] 2011-08-04 [2011-08-04,2011-11-09] 2012-05-31 [2012-02-24,2014-01-01]	2009-07-27 [2009-07-24,2009-08-25] 2010-04-02 [2010-02-16,2010-08-05] 2011-02-28 [2010-10-25,2013-12-06]	-	-

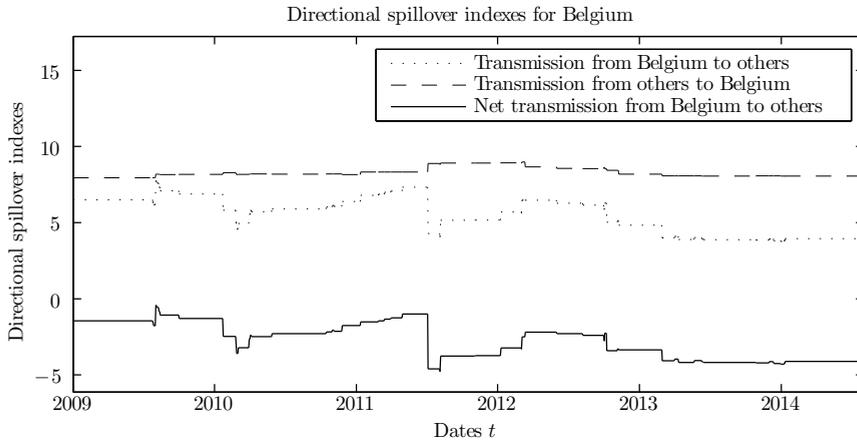
Note: Left part of the structural break table in the correlations (This part and the right part shown in Table 3.5 should be read jointly). Correlation changes obtained by applying the STP of Section 3.2. Break date point estimates, and 95% confidence intervals obtained by a block bootstrap, are reported in the lower triangle of the full table. The upper triangle of the full table reports the correlation before the first break, the changes in correlations corresponding to the breaks and the correlation after the last break.

Table 3.5: Structural breaks for correlations part 2/2

	IRE	ITA	NET	POR	SPA
AUS	0.547 <u>-0.244</u> +0.191 <u>-0.276</u> 0.217	0.511	0.561	0.611 <u>+0.064</u> <u>-0.375</u> <u>-0.068</u> 0.232	0.464 <u>+0.294</u> <u>-0.281</u> <u>-0.154</u> 0.323
BEL	0.436	0.641 <u>+0.157</u> <u>-0.053</u> <u>-0.282</u> 0.463	0.741 <u>-0.229</u> <u>+0.088</u> <u>-0.154</u> 0.446	0.434	0.643 <u>+0.217</u> <u>-0.176</u> <u>-0.264</u> 0.421
FIN	0.262	0.244 <u>+0.392</u> <u>-0.352</u> <u>-0.019</u> 0.265	0.294 <u>+0.535</u> <u>-0.224</u> <u>-0.256</u> 0.348	0.194 <u>+0.388</u> <u>-0.401</u> 0.181	0.206 <u>+0.443</u> <u>-0.247</u> <u>-0.196</u> 0.206
FRA	0.557 <u>-0.251</u> <u>+0.234</u> <u>-0.291</u> 0.249	0.555	0.708 <u>-0.036</u> <u>-0.041</u> <u>-0.159</u> 0.472	0.621 <u>-0.165</u> <u>-0.335</u> <u>+0.215</u> 0.337	0.509
GRE	0.375	0.410	0.205	0.466	0.389
IRE		0.553	0.465 <u>-0.319</u> <u>+0.340</u> <u>-0.289</u> 0.197	0.543	0.544
ITA	-		0.521 <u>+0.227</u> <u>-0.259</u> <u>-0.191</u> 0.298	0.759 <u>-0.313</u> <u>+0.286</u> 0.732	0.515 <u>+0.359</u> <u>-0.058</u> <u>+0.077</u> 0.893
NET	2010-10-15 [2009-10-13,2011-06-28] 2013-04-10	2009-09-05 [2009-07-24,2009-08-05] 2010-02-25		0.504 <u>+0.153</u> <u>-0.278</u> <u>-0.205</u> 0.174	0.743 <u>-0.363</u> <u>-0.255</u> <u>+0.282</u> 0.376
POR	-	2010-11-25 [2010-08-24,2011-02-22] 2013-12-30 [2011-06-22,2014-01-09]	2009-08-11 [2009-07-24,2009-08-11] 2010-03-03 [2010-03-03,2010-08-23] 2011-03-15 [2010-09-27,2012-06-29]		0.528 <u>+0.242</u> <u>-0.259</u> <u>+0.258</u> 0.769
SPA	-	2009-10-01 [2009-08-13,2009-12-10] 2011-11-03 [2010-04-23,2013-03-14] 2013-10-28 [2013-01-21,2013-12-25]	2010-03-03 [2009-09-18,2010-05-23] 2012-11-05 [2010-09-23,2012-11-16] 2013-06-10 [2013-05-29,2014-01-09]	2009-08-13 [2009-07-24,2009-09-03] 2010-03-31 [2010-30-05,2013-04-09] 2013-12-30 [2010-11-02,2014-01-09]	

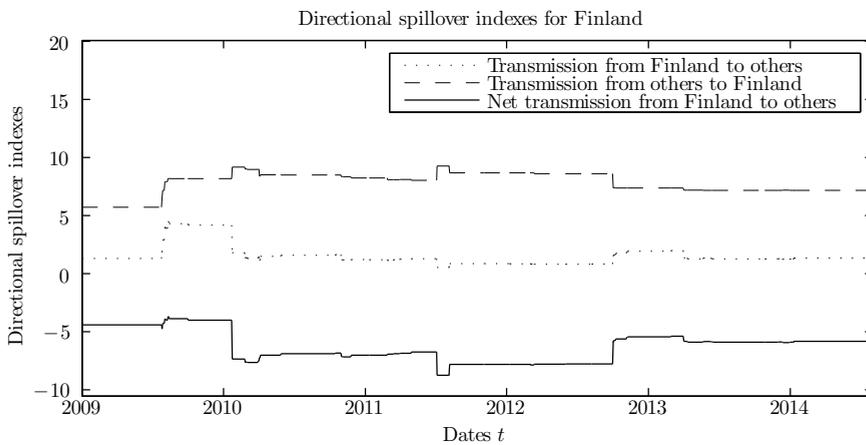
Note: Right part of the structural break table for changes in the correlations (This part and the left part shown in Table 3.4 should be read jointly). Correlation changes obtained by applying the STP of Section 3.2. Break date point estimates, and 95% confidence intervals obtained by a block bootstrap, are reported in the lower triangle of the full table. The upper triangle of the full table reports the correlation before the first break, the changes in correlations corresponding to the breaks and the correlation after the last break.

Figure 3.4: Directional spillover indexes for Belgium

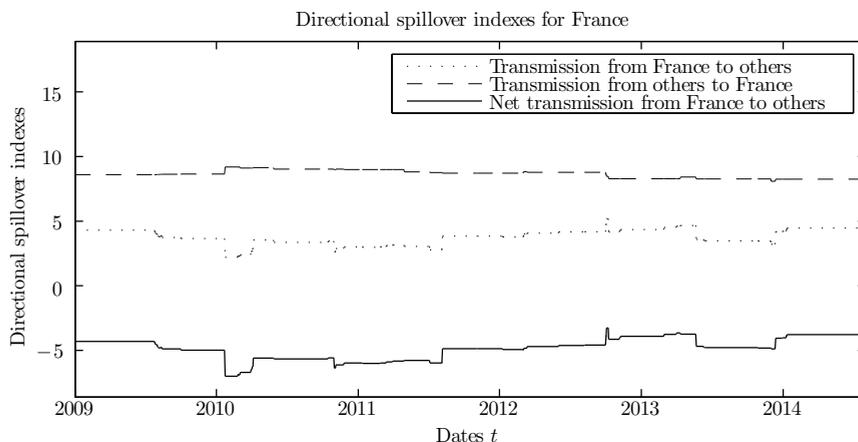


Directional spillovers of Belgium, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Belgium to other countries, spillover transmitted from other countries to Belgium, and net spillover transmitted from Belgium to other countries.

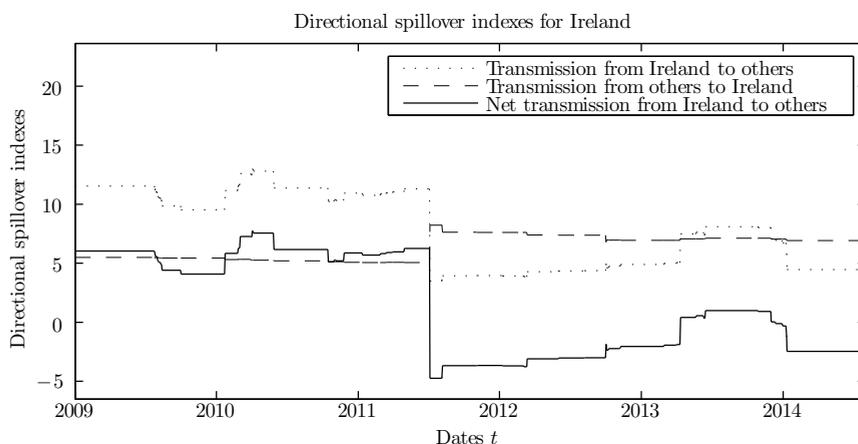
Figure 3.5: Directional spillover indexes for Finland



Note: Directional spillovers of Finland, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Finland to other countries, spillover transmitted from other countries to Finland, and net spillover transmitted from Finland to other countries.

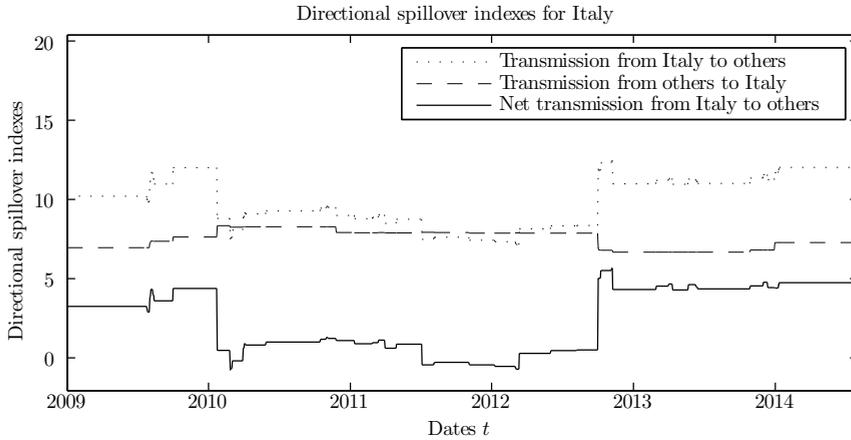
Figure 3.6: Directional spillover indexes for France

Note: Directional spillovers of France, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from France to other countries, spillover transmitted from other countries to France, and net spillover transmitted from France to other countries.

Figure 3.7: Directional spillover indexes for Ireland

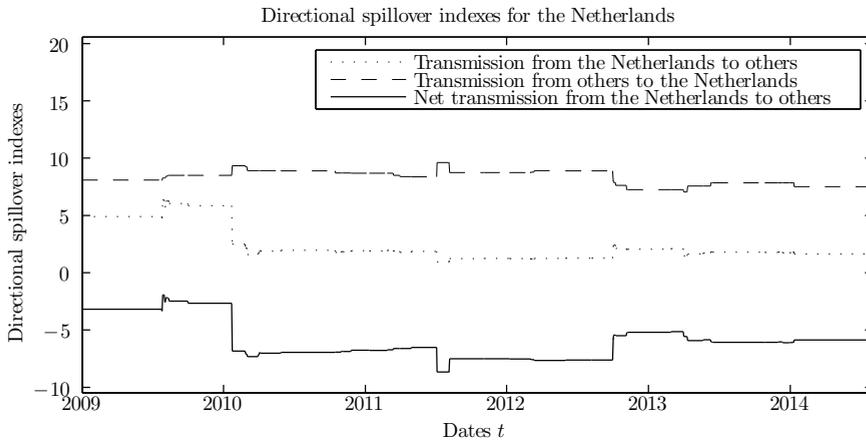
Note: Directional spillovers of Ireland, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Ireland to other countries, spillover transmitted from other countries to Ireland, and net spillover transmitted from Ireland to other countries.

Figure 3.8: Directional spillover indexes for Italy

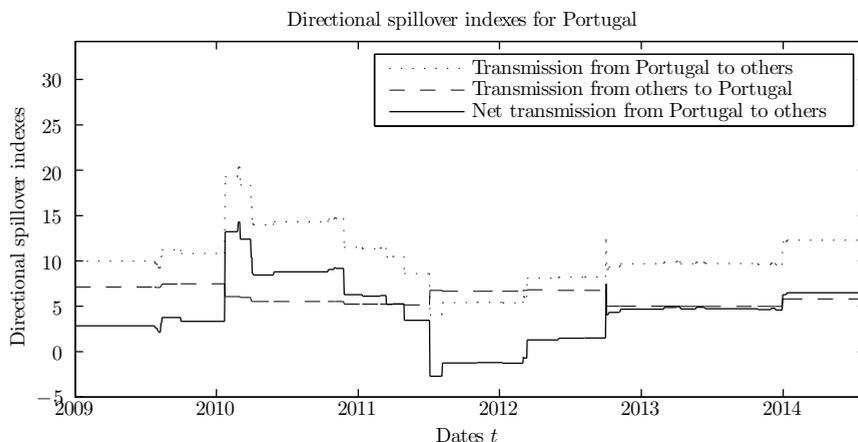


Note: Directional spillovers of Italy, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Italy to other countries, spillover transmitted from other countries to Italy, and net spillover transmitted from Italy to other countries.

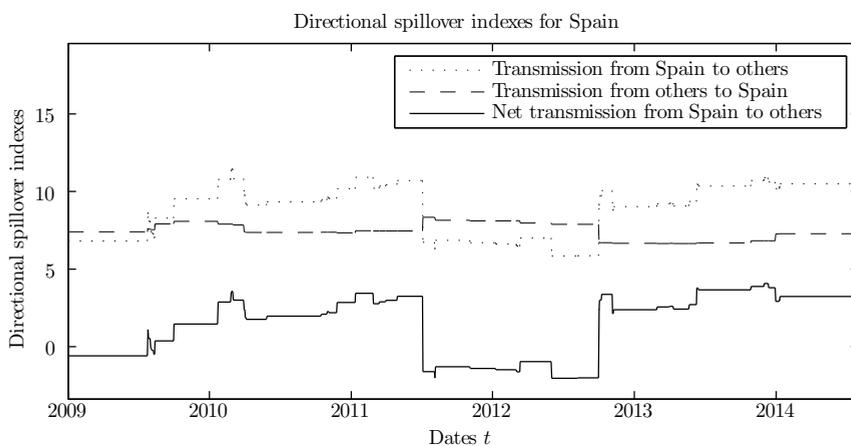
Figure 3.9: Directional spillover indexes for the Netherlands



Note: Directional spillovers of the Netherlands, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from the Netherlands to other countries, spillover transmitted from other countries to the Netherlands, and net spillover transmitted from the Netherlands to other countries.

Figure 3.10: Directional spillover indexes for Portugal

Note: Directional spillovers of Portugal, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Portugal to other countries, spillover transmitted from other countries to Portugal, and net spillover transmitted from Portugal to other countries.

Figure 3.11: Directional spillover indexes for Spain

Note: Directional spillovers of Spain, 20-step-ahead forecast, between 2009-01-02 and 2014-08-01, following a model with distinct breaks in the mean parameters B , the variances S and the correlation coefficients R according to the STP. Spillover transmitted from Spain to other countries, spillover transmitted from other countries to Spain, and net spillover transmitted from Spain to other countries.

4

A rolling window algorithm for detecting multiple structural breaks

4.1 Introduction

When testing for multiple unknown breaks in time series models, the speed of the testing algorithm can be a major concern, especially for large data samples and iterative testing. Since the location of break dates is unknown before testing, a conventional partitioning analysis must consider all possible combinations of sample divisions. This chapter proposes a rolling window algorithm that scans through the time series, one data window after the other, and performs single break optimization in each window. The collected evidence on possible break locations from this heuristic serves as a foundation for multiple break inference. Compared to conventional partitioning analysis, the computation order is reduced dramatically.

There exists a large literature on structural break tests that dates back to Chow (1960), in which a test for the significance of a structural break at a given date is presented. Refer to Chapter 2 for a summary of the developments in structural break testing. As presented in that chapter, Andrews (1993) describes the limit distribution of a group of asymptotically equivalent structural break tests that includes the maximum likelihood ratio (LR) test. Maximum LR testing has since been developed further to enable the detection of multiple structural breaks at once. For this, Bai and Perron (2003) describe the asymptotic distribution of the test statistic, which more recently has been extended by Qu and Perron (2007) to structural breaks in multivariate time series systems.

In multiple structural break testing, before the significance of a given number M of structural breaks can be assessed, break dates must be optimally distributed over the given sample data by the testing method. To evaluate and compare all possible combinations of break dates naively, without discarding certain combinations during the

optimization process, would be a heavy computational burden. Bai and Perron (2003) present a dynamic programming optimization algorithm that reduces the computation order from $\mathcal{O}(T^{M+1})$ to $\mathcal{O}(T^2)$, where T is the length of the sample. While running, the algorithm dismisses certain break date pairings that cannot be part of the final optimal solution. That way, the computational burden is reduced for multiple break detection, where $M > 1$, but the optimization algorithm detects the global optimum regardless. Related results have also been published in the signal processing literature, see for example the dynamic program with computational order $\mathcal{O}(T^2)$ of Jackson et al. (2005).

The algorithm of this chapter reduces the computation order further to a linear cost of $\mathcal{O}(T)$ by the application of rolling window break date estimation. In essence, an algorithm scans through the full sample by moving a fixed-length observation window through the data and calculating a single, local LR statistic in each window. The full set of LR statistics is used to infer multiple structural breaks. The significance is evaluated by the standard limit distribution for detecting multiple breaks in a multivariate time series system described in Qu and Perron (2007). Unlike Bai and Perron (2003), the algorithm is a heuristic that is not guaranteed to determine the global optimum. Its purpose is the reduction in execution time for $M > 1$. In a Monte Carlo study, the testing power and estimation precision and efficiency are compared to the dynamic program of Bai and Perron (2003). Due to the computational burden of the dynamic program, the size of the simulation sets is limited in this study, but the implications for even larger data sets will be clear from the reported results. The study illustrates the extent of the execution time reduction and shows that the proposed heuristic can even increase the precision of break date estimates, when parameter values shift in the same direction multiple times.

A few other studies have made computation time in structural break testing a topic. Killick et al. (2012) describe an variant of the dynamic optimal partitioning program known from Jackson et al. (2005). In this exact method, the inclusion of pruning eliminates additional partitions that cannot be optimal, which will result in a computational cost somewhere between $\mathcal{O}(T^2)$ and $\mathcal{O}(T)$, depending on the application. In Inclan and Tiao (1994), a centered CUSUM algorithm is used to find multiple changes of variance in independent observations. The power is similar to the power of a Maximum LR statistics for moderate sample sizes, and the computational burden is reduced. Apart from the restrictive assumptions, a disadvantage of the algorithm is that it must be performed repeatedly until convergence, as other structural breaks within a tested sample can bias the localization of each individual break. This issue is avoided in the method of this chapter by setting the rolling window size to twice the minimal segment length between breaks.

A low execution time can be particularly desirable in situations where large data sets must be repeatedly tested for multiple structural breaks. This situation arises for partial structural break testing, as described in Qu and Perron (2007), where parameters are restricted to remain constant to yield efficiency gains in testing. For the exact global optimum, the parameter must be estimated using the full sample, conditional on the location of break dates of the changing parameters. Conversely, break dates are estimated conditional on the constant parameter estimates. According to the authors, convergence can be achieved by an iterative procedure, which multiplies the computational burden of the detection program. A low execution time can also be useful for applications in which

data is re-sampled frequently to detect confidence intervals of the structural break test, as in a bootstrap method, or in the forecast combination method described in Pesaran and Timmermann (2007). Further, the estimation results of the rolling window algorithm can also serve as an initial guess for the dynamic program of Bai and Perron (2003) to reduce the execution time of the dynamic program. In such a rolling window-dynamic program combination method, the dynamic program is restricted to only compare the combinations of dates that lie close to the break dates found by the rolling window algorithm. An analysis of this combination method is included in this chapter's Monte Carlo study.

The chapter is completed by an empirical study on the redemption yield volatility of a large number of European long-term government bonds of the last 13 years. Generally low financing costs after the introduction of the euro, the 2008 Global Financial Crisis and the subsequent European sovereign debt crisis motivate an analysis of the volatility of government bond yields. In the early 2000's, bond returns were nearly perfectly correlated in the euro area and given the harmonization in monetary policy this did not seem surprising to market observers, according to Cappiello et al. (2006). According to Hartmann et al. (2003), small differences could not be explained by credit risk. Then the 2008 Global Financial Crisis triggered a fierce expansion of government spending in Europe and other parts of the world, causing a rapid accumulation of government debt in many countries. As a result, the joined general government gross debt of the euro area countries equaled 90.4% of GDP in 2015, up from 69.2% in 2005, according to data retrieved from Eurostat. Ever since, periphery countries of the euro area have been the focus of news and research, due to the sustainability concerns of government debt in these countries in particular.

The euro area has experienced uncertainty about the bailout responsibilities of foreign governments, and has gone through varying degrees of euroscepticism between countries and over time (Lubbers and Jaspers, 2011; Startin and Krouwel, 2013), possibly affecting yield volatility. Within the euro area, no member state is isolated from the developments in other member states, which has also been shown in the analysis of Chapter 3. Monfort and Renne (2013) find evidence of non-diversifiable risk associated with euro area bonds. For the European Union and Europe as a whole, the financial industry has become more integrated and interdependent in the last decades, with some reduction in the rate of change since the European sovereign debt crisis (Bartram and Wang, 2015; Bekaert et al., 2017). Di Cesare et al. (2012) have argued that the dynamics of interest rates during the European sovereign debt crisis indicate that new common risk factors have emerged and fears of the reversibility of the euro must be investigated. Jointly, this motivates a study with a large European data set on changes in the variation of government bond yields after controlling for global risk. Instead of investigating crisis events separately as in Abad et al. (2014), studying exogenously defined events as in Mohl and Sondermann (2013), or using a dummy for selected segments as in Chan et al. (1992), structural break tests can uncover time series dynamics endogenously. The applied multiple structural break analysis finds several changes in the yield volatility, indicating that market perceptions about bond risk are different between countries and have changed over time.

Section 4.2 describes the proposed rolling window algorithm in detail, including a formal description and an illustration of the method. The last part of the section discusses various aspects of the algorithm and describes a rolling window-dynamic program

combination method. Section 4.3 analyzes the results of a Monte Carlo study on the small sample properties of the algorithm. In this, the properties are compared to the properties of the dynamic program of Bai and Perron (2003). The empirical application is described in Section 4.4 and conclusions are found in Section 4.5.

4.2 Methodology

4.2.1 The model

Consider a univariate or multivariate time series with T data points $t = 1, \dots, T$. The data generating process is described by a linear regression model, which is subjected to a test for multiple structural breaks in its model parameters. The required assumptions for the model have been introduced in Section 2.2.1. Within a given set of dates, the linear regression model can be written as in the following,

$$\begin{aligned}
 y_t &= x_t' \beta_m + \varepsilon_t, & \varepsilon_t &\sim WN(0, \sigma_m^2) \\
 & & \text{for } t &= k_{m-1}, \dots, k_m, \\
 & & \text{for } m &= 1, \dots, M_{\max} + 1.
 \end{aligned}$$

Under the null hypothesis one set of parameters $\theta = \{\beta, \sigma^2\}$ describes the model for the full sample of the time series. Then, $M_{\max} = 0$ and the model is delimited by a set of only two ending dates, $\mathcal{T}_0 = \{0, T\}$. Under the alternative hypothesis of multiple structural breaks there are $b > 0$ pre-specified model parameters θ^b that jointly change their respective values at most M_{\max} times within the sample. Thus, under the alternative hypothesis there exist at most M_{\max} break dates k_m that separate at most $M_{\max} + 1$ regimes of the model. It is not known ex ante, at which data points the hypothesized break dates are located. Under the alternative hypothesis the set of ending dates also contains the break dates, $\mathcal{T}_M = \{0, k_1, k_2, \dots, T\}$.¹ The subset $\theta^c = \theta \setminus \theta^b$ contains all model parameters that are restricted to remain unchanged throughout the sample under both hypotheses.

Consider a minimum required data segment length ω . It says that each pair of dates from the set of ending dates must at least be separated by ω data points. This also means that the maximum number of breaks is limited to $\max(M_{\max}) = \lfloor \frac{T}{\omega} \rfloor - 1$, where $\lfloor \cdot \rfloor$ rounds a number to the next smallest integer. The possible choices for the minimum required data segment length are all $\omega = \lceil \kappa T \rceil$, for which $\kappa \in (0, 0.5]$, where $\lceil \cdot \rceil$ rounds a number to the next largest integer. A common choice in maximum LR testing is $\kappa = 0.15$, see Andrews (1993), Bai and Perron (2003), for example.

4.2.2 The rolling window algorithm

The proposed rolling window optimization algorithm involves the following 6 steps.

¹The notation follows the usual choice in structural break test literature. A break date belongs to the earlier of the two data segments that it separates. This is also why the first ending date is always $t = 0$, which is not an observation in the sample.

1. For each $k \in \mathcal{J}_\omega$, where $\mathcal{J}_\omega = \{\omega, \omega + 1, \dots, T - \omega\}$ is the set of permitted break dates, evaluate one LR statistic, given as

$$LR(k) = -2 \left(\sum_{t=k-\omega+1}^k l^*(t, \theta_k^{pre}, \theta_k^c) + \sum_{t=k+1}^{k+\omega} l^*(t, \theta_k^{post}, \theta_k^c) - \sum_{t=k-\omega+1}^{k+\omega} l^*(t, \theta_k^b, \theta_k^c) \right).$$

l^* is the Gaussian log-likelihood function at date t maximized with respect to the set of parameters within the segment. Any model parameter θ depends on k , but the estimates will not be used later. Obtain and store the resulting $T - 2\omega + 1$ LR values.

2. Identify the date $k^* \in \mathcal{J}_\omega$ with the largest associated LR value and add it to the set of candidate break dates \mathcal{K}_m .
3. Update the set of permitted break dates \mathcal{J}_ω by deleting all dates within the range $[k^* - \omega + 1, k^* + \omega - 1]$ around the date k^* identified in Step 2.
4. Repeat Steps 2 and 3 until the set of permitted break dates \mathcal{J}_ω is empty or the set of candidate break dates \mathcal{K}_m contains M_{\max} dates.
5. Denote the size of the set of candidate break dates by $\hat{M} = |\mathcal{K}_m|$. Order the contained dates chronologically to obtain ordered break date estimates \hat{k}_m , $m = 1, \dots, \hat{M}$. Let $\hat{k}_0 = 0$ and $\hat{k}_{\hat{M}+1} = T$ complete the set of estimated ending dates $\hat{\mathcal{J}}_A = \{0, \hat{k}_1, \dots, \hat{k}_{\hat{M}}, T\}$. Then calculate the LR statistic

$$LR(\hat{\mathcal{J}}_M) = -2 \left[\sum_{m=1}^{\hat{M}+1} \left(\sum_{t=\hat{k}_{m-1}+1}^{\hat{k}_m} l^*(t, \theta_m, \theta^c) \right) - \sum_{t=1}^T l^*(t, \theta^b, \theta^c) \right],$$

where the set θ_m contains the estimated values for the breaking parameters in segment $m = 1, \dots, \hat{M} + 1$.

6. Let $QP_{1-\alpha}$ denotes the $(1 - \alpha)$ -th quantile of the limit distribution for detecting multiple breaks in a multivariate time series system. If $LR(\hat{\mathcal{J}}_M) > QP_{1-\alpha}(\hat{M}, b, \kappa)$, the number of breaks \hat{M} are found to be significant and the algorithm stops. Otherwise delete the date from the set of candidate break dates \mathcal{K}_m with the smallest associated LR value found in Step 1 and repeat Steps 5 and 6.

In Step 5, the algorithm assumes prior knowledge of the break dates from earlier steps. Then the full set of available data points $t = 1, 2, \dots, T$ is used to estimate the parameters of the model under the alternative hypothesis. Thereby the associated statistic $LR(\hat{\mathcal{J}}_M)$ is calculated.

In Step 6, significance of the estimated break dates and the corresponding shifts in parameter values must be established. The QP distribution has been described in Qu and Perron (2007) for multivariate data. Critical values of the test are derived from simulations of the limit distribution. They depend on the number of estimated break dates \hat{M} , the number of breaking parameters b , and the trimming parameter κ (which determines the minimum segment length $\omega = \lceil \kappa T \rceil$). As discussed in Qu and Perron (2007), covariance parameters of the disturbances should generally be included in θ^b , and allowed to change jointly with hypothesized changes in the mean parameters of a model. Also, if structural breaks involve changes of the covariance parameters of non-

normal disturbances, the limit distribution and therefore critical values are non-standard. A formal description of the QP distribution is given in Section 2.2.2 of this book, along with instructions on the determination of critical values.

The algorithm takes the maximum number of breaks M_{\max} as given. A straightforward choice for M_{\max} is the maximum permitted number $\max(M_{\max}) = \lfloor \frac{T}{\omega} \rfloor - 1$. It then tests for the significance of that number of breaks (or a smaller number, if not sufficiently many candidate break dates exist, see Step 4). If not significant, it reduces the number of breaks by one and tests for significance again. This sequential practice is motivated by Bai and Perron (1998), Qu and Perron (2007), and most recently Bataa et al. (2013). The referenced articles also provide different methods to derive M_{\max} .

The choice to use windows of length 2ω has been made in reference to the established LR based break detection methods. As mentioned, it is necessary to set a minimum required segment length ω in all break detection testing. Without, the probability to detect breaks approaches 1, as breaks would be inferred to lie at the edges of the data sample (see Andrews, 1993). So both the proposed rolling window algorithm and all global LR maximization methods set a value for ω . In case of the proposed algorithm, this implies the use of windows of length 2ω . An objection against the use of a sample with only 2ω data points to calculate the LR statistic is effectively an objection against setting the minimum segment length to a certain value of ω in any break detection method.

4.2.3 Illustration

Figure 4.1 illustrates the algorithm for detecting $M = 2$ break dates. The data has been simulated according to the following linear regression model.

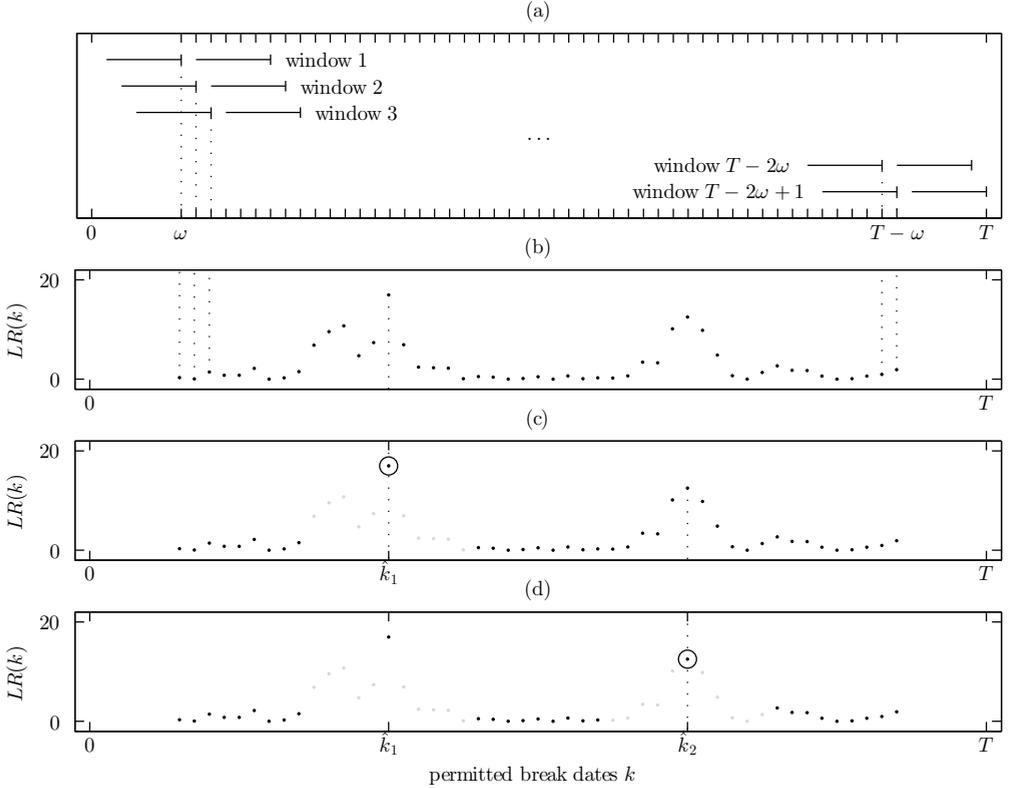
$$y_t = \begin{cases} 1 + \varepsilon_t, & \text{for } t = 1, \dots, k_1, \\ 3 + \varepsilon_t, & \text{for } t = k_1 + 1, \dots, k_2, \\ 1 + \varepsilon_t, & \text{for } t = k_2 + 1, \dots, T, \end{cases}$$

$$\varepsilon_t \sim \mathcal{N}(0, 1).$$

The test is set to find a maximum of 2 breaks in the intercept term of the model.

In its first step the algorithm runs through the data set once, as depicted in panel (a) of Figure 4.1. The algorithm begins by extracting the data of window 1, data points $t = 1, 2, \dots, 2\omega$. The data of this window is divided into two equal sized parts at the first permitted break date ω , with first segment $t = 1, 2, \dots, \omega$ and second segment $t = \omega + 1, \omega + 2, \dots, 2\omega$. Note how Figure 4.1 depicts the two ending dates ω and 2ω contained in data window 1 by ticks. The algorithm calculates the corresponding LR statistic, which measures the data fit increase of allowing the pre-specified parameters to have different values in the two data segments. The resulting LR value $LR(\omega)$ is marked as a first point in panel (b) of Figure 4.1. Then the algorithm moves to data window 2, which contains data points $t = 2, 3, \dots, 2\omega + 1$, and the next LR value is calculated, $LR(\omega + 1)$.

All considered data windows have the same length 2ω and are divided into two equal sized parts at a permitted break date. Among all permitted break dates $\mathcal{J}_\omega = \{\omega, \omega + 1, \dots, T - \omega\}$, those that generate the highest LR values within their respective data

Figure 4.1: Illustration of the rolling window algorithm

Note: The rolling window detection of $M = 2$ break dates is simulated for a small sample of $T = 60$ dates. Here, the minimum segment length is $\omega = 0.1T = 6$, so the window size is $2\omega = 0.2T = 12$. Panel (a) illustrates the progression of the rolling window algorithm to produce LR values. Panel (b) visualizes how each LR value corresponds to a unique window. Panel (c) shows the selection of the largest LR value to produce \hat{k}_1 and dismissal of surrounding LR values. Panel (d) shows the same for \hat{k}_2 .

windows are collected, as shown in panels (c) and (d) of Figure 4.1. Panel (c) depicts how the largest LR value is identified at the date that becomes break date estimate \hat{k}_1 . To satisfy the minimum segment length, the $\omega - 1$ data points closest to the break date estimate on either side are eliminated from the set of permitted break dates, and their LR values ignored in the following step. Panel (d) shows the identification of the second break date estimate.

4.2.4 Testing power and method combination

The selection of the final set of break dates depends on individual LR values at the center of each data window, and structural breaks are selected sequentially by the size of that LR value. It may be impossible to add another date to the set of break dates candidates \mathcal{K}_m , before M_{\max} breaks have been selected, depending on the distribution of break date candidates so far. This can prevent the algorithm from finding all existing

breaks, especially if they are all separated by ω dates. It also prevents the algorithm from selecting break date estimates separated by ω dates, when some of the true breaks are separated by less than ω dates. The algorithm prioritizes local evidence for a structural break, not the optimal distribution of the maximal number of breaks. Also, the order of selection into set \mathcal{K}_m can benefit the interpretation of structural breaks, by indicating which structural break has been selected first and which last. The order of selection will be discussed in the empirical application of this chapter. For an assessment of the significance of the ordering of structural break selection, a more advanced method than the current RW heuristic is required that incorporates the empirical context of each break date estimate.

In the following section of Monte Carlo simulations, the performance of the rolling window algorithm is compared to the dynamic LR maximization method of Bai and Perron (2003). In the simulations, both methods are set to begin with $M = M_{\max}$ hypothesized break dates. Only if not found to be significant, each method considers one less break date, until at last $M = 1$ break date is considered. By setting the competing programs to run the same way, conceptually, the performance evaluation can focus on differences in how the methods perform that same task. Other simulations have been run in which the dynamic LR maximization method has been set to start with searching for one break, and to only move to a two-break hypothesis if a single break is found to be significant. The results are available, but not reported in the study. If the dynamic program starts with one hypothesized break first, execution time will be longer and testing power will be lost. The reason is that the method can get stuck, when only two breaks are significant, for example, but no single break in the full sample is significant. This can happen for shift reversals, where a parameter first increases in value and then decreases. The proposed rolling window algorithm can uncover such cases immediately by producing local LR values for each data window.

If structural break detection is initiated with the correct number of breaks, the rolling window algorithm is not guaranteed to find the global LR maximum. This is because the model parameters are determined by maximum likelihood estimation that is conditional on the set of detected break dates, where the set of break dates has been determined by a heuristic. In contrast, the global LR maximization method of Bai and Perron (2003) is guaranteed to detect the data points with the strongest empirical indication of structural breaks. Subject to the same critical values, the global LR maximization method must at least be as powerful as the proposed rolling window heuristic. This probable reduction in test power is the main trade-off for the heuristic's substantial decrease in computational order and execution time. The following section of Monte Carlo simulations will study the impact on testing power for this scenario.

A natural extension of the rolling window method is to use it as an initial guess method before applying the dynamic LR maximization method of Bai and Perron (2003). For this, Step 5 of the algorithm is replaced. At the end of Step 4, for each date k_m^* in the set of candidate break dates \mathcal{K}_m , a small interval of adjacent dates is selected, $[k_m^* - \lambda, k_m^* + \lambda]$. Then the LR maximization method of Bai and Perron (2003) is applied to detect optimal break dates $\hat{\mathcal{J}}_M$ among these restricted intervals of permitted break dates. The algorithm continues with Step 6 to assess the significance of the structural break model.

In this combination of methods, a restriction is made on the set of permitted break dates, but then the full data is used to evaluate the positioning of possible breaks, not only data windows. If the rolling window algorithm locates the breaks sufficiently close to the globally optimal break dates, the combined method produces the same estimation results as the global optimization method. Note that if $\lambda > 0.5\omega$, each resulting set of permitted break dates is larger than the minimum required segment length. Then, these sets could overlap, if the initially guessed break dates were positioned close to one another. This is not an issue, if the minimum segment rule is consistently applied in the LR maximization. In practice, smaller values of λ are more efficient in terms of execution time. The performance of the combined method is included in the results of the Monte Carlo study in the following section.

4.3 Monte Carlo simulations

4.3.1 Introduction

To determine the performance of the proposed rolling window algorithm in small samples, data has been simulated using various specifications of the data generating process (DGP). For each specification, 1000 data sets have been simulated and all simulated DGP's can be described by a simple linear regression model of the following form,

$$\begin{aligned} y_{i,t} &= x_t \beta_m + \varepsilon_{i,t}, \\ x_t &\sim \mathcal{N}(0, 1), \\ \varepsilon_{i,t} &\sim \mathcal{N}(0, 1), \end{aligned}$$

where $i = 1, \dots, n$, and n is the number of contemporaneous observations, where $t = 1, \dots, T$, and T is the number of observations per series, and where $m = 1, \dots, M+1$, and M is the number of structural breaks. All breaks occur in the coefficient β of the predictor vector x . The segment ending dates are given by $\mathcal{J} = k_0, k_1, \dots, k_M, k_{M+1}$, where $k_0 = 0$ and $k_{M+1} = T$ delimit the full sample. In the m -th data segment, which contains the interval of dates $[k_{m-1} + 1, k_m]$, all n series share the same coefficient value β_m . At each break date between two regimes, all coefficient values shift by $\Delta_m = \beta_{m+1} - \beta_m$. If $n > 1$, the correlation between distinct disturbance terms is $\rho = 0$.

Data for each considered DGP specification is repeatedly simulated for this study.² Each one of the resulting data samples is subjected to three methods of break detection. As a setting, all detection methods correctly identify that the structural breaks occur in the mean parameter β . They also assume prior knowledge of the covariance of the disturbance terms. Let $M_{\max} = M$, so that the maximum considered number of breaks equals the true number of breaks. The three considered methods are described in the following.

²The simulations and tests were programmed in MATLAB 2014a, The MathWorks, Inc., Natick, Massachusetts, United States. In all types of tests the parameters of a given data segment were estimated using local optimizer function `fminunc`, with quasi-Newton optimization algorithm and full sample OLS estimates as initial guesses. The codes were run on a 64-bit Microsoft Windows 7 system using a single processor of an Intel quad-core CPU with 3.2GHz and 8GB RAM.

1. DP: For a given number of hypothesized break dates M , the algorithm proposed in Bai and Perron (2003) recursively optimizes the positioning of break dates k_m , $m = 1, \dots, M$, using a dynamic program.
2. RW: The proposed rolling window algorithm, described in Section 4.2.
3. RW+DP: In this combination of methods, break date candidates from the rolling window method are used as initial guesses for the recursive dynamic program of Bai and Perron (2003). For each initial break date guess k_m^* , the set of permitted break dates $[k_m^* - \lambda, k_m^* + \lambda]$ contains a small interval of data points surrounding the respective rolling window estimate.
 - $\lambda = 0.1\omega$.
 - $\lambda = 0.5\omega$.
 - $\lambda = \omega$.

The study reports a number of performance indicators, divided into three categories: Testing power, break estimation and execution time. A first indicator is the break-detection frequency I^A , which is given for the usual confidence levels $1 - \alpha = 0.99, 0.95, 0.90$. It indicates how often a break detection method could find the correct number of structural breaks M , relative to the number of simulations. All remaining indicators (such as $\text{mean}(\hat{k}_m)$) are calculated using qualified simulations, only. This means that only simulations are used for which each method has detected the correct number of breaks at 95% confidence.

4.3.2 General results

In this section, simulation results of only one DGP specification are reported, so that the results can be analyzed in detail. The results will show that RW is typically less powerful than DP, but that the quality of structural break estimates is similar between methods, and that the execution time of the RW algorithm is relatively small even in small data samples. Data with two small shifts $|\Delta_m| = 0.4$ has been simulated. Formally, it holds that $T = 200$, $n = 1$, $M = 2$, $k_1 = [T/3] = 67$, $k_2 = [2T/3] = 133$, $\beta_1 = 1$, $\beta_2 = 1.4$, $\beta_3 = 1$. The trimming parameter is $\kappa = 0.15$. This reference specification will later be adjusted in the individual properties to evaluate the individual impact. For now, Table 4.1 reports how the three methods have performed for data generated and tested according to the reference specification.

The first observation is that DP is a more powerful detection method than RW. DP detects all $M = 2$ breaks in $I_{0.95}^A = 55.7\%$ of the simulations at a confidence level of 95%, while $I_{0.95}^A = 46.4\%$ for RW. The difference in power is similar at other confidence levels and can be explained by looking at the following four rows of the table. Values of $LR(\hat{\mathcal{T}}_A)$ are lower for RW than for DP in the mean, median, minimum and maximum. RW is not guaranteed to find the LR maximum in a given simulation. Given the relatively small parameter shift, this affects the overall detection power of RW.

RW+DP is at least as powerful as RW for any positive value of λ . For larger λ , RW+DP can select from larger sets of potential break dates. This increasingly enables RW+DP to reposition the RW break dates towards the globally optimal positions, so

Table 4.1: General testing results

			DP	RW	RW+DP		
					$\lambda = 0.1\omega$	0.5ω	ω
Power	I^A	$\alpha = 0.01$	0.295	0.230	0.250	0.279	0.285
		0.05	0.557	0.464	0.484	0.523	0.537
		0.10	0.689	0.584	0.611	0.649	0.664
-----			-----			-----	
		mean($LR(\hat{\mathcal{J}}_M)$)	18.61	17.19	17.64	18.31	18.49
		median($LR(\hat{\mathcal{J}}_M)$)	17.81	16.42	16.90	17.46	17.77
		maximum($LR(\hat{\mathcal{J}}_M)$)	37.50	36.79	36.79	37.50	37.50
		minimum($LR(\hat{\mathcal{J}}_M)$)	12.17	12.08	12.08	12.17	12.17
Time		mean(time) in sec	17.24	0.96	1.11	3.34	9.04
		median(time) in sec	17.07	0.94	1.10	3.54	9.45
		maximum(time) in sec	24.54	1.19	1.45	4.26	12.59
		minimum(time) in sec	16.51	0.91	1.03	1.53	2.69
Estimation	mean($\hat{\Delta}_m$)	$m = 1$	0.64	0.59	0.60	0.61	0.62
		2	-0.64	-0.59	-0.60	-0.62	-0.63
	sd($\hat{\Delta}_m$)	$m = 1$	0.24	0.27	0.27	0.28	0.26
		2	0.24	0.27	0.27	0.27	0.26
	mean(\hat{k}_m)	$m = 1$	67.632	66.547	66.664	66.785	67.403
		2	131.782	132.948	132.836	132.720	132.545
	median(\hat{k}_m)	$m = 1$	67	67	67	67	67
		2	133	133	133	133	133
	sd(\hat{k}_m)	$m = 1$	16.558	17.737	17.638	17.567	16.940
		2	17.653	18.544	18.714	19.060	19.092
	$\hat{k}_m = \hat{k}_m(DP)$	$m = 1$	1	0.353	0.591	0.875	0.950
		2	1	0.364	0.625	0.873	0.927
	$\hat{k}_m = \hat{k}_m(RW)$	$m = 1$	0.353	1	0.515	0.397	0.369
		2	0.364	1	0.504	0.375	0.362
	$\hat{k}_m = k_m$	$m = 1$	0.080	0.067	0.073	0.078	0.080
		2	0.047	0.045	0.052	0.056	0.050
$ \hat{k}_m - k_m < \hat{k}_m(DP) - k_m $	$m = 1$	0	0.304	0.192	0.047	0.022	
	2	0	0.323	0.179	0.054	0.019	
$ \hat{k}_m - k_m < \hat{k}_m(RW) - k_m $	$m = 1$	0.343	0	0.254	0.317	0.334	
	2	0.313	0	0.228	0.295	0.300	

Note: Measuring the performance of three break detection and estimation methods in a reference DGP specification of $T = 200$, $n = 1$, $M = 2$, $k_1 = [T/3] = 67$, $k_2 = [2T/3] = 133$, $\beta_1 = 1$, $\beta_2 = 1.4$, $\beta_3 = 1$. The trimming parameter is $\kappa = 0.15$. The performance indicators are divided into the categories of testing power, execution time and structural break estimation. For the calculation of all measures below the dashed line, only data from simulations where all methods have identified the correct number of breaks at confidence level $1 - \alpha = 0.95$ has been used.

that the correct number of structural breaks is detected in more simulations. When λ is as large as the minimum segment length ω , the difference in detection frequencies between DP and RW+DP shrinks to $\Delta I_{0.95}^A = 2.0\%$. Correspondingly, the largest LR value among all simulations, $LR(\hat{\mathcal{J}}_A) = 37.50$, is not found in RW, but in DP and in RW+DP at $\lambda = 0.5\omega, \omega$. In that particular simulation, RW+DP at $\lambda = 0.5\omega$ has adjusted the estimation of the break dates towards the global optimum, because the first break date guesses derived from RW had been sufficiently close.

