

DISSERTATION

INFERENCE FOR FUNCTIONAL TIME SERIES WITH APPLICATIONS TO YIELD
CURVES AND INTRADAY CUMULATIVE RETURNS

Submitted by

Gabriel J. Young

Department of Statistics

In partial fulfillment of the requirements

For the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2016

Doctoral Committee:

Advisor: Piotr S. Kokoszka

Hong Miao

F. Jay Breidt

Wen Zhou

Copyright by Gabriel J. Young 2016

All Rights Reserved

ABSTRACT

INFERENCE FOR FUNCTIONAL TIME SERIES WITH APPLICATIONS TO YIELD CURVES AND INTRADAY CUMULATIVE RETURNS

Econometric and financial data often take the form of a functional time series. Examples include yield curves, intraday price curves and term structure curves. Before an attempt is made to statistically model or predict such series, we must address whether or not such a series can be assumed stationary or trend stationary. We develop extensions of the KPSS stationarity test to functional time series. Motivated by the problem of a change in the mean structure of yield curves, we also introduce several change point methods applied to dynamic factor models. For all testing procedures, we include a complete asymptotic theory, a simulation study, illustrative data examples, as well as details of the numerical implementation of the testing procedures. The impact of scheduled macroeconomic announcements has been shown to account for sizable fractions of total annual realized stock returns. To assess this impact, we develop methods of derivative estimation which utilize a functional analogue of local-polynomial smoothing. The confidence bands are then used to find time intervals of statistically increasing cumulative returns.

ACKNOWLEDGEMENTS

I am grateful for my doctoral committee members: Professors Piotr Kokoszka, Hong Miao, F. Jay Breidt and Wen Zhou. Without my committee, I would have never been able to accomplish this amazing task.

Out of everyone in my academic career, I cannot express enough thanks to my advisor Professor Piotr Kokoszka. His guidance, support and encouragement have been immeasurable. He has helped me grow into a confident researcher and working with him has been a high honor in itself. On a personal note, Professor Piotr Kokoszka has always provided me with great advice outside of academia and has proven to be a man of great character.

I give special thanks to Professor Hong Miao. As my outside committee member, he has played a pivotal role in my graduate career. He provided all relevant data sets that were analyzed in my dissertation and has given me the opportunity to present our research to the finance community. Professor Hong Miao has also been a great support system throughout my graduate career.

One project in my dissertation is a collaboration with faculty from the Department of Mathematics at the University of Utah. I thank Professor Lajos Horváth for his expertise in functional time series and theoretical contributions to our research. It is an exceptionally high honor to have worked with Professor Lajos Horváth. I thank Patrick Bardsley for his technical expertise and contributions to our collaborative project.

I am grateful to many professors at Colorado State University who contributed to my knowledge in statistics and mentored me in teaching. I give thanks to Professor Chihoon Lee for serving on my doctoral committee for several years and teaching me the fundamentals of probability theory. I am grateful for Professor Phillip Turk, who has been a good friend and mentor. I thank Professor Jean Opsomer for being an outstanding chair of the Department of Statistics and for granting me

numerous teaching opportunities. I am also particularly grateful for Professor Mary Meyer, who is a great teacher and mentor.

I also want to thank my fellow graduate students. Working together and becoming friends was an amazing experience, which has helped contribute to my success.

Lastly I wish to thank my family and friends. I am fortunate to say there are too many people to list entirely. To name a few, I thank my mother Cindy Cavaleri, my father Cliff Young and my brother Zach Young. Through good times and bad, thank you for the love and support.

DEDICATION

This dissertation is dedicated to my mother, for her continued love and support, and for taking care of Grandpa Joe all the way to the end.

TABLE OF CONTENTS

ABSTRACT	ii
ACKNOWLEDGEMENTS	iii
DEDICATION	v
LIST OF TABLES	x
LIST OF FIGURES	xiii
Chapter 1 - Introduction	1
1.1 Background	1
1.2 Functional Time Series	12
1.3 KPSS Test for Functional Time Series	14
1.4 Change Point Tests in Functional Factor Models	15
1.5 Determination of the Interval of Increasing Cumulative Returns Prior to Macroeconomic Announcements	16
Chapter 2 - KPSS Test for Functional Time Series	19
2.1 Introduction	19
2.2 Problem Statement, Definitions and Assumptions	22
2.3 Test Statistics and Their Limit Distributions	26
2.4 Application to Yield Curves and a Simulation Study	29
2.5 Proofs of the Results of Section 2.3	32
Chapter 3 - Testing Trend Stationarity of Functional Time Series with Application to Yield and Daily Price Curves	57

3.1	Introduction	57
3.2	Preliminaries	59
3.3	Large sample limits	62
3.4	Algorithmic description of the test procedures	64
3.5	Application to yield and daily price curves	66
Chapter 4 - Change Point Tests in Functional Factor Models with Application to Yield Curves		80
4.1	Introduction	80
4.2	Functional Factor Model	83
4.3	Detection Through Projections onto Factors	87
4.4	A Nonparametric Functional Approach	91
4.5	Finite Sample Performance	96
4.6	Summary	110
4.7	Proofs of the Asymptotic Results	111
4.8	Details of the Numerical Implementation of the Tests	116
Chapter 5 - Determination of the Interval of Increasing Cumulative Returns Prior to Macroeconomic Announcements		120
5.1	Introduction	120
5.2	Statistical Model and Assumptions	124
5.3	Estimation of the Derivative	125
5.4	Asymptotic and Bootstrap Procedures	130
5.5	Data	131
5.6	Empirical Applications	133
5.7	Simulation study	139
5.8	Conclusion	141

5.9 Proofs	153
Bibliography	155

LIST OF TABLES

2.1 Empirical sizes for functional time series generated using Algorithm 13. 32

2.2 Empirical power based on the DGP (2.4.3) and $h = N^{2/5}$ 33

3.1 Critical values of the distribution of the variable $R^0(d)$ given by (3.4.1). 65

3.2 P-values of the tests of Section 3.4 applied to **Treasury Yield Curves**. The data are shown in Figure 3.2. 76

3.3 P-values of the tests of Section 3.4 applied to **S&P500 index**. The data are shown in Figure 3.3. 77

3.4 P-values of the tests of Section 3.4 applied to **U.S. dollar index**. The data are shown in Figure 3.4. 78

3.5 P-values of the tests of Section 3.4 applied to **Oil Futures**. The data are shown in Figure 3.5. 79

4.1 Testing procedures 96

4.2 Application of the test procedures to yield curves over three sampling periods. We expect small P-values in periods (1)–(2), large in period (3). 98

4.3 Empirical size and power for the **ProjSim** approach. 105

4.4 Empirical size and power for the **ProjEigen** approach. 106

4.5 Empirical size and power for the **NFEigen** approach. 107

4.6 Empirical size and power for the **ProjSim** approach with misspecified factors. 108

4.7 Empirical size and power for the **ProjEigen** approach with misspecified factors. 109

5.1 Tickers and Abbreviation 143

5.2 Sampling Periods and Sizes 143

5.3 MSE for Simulations 144

5.4	FOMC and S&P One Day Increasing Return Intervals: 1994 to 2011	145
5.5	FOMC and S&P One Day Increasing Return Intervals: 1994 to 2012	146
5.6	FOMC and S&P Three Days Increasing Return Intervals: Asymptotic Procedure . . .	147
5.7	FOMC and S&P Three Days Increasing Return Intervals: Bootstrap Procedure	148
5.8	EIA and CL One Day Increasing Return Intervals	149
5.9	EIA and CL One Day Increasing Return Intervals: Negative Inventory Shock	150
5.10	EIA and CL One Day Increasing Return Intervals: Positive Inventory Shock	151
5.11	CM and S&P Three Days Increasing Return Intervals: Bootstrap Procedure	152

LIST OF FIGURES

2.1 Right Panel: Ten consecutive yield curves of bonds issued by the United States Federal Reserve; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period. 20

2.2 Left Panel: Ten consecutive price curves of the SP500 index; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period 21

3.1 Left: five consecutive yield curves. Right:prices of the S&P 500 index over five consecutive days. 68

3.2 Consecutive yield curves over two time periods. Vertical lines show the location of sample sizes $N = 50, 100, 150, 200, 250$ 68

3.3 S&P 500 index for two different time periods. 69

3.4 U.S. dollar index for two different time periods. 70

3.5 Light crude oil futures over two different time periods. 70

3.6 P-values of the test based on Algorithm 17 applied to the **Treasury Yield Curves**. The plot on the left shows thirty 100 day periods and the plot on the right shows ten 300 day periods. 72

3.7 P-values of the test described in Algorithm 17 applied to the **S&P500 index**. The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods. 73

3.8 P-values of the test described Algorithm 17 applied to the **U.S. dollar index**. The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods. 74

3.9	P-values of the test described Algorithm 17 applied to the Oil Futures . The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods.	75
4.1	US yield curves over $N = 100$ business days form July-08-2008 to November-28-2008. The central time point corresponds to the Lehman Brothers collapse.	82
4.2	US yield curves at five business days around the Lehman Brothers collapse	97
4.3	Yield curves on $N = 100$ business days form October-18-2005 to March-14-2006.	99
4.4	Yield curves on $N = 100$ business days from June-4-2012 to October-24-2012.	99
4.5	Nelson–Siegel factors $f_1(t, \lambda), f_2(t, \lambda), f_3(t, \lambda)$. Left panel: curves with $\lambda = 0.0609$ corresponding to the domain of real yield curves. Right panel: curves with $\lambda = 21.52$ corresponding to the unit interval.	100
4.6	The left panel shows the residuals of the dynamic Nelson-Siegel model estimated over $N = 250$ business days from March-20-2008 to March-19-2008. The right panel shows $N = 100$ error curves simulated using Equation (4.5.3).	101
4.7	Both panels show simulated yield curves under the alternative hypothesis (4.5.5). In the left panel, including the break point leads to P-value = 0.5%, without it, P-value = 38.1%. In the right panel, when the break point is included, P-value = 0.0%, otherwise P-value = 27.8%.	102
4.8	Intelligible factor curves. Left: corresponding to real yield curves; Right: transformed to the unit interval.	104
5.1	Intraday Cumulative Returns of S&P500 on the FOMC Days	123
5.2	Three Day Window SP 95% Confidence Bands: FOMC Days	135
5.3	One Day Window CL 95% Confidence Bands: FOMC Days	136
5.4	One Day Window CL Asymptotic 95% Confidence Bands: CI Days	138

5.5	Three Day Window SP Asymptotic 95% Confidence Bands: CM Days	138
5.6	$N = 100$ simulated curves using refinement $J = 390$ minutes. The BS and AS procedures are used to estimate when the true maximum occurs with 95% confidence. The true maximum is located at 240 minutes and the true amplitude of the mean $\mu(t)$ is $1/2$. For this realization, the BS confidence band crosses at 223 minutes and the AS confidence band crosses at 227 minutes.	140

Chapter 1

INTRODUCTION

The goal of this chapter is to motivate and introduce the major research topics presented in this dissertation. Each chapter, except for the introduction represents a unique collaborative project, resulting in four papers. The central theme encompasses functional time series with applications to yield curves and intraday cumulative returns. Chapter 2 extends the extensively used KPSS stationarity test of scalar time series to a functional time series setting. Chapter 3 is a comprehensive application of the theoretical developments presented in Chapter 2. The procedures are applied to common financial indices, chosen to represent the market as a whole. Daily treasury yield curves are a well studied example of functional time series. Chapter 4 includes a formal development of a change point test for functional factor models, which have recently become state of the art method of modeling yield curves. Chapter 5 includes a data-driven approach for detecting intervals of statistically increasing cumulative returns in the presence of macroeconomic announcements. Before introducing the individual research topics in more detail, we describe relevant topics of scalar time series analysis, local polynomial smoothing and give an introduction to functional time series.