The next category of performance indicators concerns the time needed to execute the break detection programs. Execution time includes break date positioning, significance testing with given critical values and model estimation, not the generation of the data sample. As usual, only simulations are included where the correct number of breaks is identified, so that the time does not reflect a potential re-run of the detection program. The execution time of RW+DP includes the execution time of the initial guess estimation according to RW. Table 4.1 reports mean, median, maximum and minimum execution times in seconds. The most striking feature of the execution time statistics are the differences between methods. They can be ranked, from fastest to slowest method according to the mean execution time:

1. RW, with $\text{mean}(\text{time}) = 0.96$ sec, or 5.54% of DP.
2. RW+DP at $\lambda = 0.1\omega$, with $\text{mean}(\text{time}) = 1.11$ sec, or 6.44% of DP.
3. RW+DP at $\lambda = 0.5\omega$, with $\text{mean}(\text{time}) = 3.34$ sec, or 19.37% of DP.
4. RW+DP at $\lambda = \omega$, with $\text{mean}(\text{time}) = 9.04$ sec, or 52.44% of DP.
5. DP, with $\text{mean}(\text{time}) = 17.24$ sec.

Using RW instead of DP reduces the execution time significantly. The computation order is drastically reduced, even though the sample size is relatively small at $T = 200$.

The lower section of Table 4.1 contains measures that relate to the accuracy and precision of the structural break estimation. Each measure is reported for the structural break at $k_1 = 67$ and for the second break at $k_2 = 133$, separately. It is noticeable that all methods lead to a substantial overestimation of the parameter shift size, in the mean. The smallest overestimation is recorded for RW, where both parameter shift estimates are 47.5% larger than the true shifts $|\Delta_m| = 0.4$. There is some variation between simulations, as values for the standard deviation of parameter shift estimates $\text{sd}(\hat{\Delta}_m)$ range between 0.24 and 0.27. Overestimation is explained as a result of the data fit optimization method, as the maximum LR statistic selects dates with the strongest empirical indication of a parameter shift.

The next set of indicators of Table 4.1 show that break date estimation is generally unbiased, but specific estimates regularly differ from the true break date. The values for $\text{mean}(\hat{k}_m)$ are close to $k_1 = 67$ and $k_2 = 133$ for all methods and all median(\hat{k}_m) are equal to the correct break date. RW is slightly less precise than DP in the reported simulations, with larger values for the standard deviation of break date estimates $\text{sd}(\hat{k}_m)$. An important conclusion, however, is that the simulations do not indicate that break date estimation precision suffers much from an application of RW instead of DP.

The next two descriptive measures of Table 4.1 illustrate how the combination of methods RW+DP repositions initial RW estimates. In 35.3% of qualified simulations, DP and RW produce the same estimate \hat{k}_1 , and in 36.4% they produce the same estimate \hat{k}_2 . With larger λ , RW+DP increasingly detects the same breaks as DP. When $\lambda = \omega$,

$\hat{k}_1 = \hat{k}_1(DP)$ in 95.0% and $\hat{k}_2 = \hat{k}_2(DP)$ in 92.7% of qualified simulations. This result is intuitive, as larger values of λ increase the domain of dates that may include the globally optimal break dates found by DP. Notably, even for small intervals λ , the equality of break date estimates increases sharply. For example, it increases from 35.3% to 59.1% for k_1 , for $\lambda = 0.1\omega$. This observation shows that a combination method RW+DP with a small value of λ can be associated with an execution time that is much shorter than the execution time of DP and with a large fraction of equal break date estimates.

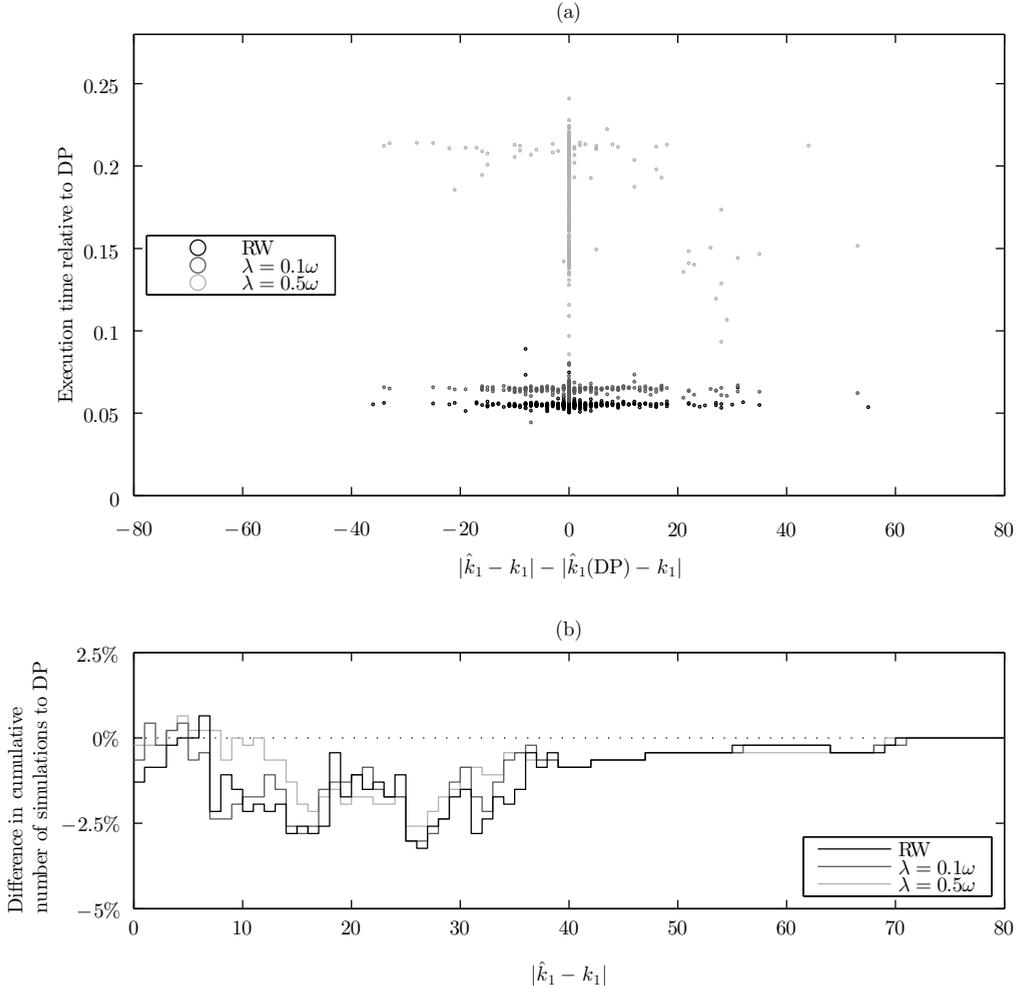
The last three descriptive measures of Table 4.1 enable a closer look at how RW and DP compare with respect to estimating the correct break date. DP estimates the correct break position slightly more often than RW, as $\hat{k}_1 = k_1$ is true in 8.0% of qualified simulations for DP and in 6.7% of qualified simulations for RW. RW+DP at $\lambda = \omega$ has equally often as DP identified that first break date. Interestingly, $k_2 = 133$ has been most often identified correctly by RW+DP at $\lambda = 0.5\omega$ with 5.6% of qualified simulations. The final rows of Table 4.1 enable a comparison of the frequency that a method has estimated break dates that lie closer to the true position than DP, and RW, respectively. RW has produced a more accurate estimate than DP in 30.4% of k_1 estimations, and in 32.3% of k_2 . It can be seen that DP has slightly less frequently than RW located k_2 closer to the true break date. Although the differences are not significant, the results at least show that DP is not generally more accurate than RW.

Figure 4.2 illustrates some of the previous findings and provides additional detail about individual simulation results for the first structural break at $k_1 = 67$. A corresponding figure for the second break can be found in the Appendix 4.A. The scatter plot in panel (a) of the figure illustrates three main findings. First, RW is a lot quicker than DP in all simulations, as all black markers lie below the 10% value, and an increase in λ leads to an increase in execution time for the combined method RW+DP. For $\lambda = 0.5\omega$, the algorithm execution still only takes a fraction of the time it takes to execute DP. Second, no method is guaranteed to produce more accurate break date estimates than the others. In a substantial amount of simulations, RW-based methods have found an estimate closer to the true break date k_2 than DP, as can be seen by observations that lie to the left of 0 in the scatter plot. Third, often DP and RW-based methods find break dates that have equal distance to k_1 , as can be seen from the accumulation of observations at 0 in the scatter plot. This does not mean, however, that these equal break date estimates match the true break date.

Lower panel (b) illustrates that DP typically finds break dates that lie closer to k_1 than RW-based break date estimates do. For most of the absolute differences $|\hat{k}_1 - k_1|$, the plotted values have a negative value on the y-axis. This means that there are fewer simulations in which RW-based methods have detected a break date with at most a distance of $|\hat{k}_1 - k_1|$ from the true break than there are simulations in which DP has detected a break date with at most that same distance. However, the difference in the number of simulations is relatively small. The largest difference for RW is recorded for $|\hat{k}_1 - k_1| = 26$ with -3.23% . That is, in 86.42% of the simulations, DP has detected break dates that lie at most 26 dates away from k_1 , compared to RW with $86.42\% - 3.23\% = 83.19\%$. Conversely, RW has found break dates that lie at most $|\hat{k}_1 - k_1| = 6$ dates away from k_1 in 0.65% more simulations than DP.

To summarize the results in the simulation study so far, DP is more powerful in finding structural breaks, and DP produces statistically more often more accurate break date

Figure 4.2: Differences in break date estimation with shift reversal 1/2



Note: Differences in break date estimation between DP and RW-based methods, when data is simulated according to the reference DGP specification of $T = 200$, $n = 1$, $M = 2$, $k_1 = \lceil T/3 \rceil = 67$, $k_2 = \lceil 2T/3 \rceil = 133$, $\beta_1 = 1$, $\beta_2 = 1.4$, $\beta_3 = 1$. The trimming parameter is $\kappa = 0.15$. Upper panel (a) shows a scatter plot, with the difference in absolute deviations of \hat{k}_1 from k_1 between RW-based methods and DP on the x-axis, and with the execution time relative to DP on the y-axis. When a dot lies to the left of zero, the RW-based method has found a break date closer to k_1 than DP. Lower panel (b) plots the difference between RW-based methods and DP in the number of simulations that k_1 has been estimated to lie as close to the true value as specified on the x-axis, or closer. If a y-value is above 0%, then the RW-based method has more often than DP produced break date estimates that lie at most as many dates away from k_1 as indicated on the x-axis.

Table 4.2: Mean execution time comparison

	DP	RW	RW+DP		
			$\lambda = 0.1\omega$	0.5ω	ω
$T = 100$	4.48	0.48	0.57	1.18	2.47
200	17.24	0.96	1.11	3.34	9.04
1000	415.93	4.85	7.45	62.58	217.74
$\kappa = 0.10$	27.34	1.07	1.16	2.22	5.06
0.15	17.24	0.96	1.11	3.34	9.04
0.20	9.13	0.81	1.05	4.45	11.06
$n = 1$	17.24	0.96	1.11	3.34	9.04
2	19.41	1.13	1.32	3.90	10.47
4	21.86	1.42	1.63	4.64	12.48
$M = 2$	17.24	0.96	1.11	3.34	9.04
3	18.63	0.96	1.24	5.22	14.45
4	11.12	0.98	1.35	6.36	17.90

Note: Mean execution time in seconds for the three break detection and estimation methods. The simulation specification is adjusted with respect to the reference specification in selected properties T , κ , n , and M , where $k_m = \lfloor mT/(M+1) \rfloor$. the calculation of the mean execution time, only data from simulations where all methods have identified the correct number of breaks at confidence level $1 - \alpha = 0.95$ has been used.

estimates than RW-based methods. The differences in break date estimation are mostly small, uncertain and tend to decrease with larger values of λ , when RW is combined with DP. As expected, the clear advantage of the RW heuristic lies in the low relative execution time, which will be more closely studied in the following part for data that is simulated with different DGP specifications.

4.3.3 Execution time

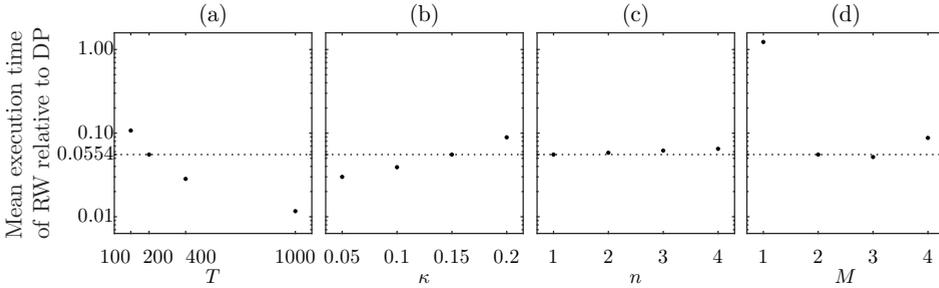
In this section some properties of the reference setting are adjusted one by one to study their respective impact on mean execution time. The section will illustrate how much execution time can be saved by the application of RW, especially when data sets become larger than $T = 200$. The size of the simulation sets is limited in this study, because of the computational burden of DP, but the implications for even larger data sets will be clear from the reported results. Table 4.2 reports the mean execution time for variations of the reference specification in the following properties: Sample length T , data trimming κ , number of series n , and number of breaks M . The table reports two adjustment of each parameter, and the reference specification is included in each group for ease of comparison.

A reduction in the sample length to $T = 100$ reduces the mean execution time, generally. Conversely, an increase in T increases the mean execution time, as a larger set of data points increases the number of operations in any method. For the recursive dynamic program optimization of DP the considered partitions become larger in the mean, but also the number of relevant combinations of partitions becomes larger. At computation order $\mathcal{O}(T^2)$, increasing the sample size by a factor of 5 from $T = 200$ to $T = 1000$, causes an increase in DP mean execution time by roughly a factor of $(1000/200)^2 = 25$ from 17.24 seconds to 415.93 seconds. In contrast, at computation order $\mathcal{O}(T)$, RW sees an increase in the mean execution time by only roughly a factor of $(1000/200) = 5$. The quadratic effect seen for the recursive dynamic program optimization can also be observed for RW+DP, albeit, on a smaller level, because the domain of potential break dates is reduced by the initial guess.

Consider a reduction of the test trimming from $\kappa = 0.15$ to $\kappa = 0.10$, as reported in the second group of mean execution times of Table 4.2. The range of potential break dates is increased, when κ is reduced. This increases the mean execution time of DP by 58.6%. For the rolling window algorithm, there are two opposing effects on the execution time. The number of rolling windows $T - 2\omega + 1$ increases, but the size of each window 2ω decreases, as $\omega = \lceil \kappa T \rceil$. For RW, the resulting increase in execution time is relatively small at 11.5%, and so the reduction of execution time relative to DP increases with smaller κ . Although the execution time of RW+DP includes the execution time of RW, for $\lambda = 0.5\omega$, ω the mean decreases by the reduction in κ . The additional effect relates to the choice of λ as a fraction of ω . A reduction of κ reduces the domain of permitted break dates around each initial guess $[\hat{k}_m - \lambda, \hat{k}_m + \lambda]$. It is particularly recognizable for $\lambda = \omega$, where the mean execution time is decreased by 44.0% from 9.04 seconds for $\kappa = 0.15$ to 5.06 seconds for $\kappa = 0.10$.

The impact of increasing the number of series n on the execution time is similar for all methods. As the data system becomes bigger, the mean execution time increases. The increase is roughly proportional to the increase in n for all methods, and the ranking remains the same as in the reference setting where $n = 1$.

Figure 4.3: Relative execution time of the rolling window algorithm



Note: Mean execution time of RW relative to the mean execution time of DP. The execution includes break date positioning, significance testing, and model estimation. The y-axis is logarithmic. Apart from the explicit adjustments, data has been simulated with reference specification $T = 200$, $n = 1$, $M = 2$, and $\kappa = 0.15$. This reference setting is marked by a dotted line.

Table 4.2 further reports the mean execution times when $M = 3$ and $M = 4$, in addition to the familiar setting that $M = 2$. For both $M = 3$ and $M = 4$, RW is still the fastest break detection method. The first observation is that RW is hardly affected by the number of breaks. The mean execution time increases remains 0.96 seconds for $M = 3$ and then increases to 0.98 seconds for $M = 4$. Evidently, the largest fraction of the execution time is spent for step 1 of the algorithm, the calculation of LR statistics in rolling data windows. The subsequent determination of M break dates candidates takes only a fraction of the overall time. If in a given application the maximum permitted number of breaks are not found to be significant, the hypothesized number of breaks could be adjusted downward very quickly. The second interesting result is the substantial reduction in mean execution time of DP from 18.63 seconds for $M = 3$ breaks to 11.12 seconds for $M = 4$ breaks. It is the only situation where DP becomes a quicker method than RW+DP at $\lambda = \omega$. An explanation is that the mean length of

potential data partitions becomes relatively small when $M = 4$ and fewer segments of the sample can be combined such that each segment still has minimal length. This lowers the computational burden for DP, but still, RW requires only a fraction of the execution time.

The comparison between DP and RW is illustrated in Figure 4.3, where the mean execution time of RW relative to DP is plotted. The four panels of the figure cover variations of the same properties as reported in Table 4.2, with (a) sample length T , (b) trimming κ , (c) number of series n , and (d) number of breaks M . The graphs underline the relative computational efficiency of RW as described before.

4.3.4 Structural break estimation

This part of the analysis of the Monte Carlo simulation study analyzes the differences in structural break estimation between DP and RW in a large set of different DGP specifications. The combined method RW+DP is not studied in this section, because it is known to reposition a fraction of the break date estimates of RW towards the DP locations, independent of the DGP specification. It will be shown that DP continues to find more accurate break dates than RW in most cases, but that RW outperforms DP in the specific situation where parameters shift multiple times in the same direction. Table 4.3 reports results for individual adjustments of $T = 200$, $\kappa = 0.15$, $n = 1$, $M = 2$, $\Delta_m = 0.4$ for odd m and $\Delta_m = -0.4$ for even m .

In its upper part, Table 4.3 reveals that both methods become more powerful for larger values of T and that the difference in power stays constant until 100% of the breaks are detected. It is important that the power of RW goes up as quickly, as larger data samples are precisely the situations where RW drastically reduces the execution time of structural break detection. Another observation concerns the mean estimated parameter shift $\text{mean}(\hat{\Delta}_m)$, which tends to be notably larger than the true shift. In the reference setting with $\Delta_1 = 0.4, \Delta_2 = -0.4$ and $T = 200$, the estimates are more than 60% larger. A larger number of observations T reduces this overestimation, so that $\hat{\Delta}_1 = 0.41, \hat{\Delta}_2 = -0.42$ for RW at $T = 1000$.

Mean break date estimates are close to the true breaks $k_1 = [T/3]$ and $k_2 = [2T/3]$ for all T and for either method. The standard deviation $\text{sd}(\hat{k}_m)$ measures the change in precision for the two methods for larger data sets. While the absolute values for $\text{sd}(\hat{k}_m)$ increase in T , the standard deviation becomes relatively smaller compared to the sample length. For $T = 100$, the standard deviation is larger than 10 dates in either method, or more than 10% of the sample size. For $T = 1000$, the standard deviation has reduced to below 5% of the sample size in case of RW, and even below 4% in case of DP. The measure in the last columns of the table, I_m^* , counts the number of times that a method has detected a break date closer to the true date than the other. As already analyzed for the reference DGP specification with $T = 200$, DP typically produces more accurate estimates more often than RW. There appears to be a weak trend for DP to become increasingly often more accurate than RW for larger T , while the relative number of times that RW is more accurate than DP remains constant. Overall, because RW uses only a fraction of each sample to evaluate break date candidates, it benefits less quickly than DP from an increase in the sample size.

CHAPTER 4. A ROLLING WINDOW ALGORITHM FOR DETECTING MULTIPLE STRUCTURAL BREAKS

Table 4.3: Structural break estimation comparison

	m	$I_{0.95}^A$		mean($\hat{\Delta}_m$)		mean(\hat{k}_m)/ T		sd(\hat{k}_m)		I_m^*	
		DP	RW	DP	RW	DP	RW	DP	RW	DP	RW
$T = 100$	1	0.288	0.213	0.74	0.70	0.35	0.35	12.17	12.12	0.235	0.263
	2			-0.75	-0.71	0.66	0.67	10.75	10.93	0.249	0.225
200	1	0.557	0.464	0.64	0.59	0.34	0.33	16.56	17.74	0.343	0.304
	2			-0.64	-0.59	0.66	0.66	17.65	18.54	0.313	0.323
400	1	0.871	0.811	0.51	0.47	0.34	0.33	27.00	30.94	0.370	0.319
	2			-0.51	-0.47	0.66	0.66	26.41	30.14	0.370	0.330
1000	1	1	0.996	0.43	0.41	0.33	0.34	39.22	48.99	0.433	0.312
	2			-0.43	-0.42	0.67	0.67	38.29	45.63	0.418	0.320
$\kappa = 0.10$	1	0.533	0.343	0.64	0.59	0.35	0.35	21.94	24.55	0.359	0.265
	2			-0.64	-0.60	0.66	0.66	22.14	24.73	0.318	0.318
0.15	1	0.557	0.464	0.64	0.59	0.34	0.33	16.56	17.74	0.343	0.304
	2			-0.64	-0.59	0.66	0.66	17.65	18.54	0.313	0.323
0.20	1	0.616	0.545	0.63	0.60	0.34	0.34	13.23	13.81	0.360	0.250
	2			-0.62	-0.60	0.66	0.67	13.70	13.42	0.283	0.321
$n = 1$	1	0.557	0.464	0.64	0.59	0.34	0.33	16.56	17.74	0.343	0.304
	2			-0.64	-0.59	0.66	0.66	17.65	18.54	0.313	0.323
2	1	0.784	0.667	0.50	0.46	0.34	0.34	14.20	17.40	0.385	0.268
	2			-0.51	-0.46	0.66	0.66	14.55	17.03	0.363	0.307
3	1	0.905	0.809	0.46	0.43	0.34	0.34	12.50	14.41	0.383	0.274
	2			-0.46	-0.43	0.67	0.66	12.66	14.98	0.389	0.305
4	1	0.956	0.882	0.44	0.41	0.34	0.34	10.61	14.04	0.409	0.287
	2			-0.44	-0.41	0.67	0.67	10.02	12.99	0.405	0.256
$M = 1$	1	0.578	0.486	0.58	0.57	0.50	0.51	22.85	25.36	0.335	0.309
	2			-0.58	-0.57	0.66	0.66	22.85	25.36	0.335	0.309
2	1	0.557	0.464	0.64	0.59	0.34	0.33	16.56	17.74	0.343	0.304
	2			-0.64	-0.59	0.66	0.66	17.65	18.54	0.313	0.323
3	1	0.649	0.559	0.62	0.59	0.25	0.25	11.93	11.70	0.308	0.293
	2			-0.65	-0.61	0.50	0.50	12.83	12.98	0.284	0.279
	3			0.61	0.58	0.74	0.74	12.84	12.55	0.275	0.306
4	1	0.677	0.549	0.61	0.57	0.20	0.21	7.69	7.31	0.290	0.244
	2			-0.67	-0.62	0.40	0.40	8.21	8.14	0.290	0.248
	3			0.68	0.63	0.60	0.60	8.30	7.99	0.264	0.313
	4			-0.63	-0.60	0.80	0.80	7.13	7.36	0.297	0.260
$\Delta_m = +0.2$	1	0.205	0.146	0.51	0.47	0.38	0.37	24.93	24.42	0.240	0.308
	2			-0.52	-0.50	0.66	0.66	22.03	22.75	0.247	0.240
$+0.4$	1	0.557	0.464	0.64	0.59	0.34	0.33	16.56	17.74	0.343	0.304
	2			-0.64	-0.59	0.66	0.66	17.65	18.54	0.313	0.323
$+0.8$	1	0.997	0.983	0.87	0.83	0.33	0.33	9.84	11.09	0.318	0.283
	2			-0.87	-0.84	0.67	0.66	9.06	11.05	0.357	0.265
$+0.4$	1	0.967	0.949	0.43	0.42	0.33	0.33	23.59	22.98	0.336	0.340
	2			0.45	0.44	0.67	0.66	23.34	23.57	0.345	0.316
$\Delta_m = +0.4$	1	0.999	0.999	0.45	0.42	0.25	0.24	15.45	12.78	0.305	0.362
	2			0.35	0.39	0.50	0.50	18.14	14.71	0.298	0.407
	3			0.47	0.45	0.75	0.75	16.27	13.18	0.274	0.398

Note: A comparison of DP and RW for several DGP specifications. Individual adjustments to the reference specification are made for $T = 200$, $\kappa = 0.15$, $n = 1$, $M = 2$, and $\Delta_m = 0.4$ for odd m and $\Delta_m = -0.4$ for even m , given that $k_m = \lceil mT/(M+1) \rceil$. For the calculation of all measures to the right of the dashed line, only data from simulations where both methods have identified the correct number of breaks at confidence level $1 - \alpha = 0.95$ has been used. I_m^* gives the relative frequency that the absolute difference between \hat{k}_m and k_m is smaller for the given method than for the other method.

The middle section of Table 4.3 reports the impact of trimming parameter κ , number of co-breaking series n , and number of breaks M on testing power and estimation. Setting $\kappa = 0.20$ instead of $\kappa = 0.15$ makes both tests more powerful, but in an empirical application this comes at the cost of dismissing more break date candidates *ex ante*. When $\kappa = 0.10$, the break-detection frequency of RW decreases drastically to $I_{0.95}^A = 34.3\%$. Given that $T = 200$, the implied reduction of the rolling window size to $2\omega = 0.2T = 40$ adversely affects the performance of RW. An increase in n benefits the testing power and break estimation efficiency of both DP and RW. RW is positively affected by the additional information that is contained in the additional series, so that $\text{sd}(\hat{k}_m)$ shrinks. However, the impact appears to be stronger for DP. Table 4.3 further indicates that larger M tend to increase the testing power of either method. Note that in the simulation, every even-numbered structural break involves a shift reversal. This explains why the testing power can be smaller for an additional structural break, if $M = 2$ instead of $M = 1$, or if $M = 4$ instead of $M = 3$.

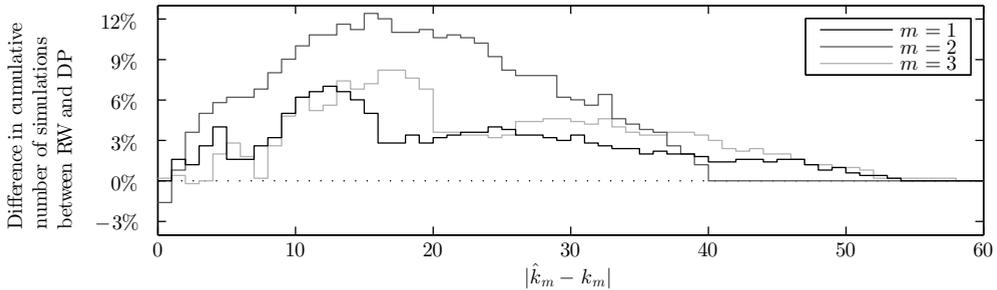
To conclude the middle section of Table 4.3, no clear preference for DP or RW can be derived from variations in the number of breaks M . Further, the use of RW instead of DP is less easily justified for larger n . Finally, if RW is applied in moderate sample sizes as discussed here, very small κ lead to a large drop in power. For either method, setting $\kappa = 0.15$ instead of $\kappa = 0.20$ prevents an unnecessary limitation in the set of permitted break dates.

When studying different values for Δ_m as shown in the lower section of Table 4.3, the first observation is that an increase in the absolute size of parameter shifts increases power and efficiency of either break detection method. For example, for RW, the break-detection frequency rises from $I_{0.95}^A = 46.4\%$ to $I_{0.95}^A = 98.3\%$, when $\Delta_1 = 0.8$ and $\Delta_2 = -0.8$ instead of $\Delta_1 = 0.4$ and $\Delta_2 = -0.4$. Notably, when the absolute size of the parameter shift is still 0.4, but both shifts are positive, RW becomes almost as powerful as DP. In this case, however, $\text{sd}(\hat{k}_m)$ is almost as large in both methods as for the shift reversal case with 50% smaller shifts, such that $|\Delta_m = 0.2|$. Since $\beta_3 \neq \beta_1$, both methods infer the existence of two breaks more easily than before, but with less certainty about the true break location. When comparing DP and RW, detection power and break estimation efficiency are very similar.

The claim that RW is a valid alternative to DP, not only to save time, is supported by the results reported for the DGP specification at the bottom of Table 4.3. Here, $M = 3$ breaks are simulated, and at each break date the parameter value increases by $\Delta_m = 0.4$. Again, all $\text{mean}(\hat{k}_m)$ lie reasonably close to the respective true break date locations $k_1 = 0.25T$, $k_2 = 0.5T$ and $k_3 = 0.75T$. But this time, all three RW break date estimates are much less dispersed than DP estimates, and all RW values of I_m^* are larger than the corresponding DP values. As an illustration of the break estimation results, Figure 4.4 shows the difference between RW and DP in the cumulative number of simulations that a break date has been detected with at most the given absolute deviation to k_m , for all three $m = 1, 2, 3$. With few exceptions the values are positive, and positive values indicate that RW has more often than DP estimated break dates to lie within the given range of k_m . For all breaks $m = 1, 2, 3$, break date estimates within 10, 20, 30, 40 dates of the true break occur more often in RW than in DP.

The overall results for a scenario of multiple parameter shifts in the same direction are favorable for RW. Break dates estimation appears to benefit from local optimization,

Figure 4.4: Differences in break date estimation for three upward shifts in one series



Note: Difference in break date estimation between DP and RW, when data is simulated with a DGP specification of $T = 200$, $n = 1$, $M = 3$, $k_1 = 50$, $k_2 = 100$, $k_3 = 150$, $\beta_1 = 1$, $\beta_2 = 1.4$, $\beta_3 = 1.8$ and $\beta_4 = 2.2$. The trimming parameter is $\kappa = 0.15$. For each $m = 1, 2, 3$, a plot shows the difference between RW and DP in the number of simulations that k_m has been estimated to lie as close to the true value as specified on the x-axis, or closer. If a y-value is above 0%, then RW has more often than DP produced break date estimates that lie at most as many dates away from k_m as indicated on the x-axis.

when parameters shift several times in the same direction. Yet, to conclude, no method can be recommended over the other method based on the reported break estimation properties of this section, since the direction of a potential parameter shift is not known prior to testing. The essential finding remains that both methods produce similar break estimates, both methods can be a valid choice in practice, the testing power of DP is larger, but it approaches 100% for both methods with larger data sets or large parameter shifts, and RW is significantly faster than DP.

4.4 Empirical Application

The empirical study will illustrate how the RW algorithm can be used in practice to detect multiple changes in the volatility of European government bonds. The RW method can uncover statistical evidence for episodes of different volatility specific to the daily price changes of a given series. Before looking at the structural break results, the section introduces the data set. Regression models are estimated to control for variation that can be explained by the development in other financial series. Reasons are given for the selection of financial series that are available for the linear regression models of the European bond series. As the models will not control for idiosyncratic, exogenous factors that measure credit or liquidity risk, developments in these risk types will influence the patterns of the error term variation of the models. The section will focus on the different aspects of using the RW algorithm and on the analysis of volatility changes in the residual observations of the models.

The daily redemption yield series of bonds of 18 European central governments with a maturity of 10 years is studied in this section of the chapter. The countries are Germany *BD*, Belgium *BG*, Czech Republic *CZ*, Denmark *DK*, Spain *ES*, Finland *FN*, France *FR*, Greece *GR*, Ireland *IR*, Italy *IT*, the Netherlands *NL*, Norway *NW*, Austria *OE*, Poland *PO*, Portugal *PT*, Sweden *SD*, Switzerland *SW*, United Kingdom *UK*. The data set starts on 2003-1-9, includes 3640 observations until 2016-12-21 and has been retrieved

from Thomson Reuters datastream. Beside the redemption yields of 18 European central governments' bonds r_i with $i = BD, BG, \dots, UK$, the set contains benchmark 10-year U.S. government redemption yields r_{US} , the difference between Euribor and EONIA r_{EX} , three stock exchange indices, S&P 500 s_{US} , MSCI World s_{WR} , and MSCI Emerging Markets Index s_{EM} , two commodity prices, LBMA Gold Price p_{gld} , and Henry Hub Natural Gas Price p_{gas} , and two economic indicators, the CBOE Volatility Index i_{vix} , and the Citigroup Economic Surprise Index for the Eurozone i_{ez} . All prices are denoted in Euros, all yields are denoted in percentage points. Each series has been tested by the augmented Dickey-Fuller test for a unit root against covariance-stationarity, with or without drift, with or without trend, with up to 4 lags. This has motivated using first differences in natural logs for stock exchange indexes and gold and gas prices, and first differences for the yields and economic indicators. Using lagged observations of each series ensures that all covariates are predetermined during model estimation. From the full set of financial series, a statistical selection process is used to determine significant covariates for each European bond yield series separately.

The initial data set is as large and varied, because various types of risks may contribute to the observed variation, either associated with the country, or the type of asset or global financial markets generally. Eventually, the models of the different countries may contain different sets of variables. For example, the impact and importance of developments in the equity market on government bond yields can differ between countries. Ericsson et al. (2009) point out that government bonds can either be seen as a relatively safe, alternative investment to corporate papers. Alternatively, the values of government bonds and corporate papers within the same country can also be seen as positively co-dependent. Manganelli and Wolswijk (2009) contains an overview of various variables included in the models of several related studies. Part of the variation in bond prices can usually be explained by international risk factors, see previous studies on sovereign bond yield spreads such as Attinasi et al. (2009) and Barrios et al. (2009). International risk is expected to be a common component in the regressions and may be proxied by U.S. bond yields, such as in Codogno et al. (2003) and Maltritz (2012). High global market volatility can be captured by the CBOE Volatility Index, see for example Beber et al. (2008).

The data set includes several global variables, but lacks variables that capture idiosyncratic effects, apart from lagged observations of the dependent series that may capture intra-market autocorrelation. Macroeconomic data is available in lower-frequency models but not for daily models, see for example Codogno et al. (2003), or Bernoth and Erdogan (2012) who measure the time-varying importance of debt to GDP ratios and projected deficit on interest differentials. Conclusions about the impact of idiosyncratic credit risk measures in Europe have changed since the sovereign debt crisis. On the one hand, Geyer et al. (2004) find that idiosyncratic country factors have almost no explanatory power in their model of weekly data for euro area countries, whereas global factors relating to differences between private and public sector are more important to include. On the other hand, Favero et al. (2010) note that the provisions of the Stability Pact do not fully eliminate differences in solvency risk and Sibbertsen et al. (2014) confirm the existence of sovereign credit risk that appears to have increased between 2006 and 2008 according to their structural break analysis on four euro area countries. In a panel discussion in Manganelli and Wolswijk (2009), Hans-Werner Sinn refers to Iceland as an example for the importance of default risk.

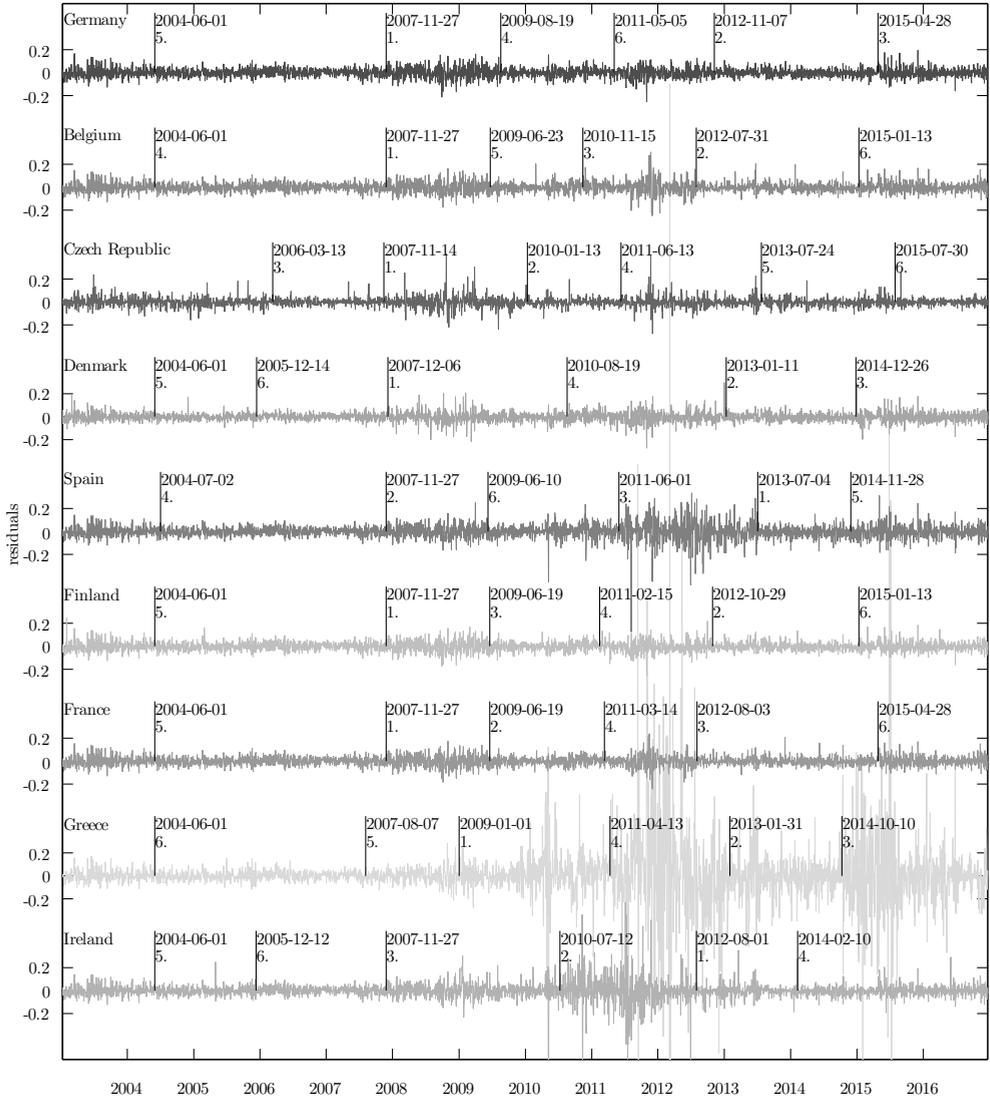
Each dependent series $\Delta(r_i)_t$ is initially regressed on an intercept term, on its own first five lags, on first lags of the other return yield differences, and on first lags of the other series in the data set. The regression models are linear and estimated separately using weighted least squares. For this, Cook's Distance is calculated to account for high-leverage points and outliers. Relevant predictor variables of the respective regressions are chosen statistically. For this, the variables are eliminated in a simple model selection approach; Those coefficients of variables that do not add significant explanation are set to zero one-by-one, see for example Hocking (1976). The remaining covariates are jointly significant in each regression model at the usual confidence levels. Full regression results can be seen in the Appendix of this chapter in Table 4.7. As a short summary, part of the variation of each series can be explained by global developments. A common observation is the positive coefficient of lagged return-differences of U.S. government bond yields, except for Greece. Further, $\Delta(\ln s_{WR})_{t-1}$ has a negative correlation with many European bond series, while $\Delta(\ln s_{US})_{t-1}$ typically has a positive correlation. In general, European return-differences are self-correcting between trading days, with either negative coefficients for the first lag of the dependent variable or negative coefficients of the lagged German series.

The residuals of each series are plotted in Figures 4.5 and 4.6. The plots all use the same scale, which highlights several episodes of high volatility in Greece, Portugal and other euro area periphery countries in particular. The figures already incorporate the results from testing the standard deviation of the error term of weighted observations of each model for multiple breaks. To evaluate each structural break hypothesis, the proposed RW test is applied to the data samples of weighted observations individually. The Gaussian distribution function is used for the calculation of likelihood function values. Trimming is set to $\kappa = 0.10$, which implies minimum segment length of $\omega = \lceil 0.1 \cdot 3640 \rceil = 364$, which is sufficiently large for well-behaved test properties according to the Monte Carlo Study. The Kolmogorov-Smirnov test generally rejects the normality hypothesis of the residuals of the full sample model of the series at 90% confidence. Therefore, critical values of the multiple break hypothesis are derived according to Qu and Perron (2007), using a Newey-West estimator to account for conditional heteroscedasticity.

The algorithm detects six breaks in all series, in each case with $p < 0.001$, as can also be seen in Tables 4.4, 4.5 and 4.6.³ The full sample maximum LR values can be found in the first columns of the tables. The tables further report each estimated structural break date along with the corresponding $LR(k)$ value, estimates of σ_m as well as the estimated absolute and relative parameter changes Δ_m .⁴ Further, the positions of all structural breaks are visualized in the plots of the residuals in Figures 4.5 and 4.6.

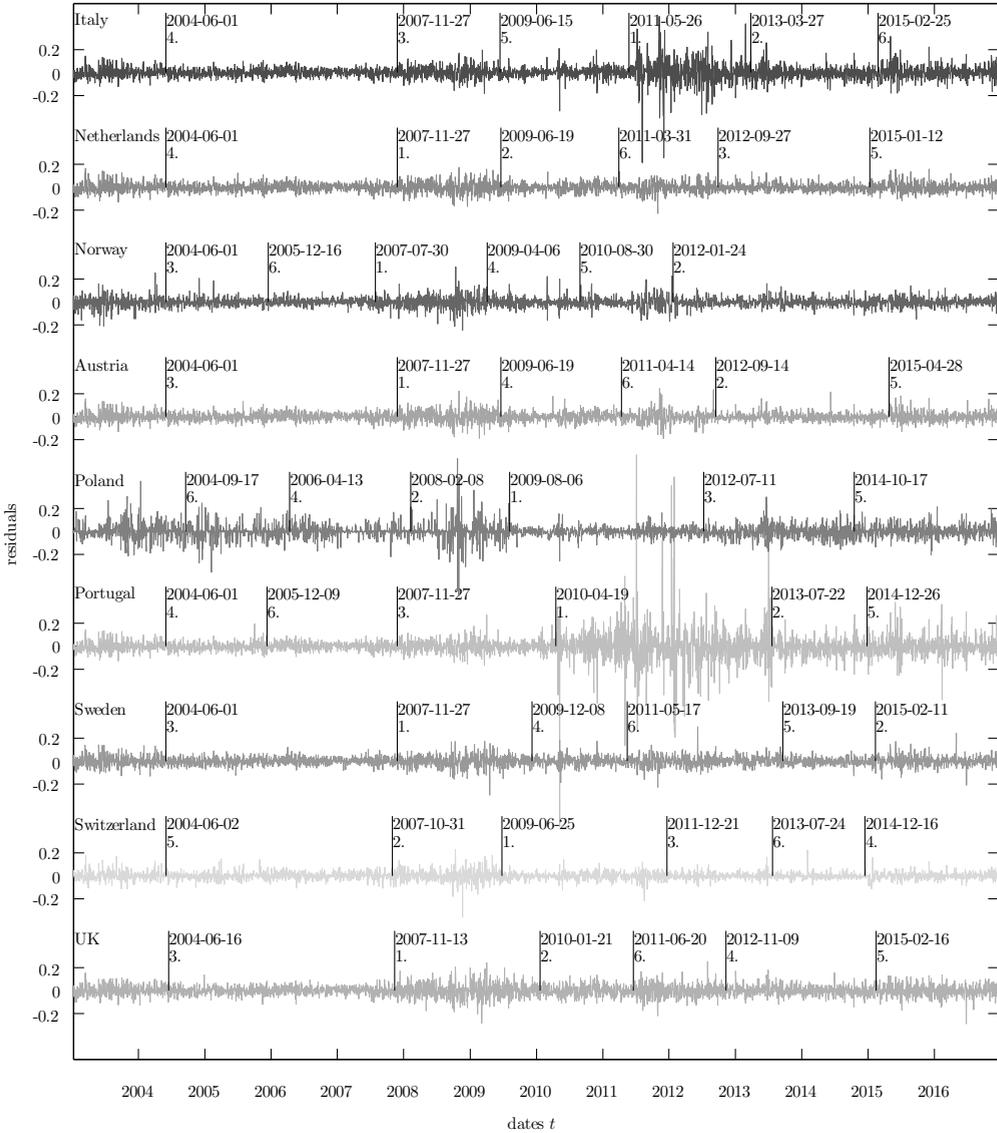
³In some series, a test allowing more breaks has lead to seven or eight structural breaks being found. For ease of comparisons, only the results with six breaks were included here. Since the RW method orders the breaks by their respective LR value, the missing seventh or eighth breaks are associated with the smallest LR values and have relatively minor impact on the overall model fit. For additional testing of the significance of additional structural breaks in multiple structural break testing, refer to Bai and Perron (2003).