1.1 Background

Functional data analysis is a growing field of statistics that has applications spanning many disciplines, e.g., biology, physics, geophysics, economics and finance are just a few. Functional time series is a specialized subfield of functional data analysis and has many applications arising in finance and economics. Among recent contributions, we note Antoniadis et al. (2006), Kargin and Onatski (2008), Horváth et al. (2010), Müller et al. (2011), Panaretos and Tavakoli (2013), Kokoszka and Reimherr (2013a), Hörmann et al. (2015), Aue et al. (2015), which account for only a

small list of recent publications. Before introducing relevant definitions of functional data analysis, consider the following motivating example.

1.1.1 Scalar Time Series and Stationarity

Suppose an analyst is investigating the impact monetary policy decisions have on particular financial indices. A common model used in this scenario is

$$r_i = \beta_0 + \beta_1 \mathbf{1}_i(A) + \eta_i, \quad i = 1, 2, \dots, n, \quad (1.1.1)$$

where the η_i 's are independent and identically distributed or temporally dependent random variables, each having mean zero and common variance. The variable $\mathbf{1}_i(A)$ is equal to one during the announcement days and zero otherwise. To control for additional sources of variation, other covariates could also be included in the model. In a recent paper, Lucca and Moench (2015), use a similar model to detect the impact the scheduled Federal Open Market Committee (FOMC) meetings have on the Standard & Poor's 500 financial index over a prespecified sampling period. In nearly every financial or econometric setting, the errors η_i 's do exhibit temporal dependence. Modeling the dependence structure of the errors η_i is vital so that inferential procedures and predictions are well founded.

To effectively model the time series $\{\eta_i, i \in \mathbb{Z}\}$, we require that the series be stationary. This crucial property informally gives insight on the predictability of the time series. If a time series model exhibits stationary behavior, then classical inferential procedures can be performed. Hence stationarity and predictability are often synonymously used in financial and econometric literature. Strict stationarity is defined below:

Definition 1. *Let $\{\eta_i, i \in \mathbb{Z}\}$ be a scalar time series and let $F(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_k})$ be the joint distribution of any size k -subset of this series. The time series is strictly stationary if for any lag h ,*

$$F(\eta_{t_1+h}, \eta_{t_2+h}, \dots, \eta_{t_k+h}) = F(\eta_{t_1}, \eta_{t_2}, \dots, \eta_{t_k}).$$

If a time series is stationary, the joint distribution does not change with any shift in time. It is often convenient to work with a weaker form of stationarity.

Definition 2. *A scalar time series $\{\eta_i, i \in \mathbb{Z}\}$ is weakly stationary if*

1. $E(\eta_i)$ is independent of i ,
2. $\text{Cov}(\eta_{i+h}, \eta_i)$ is independent of i for each h .

Assuming the variances exist, strict stationarity implies weak stationarity but the converse is not true. The canonical example of a stationary time series is the autoregressive process

$$\eta_{i+1} = \varphi\eta_i + u_i, \tag{1.1.2}$$

where the u_i 's are iid mean zero random variables each sharing common variance σ_u^2 . Stationarity holds provided $|\varphi| < 1$, see e.g. page 79 of Brockwell and Davis (1991). If $\varphi = 1$, the autoregressive process becomes a random walk which is the canonical example of a nonstationary time series. The random walk is thus defined by

$$\eta_{i+1} = \eta_i + u_i, \tag{1.1.3}$$

where the u_i 's are iid mean zero random variables each sharing the common variance σ_u^2 . Here the second condition of Definition 2 is violated since the covariance function of a random walk clearly depends on i for each lag h .

The concept of stationarity can then be imposed on common financial and statistical models. For example, one can assume the errors η_i 's from model (1.1.1) form a stationary time series. This would encompass a large class of error structures and consequently produce valid inferential procedures for testing relevant parameters and forecasting returns.

One key distinction between an iid sequence of random variables and a stationary sequence of random variables is the covariance structure. For an iid sequence of mean zero random variables $\{\eta_i, i \in \mathbb{Z}\}$, each case shares the same variance $\sigma^2 = E[\eta_1^2]$. If the time series $\{\eta_i, i \in \mathbb{Z}\}$ is stationary, then the common dispersion parameter of this sequence is coined as the long-run variance.

Definition 3. Suppose $\{\eta_i, i \in \mathbb{Z}\}$ is a stationary time series with covariances $\text{Cov}(\eta_i, \eta_{i-j}) = \gamma_j$.

Then, the long-run variance σ^2 is defined by

$$\sigma^2 = \sum_{j=-\infty}^{\infty} \gamma_j = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j, \quad (1.1.4)$$

provided the infinite series converges.

One way to motivate this concept comes from the Central Limit Theorem for strictly stationary sequences. Under many forms of assumptions, including mixing and cumulant type assumptions,

$$\sqrt{n}(\bar{\eta}_n - E\eta_1) \xrightarrow{D} N(0, \sigma^2),$$

where $\bar{\eta}_n$ is the sample mean of the time series and σ^2 is the long-run variance defined in Equation (1.1.4). For stationary series, the variance of the sample mean is thus asymptotically approximated by σ^2/n , where σ^2 is the long-run variance rather than the variance of η_1 , as would be the case for iid η_i . To make the idea of the long-run variance more tangible, consider the autoregressive process defined in Equation (1.1.2). The covariances of the autoregressive model are given by $\gamma_j = \sigma_u^2 \varphi^{|j|} / (1 - \varphi^2)$. Consequently, the long-run variance of the autoregressive model is

$$\sigma^2 = \gamma_0 + 2 \sum_{j=1}^{\infty} \gamma_j = \sigma_u^2 / (1 - \varphi)^2.$$

Kernel estimation is one popular technique used for estimating σ^2 . Estimators of this type take the form

$$\hat{\sigma}^2 = \hat{\gamma}_0 + 2 \sum_{j=1}^{n-1} K\left(\frac{j}{h_n}\right) \hat{\gamma}_j, \quad (1.1.5)$$

where $K(\cdot)$ is a kernel and h_n is a bandwidth which, in asymptotic theory, depends on the sample size n . Suitable assumptions are imposed on $K(\cdot)$ and $\{h_n\}$, which guarantee the consistency of this estimator, as n approaches infinity, e.g. Newey and West (1987). The standard Newey-West estimator uses the Barlett kernel defined by

$$K(x) = \begin{cases} 1 - |x| & |x| \leq 1 \\ 0 & \text{otherwise.} \end{cases}$$

1.1.2 Testing for Stationarity of Scalar Time Series

There have been many testing procedures developed to assess whether or not a scalar valued time series exhibits stationarity. In economic research, procedures commonly known as unit root tests have gained greatest importance. The most widely known procedures of this type are the unit root test developed by (Dickey and Fuller 1979, 1981) and Said and Dickey (1984). If we consider autoregressive model (1.1.2), the Dickey-Fuller procedure formally tests the null alternative pair $H_0 : \varphi = 1$ versus $H_1 : |\varphi| < 1$. Heuristically, if we retain the null hypothesis, we cannot rule out the possibility that autoregressive model (1.1.2) exhibits random walk behavior and hence, a unit root. Another popular tool for testing stationarity of scalar time series is the celebrated KPSS test of Kwiatkowski et al. (1992). The work of Kwiatkowski et al. (1992) was in fact motivated by the unit root tests. Since traditional unit root tests have low power in samples of sizes occurring in many applications, Kwiatkowski et al. (1992) proposed that stationarity should be considered as the null hypothesis, and the unit root should be the alternative. The KPSS test has consequently become a standard tool in time series analysis applied to econometric and financial data which is applied alongside unit root tests.