⁴Estimates of σ_m are calculated using the original series, while the structural breaks have been detected using the weighted data set. The unreported values of σ_m for the weighted observations are all slightly smaller, because the weight of outliers and high-leverage points is lower than 1.

Figure 4.5: Residual plots and significant changes in volatility 1/2

Note: Residuals of the first set of series over time, $i = BD, BE, \dots, IR$, all using the same scaling. The residuals $\hat{\varepsilon}_{i,t}$ have been obtained from the original daily differences and the fitted model using only significant covariates and WLS, $\hat{\varepsilon}_{i,t} = \Delta(r_i)_t - x_{i,t}\hat{\beta}_i$. In each plot, the estimated break dates \hat{k} from the application of RW to find structural breaks in the error standard deviation are marked by vertical lines. The marks show the estimated break date as well as the order, by which the RW algorithm has detected the breaks.

Figure 4.6: Residual plots and significant changes in volatility 2/2



Note: Residuals of the second set of series over time, $i = IT, NL, \dots, UK$, all using the same scaling. The residuals $\hat{\varepsilon}_{i,t}$ have been obtained from the original daily differences and the fitted model using only significant covariates and WLS, $\hat{\varepsilon}_{i,t} = \Delta(r_i)_t - x_{i,t}\hat{\beta}_i$. In each plot, the estimated break dates \hat{k} from the application of RW to find structural breaks in the error standard deviation are marked by vertical lines. The marks show the estimated break date as well as the order, by with the RW algorithm has detected the breaks.

Table 4.4: RW test results for volatility changes 1/3

i	$LR(\hat{\mathcal{T}}_M)$	m	1	2	3	4	5	6
BD	215.1046	\hat{k}_m	2004-06-01	2007-11-27	2009-08-19	2011-05-05	2012-11-07	2015-04-28
		$LR(k)$	29.680	82.346	36.763	11.660	47.473	40.618
		$\hat{\sigma}_m$	0.043	0.031	0.056	0.037	0.048	0.031
		$\hat{\Delta}_m$	-0.011	0.024	-0.019	0.011	-0.018	0.013
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-26.39	77.38	-33.34	30.52	-37.09	42.01
		$\hat{\sigma}_{m+1}$	0.031	0.056	0.037	0.048	0.031	0.043
BG	290.5139	\hat{k}_m	2004-06-01	2007-11-27	2009-06-23	2010-11-15	2012-07-31	2015-01-13
		$LR(k)$	38.742	101.130	35.065	40.405	93.849	22.362
		$\hat{\sigma}_m$	0.043	0.032	0.054	0.039	0.066	0.034
		$\hat{\Delta}_m$	-0.012	0.023	-0.016	0.028	-0.032	0.009
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-27.17	71.52	-28.97	71.35	-48.97	26.34
		$\hat{\sigma}_{m+1}$	0.032	0.054	0.039	0.066	0.034	0.043
CZ	290.5403	\hat{k}_m	2006-03-13	2007-11-14	2010-01-13	2011-06-13	2013-07-24	2015-07-30
		$LR(k)$	57.077	144.890	79.597	44.816	29.544	19.328
		$\hat{\sigma}_m$	0.039	0.026	0.056	0.030	0.049	0.033
		$\hat{\Delta}_m$	-0.013	0.029	-0.026	0.018	-0.015	-0.007
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-32.26	111.35	-45.75	61.07	-31.34	-19.84
		$\hat{\sigma}_{m+1}$	0.026	0.056	0.030	0.049	0.033	0.027
DK	340.4974	\hat{k}_m	2004-06-01	2005-12-14	2007-12-06	2010-08-19	2013-01-11	2014-12-26
		$LR(k)$	30.761	11.065	100.570	31.891	67.327	63.165
		$\hat{\sigma}_m$	0.037	0.027	0.025	0.046	0.052	0.031
		$\hat{\Delta}_m$	-0.010	-0.001	0.020	0.006	-0.021	0.016
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-27.15	-5.34	80.57	13.24	-40.91	53.28
		$\hat{\sigma}_{m+1}$	0.027	0.025	0.046	0.052	0.031	0.047
ES	808.6617	\hat{k}_m	2004-07-02	2007-11-27	2009-06-10	2011-06-01	2013-07-04	2014-11-28
		$LR(k)$	39.463	119.020	15.135	92.605	158.600	27.234
		$\hat{\sigma}_m$	0.043	0.031	0.055	0.054	0.110	0.044
		$\hat{\Delta}_m$	-0.012	0.024	-0.001	0.056	-0.066	0.012
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-27.86	77.88	-1.55	102.69	-59.68	27.71
		$\hat{\sigma}_{m+1}$	0.031	0.055	0.054	0.110	0.044	0.057
FN	208.9358	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-02-15	2012-10-29	2015-01-13
		$LR(k)$	32.739	74.731	44.049	32.980	51.596	28.739
		$\hat{\sigma}_m$	0.045	0.032	0.055	0.036	0.052	0.032
		$\hat{\Delta}_m$	-0.012	0.023	-0.020	0.017	-0.020	0.009
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-27.60	71.52	-35.70	46.99	-38.93	29.09
		$\hat{\sigma}_{m+1}$	0.032	0.055	0.036	0.052	0.032	0.041

Note: For a first set of series i , estimates resulting from an application of RW to detect multiple structural breaks in the error term standard deviation σ_ε , given that $\kappa = 0.10$. The second column shows the maximized value of $LR(\hat{\mathcal{T}}_M)$, which is associated with $p < 0.001$ for all series i . For each series i and each estimated break m , the table reports break date estimate \hat{k}_m , single-break LR values of the date $LR(k)$, the estimate before the break $\hat{\sigma}_m$, absolute and relative parameter shift estimates, $\hat{\Delta}_m$ and $\%(\hat{\Delta}_m/\hat{\sigma}_m)$, and the estimate after the break $\hat{\sigma}_{m+1}$. For the detection, weighted data sets are used, while each $\hat{\sigma}_m$ is calculated using the original data.

CHAPTER 4. A ROLLING WINDOW ALGORITHM FOR DETECTING MULTIPLE STRUCTURAL BREAKS

Table 4.5: RW test results for volatility changes 2/3

i	$LR(\hat{\mathcal{J}}_M)$	m	1	2	3	4	5	6
<i>FR</i>	266.797	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-14	2012-08-03	2015-04-28
		$LR(k)$	38.104	106.460	63.510	47.775	57.155	22.966
		$\hat{\sigma}_m$	0.043	0.031	0.055	0.035	0.056	0.033
		$\hat{\Delta}_m$	-0.012	0.023	-0.020	0.021	-0.023	0.011
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-26.89	73.06	-36.12	60.84	-41.14	32.93
		$\hat{\sigma}_{m+1}$	0.031	0.055	0.035	0.056	0.033	0.044
<i>GR</i>	5383.176	\hat{k}_m	2004-06-01	2007-08-07	2009-01-01	2011-04-13	2013-01-31	2014-10-10
		$LR(k)$	27.219	110.100	631.370	170.420	474.900	412.210
		$\hat{\sigma}_m$	0.047	0.036	0.059	0.238	1.413	0.156
		$\hat{\Delta}_m$	-0.011	0.022	0.179	1.175	-1.257	0.269
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-23.35	61.79	306.37	493.97	-88.94	172.12
		$\hat{\sigma}_{m+1}$	0.036	0.059	0.238	1.413	0.156	0.425
<i>IR</i>	1577.825	\hat{k}_m	2004-06-01	2005-12-12	2007-11-27	2010-07-12	2012-08-01	2014-02-10
		$LR(k)$	19.330	11.228	160.810	229.960	244.730	20.062
		$\hat{\sigma}_m$	0.042	0.035	0.031	0.071	0.146	0.052
		$\hat{\Delta}_m$	-0.007	-0.004	0.040	0.075	-0.094	-0.009
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-17.34	-11.35	130.96	105.06	-64.59	-17.54
		$\hat{\sigma}_{m+1}$	0.035	0.031	0.071	0.146	0.052	0.043
<i>IT</i>	995.932	\hat{k}_m	2004-06-01	2007-11-27	2009-06-15	2011-05-26	2013-03-27	2015-02-25
		$LR(k)$	37.446	109.000	33.345	231.950	145.960	16.593
		$\hat{\sigma}_m$	0.043	0.032	0.052	0.043	0.120	0.052
		$\hat{\Delta}_m$	-0.011	0.021	-0.009	0.077	-0.068	0.004
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-26.55	66.25	-17.63	178.71	-56.77	7.95
		$\hat{\sigma}_{m+1}$	0.032	0.052	0.043	0.120	0.052	0.056
<i>NL</i>	164.874	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-31	2012-09-27	2015-01-12
		$LR(k)$	30.678	69.710	36.104	10.540	35.799	28.423
		$\hat{\sigma}_m$	0.044	0.032	0.055	0.036	0.048	0.031
		$\hat{\Delta}_m$	-0.012	0.023	-0.019	0.011	-0.016	0.009
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-27.45	72.96	-33.95	30.84	-34.56	29.60
		$\hat{\sigma}_{m+1}$	0.032	0.055	0.036	0.048	0.031	0.040
<i>NW</i>	346.920	\hat{k}_m	2004-06-01	2005-12-16	2007-07-30	2009-04-06	2010-08-30	2012-01-24
		$LR(k)$	63.923	22.287	138.890	43.510	23.670	64.283
		$\hat{\sigma}_m$	0.052	0.034	0.027	0.057	0.040	0.052
		$\hat{\Delta}_m$	-0.019	-0.007	0.030	-0.017	0.012	-0.017
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-35.77	-19.31	108.80	-29.97	29.87	-32.07
		$\hat{\sigma}_{m+1}$	0.034	0.027	0.057	0.040	0.052	0.035

Note: For a second set of series i , estimates resulting from an application of RW to detect multiple structural breaks in the error term standard deviation σ_ε , given that $\kappa = 0.10$. The second column shows the maximized value of $LR(\hat{\mathcal{J}}_M)$, which is associated with $p < 0.001$ for all series i . For each series i and each estimated break m , the table reports break date estimate \hat{k}_m , single-break LR values of the date $LR(k)$, the estimate before the break $\hat{\sigma}_m$, absolute and relative parameter shift estimates, $\hat{\Delta}_m$ and $\%(\hat{\Delta}_m/\hat{\sigma}_m)$, and the estimate after the break $\hat{\sigma}_{m+1}$. For the detection, weighted data sets are used, while each $\hat{\sigma}_m$ is calculated using the original data.

Table 4.6: RW test results for volatility changes 3/3

i	$LR(\hat{\mathcal{T}}_M)$	m	1	2	3	4	5	6
<i>OE</i>	173.272	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-04-14	2012-09-14	2015-04-28
		$LR(k)$	29.825	69.323	27.436	11.493	40.979	25.466
		$\hat{\sigma}_m$	0.043	0.032	0.056	0.038	0.052	0.032
		$\hat{\Delta}_m$	-0.011	0.024	-0.019	0.014	-0.020	0.011
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-25.69	75.65	-33.03	38.27	-38.85	34.43
		$\hat{\sigma}_{m+1}$	0.032	0.056	0.038	0.052	0.032	0.043
<i>PO</i>	546.392	\hat{k}_m	2004-09-17	2006-04-13	2008-02-08	2009-08-06	2012-07-11	2014-10-17
		$LR(k)$	4.844	63.866	212.880	274.800	74.620	14.188
		$\hat{\sigma}_m$	0.070	0.066	0.044	0.101	0.037	0.059
		$\hat{\Delta}_m$	-0.004	-0.022	0.056	-0.064	0.023	-0.006
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-5.71	-33.10	127.48	-63.65	61.66	-9.97
		$\hat{\sigma}_{m+1}$	0.066	0.044	0.101	0.037	0.059	0.053
<i>PT</i>	2099.002	\hat{k}_m	2004-06-01	2005-12-09	2007-11-27	2010-04-19	2013-07-22	2014-12-26
		$LR(k)$	32.187	7.147	114.840	313.290	162.170	17.214
		$\hat{\sigma}_m$	0.044	0.034	0.031	0.051	0.194	0.069
		$\hat{\Delta}_m$	-0.010	-0.003	0.020	0.143	-0.124	0.014
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-22.08	-8.65	62.62	279.74	-64.21	20.51
		$\hat{\sigma}_{m+1}$	0.034	0.031	0.051	0.194	0.069	0.084
<i>SD</i>	182.631	\hat{k}_m	2004-06-01	2007-11-27	2009-12-08	2011-05-17	2013-09-19	2015-02-11
		$LR(k)$	37.588	92.404	36.025	19.231	28.748	39.545
		$\hat{\sigma}_m$	0.046	0.031	0.053	0.039	0.045	0.030
		$\hat{\Delta}_m$	-0.015	0.022	-0.014	0.006	-0.015	0.013
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-33.17	72.38	-26.87	15.39	-33.42	43.84
		$\hat{\sigma}_{m+1}$	0.031	0.053	0.039	0.045	0.030	0.043
<i>SW</i>	214.119	\hat{k}_m	2004-06-02	2007-10-31	2009-06-25	2011-12-21	2013-07-24	2014-12-16
		$LR(k)$	15.500	52.035	67.259	36.509	12.508	27.918
		$\hat{\sigma}_m$	0.040	0.031	0.053	0.034	0.028	0.024
		$\hat{\Delta}_m$	-0.009	0.023	-0.020	-0.006	-0.004	0.006
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-22.04	73.07	-37.01	-16.79	-13.41	26.01
		$\hat{\sigma}_{m+1}$	0.031	0.053	0.034	0.028	0.024	0.031
<i>UK</i>	304.508	\hat{k}_m	2004-06-16	2007-11-13	2010-01-21	2011-06-20	2012-11-09	2015-02-16
		$LR(k)$	33.502	157.540	45.513	18.768	28.952	23.689
		$\hat{\sigma}_m$	0.045	0.033	0.065	0.044	0.058	0.042
		$\hat{\Delta}_m$	-0.011	0.032	-0.021	0.014	-0.016	0.010
		$\%(\hat{\Delta}_m/\hat{\sigma}_m)$	-25.39	96.48	-32.01	31.18	-27.57	22.96
		$\hat{\sigma}_{m+1}$	0.033	0.065	0.044	0.058	0.042	0.052

Note: For a third set of series i , estimates resulting from an application of RW to detect multiple structural breaks in the error term standard deviation σ_ε , given that $\kappa = 0.10$. The second column shows the maximized value of $LR(\hat{\mathcal{T}}_M)$, which is associated with $p < 0.001$ for all series i . For each series i and each estimated break m , the table reports break date estimate \hat{k}_m , single-break LR values of the date $LR(k)$, the estimate before the break $\hat{\sigma}_m$, absolute and relative parameter shift estimates, $\hat{\Delta}_m$ and $\%(\hat{\Delta}_m/\hat{\sigma}_m)$, and the estimate after the break $\hat{\sigma}_{m+1}$. For the detection, weighted data sets are used, while each $\hat{\sigma}_m$ is calculated using the original data.

The uncovered changes in volatility can have a variety of interpretations in terms of risk. It seems unlikely that the changes are associated with changes in global risk or global investor behavior, as these factors have been controlled by allowing the inclusion of appropriate covariates. Liquidity risk may affect bond prices through trading costs and the possibility of changing market availability. Empirical evidence for this is mixed, see Beber et al. (2008) and Favero et al. (2010), who offer different results depending on the inclusion of other idiosyncratic factors. Arguably, the relative importance of liquidity for long-term European central government bonds is low due to the depth of the secondary market. As pointed out in Barrios et al. (2009), the remaining sources of risk are idiosyncratic macroeconomic factors, such as debt sustainability or socio-economic and political uncertainty, and affect credit risk. As these were not controlled by the regression model, changes in error term volatility may largely be explained by such unobserved developments. Conversely, if the European sovereign debt crisis has increased the market sensitivity to the fundamentals of euro area countries, as argued by Giordano et al. (2013), then their developments should be visible in the error term volatility.

Barrios et al. (2009) find a significant increase in the impact of domestic factors on bond yield spreads in Europe starting in August 2007, when their weekly data is tested for exogenously chosen structural breaks. In the present RW study, structural breaks in late 2007 are a common feature as well. The preceding segment is the least or second-to-least volatile segment for all countries. The following data segment covers the Global Financial Crisis and is associated with higher volatility. For all series, the residuals have at least 60% higher standard deviation in the Global Financial Crisis segment than before. For ten series, $i = BD, CZ, FN, NL, NW, OE, PO, SD, SW, UK$, none of the remaining sections will be as volatile as the segment of the Global Financial Crisis, including data segments that cover the European sovereign debt crisis. For these ten countries, this means that the debt crisis has not been associated with such large credit risk increase, that it would have elevated volatility above Global Financial Crisis levels for a prolonged time. The result supports the evidence found by Ehrmann and Fratzscher (2017) for flight to quality during the height of the sovereign debt crisis, which holds for countries outside the euro area but also a few central euro area countries.

The respective late-2007 break is the one that the RW method selects first for eleven countries, $i = BD, BG, CZ, DK, FN, FR, NL, NW, OE, SD, UK$. The selection order is based on the respective $LR(k)$ values, which are reported for all estimated break dates \hat{k}_m in Tables 4.4–4.6. The ordering is also noted under each structural break date in the plots of Figures 4.5 and 4.6. Out of the other seven countries, for Poland and Switzerland, only a break in summer 2009 has been selected quicker by RW. The Polish series experiences its most calm segment after the 2009 break for three years, whereas the 2009 break initiates a sequential reduction in volatility in the Swiss bond market. The remaining five countries, the euro area periphery countries $i = ES, GR, IR, IT, PT$, have been at the center of debt sustainability problems during the European sovereign debt crisis. Accordingly, the LR value ordering indicates that later parts of the sample are associated with more extreme changes in volatility than the 2007 initiation of the Global Financial Crisis.

The outbreak of the European sovereign debt crisis can be put in the second quarter of 2010, as the Greek government requested international financial aid on 2010-04-23

and the European Commission agreed on the European Financial Stability Facility on 2010-05-09 to provide financial assistance to euro area countries in distress. The period between 2010 and 2012 is regarded as a period of large uncertainty, and the volatility of sovereign bond yields is relatively high throughout the euro area, as studied in Chapter 3 as well. It is also a period of many debates about financial assistance, austerity measures and about the mandate of the European Central Bank (ECB). Its introduction of the Securities Markets Programme was subject to significant controversy, as described in Eser and Schwaab (2016). Key features of the program were not disclosed. Also, its goal differed from quantitative easing programs in other countries, in that the program did not aim to make monetary policy more accommodative, since the ECB followed the goal of price stability by targeting a stable money growth rate. Instead, the program addressed the malfunctioning of certain government bond markets for the transmission of policy. In the following months and years, other periphery countries were directly affected by the debt crisis, such as Ireland, whose government received financial support on 2010-11-29. According to the RW results, the highest volatility segments for $i = ES, GR, IR, IT, PT$ all cover the period between 2011-06-03 and 2012-08-01, with the earliest starting date for Portugal on 2010-04-20 and the last ending date for Spain on 2013-07-04.

For Belgium, Ireland and France, the significant reduction in volatility is estimated within a maximum of eight days of 2012-07-26. On that day, Ireland had its first market sale of government bonds since 2010 and also ECB president Mario Draghi gave a speech to do “whatever it takes” to preserve the euro within the mandate of the ECB. The following adjustments of the institutions of the European Monetary System are associated with volatility reductions in other euro area markets. Austria and the Netherlands are estimated to enter a lower volatility segment shortly after the ECB announced its Outright Monetary Transactions program on 2012-09-06. The reduction of volatility in Finland and Germany is estimated to occur after European leaders agreed on the commencement of a single banking supervisor for the euro area in early 2013. For all periphery euro area countries as well as many other countries, structural breaks associated with the calming of the European financial markets in 2012/2013 have been selected very quickly by the RW algorithm, even first for Ireland and Spain. This highlights how extreme the reduction of volatility has been between the height of the crisis and afterwards.

As previously announced in the winter of 2014/2015, on 2015-03-05 the governing council of the ECB declared that an outright purchase program that involves government bonds be initiated on 2015-03-09. In the RW application, the volatility of all euro area countries except Ireland is found to be significantly higher in the last part of the sample than in the segment before the winter of 2014/2015. In contrast with earlier programs, the new program’s purpose is to make monetary policy more expansionary, so that inflation rates increase. The program breaks with ECB tradition, and so the volatility changes may be associated with fears of facilitating bubbles and causing excessive inflation, or with enabling excessive government spending and increased credit risk. Yet, it meant that ECB policy finally aligned with the quantitative easing policies of American, British and Japanese central banks. It is worth investigating this period of the sample more, such as in the study on the predictive value of U.S. government bond yields in Chapter 6 of this book.

To conclude this empirical illustration of using the RW algorithm, six structural breaks have been detected that involve significant changes in the error term volatility

of 18 European government bond yield models between 2003 and 2016. For most euro area countries and the other European countries in the set, the European sovereign debt crisis is associated with higher volatility, but usually not with volatility as high as during the Global Financial crisis. In five periphery countries of the euro area, credit risk has risen during the debt crisis and appears to have contributed to prolonged periods of severely elevated volatility. In the study, between the two breaks that start and end the most volatile segments of each country, it is interesting that the RW algorithm has typically selected the end break first. Yet, the RW is designed as a time-saving heuristic for detecting a set of structural breaks and the order of selection is not formally tested for significance. Often, the estimated breaks occur close to specific events that affected the financial markets during the recent crises. Note, however, that break date estimates do not necessarily have to coincide with individually significant, one-time events. When a multiple structural break hypothesis is found to be significant, it means that several segments of the data are structurally different, not that individual events had a significant impact. Chapter 5 will investigate the abruptness of the estimated volatility transitions between the segments.

4.5 Conclusion

The chapter has presented an optimization heuristic for finding multiple structural breaks in time series models. The optimization algorithm scans through the data in rolling windows that are a fraction in length of the full sample. Within these windows, a single likelihood ratio is calculated that evaluates the data fit of a structural break in the middle of the window. The collection of likelihood ratios of all data windows enables the testing for a given number of structural breaks of the overall sample. For this, limit distributions and critical values described in previous literature on maximum likelihood ratio testing apply, see Andrews (1993) or Qu and Perron (2007).

With the rolling window algorithm, multiple structural break detection can be performed under linear instead of quadratic computation order. A Monte Carlo study confirms that the algorithm reduces the execution time substantially compared to the dynamic optimization program of Bai and Perron (2003). In the study, the rolling window algorithm is more precise in estimating multiple parameter shifts in the same direction. This can be explained by the local likelihood ratio evaluation of each break at a time, independent of the data outside the respective window. Since the direction is not known prior to testing, however, no preference of one method over the other can be expressed based on these performance results. The testing power of the rolling window algorithm is lower in samples with relatively few observations and/or relatively small parameter shifts. If data samples are large, low testing power is not a concern and the increasing reduction in execution time can be a great asset of the rolling window algorithm.

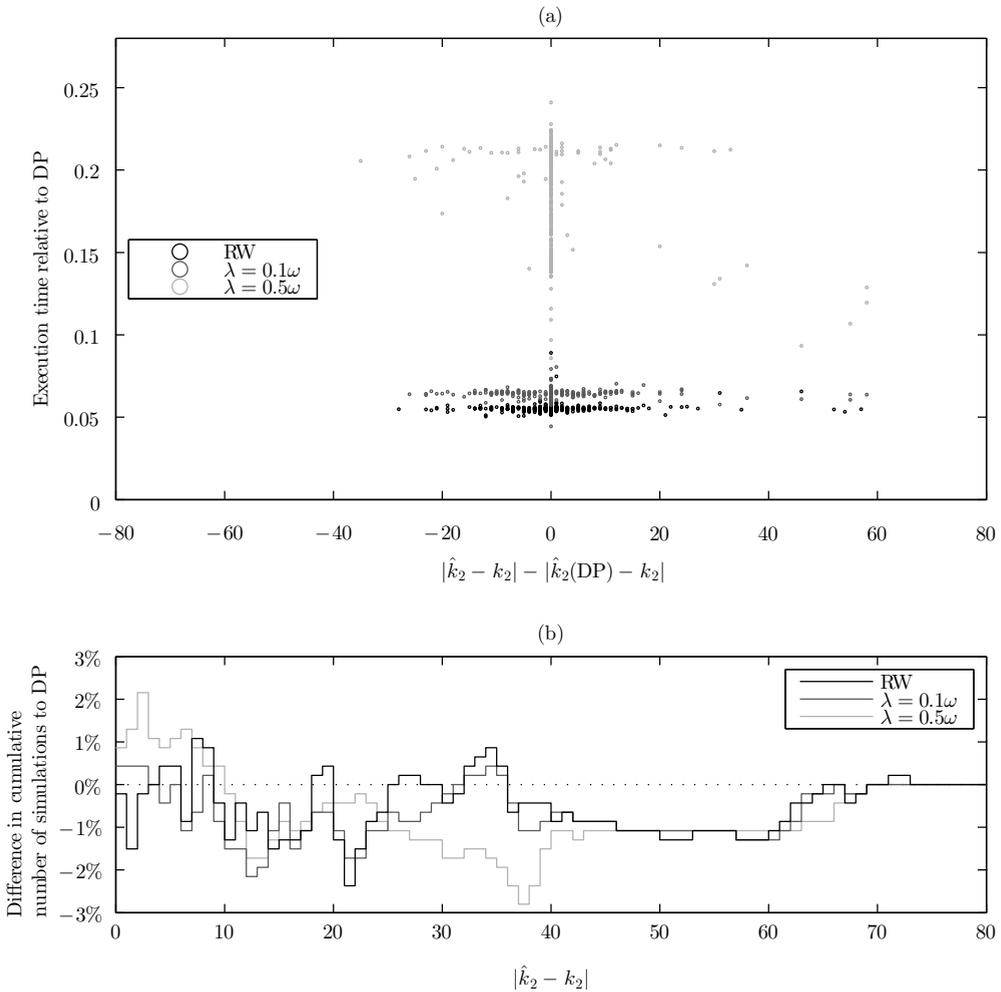
An empirical application shows that the rolling window algorithm can be used to detect structural breaks in the variation of differences in eighteen long-term European government redemption yields. Only volatility that remains unexplained by global risk measures is tested for structural breaks, so that country-specific patterns can be uncovered. The results closely match the results of earlier studies, in that the 2008 Global Financial Crisis and the subsequent European sovereign debt crisis are associated with

higher volatility regimes. The results differ by country, with euro area periphery countries experiencing particularly high volatility the stages of the European sovereign debt crisis. The study shows how likelihood ratio values from each data window augment the analysis by highlighting particularly important break dates. Future research could investigate in which way the likelihood ratio values from each data window can be used for a systematic ordering of structural breaks, or for estimating the number of structural breaks in a sample.

The chapter includes an extension suggestion, in which the potential break dates according to the rolling window algorithm are used as initial guesses for a subsequent application of the dynamic optimization program of Bai and Perron (2003). As the domain of potential break dates is restricted this way, the execution time of the dynamic optimization program is reduced. Another extension idea is to transfer the rolling window concept to other tests, such as the cross-validated error estimation test of Chapter 6.

4.A Appendix

Figure 4.7: Differences in break date estimation with shift reversal 2/2



Note: Differences in break date estimation between DP and RW-based methods, when data is simulated according to the reference DGP specification of $T = 200$, $n = 1$, $M = 2$, $k_1 = \lceil T/3 \rceil = 67$, $k_2 = \lfloor 2T/3 \rfloor = 133$, $\beta_1 = 1$, $\beta_2 = 1.4$, $\beta_3 = 1$. The trimming parameter is $\kappa = 0.15$. Upper panel (a) shows a scatter plot, with the difference in absolute deviations of \hat{k}_2 from k_2 between RW-based methods and DP on the x-axis, and with the execution time relative to DP on the y-axis. When a dot lies to the left of zero, the RW-based method has found a break date closer to k_2 than DP. Lower panel (b) plots the difference between RW-based methods and DP in the number of simulations that k_2 has been estimated to lie as close to the true value as specified on the x-axis, or closer. If a y-value is above 0%, then the RW-based method has more often than DP produced break date estimates that lie at most as many dates away from k_2 as indicated on the x-axis.

Table 4.7: Full sample regression models

covariates	dependent series $\Delta(r_t)_t$																		
	$i = BD$	BG	CZ	DK	ES	FN	FR	GR	IR	IT	NL	NW	OE	PO	PT	SD	SW	UK	
const.													-1.5e-3 (6.5e-4)						
$\Delta(r_t)_{t-1}$	-0.308 (0.053)	0.194 (0.048)	0.051 (0.023)	-0.192 (0.031)	0.229 (0.034)				0.206 (0.048)	0.136 (0.040)					0.281 (0.051)	-0.057 (0.026)	-0.137 (0.024)	-0.070 (0.035)	
$\Delta(r_t)_{t-3}$			0.056 (0.021)																
$\Delta(r_t)_{t-5}$																			
$\Delta(r_{BD})_{t-1}$		-0.259 (0.042)				-0.228 (0.039)	-0.201 (0.035)	-1.034 (0.192)	-0.432 (0.080)	-0.164 (0.040)	-0.290 (0.054)				-0.446 (0.072)			-0.115 (0.038)	
$\Delta(r_{BG})_{t-1}$	0.083 (0.036)		0.107 (0.031)	0.072 (0.034)		0.098 (0.029)	0.108 (0.034)		0.171 (0.080)		0.090 (0.034)					0.058 (0.022)			
$\Delta(r_{DK})_{t-1}$	0.076 (0.028)					0.058 (0.028)					0.063 (0.027)			0.080 (0.032)					
$\Delta(r_{ES})_{t-1}$								0.671 (0.218)							0.132 (0.060)				
$\Delta(r_{FN})_{t-1}$	0.124 (0.051)			0.167 (0.043)	-0.235 (0.039)						0.121 (0.053)	0.154 (0.023)		0.040 (0.014)					
$\Delta(r_{IR})_{t-1}$																			
$\Delta(r_{IT})_{t-1}$	-0.049 (0.016)																		
$\Delta(r_{NL})_{t-1}$																			0.078 (0.024)
$\Delta(r_{PT})_{t-1}$																			
$\Delta(r_{US})_{t-1}$	0.203 (0.018)	0.166 (0.020)	0.134 (0.018)	0.223 (0.017)	0.140 (0.024)	0.190 (0.018)	0.197 (0.018)		0.122 (0.032)	0.160 (0.025)	0.184 (0.018)	0.191 (0.017)		0.194 (0.018)	0.097 (0.024)	0.135 (0.041)	0.200 (0.017)	0.139 (0.014)	0.243 (0.020)
$\Delta(\ln swr)_{t-1}$	-0.721 (0.224)	-0.296 (0.104)	-0.811 (0.213)	-0.698 (0.209)		-0.687 (0.224)	-0.245 (0.097)			-0.271 (0.133)	-0.772 (0.218)			-0.732 (0.227)	-0.620 (0.233)				
$\Delta(\ln sus)_{t-1}$	0.470 (0.172)			0.493 (0.164)	0.423 (0.164)	0.389 (0.176)					0.402 (0.171)			0.362 (0.178)					
$\Delta(\ln sem)_{t-1}$																			
$\Delta(i_{vix})_{t-1}$																			
$\Delta(i_{ez})_{t-1}$			2.4e-4 (1.0e-4)																
$\Delta(\ln pga)_{t-1}$		0.194 (0.074)		0.157 (0.074)						0.207 (0.090)									
$\Delta(\ln pgas)_{t-1}$	0.031 (0.016)																		
F -stat	32.71	59.36	46.37	89.11	63.54	38.74	56.09	24.97	68.99	18.56	30.92	118.00	34.29	19.24	45.59	103.19	96.73	70.64	
adjusted R^2	0.073	0.074	0.081	0.145	0.065	0.059	0.057	0.013	0.070	0.028	0.062	0.114	0.068	0.025	0.069	0.078	0.073	0.055	

Note: Coefficient estimates and estimated standard errors of full sample regression models for the daily difference in redemption yields of 18 European government bonds. Coefficients are estimated in univariate regressions using weighted least squares. Significance is assessed with heteroscedasticity-consistent estimates of the standard errors. Backward elimination is performed to reduce the number of coefficients in each model.

5

Testing parameter change abruptness with a sum of likelihood ratios

5.1 Introduction

Structural break tests are designed to investigate the stability of estimated model parameters. Among the most cited publications in recent years is Andrews (1993), in which a maximum likelihood ratio test, a maximum Lagrange Multiplier test and a maximum Wald test for the endogenous detection of a single break in linear regression models are presented. The test statistics follow a simple basic concept: Should a set of parameters of the time series model be allowed to have different values before and after a break date? If the test result is significant, stability of the tested set of parameters is rejected in favor of a structural break with an abrupt parameter shift at the break date. A model with constant parameters is declared inadequate by the rejection of the stability hypothesis, but the adequacy of an abrupt parameter transition can be investigated further. Andrews (1993) advises to view his break tests as general tests of model adequacy. He argues that the tests exhibit power against more general forms of instability, not only abrupt parameter shifts.

The maximum likelihood ratio (MLR) break detection test is designed to evaluate the data fit of two options of modeling a data generating process. The two options are a model that contains a structural break in the form of an abrupt shift in a set of parameters and a model that does not. Consider a data generating process that contains a property that changes gradually over time. An appropriate model representation involves a set of parameters that transition gradually in value, but this is not an option in MLR testing. Between the two available options, the model with abruptly shifting parameters is favored over the model with constant parameters. Permitting a parameter shift as opposed to constant parameter values will improve the data fit, because different parameter values

can be assigned to the earlier part of the data and to the later part. If the overall magnitude of the parameter shift is sufficiently large, the stability hypothesis will be rejected, regardless of the speed of parameter transition.

This chapter proposes a follow-up analysis of the data collected during structural break testing. A requirement for the analysis is that a structural break has been detected by the MLR test according to Andrews (1993). The attention is then focused on data associated with the dates that surround the estimated break date. Specifically, a sum of likelihood ratios (SLR) statistic adds up the likelihood ratio (LR) values that surround the LR maximum to calculate an average. The central idea of the method is that a steeper peak in LR values at the maximum indicates a more abrupt transition of the changing parameters, all other things equal. A small LR average close to the break is associated with abrupt parameter change. Thus, the SLR statistic is used for a one-sided test, in which relatively large values indicate parameter transitions that are non-abrupt. Critical values can be derived from an empirical bootstrap in which data is simulated that satisfies the null hypothesis of abrupt change. A Monte Carlo study demonstrates the power of the proposed test in practice.

Gradual instead of abrupt change is considered in an empirical study by McConnell and Perez-Quiros (1998), and investigated in Jung and Maderitsch (2014). Stock and Watson (2003) and Paye and Timmermann (2006) recognize the power of structural break tests against general forms of instability. The analysis in Kleimeier and Sander (2000) detects structural breaks for several series during the early 1990's, and interprets a three-year period as a transitional period that divides two sample segments, but the possibility of gradual change in each series is not explicitly addressed. The extensive analysis of growth rate changes of Berg et al. (2012) includes the visualization of growth dynamics and studies the economic meaningfulness of statistical break test results, but does not challenge the concept of abrupt break points itself. Most recently, for the analysis of the European sovereign debt crisis, Gómez-Puig and Sosvilla-Rivero (2014) consider a wide range of break and causality measures, but do not mention the possibility that their endogenously detected structural breaks could indicate gradual change or another form of model instability. Similarly, Kunze and Gruppe (2014) and Gruppe and Lange (2014) each focus on the point estimate of a break date and do not explicitly consider non-abrupt change.

Other parts of the literature focus on the theory of gradual parameter transitions itself. Johansen et al. (2000) consider piecewise linear trends in a cointegration analysis, but assume prior knowledge of the break point. Smooth transition models have been established in Teräsvirta (1994) and Lin and Teräsvirta (1994). The models have been included in educational books on the topic of time series, such as Lütkepohl and Krätzig (2004). In these models parameters shift their values from one set of constants to another, following a specified functional form over an estimated amount of time. Abrupt change corresponds to one extreme case of the smooth transition models, in which the transition period is as short as an instant. In the Monte Carlo study of this chapter, the smooth transition model is used for the simulation of data. Both Lin and Teräsvirta (1994) and Chen and Hong (2012) contain tests within the smooth transition models framework, with which parameter stability can be rejected in favor of a smooth transition alternative hypothesis. By the more general hypothesis, the goal of these test is to decrease the probability of maintaining a false stability assumption.

In contrast, the goal of SLR testing is to decrease the probability of maintaining a false abruptness assumption. The test is not concerned with limited power of the MLR test and it does not assume a certain shape of transition. It is designed to provide a non-parametric and straightforward refinement tool to test the abruptness validity with the empirical data already collected for the MLR test. Then, a rejection of the abruptness hypothesis in practice can motivate the application of further methods found in the literature. A smooth transition model can be estimated to parameterize the gradual transition. Also, rolling window estimation of the model parameters may help to illustrate the change visually. Even if the form of the parameter transition is not studied further, the result must be considered when future observations are forecasted with the post-break model. As pointed out by Elliott and Müller (2014), parameter estimates of the model will be distorted, if they are derived from a data sample that does not fully belong to the post-break data generating process.

The topic of the European sovereign debt crisis is shortly adopted in this chapter. The proposed abruptness test is applied to European government bond yield regression models that covers daily data between 2003 and 2016. Specifically, the abruptness of multiple shifts in the error term variance as detected by the rolling window method of Chapter 4 is tested, one break at a time. Blocks of the empirical data are resampled in a semi-parametric bootstrap to assess the significance of the applied non-abruptness test. For some structural breaks, the parameter shift abruptness hypothesis cannot be rejected at usual confidence levels, but a large number of structural breaks are associated with more gradual change according to SLR testing. The study will highlight that it is important, in which range likelihood ratio values around the maximum are taken into account, an observation that is also made in the Monte Carlo simulation section of the chapter.

The rest of the chapter begins with a description of model and methodology in Section 5.2. Since the method builds on the design of the MLR test, that test is summarized first, followed by a description and an illustration of the proposed non-abruptness testing. Section 5.3 contains various Monte Carlo simulations on transitions in normal data, for a simple and fundamental illustration of the necessity and validity of the described method. The empirical study on European government bond yield breaks is found in Section 5.4, and conclusions are found in Section 5.5.

5.2 Methodology

5.2.1 The model

Consider a time series with T data points $t = 1, \dots, T$. The dependent variable has n observations at each point in time. Its data generating process (DGP) can be described by a regression model, $E(y|x) = f(x, \theta^c, \theta^b)$. Among the parameters that describe the model is the set θ^c that contains $c \geq 0$ parameters that are constant for all t . Another set of $b \geq 1$ parameters θ^b is dynamic: Each parameter θ_j in the set can be described by its starting value θ_j^{start} , its ending value θ_j^{end} and a monotonically increasing function in t with target set $[0, 1]$. Let $MT_j(t, \gamma_j, k_0)$ be this monotonically increasing function in t ,

so that

$$\begin{aligned} \theta_j(t) &= \theta_j^{start} + MT_j(t, \gamma_j, k_0) \cdot (\theta_j^{end} - \theta_j^{start}), \quad \text{for } j = 1, \dots, b, \\ MT_j(t, \gamma_j, k_0) &: \mathbb{R} \mapsto \mathbb{R} \in [0; 1], \quad \text{for } j = 1, \dots, b. \end{aligned}$$

Then $\gamma_j \in \mathbb{R}^+$ is a transition speed parameter. All parameter transitions share a single location parameter $k_0 \in \mathcal{T}_\kappa$, where $\mathcal{T}_\kappa = [\omega, T - \omega]$ is the interval of permitted break dates. The limiting dates are determined from the trimming parameter κ introduced by Andrews (1993) for the permitted position of structural breaks, such that $\kappa \in (0, 0.5]$ and $\omega = \lceil \kappa T \rceil$, and where $\lceil \cdot \rceil$ denotes an integer part.

Apart from the possibility of gradual parameter transition, the DGP fulfills the assumptions stated in prior literature on maximum likelihood ratio tests for structural breaks. Appendix 2.2.1 reports this set of general assumptions. Importantly, the time series must be weakly stationary, but conditional heteroscedasticity in the covariance matrix Σ_ε of the error term ε is permitted. Abrupt change in the value of parameter θ_j can be expressed by an extreme case of the above model, in which $MT_j = 0$ for $t = 1, \dots, k$, $MT_j = 1$ for $t = k_0 + 1, \dots, T$ and k_0 identifies the break date.

One example of the function MT_j is the smooth transition function, such that

$$ST(t, \gamma, k_0) = \frac{1}{1 + e^{-\gamma(t-k_0)}}, \quad \text{for } t = 1, \dots, T,$$

where $\gamma \in (0, \infty)$, such that $\tau = \frac{10}{\gamma T}$ is the transition duration relative to the data length. Here, $k_0 \in \mathcal{T}_\kappa$ is the center of the transition. Then, the abrupt change case corresponds to $\gamma \rightarrow \infty \Leftrightarrow \tau \rightarrow +0$. This smooth transition function will be used for illustration purposes in this methodology section and will be used in the Monte Carlo study in Section 5.3.

5.2.2 Structural break testing

A requirement of the proposed non-abruptness test is that a joint structural break has been detected in the set of dynamic parameters. The hypotheses of the structural break test only compare the starting and ending values of the relevant parameters. That is,

$$\begin{aligned} H_0 &: \theta_j^{start} = \theta_j^{end}, \quad \text{for all } j = 1, \dots, b. \\ H_A &: \theta_j^{start} \neq \theta_j^{end}, \quad \text{for at least one } j \in \{1, \dots, b\}. \end{aligned}$$

If the null hypothesis holds, the relevant parameters are constant in t and so the transition function MT_j that is used to describe the model becomes irrelevant. Given that the null hypothesis has been rejected, however, a structural break date has been estimated and the type of parameter transition can be studied.

The maximum likelihood ratio (MLR) test detects a structural break that involves the specified parameters and locates the break endogenously. For this, a likelihood ratio is computed for all permitted break dates $k \in \mathcal{T}_\kappa$. The date \hat{k} that corresponds to the largest recorded likelihood ratio is the estimate for the true break date k_0 , if the test

statistic is significant. That is,

$$\hat{k} = \arg \max_{k \in \mathcal{T}_\kappa} LR(k),$$

$$LR(k) = 2(l_A^*(k) - l_0^*).$$

A maximized function value is denoted by an asterisk. For $LR(k)$, a restricted log-likelihood function l_0 is maximized with respect to a set of only constant parameters, such that $\theta_j^{start} = \theta_j^{end}$, for all $j = 1, \dots, b$. Further, an unrestricted log-likelihood function l_A is maximized with respect to a constant parameter set θ^c and a dynamic set of parameters θ^b . Each parameter θ_j contained in the dynamic set is free to have one constant value before date k and another constant value from date $k + 1$ onwards.

Note that the considered transition model contains a single parameter transition, only. Current methods permit the detection of multiple structural breaks at once. Chapter 2 provides an overview of MLR for multiple breaks in multivariate data, including a description of the limiting distribution of the test statistic. The limiting distribution can be used for the derivation of critical values, also in the single break case considered in this chapter, setting $M = 1$. If multiple breaks have been detected in an empirical application, one option is to split the sample such that only a single break exists in each resulting sample segment. Then, each segment can be investigated separately, a practice that is shown in the empirical application of Section 5.4.

The detection of a structural break by MLR produces pre-break and post-break estimates of each dynamic parameter, $\hat{\theta}_j^{pre}$ and $\hat{\theta}_j^{post}$, for all $j = 1, \dots, b$. These are only valid estimates of θ_j^{start} and θ_j^{end} , respectively, in case the abruptness hypothesis holds. If the parameter change is non-abrupt, the pre-break estimate will reflect both the starting value of a parameter and its values during the first part of the transition. The analogous implication holds for the post-break estimate. As a result, each estimated parameter shift $\hat{\Delta}_{\theta_j}$ is expected to underestimate the overall magnitude of the parameter transition from θ_j^{start} to θ_j^{end} , when the transition is non-abrupt. Effectively, the MLR test makes the abrupt change assumption in its test statistic, but the alternative test hypothesis is not equally specific. This motivates the further analysis of the assumed abruptness of a detected parameter change in the next section.

5.2.3 Proposed non-abruptness test

At this point, it is assumed that the hypothesis has been rejected that $\theta_j^{start} = \theta_j^{end}$, for all $j = 1, \dots, b$. Building on this structural break assumption, the next test concerns the speed of parameter transition. Under the new null hypothesis, the transition in all b parameters that have been free to shift their value happens abruptly between date k_0 and date $k_0 + 1$. Under the alternative hypothesis, this restriction is not made, so that the transition can take a more general form as described formally in the model of Section 5.2.1. That is,

$$H_0 : \theta_j = \begin{cases} \theta_j^{start}, & \text{for } t = 1, \dots, k_0, \quad \text{for all } j = 1, \dots, b, \\ \theta_j^{end}, & \text{for } t = k_0 + 1, \dots, T, \quad \text{for all } j = 1, \dots, b, \end{cases}$$

$$H_A : \theta_j(t) = (\theta_j^{start}, \theta_j^{end}), \quad \text{for } t = a_0 + 1, \dots, b_0, \quad \text{for at least one } j \in \{1, \dots, b\},$$

where $a_0 = k_0 - \lceil v_0 T/2 \rceil$, $b_0 = a_0 + v_0 T$, with true transition duration $v_0 \geq 1/T$, and where $\lceil \cdot \rceil$ rounds a number to the next largest integer. Under the alternative hypothesis, there exists at least one date k_0 , for which at least one parameter θ_j lies between values θ_j^{start} and θ_j^{end} .

Having applied the MLR test, all necessary values are available to calculate the SLR statistic. The SLR statistic Υ_v determines the average LR value in a range of dates surrounding the maximum at date \hat{k} , relative to the maximum LR itself. That is,

$$\begin{aligned} \Upsilon_v &= \frac{1}{\lceil vT \rceil + 1} \sum_{t=a}^b \frac{LR(k)}{LR(\hat{k})}, \\ a &= \max \left(\omega, \min \left(\hat{k} - \lceil vT/2 \rceil, T - \omega - \lceil vT \rceil \right) \right), \\ b &= \min \left(T - \omega, \max \left(\hat{k} + \lceil vT/2 \rceil, \omega + \lceil vT \rceil \right) \right). \end{aligned} \quad (5.1)$$

The division by $LR(\hat{k})$ normalizes the LR values, such that $\frac{LR(k)}{LR(\hat{k})} \leq 1$ for all $k \in \mathcal{T}_\kappa$. This also normalizes the test statistic, such that $\Upsilon_v < 1$ for $v \geq 1/T$. v is the summation range, which specifies the number of LR values that are added. A straightforward choice for v is v_0 , but the true transition duration is not known prior to testing. The upper limitation for values of v is the availability of LR values, so that v is restricted by the trimming parameter κ used in MLR testing, $v \in [1/T, 1 - 2\kappa]$. The parameters a and b determine the first and last date of the summation, respectively. When \hat{k} lies sufficiently close to the center of the sample and v is sufficiently small, then LR values from the $\lceil vT/2 \rceil$ dates before and from the $\lceil vT/2 \rceil$ dates after \hat{k} are used in the test statistic. Otherwise, the summation range is restricted at ω and at $T - \omega$, which are the first and the last date with an associated LR value, respectively. If $v = 1 - 2\kappa$, all LR values available from MLR testing are used in the proposed test statistic Υ_v . Then also, $a = \omega$ and $b = T - \omega$ must hold.

The null hypothesis of an abrupt parameter change is rejected in favor of the alternative hypothesis, if Υ_v produces values that are larger than those typically calculated for time series with an abrupt parameter change. Critical values of the test must be obtained empirically from data that contains an abrupt parameter change. One option is to perform a parametric, residual resampling bootstrap, see also Chapter 2. For this, the model parameters $\{\hat{\theta}^c, \hat{\theta}^{pre}, \hat{\theta}^{post}\}$ associated with \hat{k} from the estimation of a structural break model during MLR testing are used. In a given bootstrap simulation, a synthetic observation y_t^* is obtained as in the following,

$$\begin{aligned} y_t^* &= \begin{cases} \hat{f}(x_t; \hat{\theta}^c, \hat{\theta}^{pre}) + \hat{\varepsilon}_i, & \text{for } t = 1, \dots, \hat{k}, \\ \hat{f}(x_t; \hat{\theta}^c, \hat{\theta}^{post}) + \hat{\varepsilon}_j, & \text{for } t = \hat{k} + 1, \dots, T, \end{cases} \\ \hat{\varepsilon}_i &\in \{\hat{\varepsilon}_1, \dots, \hat{\varepsilon}_{\hat{k}}\}, \\ \hat{\varepsilon}_j &\in \{\hat{\varepsilon}_{\hat{k}+1}, \dots, \hat{\varepsilon}_T\}, \end{aligned}$$

where \hat{f} is the estimated functional form of the model and where $\hat{\varepsilon}_i, \hat{\varepsilon}_j$ are randomly drawn (with replacement). Another option is to use a non-parametric, case-resampling bootstrap. Here, the data before and after the estimated break date is randomly dispersed

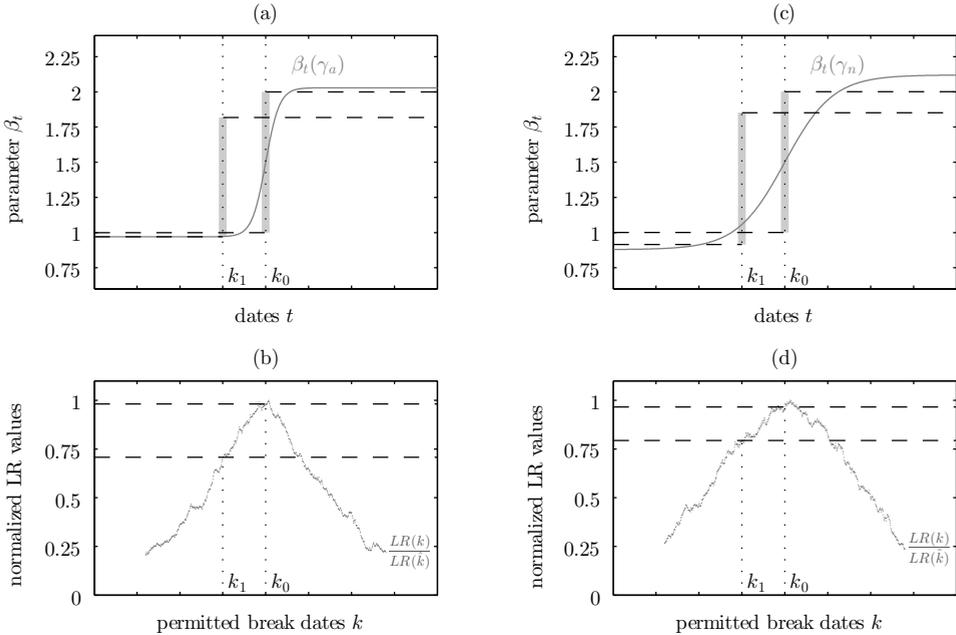
to erase any gradual transition, in expectation. Per bootstrap simulation, this involves random draws and a resampling of the data that lies before the estimated break date, and random draws and a resampling of the data that lies behind the estimated break date. That is,

$$(y_t, x_t)^* = (y_j, x_j), \text{ where } j \in \begin{cases} \{1, \dots, \hat{k}\}, & \text{for } t = 1, \dots, \hat{k}, \\ \{\hat{k} + 1, \dots, T\}, & \text{for } t = \hat{k} + 1, \dots, T. \end{cases}$$

If the original data has conditionally heteroscedastic disturbances, it is advisable to mimic the volatility clustering in either bootstrap method. If the clusters are not parametrically simulated, the original clusters can be preserved by building blocks of data from which to sample. In any case, data is simulated a large number of times, the corresponding values of the SLR statistic Υ_v are recorded, and critical values are derived.

5.2.4 Illustration of the test statistic

Figure 5.1: Visualization of two parameter transitions



Note: Smooth parameter transition over time in a univariate time series with intercept term β_t and with standard normal errors. The sample has length $T = 1000$, and the central transition date is $k_0 = 0.5T = 500$. Left panels (a) and (b) show a relatively abrupt transition of β_t with $\gamma_a = 1/10 \Leftrightarrow \tau_a = 0.1$, while right panels (c) and (d) show a non-abrupt transition with $\gamma_n = 1/70 \Leftrightarrow \tau_n = 0.7$. The upper panels (a) and (c) show plots of $\beta_t(\gamma)$ over time, and mark pre-break and post-break means of β_t for separations at $k_0 = 500$ and $k_1 = 375$ as dashed lines. The lower panels (b) and (d) show the corresponding scatterplots of $\frac{LR(k)}{LR(\hat{k})}$ for all permitted break dates k , and mark values of $\frac{LR(k)}{LR(\hat{k})}$ for $k_0 = 500$ and $k_1 = 375$ as dashed lines.

Consider data generated from a simple univariate regression model, where an intercept term β_t transitions smoothly between two values. Figure 5.1 compares two such transition specifications, one transition more abrupt (left panels) than the other (right panels).¹ In both cases the series contain $T = 1000$ dates and the central transition date is $k_0 = 500$. The two upper panels (a) and (c) show the respective plot of β_t over time t . The dashed horizontal lines show the mean parameter values, when the sample is divided by either the central transition date $k_0 = 500$ or some earlier date $k_1 = 375$. The lower panels (b) and (d) plot the corresponding normalized LR values $\frac{LR(k)}{LR(\hat{k})}$ for all permitted break dates k , including dates k_0 and k_1 . The LR values are obtained from a simulation of the simple regression model, with standard normal error term. The trimming parameter $\kappa = 0.15$ prevents the computation of LR values for possible break dates at the edges of the sample, as they are not in the set of permitted break dates \mathcal{J}_κ .

Consider first panel (a) of Figure 5.1 with a relatively abrupt parameter change, $\gamma_a = 1/10 \Leftrightarrow \tau_a = 0.1$. The mean parameter value is $\beta^{pre} = 1$ before k_0 and $\beta^{post} = 2$ after k_0 . The difference in means is $\Delta_\beta = 1$, while the difference between starting and ending values is slightly larger. If the sample is split at k_1 instead, β^{pre} decreases minimally, because a few observations during the transition segment are not included in the mean calculation. Now that a significant amount of observations from the pre- k_0 segment are included in the mean calculation for β^{post} , its value decreases and the difference in means decreases by more than 15% to $\Delta_\beta = 0.848$. The corresponding normalized LR values in panel (b) also reflect this decrease in parameter difference. The LR value for k_0 is almost equal to the maximum LR value at $\frac{LR(k_0)}{LR(\hat{k})} = 0.982$, while the LR value for k_1 is at $\frac{LR(k_1)}{LR(\hat{k})} = 0.707$ of the maximum.

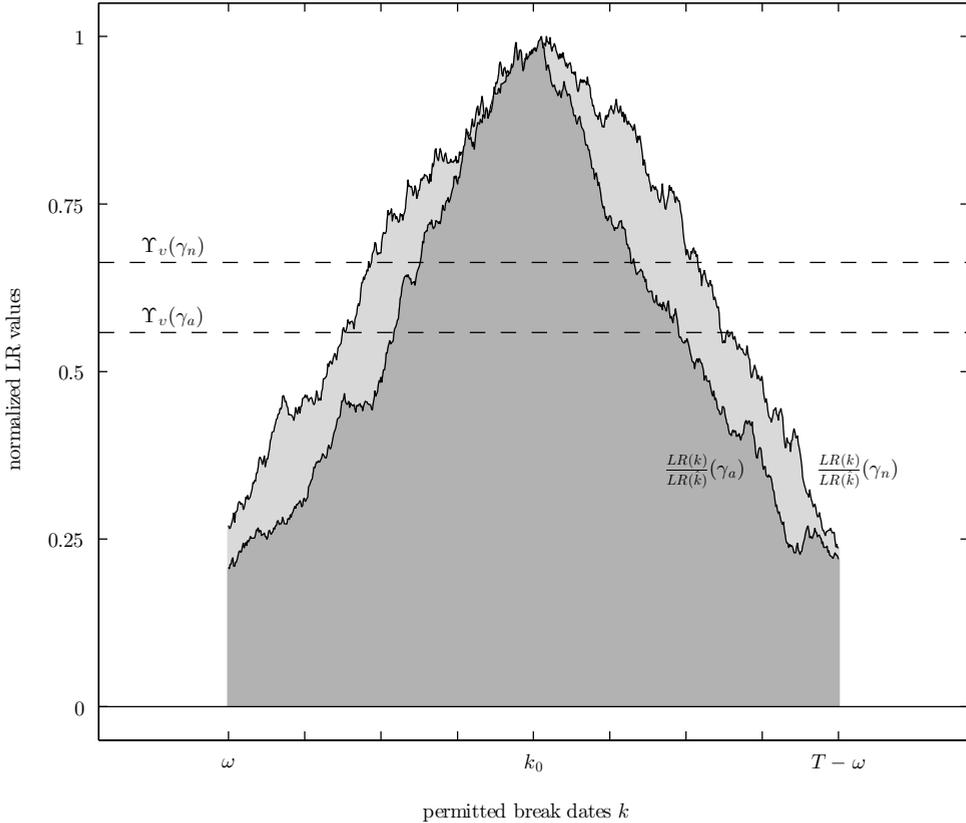
If the parameter transitions more slowly with $\gamma_n = 1/70 \Leftrightarrow \tau_n = 0.7$, the normalized LR value for k_0 is still close to 1 at $\frac{LR(k_0)}{LR(\hat{k})} = 0.966$, but now the normalized LR value for k_1 is relatively large at $\frac{LR(k_1)}{LR(\hat{k})} = 0.794$. This can be seen in panel (d) of Figure 5.1. The reason is visualized in panel (c) above. The starting and ending values of β_t have been set such that for k_0 the pre-break mean is $\beta^{pre} = 1$ and the post-break mean is $\beta^{post} = 2$, as before. So the largest difference in means is $\Delta_\beta = 1$, as before. When this series is separated at k_1 , more small values of β_t are included in the post-break mean and so β^{post} decreases, as before. However, the slower transition implies that β^{post} decreases less than before. Also, the decrease in β^{pre} is now very apparent, which results from the earlier separation at k_1 . Because the transition is slower, a larger fraction of relatively large values of β_t are now excluded from the pre-break segment, so that the mean value decreases strongly. As a joint consequence of the smaller decrease in β^{post} and the larger decrease in β^{pre} , the difference in means decreases by less, from $\Delta_\beta = 1$ at k_0 to $\Delta_\beta = 0.936$ at k_1 . This explains why the LR value for k_1 will tend to be relatively close to 1, if the parameter transition is slower.

Note at this point that the LR statistic is not necessarily maximized at the point where mean parameter values before and after the considered break date differ the most. If that were true, $\frac{LR(k_0)}{LR(\hat{k})} = 1$ had to hold by definition in either plot of Figure 5.1. LR

¹For the purpose of the illustration it is sufficient to consider two transitions with different duration. The choice that neither is fully abrupt is deliberate, to show that the studied statistics are also gradually affected by the transition duration.

values reflect estimated likelihood function values and as such reflect the uncertainty of parameter estimation. The erratic plots of normalized LR values in the figure show that likelihood ratios contain information that is different from mean parameter estimates. Still, mean parameter values are a visually appealing way to connect the parameter plots to LR value plots. For an analysis of the asymptotic equivalence of the maximum LR statistic and a statistic that finds the largest estimated parameter shift, given the uncertainty of estimation, see Andrews (1993).

Figure 5.2: Visualization of test results for two parameter transitions



Note: The graph shows normalized LR values $\frac{LR(k)}{LR(\bar{k})}$ and SLR test statistics Υ_v for two simulations with different parameter transitions over time. A relatively abrupt, smooth transition of β_t with $\gamma_a = 1/10 \Leftrightarrow \tau_a = 0.1$ is compared to a non-abrupt, smooth transition with $\gamma_n = 1/70 \Leftrightarrow \tau_n = 0.7$. The data is simulated as univariate series with standard normal errors. For both simulations, the sample length is $T = 1000$, and the central transition date is $k_0 = 0.5T = 500$.

A test statistic that is set to detect the non-abruptness of a structural break can make use of the information contained in the normalized LR values around the maximum. The summation of normalized LR values is both simple and has a visual appeal. In the limit, the sum measures the area below the plot of normalized LR values. Figure 5.2 depicts the normalized LR values that are familiar from Figure 5.1 in one picture. The area

below the (linearly interpolated) LR values is shaded for an approximate visualization of the idea. Evidently, the area below the plot with the slower parameter transition is larger than the area below the plot with a more abrupt transition.

In its final form, the proposed statistic Υ_ν of the non-abruptness test computes the average, normalized LR value, as seen in equation (5.1). The respective values of Υ_ν are depicted as horizontal, dashed lines in Figure 5.2. The summation range has been set to $\nu = 0.7$, which means that all available LR values were used in the statistic. The test statistic produces a larger value when the transition of β_t is less-abrupt, $\Upsilon_\nu(\gamma_n) = 0.662$, than when the transition is relatively abrupt, $\Upsilon_\nu(\gamma_a) = 0.558$.

5.3 Monte Carlo simulations

5.3.1 Introduction

To analyze the small sample properties of the sum of likelihood ratios (SLR) test statistic, data has been simulated using various DGP specifications. The simulated series all satisfy the following description of a univariate regression model with smooth transition of the regression coefficients, that is

$$\begin{aligned}
 y_t &= x_t' \beta_t + \varepsilon_t, & \text{for } t = 1, \dots, T, \\
 \beta_t &= \beta^{start} + ST(t, \gamma, k_0) \cdot (\beta^{end} - \beta^{start}) & \text{for } t = 1, \dots, T, \text{ for } j = 0, \dots, q, \\
 ST(t, \gamma, k_0) &= \frac{1}{1 + e^{-\gamma(t-k_0)}}, & \text{for } t = 1, \dots, T, \\
 x_{0,t} &= 1, & \text{for } t = 1, \dots, T, \\
 x_{j,t} &\sim \mathcal{N}(0, 1), & \text{for } t = 1, \dots, T, \text{ for } j = 1, \dots, q, \\
 \varepsilon_t &\sim \mathcal{N}(0, \sigma_\varepsilon^2), & \text{for } t = 1, \dots, T.
 \end{aligned}$$

The series involve T dates, and the vector of covariates x_t contains the scalar 1 and q additional stochastic variables. $q + 1$ regression coefficients measure the linear impact of the vector of covariates x_t on the dependent variable y_t . Each regression coefficient smoothly transitions from β^{start} to β^{end} , according to the smooth transition function $ST = ST(t, \gamma, k_0)$, which is centered at date k_0 . The parameter that determines the speed of transition γ can be transformed to obtain a measure of the transition duration $\tau = \frac{10}{\gamma T}$. Parameters β^{start} and β^{end} are computed such that the pre- k_0 mean of β_t equals 1, and the difference in parameter means before and after k_0 is a pre-set value Δ_β as well, such that

$$\begin{aligned}
 \beta^{start} &= 1 - \Delta_\beta \cdot \frac{\sum_{t=1}^{k_0} ST}{\frac{k_0}{T-k_0} \sum_{t=k_0+1}^T ST - \sum_{t=1}^{k_0} ST}, \\
 \beta^{end} &= 1 + \Delta_\beta \cdot \frac{k_0 - \sum_{t=1}^{k_0} ST}{\frac{k_0}{T-k_0} \sum_{t=k_0+1}^T ST - \sum_{t=1}^{k_0} ST}.
 \end{aligned}$$

For this Monte Carlo (MC) study, simulated time series samples are first subjected to the maximum likelihood ratio (MLR) statistic, to test whether the stability hypothesis

is rejected in favor of the structural break hypothesis. If so, the sample is subjected to SLR, to test whether the abruptness hypothesis is rejected in favor of the hypothesis of non-abrupt change. For each individual DGP specification, $S = 10000$ data samples have been simulated. During the application of MLR and SLR tests, prior knowledge of σ_ε is assumed and the structural breaks are correctly identified to occur in the mean parameter β .

Critical values are obtained empirically, but not from resampling the simulated series according to a bootstrap. Instead, time series samples are simulated with corresponding DGP properties, but where the tested property is known to satisfy the respective null hypothesis. This is possible because all properties are known. Specifically, to find critical values of MLR, time series are simulated where the coefficients do not shift in value, while the remaining DGP properties are equal to the series with a shift. To find critical values of SLR, time series are simulated where the coefficients abruptly shift in value, while the remaining DGP properties are equal to the series with smooth transition. As discussed in Sections 5.2.2 and 5.2.4, the MLR test produces an estimate of the parameter shift $\hat{\Delta}_\beta$, which is not an estimate of the overall transition magnitude $\beta^{end} - \beta^{start}$ for non-abrupt transitions. Therefore, smooth transition series are simulated first, then $\hat{\Delta}_\beta$ is obtained from MLR testing, then $\text{mean}(\hat{\Delta}_\beta)$ is used to simulate abrupt transition data, which then leads to critical values of the SLR test for the original smooth transition data.

5.3.2 General results

A first specification of the DGP is given by $T = 100$ and $\sigma_\varepsilon = 1$. Further, $q = 0$, so that the model only contains an intercept term. The transition is centered at $k_0 = 0.5T = 50$. The transition duration is $\tau = 0.3 \Leftrightarrow \gamma = \frac{1}{3}$. The pre- k_0 mean of β_t is set to 1, and the difference in parameter means before and after k_0 is $\Delta_\beta = 1$. With $\gamma = 1/3$, this implies $\beta^{end} - \beta^{start} = 1.091$. The data trimming for structural break testing is set to $\kappa = 0.15$, a standard value in MLR testing, see Andrews (1993) or Bai and Perron (2003). Further, with $v = 0.7$ the sum of LR values is taken over the full set of 71 permitted break dates in the interval $[15, 85]$.

The results of applying MLR and SLR to the simulated series are reported in Table 5.1. The first column lists the results when no break is simulated. As these simulations are taken as the basis for deriving critical values for MLR testing, each break-detection frequency $I_{1-\alpha}^{bd}$ must be equal to the nominal size α of the test. Between the 5% of simulations that have led to the rejection of the null hypothesis of parameter stability at a confidence level of $1 - \alpha = 0.95$, MLR estimates the break to lie at $k = 50.670$ on average, with a substantial spread of individual estimates according to the standard deviation $\text{sd}(\hat{k}) = 22.949$.

When there is an abrupt parameter shift of $\beta^{end} - \beta^{start} = 1.091$, the power of MLR is close to or equal to 1 at all reported confidence levels, according to the second column of Table 5.1. The remaining results are unsurprising, as $\text{mean}(\hat{k}) = 49.952$ lies close to the true break date $k_0 = 50$, $\text{sd}(\hat{k}) = 5.177$, and the mean estimated parameter shift magnitude is close to its true value at $\text{mean}(\hat{\Delta}_\beta) = 1.131$. Shift size Δ_β is slightly overestimated, which can be explained as MLR is expected to locate the break at the

Table 5.1: General testing results

		Transition type		None	Abrupt	Smooth	
		Δ_β		0	*1.091	1	
		$\beta^{end} - \beta^{start}$		0	*1.091	1.091	
MLR	Power	$I_{1-\alpha}^{bd}$	$\alpha = 0.01$	0.010	0.986	0.983	
			0.05	0.050	0.998	0.997	
			0.10	0.100	1.000	0.999	
	Location	mean(\hat{k})		50.670	49.952	49.517	
		sd(\hat{k})		22.949	5.177	7.252	
	Shift	mean($\hat{\Delta}_\beta$)		-0.029	1.131	*1.091	
		sd($\hat{\Delta}_\beta$)		0.744	0.191	0.187	
	SLR	Power	$I_{1-\alpha}^{ar}$	$\alpha = 0.01$	-	0.010	0.028
				0.05	-	0.050	0.106
0.10				-	0.100	0.192	

Note: Results of testing for the existence of a structural break with MLR, as well as results of testing for non-abruptness of the parameter change with SLR, where $v = 0.7$. The data follows a regression model with $T = 100$, $\sigma_\varepsilon = 1$, $q = 0$, $k_0 = 50$. For the smooth transition, $\Delta_\beta = 1$, $\beta^{end} - \beta^{start} = 1.091$, and $\gamma = 1/3 \Leftrightarrow \tau = 0.3$. For the abrupt transition, $\Delta_\beta = \beta^{end} - \beta^{start}$ is set equal to the estimated shift in the non-abrupt setting mean($\hat{\Delta}_\beta$), all marked with an asterisk. For MLR location and shift measures, and for SLR power indicators, only simulations are used in which MLR has detected the structural break with 95% confidence.

largest estimated parameter shift in the sample.² Finally, moving to the lower part of Table 5.1, the simulations are used as a basis for the derivation of critical values for the SLR test, so power equals test size by definition. This can be seen by the values of the abruptness-rejection frequencies $I_{1-\alpha}^{ar}$, for all $\alpha = 0.01, 0.05, 0.10$. Note that here, as in all reported results of the MC study, abruptness-rejection frequencies are computed for simulations that have led to the rejection of parameter stability via MLR at 95% confidence, only.

The last column of Table 5.1 provides the results when the parameter transitions smoothly, as specified before, with $\tau = 0.3$. MLR testing leads to the detection of a structural break almost equally often as when the break is abrupt. At a 95% confidence level, for example, the structural break is detected in $I_{0.95}^{bd} = 99.8\%$ of simulations when the break is abrupt, while it is detected in $I_{0.95}^{bd} = 99.7\%$ of simulations when the break is not abrupt. The break date is located slightly before the central transition date $k_0 = 50$ in the mean, as mean(\hat{k}) = 49.517. The estimated parameter shift, mean($\hat{\Delta}_\beta$) = 1.091, is higher than the difference in parameter means before and after k_0 , $\Delta_\beta = 1$. As explained before, the estimate is used to calibrate the magnitude of the abrupt-transition simulations. The values have been marked with an asterisk in the table to emphasize the equality.³

Most importantly, the proposed SLR method is shown to have some power for the detection of the non-abrupt parameter transition. Abruptness-rejection frequencies are reported as $I_{0.99}^{ar} = 0.028$, $I_{0.95}^{ar} = 0.106$, and $I_{0.90}^{ar} = 0.192$. The test clearly fulfils the

²When the mean of estimated parameter shifts is taken over all simulations, including the ones that did not lead to the rejection of the null hypothesis, the positive bias still exists at mean($\hat{\Delta}_\beta$) = 1.130. As the break detection method optimizes the data fit, the model may be overfit to the specific data at hand.

³The estimate is also equal to $\beta^{end} - \beta^{start}$ of the smooth transition itself, but only to the third digit. This will not be a general observation, but can coincidentally occur because MLR tends to overestimate parameter shifts.

Table 5.2: Test results for different specifications of the DGP

DGP specifications			MLR		SLR		
	Δ_β	$\beta^{end} - \beta^{start}$	$I_{0.95}^{bd}$	$\text{mean}(\hat{\Delta}_\beta)$	$I_{0.99}^{ar}$	$I_{0.95}^{ar}$	$I_{0.90}^{ar}$
$\tau = 0.1$	1	1.029	0.994	1.069	0.013	0.062	0.114
0.2	1	1.059	0.996	1.081	0.016	0.076	0.131
0.3	1	1.091	0.997	1.091	0.028	0.106	0.192
0.4	1	1.125	0.997	1.097	0.031	0.125	0.222
0.5	1	1.161	0.998	1.107	0.045	0.167	0.281
0.6	1	1.200	0.998	1.117	0.058	0.199	0.328
0.7	1	1.241	0.999	1.124	0.062	0.238	0.382
$\Delta_\beta = -0.5$	-0.5	-0.546	0.594	-0.769	0.013	0.065	0.131
0.5	0.5	0.546	0.602	0.768	0.014	0.066	0.126
1	1	1.091	0.997	1.091	0.028	0.106	0.192
2	2	2.182	1.000	2.057	0.108	0.329	0.476
$\sigma_\varepsilon = 0.25$	1	1.091	1.000	1.028	0.106	0.313	0.456
0.5	1	1.091	1.000	1.028	0.102	0.287	0.437
1	1	1.091	0.997	1.091	0.028	0.106	0.192
2	1	1.091	0.622	1.526	0.014	0.062	0.122
$T = 50$	1	1.092	0.899	1.202	0.018	0.074	0.142
100	1	1.091	0.997	1.091	0.028	0.106	0.192
200	1	1.091	1.000	1.052	0.046	0.159	0.272
500	1	1.091	1.000	1.029	0.132	0.367	0.525
1000	1	1.091	1.000	1.019	0.376	0.672	0.795
$q = 0$	1	1.091	0.997	1.091	0.028	0.106	0.192
1	1	1.091	1.000	1.043	0.027	0.136	0.239
2	1	1.091	1.000	1.032	0.036	0.154	0.272
$k_0 = 30$	1	1.116	0.989	1.110	0.017	0.082	0.160
40	1	1.098	0.996	1.098	0.024	0.099	0.188
50	1	1.091	0.997	1.091	0.028	0.106	0.192
60	1	1.093	0.996	1.090	0.020	0.100	0.182
70	1	1.105	0.989	1.101	0.019	0.080	0.164

Note: Results of testing for the existence of a structural break with MLR, as well as results of testing for non-abruptness of the parameter change with SLR, where $v = 0.7$. Parameter values of the DGP are sequentially adjusted, with the remaining parameters always given by $\tau = 0.3$, $\Delta_\beta = 1$, $\sigma_\varepsilon = 1$, $T = 100$, $q = 0$, and $k_0 = 0.5T$. Δ_β determines $\beta^{end} - \beta^{start}$. For $\text{mean}(\hat{\Delta}_\beta)$ and $I_{1-\alpha}^{ar}$, only simulations are used in which MLR has detected the structural break with 95% confidence.

intended task, but only in relatively few simulations. It is intuitive that the test is less powerful than MLR, because SLR seeks to distinguish features that are joined under one hypothesis in structural break testing. The low testing power is also intentional in so far as to clearly illustrate how changing certain simulation parameters will affect the power in the upcoming simulations.

5.3.3 Testing power

In the following, the power of SLR is analyzed for a variety of DGP specifications, as seen in Table 5.2. Each individual property is adjusted with respect to the reference setting of $\tau = 0.3$, $\Delta_\beta = 1$, $\sigma_\varepsilon = 1$, $T = 100$, $q = 0$, and $k_0 = 0.5T$, so that the individual impact on testing is isolated. The difference in means at k_0 , Δ_β , and the parameter shift magnitude, $\beta^{end} - \beta^{start}$, are included in the table. Further, two indicators from MLR testing are reported that relate to power and shift estimation, respectively. The three main indicators in the table are the abruptness-rejection frequencies $I_{1-\alpha}^{ar}$ at confidence levels $1 - \alpha = 0.99, 0.95, 0.90$, when the simulated series are subjected to SLR testing with $v = 0.7$.

The first and most important result is that the increase in the parameter shift duration τ increases the power of SLR testing. Abruptness-rejection frequencies rise for all confidence levels when τ increases from $\tau = 0.1$ to $\tau = 0.7$. As discussed before, $\Delta_\beta = 1$ has been set in all of these simulations. The intention is to imitate the practical setting, in which MLR estimates values of pre- and post-break parameters before the SLR test is applied. Given $\Delta_\beta = 1$, the parameter shift magnitude $\beta^{end} - \beta^{start}$ must rise for larger transition durations. Also, the mean estimated parameter shift rises from $\text{mean}(\hat{\Delta}_\beta) = 1.069$ to $\text{mean}(\hat{\Delta}_\beta) = 1.124$. To separate magnitude and transition duration effects, additional data has been simulated with a relatively abrupt transition of $\tau = 0.3$, but with the highest reported magnitude of $\beta^{end} - \beta^{start} = 1.241$. In these simulations, SLR has power of only $I_{0.95}^{ar} = 0.115$. Compared to the other values reported for $\tau = 0.3$ with $\Delta_\beta = 1$ and $\tau = 0.7$ with $\Delta_\beta = 1$, only an increase of $0.115 - 0.106 = 0.009$ in the overall increase of power can be attributed to the adjustment of starting and ending values of β_t . At least the remaining increase of $0.238 - 0.115 = 0.123$ is attributed purely to the increase in parameter transition duration.

To further see how the mean parameter shift affects the power of SLR when $\tau = 0.3$ stays unchanged, $\Delta_\beta = -0.5, 0.5, 1, 2$ are considered in the next section of Table 5.2. An increase in the absolute value from $\Delta_\beta = 0.5$ to $\Delta_\beta = 2$ increases the break-detection frequency from $I_{0.95}^{bd} = 0.602$ to $I_{0.95}^{bd} = 1$ and also the abruptness-rejection frequency from $I_{0.95}^{ar} = 0.066$ to $I_{0.95}^{ar} = 0.329$. The larger shift increases the ability of the tests to detect the transition and to detect the non-abruptness against the noise in the data, given the limited number of observations $T = 100$. The overestimation of the parameter shift is also reduced when $\Delta_\beta = 2$, so that $\text{mean}(\hat{\Delta}_\beta)$ is only 0.057 above the true difference in means, compared to $\Delta_\beta = 1$ where $\text{mean}(\hat{\Delta}_\beta)$ is 0.091 above the true difference. The sign of the parameter change has no apparent impact.

The remaining results mostly follow along the lines of the results discussed so far. The effects of decreasing σ_ε or increasing T on the power of SLR testing are comparable to the effects of increasing Δ_β . Everything else equal, abruptness-rejection frequencies increase with a reduction of the noise in the data, and increase with an increase in the number of available observations, as well. A similar observation is made for increases in the number of covariates q . The addition of stochastic predictor variable with equally dynamic coefficients β_t tends to increase the testing power of SLR. The impact is, however, relatively small. Going from $q = 0$ to $q = 1$ increases the abruptness-rejection frequency from $I_{0.95}^{ar} = 0.106$ to $I_{0.95}^{ar} = 0.136$ at 95% and even decreases $I_{0.99}^{ar} = 0.028$ to $I_{0.99}^{ar} = 0.027$ at 99%.

Moving the central transition date k_0 towards the edges of the permitted break date set lowers the abruptness-rejection frequency by $0.106 - 0.082 = 0.024$ when $k_0 = 30$, and by $0.106 - 0.080 = 0.026$ when $k_0 = 70$. It is not surprising that SLR performs worse at off-center transition positions; a comparable reduction in power is recorded for break detection tests such as MLR in other small sample studies (see, for example, Chapter 4). The following final section of the Monte Carlo study answers the question, whether the testing power of SLR can be increased by an adjustment of the summation range v .

Table 5.3: The impact of v on SLR power

Central summation date		k_0				\hat{k}			
		$\tau = 0.1$	0.3	0.5	0.7	$\tau = 0.1$	0.3	0.5	0.7
$I_{0.99}^{ar}$	$v = 0.1$	*0.014	0.013	0.016	0.014	0.011	0.013	0.015	0.015
	0.2	0.014	0.024	0.035	0.035	0.013	0.019	0.026	0.029
	0.3	0.012	0.027	0.046	0.056	*0.014	0.025	0.039	0.043
	0.4	0.011	0.033	0.050	*0.079	0.013	0.028	0.040	0.052
	0.5	0.012	*0.034	0.049	0.078	0.013	0.028	0.040	0.054
	0.6	0.013	0.031	*0.050	0.068	0.012	*0.030	*0.049	*0.063
	0.7	0.013	0.028	0.045	0.062	0.013	0.028	0.045	0.062
$I_{0.95}^{ar}$	$v = 0.1$	0.063	0.069	0.070	0.054	0.055	0.065	0.073	0.069
	0.2	0.060	0.100	*0.128	0.134	0.058	0.087	0.115	0.118
	0.3	0.061	0.115	0.170	0.198	0.061	0.105	0.146	0.157
	0.4	0.062	*0.115	*0.179	0.224	0.061	0.102	0.153	0.184
	0.5	0.062	0.113	0.176	0.236	0.061	0.103	0.162	0.210
	0.6	*0.065	0.110	0.172	*0.241	*0.063	0.106	*0.170	0.235
	0.7	0.062	0.106	0.167	0.238	0.062	*0.106	0.167	*0.238
$I_{0.90}^{ar}$	$v = 0.1$	*0.121	0.131	0.121	0.100	0.115	0.130	0.141	0.137
	0.2	0.115	0.192	0.218	0.219	0.110	0.177	0.200	0.216
	0.3	0.111	*0.206	0.270	0.302	0.111	0.186	0.242	0.270
	0.4	0.114	0.206	*0.292	0.353	0.113	0.190	0.264	0.320
	0.5	0.118	0.200	0.291	0.372	0.111	*0.197	0.279	0.354
	0.6	0.115	0.194	0.292	*0.382	*0.116	0.194	*0.286	0.378
	0.7	0.114	0.192	0.281	0.382	0.114	0.192	0.281	*0.382

Note: Abruptness-rejection frequencies for different values of summation range v . $S = 10000$ samples are simulated for each DGP specification. All data follows a regression model with $T = 100$, $\sigma_\varepsilon = 1$, $q = 0$, $\Delta_\beta = 1$, and $k_0 = 0.5T = 50$. Results are shown for different values of the transition duration τ , using k_0 or \hat{k} as a central summation date. In each case, the v that has led to the largest value of $I_{1-\alpha}^{ar}$ is marked with an asterisk.

5.3.4 Impact of the summation range parameter

In earlier simulations, all results have been based on setting the summation range parameter to $v = 0.7$. Together with $\kappa = 0.15$ this implies that the full set of available LR values is included in the summation. Only then the central summation date is invariably given by $t = 0.5T$, but for other values of v the central summation date may change. This section describes the impact of v on SLR testing power.

Consider first that the summation of LR values is centered at k_0 , so that its value is assumed prior knowledge. The left half of Table 5.3 reports the corresponding abruptness-rejection frequencies $I_{1-\alpha}^{ar}$. The best respective settings of v are identified by the largest associated value of $I_{1-\alpha}^{ar}$, each marked by an asterisk in the table. A positive correlation between τ and the best setting for v is revealed. It is intuitive that the summation of LR values is most powerful, when the included LR values correspond to the dates during which the transition occurs. Since the duration of the transition is not known before testing, this result only underlines the appropriateness of the test statistic, but has no practical implications.

In practice the central transition date is unknown. The right half of Table 5.3 reports abruptness-rejection frequencies when SLR takes \hat{k} of each simulation as its central summation date. Recall that there is some variation in break date estimation, for example $\text{sd}(\hat{k}) = 7.252 \approx 0.073T$ for $\tau = 0.3$ (see Table 5.1). A first result is that SLR testing power tends to be smaller when k_0 must be estimated. The table also indicates that the increase in SLR power for larger transition durations is most apparent for $v = 0.7$. Generally, the highest detection frequencies are now reported for larger values of v than

Table 5.4: The impact of v on SLR power for additional DGP specifications

v	$\sigma_\varepsilon = 0.5$			$T = 500$			$k_0 = 30$		
	$\tau = 0.3$	0.5	0.7	$\tau = 0.3$	0.5	0.7	$\tau = 0.3$	0.5	0.7
0.1	0.315	0.365	0.388	0.368	0.434	0.462	0.073	0.079	0.084
0.2	*0.399	0.579	0.662	*0.501	0.707	0.798	0.097	0.125	0.141
0.3	0.388	0.656	0.773	0.491	*0.768	0.890	*0.106	0.165	0.200
0.4	0.361	*0.664	0.818	0.457	0.763	0.916	0.098	*0.167	*0.204
0.5	0.320	0.642	0.826	0.418	0.747	*0.921	0.083	0.165	0.203
0.6	0.301	0.621	*0.832	0.391	0.726	0.917	0.086	0.148	0.188
0.7	0.287	0.606	0.831	0.367	0.701	0.912	0.082	0.146	0.181

Note: Abruptness-rejection frequencies at 95% confidence for different values of summation range v . For each specification, $S = 10000$ data samples have been simulated. All data has $q = 0$ and $\Delta_\beta = 1$, and if not explicitly stated differently, also $T = 100$, $\sigma_\varepsilon = 1$, and $k_0 = 0.5T = 50$. Transition durations are either value of $\tau = 0.3, 0.5, 0.7$ and k of each simulation is used as the central summation date. In each case, the v that has led to the largest value of $I_{0.95}^{ar}$ is marked with an asterisk.

when the central summation date is k_0 . For every considered case, the best v has remained the same, or increased in value. In a few cases $v = 0.7$ is the best setting, and when not, the difference in $I_{1-\alpha}^{ar}$ between the best v and $v = 0.7$ is at most equal to 0.004 (for $\tau = 0.3$ and $\tau = 0.5$ at 90% confidence). There is support for setting $v = 0.7$, or $v = 1 - 2\kappa$ for general values of the trimming parameter κ . This is a straightforward setting, where the full range of available LR data is used. Yet, no final conclusions can be drawn before simulations with different DGP specifications are analyzed.

Table 5.4 reports abruptness-detection frequencies at 95% confidence for additional specifications of the DGP. Again, different settings of v are compared for three different transition durations τ . The first case involves a reduction of the noise in the data, such that the disturbance standard deviation equals $\sigma_\varepsilon = 0.5$. The second case involves an increase in the amount of observations, as the sample length is set to $T = 500$. Both cases represent situations that make structural break testing more powerful in general, and SLR testing in particular. However, in none of these specifications $v = 0.7$ is the best setting, not even for $\tau = 0.7$. The results indicate that now lower values of v must be considered. It can be argued that for $T = 500$, and given that the transition duration is unknown ex ante, a moderate summation range such as $v = 0.4$ is the best choice. When $\tau = 0.7$, $v = 0.4$ is almost as powerful as best setting $v = 0.5$, with a difference in $I_{0.95}^{ar}$ equal to 0.005 only. When $\tau = 0.3$, $v = 0.4$ is significantly more powerful than $v = 0.5$, with a difference in $I_{0.95}^{ar}$ equal to 0.029.

When $k_0 = 30$, so that the central transition date is off-center in the sample of $T = 100$ observations, SLR testing also benefits from smaller values for v . Recall that Table 5.2 reported that at $\tau = 0.3$, abruptness-detection frequencies decreased, when the central transition date moved off-center. In Table 5.4, it can be seen that $I_{0.95}^{ar} = 0.106$ is achieved for $k_0 = 30$, which is just as high as the highest abruptness rejection frequency for data with central transition date. This is only achieved when the summation range is lowered to $v = 0.3$. An off-center location of the break date need not be associated with a decrease in testing power, if the summation range is moderately sized.

The overall implication of this section of the Monte Carlo study is that larger summation ranges are appropriate, when a structural break has been detected in the middle of the sample, with relatively low confidence. Additional simulations have revealed that there are situations where SLR testing benefits from using moderate summation ranges v instead of large ranges. Situations that are favorable to break testing in general are

one such category, with many observations and little noise. In these situations, if the parameter transition is slow, testing power is relatively large for any setting of v . However, if the parameter transitions relatively quickly, crucial testing power can be gained from setting lower values for v . The other situation is when the central transition date lies close to the edge of the data sample. Then, again, crucial testing power can be gained from moderate summation ranges such as $v = 0.3$ when the transition is quick, without causing a large reduction in testing power when the transition is slow.

5.4 Empirical Application

Chapter 4 on the detection of structural breaks with a rolling window algorithm contains an application to a sample of daily redemption yield observations for 18 European 10-year government bonds. The data covers 3640 dates between 2003-01-09 and 2015-12-21. Linear regression models for differences in the redemption yield, Δr_i , are estimated, see Table 4.7. For most of the countries the corresponding model contains an autoregressive component and other significant, predetermined financial series, such as U.S. bond yield differences $\Delta(r_{US})_{t-1}$, German bond yield differences $\Delta(r_{BD})_{t-1}$ and global series such as the MSCI World stock exchange index, $\Delta(\ln s_{WR})_{t-1}$. For each of the 18 countries, the detection algorithm finds six significant structural breaks in the standard deviation of the error term. In the following, the abruptness of the detected structural breaks is investigated.

Each break date is analyzed separately, for $m = 1, \dots, 6$. For this, data set segments are taken from the full sample of a series, each segment centered around a break date estimate. Each data segment is chosen to contain the same number of observations as a data window in the break detection method, $T_k = 2\omega = 2(\lceil 0.1 \cdot 3640 \rceil) = 728$. If the same segments had been used during separate single break detection runs before, the LR values surrounding the maximum would have already been available for the SLR method. However, surrounding LR values were calculated in rolling windows of data, so now they will be calculated anew.⁴ The trimming parameter for calculating new LR values is set to $\kappa = 0.15$, as the segments are relatively short.

To evaluate the SLR results for each break, a semi-parametric bootstrap of resampling data blocks is performed and the SLR test is applied to the bootstrapped observations. For this, each segment of $T_k = 728$ observations is separated at \hat{k}_m into a pre- and a post-break segment, and the observations are resampled 1000 times. As the given application concerns error term volatility breaks, the residuals of the model are resampled, directly. The bootstrap method is not fully parametric, as the disturbances are not simulated anew, using the estimated pre- and post-break standard deviation values. Instead, the observed residuals are used to preserve unspecified properties of the empirical data. The residuals are, however, standardized within blocks of 15 observations to have the estimated standard deviation.

Table 5.5 reports structural break date estimates and the respective p -value of the null hypothesis of an abrupt parameter shift as tested by SLR. The summation range

⁴One could also choose to continue with the rolling window approach. Then the already calculated rolling window LR values for single breaks could be used, and instead the bootstrap method to derive critical values would be adjusted to use rolling windows as well. This method has not been tested for this study.

Table 5.5: SLR test results for a moderate summation parameter value

<i>i</i>	<i>m</i>	1	2	3	4	5	6
<i>BD</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-08-19	2011-05-05	2012-11-07	2015-04-28
	<i>p</i> -value	0.002	< 0.001	0.961	1	0.644	1
<i>BG</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-23	2010-11-15	2012-07-31	2015-01-13
	<i>p</i> -value	0.002	< 0.001	0.832	0.751	1	1
<i>CZ</i>	\hat{k}_m	2006-03-13	2007-11-14	2010-01-13	2011-06-13	2013-07-24	2015-07-30
	<i>p</i> -value	0.026	0.769	0.051	0.999	0.002	1
<i>DK</i>	\hat{k}_m	2004-06-01	2005-12-14	2007-12-06	2010-08-19	2013-01-11	2014-12-26
	<i>p</i> -value	< 0.001	0.020	< 0.001	< 0.001	0.633	1
<i>ES</i>	\hat{k}_m	2004-07-02	2007-11-27	2009-06-10	2011-06-01	2013-07-04	2014-11-28
	<i>p</i> -value	< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001
<i>FN</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-02-15	2012-10-29	2015-01-13
	<i>p</i> -value	< 0.001	< 0.001	0.611	0.992	1	0.999
<i>FR</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-14	2012-08-03	2015-04-28
	<i>p</i> -value	0.002	< 0.001	< 0.001	1	1	1
<i>GR</i>	\hat{k}_m	2004-06-01	2007-08-07	2009-01-01	2011-04-13	2013-01-31	2014-10-10
	<i>p</i> -value	0.031	< 0.001	0.001	0.999	< 0.001	< 0.001
<i>IR</i>	\hat{k}_m	2004-06-01	2005-12-12	2007-11-27	2010-07-12	2012-08-01	2014-02-10
	<i>p</i> -value	0.100	0.011	< 0.001	< 0.001	< 0.001	1
<i>IT</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-15	2011-05-26	2013-03-27	2015-02-25
	<i>p</i> -value	0.002	< 0.001	1	1	< 0.001	< 0.001
<i>NL</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-31	2012-09-27	2015-01-12
	<i>p</i> -value	0.001	< 0.001	0.930	0.020	0.999	1
<i>NW</i>	\hat{k}_m	2004-06-01	2005-12-16	2007-07-30	2009-04-06	2010-08-30	2012-01-24
	<i>p</i> -value	< 0.001	0.041	< 0.001	1	1	1
<i>OE</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-04-14	2012-09-14	2015-04-28
	<i>p</i> -value	0.068	< 0.001	0.914	1	1	1
<i>PO</i>	\hat{k}_m	2004-09-17	2006-04-13	2008-02-08	2009-08-06	2012-07-11	2014-10-17
	<i>p</i> -value	1	< 0.001	0.742	< 0.001	< 0.001	0.997
<i>PT</i>	\hat{k}_m	2004-06-01	2005-12-09	2007-11-27	2010-04-19	2013-07-22	2014-12-26
	<i>p</i> -value	0.007	0.057	< 0.001	< 0.001	1	0.808
<i>SD</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-12-08	2011-05-17	2013-09-19	2015-02-11
	<i>p</i> -value	0.402	< 0.001	0.019	1	< 0.001	0.795
<i>SW</i>	\hat{k}_m	2004-06-02	2007-10-31	2009-06-25	2011-12-21	2013-07-24	2014-12-16
	<i>p</i> -value	0.838	< 0.001	0.986	0.723	0.979	1
<i>UK</i>	\hat{k}_m	2004-06-16	2007-11-13	2010-01-21	2011-06-20	2012-11-09	2015-02-16
	<i>p</i> -value	0.027	< 0.001	< 0.001	1	0.026	< 0.001

Note: For all series $i = BD, \dots, UK$, results of testing the abruptness of six shifts in σ_ε that have been detected by a multiple structural break test. The table reports *p*-values for the null hypothesis that the corresponding structural break *m* is abrupt, according to the SLR test, using summation range parameter $\nu = 0.4$.