The KPSS procedure assumes stationary under the null, giving the level and trend models:

$$H_0 : x_i = \mu + \eta_i, \quad i = 1, 2, \dots, n \quad (1.1.6)$$

and

$$H_0 : x_i = \mu + \xi i + \eta_i, \quad i = 1, 2, \dots, n, \quad (1.1.7)$$

where the errors η_i 's form a mean zero stationary time series. Model (1.1.6) represents level-stationarity while model (1.1.7) represents trend-stationarity. Model (1.1.7) can be viewed as a stationary time series with a deterministic drift. The alternative hypothesis can include any non-stationary process, but the usual choice is the random walk model. This is because random walk

implies the existence of a dominating nonpredictable component. Thus, the KPSS procedure essentially tests stationarity, or trend stationarity, against the random walk alternative. The random walk is commonly called the unit root process. To test the presence of a unit root, we use the the cumulative sum (CUSUM) process defined by

$$S_k = \sum_{i=1}^k e_i, \quad (1.1.8)$$

where the random variables e_i are the residuals, i.e., $e_i = x_i - \hat{x}_i$. The CUSUM process produces the KPSS test statistic

$$\text{KPSS}_n = \frac{1}{n^2 \hat{\sigma}^2} \sum_{k=1}^n S_k^2, \quad (1.1.9)$$

where $\hat{\sigma}^2$ is the kernel estimator (1.1.5) of the long-run variance σ^2 . For both level and trend cases, the residuals are estimated via least squares. For example, under model (1.1.6),

$$\text{KPSS}_n = \frac{1}{n^2 \hat{\sigma}^2} \sum_{k=1}^n \left\{ \sum_{i=1}^k (x_i - \bar{x}) \right\}^2.$$

Deviations far from H_0 produce large values of (1.1.9), which supports the random walk alternative. Under H_0 , Kwiatkowski et al. (1992) derive limiting distributions of the statistic KPSS_n . For the level-stationary case, test statistic (1.1.9) converges in distribution to a continuous functional of a Brownian bridge $B(r)$:

$$\text{KPSS}_n \xrightarrow{\mathcal{D}} \int_0^1 B^2(r) dr. \quad (1.1.10)$$

Similarly, for the trend-stationary model, test statistic (1.1.9) converges in distribution to a continuous functional of a second level Brownian bridge $V(r)$:

$$\text{KPSS}_n \xrightarrow{\mathcal{D}} \int_0^1 V^2(r) dr. \quad (1.1.11)$$

The second level Brownian bridge is defined by

$$V(r) = W(r) + (2r - 3r^2)W(1) + (-6r + 6r^2) \int_0^1 W(s) ds, \quad 0 \leq r \leq 1,$$

where $W(r)$ is a standard Brownian motion (Wiener process).

1.1.3 Scalar Change Point Analysis

Although the random walk is the classic example of a nonstationary process, another common violation of Definition 2 is when a time series contains a change point.

Definition 4. Let $\{\eta_i, i \in \mathbb{Z}\}$ be a stationary time series. Define a new time series $\{x_i, i \in \mathbb{Z}\}$ by

$$x_i = \begin{cases} \mu_1 + \eta_i, & i \leq k^* \\ \mu_2 + \eta_i, & i > k^*. \end{cases}$$

If $\mu_1 \neq \mu_2$, the integer k^* is the change point (in mean) of the time series $\{x_i, i \in \mathbb{Z}\}$.

A change point intuitively represents a change in the mean structure located at discrete time k^* . A change point violates the first condition of Definition 2. If a time series contains a change point, then model misspecification can introduce problems with both prediction and inference on relevant parameters. Definition 4 can also be generalized to a finite set of change points. For many theoretical purposes, there is no loss in generality using the single change point model. To formally test if the time series $\{x_i, i \in \mathbb{Z}\}$ contains a change point, the null hypothesis is $H_0 : \mu_1 = \mu_2$. This is equivalent to model (1.1.6), which implies the null hypothesis can be expressed as $H_0 : x_i = \mu + \eta_i, i = 1, 2, \dots, n$. Similarly to the scalar KPSS procedure, testing for the presence of a change point uses the same cumulative sum process from Equation (1.1.8). A possible change point test statistic is the $KPSS_n$ statistic. Violations of the change point null hypothesis will result in large values of this statistic. Under H_0 , the limiting distribution is the same as the KPSS procedure.

The problem of detecting a change point has been extensively studied, see e.g. Csörgő and Horváth (1997). In Chapter 4, we develop a change point testing procedure for functional factor models, and apply it to yield curves.

1.1.4 Local Polynomial Smoothing

A review of local polynomial smoothing will facilitate the understanding of Chapter 5, which extends these ideas to a functional data analysis setting. First consider the model

$$x_i = f(s_i) + \eta_i, \quad i = 1, 2, \dots, n,$$

in which f is an unknown smooth function observed at time points s_i in an interval, with a measurement errors, or noise, η_i . In contrast to previous models, we do not impose temporal dependence on the errors, i.e., assume the errors η_i 's are an iid sequence of mean zero random variables. The goal of local polynomial smoothing is to estimate the function f without postulating any specific parametric form. It thus falls into the field of nonparametric statistics. Using this method also allows us to estimate the derivatives of f . An introduction to this technique is presented in Chapter 21 of Ruppert (2011), a detailed and comprehensive treatment is given in Fan and Gijbels (1996). We merely present a few formulas to aid the understanding of Chapter 5.

To estimate the function f and its derivatives, for every t , we minimize the objective function

$$\sum_{i=1}^N \{x_i - (\alpha_0 + \alpha_1(s_i - t) + \dots + \alpha_{p-1}(s_i - t)^{p-1})\}^2 K\left(\frac{s_i - t}{h}\right),$$

with respect to $\alpha_0, \alpha_1, \dots, \alpha_{p-1}$. The smoothed curve at time t is then given by the estimated intercept $\hat{f}(t) = \hat{\alpha}_0$, and the estimated derivative by $\hat{f}'(t) = \hat{\alpha}_1$. In conjunction with the kernel $K(\cdot)$, the bandwidth h governs the level of smoothness. A large bandwidth results in over-smoothing, while a small bandwidth results in over-fitting the curve. Closed form formula for the estimator of the parameter vector $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_{p-1}]'$ follows easily using weighted least squares. For fixed t , define the response vector and design matrix respectively by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{U}_t = \begin{bmatrix} 1 & (s_1 - t) & (s_1 - t)^2 & \dots & (s_1 - t)^{p-1} \\ 1 & (s_2 - t) & (s_2 - t)^2 & \dots & (s_2 - t)^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (s_n - t) & (s_n - t)^2 & \dots & (s_n - t)^{p-1} \end{bmatrix}.$$

Define the diagonal weight matrix by

$$\boldsymbol{\Omega}_{t,h} = \text{diag}\left\{K\left(\frac{s_1 - t}{h}\right), K\left(\frac{s_2 - t}{h}\right), \dots, K\left(\frac{s_n - t}{h}\right)\right\}.$$

The estimated parameter vector is given by the weighted least squares solution

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_{p-1} \end{bmatrix} = (\mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{x}.$$

1.1.5 Functional Data

In models (1.1.1), (1.1.6) and (1.1.7), the observations x_i are scalars. Each model can be extended to a multivariate setting, e.g., each x_i could be a vector simultaneously representing two or more financial indexes. However, several important data structures studied in finance and economics can be most naturally viewed as smooth curves that exhibit certain patterns in their shape. These shapes evolve with time. Such a curve on each trading day could be treated as a single functional observation and the analyst could test whether or not the time series of functions is stationary. While a more general framework is possible, Bosq (2000) and Horváth and Kokoszka (2012), we illustrate the fundamental concepts using minute by minute scalar data on a trading day.

Let J be the number of minutes in a typical trading day. A functional data set can take on the form

$$X_i(t_j), \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, J.$$

Notice in a functional data analysis setting, it is common to use N as the sample size as compared to n . Even with refinement J , we still only have N observations, which represents a significant dimension reduction. Typically each function $X_i(t)$ is assumed to be square integrable over the unit interval, which can be achieved with an affine transformation. The space $L^2 = L^2([0, 1])$ is the set of all measurable real-valued functions defined on the unit interval such that $\int_0^1 x^2(t) dt < \infty$.

The space L^2 is a separable Hilbert space with inner product

$$\langle x, y \rangle = \int x(t)y(t)dt.$$

Note that any integral without limits of integration is assumed to be over the unit interval $[0, 1]$.

The inner product generates the norm

$$\|x(t)\| = \sqrt{\langle x, x \rangle}.$$

The most relevant class of operators in functional data analysis are integral operators defined by

$$\Psi(x)(t) = \int \psi(t, s)x(s)ds, \quad x \in L^2. \quad (1.1.12)$$

The kernel of the integral operator is the real-valued function $\psi(\cdot, \cdot)$. For functional data, all functions X are random elements of L^2 . To define a random curve X , it must be equipped with the Borel σ -algebra and X is said to be integrable if $E\|X\| = E[\int X^2(t)dt]^{1/2} < \infty$. Functional parameters can be defined analogously to the scalar case. If observations X_1, \dots, X_N are iid functions in L^2 such that $E\|X_1\|^2 = E \int X_1^2(t)dt < \infty$, then define the mean function, covariance function and covariance operator respectively by

$$\mu(t) = E[X_1(t)], \quad (1.1.13)$$

$$c(t, s) = E[(X_1(t) - \mu(t))(X_1(s) - \mu(s))], \quad (1.1.14)$$

and

$$C = E[\langle (X_1 - \mu), \cdot \rangle (X_1 - \mu)]. \quad (1.1.15)$$

The covariance operator (1.1.15) is in fact an integral operator defined similarly to (1.1.12). For arbitrary random element X in L^2 such that $EX = 0$ and $E\|X\|^2 = E \int X^2(t)dt < \infty$, the covariance operator C is often written as

$$C(y) = E[\langle X, y \rangle X] = \int c(t, s)y(s)ds, \quad y \in L^2, \quad (1.1.16)$$

where the kernel of C is the centered covariance function $c(s, t) = E[X(t)X(s)]$. All functional parameters can be estimated by their sample equivalents, i.e., the sample mean function, sample covariance function and sample covariance operator are defined respectively by

$$\bar{X}_N(t) = \frac{1}{N} \sum_{i=1}^N X_i(t), \quad (1.1.17)$$

$$\hat{c}(t, s) = \frac{1}{N} \sum_{i=1}^N (X_i(t) - \bar{X}_N(t))(X_i(s) - \bar{X}_N(s)), \quad (1.1.18)$$

and

$$\hat{C}(x) = \frac{1}{N} \sum_{i=1}^N \langle X_i - \bar{X}_N, x \rangle (X_i - \bar{X}_N). \quad (1.1.19)$$

The sample statistics displayed above are important quantities to synthesize the scalar framework with the functional data analysis setting. For a more comprehensive introduction to functional data analysis, see Horváth and Kokoszka (2012).

1.1.6 Eigenvalues and Eigenfunctions

Many theoretical results in functional data analysis use well-known properties of linear operators in the space L^2 . Specifically, eigenvalues and eigenfunctions of the covariance operator (1.1.16) are essential items used in asymptotic theory.

Definition 5. *An eigenvector $\nu(t) \in L^2$ of operator Ψ is a non-zero function such that*

$$\Psi(\nu(t)) = \lambda\nu(t),$$

where λ is a real scalar. λ is called the eigenvalue of Ψ corresponding to eigenfunction $\nu(t)$.