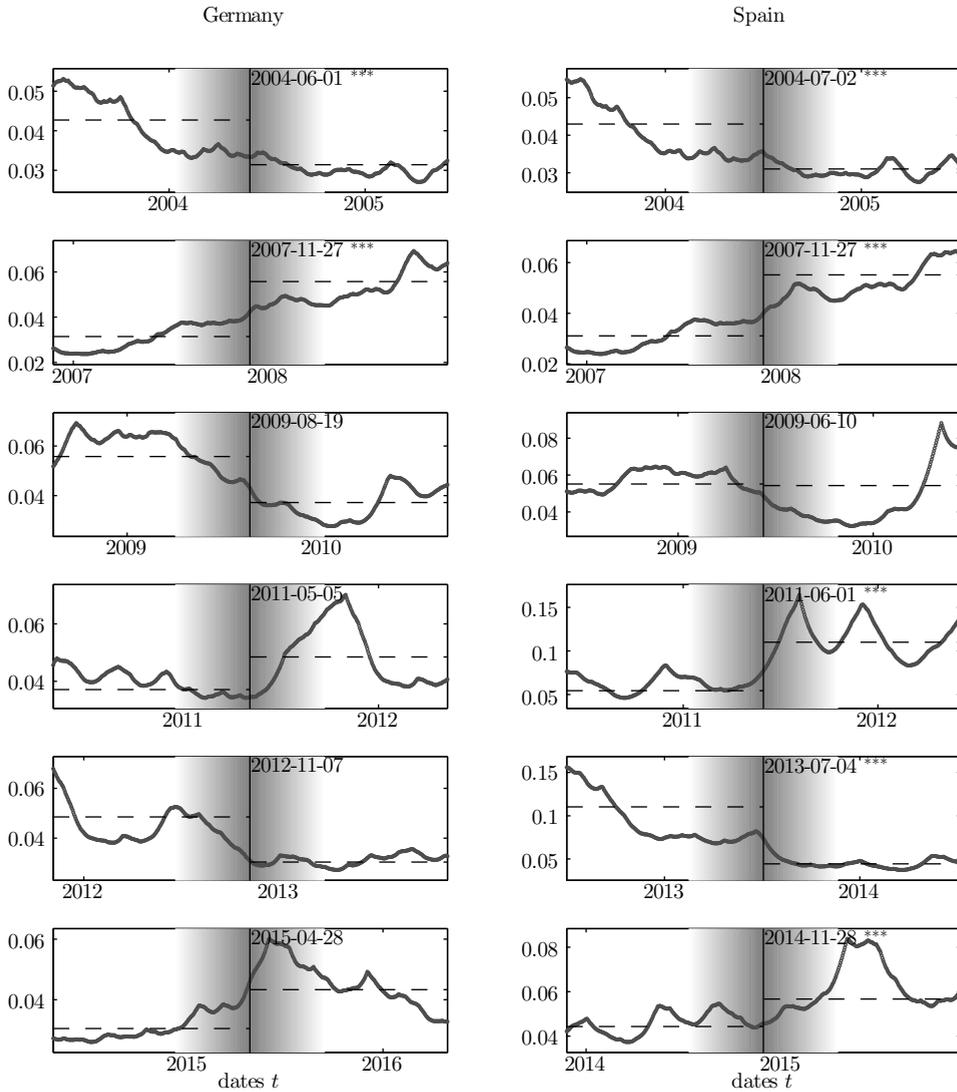
is set to a moderate $v = 0.4$ according to the Monte Carlos study recommendation for large data sets with highly significant structural breaks. Results for different values of v are found in the Appendix 5.A. The results for $v = 0.4$ are vast and mostly differ between countries, but they jointly show the importance of following up structural break testing by abruptness testing. Usually, structural breaks before 2009 are associated with volatility shifts whose abruptness is rejected by the SLR test. All structural breaks between 2004 and 2006 are associated with a reduction in volatility, and typically low p -values of the SLR test indicate that these reductions have happened gradually over time. The reason for the significance of these structural breaks is not an abruptly lowered volatility. Instead, the respective model fit is significantly improved by allowing different volatility estimates before and after the structural break, because volatility is lowered gradually and steadily.

The next noteworthy result concerns the general increase of volatility at the end of 2007, following the initial financial turmoil that would develop into the Global Financial Crisis. All 18 countries experience a structural break, as early as on 2007-07-30 (Norway) and 2007-08-07 (Greece) or late on 2008-02-08 (Poland), with the other 15 structural breaks between 2007-10-31 and 2007-12-06. Despite the small variation in structural break date estimates, the volatility increase does not seem to occur abruptly in or around November 2007. With $v = 0.4$, volatility change abruptness is rejected at 99% confidence for all series except Czech Republic and Poland. As reported in Appendix 5.A, it is rejected for all 18 series when $v = 0.6, 0.7$. This means that abruptness is confidently rejected when LR values are summed over at least $0.6 \cdot 728 \approx 737$ working days, or roughly 44 weeks before and 44 weeks after the estimated break date. This is an indication that volatility has not abruptly risen in November 2007. Instead, structural breaks at the end of 2007 are significant, because volatility slowly increased and was much higher at the end of 2008 than at the beginning of 2007.

Structural breaks towards the end of the sample are often associated with volatility shifts whose abruptness is not rejected by SLR testing. Instead of a Global Financial Crisis that has an indirect impact on government bond prices due to general financial instability and government spending, the European sovereign debt crisis directly impacts European bond markets due to concerns of sustainability and shared responsibility. Volatility abruptly declines in all central euro area countries following the announcement of European Central Bank president Mario Draghi on 2012-07-26 to do “whatever it takes” (Belgium, France, and later Germany, Finland, Netherlands, and Austria). In contrast, the volatility reduction is a non-abrupt process according to the SLR test results in euro area periphery countries Spain, Greece, Ireland, Italy and Portugal. For these, only the structural break of Ireland is estimated to occur close to Draghi’s speech. However, the SLR test results of non-abrupt change show that for the periphery countries, no individual event appears to have produced the overall change in volatility. Instead, a series of developments over time have led to a significant volatility reduction, one of which may have been the speech several weeks before the estimated break date. Clearly, the result underlines the limitations of structural break analysis without follow-up analysis such as the SLR test. Interpretations of structural breaks should not focus solely on the structural break date estimates.

To illustrate the parameter changes during the segments, estimates of σ_ε can be plotted over time. Figure 5.3 shows plots for Germany and Spain, and plots for the

Figure 5.3: Error term standard deviation over time 1/9



Note: For $i = BD, ES$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $\nu = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

remaining countries can be found in the Appendix 5.A. The graphs are each centered at one of six estimated break dates of the country. Significance of the SLR test of non-abruptness is indicated with *, **, *** for 90%, 95%, 99% confidence, respectively, when $\nu = 0.4$. Estimates associated with the segments between structural breaks are shown as dashed horizontal lines. These are calculated from the actual unexplained variation of Δr_i , not from the weighted observations used in the structural break detection. They are compared to moving average estimates $\bar{\sigma}_\varepsilon$ calculated from rolling windows of regression residuals with, such that

$$\bar{\sigma}_{\varepsilon,t} = \frac{\left(\sum_{j=t-99}^{t+99} \omega_j (\hat{\varepsilon}_j - \bar{\varepsilon})^2\right)^{0.5}}{\sum_{j=t-99}^{t+99} \omega_j},$$

where $\bar{\varepsilon} = 1/199 \sum_{j=t-99}^{t+99} \hat{\varepsilon}_j$ and where weights are discounted such that $\omega_j = 0.95^{|j-t|}$. The shaded bands mark all moving average estimates that are calculated using the observation of the respective break date estimate \hat{k}_m itself. Generally, the plots visually aide and support the SLR test findings. Often, when breaks are associated with significantly non-abrupt standard deviation change, the change in the corresponding moving averages is relatively small surrounding the break date compared to the difference in estimated pre- and post-break standard deviations.

In the first two plots for Germany, moving average estimates follow the same trend for roughly one year before and after the estimated break dates. Accordingly, the volatility change is not abrupt at 99% confidence in either case, using a moderate summation parameter of $\nu = 0.4$. Consider then the estimates of the standard deviation for each full segment between structural breaks and especially the difference between these segment-estimates. For the first two structural breaks, the moving average estimates within a shaded band barely change by half the difference in segment-estimates, further supporting the claim that the change that is captured by the structural break has not occurred abruptly. For the four later structural breaks of Germany, in contrast, the change in recursive residuals is almost as large or larger than the difference in segment-estimates. Accordingly, abruptness has not been rejected for any of these structural breaks at $\nu = 0.4$. The only instance of abruptness-rejection in these four structural breaks is for \hat{k}_5 on 2012-11-07 when $\nu = 0.6$, so when the summation includes the very low spring-2012 volatility in the market but not the high volatility between the years 2011/2012.

For Spain, five of the six volatility changes are significantly non-abrupt. The exception is the break at 2009-06-10, which can be explained by the small decrease in estimated standard deviation of only 1.55%. Given the small change between data segments, the change has occurred relatively quickly. The volatility increase of 2011-06-01 appears to be relatively abrupt, judging from the moving average estimates, but this conclusion depends on the frame of reference. It is confirmed by the SLR test for short summation ranges $\nu = 0.1, 0.2$, where the null hypothesis of abrupt parameter change is not rejected. In this case, the rapid increase to a peak volatility of the second half of 2011 is influential. In contrast, given a moderate setting of $\nu = 0.4$, the transition between data regimes is significantly non-abrupt according to the SLR test. This can be explained, as the extremely high volatility values in the second half of 2011 were only temporary.

Table 5.6: Full SLR test results for Germany and Spain

i	m	1	2	3	4	5	6	
<i>BD</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-08-19	2011-05-05	2012-11-07	2015-04-28	
	$v = 0.1$	0.102	0.027	0.920	1	0.140	1	
	0.2	0.202	< 0.001	0.508	1	0.351	1	
	0.3	0.374	< 0.001	0.681	1	0.795	1	
	0.4	0.002	< 0.001	0.961	1	0.644	1	
	0.5	< 0.001	< 0.001	0.998	1	0.365	1	
	0.6	< 0.001	< 0.001	1	1	0.071	1	
	0.7	< 0.001	< 0.001	1	1	0.128	1	
	<i>ES</i>	\hat{k}_m	2004-07-02	2007-11-27	2009-06-10	2011-06-01	2013-07-04	2014-11-28
		$v = 0.1$	0.006	< 0.001	1	0.619	0.954	0.998
0.2		< 0.001	< 0.001	1	0.942	0.319	0.939	
0.3		< 0.001	< 0.001	1	< 0.001	< 0.001	0.510	
0.4		< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001	
0.5		< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001	
0.6		< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001	
0.7		< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001	

Note: For a series $i = BD, ES$, results of testing the abruptness of six shifts in σ_ε that have been detected by a multiple structural break test. The table reports p -values for the null hypothesis that the corresponding structural break m is abrupt, according to the SLR test, using several values of the summation range parameter v .

To summarize the results, the growth in unexplained volatility of the European bond price series at the dawn of the Global Financial Crisis has occurred relatively slowly, much like the decrease in volatility before. In the later parts of the sample, which cover the European sovereign debt crisis, abruptness of volatility changes in Germany and other central European countries is typically not rejected. Such results encourage considerations of specific events that may be associated with sudden volatility changes. For Spain and other periphery countries of the euro area, significant structural breaks typically involve relatively large changes in volatility. SLR testing has indicated that these changes rarely occur rapidly, but instead accumulate gradually over time until the full extent of the change is realized. Accordingly, the use of SLR with several values of the summation range parameter v can help in the determination of type and speed of change. In contrast to the Monte Carlo simulation results, the use of small values for v is not discouraged, at least to augment other settings. Overall, the empirical results are quite diverse, but an in-depth analysis is beyond the scope of the chapter.

Evidently, the specific interpretation of SLR test results in practice benefits from additional graphical tools. In this empirical application, moving average estimates of the standard deviation were plotted, which uncovered the dynamics of each series. Conversely, the evaluation of a graphical output benefits from the simple, scalar values of the SLR test. This can be seen in this study, where individual parameters are analyzed separately, but the advantage is even more relevant when the transition of more than one parameter must be judged at once.

5.5 Conclusion

The chapter presents a test for the investigation of parameter change abruptness, when a structural break has been detected in a regression model. It is a follow-up method that assumes that the structural break has been detected by the popular maximum likelihood

ratio test. The test statistic involves the summation of the likelihood ratio values typically available from maximum likelihood ratio testing. With this summation statistic, the abruptness of the detected parameter shift is tested against other forms of non-abrupt parameter transition. The test can be easily applied to support the interpretation of break detection outcomes.

The concept of the test is described and illustrated, before its properties are analyzed in a Monte Carlo simulation study. In the simulations, regression parameters gradually change their values according to a smooth transition specification. The results confirm the validity of the test in two ways. First, they show that the maximum likelihood ratio statistic is sensitive to parameter changes that are not abrupt. This highlights the importance of having to consider types of structural change that are different from the abrupt transition case, once a break detection method infers a structural break. Second, the study confirms that the sum of likelihood ratio statistic can uncover the non-abruptness of parameter transitions in small samples. Specifically, the test becomes more powerful when parameter transitions take longer. The test shares some properties with related methods of structural break testing and statistical methods in general: Its ability to reject the hypothesis of abruptness increases with the number of observations and with the reduction of noise, relative to the strength of the signal.

The sum of likelihood ratio test is applied to several data sets of European government bond yields that each contain multiple significant changes in the error term of a linear regression model. The test is applied one structural break at a time. The unexplained variation of the series is found to transition non-abruptly in many cases, at conventional confidence levels. The increases in volatility during 2007 and the first months of 2008, at the dawn of the Global Financial Crisis, are generally not abrupt according to the test. In such a case, the original structural break test indicates that the model is improved by allowing a parameter shift compared to no change, but the sum of likelihood ratio test shows that a non-abrupt transition model is significantly more appropriate.

The proposed test has a single parameter that must be set by its applicant, the summation range v . There is no clear support for any specific value, but the use of small summation ranges is discouraged. In data samples with few observations, relatively small parameter shifts, and a structural break date that lies close to the center of the sample, the summation of a large range of available likelihood ratio data benefits the test. In any other case, more moderate values of v lead to higher testing power. Generally, the results indicate a positive correlation between the most favorable test value of v and the parameter transition duration itself. Since the duration is unknown beforehand, this observation serves no practical purpose at this point. When using the test in practice, several settings of the summation range could be tried. In future research, the positive correlation could be used as the basis of a test that investigates the speed of parameter transition.

The described test properties are valid for the given specifications of the Monte Carlo study, with simple regression samples and smooth parameter transition. The test is shown to augment simple structural break tests as well as graphical tools of parameter change with a scalar statistic. A study that analyzes the implications of allowing a wider range of data types could be a potential extension. Another possibility would be to study Wald test results and their summation, for example, instead of likelihood ratio results.

The study of asymptotic properties of the test would provide another angle, and may result in the description of a limit distribution of the test.

5.A Appendix

Table 5.7: Full SLR test results 1/3

i	m	1	2	3	4	5	6
BD	\hat{k}_m	2004-06-01	2007-11-27	2009-08-19	2011-05-05	2012-11-07	2015-04-28
	$v = 0.1$	0.102	0.027	0.920	1	0.140	1
	0.2	0.202	< 0.001	0.508	1	0.351	1
	0.3	0.374	< 0.001	0.681	1	0.795	1
	0.4	0.002	< 0.001	0.961	1	0.644	1
	0.5	< 0.001	< 0.001	0.998	1	0.365	1
	0.6	< 0.001	< 0.001	1	1	0.071	1
	0.7	< 0.001	< 0.001	1	1	0.128	1
BG	\hat{k}_m	2004-06-01	2007-11-27	2009-06-23	2010-11-15	2012-07-31	2015-01-13
	$v = 0.1$	0.046	< 0.001	0.709	0.047	0.999	0.924
	0.2	0.038	< 0.001	0.289	0.160	1	0.963
	0.3	0.170	< 0.001	0.432	0.955	1	1
	0.4	0.002	< 0.001	0.832	0.751	1	1
	0.5	< 0.001	< 0.001	0.986	0.049	1	1
	0.6	< 0.001	< 0.001	0.999	< 0.001	1	1
	0.7	0.001	< 0.001	1	< 0.001	1	1
CZ	\hat{k}_m	2006-03-13	2007-11-14	2010-01-13	2011-06-13	2013-07-24	2015-07-30
	$v = 0.1$	0.793	1	0.980	0.813	1	1
	0.2	0.080	1	0.577	0.804	0.806	1
	0.3	0.090	0.950	0.405	0.997	0.109	1
	0.4	0.026	0.769	0.051	0.999	0.002	1
	0.5	0.004	0.101	0.008	1	0.063	1
	0.6	< 0.001	< 0.001	0.002	1	0.013	1
	0.7	< 0.001	< 0.001	< 0.001	1	< 0.001	1
DK	\hat{k}_m	2004-06-01	2005-12-14	2007-12-06	2010-08-19	2013-01-11	2014-12-26
	$v = 0.1$	0.012	0.569	0.039	0.991	0.996	1
	0.2	< 0.001	0.648	< 0.001	0.690	0.889	1
	0.3	< 0.001	0.197	< 0.001	0.035	0.720	1
	0.4	< 0.001	0.020	< 0.001	< 0.001	0.633	1
	0.5	< 0.001	0.009	< 0.001	< 0.001	0.647	1
	0.6	< 0.001	0.004	< 0.001	< 0.001	0.270	1
	0.7	< 0.001	0.001	< 0.001	< 0.001	0.018	1
ES	\hat{k}_m	2004-07-02	2007-11-27	2009-06-10	2011-06-01	2013-07-04	2014-11-28
	$v = 0.1$	0.006	< 0.001	1	0.619	0.954	0.998
	0.2	< 0.001	< 0.001	1	0.942	0.319	0.939
	0.3	< 0.001	< 0.001	1	< 0.001	< 0.001	0.510
	0.4	< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001
	0.5	< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001
	0.6	< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001
	0.7	< 0.001	< 0.001	1	< 0.001	< 0.001	< 0.001
FN	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-02-15	2012-10-29	2015-01-13
	$v = 0.1$	0.013	0.026	0.727	0.975	1	0.964
	0.2	0.016	0.003	0.165	0.999	1	0.912
	0.3	0.042	0.001	0.266	0.995	1	0.994
	0.4	< 0.001	< 0.001	0.611	0.992	1	0.999
	0.5	< 0.001	< 0.001	0.885	0.715	1	1
	0.6	< 0.001	< 0.001	0.987	0.028	1	1
	0.7	< 0.001	< 0.001	0.996	< 0.001	1	1

Note: For a first set of series $i = BD, \dots, FN$, results of testing the abruptness of six shifts in σ_ε that have been detected by a multiple structural break test. The table reports p -values for the null hypothesis that the corresponding structural break m is abrupt, according to the SLR test, using several values of the summation range parameter v .

CHAPTER 5. TESTING PARAMETER CHANGE ABRUPTNESS WITH A SUM OF LIKELIHOOD RATIOS

Table 5.8: Full SLR test results 2/3

i	m	1	2	3	4	5	6
<i>FR</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-14	2012-08-03	2015-04-28
	$v = 0.1$	0.064	0.005	0.073	0.373	0.924	1
	0.2	0.052	< 0.001	< 0.001	0.846	1	1
	0.3	0.202	< 0.001	< 0.001	1	1	1
	0.4	0.002	< 0.001	< 0.001	1	1	1
	0.5	< 0.001	< 0.001	0.025	1	1	1
	0.6	< 0.001	< 0.001	0.415	1	1	1
	0.7	< 0.001	< 0.001	0.873	0.832	1	1
<i>GR</i>	\hat{k}_m	2004-06-01	2007-08-07	2009-01-01	2011-04-13	2013-01-31	2014-10-10
	$v = 0.1$	0.701	< 0.001	< 0.001	1	0.001	< 0.001
	0.2	0.482	< 0.001	0.001	0.989	< 0.001	< 0.001
	0.3	0.134	< 0.001	0.358	0.981	< 0.001	< 0.001
	0.4	0.031	< 0.001	0.001	0.999	< 0.001	< 0.001
	0.5	0.046	< 0.001	< 0.001	1	< 0.001	< 0.001
	0.6	0.085	< 0.001	< 0.001	1	< 0.001	< 0.001
	0.7	0.145	< 0.001	< 0.001	1	< 0.001	< 0.001
<i>IR</i>	\hat{k}_m	2004-06-01	2005-12-12	2007-11-27	2010-07-12	2012-08-01	2014-02-10
	$v = 0.1$	0.047	0.002	0.001	0.140	1	1
	0.2	0.064	0.015	< 0.001	< 0.001	0.999	0.999
	0.3	0.278	0.015	< 0.001	< 0.001	0.026	1
	0.4	0.100	0.011	< 0.001	< 0.001	< 0.001	1
	0.5	0.157	0.034	< 0.001	< 0.001	< 0.001	1
	0.6	0.389	0.034	< 0.001	< 0.001	< 0.001	1
	0.7	0.536	0.049	< 0.001	< 0.001	< 0.001	1
<i>IT</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-15	2011-05-26	2013-03-27	2015-02-25
	$v = 0.1$	0.690	< 0.001	0.997	1	0.351	1
	0.2	0.517	< 0.001	0.999	0.994	< 0.001	0.961
	0.3	0.204	< 0.001	1	0.859	< 0.001	0.120
	0.4	0.002	< 0.001	1	1	< 0.001	< 0.001
	0.5	< 0.001	< 0.001	1	1	< 0.001	< 0.001
	0.6	< 0.001	< 0.001	1	1	< 0.001	0.005
	0.7	< 0.001	< 0.001	1	1	< 0.001	0.010
<i>NL</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-03-31	2012-09-27	2015-01-12
	$v = 0.1$	0.105	0.010	0.612	0.624	0.413	1
	0.2	0.163	0.001	0.382	0.246	0.830	1
	0.3	0.305	< 0.001	0.637	0.169	0.988	1
	0.4	0.001	< 0.001	0.930	0.020	0.999	1
	0.5	< 0.001	< 0.001	0.995	< 0.001	0.995	1
	0.6	< 0.001	< 0.001	1	< 0.001	0.982	1
	0.7	0.001	< 0.001	1	< 0.001	0.962	1
<i>NW</i>	\hat{k}_m	2004-06-01	2005-12-16	2007-07-30	2009-04-06	2010-08-30	2012-01-24
	$v = 0.1$	1	0.950	< 0.001	0.959	0.588	1
	0.2	0.880	0.816	< 0.001	0.995	1	1
	0.3	0.300	0.256	< 0.001	1	1	1
	0.4	< 0.001	0.041	< 0.001	1	1	1
	0.5	< 0.001	< 0.001	< 0.001	1	1	1
	0.6	< 0.001	< 0.001	< 0.001	1	1	1
	0.7	< 0.001	< 0.001	< 0.001	1	1	1

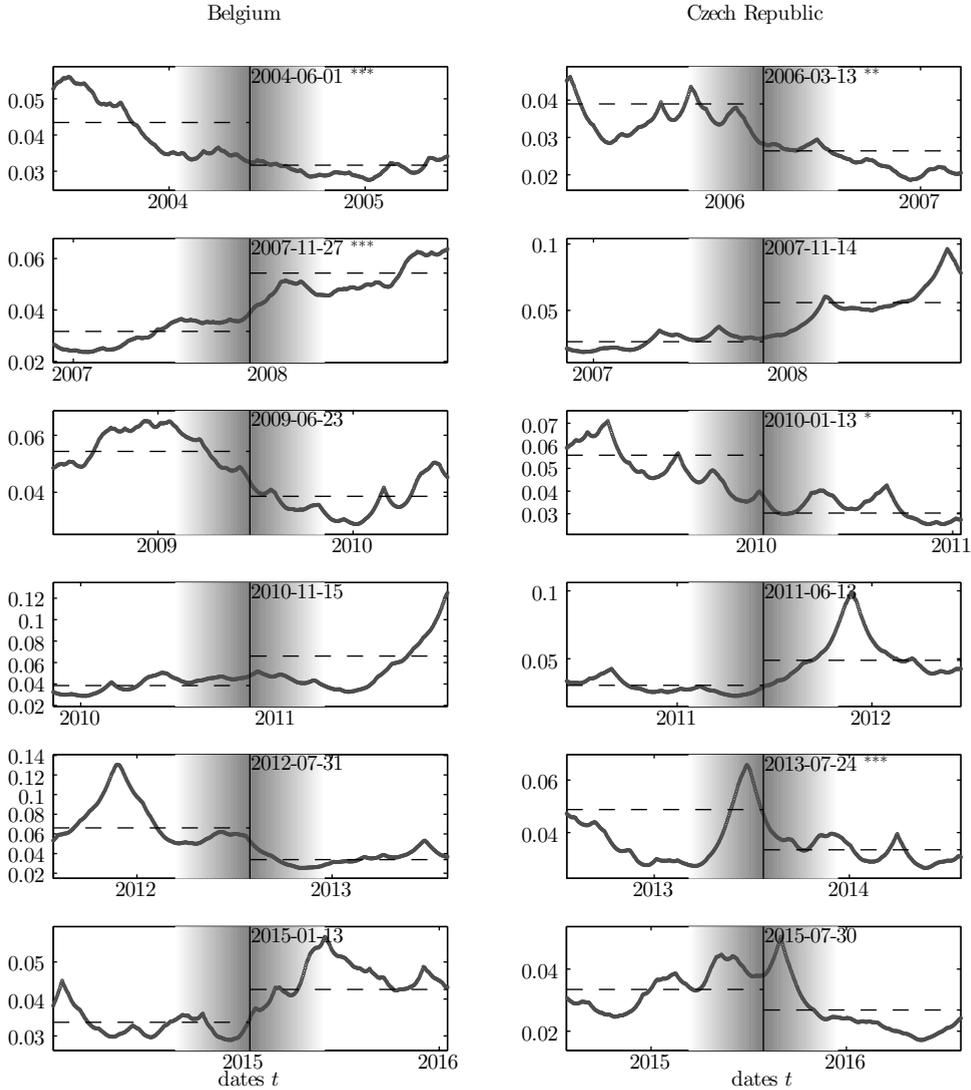
Note: For a second set of series $i = FR, \dots, NW$, results of testing the abruptness of six shifts in σ_ε that have been detected by a multiple structural break test. The table reports p -values for the null hypothesis that the corresponding structural break m is abrupt, according to the SLR test, using several values of the summation range parameter v .

Table 5.9: Full SLR test results 3/3

i	m	1	2	3	4	5	6
<i>OE</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-06-19	2011-04-14	2012-09-14	2015-04-28
	$v = 0.1$	0.198	< 0.001	0.882	0.967	0.995	1
	0.2	0.308	< 0.001	0.618	0.991	1	1
	0.3	0.683	< 0.001	0.600	1	1	1
	0.4	0.068	< 0.001	0.914	1	1	1
	0.5	0.005	< 0.001	0.995	1	1	1
	0.6	0.002	< 0.001	0.999	1	0.999	1
	0.7	0.003	< 0.001	0.999	1	0.999	1
<i>PO</i>	\hat{k}_m	2004-09-17	2006-04-13	2008-02-08	2009-08-06	2012-07-11	2014-10-17
	$v = 0.1$	0.879	0.001	< 0.001	0.066	< 0.001	1
	0.2	0.999	< 0.001	< 0.001	0.003	< 0.001	1
	0.3	1	< 0.001	0.115	< 0.001	< 0.001	1
	0.4	1	< 0.001	0.742	< 0.001	< 0.001	0.997
	0.5	0.999	< 0.001	0.001	< 0.001	< 0.001	0.790
	0.6	0.919	< 0.001	< 0.001	0.001	< 0.001	0.689
	0.7	0.748	< 0.001	< 0.001	0.008	< 0.001	0.711
<i>PT</i>	\hat{k}_m	2004-06-01	2005-12-09	2007-11-27	2010-04-19	2013-07-22	2014-12-26
	$v = 0.1$	0.326	0.002	0.001	0.997	1	1
	0.2	0.399	0.033	< 0.001	0.001	1	1
	0.3	0.253	0.060	< 0.001	< 0.001	1	0.993
	0.4	0.007	0.057	< 0.001	< 0.001	1	0.808
	0.5	0.001	0.062	< 0.001	< 0.001	0.970	0.774
	0.6	0.002	0.045	< 0.001	< 0.001	0.721	0.958
	0.7	0.012	0.050	< 0.001	< 0.001	0.377	0.932
<i>SD</i>	\hat{k}_m	2004-06-01	2007-11-27	2009-12-08	2011-05-17	2013-09-19	2015-02-11
	$v = 0.1$	0.583	0.391	0.011	0.999	0.873	0.670
	0.2	0.741	< 0.001	0.006	1	0.271	0.912
	0.3	0.706	< 0.001	0.016	1	0.006	0.727
	0.4	0.402	< 0.001	0.019	1	< 0.001	0.795
	0.5	0.046	< 0.001	0.080	1	< 0.001	0.905
	0.6	0.005	< 0.001	0.046	1	< 0.001	0.988
	0.7	< 0.001	< 0.001	0.006	1	< 0.001	0.996
<i>SW</i>	\hat{k}_m	2004-06-02	2007-10-31	2009-06-25	2011-12-21	2013-07-24	2014-12-16
	$v = 0.1$	0.998	0.275	< 0.001	0.903	0.726	1
	0.2	1	0.004	0.166	0.694	0.995	1
	0.3	0.965	< 0.001	0.797	0.701	1	1
	0.4	0.838	< 0.001	0.986	0.723	0.979	1
	0.5	0.888	< 0.001	1	0.487	0.915	1
	0.6	0.901	< 0.001	1	0.205	0.360	1
	0.7	0.661	< 0.001	1	0.117	0.006	1
<i>UK</i>	\hat{k}_m	2004-06-16	2007-11-13	2010-01-21	2011-06-20	2012-11-09	2015-02-16
	$v = 0.1$	0.438	0.033	0.004	0.994	0.229	0.366
	0.2	0.101	< 0.001	< 0.001	1	0.589	0.004
	0.3	0.079	< 0.001	< 0.001	1	0.276	< 0.001
	0.4	0.027	< 0.001	< 0.001	1	0.026	< 0.001
	0.5	< 0.001	< 0.001	< 0.001	1	0.003	< 0.001
	0.6	< 0.001	< 0.001	< 0.001	1	< 0.001	< 0.001
	0.7	< 0.001	< 0.001	< 0.001	1	< 0.001	< 0.001

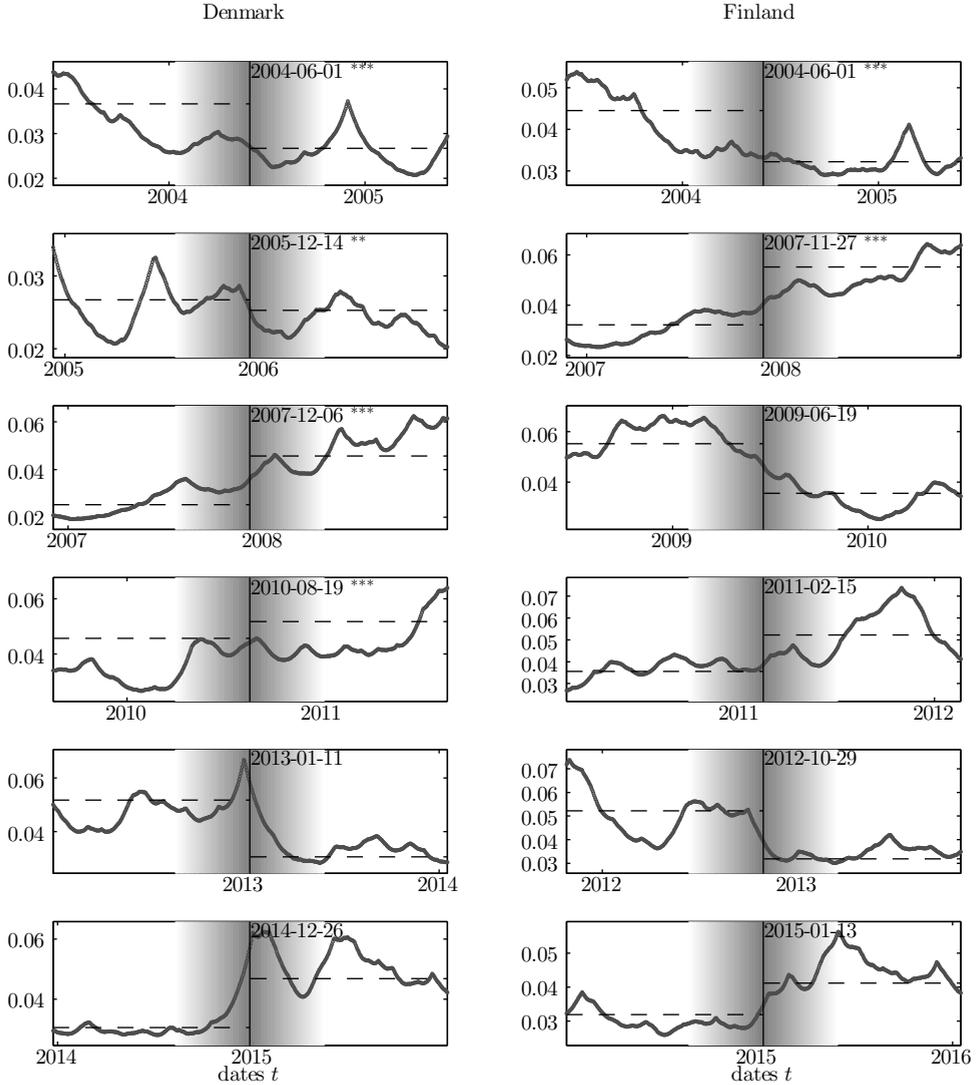
Note: For a third set of series $i = OE, \dots, UK$, results of testing the abruptness of six shifts in σ_ε that have been detected by a multiple structural break test. The table reports p -values for the null hypothesis that the corresponding structural break m is abrupt, according to the SLR test, using several values of the summation range parameter v .

Figure 5.4: Error term standard deviation over time 2/9



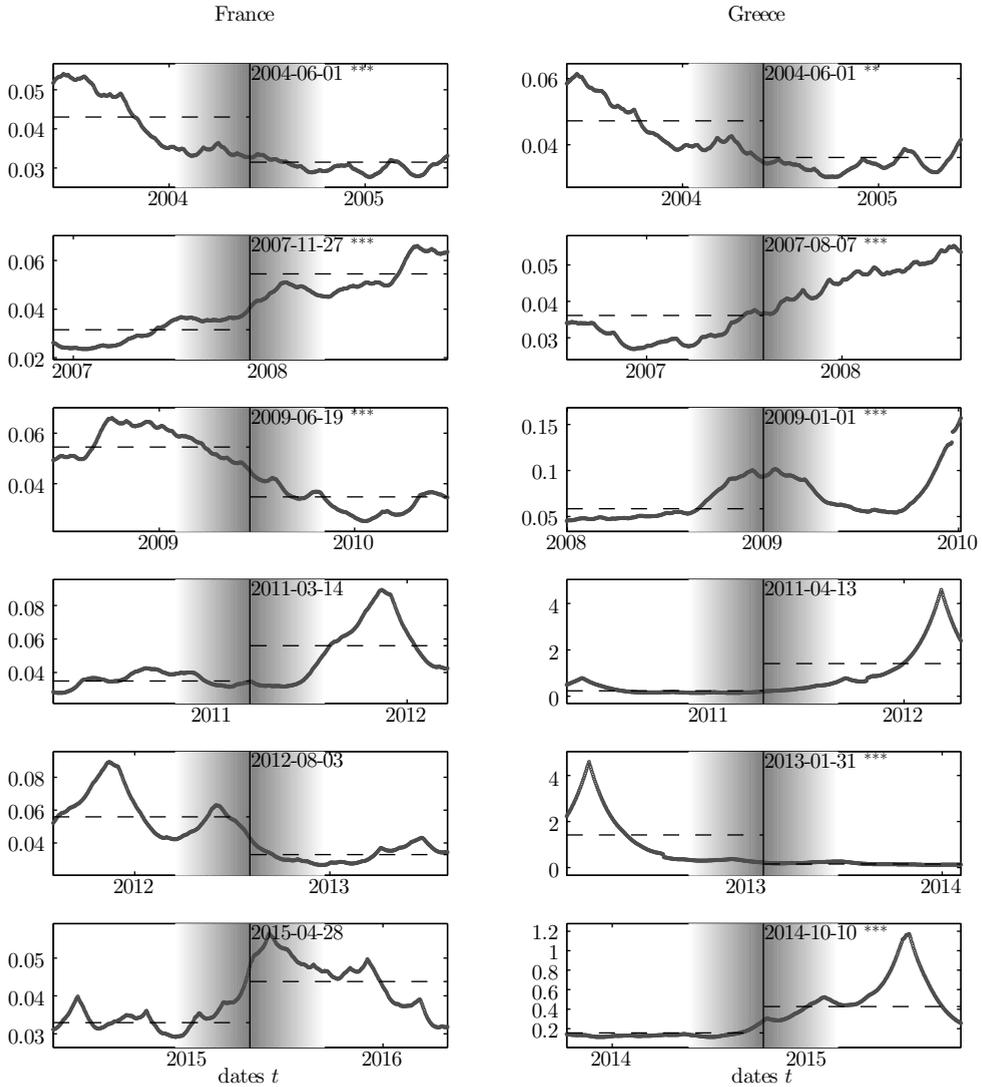
Note: For $i = BG, CZ$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.5: Error term standard deviation over time 3/9



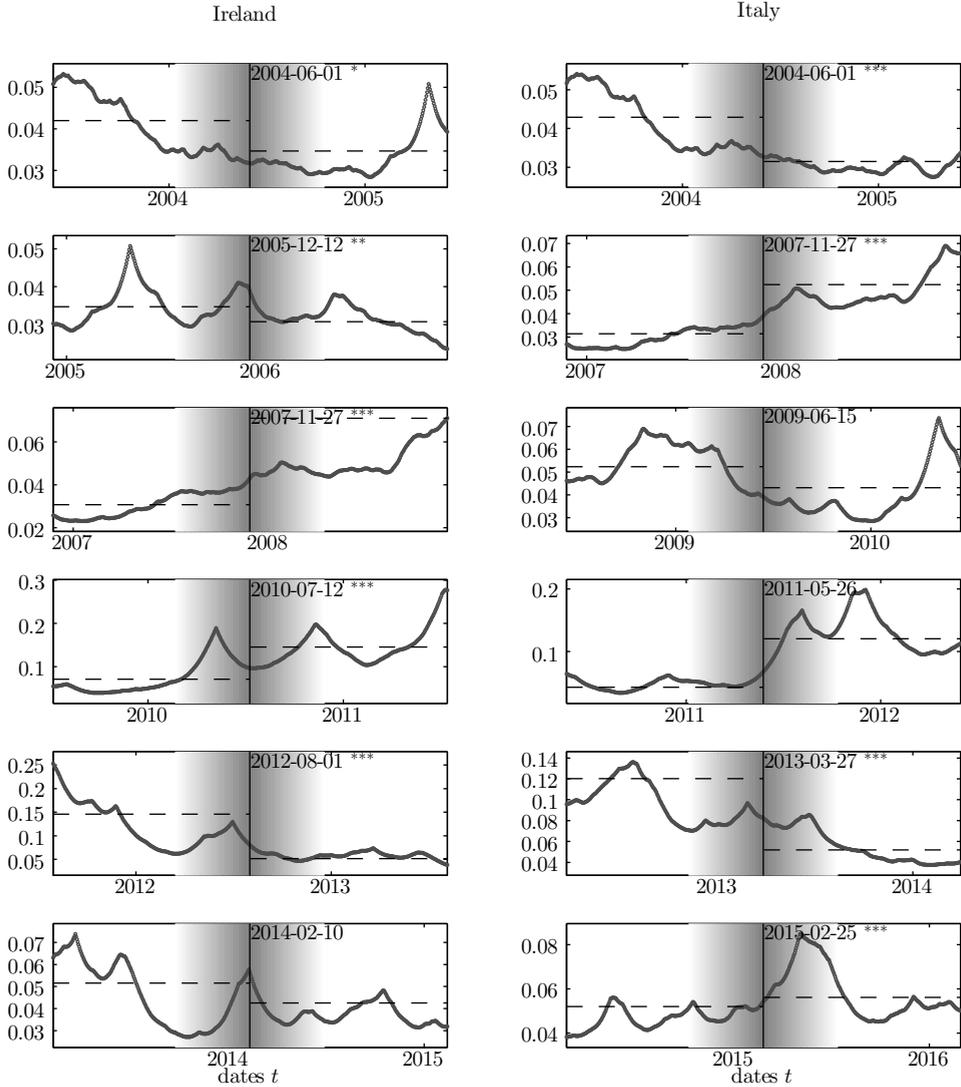
Note: For $i = DK, FN$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.6: Error term standard deviation over time 4/9



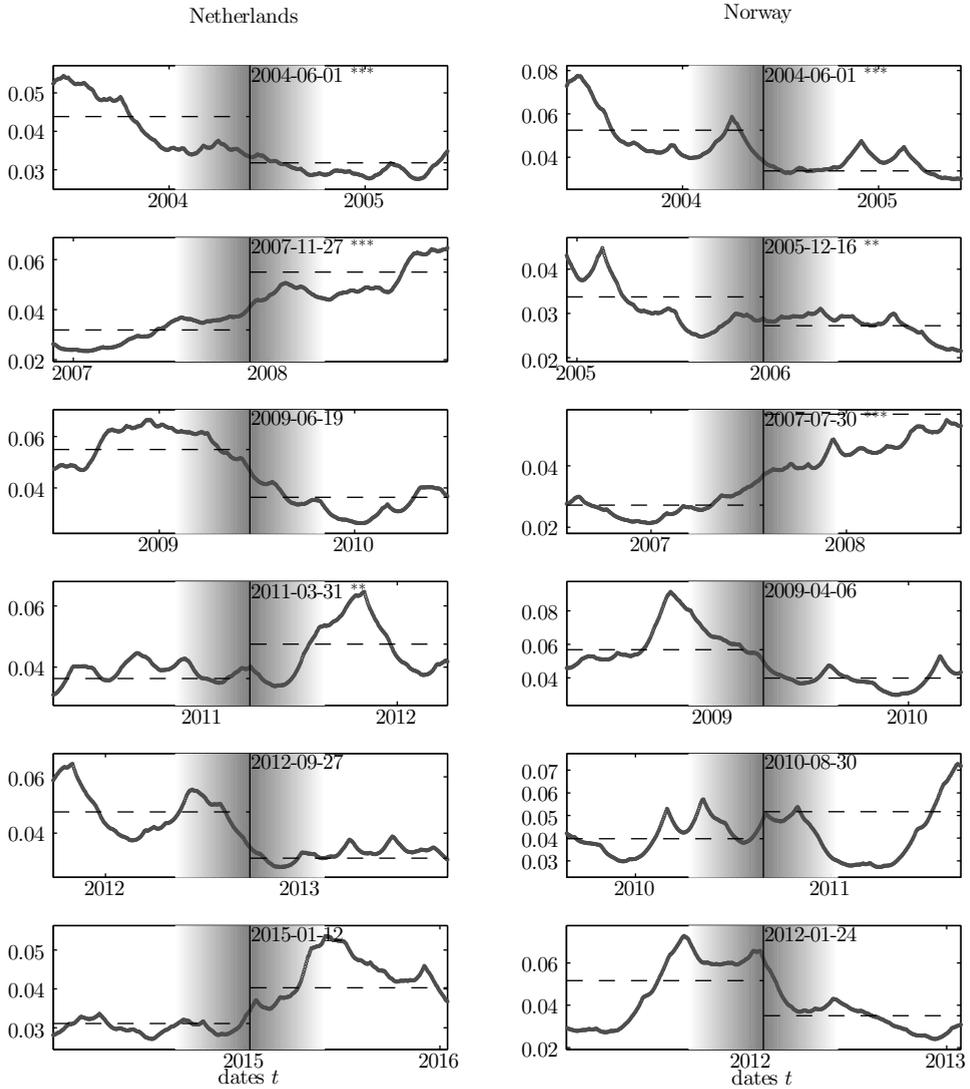
Note: For $i = FR, GR$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.7: Error term standard deviation over time 5/9