For the space L^2 , there are countably many eigenvalues and the common operator used in functional data analysis is the covariance operator C defined in (1.1.16). Therefore, for all $i \geq 1$, we say $\lambda_i \in \mathbb{R}$ is an eigenvalue of covariance operator C corresponding to non-zero eigenvector ν_i provided that

$$C(\nu_i) = \lambda_i\nu_i.$$

The eigenfunctions are traditionally referred to as the principal components of covariance operator C .

1.2 Functional Time Series

In agreement with the scalar case, temporal dependence can also be imposed on random elements of L^2 . Studying the behavior of random curves over time is the subfield of functional time series. In functional time series, the index t is used to denote “time” within a function. For example, for price curves, t is the time (minutes) within a trading day. Functional observations are indexed by i ; it is convenient to think of i as a trading day. Under this convention, consider the functional analogue of model (1.1.6),

$$X_i(t) = \mu(t) + \eta_i(t), \quad i = 1, 2, \dots, N. \quad (1.2.1)$$

Here, the functional parameter μ is a deterministic element in L^2 and the series η_1, η_2, \dots is a stationary time series of functions existing in L^2 . To rigorously prove asymptotic theory related to functional time series, we must formally define stationarity in a functional data analysis setting. One method of defining these objects was introduced in Hörmann and Kokoszka (2010, 2012), and Chapter 16 of Horváth and Kokoszka (2012). The dependence imposed on the errors η_i 's are called Bernoulli shifts and their structure is summarized in Definition 6.

Definition 6. *A sequence of random functions $\{\eta_j, j \in \mathbb{Z}\}$ are Bernoulli shifts if for some measurable function $g : S^\infty \rightarrow L^2$ and iid functions $\epsilon_j, -\infty \leq j \leq \infty$, with values in a measurable space S , each function admits representation $\eta_j = g(\epsilon_j, \epsilon_{j-1}, \dots)$. The functions $(t, \omega) \mapsto \eta_j(t, \omega)$ are product measurable, $E\eta_0 = 0$ in L^2 and $E\|\eta_0\|^{2+\delta} < \infty$ for some $0 < \delta < 1$. The sequence $\{\eta_n\}_{n=-\infty}^\infty$ can be approximated by ℓ -dependent sequences $\{\eta_{n,\ell}\}_{n=-\infty}^\infty$ in the sense that*

$$\sum_{\ell=1}^{\infty} (E\|\eta_n - \eta_{n,\ell}\|^{2+\delta})^{1/\kappa} < \infty \text{ for some } \kappa > 2 + \delta, \quad (1.2.2)$$

where $\eta_{n,\ell}$ is defined by

$$\eta_{n,\ell} = g(\epsilon_n, \epsilon_{n-1}, \dots, \epsilon_{n-\ell+1}, \epsilon_{n-\ell}^*, \epsilon_{n-\ell-1}^*, \dots),$$

and the ϵ_k^* are independent copies of ϵ_0 , independent of $\{\epsilon_i, -\infty < i < \infty\}$.

The central theme of Definition 6 is that the dependence of the function g on the iid innovations ϵ_j far in the past decays so fast that these innovations can be replaced by their independent copies. This replacement is asymptotically negligible using relation (1.2.2).

Similar to scalar valued time series, the CUSUM process will give insight on the temporal behavior of model (1.2.1). Define the CUSUM process as

$$\frac{1}{\sqrt{N}} \left(\sum_{1=1}^{[Nx]} X_i(t) - \frac{[Nx]}{N} \sum_{i=1}^N X_i(t) \right). \quad (1.2.3)$$

Continuous functionals of Equation (1.2.3) will produce reasonable test statistics to formally assess whether a functional time series is stationary.

We also define a covariance structure that accounts for the long-run behavior of functional time series $\{\eta_i, i \in \mathbb{Z}\}$.

Definition 7. Let $\{\eta_i, i \in \mathbb{Z}\}$ be a stationary functional time series each having mean zero. The long-run covariance kernel of the time series is

$$c(t, s) = E\eta_0(t)\eta_0(s) + \sum_{i=1}^{\infty} (E\eta_0(t)\eta_i(s) + E\eta_0(s)\eta_i(t)), \quad 0 \leq s, t \leq 1. \quad (1.2.4)$$

The above definition is analogous to Definition 3. The function $c(t, s)$ is positive definite which implies the existence of eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, and orthonormal eigenfunctions $\phi_i(t)$, $0 \leq t \leq 1$, satisfying

$$\lambda_i \phi_i(t) = \int c(t, s) \phi_i(s) ds, \quad 0 \leq i \leq \infty. \quad (1.2.5)$$

The above relation yields the principal components of the long-run covariance operator C . The spectral properties are crucial in asymptotic results related to functional time series. Estimation of the long-run covariance kernel is similar to the scalar case. Define the estimated auto-covariances by

$$\hat{\gamma}_i(t, s) = \frac{1}{N} \sum_{j=i+1}^N e_j(t) e_{j-i}(s), \quad 0 \leq i \leq N-1, \quad (1.2.6)$$

where e_j is the j th functional residual. Define the estimated long-run covariance kernel by

$$\hat{c}(t, s) = \hat{\gamma}_0(t, s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) (\hat{\gamma}_i(t, s) + \hat{\gamma}_i(s, t)). \quad (1.2.7)$$

The empirical eigenvalues and empirical principal components are computed using the estimated long-run covariance kernel $\hat{c}(t, s)$.

1.3 KPSS Test for Functional Time Series

This section introduces the topics presented in Chapter 2 and Chapter 3. The research presented in these two chapters is a collaborative project with authors Piotr Kokoszka and Gabriel Young. As mentioned in Section 1.1.2, the KPSS test of Kwiatkowski et al. (1992) has become one of the standard tools in the analysis of econometric time series. In Chapter 2, we extend the trend-stationary KPSS test procedure to a functional data analysis setting, which includes rigorous theory and a simulation study. A contribution most closely related to Chapter 2 is Horváth et al. (2014) who developed a test of level stationarity. The functional time series that motivate this work are visually not level stationary, but can be potentially trend stationary. Incorporating a possible trend changes the structure of functional residuals and leads to different limit distributions. The null hypothesis of trend stationarity is stated as follows:

$$H_0 : \quad X_i(t) = \mu(t) + i\xi(t) + \eta_i(t). \quad (1.3.1)$$

The functions μ and ξ correspond, respectively, to the intercept and slope. The errors η_i are also functions and obey the dependence structure from Definition 6. Under the alternative, the model contains a random walk component:

$$H_1 : \quad X_i(t) = \mu(t) + i\xi(t) + \sum_{k=1}^i u_k(t) + \eta_i(t). \quad (1.3.2)$$

Using a CUSUM process similar to (1.2.3), we develop a test statistic that extends the trend-stationarity procedure from Kwiatkowski et al. (1992) to a functional time series setting. Under

H_0 , we develop rigorous asymptotic theory which has limiting distributions relating to (1.1.11). We also show that the test statistic is consistent under the random walk alternative (1.3.2). A simulation study is conducted to find empirical size under H_0 and empirical power under the random walk alternative (1.3.2).

Chapter 3 is an extensive application of the functional trend stationarity test developed in Chapter 2. This chapter is a joint paper with authors Piotr Kokoszka and Gabriel Young. Many financial data sets form functional time series. The best known and most extensively studied data of this form are yield curves. Even though they are reported at discrete maturities, in financial theory and practice they are viewed as continuous functions, one function per month or per day. Other well known examples include intraday price, volatility or volume curves. For this application, we consider daily yield curves, daily price curves of the S&P 500 financial index, daily price curves of the U.S. dollar index and daily price curves of light crude oil futures. The main objective of the empirical analysis is to uncover commonalities and differences between the various classes of assets with respect to the trend behavior of specific daily functions. Functional time series of this type often exhibit a visual trend, and obviously cannot be treated as stationary. The question is if trend plus stationarity is enough or if a nonstationary component must be included. If the time period is sufficiently long, trend stationarity will not be a realistic assumption due to periods of growth and recession and changes in monetary policy of central banks.

1.4 Change Point Tests in Functional Factor Models

Chapter 4 is a joint project with authors Patrick Bardsley, Lajos Horváth, Piotr Kokoszka and Gabriel Young. In this paper, we introduce several methods to test the null hypothesis that the mean structure of a time series of curves does not change. Yield curve modeling has been an important direction of economic research over many decades. An approach that has gained wide acceptance in recent years is the Nelson–Siegel model and its dynamic modification, Diebold and Rudebusch (2013). The most direct motivation for this project comes from the work of Chen and

Niu (2014) who show that accounting for possible change points in the term structure improves yield curve predictions. Our paper is concerned with the detection of change points in models that generalize the dynamic Nelson–Siegel model.

The general form of the dynamic Nelson–Siegel model can be written as

$$X_i(t_j) = \sum_{k=1}^K \beta_{i,k} f_k(t_j) + \eta_i(t_j), \quad (1.4.1)$$

where the $t_j, 1 \leq j \leq J$, denotes the maturities ordered from the shortest (1 month) to the longest (10 years). The index i refers to time periods at which the curves are available, it typically indexes days or months. The functions f_k are postulated to have a specific parametric form. The attribute “dynamic” stems from the fact that the weights $\beta_{i,k}$ are time series; in a static model $\beta_{i,k} = \beta_k$ does not depend on period i . The K -series $\{\beta_{i,k}\}$ and errors η_i are assumed to be joint stationary time series following a dependence structure analogous to Definition 6.

The objective of this work is to develop significance tests whose null hypothesis is that the mean structure of the K -series $\{\beta_{i,k}\}$ is constant over a time period under consideration against the alternative that it changes at unknown change points. We consider a projection approach and a fully functional approach. Test statistics are formulated using continuous functionals of cumulative sum processes. The fully functional approach exploits the CUSUM process from Equation (1.2.3). We develop formal tests of significance, simulate empirical size, simulate empirical power and apply the developed procedures to yield curves over periods of economic expansion and contraction.

1.5 Determination of the Interval of Increasing Cumulative Returns Prior to Macroeconomic Announcements

Chapter 5 is joint project with authors Piotr Kokoszka, Hong Miao and Gabriel Young. The goal of this project is to identify the time intervals over which an average increase of high frequency cumulative returns can be claimed under the presence of scheduled macroeconomic announcements.

High–frequency, and intraday financial data in general, have been an important focus of research in finance, econometrics and statistics for over two decades. The literature is enormous; an

introduction is given in Chapters 5 and 6 of Tsay (2005), and to list a few influential publications, which contain literature overview, we cite Engle and Russel (2004), Barndorff-Nielsen and Shephard (2004), Hayashi and Yoshida (2005), Wang and Zou (2010), Andersen et al. (2012).

The impact of scheduled macroeconomic news on assets has been an important research topic over the last few years. There have been several studies that established the impact or lack thereof scheduled announcements on various types of assets, see e.g. Elder et al. (2011) for an enquiry of this type which also contains references to earlier work. The most direct motivation for this project comes from the work of Lucca and Moench (2015). As observed by Lucca and Moench (2015), an interesting phenomenon occurs on the day of the scheduled Federal Open Market Committee (FOMC) announcement. The authors document large average excess returns on U.S. equities in anticipation of monetary policy decisions made at scheduled meetings of the Federal Open Market Committee (FOMC). This phenomenon, which was observed over the last few decades is coined as the “pre-FOMC drift”. These pre-FOMC returns have increased over time and account for sizable fractions of total annual realized stock returns. To infer upon this phenomenon, Lucca and Moench (2015) use a model similar to Equation (1.1.1). Although the “dummy” model approach is common practice in finance, the method lacks novelty and the ability to detect precisely where the cumulative returns are statistically increasing and decreasing.

We develop a data-driven approach for detecting intervals of statistically increasing cumulative returns in the presence of macroeconomic announcements. This is done by finding an interval with a positive derivative and extending the ideas developed by Liu and Müller (2009). We consider Cumulative Intradaily Returns (CIDR’s) of the assets. The CIDR’s are defined by

$$R_i(t) = 100\{\log P_i(t) - \log P_i(0)\}, \quad (1.5.1)$$

where $P_i(t)$ is the price of the asset at time t on day i . Since $R_i(t) \approx 100(P_i(t) - P_i(0))/P_i(0)$, the CIDR’s reflect the shape of the daily price curves. We assume that the observed CIDR curves,

$R_1(t), \dots, R_N(t)$, obey the model

$$R_i(t) = \mu(t) + \varepsilon_i(t), \quad i = 1, \dots, N, \quad t \in [0, 1], \quad (1.5.2)$$

where μ is the unknown mean function, $\mu(t) = ER_i(t)$, and the curves ε_i are independent identically distributed random functions with $E\varepsilon_1(t) = 0$. In model (1.5.2), the mean function μ is a parameter to which inference applies. We are specifically interested in estimating the derivative $\mu'(t)$ at any given time point t and finding a confidence interval for this derivative. We use a framework akin to the local polynomial smoothing, see Section 1.1.4, but suitably modified to apply to model (1.5.2).

A local polynomial estimator is calculated by minimizing the objective function

$$\sum_{i=1}^N \int_0^1 \{R_i(s) - (\alpha_0 + \alpha_1(s-t) + \dots + \alpha_{p-1}(s-t)^{p-1})\}^2 K\left(\frac{s-t}{h}\right) ds, \quad (1.5.3)$$

with respect to $\alpha_0, \alpha_1, \dots, \alpha_{p-1}$. We denote the estimated derivative by $\hat{\mu}'(t) = \hat{\alpha}_1$. Minimizing (1.5.3) is a different procedure than the standard local polynomial smoothing of a single function observed with noise. It makes sense only in the context of model (1.5.2), and is similar to the approach of Liu and Müller (2009).

Chapter 2

KPSS TEST FOR FUNCTIONAL TIME SERIES

2.1 Introduction

The KPSS test of Kwiatkowski et al. (1992) has become one of the standard tools in the analysis of econometric time series. Its null hypothesis is that the series follows the model $x_t = \alpha + \beta t + \eta_t$, where $\{\eta_t\}$ is a stationary time series. The alternative is the model that includes a random walk: $x_t = \alpha + \beta t + \sum_{i \leq t} u_i + \eta_t$, which then dominates the long term behavior of the series. The alternative is thus a series known in econometrics as a unit root or an integrated series. The work of Kwiatkowski et al. (1992) was in fact motivated by the unit root tests of Dickey and Fuller (1979, 1981) and Said and Dickey (1984). In these tests, the null hypothesis is that the series has a unit root. Since such tests have low power in samples of sizes occurring in many applications, Kwiatkowski et al. (1992) proposed that trend stationarity should be considered as the null hypothesis, and the unit root should be the alternative. Rejection of the null hypothesis could then be viewed as a convincing evidence in favor of a unit root. It was soon realized that the KPSS test has a much broader utility. For example, Lee and Schmidt (1996) and Giraitis et al. (2003) used it to detect long memory, with short memory as the null hypothesis; de Jong et al. (1997) developed a robust version of the KPSS test. The work of Lo (1991) is crucial because he observed that under temporal dependence, to obtain parameter free limit null distributions, statistics similar to the KPSS statistic must be normalized by the long run variance rather than by the sample variance.

We develop extensions of the KPSS test to time series of curves, which we call functional time series (FTS). Many financial data sets form FTS. The best known and most extensively studied data of this form are yield curves. Even though they are reported at discrete maturities, in financial

theory and practice they are viewed as continuous functions, one function per month or per day, see Figure 2.1. This is because fractions of bonds can be traded which can have an arbitrary maturity up to 30 years. Other well known examples include intraday price, volatility or volume curves. Intraday price data are smooth, volatility and volume data are noisy, and must be smoothed before they can be effectively treated as curves. Figure 2.2 shows series of price curves. Over a specific period of time, FTS of this type often exhibit a visual trend, and obviously cannot be treated as stationary. The question is if trend plus stationarity is enough or if a nonstationary component must be included. If the time period is sufficiently long, trend stationarity will not be a realistic assumption due to periods of growth and recession and changes in monetary policy of central banks. As in the context of scalar time series, the question is if a specific finite realization can be assumed to be generated by a trend stationary model.

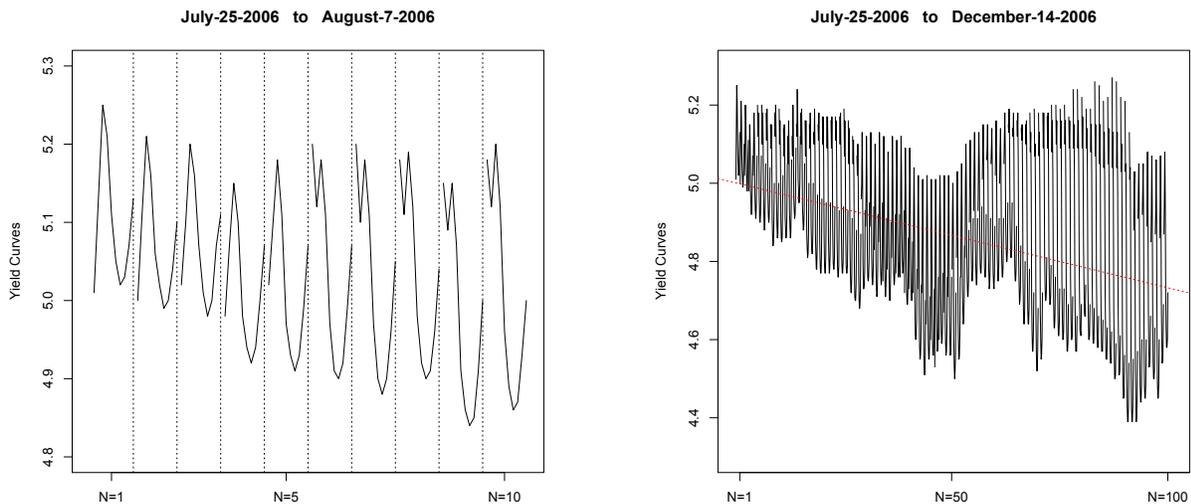


Figure 2.1: Right Panel: Ten consecutive yield curves of bonds issued by the United States Federal Reserve; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period.

We develop the required theory in the framework of functional data analysis (FDA). Application of FDA to time series analysis and econometrics is not new. Among recent contributions, we note Antoniadis et al. (2006), Kargin and Onatski (2008), Horváth et al. (2010), Müller et al. (2011),

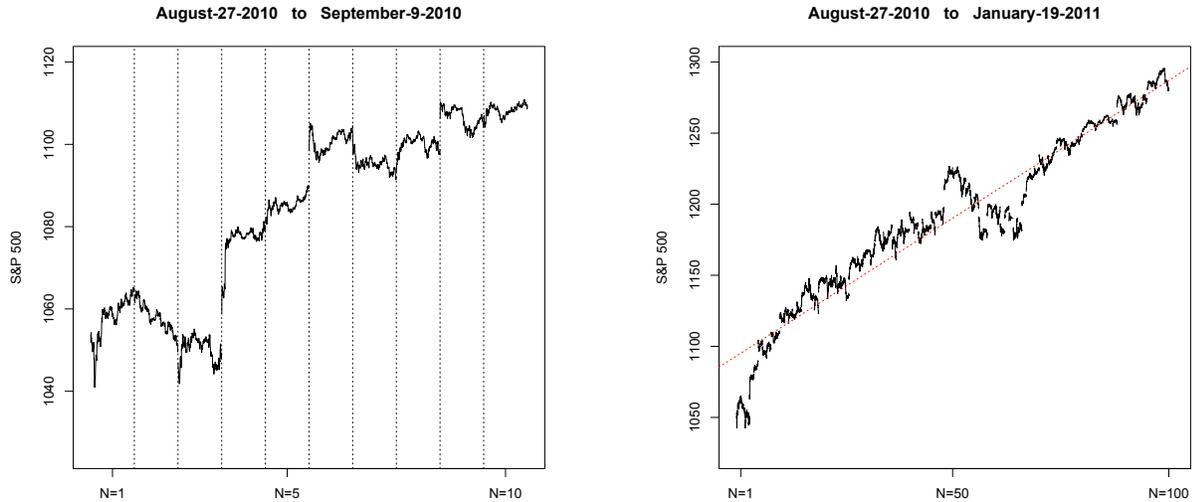


Figure 2.2: Left Panel: Ten consecutive price curves of the SP500 index; Right Panel: a series of 100 of these curves. The red trend line is added for illustration only; the model under H_0 assumes that a function is added at each time period

Panaretos and Tavakoli (2013), Kokoszka and Reimherr (2013a), Hörmann et al. (2015), Aue et al. (2015), with a strong caveat that this list is far from exhaustive. A contribution most closely related to the present work is Horváth et al. (2014) who developed a test of level stationarity. The FTS that motivate this work are visually not level stationary, but can be potentially trend stationary. Incorporating a possible trend changes the structure of functional residuals and leads to different limit distributions. It also requires the asymptotic analysis of the long run variance function of these residuals, which was not required in the level stationary case. A spectral approach to testing stationarity of multivariate time series has recently been developed by Jentsch and Subba Rao (2015). It is possible that it could be extended to a test of *trend* stationarity of FTS, but in this paper we focus on the time domain approach in the spirit of the original work of Kwiatkowski et al. (1992).