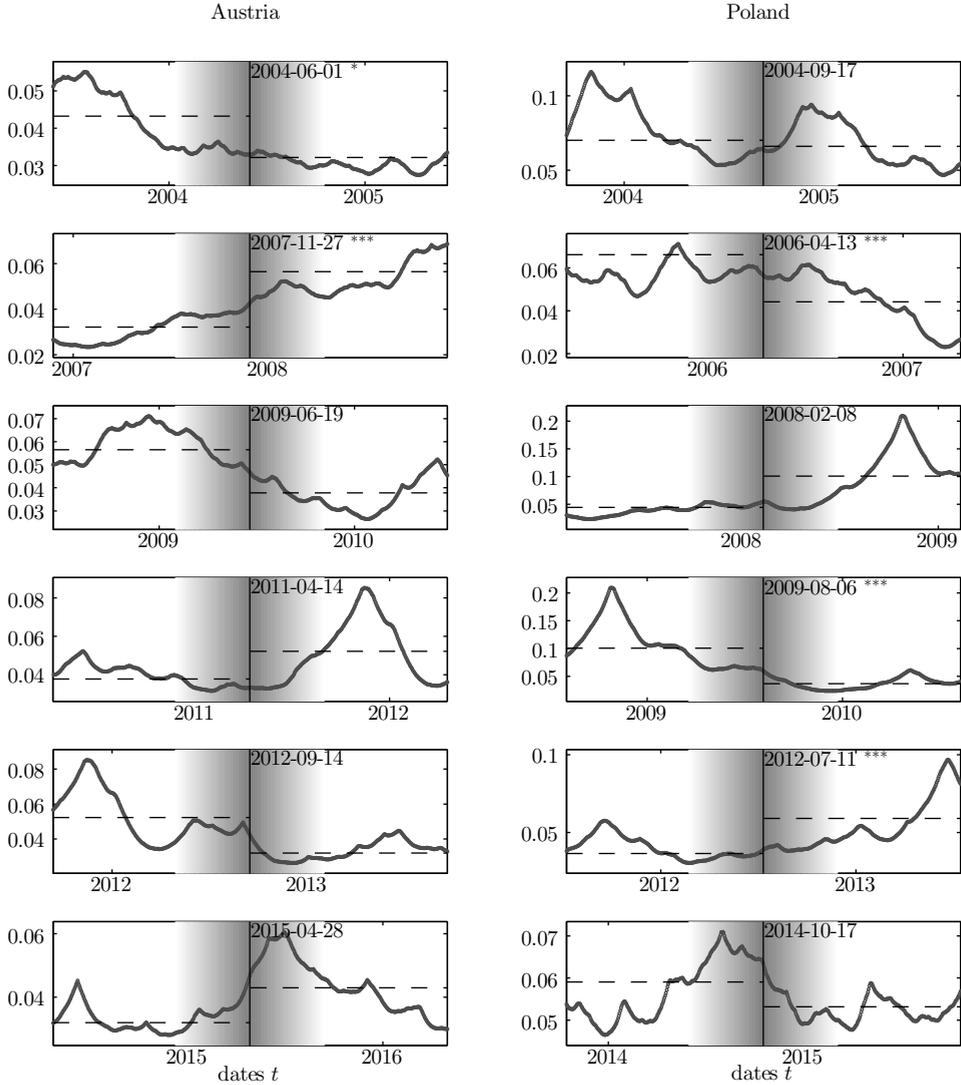
Note: For $i = IR, IT$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.8: Error term standard deviation over time 6/9



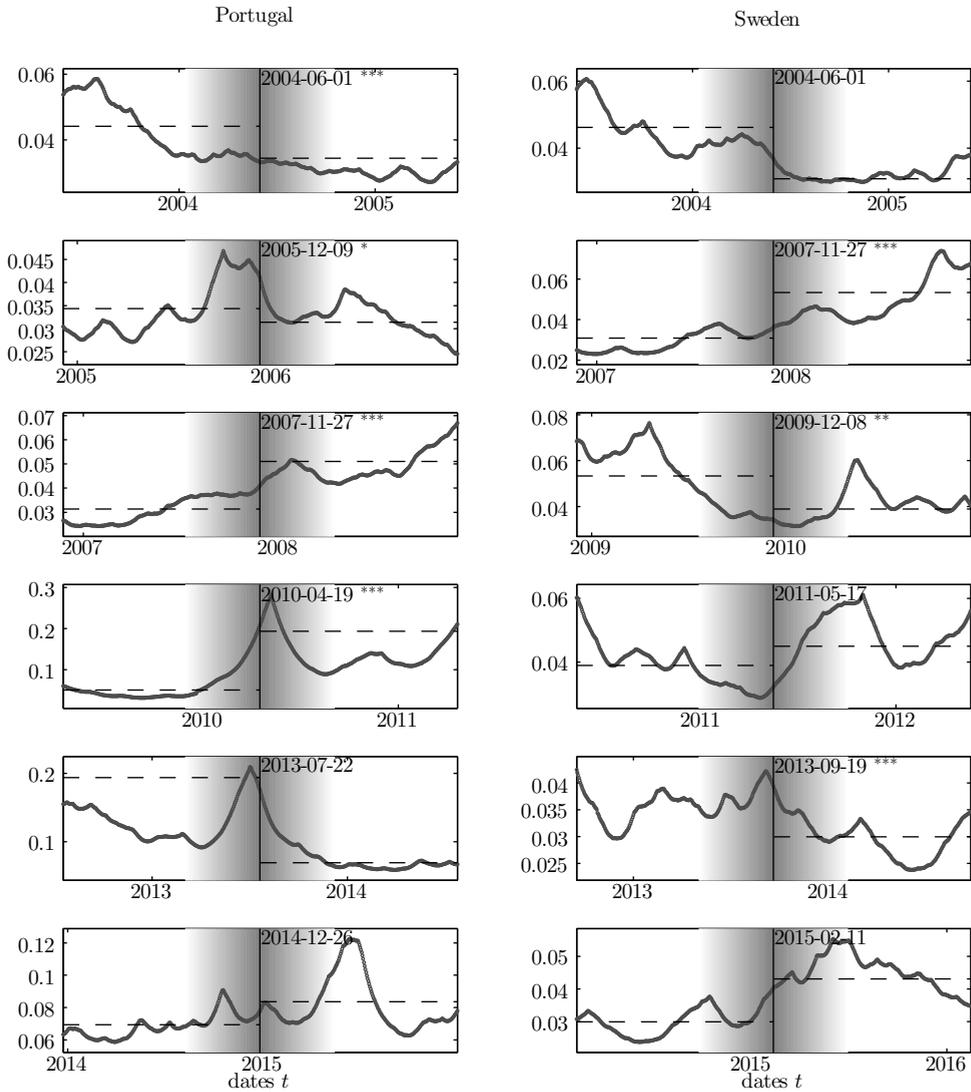
Note: For $i = NL, NW$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.9: Error term standard deviation over time 7/9



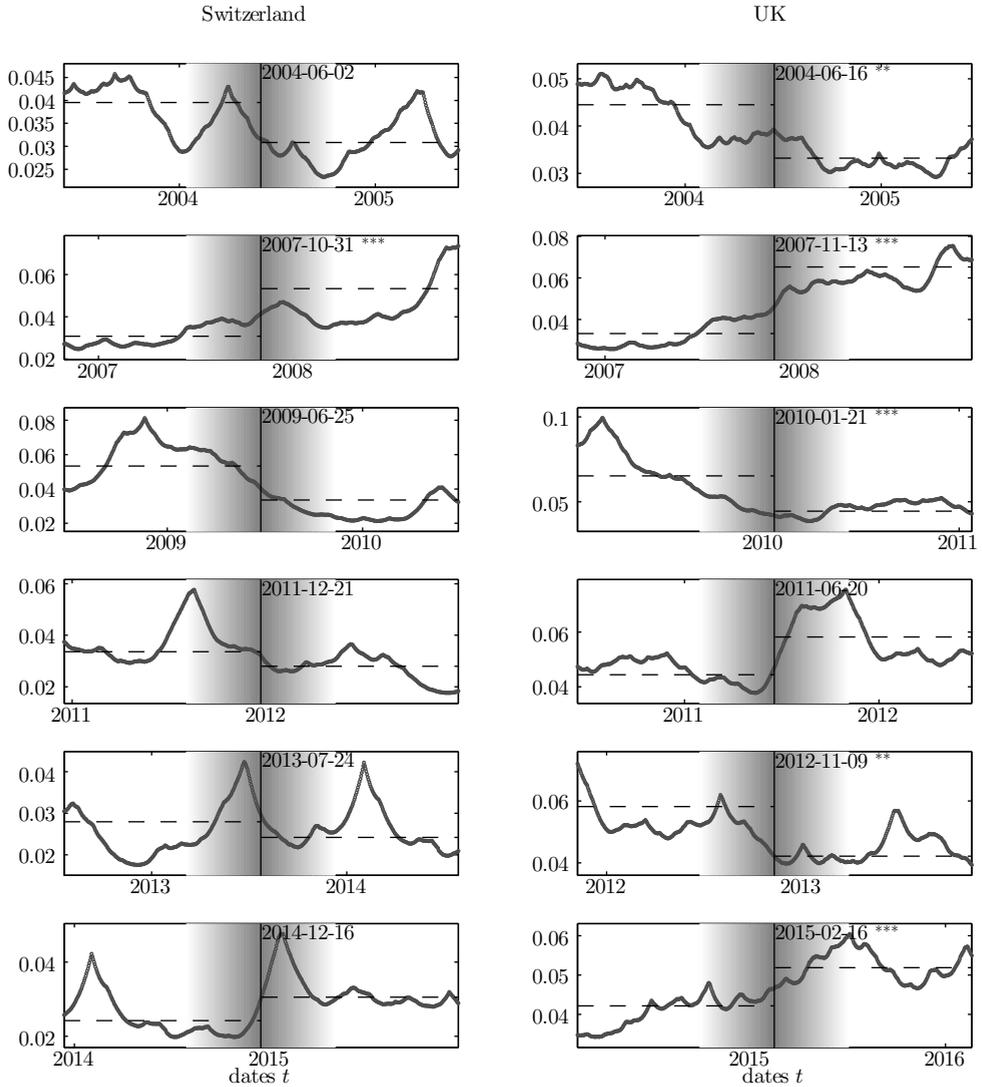
Note: For $i = OE, PO$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.10: Error term standard deviation over time 8/9



Note: For $i = PT, SD$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

Figure 5.11: Error term standard deviation over time 9/9



Note: For $i = SW, UK$, and for breaks $m = 1, \dots, 6$, the panels show estimated values of σ_ε over time, within the respective data segment of SLR testing. In each panel, moving average estimates of σ_ε derived from rolling windows of 199 data points are shown by a scatter plot, using residuals of the regression of Δr_i . A vertical black line marks the estimated break date \hat{k}_m according to a multiple break testing in the full data set. *, **, *** are used to mark the breaks that are not abrupt according to SLR testing at $v = 0.4$ at 90%, 95%, 99% confidence, respectively. Horizontal, dashed lines indicate estimated values $\hat{\sigma}_m$ according to the multiple break testing method. The shaded area around each break date \hat{k}_m estimate marks all moving average estimates that have been calculated using a data window that includes \hat{k}_m , with lighter shading indicating a larger discounting.

6

Detecting structural breaks by cross-validated error estimation

6.1 Introduction

This chapter introduces a novel way of detecting structural breaks in time series data. Structural break tests commonly incorporate a measure of model improvement as a test statistic, a design that is also adopted in the proposed break detection method. The novelty of the test comes from the use cross-validated error estimates in the test statistic. When estimating the model of the time series, each observation of the sample is removed exactly one time. The removed observations are used to evaluate how well the current model predicts out-of-sample data. Model estimation is repeated a number of times C , and each time a different validation set of observations is removed from the data that is used in the estimation. Due to this C -fold cross-validation, the detection of a structural break is based on the optimization of the out-of-sample (prediction) fit of a regression model.

The proposed structural break test shares the conceptual design of other structural break tests. It compares two nested versions of a model, in which one version of the model is less restricted than the other. Specifically, in the less restricted version the parameters of the model can change their value at some point in the data. If this leads to a significant improvement, as measured by the test statistic, a structural break is found. In this case, the test result indicates that allowing the parameters to change their value leads to a significantly better structural modeling of the time series. Differently stated, a model that contains a structural break and one that does not are fitted to the available data and the impact is measure by some test statistic. If this impact is larger than what is typically observed when a structural break is known to be absent, a structural break is inferred.

The proposed test uses cross-validated estimates of prediction errors to validate the separation of the data at a break date. In contrast to other tests, such as the popular maximum likelihood ratio (MLR) test of Andrews (1993), this method acknowledges the role of the sample and its limited representation of the data generating process. Small prediction errors indicate that a model represents current data as well as not (yet) sampled observations. That way, structural break detection is not based on optimally fitting a model to the available data, but on estimating an optimal model fit to the data generating process. To evaluate the performance of the resulting test, small sample time series have been simulated in a Monte Carlo study. In the study, the test is compared to the MLR test of Andrews (1993), due to its popularity in structural break testing. The proposed test is shown to have similar testing power, but often finds break dates that lie closer to the true break date than the MLR test.

Like other break tests, the proposed test is one that detects breaks; it finds breaks at previously unknown dates in the time series. No prior hypothesis for the position of the break is required, instead the method finds the date that is associated with the structural break endogenously. The most likely breaks are located at those dates, where the hypothetical structural break leads to the largest degree of improvement in the model. This selection process leads to a difficulty in the determination of critical values that is shared by comparable break detection methods: The maximum of the break test statistic has a larger expectation than the break test statistic itself. Even in the absence of a structural break in the data generating process, the most likely break date candidate can seem to indicate model improvements. So by implication, a break is only significant, if the empirically measured improvement of data fit surpasses such values typically attained when a data generating process is known to not contain a structural break. The chapter includes a description of how to derive critical values from the available data by the application of resampling bootstraps.¹

It is widely accepted that recognizing structural breaks is important for reducing prediction errors in time series forecasting, see for example Stock and Watson (1996). More recently Clark and McCracken (2009) describe potential forecasting improvements of combining forecasts derived from the most recent model after a structural break and a model that covers a larger amount of data. Here as well, reliably identifying the structural patterns of a data generating process, including breaks, will lower expected errors in forecasting. In Jiang et al. (2013) structural breaks are not only used to update the appropriate model for the most recent observations. Also, the frequency of structural breaks is used as an indicator of the probability of upcoming structural breaks to reduce out-of-sample prediction errors in expectation.

Conversely, the way that prediction error estimation can enhance model selection is increasingly being studied within the field of statistical learning. Estimated prediction errors are used to choose between two competing models in Van de Wiel et al. (2009),

¹Data fit measures of popular testing methods determine to which degree the parameter change reduces the unexplained variation in the model. An increase in the degrees of freedom in nested models cannot decrease the data fit. In other words, a structural break cannot worsen the ability of the model to represent the recorded data. In contrast, the prediction error is not necessarily reduced by a structural break, because of the variance-bias-tradeoff. The structural break increases the degrees of freedom for the same amount of observations, which increases the probability of over-fitting, and an overfit model has a large prediction error. Nevertheless, in either method the selection of a maximum has the described implication.

but the idea is not applied to time series data and structural break detection testing. Goodfellow et al. (2016) explain how cross-validated errors are used to derive confidence intervals for mean prediction errors, so that a machine learning algorithm can be declared better than its competing algorithms. Pesaran and Timmermann (2007) describe how cross-validation can be used to find confidence intervals for structural breaks. Then, forecasts can become a weighted average of several forecasts that each correspond to a break date within the confidence interval. Yet, the applied structural break test statistic does not itself consist of prediction error estimates.

The test developed in Boot and Pick (2017), which is based on forecast accuracy comparison as described in Clark and McCracken (2001), also uses prediction errors to detect a structural break. Specifically, h -step ahead forecasts are calculated, so that the accumulated forecast error between a model with structural break and one without can be compared. The cross-validated error test of this chapter is different, as the predictions of each validation set are derived from model parameters that are estimated from the full set of remaining data. Further, the test statistic Boot and Pick (2017) only compares no-break predictions with post-break predictions, not with pre-break predictions. Its limited use of the available data may explain the lower testing power reported by the authors, especially in practical application.

Other publications also use prediction errors to test for parameter stability (or model adequacy, in general), but the tests do not share the form and aim of the proposed test. Chinn et al. (2014) apply a structural break test that is based on forecast errors. Here structural breaks are recursively inferred when observations lie beyond the confidence bounds of their own forecasts. For this, the authors exogenously select the last observation to be included in the model estimation, and evaluate the model's ability to forecast later observations. The authors apply a concept that is reminiscent of the recursive residual tests originally described in Brown et al. (1975), which are based on cumulative sums of (squared) forecast errors. A cumulative sum test can uncover model instability, when data is added to the end of a sample and the sum of estimated (squared) forecast errors deviates sufficiently from the expected value. The test uses a series of adjacent observations, it will depend on the first date that is chosen to be forecasted, and it can be triggered by individual outlier observations. Unlike the cumulative sum test, the proposed test of this chapter will estimate and use prediction errors of all available observations to detect parameter instability. Structural breaks are only found when whole segments of the data are structurally different from another, not when consecutive observations are unusual. As an additional feature, the proposed test provides an estimate of the location of the structural break in the sample endogenously.

In the proposed test, C -fold cross-validation is used to estimate the mean squared error of competing models, whereby the model with a structural break has more degrees of freedom than the model without break. The minimum of the mean squared error identifies the degrees of freedom that reduce the expected prediction error as far as possible. Minimizing the estimated mean squared errors is therefore a common strategy in model selection, see for example Hastie et al. (2009) or Goodfellow et al. (2016). As a non-parametric method of cross-validated estimation errors with bootstrapped critical values, the assumptions are very mild. Cross-validation requires that data is identically distributed apart from the structural break that is tested, see Arlot et al. (2010) for an overview. As the test is presented in this chapter, disturbances of the model must

not be serially dependent. The assumption could be relaxed by using modified cross-validation methods, or otherwise iterative forecasting such as in Boot and Pick (2017). By construction, the proposed test looks for breaks in all parameters of the model, including the error term. Partial structural break hypotheses that concern a subset of covariates, only, can be tested by formulating a separate model before testing. This can be done by orthogonally projecting the subset of covariates off the remaining covariates and applying the test to the residual observations.

An empirical study on the redemption yield of several European central government bonds with a maturity of 10 years demonstrates the procedure in a practical setting. The study extends the results from earlier Chapters 4 and 5 of this book, as the cross-validated error test is applied to a separated model that only includes one predictor variable, lagged U.S. treasury bond yields. For six central euro area countries and Denmark a structural break is detected in April 2015, after which the estimated linear impact of lagged U.S. yield differences on the expected respective European yield difference is increased. The introduction of a quantitative easing program by the ECB is interpreted to have increased the perceived similarity between some European and American bonds. Until then, the ECB had conducted mostly credit easing and had sterilized its unconventional monetary policy to maintain constant monetary base growth, see Borio and Disyatat (2010) for a comparison between euro area, U.S. and the UK. Meanwhile, quantitative easing has been attributed a stabilizing role in the U.S. economic downturn following the Global Financial Crisis, see for example Kapetanios et al. (2012). Claeys et al. (2015) argue that an earlier adoption of quantitative easing could have lowered the debt burden of euro area periphery countries significantly. Faced with the danger of deflation, in 2015 the ECB deviated from its monetary target and initiated the public-sector purchase programme. This outright asset purchase program more closely resembles the quantitative easing policy of the U.S. The hypothesis of the empirical study is that the explanatory value of U.S. government bond prices for European bond prices has increased since 2014, due to the alignment of euro area and U.S. monetary policies and the role of government bonds in the policies.

The relation between American and European financial market developments is explored by several contemporary researchers. Ehrmann and Fratzscher (2005) analyze the impact of macroeconomic announcements on interbank rates in the euro area and the U.S. Neely and Rapach (2011) study the impact of unconventional monetary policy announcements in the U.S. on international yields. Studying the impact of exogenously chosen events on bond yields is a common approach, see also Krishnamurthy et al. (2014). In Eser and Schwaab (2016) unconventional monetary policy measures of the ECB are studied in event studies around program announcement dates, similar to the studies of Gagnon et al. (2011) and Von Borstel et al. (2016) for the U.S. and of Joyce et al. (2011) for the UK. Theoretical explanations are offered for the way unconventional monetary policy can lower lending rates, as the policy directly increases bond demand, increases market liquidity and signals long-term commitment to expansionary policy. The study of this chapter does not explore the channels by which unconventional monetary policy helps lower lending rates. Instead, it is concerned with the extent that a predetermined financial series can explain European developments and whether this changes. As argued by Altavilla and Giannone (2017), event studies are powerful for quantifying the immediate effects of policy, but not for assessing the persistence of policy. Structural

break tests can detect significant differences that govern long periods of the data. With a similar method, monetary policy is found to have long-lasting effects on forward real interest rates in Hanson and Stein (2015). In this chapter, however, change is detected endogenously as a significant break that divides the data set.

Section 6.2 will carefully introduce the test and its components. A detailed description of the Monte Carlo study results can be found in Section 6.3. The application of the cross-validated error estimation test is illustrated by the empirical study in Section 6.4. Finally, the chapter concludes in Section 6.5.

6.2 Methodology

6.2.1 The model

Consider a time series with T dates $t = 1, \dots, T$. The dependent time series y_t is univariate. Its data generating process is described by a linear regression model. A number $q \geq 0$ of predictor variables and the intercept term are jointly given by $x_t = [1, x_{1,t}, \dots, x_{q,t}]'$. That is,

$$\begin{aligned} y_t &= x_t' \beta_{pre} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_{pre}^2) && \text{for } t = 1, \dots, k_0, \\ y_t &= x_t' \beta_{post} + \varepsilon_t, \quad \varepsilon_t \sim WN(0, \sigma_{post}^2) && \text{for } t = k_0 + 1, \dots, T, \end{aligned}$$

Date k_0 divides the DGP in two segments. Parameter vector β_{pre} contains $q + 1$ regression coefficients, which measure the linear impact of the vector of covariates x_t on the dependent observation y_t in the first segment until date k_0 . Parameter vector β_{post} does the same for the second segment starting at date $k_0 + 1$.

The proposed test is based on non-parametric estimation, only, and the distribution of the test statistic under the null hypothesis is empirically derived. That means that the model must satisfy few assumptions. The first assumption is that the data must be identically distributed within segments. The given linear regression is just one possible form, which is presented here as it is studied in in the Monte Carlo Section 6.3 of this chapter. The second assumption is that the error term is a white noise process. Serial dependence of the error term would prevent the use of cross-validation in its standard form, see Opsomer et al. (2001), for example. According to Chu and Marron (1991), modified cross-validation methods can be used in case of serial error dependence.

The relevant test hypotheses compare the situation where the set of parameters $\{\beta, \sigma\}$ has the same value in both segments to the situation where at least one of the parameters shifts its value. That is,

$$\begin{aligned} H_0 &: \{\beta_{pre}, \sigma_{pre}\} = \{\beta_{post}, \sigma_{post}\}, \\ H_A &: \{\beta_{pre}, \sigma_{pre}\} \neq \{\beta_{post}, \sigma_{post}\}, \end{aligned}$$

A structural break test must involve a test statistic that evaluates the resulting model alternatives. It will consider each potential break date k in the sample and measure the impact of allowing different parameter values before and after that date. In the end, it must estimate the most likely break date \hat{k} and provide a p -value that gives the

probability of observing structural break indication at least as strong as the measured indication, assuming that no structural break exists.

Consider a minimum required data segment length $\omega = [\kappa T]$, where $[\cdot]$ gives the closest integer value, and where $\kappa \in (0; 0.5]$. The minimum segment length restricts the break date from being very close to the beginning or end of the sample, to ensure a well-behaved test statistic. This means that break dates can only be correctly identified, if $k_0 \in [\omega, T - \omega]$ holds. A common choice in MLR testing is $\kappa = 0.15$, see Andrews (1993), Bai and Perron (2003), which is adapted in this article.

6.2.2 Cross-validated error estimation

The cross-validates error estimation test (CVE) makes use of C -fold cross-validation to estimate mean squared errors. Consider that for each date t , a vector of observations x_t is invariantly linked to the observation of the dependent variable y_t , jointly denoted as the vector of linked observations $[y_t, x_t']'$. All T vectors of linked observations $[y_t, x_t']'$ are randomly assigned to one of the C validation sets, without replacement. Let \mathcal{J}_c be the set of dates t , whose associated vectors of linked observations $[y_t, x_t']'$ have been assigned to the c -th validation set. The size of set \mathcal{J}_c is T_c , the number of dates in the set. Then $T_c \in [1, T - C + 1]$, as each validation set contains at least 1 vector of linked observations and at most $T - C + 1$ vectors.

For each $c = 1, \dots, C$, one estimates the model parameters $\hat{\beta}_c$ by using all observations except for those that have been assigned to the validation set \mathcal{J}_c . That way the observations in the validation set can be predicted, using parameter estimates that have been obtained without the observations of the validation set themselves. Next, the mean squared error of the validation set \mathcal{J}_c is estimated, by comparing the actual dependent variable observations in the set with their respective predictions, as in

$$m\hat{se}_c = \frac{1}{T_c} \sum_{t \in \mathcal{J}_c} (y_t - x_t' \hat{\beta}_c)^2.$$

As a result one obtains the cross-validated error estimate of the whole sample of observations, given the random assignment of C validation sets. It is a weighted average of mean squared errors, such that

$$m\hat{se} = \frac{1}{T} \sum_{c=1}^C T_c \cdot m\hat{se}_c.$$

The use of mean squared errors is motivated because of its meaning with respect to the bias-variance tradeoff. Part of the expected prediction error of a model is irreducible, as it derives from the stochastic part of the data generating process. The other part can be minimized by appropriate feature selection. In expectation, prediction will be negatively affected by either too few degrees of freedom (underfitting, high bias) or too many degrees of freedom (overfitting, high variance). According to James et al. (2013), the formal decomposition of the expected prediction error into squared bias, variance

and irreducible error variance σ_ε^2 is given as

$$E((y - \hat{y})^2) = Bias(\hat{y})^2 + Var(\hat{y}) + \sigma_\varepsilon^2.$$

The mean squared error is an estimator for the expected prediction error. At the minimum of the mean squared error, the optimal capacity of a model is found, which makes mean squared error minimization a common tool of model selection.

6.2.3 Structural breaks

The structural break test statistic will measure the impact of each structural break date candidate k on error estimates. To measure this impact, mean squared errors are estimated separately for two segments of the sample: An error is estimated in the pre- k segment and an error is estimated in the post- k segment. In each of the segments, the error is derived from building cross-validation sets and calculating mean squared errors. Then, the pre- and post- k estimates are joined in a weighted average. This statistic is an indication of the validity of having a structural break in the time series model at date k .²

The determination of the weighted average of pre- and post-break error estimates is described in the following. As the possible break date is unknown ex ante, all potential break dates are considered. For any potential break at date k , the pre-break sample consists of the vectors of linked observations $[y_t, x_t']'$ associated with dates $t = 1, \dots, k$, and the number of pre-break validation sets is $C_k = [\frac{k}{T}C]$. Accordingly, the corresponding post-break sample covers dates $t = k + 1, \dots, T$, and $C_{T-k} = [\frac{T-k}{T}C]$.³ Then, for all $k = \omega, \omega + 1, \dots, T - \omega$, the errors are estimated and joined in a weighted average, such that

$$\begin{aligned} \hat{mse}_{pre}(k) &= \frac{1}{k} \sum_{c=1}^{C_k} T_c \cdot \hat{mse}_c, \\ \hat{mse}_{post}(k) &= \frac{1}{T-k} \sum_{c=1}^{C_{T-k}} T_c \cdot \hat{mse}_c, \\ \hat{mse}(k) &= \frac{1}{T} (k \cdot \hat{mse}_{pre}(k) + (T-k) \cdot \hat{mse}_{post}(k)), \\ &= \frac{1}{T} \left(\sum_{c=1}^{C_k} T_c \cdot \hat{mse}_c + \sum_{c=1}^{C_{T-k}} T_c \cdot \hat{mse}_c \right), \end{aligned}$$

where T_c still refers to the number of dates in the respective validation set c .

²A different formulation has been tried, where cross-validation sets were not separated into pre- and post-break sets. Instead errors were estimated for the full sample, recognizing parameter shifts at a given break date candidate k . This formulation has the advantage of enabling partial structural break testing, but the test performance in simulations worsened.

³Note that potential break dates as early as $k = \omega = 0.15T$ are considered, see Section 6.2.1. This leads to a pre-break validation set number of $C_k = [0.15C]$. To guarantee a moderate number of validation sets in the smallest possible data segment, for example $C_k \geq 3$, the total number of validation sets must be set to a sufficiently large number, $C \geq 17$.

If all break date candidates are considered, there are $T - 2\omega + 1$ subsample error estimates $\hat{m}se(k)$. A low prediction error indicates a high prediction quality of the associated model. Thus, the date k^* that corresponds to the lowest error estimate is taken as a best guess for the possible break date, that is

$$k^* = \arg \min_{k \in [\omega, T-\omega]} \hat{m}se(k), \quad (6.1)$$

$$\hat{m}se^* = \hat{m}se(k^*). \quad (6.2)$$

6.2.4 Test statistic

Each estimation of mean squared errors by the method of cross-validation is affected by the random assignment of validation sets. In effect, also the best guess for the possible break date k^* from equation (6.1) and the lowest error estimate $\hat{m}se^*$ from equation (6.2) depend on the outcome of the random assignment. To diminish the impact of the random draw, the random assignment of validation sets is repeated several times. For each repetition $p = 1, \dots, P$, one obtains specific error estimates for all possible pre- and post-break samples. In effect, each repetition $p = 1, \dots, P$ is associated with a best guess for a possible break date k_p^* from equation (6.1) and a resulting error estimate $\hat{m}se_p^*$ from equation (6.2).

The distribution of outcomes itself is an interesting result worth exploring. The distribution of possible break dates contains information about the type of structural break and the certainty of its position. To obtain a single test statistic result a measure of central tendency can be applied to the error estimates and associates dates from the P repetitions. This is done to obtain the proposed CVE test statistic. Here, the use of the median is selected instead of the arithmetic mean to reduce the impact of outliers. Analogously, the single break date estimate \hat{k} is obtained by applying a measure of central tendency to the P candidate break dates k_p^* from the repetitions. For the same reason as before, the median is favored over the mean. The median is also favored over the mode to incorporate the distribution of all P mean squared errors to an extent.

$$\hat{c}ve = \text{median}_p(\hat{m}se_p^*), \quad (6.3)$$

$$\hat{k} = \text{median}_p(k_p^*).$$

The choice of a measure can lead to different break date estimates in practice, as shown in the empirical study of Section 6.4. The study shows that it is advisable to consider various measures, or even the full range of break dates of the estimation repetitions.

Stability of the tested model is rejected for small values of $\hat{c}ve$. Critical values for the test can be obtained from a bootstrap method, for example a parametric, residual resampling bootstrap, or a non-parametric, case resampling bootstrap, see also Chapter 2. For the residual resampling bootstrap, the series of covariates x_t remain fixed and a synthetic series of the dependent variable y_t is created. The no-break estimates $\hat{\beta}$ are used to calculate fitted values, and the residuals of the model are randomly added to the fitted values (with replacement). That is,

$$y_t^* = x_t' \hat{\beta} + \hat{\varepsilon}_j, \text{ where } j \in \{1, \dots, T\}, \text{ for } t = 1, \dots, T.$$

For case resampling, observations (y_t, x_t) are jointly resampled (with replacement), such that

$$(y_t, x_t)^* = (y_j, x_j), \text{ where } j \in \{1, \dots, T\}, \text{ for } t = 1, \dots, T.$$

Either way, for each bootstrap sample $b = 1, \dots, B$, the prediction error is estimated repeatedly and then the median error $c\hat{v}e$ is determined according to equation (6.3). The collection of all bootstrapped values of the test statistic gives the distribution of the test statistic under the null hypothesis, and critical values can be derived. If the application of the test to empirical data leads to a test statistic value that is smaller than $(1 - \alpha)\%$ of the values observed in the bootstrap simulation, then the null hypothesis of parameter stability can be rejected in favor of the structural break hypothesis, with corresponding confidence.

6.3 Monte Carlo simulations

6.3.1 Introduction

For the evaluation of the proposed test statistic, data has been simulated using various DGP specifications. The simulated series all satisfy the following description of a univariate regression model, that is

$$\begin{aligned} y_t &= x_t' \beta_{pre} + \varepsilon_t, & \text{for } t = 1, \dots, k_0, \\ y_t &= x_t' \beta_{post} + \varepsilon_t, & \text{for } t = k_0 + 1, \dots, T, \\ x_{0,t} &= 1, & \text{for } t = 1, \dots, T, \\ x_{j,t} &\sim \mathcal{N}(0, 1), & \text{for } t = 1, \dots, T, \text{ for } j = 1, \dots, q, \\ \varepsilon_t &\sim \mathcal{N}(0, \sigma_\varepsilon^2), & \text{for } t = 1, \dots, T. \end{aligned}$$

T is the number of dates, and k_0 divides the DGP in two data segments of linear regression models. The parameter vector β contains $q + 1$ regression coefficients, which measure the linear impact of the vector of covariates $x_t = [1, x_{1,t}, \dots, x_{q,t}]'$ on the dependent observation y_t . Each coefficient β_j for $j = 0, \dots, q$ may shift its value after date k_0 . The magnitude of the parameter shift is $\Delta_\beta = \beta_{j,post} - \beta_{j,pre}$.

One specification of the DGP has the properties $T = 100$ and $\sigma_\varepsilon = 1$. Further, $k_0 = 0.5T$, so that the break lies at the center of the sample, and $q = 0$, so that that the model only contains an intercept term. The value of the coefficient before the break is $\beta_{j,pre} = 1$, and the parameter shift has a small size $\Delta_\beta = 0.2$. The properties of this reference specification are changed one after the other so as to study their impact on the test results.

For each considered DGP specification, $S = 10000$ data samples have been simulated. In this Monte Carlo simulation study the proposed cross-validated error estimation test (CVE) competes against the established maximum likelihood ratio test (MLR). Any one simulated data series is subjected to both detection tests. During the Monte Carlo simulations, both detection methods correctly identify the structural break to involve the mean parameter β and assume prior knowledge of σ_ε . Further, for each considered DGP specification, another $S = 10000$ data samples are simulated that do not contain

a structural break with changes in β , but share all other specification properties. That way, the null distribution of each test is empirically observed, critical test values can be derived, and both tests have nominal size. For CVE then, specifically, the distribution does not need to be derived from a bootstrap.

Before an application of either break detection test, a few test setting decisions must be made. In both MLR and CVE, the data trimming is set to $\kappa = 0.15$ in the standard setting. Larger values are found to increase the testing power of both tests slightly at the expense of a wider range of possible break locations. For CVE the number of validation sets is set to $C = 30$, the number of error estimation repetitions is set to $P = 5$, and the median is used to determine mean squared error estimate averages as well as break date estimate averages. Appendix 6.A contains results for using different values of C , P and using the arithmetic mean instead, all with insignificant effect on testing power, mean break date estimates and the standard deviation of break date estimates. The results further indicate that $P = 1$ is a valid option, but this must be seen in the context of the Monte Carlo study with summarized results of 10000 simulation series. In a practical application to a definite sample, larger values of P reduce the impact of each specific draw of random validation sets. Finally, the chosen number of validation sets C only needs to reflect the technical limitations as described before in Section 6.2.3.

6.3.2 Testing power

This section focuses on the power of CVE compared to MLR. Table 6.1 reports results for a large number of different DGP specifications. The table contains the break detection frequency $I_{0.95}^A$ at 95% confidence. Since the exact same simulations are subjected to both tests, it is possible to say which test could detect a break in a given simulation, when the other could not. The second indicator $I_{0.95}^*$ reports the number of such instances, relative to the number of simulations, for a confidence level of 95%. By definition, the difference between the measures must be equal, apart from a rounding error. $I_{0.95}^A(CVE) - I_{0.95}^A(MLR) = I_{0.95}^*(CVE) - I_{0.95}^*(MLR)$. Consider the reference specification, with $\Delta_\beta = 0.2$, $T = 100$, $\sigma_\varepsilon = 1$, $q = 0$, $k_0 = 0.5T$, which appears several times in the table. In $I_{0.95}^* = \frac{98}{10000} \approx 1.0\%$ of the simulations MLR detects the structural break with 95% confidence when CVE does not, and in $I_{0.95}^* = \frac{102}{10000} \approx 1.0\%$ it is the other way around. This difference between MLR and CVE is minimal, but it is worth pointing out that a statistically more powerful method does not necessarily detect each structural break that is detected by the less powerful method.

The implications of changes to the magnitude of the parameter shift Δ_β are shown in the first rows of Table 6.1, which has the expected overall result. The tests become both more powerful, as indicated by larger values for $I_{0.95}^A$. At the largest reported shift, $\Delta_\beta = 0.8$, both MLR and CVE score $I_{0.95}^A = 0.936$. Note again that $I_{0.95}^* = 0.006$ means that both tests have detected structural breaks in a few instances when the other test has not. The higher power for larger Δ_β is interpreted to be caused by a stronger relative break signal against the noise in the series, given a limited number of observations. This interpretation is supported by the break detection frequencies reported in the following rows of Table 6.1. An increase in the number of observations T increases the power, while an increase in the standard deviation of the error term σ_ε decreases the power.

Table 6.1: Testing power for various DGP specifications

				$I_{0.95}^A$		$I_{0.95}^*$	
				MLR	CVE	MLR	CVE
$T = 100, k_0 = 50, \sigma_\varepsilon = 1, q = 0$	$\Delta_\beta = 0.1$		0.1	0.062	0.064	0.005	0.007
			0.2	0.114	0.114	0.010	0.010
			0.4	0.371	0.373	0.017	0.019
			0.8	0.936	0.936	0.006	0.006
$k_0 = 0.5T, \sigma_\varepsilon = 1, q = 0, \Delta_\beta = 0.2$	$T = 50$		50	0.085	0.085	0.009	0.009
			100	0.114	0.114	0.010	0.010
			200	0.201	0.200	0.014	0.013
$T = 100, k_0 = 50, q = 0, \Delta_\beta = 0.2$	$\sigma_\varepsilon = 0.5$		0.5	0.377	0.377	0.019	0.019
			1	0.114	0.114	0.010	0.010
			2	0.069	0.067	0.008	0.007
$T = 100, k_0 = 50, \sigma_\varepsilon = 1, \Delta_\beta = 0.2$	$q = 0$		0	0.114	0.114	0.010	0.010
			1	0.156	0.158	0.018	0.021
			2	0.182	0.185	0.023	0.026
			3	0.203	0.204	0.032	0.033
$T = 100, \sigma_\varepsilon = 1, q = 0$	$\Delta_\beta = 0.2$	$k_0 = 0.3T$	0.3T	0.103	0.106	0.008	0.011
			0.5T	0.114	0.114	0.010	0.010
			0.7T	0.105	0.109	0.008	0.012
	$\Delta_\beta = 0.8$	$k_0 = 0.3T$	0.3T	0.877	0.878	0.009	0.010
			0.4T	0.925	0.926	0.006	0.007
			0.5T	0.936	0.936	0.006	0.006
			0.6T	0.924	0.926	0.006	0.008
			0.7T	0.881	0.880	0.010	0.009

Note: Break detection power of MLR and CVE for different DGP specifications of parameter shift magnitude Δ_β , sample length T , error standard deviation σ_ε , the number of stochastic covariates q , and the true break date location k_0 . The coefficient value before the break is $\beta_{j,pre} = 1$, and CVE settings are $P = 5$ and $C = 30$. There are $S = 10000$ simulations. $I_{0.95}^A$ measures the number of null hypothesis rejections at 95% confidence. $I_{0.95}^*$ is the relative number of simulations in which the test rejects the null hypothesis at 95% confidence and the other test does not.

Table 6.2: Break date estimation for various DGP specifications

		mean(\hat{k})		median(\hat{k})		I_k^*		$m\hat{a}e(\hat{k})$		CVE \succ MLR
		MLR	CVE	MLR	CVE	MLR	CVE	MLR	CVE	
$k_0 = 50$	$\Delta_\beta = 0.1$	50.22	50.26	50	50	0.184	0.193	18.48	18.38	0.562
		50.04	50.31	50	50	0.188	0.214	15.56	15.43	0.578
		50.14	50.13	50	50	0.185	0.216	10.52	10.36	0.599
		49.96	49.92	50	50	0.160	0.185	5.14	5.08	0.555
$\Delta_\beta = 0.2$	$k_0 = 30$	40.81	40.89	34	35	0.160	0.210	16.32	16.28	0.516
		50.04	50.31	50	50	0.188	0.214	15.56	15.43	0.577
		58.07	58.38	65	65	0.185	0.229	17.30	17.03	0.607
$\Delta_\beta = 0.8$	$k_0 = 30$	31.74	31.75	30	30	0.168	0.185	5.16	5.09	0.552
		40.47	40.52	40	40	0.161	0.175	5.13	5.11	0.517
		49.96	49.92	50	50	0.160	0.185	5.14	5.08	0.556
		59.23	59.27	60	60	0.157	0.182	5.17	5.08	0.578
		68.46	68.45	70	70	0.162	0.186	5.11	5.05	0.555

Note: Break date estimation of MLR and CVE for different DGP specifications of parameter shift magnitude Δ_β , and the true break date location k_0 . Sample length is $T = 100$, error standard deviation is $\sigma_\varepsilon = 1$, the number of stochastic covariates is $q = 0$, the coefficient value before the break is $\beta_{j,pre} = 1$, and CVE settings are $P = 5$ and $C = 30$. There are $S = 10000$ simulations. The values are calculated for simulations in which both methods have detected a break at 95% confidence, only. I_k^* is the number of times that a method has positioned the break closer to the true break date than the other method, relative to the number of simulations in which both methods have detected a break at 95% confidence. The mean absolute error estimate $m\hat{a}e = \text{mean}(|\hat{k} - k_0|)$ is reported. Finally, CVE \succ MLR gives the confidence level of preference for CVE over MLR based on mean absolute errors that have been estimated in a bootstrap resampling of the break date estimates.

Increasing the number of coefficients that change their value also increases the power of either method, as shown for $q = 0, 1, 2, 3$. As before, an enhanced break signal benefits both methods equally. The next DGP property that is adjusted is the break date k_0 , given that $T = 100$, and initially for the reference parameter shift of $\Delta_\beta = 0.2$. The power of both tests is negatively affected by off-center break dates. This reduction in power is also observed when the parameter shift is larger, such as $\Delta_\beta = 0.8$. In this case, both tests see a decrease in $I_{0.95}^A$ from 93.6% for $k_0 = 50$ to values close to 88% when the structural break lies at either $k_0 = 30$ or $k_0 = 70$. Typically, for off-center structural breaks CVE is at least as powerful as MLR, but the differences are not significant.

To summarize this section of the Monte Carlo study, the testing power of both MLR and CVE is affected, as expected, by adjustments of Δ_β , T , σ_ε , and q . When structural breaks lie closer to the edges of the sample the testing power of either test is reduced. Further, the results so far do not allow a clear ranking of the two competing methods. MLR and CVE alternate non-systematically in having the largest detection frequency, with minimal differences between them.

6.3.3 Break estimation

This next section of the Monte Carlo study compares the estimation of model parameters. It will focus on the ability of the methods to locate the structural break accurately at its true date k_0 in the data. Appendix 6.A contains the results of comparing the methods with respect to the estimation of the size of the parameter shift Δ_β . Both methods overestimate the size of the parameter shift for small shifts, which is not analyzed in more detail, as both methods perform equally well.

For the evaluation of the break location estimation, consider first the upper four rows of Table 6.2, which compare the break estimation results for different values of shift size Δ_β , given a central structural break at $k_0 = 0.5T = 50$. The shift size does not appear to affect the first indicator, $\text{mean}(\hat{k})$, systematically. A difference between MLR and CVE becomes apparent when looking at I_k^* in the central columns of Table 6.2. This indicator counts the number of times that a method has positioned the break closer to the true break date than the other method, relative to the number of qualified simulations. Qualified means that both methods have detected a break at 95% confidence. The values are relatively stable within method, but CVE values are always higher than MLR values. The smallest difference is reported for $\Delta_\beta = 0.2$, where MLR locates the break more accurately than CVE in 18.4% of qualified simulations, while CVE locates it more accurately in 19.3% of the simulations.

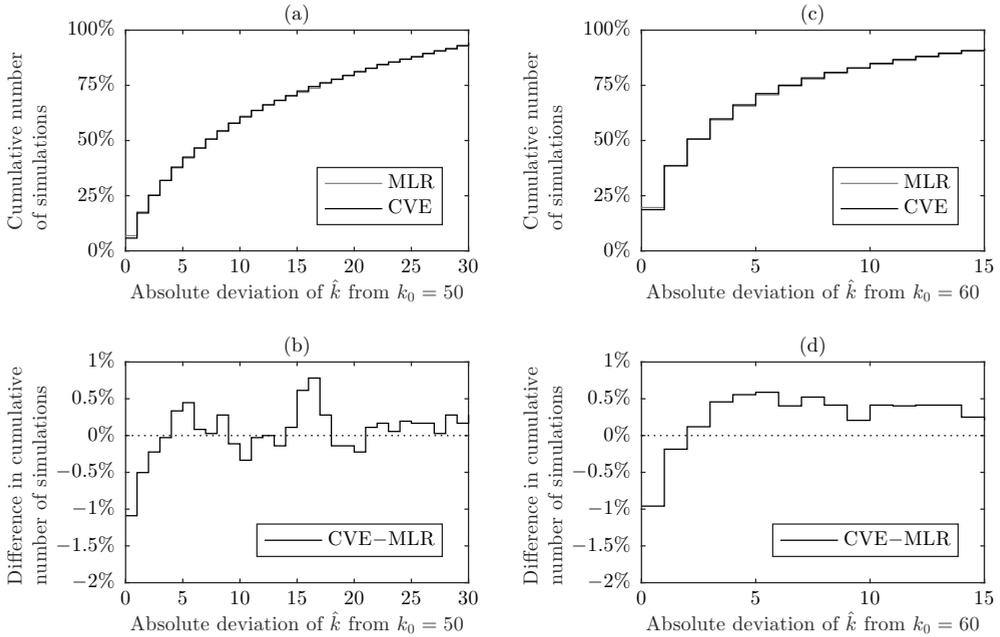
To extend on this result, the mean absolute error of estimating k_0 is estimated for either method by $m\hat{a}e(\hat{k}) = 1/S|\hat{k} - k_0|$. It measures the amount of dates that break date estimation deviates from the correct date on average. As expected, both methods become more precise when the parameter shift increases. In comparison, the mean absolute error is estimated to be smaller for CVE than for MLR for all Δ_β . To see with which confidence CVE is better than MLR based on this measure, full simulation samples have been resampled in a bootstrap. In the case of $\Delta_\beta = 0.4$, the 59.9% lowest mean absolute error estimates of CVE are lower than the 59.9% highest estimates of MLR (while only the 41.1% lowest mean absolute error estimates of MLR are lower than the 41.1% highest estimates of CVE). The last column of Table 6.2 that CVE is generally better than MLR, but the result is not significant.