The remainder of the paper is organized as follows. Section 2.2 introduces the problem and assumptions. Test statistics and their asymptotic distributions are presented in Section 2.3.

Section 2.4 contains an application to yield curves and a small simulation study. Proofs of the theorems stated in Section 2.3 are developed in Section 2.5.

2.2 Problem Statement, Definitions and Assumptions

In FDA, the index t is used to denote “time” within a function. For example, for price curves, t is the time (e.g. in minutes) within a trading day; for yield curves, t is time to maturity. Functional observations are indexed by n ; it is convenient to think of n as a trading day. Using this convention, the null hypothesis of trend stationarity is stated as follows:

$$H_0 : \quad X_n(t) = \mu(t) + n\xi(t) + \eta_n(t). \quad (2.2.1)$$

The functions μ and ξ correspond, respectively, to the intercept and slope. The errors η_n are also functions. Under the alternative, the model contains a random walk component:

$$H_A : \quad X_n(t) = \mu(t) + n\xi(t) + \sum_{i=1}^n u_i(t) + \eta_n(t). \quad (2.2.2)$$

Our theory requires only that the sequences $\{\eta_n\}$ and $\{u_i\}$ be stationary in a function space, they do not have to be iid. Our tests have power against other alternatives, for example change-points or heteroskedasticity. We focus on the alternative (2.2.2) to preserve the context of the scalar KPSS test.

All random functions and deterministic functional parameters μ and ξ are assumed to be elements of the Hilbert space $L^2 = L^2([0, 1])$ with the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$. This means that the domain of all functional observations, e.g. of the daily price or yield curves, has been normalized to be the unit interval. If the limits of integration are omitted, integration is over the interval $[0, 1]$. All random functions are assumed to be square integrable, i.e. $E \|\eta_n\|^2 < \infty$, $E \|u_n\|^2 < \infty$. Further background on random elements of L^2 is given in Chapter 2 of Horváth and Kokoszka (2012); a more extensive theoretical treatment is presented in Hsing and Eubank (2015).

We quantify the weak dependence of the errors via the following assumption:

Assumption 8. *The errors η_j are Bernoulli shifts, i.e. $\eta_j = g(\epsilon_j, \epsilon_{j-1}, \dots)$ for some measurable function $g : S^\infty \rightarrow L^2$ and iid functions ϵ_j , $-\infty \leq j \leq \infty$, with values in a measurable space S .*

The functions $(t, \omega) \mapsto \eta_j(t, \omega)$ are product measurable.

$E\eta_0 = 0$ in L^2 and $E\|\eta_0\|^{2+\delta} < \infty$ for some $0 < \delta < 1$.

The sequence $\{\eta_n\}_{n=-\infty}^\infty$ can be approximated by ℓ -dependent sequences $\{\eta_{n,\ell}\}_{n=-\infty}^\infty$ in the sense that

$$\sum_{\ell=1}^{\infty} (E\|\eta_n - \eta_{n,\ell}\|^{2+\delta})^{1/\kappa} < \infty \text{ for some } \kappa > 2 + \delta, \quad (2.2.3)$$

where $\eta_{n,\ell}$ is defined by

$$\eta_{n,\ell} = g(\epsilon_n, \epsilon_{n-1}, \dots, \epsilon_{n-\ell+1}, \epsilon_{n-\ell}^*, \epsilon_{n-\ell-1}^*, \dots)$$

where the ϵ_k^ are independent copies of ϵ_0 , independent of $\{\epsilon_i, -\infty < i < \infty\}$.*

Assumption 8 has been shown to hold for all known models for temporally dependent functions, assuming the parameters of these models satisfy nonrestrictive conditions, see Hörmann and Kokoszka (2010, 2012), or Chapter 16 of Horváth and Kokoszka (2012). Its gist is that the dependence of the function g on the iid innovations ϵ_j far in the past decays so fast that these innovations can be replaced by their independent copies. Such a replacement is asymptotically negligible in the sense quantified by (2.2.3). For scalar time series, conditions similar in spirit were used by Pötscher and Prucha (1997), Wu (2005), Shao and Wu (2007) and Berkes et al. (2011), to name just a few references. In this paper, Assumption 8 is needed to ensure that the partial sums (2.3.2) can be approximated by a two-parameter Gaussian process. In particular, (2.2.3) is not used directly; it is a condition used by Berkes et al. (2013a) to prove Theorem 15. To establish the results of Section 2.3, one can, in fact, replace Assumption 8 by the conclusions of Theorem 15 and the existence of an estimator $\hat{c}(t, s)$ such that

$$\iint [\hat{c}(t, s) - c(t, s)]^2 dt ds \xrightarrow{P} 0, \quad \text{as } N \rightarrow \infty, \quad (2.2.4)$$

with the kernel c defined by (2.2.5). Assumption 8 is a general weak dependence condition under which these conclusions hold. While we expect that our limit results can be proven under different weak dependence conditions, the general theorems we use have been so far proven only under Assumption 8.

We now define the bivariate functions appearing in (2.2.4). The long-run covariance function of the errors η_n is defined as

$$c(t, s) = E\eta_0(t)\eta_0(s) + \sum_{i=1}^{\infty} (E\eta_0(t)\eta_i(s) + E\eta_0(s)\eta_i(t)). \quad (2.2.5)$$

The series defining the function $c(t, s)$ converges in $L^2([0, 1] \times [0, 1])$, see Horváth et al. (2013). The function $c(t, s)$ is positive definite. Therefore there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, and orthonormal eigenfunctions $\phi_i(t)$, $0 \leq t \leq 1$, satisfying

$$\lambda_i \phi_i(t) = \int c(t, s) \phi_i(s) ds, \quad 0 \leq i \leq \infty. \quad (2.2.6)$$

To ensure that the ϕ_i corresponding to the d largest eigenvalues are uniquely defined (up to a sign), we assume that

$$\lambda_1 > \lambda_2 > \dots > \lambda_d > \lambda_{d+1} > 0. \quad (2.2.7)$$

The eigenvalues λ_i play a crucial role in our tests. They are estimated by the sample, or empirical, eigenvalues defined by

$$\hat{\lambda}_i \hat{\phi}_i(t) = \int \hat{c}(t, s) \hat{\phi}_i(s) ds, \quad 0 \leq i \leq N, \quad (2.2.8)$$

where $\hat{c}(\cdot, \cdot)$ is an estimator of (2.2.5). We use a kernel estimator similar to that introduced by Horváth et al. (2013), but with suitably defined residuals in place of the centered observations X_n . To define model residuals, consider the least squares estimators of the functional parameters $\xi(t)$ and $\mu(t)$ in model (2.2.1):

$$\hat{\xi}(t) = \frac{1}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) X_n(t) \quad (2.2.9)$$

with

$$s_N = \sum_{n=1}^N \left(n - \frac{N+1}{2} \right)^2 \quad (2.2.10)$$

and

$$\hat{\mu}(t) = \bar{X}(t) - \hat{\xi}(t) \left(\frac{N+1}{2} \right). \quad (2.2.11)$$

The functional residuals are therefore

$$e_n(t) = (X_n(t) - \bar{X}(t)) - \hat{\xi}(t) \left(n - \frac{N+1}{2} \right), \quad 1 \leq n \leq N. \quad (2.2.12)$$

Defining their empirical autocovariances by

$$\hat{\gamma}_i(t, s) = \frac{1}{N} \sum_{j=i+1}^N e_j(t) e_{j-i}(s), \quad 0 \leq i \leq N-1, \quad (2.2.13)$$

leads to the kernel estimator

$$\hat{c}(t, s) = \hat{\gamma}_0(t, s) + \sum_{i=1}^{N-1} K \left(\frac{i}{h} \right) (\hat{\gamma}_i(t, s) + \hat{\gamma}_i(s, t)). \quad (2.2.14)$$

The following assumption is imposed on kernel function K and the bandwidth h .

Assumption 9. *The function K is continuous, bounded, $K(0) = 1$ and $K(u) = 0$ if $|u| > c$, for some $c > 0$. The smoothing bandwidth $h = h(N)$ satisfies*

$$h(N) \rightarrow \infty, \quad \frac{h(N)}{N} \rightarrow 0, \quad \text{as } N \rightarrow \infty. \quad (2.2.15)$$

The assumption that K vanishes outside a compact interval is not crucial to establish (2.2.4).

It is a simplifying condition which could be replaced by a sufficiently fast decay condition, at the cost of technical complications in the proof of (2.2.4).

Recall that if $\{W(x), 0 \leq x \leq 1\}$ is a standard Brownian motion (Wiener process), then the Brownian bridge is defined by $B(x) = W(x) - xW(1)$, $0 \leq x \leq 1$. The second-level Brownian bridge is defined by

$$V(x) = W(x) + (2x - 3x^2)W(1) + (-6x + 6x^2) \int_0^1 W(y) dy, \quad 0 \leq x \leq 1. \quad (2.2.16)$$

Both the Brownian bridge and the second-level Brownian bridge are special cases of the generalized Brownian bridge introduced by MacNeill (1978) who studied the asymptotic behavior of partial sums of polynomial regression residuals. Process (2.2.16) appears as the null limit of the KPSS statistic of Kwiatkowski et al. (1992). We will see in Section 2.3 that for functional data the limit involves an infinite sequence of independent and identically distributed second-level Brownian bridges $V_1(x), V_2(x), \dots$

2.3 Test Statistics and Their Limit Distributions

We will work with the partial sum process of the curves $X_1(t), X_2(t), \dots, X_N(t)$ defined by

$$S_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t), \quad 0 \leq t, x \leq 1, \quad (2.3.1)$$

and the partial sum process of the unobservable errors defined by

$$V_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t), \quad 0 \leq t, x \leq 1. \quad (2.3.2)$$

Test statistic are based on the partial sum process of residuals (2.2.12). Observe that

$$\begin{aligned} \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} e_n(t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \left((X_n(t) - \bar{X}(t)) - \hat{\xi}(t)(n - (N + 1)/2) \right) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t) - \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \bar{X}(t) - \frac{\hat{\xi}(t)}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} (n - (N + 1)/2) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t) - \frac{\lfloor Nx \rfloor}{N} \frac{1}{\sqrt{N}} \sum_{n=1}^N X_n(t) - \frac{\hat{\xi}(t)}{2\sqrt{N}} \left(\lfloor xN \rfloor (\lfloor xN \rfloor - N) \right) \\ &= S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor xN \rfloor}{N} \left(\frac{\lfloor xN \rfloor}{N} - 1 \right) \right). \end{aligned}$$

A suitable test statistic is therefore given by

$$R_N = \iint Z_N^2(x, t) dt dx = \int \|Z_N(x, \cdot)\|^2 dx, \quad 0 \leq t, x \leq 1, \quad (2.3.3)$$

where

$$Z_N(x, t) = S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \quad (2.3.4)$$

and $S_N(x, t)$ and $\hat{\xi}(t)$ are, respectively, defined in equations (2.3.1) and (2.2.9). The null limit distribution of test statistic (2.3.3) is given in the following theorem.

Theorem 10. *If Assumption 8 holds, then under null model (2.2.1),*

$$R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx,$$

where $\lambda_1, \lambda_2, \dots$, are eigenvalues of the long-run covariance function (2.2.5), and V_1, V_2, \dots are iid second-level Brownian bridges.

The proof of Theorem 10 is given in Section 2.5. We now explain the issues arising in the functional case by comparing our result to that obtained by Kwiatkowski et al. (1992). If all curves are constant functions ($X_i(t) = X_i$ for $t \in [0, 1]$), the statistic R_N given by (2.3.3) is the numerator of the KPSS test statistic of Kwiatkowski et al. (1992), which is given by

$$\text{KPSS}_N = \frac{1}{N^2 \hat{\sigma}_N^2} \sum_{n=1}^N S_n^2 = \frac{R_N}{\hat{\sigma}_N^2},$$

where $\hat{\sigma}_N^2$ is a consistent estimator of the long-run variance σ^2 of the residuals. In the scalar case, Theorem 10 reduces to $R_N \xrightarrow{d} \sigma^2 \int_0^1 V^2(x) dx$, where $V(x)$ is a second-level Brownian bridge. If $\hat{\sigma}_N^2$ is a consistent estimator of σ^2 , the result of Kwiatkowski et al. (1992) is recovered, i.e. $\text{KPSS}_N \xrightarrow{d} \int_0^1 V^2(x) dx$. In the functional case, the eigenvalues λ_i can be viewed as long-run variances of the residual curves along the principal directions determined by the eigenfunctions of the kernel $c(\cdot, \cdot)$ defined by (2.2.5). To obtain a test analogous to the scalar KPSS test, with a parameter free limit null distribution, we must construct a statistic which involves a division by consistent estimators of the λ_i . We use only d largest eigenvalues in order not to increase the variability of the statistic caused by division by small empirical eigenvalues. A suitable statistic is

$$R_N^0 = \sum_{i=1}^d \frac{1}{\hat{\lambda}_i} \int_0^1 \langle Z_N(x, \cdot), \hat{\phi}_i \rangle^2 dx, \quad (2.3.5)$$

where the sample eigenvalues $\hat{\lambda}_i$ and eigenfunctions $\hat{\phi}_i$ are defined by (2.2.8). Statistic (2.3.5) extends the statistic KPSS_N . Its limit distribution is given in the next theorem.

Theorem 11. *If Assumptions 8, 9 and (2.2.7) hold, then under null model (2.2.1),*

$$R_N^0 \xrightarrow{\mathcal{D}} \sum_{i=1}^d \int_0^1 V_i^2(x) dx,$$

with the $V_i, 1 \leq i \leq d$, the same as in Theorem 10.

Theorem 11 is proven in Section 2.5. Here we only note that the additional Assumption 9 is needed to ensure that (2.2.4) holds which is known to imply $\hat{\lambda}_i \xrightarrow{P} \lambda_i, 1 \leq i \leq d$.

We conclude this section by discussing the consistency of the tests based on the above theorems. Theorem 12 implies that under H_A statistic R_N of Theorem 10 increases like N^2 . The critical values increase at the rate not greater than N . The test based on Theorem 10 is thus consistent. The exact asymptotic behavior under H_A of the normalized statistic R_N^0 appearing in Theorem 11 is more difficult to study due to almost intractable asymptotics (under H_A) of the empirical eigenvalues and eigenfunctions of the kernel $\hat{c}(\cdot, \cdot)$. The precise asymptotic behavior under H_A is not known even in the scalar case, i.e. for the statistic KPSS $_N$. We therefore focus on the asymptotic limit under H_A of the statistic R_N whose derivation is already quite complex. This limit involves iid copies of the process

$$\Delta(x) = \int_0^x W(y) dy + (3x^2 - 4x) \int_0^1 W(y) dy + (-6x^2 + 6x) \int_0^1 yW(y) dy, \quad 0 \leq x \leq 1, \quad (2.3.6)$$

where $W(\cdot)$ is a standard Brownian motion.

Theorem 12. *If the errors u_i satisfy Assumption 8, then under the alternative (2.2.2),*

$$\frac{1}{N^2} R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x) dx,$$

where R_N is the test statistic defined in (2.3.3) and $\Delta_1, \Delta_2(x), \dots$ are iid processes defined by (2.3.6). The weights τ_i are the eigenvalues of the long-run covariance kernel of the errors u_i defined analogously to (2.2.5) by

$$c_u(t, s) = E[u_0(t)u_0(s)] + \sum_{l=1}^{\infty} E u_0(t) u_l(s) + \sum_{l=1}^{\infty} E u_0(s) u_l(t). \quad (2.3.7)$$

The proof of Theorem 12 is given in Section 2.5.

2.4 Application to Yield Curves and a Simulation Study

In this section, we illustrate the theory developed in this paper with an application to yield curves followed by a simulation study. Applications to other asset classes, including currency exchange rates, commodities and equities are presented in Kokoszka and Young (2015b), which also contain details of numerical implementation.

We consider a time series of daily United States Federal Reserve yield curves constructed from discrete rates at maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months. Yield curves are discussed in many finance textbooks, see e.g. Chapter 10 of Campbell et al. (1997) or Diebold and Rudebusch (2013). The left panel of Figure 2.1 shows ten consecutive yield curves. Following the usual practice, each yield curve is treated as a single functional observation, and so the yield curves observed over a period of many days form a functional time series. The right panel of Figure 2.1 shows the sample period we study, which covers 100 consecutive trading days. It shows a downward trend in interest rates, and we want to test if these curves also contain a random walk component. The tests were performed using $d = 2$. The first two principal components of \hat{c} explain over 95% of variance and provide excellent visual fit. Our selection thus uses three principal shapes to describe the yield curves, the mean function and the first two principal components. It is in agreement with recent approaches to modeling the yield curve, cf. Hays et al. (2012) and Diebold and Rudebusch (2013), which are based on the three component Nelson–Siegel model.

We first apply both tests to the time series of $N = 100$ yield curves shown in the right panel of Figure 2.1. The test based on statistic R_N , yields the P–value of 0.0282 and the test based on R_N^0 , 0.0483, indicating the presence of random walk in addition to a downward trend. Extending the sample by adding the next 150 business days, so that $N = 250$, yields the respective P–values 0.0005 and 0.0013. In all computation the bandwidth $h = N^{2/5}$ was used. Examination of different periods shows that trend stationarity does not hold if the period is sufficiently long. This agrees with the empirical finding of Chen and Niu (2014) whose method of yield curve prediction, based on

utilizing periods of approximate stationarity, performs better than predictions based on the whole sample; random walk is not predictable. Even though our tests are motivated by the alternative of a random walk component, they reject any serious violation of trend stationarity. Broadly speaking, our analysis shows that daily yield curves can be treated as a trend stationary functional time series only over certain short periods of time, generally not longer than a calendar quarter.

We complement our data example with a small simulation study. There is a multitude of data generating process that could be used. The following quantities could vary: shapes of the mean and the principal components functions, the magnitude of the eigenvalues, the distribution of the scores and their dependence structure. In this paper, concerned chiefly with theory, we present a limited simulation study that validates the conclusions of the data example. We therefore attempt to simulate curves whose shapes resemble those of the real data, and for which either the null or the alternative holds. The artificial data is therefore generated according to the following algorithm.

Algorithm 13. *[Yield curves simulation under H_0]*

1. *Using real yield curves, calculate the estimates $\hat{\xi}(t)$ and $\hat{\mu}(t)$ defined, respectively, by (2.2.9) and (2.2.11). Then compute the residuals $e_n(t)$ defined in (2.2.12).*
2. *Calculate the first two empirical principal components $\hat{\phi}_1(t)$ and $\hat{\phi}_2(t)$ using the empirical covariance function*

$$\hat{\gamma}_0(s, t) = \frac{1}{N} \sum_{n=1}^N (e_n(t) - \bar{e}(t))(e_n(s) - \bar{e}(s)). \quad (2.4.1)$$

This step leads to the approximation

$$e_n(t) \approx a_{1,n}\hat{\phi}_1(t) + a_{2,n}\hat{\phi}_2(t), \quad n = 1, 2, \dots, N,$$

where $a_{1,n}$ and $a_{2,n}$ are the first two functional scores. The functions $\hat{\phi}_1(t)$ and $\hat{\phi}_2(t)$ are treated as deterministic, while the scores $a_{1,n}$ and $a_{2,n}$ form random sequences indexed by n .

3. To simulate temporally independent residuals e_n , generate independent in n scores $a'_{1,n} \sim N(0, \sigma_{a_1}^2)$ and $a'_{2,n} \sim N(0, \sigma_{a_2}^2)$, where $\sigma_{a_1}^2$ and $\sigma_{a_2}^2$ are the sample variances of the real scores, and set

$$e'_n(t) = a'_{1,n} \hat{\phi}_1(t) + a'_{2,n} \hat{\phi}_2(t), \quad n = 1, 2, \dots, N.$$

To simulate dependent residual curves, generate scores $a'_{1,n}, a'_{2,n} \sim AR(1)$, where each autoregressive process has parameter 0.5.