Results for detecting off-center break dates are shown in the lower parts of Table 6.2, initially for $\Delta_\beta = 0.2$ and then for $\Delta_\beta = 0.8$ below. A property shared by both tests, especially for the smaller parameter shift, is that the absolute deviation of $\text{mean}(\hat{k})$ from the true date k_0 increases substantially, when the break is located more closely to either end of the sample. For the small parameter shift, both methods locate the structural break of $k_0 = 30$ at approximately 40 in the mean. For any DGP specification in the table, CVE is more often than MLR more accurate according to measures I_k^* and $m\hat{a}e(\hat{k})$. The bootstrap repetition indicate that CVE is in 50 – 60% the preferred method over MLR.

To provide more detail on this result, panel (a) of Figure 6.1 plots the frequencies of MLR and CVE estimating break dates at most as many dates away from $k_0 = 50$ as indicated on the horizontal axis. As $T = 100$, and $\kappa T = 15$, a break date estimate can at most lie 35 dates to either side of the true break date $k_0 = 50$. The relatively large parameter shift of $\Delta_\beta = 0.8$ is considered. The plots show that the methods estimate the break to lie exactly at $k = 50$ in close to 6% and 7% of all qualified simulations (in which both methods have detected a structural break with 95% confidence). In close to 50% of qualified simulations, the methods estimate the break to lie at most 7 dates away from $k_0 = 50$. This means that both methods locate the break date within the interval $[43, 57]$ in close to 50% of qualified simulations.

Panel (b) of Figure 6.1 focuses on the differences between the plots of panel (a). A positive number means that CVE has more often than MLR estimated the break location at most as far away from the true break as the value given on the lower axis. On the one hand, the plots show a negative value of approximately -1.1% for an absolute deviation

Figure 6.1: Absolute deviation in break date estimation



Note: Upper panels (a) and (c) plot the summed-up number of simulations in which MLR and CVE have estimated a break to lie as close or closer to the true break date as indicated by the number of dates on the horizontal axis. In other words, for a range of absolute differences on the x-axis, (a) and (c) plot the relative number of times that break estimates lie as close or closer to the true break date. Left panel (a) provides results for $k_0 = 50$, right panel (c) for $k_0 = 60$. The remaining properties of the DGP in the simulations are $T = 100$, $\sigma_\varepsilon = 1$, $q = 0$, $\beta_{j,pre} = 1$, $\Delta_\beta = 0.8$. CVE settings are $P = 5$ and $C = 30$. From the original $S = 10000$ simulations, only the ones where both MLR and CVE have detected a break at 95% confidence have been used, resulting in 9304 simulations for $k_0 = 50$, and 9178 simulations for $k_0 = 60$. Lower panels (b) and (d) plot the respective difference between the CVE and the MLR plot above, for $k_0 = 50$ and $k_0 = 60$. For the plots in (b) and (d), if the vertical axis value is positive, CVE has more often than MLR detected break dates with a maximum distance to k_0 as given by the value on the horizontal axis.

of 0 dates. This means that MLR has estimated $\hat{k} = k_0 = 50$ in approximately 1.1% more simulations than CVE. However, the plot shows a positive value for an absolute deviation of 4 dates, which means that CVE has more often than MLR estimated the break date to lie at most 4 dates away from $k_0 = 50$. The plot of panel (b) explains the earlier result that MLR has similar values for $\text{mean}(\hat{k})$ and $\text{median}(\hat{k})$, but that CVE is more often more accurate than MLR. MLR has a slightly higher rate of finding the true break date, but CVE can locate the break date more often within a range of a few dates around the break date.

The two right panels (c) and (d) of Figure 6.1 show the same types of plots, but for an off-center parameter shift of $\Delta_\beta = 0.8$ at $k_0 = 60$. The results confirm the earlier result, in that CVE is more often more accurate than MLR, but MLR has few more exact estimates at $\hat{k} = k_0$. This result is confirmed for the other DGP specifications, although less clearly when the parameter shift is as small as $\Delta_\beta = 0.2$.

To conclude the Monte Carlo study, neither method can confidently be referred to as the single best method, as neither method is always more powerful or accurate or

efficient than the other. The application of CVE does not lead to a higher frequency of correct break date estimates, but it can estimate structural breaks to lie relatively close to the true break more often than MLR. This result may be explained by the use of the median date in the CVE test, so that relatively extreme break date estimates can only be inferred, if they are estimated in many repetitions of the cross-validation method.

6.4 Empirical Application

Chapter 4 has introduced an empirical study on more than a decade of daily redemption yield observations of 18 European central government bonds with a maturity of 10 years. The estimation of linear regression models for differences in the redemption yield, Δr_i , has uncovered a variety of financial series that can explain some of the variation in the different European series. The individual models have each different sets of significant covariates, ranging from own lagged values to global stock index data and gold prices.

The most prevalent covariate is included in all but one model, the lagged difference in U.S. 10-year treasury bond redemption yields, $\Delta(r_{US})_{t-1}$. Given the available information and other available variables, the variation in all considered European yield series can partly be explained by developments in the U.S. bond yields one trading day earlier. Greece is the exception in this application, for which the U.S. was not found to contribute significantly and which is therefore not covered in this study. The coefficients, estimated for data samples between 2003-1-9 and 2016-12-21, are all positive and range between 0.097 for Portugal and 0.243 for the United Kingdom. Long-term U.S. government bonds are often seen as practically risk-free assets, and so it is interesting to further investigate the recent predictive value of $\Delta(r_{US})_{t-1}$. The chosen starting point for the study is the summer of 2014, as the year can be seen as a turning point towards economic recovery in the euro area following the European sovereign debt crisis. Ireland and Portugal exited financial support programs in July 2014, and 2014 annual GDP growth rates rose above zero again, in the Netherlands and elsewhere.

In the following, the CVE test is applied to detect a structural break in each linear relation between weighted observations of yield series of countries i and the U.S., with $i = BD, BG, CZ, DK, ES, FN, FR, IR, IT, NL, NW, OE, PO, PT, SD, SW, UK$. Greece is not studied in this empirical application. The samples span $T = 600$ trading days between 2014-08-04 and 2016-12-21. The analysis is performed separately for each country i , using weights w_t according to Cook's distance, and conditional on no-break relations between $\Delta(r_i)_t$ and other covariates. For this, the Frisch-Waugh-Lovell Theorem is applied. Instead of determining covariate coefficients in multivariate models jointly with the other coefficients, the studied covariate coefficients are separately estimated and can afterwards also be tested for structural breaks by CVE, separately. In practice, each weighted series is orthogonally projected off the remaining, weighted covariates, WX , to obtain residual, weighted observations $M_{WX}w_t\Delta(r_i)_t$ and $M_{WX}w_t\Delta(r_{US})_{t-1}$. Then, a simple linear regression can be performed for each European country i in the set, so that

$$M_{WX}w_t\Delta(r_i)_t = \beta_i \cdot M_{WX}w_t\Delta(r_{US})_{t-1} + u_t.$$

Table 6.3: Structural break estimation for European countries, lagged U.S. covariate

i	full sample		CVE						MLR		\hat{k}
	$\hat{\beta}$	$\hat{\sigma}^{resid}$	$\hat{\beta}_{pre}$	$\hat{\beta}_{post}$	$\hat{\sigma}_{pre}^{resid}$	$\hat{\sigma}_{post}^{resid}$	p -value	median(k^*)	mode(k^*)	p -value	
<i>BD</i>	0.210	0.0014	0.001	0.315	0.0008	0.0016	0.011	2015-04-20	2015-04-17	0.003	2015-04-28
<i>BG</i>	0.222	0.0015	0.019	0.327	0.0011	0.0016	0.015	2015-04-14	2015-04-10	0.123	
<i>CZ</i>	0.084	0.0008					0.806			0.031	2015-09-29
<i>DK</i>	0.197	0.0018	-0.020	0.305	0.0017	0.0017	0.024	2015-04-10	2015-04-06	0.039	2014-12-31
<i>ES</i>	0.158	0.0029					0.172			0.271	
<i>FN</i>	0.204	0.0014	-0.007	0.312	0.0009	0.0015	0.009	2015-04-14	2015-04-09	0.048	2015-04-28
<i>FR</i>	0.238	0.0016	0.038	0.346	0.0010	0.0018	0.013	2015-04-15	2015-04-13	0.013	2015-04-28
<i>IR</i>	0.153	0.0019					0.182			0.338	
<i>IT</i>	0.223	0.0028					0.395			0.613	
<i>NL</i>	0.213	0.0013	0.043	0.302	0.0009	0.0015	0.046	2015-04-17	2015-04-20	0.019	2015-04-28
<i>NW</i>	0.219	0.0013					0.475			0.047	2016-03-01
<i>OE</i>	0.232	0.0014	0.020	0.339	0.0009	0.0016	0.009	2015-04-15	2015-04-13	0.010	2015-04-28
<i>PO</i>	0.252	0.0030					0.322			0.804	
<i>PT</i>	0.192	0.0060					0.624			0.849	
<i>SD</i>	0.216	0.0015					0.421			0.605	
<i>SW</i>	0.123	0.0008					0.404			0.008	2015-01-02
<i>UK</i>	0.329	0.0024					0.394			0.825	

Note: For each considered European country i , structural break test results for a simple regression model with dependent variable $M_{WX}w_t\Delta(r_i)_t$ and covariate $M_{WX}w_t\Delta(r_{US})_{t-1}$. First, the estimated parameters of a stable model, $\hat{\beta}$ and $\hat{\sigma}^{resid}$, are listed. Second, if CVE testing is associated with a p -value of below 0.05, pre- and post-break parameters are given for the respective series i as well as break date estimates median(k^*) and mode(k^*). Third, if MLR testing is associated with a p -value of below 0.05, the corresponding break date estimate is given.

Estimates of β_i for the full sample of $T = 600$ dates range between 0.084 for the Czech Republic and 0.329 for the United Kingdom. They can be found in the first column of Table 6.3, as well as the respective central panels of Figures 6.3, 6.4, 6.5, 6.6, 6.7 and 6.8.

Next, the CVE test is applied to the simple regression model of residual, weighted observations to investigate the existence of a change in the linear regression coefficient β . The results of setting the number of validation sets $C = 30$ are reported, but the break date estimation is robust to other settings ($C = 20, 50, 100$) in this application. The error estimation is repeated a large number of times, with $P = 4999$. Significance is assessed by a parametric, residual resampling block bootstrap. For this, blocks of residuals are randomly added to the fixed series of fitted values $\hat{\beta}_i \cdot (M_{WX}w_t\Delta(r_{US})_{t-1})$. Using a non-parametric, case-resampling block bootstrap instead increases the p -value in this application. The block size is set to 10 data points, other block sizes (5, 20) increase the p -value slightly. The results can be found in the central block of columns of Table 6.3.

Structural breaks are detected for six countries with confidence of more than 95% in each case. The group consists of six central euro area countries Germany, Belgium, Finland, France, the Netherlands and Austria, as well as Denmark, whose currency is pegged to the euro. The shaded column of Table 6.3 lists the corresponding, estimated break dates, when using the median break date of each error estimation repetition. The break dates all lie within a two-week range starting on 2015-04-10 (Denmark) and ending on 2015-04-20 (Germany). For each of the seven cases, the estimates covariate coefficient of $M_{WX}w_t\Delta(r_{US})_{t-1}$ increases at the break. Between 2014-08-04 and the respective structural break in April 2016, the coefficient estimates are at most as large as 0.043 (the Netherlands). The smallest coefficient increase due to the structural break is estimated for the Netherlands, with a shift of +0.259 to 0.302. The coefficients of two countries rise to values higher than for the United Kingdom, namely 0.339 for Austria and 0.346 for

France. Estimates of the standard deviations of the residuals of the corresponding simple linear regressions are also provided in Table 6.3. They show that the later segment of each country sample with a structural break is a bit more volatile than before, which largely corresponds with the findings of the structural break analysis of Chapter 4.4.

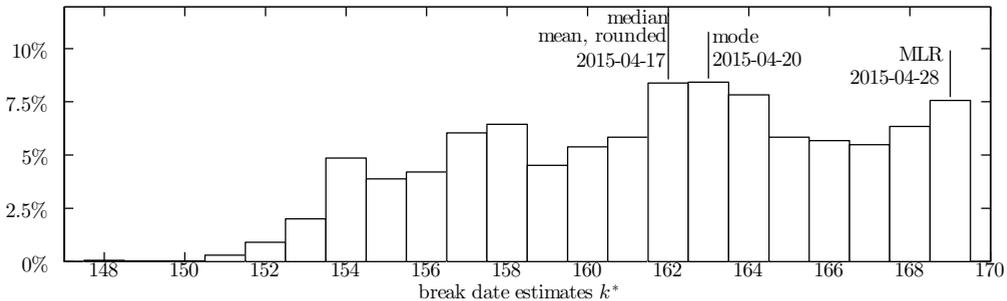
An increased, positive coefficient estimate is an indication that more than before, and all other things equal, the market has perceived the corresponding government bonds to be a similar investment as U.S. government bonds. A reasonable explanation for this development is the initiation of the public-sector purchase programme, a quantitative easing program of the euro area to raise inflation rates. In correspondence with its dual mandate of price stability and employment maximization, the U.S. had introduced its first quantitative easing program at the end of 2008 to avert liquidity shortages and reduce the negative economic impact of the Global Financial Crisis. The change in monetary policy of the ECB is a possible explanation for the larger co-movement between the two series, especially since the monetary policy change directly involves government bonds.

While the public-sector purchase programme had been announced in early 2015, it had been initiated within weeks of the structural breaks estimated by CVE. The European Central Bank (ECB) had released data on the first month of the program on 2015-04-07, just a few days before the breaks in mid-April 2015. News of the success of the programme appear to have strengthened the positive relation between the lagged price of the U.S. government bonds and the bond prices of the six central euro area countries and Denmark. This can be attributed to an increased perceived similarity between these European bonds and the American bonds as instruments for monetary policy aims.

Using MLR instead of CVE would have led to different conclusions, as shown in the final two columns of Table 6.3. The break of Belgium is not found by MLR, while the break for the other six central euro area countries is located later on 2015-04-28. The break for Denmark is located on 2014-12-31 leading up to the ECB programme announcement and three additional breaks are found spread over the sample. Since a prediction error is estimated multiple times for CVE testing, it is interesting to see how the MLR results relate to the possible break dates from the CVE repetition steps. A histogram of the possible break date estimates k^* can be plotted, as shown in Figure 6.2 for the example of the Netherlands and the $P = 4999$ repetition steps. The MLR break date estimate 2015-04-28 appears as a possible CVE break date in 7.6% of the repetitions, but in the remaining repetitions possible break dates are all located earlier in the sample and thus closer to the programme initiation and subsequent news. A first local peak is observed for $k = 152$ corresponding to 2015-04-07, the day of the first data release by the ECB.

Overall, the histogram illustrates the richness in information and the ambiguity associated with structural break testing. Clearly, the linear relation between some of the European series and the U.S. series is significantly different before and after April 2015, but the change may be located at different dates in the sample depending on the method. In Section 6.2.4 the use of the median(k^*) is suggested to account for the variety of preliminary break date estimates to some extent, without being influenced strongly by outliers. In this empirical application for the Netherlands, the rounded arithmetic mean date leads to the same result as the median, whereas the mode with 8.4% of repetitions results in a slightly different break date estimate on the next date in the series, corresponding to 2015-04-20. This result does not discourage the use of the median or any other specific

Figure 6.2: Break date estimates in prediction error repetitions, the Netherlands



Note: Histogram of the preliminary break date estimates k^* in CVE testing for the simple regression model with dependent variable $M_{WX}w_t\Delta(r_{NL})_t$ and covariate $M_{WX}w_t\Delta(r_{US})_{t-1}$, from $P = 4999$ repetitions. Median(k^*), rounded arithmetic mean(k^*), and mode(k^*) are marked, as well as the date that maximizes the likelihood ratio statistic.

metric, instead it encourages a comprehensive investigation of the full set of possible break date estimates to account for a substantial degree of uncertainty of estimation.

To further investigate the possible break date distribution for the Dutch example, the behavior of the associated LR and mean squared error statistics and implications for the resulting linear regression, consider Table 6.4. The first row of the table shows a no-break model, while the remaining rows show data corresponding to a range of candidate break dates between $k = [143, 174]$. The first column of the table lists the relative number of times that CVE has chosen a date k as a preliminary break date estimate in the $P = 4999$ estimation repetitions. Accordingly, mode(k^*) is marked where the value is highest at 8.4%. The second column reports the relative number of times that CVE has chosen date k or lower, and so median(k^*) is marked where the values surpasses 50%. The likelihood ratio of a structural break at k is reported in the third column, with its maximum of $LR(169) = 25.10$ marked accordingly. Values for $LR(k)$ steeply decline after $k = 169$, whereas the LR values for median and modal CVE break date estimates are relatively high, with $LR(162) \approx 0.92 \cdot LR(169)$ and $LR(163) \approx 0.94 \cdot LR(169)$.

To determine the preliminary estimates k^* of each prediction error estimation repetition, estimates are produced for each candidate date k . Table 6.4 summarizes the results by reporting median($m\hat{se}(k)$) for each reported k . The lowest median($m\hat{se}(k)$) is recorded for the median preliminary break date estimate, $k = 162$. MLR estimate $k = 169$ is the last date for which median($m\hat{se}(k)$) is relatively close to the minimum, median($m\hat{se}(169)$) $\approx 1.0002 \cdot$ median($m\hat{se}(162)$). In the last columns of the table, maximum likelihood estimates of the simple regression model parameters for each reported candidate break date k are reported. Within the range of possible CVE break date estimates $k = [148, 169]$ and especially between the final CVE and MLR estimates $k = 162, 163, 169$, the estimated coefficients vary only slightly. In this example, the largest difference between pre- and post-break estimates of the regression coefficient β is found at the median CVE break date estimate, with an increase by roughly $0.302 - 0.043 = 0.258$, while the largest difference between residual standard deviations $\hat{\sigma}_u$ is found at the MLR break date estimate, with an increase by roughly $0.00152 - 0.00085 = 0.00067$. These results and differences are not generally the same in all applications, but instead highlight

that different properties of the data may be picked up differently by different structural break testing methods, with minor implications for the model.

As a final illustration of the CVE test results, Figures 6.3 and 6.4 depict scatter plots for six of the tested countries, $i = DK, IR, NW, OE, PT, UK$. Appendix 6.A contains the plots of the remaining eleven countries. In the graphs, the countries' residual, weighted observations as used in the estimation of the structural break tests are plotted. For the ten countries without significant structural break according to CVE testing, the most probable structural break date candidates are used in the plots. In the figures, each left panel includes only observations before the estimated break date (or most probable break date candidate), each central panel includes the full sample, and each right panel includes only observations after the estimated break date (or most probable break date candidate). The corresponding, estimated linear regression line is added to each panel, so that the significant change in covariate coefficient (or lack thereof) is illustrated. Comparison between markets is made easy by the use of the same scaling.

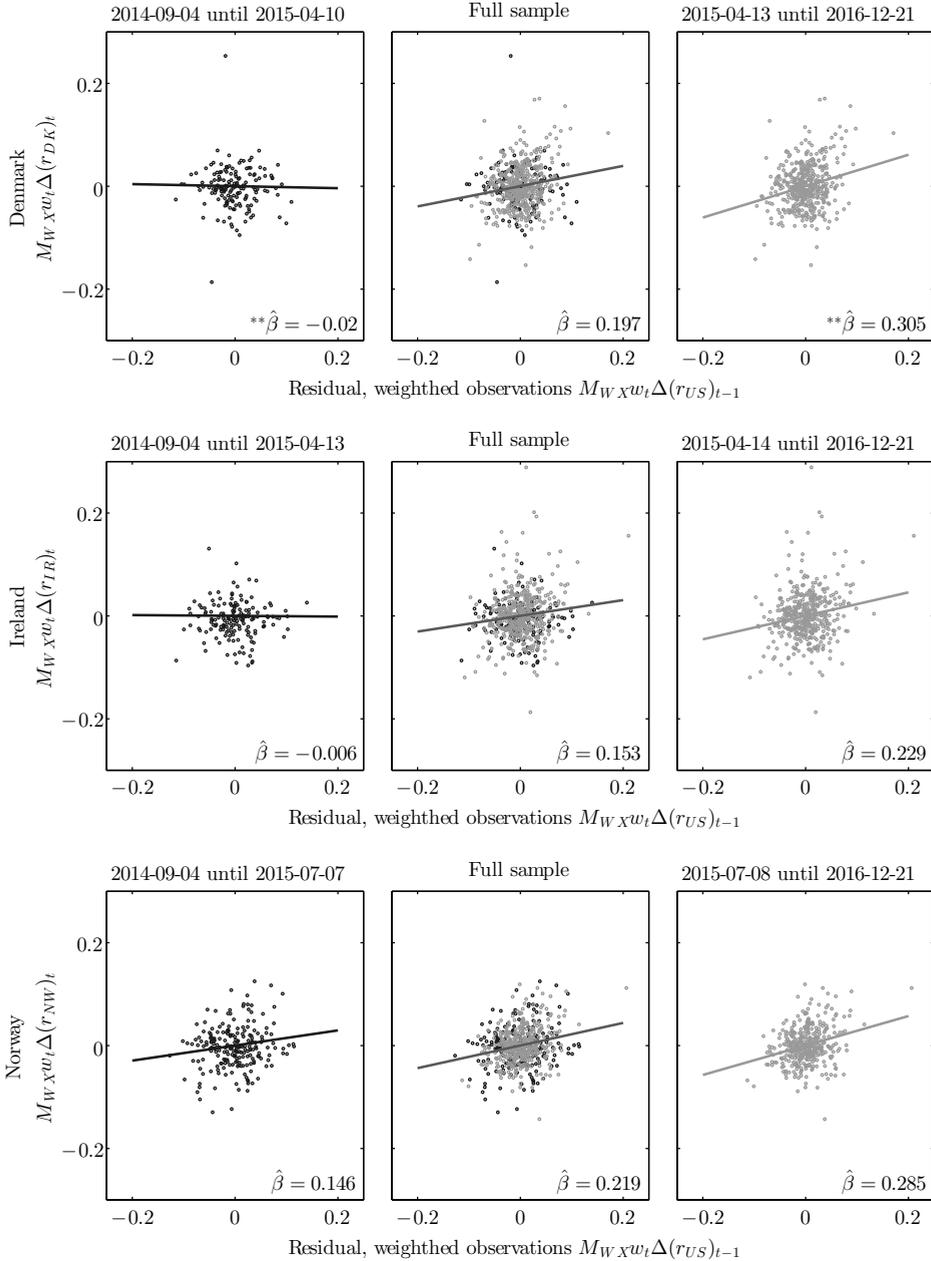
The plots underline the differences between the markets. Clearly, the slope of the regression lines are significantly steeper in the second data segment for the countries that are associated with a structural break, such as Austria in Figure 6.4. Judging from the regression lines, the four periphery countries Ireland, Italy, Portugal and Spain appear to experience a change: In the first part of the sample there is only weak evidence of correlation with the lagged U.S. values, whereas the correlation appears to have increased in the second part. However, the data is too noisy and so no structural break could be inferred at the usual significance levels. The coefficient for Norway is relatively high throughout the sample and no structural break is detected, as well as for Poland, whereas the coefficients of the Czech Republic and Switzerland stay at a low level. Interestingly, the strongest evidence for a structural break in the UK series is positioned towards the end of the sample at 2016-02-25, after which lagged U.S. developments appear to be a weaker predictor of UK developments, see Figure 6.4.

The empirical analysis concludes that the predictive value of lagged U.S. government bond yield differences for many European government bond yield differences has increased between 2014-08-04 and 2016-12-21. For six central euro area countries and Denmark, a model that differentiates two segments, separated at some point in April 2015, is favored over a model that has no structural break. The result appears to relate to the the publishing of a successful change of ECB monetary policy that is comparable to U.S. monetary policy. No significant change is recorded for euro area periphery countries, whose Government bonds are perceived differently by the market. The decision to favor the structural break model over a no-break model can be based on a significant reduction in estimated mean squared errors according to the CVE test. The true break date is not known, and since the cross-validation estimation of errors depends on the random validation set draw, a range of dates during April 2015 are possible break dates. In practice, the result interpretation benefits from a recognition of this estimation uncertainty.

Table 6.4: Estimation results for possible break dates, the Netherlands

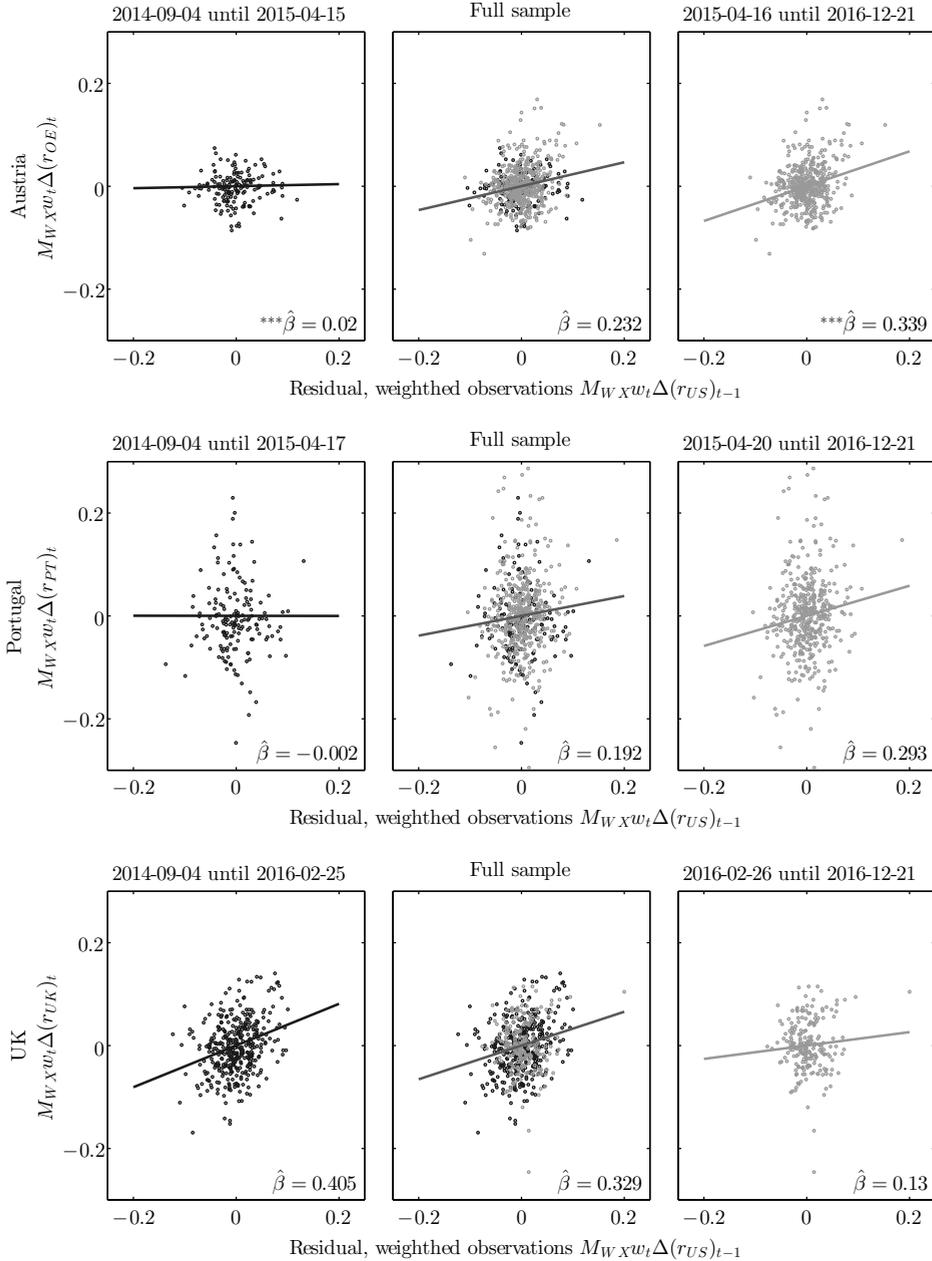
k	date	$k^* = k$	$k^* \leq k$	$LR(k)$	median $\hat{\beta}$		$\hat{\sigma}_u$		
					($mise(k)$)	pre- k	post- k	pre- k	post- k
none	-	-	-	0.00	$1.3543 \cdot 10^{-3}$	0.213	0.00135		
143	2015-03-23	0	0	14.65	$1.3471 \cdot 10^{-3}$	0.062	0.274	0.00092	0.00146
144	2015-03-24	0	0	15.14	$1.3470 \cdot 10^{-3}$	0.061	0.275	0.00092	0.00147
145	2015-03-25	0	0	15.41	$1.3474 \cdot 10^{-3}$	0.064	0.274	0.00091	0.00147
146	2015-03-26	0	0	16.51	$1.3450 \cdot 10^{-3}$	0.053	0.282	0.00091	0.00147
147	2015-03-27	0	0	17.21	$1.3445 \cdot 10^{-3}$	0.052	0.285	0.00090	0.00147
148	2015-03-30	0.001	0.001	17.68	$1.3444 \cdot 10^{-3}$	0.053	0.285	0.00089	0.00148
149	2015-03-31	0.000	0.001	17.89	$1.3448 \cdot 10^{-3}$	0.056	0.285	0.00089	0.00148
150	2015-04-01	0.000	0.001	17.00	$1.3443 \cdot 10^{-3}$	0.053	0.286	0.00091	0.00147
151	2015-04-02	0.003	0.004	17.70	$1.3430 \cdot 10^{-3}$	0.048	0.290	0.00091	0.00147
152	2015-04-03	0.009	0.013	18.40	$1.3424 \cdot 10^{-3}$	0.046	0.292	0.00091	0.00147
153	2015-04-06	0.020	0.033	19.15	$1.3417 \cdot 10^{-3}$	0.045	0.295	0.00090	0.00148
154	2015-04-07	0.049	0.082	19.96	$1.3408 \cdot 10^{-3}$	0.043	0.298	0.00090	0.00148
155	2015-04-08	0.039	0.120	20.21	$1.3409 \cdot 10^{-3}$	0.045	0.298	0.00089	0.00148
156	2015-04-09	0.042	0.162	20.73	$1.3411 \cdot 10^{-3}$	0.044	0.298	0.00089	0.00149
157	2015-04-10	0.060	0.223	21.39	$1.3407 \cdot 10^{-3}$	0.044	0.300	0.00088	0.00149
158	2015-04-13	0.064	0.287	21.94	$1.3406 \cdot 10^{-3}$	0.044	0.300	0.00087	0.00149
159	2015-04-14	0.045	0.332	22.12	$1.3408 \cdot 10^{-3}$	0.045	0.299	0.00087	0.00149
160	2015-04-15	0.054	0.386	21.95	$1.3406 \cdot 10^{-3}$	0.045	0.300	0.00087	0.00149
161	2015-04-16	0.058	0.445	22.44	$1.3407 \cdot 10^{-3}$	0.045	0.300	0.00087	0.00150
162	2015-04-17	0.084	*0.529	23.12	$1.3401 \cdot 10^{-3}$	0.043	0.302	0.00086	0.00150
163	2015-04-20	*0.084	0.613	23.64	$1.3402 \cdot 10^{-3}$	0.043	0.302	0.00086	0.00150
164	2015-04-21	0.078	0.691	23.88	$1.3403 \cdot 10^{-3}$	0.044	0.301	0.00086	0.00151
165	2015-04-22	0.058	0.749	23.02	$1.3405 \cdot 10^{-3}$	0.045	0.301	0.00087	0.00150
166	2015-04-23	0.057	0.806	23.42	$1.3406 \cdot 10^{-3}$	0.046	0.300	0.00087	0.00151
167	2015-04-24	0.055	0.861	23.93	$1.3407 \cdot 10^{-3}$	0.046	0.300	0.00086	0.00151
168	2015-04-27	0.063	0.924	24.56	$1.3405 \cdot 10^{-3}$	0.045	0.301	0.00086	0.00151
169	2015-04-28	0.076	1	*25.10	$1.3404 \cdot 10^{-3}$	0.045	0.302	0.00085	0.00152
170	2015-04-29	0	1	18.13	$1.3469 \cdot 10^{-3}$	0.078	0.287	0.00092	0.00150
171	2015-04-30	0	1	16.27	$1.3464 \cdot 10^{-3}$	0.075	0.288	0.00096	0.00149
172	2015-05-01	0	1	16.85	$1.3461 \cdot 10^{-3}$	0.074	0.290	0.00095	0.00149
173	2015-05-04	0	1	14.68	$1.3487 \cdot 10^{-3}$	0.090	0.283	0.00097	0.00149
174	2015-05-05	0	1	14.86	$1.3472 \cdot 10^{-3}$	0.083	0.288	0.00098	0.00148

Note: Estimation results associated with several break date candidates k , and the no-break case at the top, for the simple regression model with dependent variable $M_{WX}w_t\Delta(r_{NL})_t$ and regressor variable $M_{WX}w_t\Delta(r_{US})_{t-1}$. The first two columns report the number of error estimation repetitions, relative to $P = 4999$, in which the date with the smallest mean squared error estimate is equal to k , or equal to at most k . Column three lists the LR value associated with each k . Median values of error estimates are reported for each k in column four. The last four columns report the model parameter estimates, using maximum likelihood estimation with a Gaussian distribution function, for each implied model. Note that $\hat{\sigma}_u$ is an estimate of the error term standard deviation for the simple regression model of $M_{WX}w_t\Delta(r_{NL})_t$ and $M_{WX}w_t\Delta(r_{US})_{t-1}$, not for the unexplained variation of $\Delta(r_{NL})_t$ directly.

Figure 6.3: Linear regressions for pre-, no-, and post-break data, part 1/6

Note: For $i = DK, IR, NW$, scatter plots for residual, weighted observations $M_{WX}w_t\Delta(r_{US})_{t-1}$ and $M_{WX}w_t\Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

Figure 6.4: Linear regressions for pre-, no-, and post-break data, part 2/6



Note: For $i = OE, PT, UK$, scatter plots for residual, weighted observations $M_{WX} w_t \Delta(r_{US})_{t-1}$ and $M_{WX} w_t \Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

6.5 Conclusion

The chapter has described a novel idea for detecting structural breaks in univariate time series data by optimizing the estimated out-of-sample model fit. A break is detected, when the division and estimation of a data sample in separate parts yields a significant reduction in mean squared error estimates of the prediction error. Cross-validation is used in the method to acquire prediction error estimates, which make the test widely applicable, as the tested model must only fulfill minimal assumptions required for cross-validation. No standard null hypothesis distribution has yet been derived, which is a task for future research. Instead, significance is established by resampling data in a bootstrap such that the simulated data satisfies the no-break null hypothesis.

The test is compared to the popular maximum likelihood ratio break test in a Monte Carlo simulation study. In this study, the data generating process has been specified to satisfy a linear regression model with normal errors in two separate segments. The segments are divided by a structural break that changes the covariate coefficients. In the simulations, the cross-validated error estimation test is found to have approximately equal testing power as the maximum likelihood ratio test. The study further compares the relative number of times that a method detects a structural break closer to its true location in the sample than the other method. Throughout the simulations this value is larger for the cross-validated error estimation test than for the maximum likelihood ratio test. Further, the cross-validated error estimation test leads to lower absolute deviation from the true break date with 50-60% confidence in the study. This is not a significant result, but it can be concluded that the method is a valid alternative to maximum likelihood ratio testing.

A possible explanation for a higher accuracy and precision is that prediction errors cannot be measured and instead must be estimated. To avoid too strong of an influence of each individual random draw of cross-validation sets, the method is repeated several times and a measure of central tendency is used to infer breaks and break date estimates. The method incorporates the uncertainty of estimation automatically, whereas maximum likelihood ratio estimates are invariant in a given sample of observations. This can explain why the cross-validated error estimation method is more robust against large errors of structural break date estimation. Future research should investigate, how a resampling method may improve maximum likelihood ratio testing accordingly.

An empirical application to European government bond yields illustrates the test application for a partial break hypothesis. For this, a separate structural break analysis for one predictor variable is performed, conditional on the no-break model for the remaining covariates. The results indicate an increase in the linear relation between differences in seven central European government bond yields and lagged differences in the U.S government bond yield. With some variation between the countries, the break is estimated to occur in mid-April of 2015. During that time, first results were published on the success of the new public-sector purchase programme of the European Central Bank. This quantitative easing program may have increased the correlation between euro area and U.S. government bonds, as the U.S. had long been using similar monetary tools following the 2008 Global Financial Crisis.

Like other break detection methods, the proposed test is not an absolute measure of model adequacy. It can reject the structural stability hypothesis in favor of a structural

break hypothesis, but this is not to be mistaken for evidence that a structural break model is the appropriate choice. Future studies could focus on the extension of the test to multivariate data, by means of a geometric median, for example. It should be analyzed how the test performs when it is applied to cross-sectional data to detect group separations, or when it is applied to more general forms of time series and panel data. The collection of preliminary break date estimates may be used in a formalized test of parameter transition abruptness or break date estimation uncertainty. Lastly, it will be interesting to describe a standard distribution of the test under the null hypothesis.

6.A Appendix

Table 6.5: The impact of CVE settings on testing

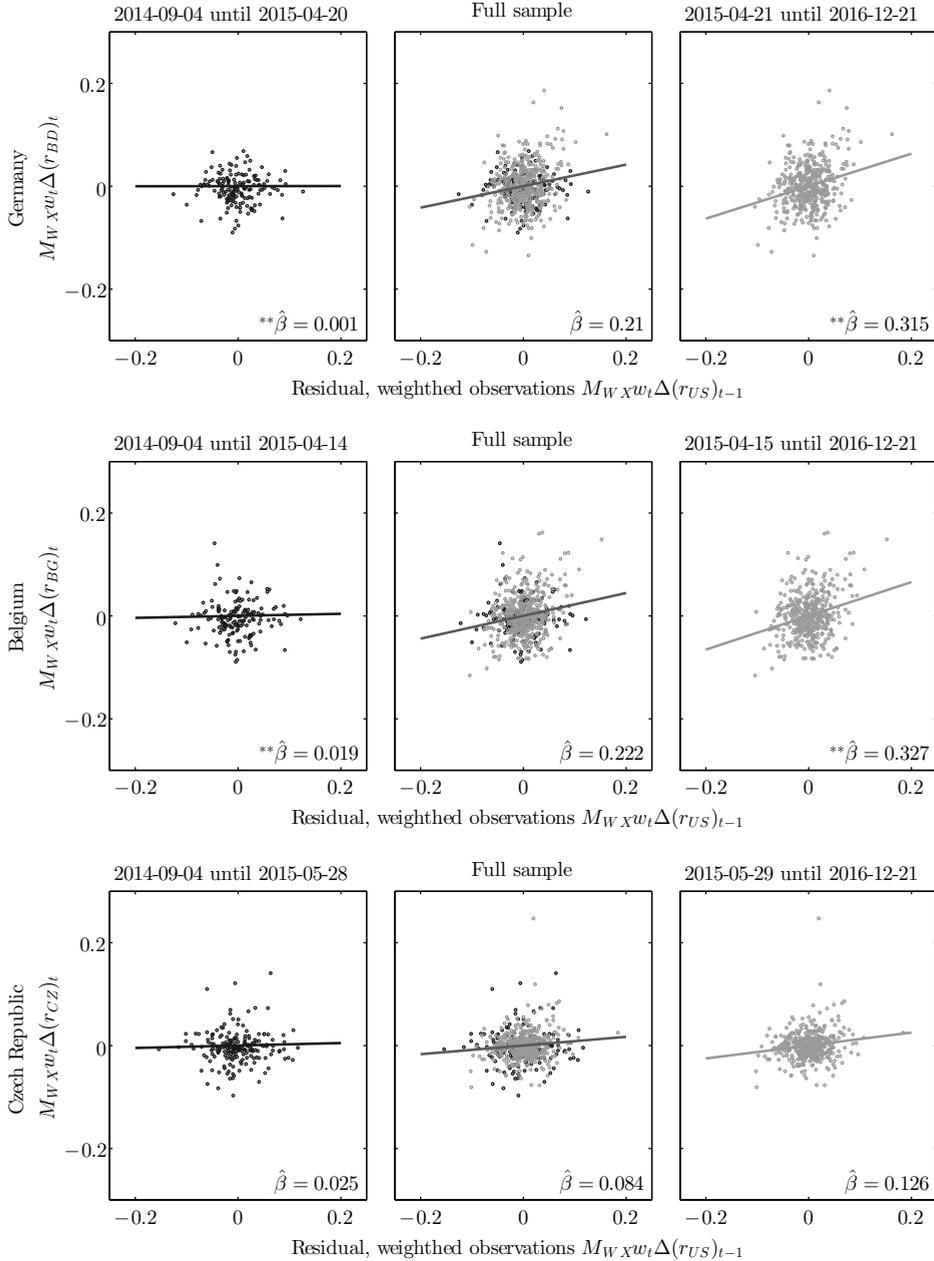
		Power		Location	
		$I_{0.95}^A$	$\text{mean}(\hat{k})$	$\text{median}(\hat{k})$	$\text{sd}(\hat{k})$
$C = 30$, median	$P = 1$	0.114	50.415	50	19.337
	3	0.113	50.587	50	18.965
	5	0.114	50.393	50	19.229
	9	0.112	50.116	50	19.036
	19	0.112	50.098	50	19.090
$P = 5$, median	$C = 20$	0.115	50.695	50	18.985
	30	0.114	50.393	50	19.229
	40	0.115	50.029	50	19.228
$C = 30$, $P = 5$	median	0.114	50.393	50	19.229
	mean	0.113	49.951	50	18.637

Note: Break detection results for CVE for different settings in the number of cross-validation sets C , the number of repetitions in prediction error difference estimation P , and the average measure. The DGP is specified with the reference properties $T = 100$, $k_0 = 0.5T$, $\sigma_\varepsilon = 1$, $q = 0$, $\beta_{j,pre} = 1$, $\Delta_\beta = 0.2$. There are $S = 10000$ simulations, where each simulated series is subjected to all considered test settings to ensure comparability. Location measures use simulations where CVE has rejected the null hypothesis at $1 - \alpha = 0.95$ confidence, only.

Table 6.6: Parameter shift estimation

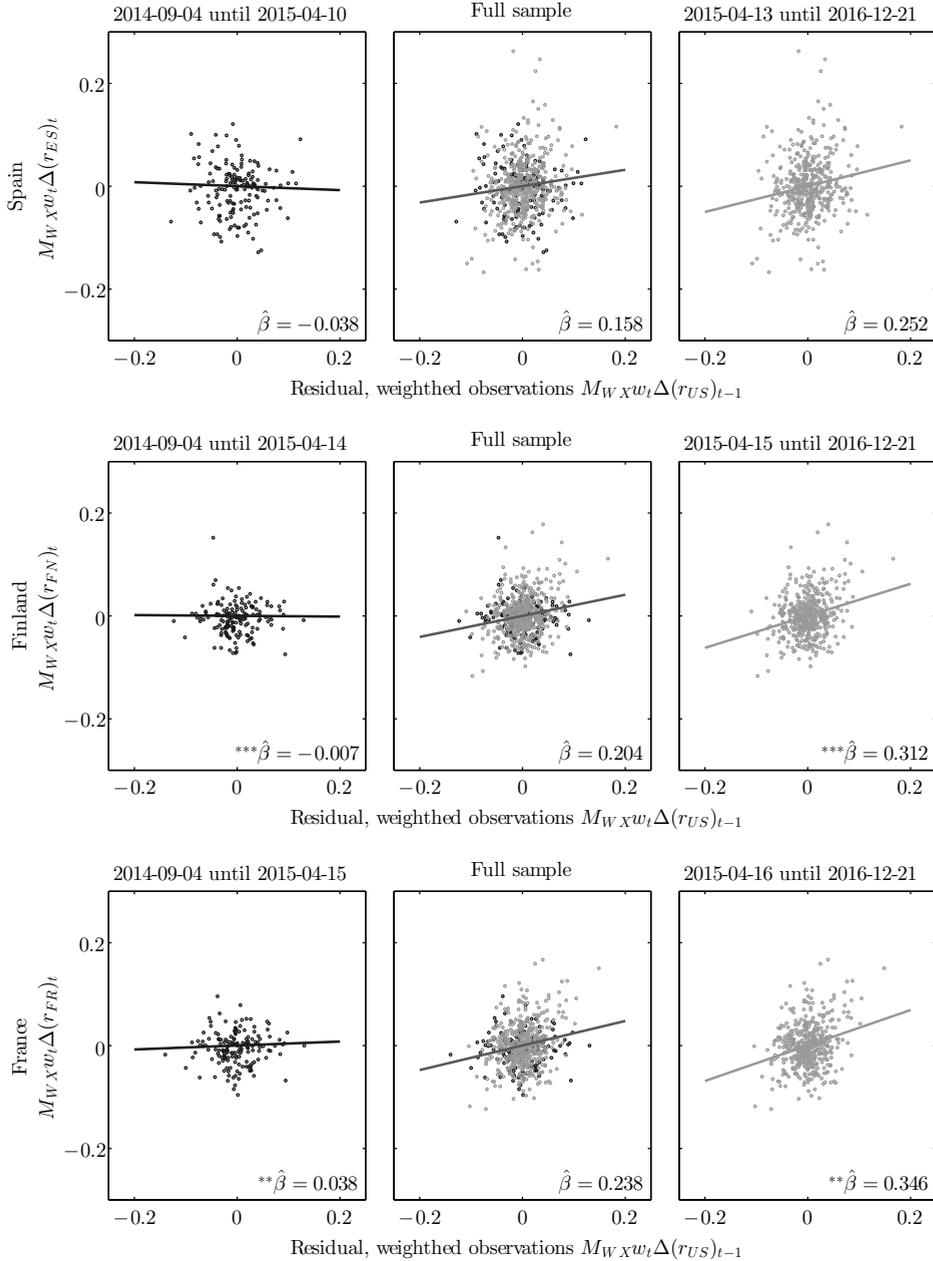
Δ_β	$\text{mean}(\hat{\Delta}_\beta)$		$\text{sd}(\hat{\Delta}_\beta)$	
	MLR	CVE	MLR	CVE
0.1	0.489	0.482	0.561	0.555
0.2	0.688	0.679	0.262	0.259
0.4	0.739	0.729	0.133	0.135
0.8	0.899	0.892	0.167	0.170

Note: Parameter shift magnitude estimation of MLR and CVE for different DGP specifications of the true shift magnitude Δ_β . Sample length is $T = 100$, break date location is $k_0 = 50$, error standard deviation is $\sigma_\varepsilon = 1$, the number of stochastic covariates is $q = 0$, the coefficient value before the break is $\beta_{j,pre} = 1$, and CVE settings are $P = 5$ and $C = 30$. There are $S = 10000$ simulations. The values are calculated for simulations in which both methods have detected a break at 95% confidence, only.

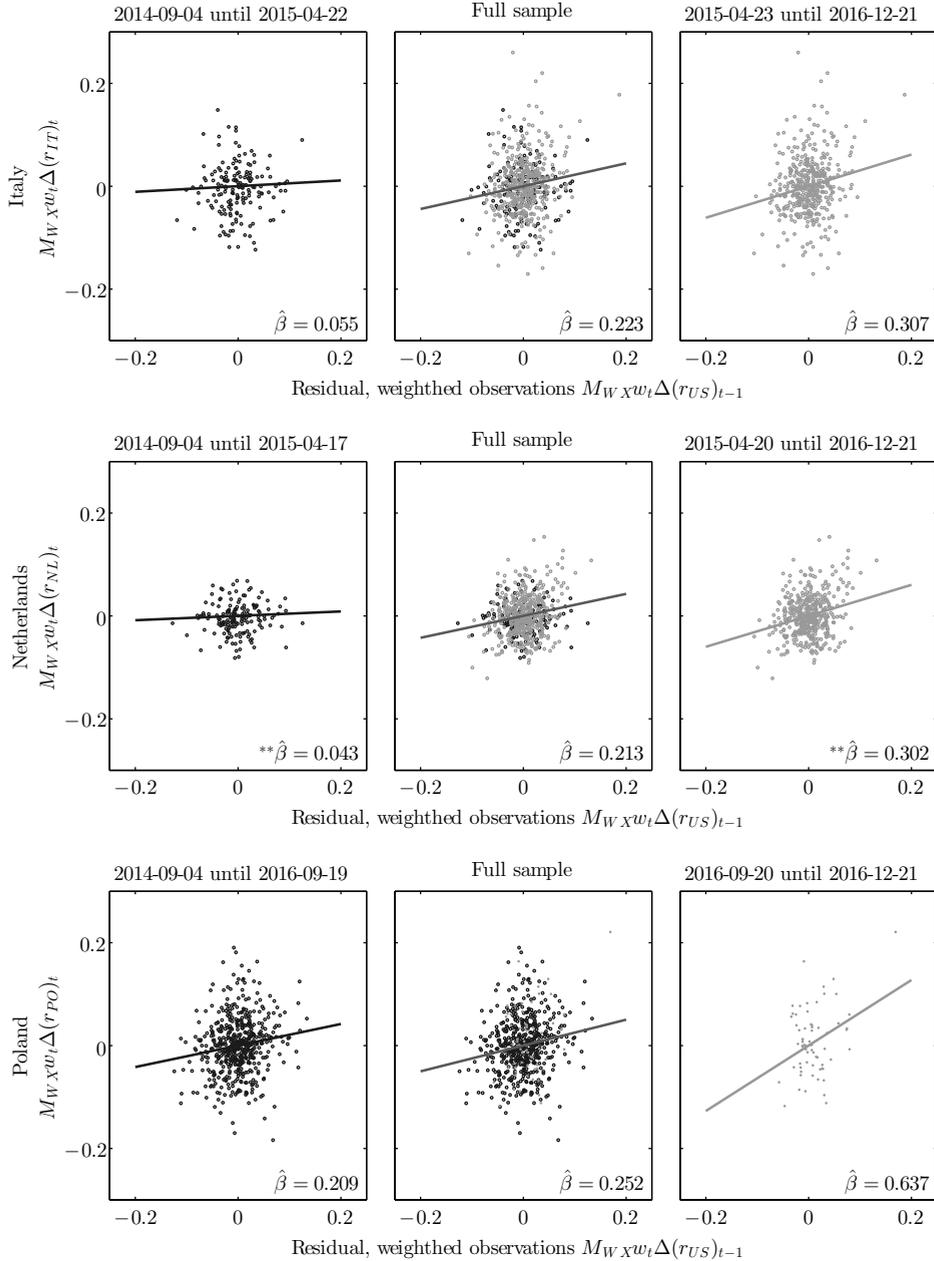
Figure 6.5: Linear regressions for pre-, no-, and post-break data, part 3/6

Note: For $i = BD, BG, CZ$, scatter plots for residual, weighted observations $M_{WX}w_t\Delta(r_{US})_{t-1}$ and $M_{WX}w_t\Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

Figure 6.6: Linear regressions for pre-, no-, and post-break data, part 4/6

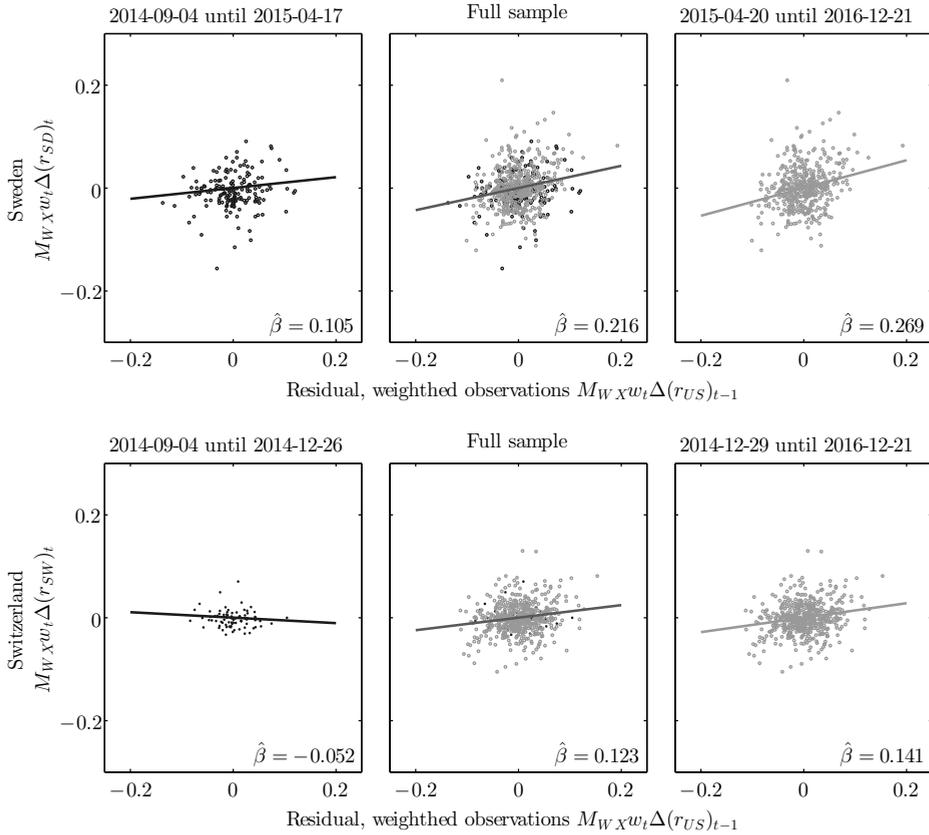


Note: For $i = ES, FN, FR$, scatter plots for residual, weighted observations $M_{WX}w_t\Delta(r_{US})_{t-1}$ and $M_{WX}w_t\Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

Figure 6.7: Linear regressions for pre-, no-, and post-break data, part 5/6

Note: For $i = IT, NL, PO$, scatter plots for residual, weighted observations $M_{WX}w_t\Delta(r_{US})_{t-1}$ and $M_{WX}w_t\Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

Figure 6.8: Linear regressions for pre-, no-, and post-break data, part 6/6



Note: For $i = SD, SW$, scatter plots for residual, weighed observations $M_{WX}w_t\Delta(r_{US})_{t-1}$ and $M_{WX}w_t\Delta(r_i)_t$. Left panels show observations until the respective estimated CVE break date and include estimated linear regression lines. Right panels show the same for data after the break dates. Central panels combine the two samples and show the corresponding, estimated linear regression line. Significance of a break is shown by *, **, ***, which are used to mark 90%, 95%, 99% significance, respectively.

7

Summary

Structural break analysis is used in Chapter 3 to trace evidence of shift-contagion in financial time series data. A sequential testing procedure is described, which uses the established maximum likelihood ratio test to detect multiple structural breaks in a multivariate data system. In a linear regression model of financial returns, shift-contagion is expected to involve a significant increase in the contemporaneous correlation between the residuals of the series. Since contagion is associated with times of crisis, volatility changes and changes in other parameters of the model can also be found by the proposed testing procedure. Applied to yield spreads of euro area governments, a large number of structural breaks are found during the period of the European sovereign debt crisis. The estimated break dates for correlation changes are diverse. Notably, volatilities change at different times than correlations, as contagion is not bound to happen at the beginning of crisis periods. The empirical study is augmented by the calculation of spillover index values to measure spillover-contagion. The level of spillover is particularly high during 2011 and 2012, and periphery countries transmit a large part of their shocks to the rest of Europe. Overall, the findings demonstrate a substantial degree of heterogeneity between euro area countries. Yet, the dynamics of European shock transmissions are shown to be rich, and both shift-contagion and spillover-contagion have occurred during the sampled period of time.

The usage of maximum likelihood ratio testing to detect multiple structural breaks in large, multivariate data sets has proven to be time consuming. This is particularly true, if various possible constellations of common structural break hypotheses are tested, and if simulation-based bootstrap methods are used to either determine critical test values empirically or to find confidence intervals for structural break date estimates. The execution time of the applied dynamic program increases quadratically with the number of observation dates. Chapter 4 presents a rolling window algorithm that can be used as a break detection heuristic. The algorithm calculates single break likelihood

ratio values in rolling data windows, which reduces the execution time substantially. A Monte Carlo simulation study shows that the rolling window algorithm has similar testing power as the dynamic program in large data samples. In simulations where parameters shift in the same direction several consecutive time, the rolling window algorithm is even shown to locate the breaks more efficiently than the dynamic program. When applied to find error volatility changes in a regression model of European central government bond yield differences between 2003 and 2016, six structural breaks are detected for each series. The chapter offers interpretation for the detected changes as they relate to the Global Financial Crisis, the European sovereign debt crisis and policy measures by the European Central Bank.

In structural break testing, when a maximum likelihood ratio statistic is used to find change, this change is captured by model parameters. That is, the structural break is inferred, if the data fit of a model is significantly improved by allowing two different sets of parameter values (before and after the structural break date). Finding a structural break means that allowing abrupt change at the break date enables a better representation of the sampled observations than not allowing change at all, but the finding does not prove that the change is actually abrupt. When interpreting empirical structural break results, such as the euro area contagion findings of Chapter 3 and the European yield volatility changes of Chapter 4, it is important to recognize this uncertainty about the type of transition. Chapter 5 further investigates the volatility changes of Chapter 4 by applying a novel abruptness test of parameter change. The abruptness test consists of the summation of likelihood ratios that lie within a specified range of the maximum likelihood ratio. In accordance with the specification under structural break model, the test assumes that the parameter change is abrupt under its null hypothesis. If the sum of likelihood ratios is larger than typically observed, the abruptness hypothesis is rejected in favor of a hypothesis with more gradual change. A Monte Carlo study underlines the validity of this test design for data that originates from smooth transition models. In the empirical application of European yields, many volatility transitions are found to be non-abrupt, especially during the European sovereign debt crisis episodes. The changes observed in periphery countries appear to accumulate gradually over time.

So far, all chapters have used the maximum likelihood ratio statistic, which is designed to find the a structural break that optimizes the model fit to the given observations. Chapter 6 advances structural break testing by introducing a cross-validated error estimation test. Here, structural breaks are found that optimize the estimated model fit to unobserved data. The test is concerned with the out-of-sample validity of the inferred model, a concept that is standard practice in statistical learning and feature selection. In a Monte Carlo study, the performance of the proposed test is systematically compared to the performance of the maximum likelihood ratio test. Either method frequently detects break dates closer to the true location in the simulated data than the other method. The new test, however, does so more frequently than the maximum likelihood ratio. For the test, prediction errors are estimated by cross-validated mean squared errors, repeatedly, to reveal the estimation uncertainty. The collection of estimated break dates can serve as the foundation of a rich structural break analysis, as shown in the empirical application of the chapter. Here, it is studied whether the association between lagged U.S. government bond yields and European government bond yields has changed between 2014 and 2016. Following the adoption of a bond purchase program by the European Central

Bank that is similar to the U.S. quantitative easing program, the association has become significantly larger after April 2015 for six euro area core countries as well as Denmark.

Bibliography

- Abad, P., Chuliá, H., Gómez-Puig, M., 2014. Time-varying integration in European government bond markets. *European Financial Management* 20 (2), 270–290.
- Altavilla, C., Giannone, D., 2017. The effectiveness of non-standard monetary policy measures: Evidence from survey data. *Journal of Applied Econometrics* 32, 952–964.
- Andrews, D. W. K., 1993. Tests for parameter instability and structural change with unknown change point. *Econometrica* 61 (4), 821–856.
- Argyrou, M. G., Kontonikas, A., 2012. The EMU sovereign-debt crisis: Fundamentals, expectations and contagion. *Journal of International Financial Markets, Institutions and Money* 22 (4), 658–677.
- Arlot, S., Celisse, A., et al., 2010. A survey of cross-validation procedures for model selection. *Statistics surveys* 4, 40–79.
- Attinasi, M.-G., Checherita-Westphal, C., Nickel, C., 2009. What explains the surge in euro area sovereign spreads during the financial crisis of 2007-09? EBC Working Paper (1131).
- Bai, J., Lumsdaine, R., Stock, J., 1998. Testing for and dating common breaks in multivariate time series. *The Review of Economic Studies* 65 (3), 395–432.
- Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. *Econometrica* 66 (1), 47–78.
- Bai, J., Perron, P., 2003. Computation and analysis of multiple structural change models. *Journal of Applied Econometrics* 18 (1), 1–22.
- Baig, T., Goldfajn, I., 1999. Financial market contagion in the Asian crisis. *IMF staff papers* 46 (2), 167–195.

BIBLIOGRAPHY

- Barrios, S., Iversen, P., Lewandowska, M., Setzer, R., 2009. Determinants of intra-euro area government bond spreads during the financial crisis. *European Economy Economic Papers* (388).
- Bartram, S. M., Wang, Y.-H., 2015. European financial market dependence: An industry analysis. *Journal of Banking & Finance* 59, 146–163.
- Bataa, E., Osborn, D. R., Sensier, M., van Dijk, D., 2013. Structural breaks in the international dynamics of inflation. *Review of Economics and Statistics* 95 (2), 646–659.
- Baur, D., Lucey, B., 2009. Flights and contagion — An empirical analysis of stock-bond correlations. *Journal of Financial stability* 5 (4), 339–352.
- Beber, A., Brandt, M. W., Kavajecz, K. A., 2008. Flight-to-quality or flight-to-liquidity? Evidence from the euro-area bond market. *The Review of Financial Studies* 22 (3), 925–957.
- Beirne, J., Fratzscher, M., 2013. The pricing of sovereign risk and contagion during the European sovereign debt crisis. *Journal of International Money and Finance* 34, 60–82.
- Bekaert, G., Harvey, C. R., Lundblad, C. T., Siegel, S., 2017. Economic and financial integration in Europe 1. *DICE Report* 15 (1), 36.
- Berg, A., Ostry, J. D., Zettelmeyer, J., 2012. What makes growth sustained? *Journal of Development Economics* 98 (2), 149–166.
- Bernoth, K., Erdogan, B., 2012. Sovereign bond yield spreads: A time-varying coefficient approach. *Journal of International Money and Finance* 31 (3), 639–656.
- Bickel, P. J., Freedman, D. A., 1981. Some asymptotic theory for the bootstrap. *The Annals of Statistics* 9 (6), 1196–1217.
- Blatt, D., Candelon, B., Manner, H., 2015. Detecting contagion in a multivariate time series system: An application to sovereign bond markets in Europe. *Journal of Banking & Finance* 59, 1–13.
- Boot, T., Pick, A., 2017. A near optimal test for structural breaks when forecasting under square error loss. *Tinbergen Institute Discussion Paper* TI 2017-039/III.
- Borio, C., Disyatat, P., 2010. Unconventional monetary policies: An appraisal. *The Manchester School* 78 (s1), 53–89.

- Brockwell, P. J., Davis, R. A., 2016. Introduction to time series and forecasting. Springer.
- Brown, R., Durbin, J., Evans, J., 1975. Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B (Methodological)*, 149–192.
- Calvo, G., Mendoza, E., 2000. Rational contagion and the globalization of securities markets. *Journal of International Economics* 51 (1), 79–113.
- Candelon, B., Bicu, A., 2013. On the importance of indirect bank linkages in Europe. *Journal of Banking & Finance* 37, 5007–5024.
- Candelon, B., Manner, H., 2010. Testing for asset market linkages: A new approach based on time-varying copulas. *Pacific Economic Review* 15 (3), 364–384.
- Caporin, M., Pelizzon, L., Ravazzolo, F., Rigobon, R., 2013. Measuring sovereign contagion in Europe. Tech. rep., National Bureau of Economic Research.
- Cappiello, L., Engle, R. F., Sheppard, K., 2006. Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial econometrics* 4 (4), 537–572.
- Carsoule, F., Franses, P. H., 2003. A note on monitoring time-varying parameters in an autoregression. *Metrika* 57 (1), 51–62.
- Chan, K. C., Karolyi, G. A., Longstaff, F. A., Sanders, A. B., 1992. An empirical comparison of alternative models of the short-term interest rate. *The journal of finance* 47 (3), 1209–1227.
- Chen, B., Hong, Y., 2012. Testing for smooth structural changes in time series models via nonparametric regression. *Econometrica* 80 (3), 1157–1183.
- Chinn, M. D., Eichengreen, B., Ito, H., 2014. A forensic analysis of global imbalances. *Oxford Economic Papers* 66 (2), 465–490.
- Chow, G., 1960. Tests of equality between sets of coefficients in two linear regressions. *Econometrica: Journal of the Econometric Society*, 591–605.
- Chu, C.-K., Marron, J. S., 1991. Comparison of two bandwidth selectors with dependent errors. *The Annals of Statistics*, 1906–1918.
- Chu, C.-S. J., Stinchcombe, M., White, H., 1996. Monitoring structural change. *Econometrica: Journal of the Econometric Society*, 1045–1065.

BIBLIOGRAPHY

- Claeys, G., Leandro, Á., Mandra, A., 2015. European central bank quantitative easing: The detailed manual. Tech. rep., Bruegel Policy Contribution.
- Claeys, P., Vašíček, B., 2014. Measuring bilateral spillover and testing contagion on sovereign bond markets in Europe. *Journal of Banking & Finance* 46 (0), 151 – 165.
- Clark, T. E., McCracken, M. W., 2001. Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics* 105 (1), 85–110.
- Clark, T. E., McCracken, M. W., 2009. Improving forecast accuracy by combining recursive and rolling forecasts. *International Economic Review* 50 (2), 363–395.
- Codogno, L., Favero, C., Missale, A., 2003. Yield spreads on EMU government bonds. *Economic Policy* 18 (37), 503–532.
- Copelovitch, M., Frieden, J., Walter, S., 2016. The political economy of the euro crisis. *Comparative Political Studies* 49 (7), 811–840.
- Davidson, R., MacKinnon, J. G., 2004. *Econometric theory and methods*. Vol. 5. Oxford University Press New York.
- De Santis, R. A., 2012. The euro area sovereign debt crisis: Safe haven, credit rating agencies and the spread of the fever from Greece, Ireland and Portugal. Tech. rep., European Central Bank.
- Di Cesare, A., Grande, G., Manna, M., Taboga, M., 2012. Recent estimates of sovereign risk premia for euro-area countries. Tech. rep., Bank of Italy, Economic Research and International Relations Area.
- DiCiccio, T. J., Efron, B., 1996. Bootstrap confidence intervals. *Statistical science*, 189–212.
- Diebold, F. X., Yilmaz, K., 2009. Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal* 119 (534), 158–171.
- Diebold, F. X., Yilmaz, K., 2012. Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting* 28 (1), 57–66.
- Dungey, M., Fry, R., Martin, V., 2004. Currency market contagion in the Asia-Pacific region. *Australian Economic Papers* 43 (4), 379–395.

- Efron, B., 1979. Bootstrap methods: Another look at the jackknife. *The annals of Statistics*, 1–26.
- Efron, B., Tibshirani, R. J., 1994. *An introduction to the bootstrap*. CRC press.
- Ehrmann, M., Fratzscher, M., 2005. Equal size, equal role? Interest rate interdependence between the euro area and the United States. *The Economic Journal* 115 (506), 928–948.
- Ehrmann, M., Fratzscher, M., 2017. Euro area government bonds—fragmentation and contagion during the sovereign debt crisis. *Journal of International Money and Finance* 70, 26–44.
- Eichengreen, B., Rose, A. K., Wyplosz, C., 1996. Contagious currency crises. Tech. rep., National Bureau of Economic Research.
- Eichengreen, B., Rose, A. K., Wyplosz, C., Dumas, B., Weber, A., 1995. Exchange market mayhem: The antecedents and aftermath of speculative attacks. *Economic Policy* 10 (2), 249–312.
- Elliott, G., Müller, U. K., 2014. Pre and post break parameter inference. *Journal of Econometrics* 180 (2), 141–157.
- Engle, R. F., 1982. A general approach to Lagrange multiplier model diagnostics. *Journal of Econometrics* 20 (1), 83–104.
- Ericsson, J., Jacobs, K., Oviedo, R., 2009. The determinants of credit default swap premia. *Journal of Financial and Quantitative Analysis* 44 (1), 109–132.
- Eser, F., Schwaab, B., 2016. Evaluating the impact of unconventional monetary policy measures: Empirical evidence from the ECB’s Securities Markets Programme. *Journal of Financial Economics* 119 (1), 147–167.
- EUobserver, 2009. Greek deficit ‘endangers’ euro, EU commission says. <https://euobserver.com/economic/29328>, accessed 2015-01-16.
- Favero, C., Pagano, M., Von Thadden, E.-L., 2010. How does liquidity affect government bond yields? *Journal of financial and quantitative analysis* 45 (1), 107–134.
- Favero, C. A., Giavazzi, F., 2002. Is the international propagation of financial shocks non-linear?: Evidence from the ERM. *Journal of International Economics* 57 (1), 231–246.

BIBLIOGRAPHY

- Forbes, K., Rigobon, R., 2002. No contagion, only interdependence: Measuring stock market comovements. *The Journal of Finance* 57 (5), 2223–2261.
- Frieden, J., Walter, S., 2017. Understanding the political economy of the Eurozone Crisis. *Annual Review of Political Science* 20.
- Gagnon, J., Raskin, M., Remache, J., Sack, B., et al., 2011. The financial market effects of the Federal Reserve’s large-scale asset purchases. *International Journal of Central Banking* 7 (1), 3–43.
- Geyer, A., Kossmeier, S., Pichler, S., 2004. Measuring systematic risk in EMU government yield spreads. *Review of Finance* 8 (2), 171–197.
- Giordano, R., Pericoli, M., Tommasino, P., 2013. Pure or wake-up-call contagion? Another look at the EMU sovereign debt crisis. *International Finance* 16 (2), 131–160.
- Goldfeld, S., Quandt, R., 1965. Some tests for homoscedasticity. *Journal of the American Statistical Association*, 539–547.
- Gombay, E., Serban, D., 2009. Monitoring parameter change in time series models. *Journal of Multivariate Analysis* 100 (4), 715–725.
- Gómez-Puig, M., Sosvilla-Rivero, S., 2014. Causality and contagion in EMU sovereign debt markets. *International Review of Economics & Finance* 33, 12–27.
- Goodfellow, I., Bengio, Y., Courville, A., 2016. *Deep learning*. MIT press, <http://www.deeplearningbook.org>.
- Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., Altman, D. G., 2016. Statistical tests, p values, confidence intervals, and power: a guide to misinterpretations. *European journal of epidemiology* 31 (4), 337–350.
- Groen, J., Kapetanios, G., Price, S., 2011. Multivariate methods for monitoring structural change. *Journal of Applied Econometrics* 28 (2), 250–274.
- Gruppe, M., Lange, C., 2014. Spain and the European sovereign debt crisis. *European Journal of Political Economy* 34, S3–S8.
- Guthery, S. B., 1974. Partition regression. *Journal of the American Statistical Association* 69 (348), 945–947.
- Hall, P., Wilson, S. R., 1991. Two guidelines for bootstrap hypothesis testing. *Biometrics*, 757–762.

- Hansen, B. E., 1992. Testing for parameter instability in linear models. *Journal of policy Modeling* 14 (4), 517–533.
- Hanson, S. G., Stein, J. C., 2015. Monetary policy and long-term real rates. *Journal of Financial Economics* 115 (3), 429–448.
- Hartley, H. O., 1950. The maximum F-ratio as a short-cut test for heterogeneity of variance. *Biometrika* 37 (3/4), 308–312.
- Hartmann, P., Maddaloni, A., Manganelli, S., 2003. The euro-area financial system: Structure, integration, and policy initiatives. *Oxford Review of Economic Policy* 19 (1), 180–213.
- Hastie, T., Tibshirani, R., Friedman, J., 2009. Overview of supervised learning. In: *The elements of statistical learning*. Springer, pp. 9–41.
- Hawkins, D., 1987. A test for a change point in a parametric model based on a maximal Wald-type statistic. *Sankhyā: The Indian Journal of Statistics, Series A*, 368–376.
- Hinkley, D., 1970. Inference about the change-point in a sequence of random variables. *Biometrika* 57 (1), 1–17.
- Hlávka, Z., Hušková, M., Kirch, C., Meintanis, S., 2016. Bootstrap procedures for online monitoring of changes in autoregressive models. *Communications in Statistics-Simulation and Computation* 45 (7), 2471–2490.
- Hocking, R. R., 1976. A Biometrics invited paper. The analysis and selection of variables in linear regression. *Biometrics* 32 (1), 1–49.
- Hušková, M., Kirch, C., 2012. Bootstrapping sequential change-point tests for linear regression. *Metrika* 75 (5), 673–708.
- Inclan, C., Tiao, G. C., 1994. Use of cumulative sums of squares for retrospective detection of changes of variance. *Journal of the American Statistical Association* 89 (427), 913–923.
- Jackson, B., Scargle, J. D., Barnes, D., Arabhi, S., Alt, A., Gioumousis, P., Gwin, E., Sangtrakulcharoen, P., Tan, L., Tsai, T. T., 2005. An algorithm for optimal partitioning of data on an interval. *IEEE Signal Processing Letters* 12 (2), 105–108.
- James, G., Witten, D., Hastie, T., Tibshirani, R., 2013. *An introduction to statistical learning*. Vol. 112. Springer.

BIBLIOGRAPHY

- Jiang, Y., Song, Z., Kusiak, A., 2013. Very short-term wind speed forecasting with Bayesian structural break model. *Renewable energy* 50, 637–647.
- Johansen, S., Mosconi, R., Nielsen, B., 2000. Cointegration analysis in the presence of structural breaks in the deterministic trend. *The Econometrics Journal* 3 (2), 216–249.
- Jones, E., Kelemen, R. D., Meunier, S., 2016. Failing forward? The euro crisis and the incomplete nature of European integration. *Comparative Political Studies* 49 (7), 1010–1034.
- Joyce, M., Lasasosa, A., Stevens, I., Tong, M., et al., 2011. The financial market impact of quantitative easing in the United Kingdom. *International Journal of Central Banking* 7 (3), 113–161.
- Jung, R., Maderitsch, R., 2014. Structural breaks in volatility spillovers between international financial markets: Contagion or mere interdependence? *Journal of Banking & Finance* 47, 331–342.
- Kahneman, D., 2011. *Thinking, Fast and Slow*. Macmillan.
- Kapetanios, G., Mumtaz, H., Stevens, I., Theodoridis, K., 2012. Assessing the economy-wide effects of quantitative easing. *The Economic Journal* 122 (564).
- Killick, R., Fearnhead, P., Eckley, I. A., 2012. Optimal detection of changepoints with a linear computational cost. *Journal of the American Statistical Association* 107 (500), 1590–1598.
- King, M., Wadhvani, S., 1990. Transmission of volatility between stock markets. *Review of Financial Studies* 3 (1), 5–33.
- Kirch, C., 2008. Bootstrapping sequential change-point tests. *Sequential Analysis* 27 (3), 330–349.
- Kleimeier, S., Sander, H., 2000. Regionalisation versus globalisation in European financial market integration: Evidence from co-integration analyses. *Journal of banking & finance* 24 (6), 1005–1043.
- Koop, G., Pesaran, M. H., Potter, S. M., 1996. Impulse response analysis in non-linear multivariate models. *Journal of Econometrics* (74), 119–147.
- Krishnamurthy, A., Nagel, S., Vissing-Jorgensen, A., 2014. ECB policies involving government bond purchases: Impact and channels. Unpublished working paper.

- Kroese, D. P., Brereton, T., Taimre, T., Botev, Z. I., 2014. Why the Monte Carlo method is so important today. *Wiley Interdisciplinary Reviews: Computational Statistics* 6 (6), 386–392.
- Kunze, F., Gruppe, M., 2014. Performance of survey forecasts by professional analysts: did the European debt crisis make it harder or perhaps even easier? *Social Sciences* 3 (1), 128–139.
- Lahiri, S. N., et al., 1999. Theoretical comparisons of block bootstrap methods. *The Annals of Statistics* 27 (1), 386–404.
- Lai, T. L., 1995. Sequential changepoint detection in quality control and dynamical systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, 613–658.
- Lee, S., Ha, J., Na, O., Na, S., 2003. The CUSUM test for parameter change in time series models. *Scandinavian Journal of Statistics* 30 (4), 781–796.
- Lin, C.-F. J., Teräsvirta, T., 1994. Testing the constancy of regression parameters against continuous structural change. *Journal of Econometrics* 62 (2), 211–228.
- Lubbers, M., Jaspers, E., 2011. A longitudinal study of euroscepticism in the Netherlands: 2008 versus 1990. *European Union Politics*, 1465116510390062.
- Lütkepohl, H., Krätzig, M., 2004. *Applied time series econometrics*. Cambridge University Press.
- Maltritz, D., 2012. Determinants of sovereign yield spreads in the Eurozone: A Bayesian approach. *Journal of International Money and Finance* 31 (3), 657–672.
- Manganelli, S., Wolswijk, G., 2009. What drives spreads in the euro area government bond market? *Economic Policy* 24 (58), 191–240.
- McConnell, M. M., Perez-Quiros, G., 1998. Output fluctuations in the United States: what has changed since the early 1980s? *FRB of New York Staff Report* (41).
- Metiu, N., 2012. Sovereign risk contagion in the Eurozone. *Economics Letters* 117 (1), 35–38.
- Missio, S., Watzka, S., 2011. Financial contagion and the European debt crisis. CESifo, Munich, Germany. Working Paper 3554.
- Mohl, P., Sondermann, D., 2013. Has political communication during the crisis impacted sovereign bond spreads in the euro area? *Applied Economics Letters* 20 (1), 48–61.

BIBLIOGRAPHY

- Monfort, A., Renne, J.-P., 2013. Decomposing euro-area sovereign spreads: credit and liquidity risks. *Review of Finance* 18 (6), 2103–2151.
- Nadeau, C., Bengio, Y., 2000. Inference for the generalization error. In: *Advances in neural information processing systems*. pp. 307–313.
- Neely, C. J., Rapach, D. E., 2011. International comovements in inflation rates and country characteristics. *Journal of International Money and Finance* 30 (7), 1471–1490.
- Newey, W., West, K., 1987. A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55 (3), 703–708.
- Opsomer, J., Wang, Y., Yang, Y., 2001. Nonparametric regression with correlated errors. *Statistical Science*, 134–153.
- Page, E., 1954. Continuous inspection schemes. *Biometrika* 41 (1/2), 100–115.
- Page, E., 1955. A test for a change in a parameter occurring at an unknown point. *Biometrika* 42 (3/4), 523–527.
- Palm, F. C., Smeekes, S., Urbain, J.-P., 2008. Bootstrap unit-root tests: Comparison and extensions. *Journal of Time Series Analysis* 29 (2), 371–401.
- Paye, B. S., Timmermann, A., 2006. Instability of return prediction models. *Journal of Empirical Finance* 13 (3), 274–315.
- Perron, P., Oka, T., 2011. Testing for common breaks in a multiple equations system. Tech. rep., Boston University-Department of Economics.
- Pesaran, M. H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. *Economics Letters* (58), 17–29.
- Pesaran, M. H., Timmermann, A., 2007. Selection of estimation window in the presence of breaks. *Journal of Econometrics* 137 (1), 134–161.
- Ploberger, W., Krämer, W., Kontrus, K., 1989. A new test for structural stability in the linear regression model. *Journal of Econometrics* 40 (2), 307–318.
- Politis, D. N., 2003. The impact of bootstrap methods on time series analysis. *Statistical Science*, 219–230.
- Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. *Econometrica* 75 (2), 459–502.

- Quenouille, M. H., 1949. Problems in plane sampling. *The Annals of Mathematical Statistics*, 355–375.
- Sibbertsen, P., Wegener, C., Basse, T., 2014. Testing for a break in the persistence in yield spreads of EMU government bonds. *Journal of Banking & Finance* 41, 109–118.
- Startin, N., Krouwel, A., 2013. Euroscepticism re-galvanized: The consequences of the 2005 French and Dutch rejections of the EU Constitution. *JCMS: Journal of Common Market Studies* 51 (1), 65–84.
- Steland, A., 2006. A bootstrap view on Dickey-Fuller control charts for AR (1) series. *Austrian Journal of Statistics* 35 (2&3), 339–346.
- Stock, J. H., Watson, M. W., 1996. Evidence on structural instability in macroeconomic time series relations. *Journal of Business & Economic Statistics* 14 (1), 11–30.
- Stock, J. H., Watson, M. W., 2003. Forecasting output and inflation: The role of asset prices. *Journal of Economic Literature* 41 (3), 788–829.
- Teräsvirta, T., 1994. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association* 89 (425), 208–218.
- The New York Times, 2009. Growing economic crisis threatens the idea of one Europe. <http://www.nytimes.com/2009/03/02/world/europe/02euro.html>, accessed 2015-01-17.
- Tonzer, L., Buchholz, M., 2014. Sovereign credit risk co-movements in the Eurozone: Simple interdependence or contagion? ZBW-Deutsche Zentralbibliothek für Wirtschaftswissenschaften, Leibniz-Informationszentrum Wirtschaft.
- Tukey, J. W., 1958. Bias and confidence in not-quite large samples. *Ann. Math. Statist.* 29, 614.
- Van de Wiel, M. A., Berkhof, J., van Wieringen, W. N., 2009. Testing the prediction error difference between 2 predictors. *Biostatistics* 10 (3), 550–560.
- Von Borstel, J., Eickmeier, S., Krippner, L., 2016. The interest rate pass-through in the euro area during the sovereign debt crisis. *Journal of International Money and Finance* 68, 386–402.

BIBLIOGRAPHY

Valorisation addendum

When investigating motives and patterns of human action and interaction, statistics are an important source of information. Because they are important, statistical results then steer human decision making, both implicitly and explicitly. Statistics are therefore deeply ingrained in human behavior as an interactive component.

Understandably, people are fascinated by the patterns of their own behavior and are determined to understand them. Scientists investigate the configurations of human interaction tirelessly. Our general curiosity for the topic can also explain the popularity of such books as *Thinking, Fast and Slow* by Kahneman (2011). We recognize the importance of being aware of rules and implications of statistics, of probabilities and heuristics. We acknowledge that many conclusions are better drawn having observed a lot of data and looking for significant effects, by *Thinking, Slow*. Statisticians are guides for learning from data and navigating common problems. The scientific contributions to the field of statistics can help us better understand ourselves and each other, and help us make decisions more carefully.

Structural break testing is one explicit form of statistical information processing. It is a form of model diagnostics. Models are used to reduce the complexity of an interaction between variables, between human actions or other forces. We observe patterns and may be able to model a process that can explain and predict what we see within limited margins of error. But do we have reason to assume that the same process holds for a full range of observations, or must we admit that there has been a change at some point? Are the patterns that we observe the outcome of a stable model or do we have to consider a structural break in the process?

The error of any model can be reduced by incorporating a structural break, as it increases the adaptability of the model to the given data. If a structural break reduces the model error significantly, the break itself can be interpreted, but the separation of data at the break also ensures that prediction of additional data is based on an appropriately estimated model. Therefore, structural break testing can serve a wide range of purposes in unlimited application areas.

Often, as in this book, structural break tests are applied to time series to find changes in macroeconomic relations, between financial series or other records of human behavior such as administrative records. Evidently, periods of financial and economic crisis such as the Global Financial Crisis and the European sovereign debt crisis impact our lives drastically. All chapters of this book illustrate how structural break tests can contribute to the evaluation of such events and thus help find appropriate policy responses. Structural break tests are, however, much more widely applicable, if variables are ordered. They can, for example, be used to find a significant change in a health indicator for a group of people sorted by age or another indicator. Another relevant function of structural break

testing is studying geospatial data, to find pattern changes in satellite images. One could also imagine a structural break test being used to detect a significant composition change in a probe of ice taken from a glacier. As soon as time is added as a dimension again, the application possibilities multiply. A test could be used to find a significant change in the area fluctuations of an ice sheet. The recording of a structural break for light sensor data would help identify movement of objects in the detection field of the sensor. Therefore, any advancement of methods used in structural break testing can have a substantial impact just by the large number of application possibilities.

Applied statistics in general can augment and improve the analysis of observations we make. Modern technology allows us to conveniently collect and track a large amount of data about ourselves, from calorie intake to time spent clicking a computer mouse and recording the locations we have been to. Structural break tests, specifically, can be employed to find changes of our behavior in these vast sets of data. Apps are created with this goal in mind. Computers enable us to reap the benefits of statistical analysis in a quick and convenient way. But a confident analysis of the suggested results requires some insight into the statistical methods being used. An understanding of structural break testing could help people make more sense of their behavior and how it has possibly changed.

It must be discussed, whether applicants of structural break testing programs understand what it means to have a significant structural break result. Structural break tests have become somewhat standard practice, a development that has led to and has been accelerated by the inclusion of structural break testing methods in popular econometric software such as Eviews and in open source distributions such as R. As with other statistical tests, the convenience of having a scalar test result with an associated p -value can also pose a thread. First, the p -value is not to be interpreted as the probability of the structural break hypothesis to be true. Secondly, structural break tests typically compare an abrupt change in the model to the no-change assumption, but they have power against much more general forms of change and instability. When a structural break is detected, all that is shown is that the assumed model without break is unlikely to have produced the observed data, but nothing is said about the base probabilities of no-break and break models and nothing is proven about the way in which the parameters of the model may have changed.

Given a positive structural break test result, the sum of likelihood ratio tests offers a way to investigate the following hypothesis that parameters of the model have changed abruptly at the estimated structural break date. The test incorporates more general forms of change as potential sources of the significance of a structural break hypothesis. This book's Chapter 5 has been written to describe the test and its concept. But further, the chapter is also intended to highlight and discuss the limited implications of a significant structural break test result in practice.

The descriptions of the tests of this book do not rely on a large amount of formal descriptions and derivations. As a relatively formal science, statistics is not a very accessible field of study for many people. When presenting the bootstrap, Efron and Tibshirani (1994) comment that the road to statistical knowledge is often blocked by a wall of mathematics. The bootstrap, a computer-based method of statistical inference, requires few formulas but answers real statistical questions. Generally, simulation based research may lower the barrier between statistics and the people interested in it.

Simulation based research also conforms to our relatively straightforward understanding of controlled experiments, and how the frequency of an observation can indicate the relevance of a pattern.

Not adding theoretical proof, when showing the validity of the discussed concepts of this book empirically, is an unusual choice in this area of research. Regardless, the concepts of this book have been explored and researched carefully and are worth sharing and discussing and exploring further. They are shown to be relevant, by simulation based means. The book is focused on conveying ideas in a illustrative and convincing way rather than their mathematical justification, similarly to the argument in Efron and Tibshirani. Ideally, the presentation will broaden the interest in the proposed concepts by making them accessible to a larger group of people.

Practicality and accessibility are central themes in this book. The intention of the sum of likelihood ratios of Chapter 5 is to make a more informed statement about the kind of change that has triggered a positive structural break test result. The new test uses data that is already available during structural break testing to test the next hypothesis: Can the change that has been discovered be assumed to have happened abruptly, or is there evidence against this assumption? Simulated or resampled observations that fulfill the abruptness assumption are used to conclude. The setup of the test is therefore very pragmatic and relatively straightforward.

Similarly, the rolling window algorithm of Chapter 4 has been developed to provide the benefit of multiple structural break testing at low computation cost. Time efficiency can be a concern in structural break testing, especially when data sets are large and when results must be presented quickly. The proposed algorithm offers a solution in such circumstances. Tradeoffs, however, have to be accepted when a heuristic such as the rolling window algorithm is used, as the resulting test is conservative and may only find a locally optimal result.

Diversity in research is a central principle of doctoral programmes. The methods of this book connect the topic of structural break testing to a range of other areas of science. They seek to incorporate insight from areas of science other than econometrics or applied statistics to enrich the literature on structural break testing. The idea to describe the rolling window algorithm of Chapter 4 has been developed when studying search algorithms, a topic of computer science. The heuristic can be discussed from the perspective of optimization theory. The need to provide an initial guess in several optimization techniques has inspired the combined method presented in the chapter, where the rolling window estimates serve as starting points for finding the global optimum. A feature selection technique has been adopted in Chapter 6, where cross-validated error estimates are used to find a structural break. Here, observations are excluded from the sample to test the existence of a structural break, a concept typically used in Machine Learning when building neural networks. Finally, topics of interest for macroeconomics and finance are discussed in all empirical examples of the presented tests, when studying the European sovereign debt crisis.

Parts of this book offer insight into a range of important aspects of structural break testing in general. The sequential testing procedure of Chapter 3 is designed specifically to capture financial contagion events in time series data. As such, it is developed to detect structural breaks that involve different types of parameters, which can pose problems of nuisance if not properly addressed. A solution that is incorporated in the se-

quential testing procedure is iterative testing, where break testing is applied conditional on previous results, and repeated until convergence of the estimates. A common break test hypothesis is included in the sequential testing procedure to account for estimation uncertainty and the possibility of joint structural breaks. Uncertainty of break estimation results is also explicitly addressed in other parts of the book. The cross-validated error estimation method of Chapter 6 uses random resampling of the observed data and thus directly incorporates a method of demonstrating the estimation uncertainty. Finally, since structural break detection test statistics typically have non-standard distributions, the book can also be used as a guide to implementing bootstrap empirical resampling methods to derive critical values of testing.

Jointly, the methods of this book can be combined into a powerful and comprehensive multiple structural break testing procedure. Cross-validated error estimation would be the central test statistic, a modern detection strategy that incorporates estimation uncertainty. The statistic would be calculated in a sequential, iterative procedure to account for nuisance, if several types of parameters had to be tested for structural breaks individually. Optionally, rolling windows of data could be used, if time-efficiency is a concern, possibly to derive an initial guess of a globally optimal solution. In the end, a summation of test statistic outcomes within specified ranges would offer additional insight into the transition duration of detected changes in the model. The product could be a package or library that is made available to a broad community of researchers using R syntax, and python scripts could be shared on open source platforms such as github.

The field of statistics can be compelling to a student of economics because its concepts defy time. Economic analysis is an exciting subject to a large degree because it depends on current definitions of economic concepts and on current understanding of economic value. This, however, makes results and conclusions of economic research somewhat fragile, and dependent on context. The methods of this book are invariant to context. The presented tests are shown to have certain properties in standard data simulations, regardless of the times in which we live.

This thesis combines timeless contributions to statistical concepts with modern topics of economics as well as modern issues of statistical testing. The rapidly increasing size and complexity of collected data imply that questions related to time efficiency in computation are a major concern. But also, rapid increases in computer processing power, Monte Carlos techniques and simulation based research enable an analysis that uses radically different statistical models than a few decades ago. Kroese et al. (2014) write that statistical analysis can be based on any probabilistic model that can be simulated on a computer. Data science is looking to significantly shape the production processes in the next years and decades. While interest in the specific contributions to statistical analysis will fluctuate, developments in statistics are certain to remain valid and widely applicable.

So far, the research of this thesis has led to a publication in an academic journal and has been discussed at four international conferences on applied econometrics and finance. Given a broad target audience, there are vast opportunities for the testing concepts of this thesis to be used to further knowledge creation in many areas of societal interest, and to inspire continued research on the specific topics and similar advancements in other areas of statistical research.

Curriculum Vitae

Dominik Blatt was born on December 5, 1984 in Koblenz, Germany. He studied Economics at the Universität zu Köln, where he graduated as a Diplom-Volkswirt in 2012.

Later in 2012, Dominik joined the Department of Economics at Maastricht University as a Ph.D. candidate under the supervision of Prof. Dr. Bertrand Candelon. At Maastricht University, Dominik's research interest quickly gravitated towards applied statistical methods in economic modeling. The results of his research are presented in this dissertation. So far, one chapter of the dissertation is published in an international, refereed academic journal. Dominik has presented his research at several international conferences and workshops.

Since 2017 Dominik has two positions at Statistics Netherlands (CBS) in Heerlen, the Netherlands. For the Information Management and Information Systems Management team he contributes to the implementation of Standard Business Reporting data channels. As a data scientist at the Center for Big Data Statistics he develops artificial neural networks for the classification of vacancy data.