4. Using the estimated functional parameters $\hat{\mu}(t)$, $\hat{\xi}(t)$ and the simulated residuals $e'_n(t)$, construct the simulated data set

$$X'_n(t) = \hat{\mu}(t) + \hat{\xi}(t)n + e'_n(t), \quad n = 1, 2, \dots, N. \quad (2.4.2)$$

Table 2.1 shows empirical sizes based on 1000 replication of the data generating process described in Algorithm 13. We use two ways of estimating the eigenvalues and eigenfunctions. The first one uses the function $\hat{\gamma}_0$ defined by (2.4.1) (in the scalar case this corresponds to using the usual sample variance rather than estimating the long-run variance). The second uses the estimated long-run covariance function (2.2.14) with the bandwidth h specified in Table 2.1. The covariance kernel $\hat{\gamma}_0(t, s)$ is appropriate for independent error curves. The bandwidth $h = N^{1/3}$ is too small, not enough temporal dependence is absorbed. The bandwidth $h = N^{2/5}$ gives fairly consistent empirical size, typically within one percent of the empirical size. The bandwidth h is not relevant when the kernel $\hat{\gamma}_0$ is used. The different empirical sizes reflect random variability due to three different sets of 1000 replications being used. This indicates that with 1000 replications, a difference of one percent in empirical sizes is not significant.

To evaluate power, instead of (2.4.2), the data generating process is

$$X'_n(t) = \hat{\mu}(t) + \hat{\xi}(t)n + \sum_{i=1}^n u_i(t) + e'_n(t), \quad n = 1, 2, \dots, N, \quad (2.4.3)$$

where the increments u_i are defined by

$$u_i(t) = aN_{i1} \sin(\pi t) + aN_{i2} \sin(2\pi t),$$

Table 2.1: Empirical sizes for functional time series generated using Algorithm 13.

N	Test statistic	R_N			R_N^0		
	DGP	iid normal	iid normal	AR(1)	iid normal	iid normal	AR(1)
	Cov-kernel	$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$	$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$
100	$h = N^{1/3}$	6.3	5.6	9.4	5.9	5.2	9.1
	$h = N^{2/5}$	5.6	4.4	6.6	5.8	3.6	6.5
	$h = N^{1/2}$	5.1	4.8	3.5	4.5	5.1	2.9
250	$h = N^{1/3}$	5.0	4.3	10.2	5.8	5.2	9.4
	$h = N^{2/5}$	5.5	4.9	7.2	4.5	4.1	5.6
	$h = N^{1/2}$	5.5	5.9	4.3	4.8	3.4	3.5
1000	$h = N^{1/3}$	4.8	4.2	7.0	5.9	5.6	7.1
	$h = N^{2/5}$	6.1	6.3	6.3	6.0	5.1	5.7
	$h = N^{1/2}$	5.8	4.9	4.6	5.6	4.7	3.9

with standard normal N_{ij} , $j = 1, 2$, $1 \leq i \leq N$, totally independent of each other. The scalar a quantifies the distance from H_0 ; $a = 0$ corresponds to H_0 . For all empirical power simulations, we use a 5% size critical value and $h = N^{2/5}$. The empirical power reported in Table 2.2 increases as the sample size and the distance from H_0 increase. It is visibly higher for iid curves as compared to dependent curves.

2.5 Proofs of the Results of Section 2.3

2.5.1 Preliminary results

For ease of reference, we state in this section two theorems which are used in the proofs of the results of Section 2.3. Theorem 14 is well-known, see Theorem 4.1 in Billingsley (1968). Theorem 15 was recently established in Berkes et al. (2013a).

Theorem 14. *Suppose Z_N, Y_N, Y are random variables taking values in a separable metric space with the distance function ρ . If $Y_N \xrightarrow{\mathcal{D}} Y$ and $\rho(Z_N, Y_N) \xrightarrow{P} 0$, then $Z_N \xrightarrow{\mathcal{D}} Y$.*

Table 2.2: Empirical power based on the DGP (2.4.3) and $h = N^{2/5}$.

Test statistic		R_N			R_N^0		
		iid normal	iid normal	AR(1)	iid normal	iid normal	AR(1)
DGP		$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$	$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$
Cov-kernel		$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$	$\hat{\gamma}_0(t, s)$	$\hat{c}(t, s)$	$\hat{c}(t, s)$
$a = 0.1$	$N = 125$	100.0	89.9	10.1	100.0	87.9	10.4
	$N = 250$	100.0	97.0	27.7	100.0	96.0	21.9
$a = 0.5$	$N = 125$	100.0	91.5	83.1	100.0	89.7	71.2
	$N = 250$	100.0	97.3	96.4	100.0	97.4	92.4

In our setting, we work in the metric space $D([0, 1], L^2)$ which is the space of right-continuous functions with left limits taking values in $L^2([0, 1])$. A generic element of $D([0, 1], L^2)$ is

$$z = \{z(x, t), 0 \leq x \leq 1, 0 \leq t \leq 1\}.$$

For each fixed x , $z(x, \cdot) \in L^2$, so $\|z(x, \cdot)\|^2 = \int z^2(x, t)dt < \infty$. The uniform distance between $z_1, z_2 \in D([0, 1], L^2)$ is

$$\rho(z_1, z_2) = \sup_{0 \leq x \leq 1} \|z_1(x, \cdot) - z_2(x, \cdot)\| = \sup_{0 \leq x \leq 1} \left\{ \int (z_1(x, t) - z_2(x, t))^2 dt \right\}^{1/2}.$$

In the following, we work with the space $D([0, 1], L^2)$ equipped with the uniform distance.

Theorem 15. *If Assumption 8 holds, then $\sum_{i=1}^{\infty} \lambda_i < \infty$, and we can construct a sequence of Gaussian processes $\Gamma_N(x, t)$ such that for every N*

$$\{\Gamma_N(x, t), 0 \leq x, t \leq 1\} \stackrel{\mathcal{D}}{=} \{\Gamma(x, t), 0 \leq x, t \leq 1\},$$

where

$$\Gamma(x, t) = \sum_{i=1}^{\infty} \lambda_i^{1/2} W_i(x) \phi_i(t), \quad (2.5.1)$$

and

$$\kappa_N = \sup_{0 \leq x \leq 1} \|V_N(x, \cdot) - \Gamma_N(x, \cdot)\| = o_p(1). \quad (2.5.2)$$

Recall that the W_i are independent standard Wiener processes, λ_i and ϕ_i are defined in (2.2.6) and $V_N(x, t)$ is defined in (2.3.2).

2.5.2 Proof of Theorem 10

The proof of Theorem 10 is constructed from several lemmas decomposing the statistic R_N into a form suitable for the application of the results of Section 2.5.1, i.e. to leading and asymptotically negligible terms. Throughout this section, we assume that the null model (2.2.1) and Assumption 8 hold.

Lemma 1. For the s_N is defined in (2.2.10),

$$\frac{s_N}{N^3} \rightarrow \frac{1}{12}, \quad \text{as } N \rightarrow \infty.$$

Proof. Identify the left-hand side with a Riemann sum. □

Lemma 2. For the functional slope estimate $\hat{\xi}$ defined by (2.2.9),

$$\hat{\xi}(t) - \xi(t) = \frac{1}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) \eta_n(t).$$

Proof. Use the identities

$$\sum_{n=1}^N \left(n - \frac{N+1}{2} \right) = 0, \quad \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) n = \sum_{n=1}^N \left(n - \frac{N+1}{2} \right)^2 = s_N.$$

□

Lemma 3. The following identity holds

$$\frac{N^{3/2}}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) \eta_n(t) = \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N \left(\frac{k}{N}, t \right) \right\},$$

where $V_N(x, t)$ is the partial sum process of the errors defined in (2.3.2).

Proof. Notice that

$$\begin{aligned} & \frac{N^{3/2}}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) \eta_n(t) \\ &= \frac{1}{N^{-3}s_N} \left\{ \frac{1}{N} \frac{1}{\sqrt{N}} \sum_{n=1}^N n \eta_n(t) - \left(\frac{N+1}{2N} \right) \frac{1}{\sqrt{N}} \sum_{n=1}^N \eta_n(t) \right\}. \end{aligned}$$

Using the relation

$$\sum_{n=1}^N n\eta_n(t) = N \sum_{n=1}^N \eta_n(t) - \sum_{k=1}^{N-1} \sum_{n=1}^k \eta_n(t),$$

we have

$$\begin{aligned} & \frac{N^{3/2}}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) \eta_n(t) \\ &= \frac{1}{N^{-3}s_N} \left\{ \frac{1}{N} \frac{1}{\sqrt{N}} \left(N \sum_{n=1}^N \eta_n(t) - \sum_{k=1}^{N-1} \sum_{n=1}^k \eta_n(t) \right) \right. \\ & \quad \left. - \left(\frac{N+1}{2N} \right) \frac{1}{\sqrt{N}} \sum_{n=1}^N \eta_n(t) \right\} \\ &= \frac{1}{N^{-3}s_N} \left\{ \frac{1}{\sqrt{N}} \sum_{n=1}^N \eta_n(t) - \frac{1}{N} \sum_{k=1}^{N-1} \frac{1}{\sqrt{N}} \sum_{n=1}^k \eta_n(t) \right. \\ & \quad \left. - \left(\frac{N+1}{2N} \right) \frac{1}{\sqrt{N}} \sum_{n=1}^N \eta_n(t) \right\} \\ &= \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\}. \end{aligned}$$

□

Lemma 4. *The process $Z_N(x, t)$ defined by (2.3.4) admits the decomposition*

$$\begin{aligned} Z_N(x, t) &= V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) \\ & \quad - \frac{1}{2} \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\} \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right). \end{aligned}$$

Proof. Notice that

$$\begin{aligned} Z_N(x, t) &= S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t) - \frac{\lfloor Nx \rfloor}{N} \frac{1}{\sqrt{N}} \sum_{n=1}^N X_n(t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\ &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} (\mu(t) + \xi(t)n + \eta_n(t)) - \frac{\lfloor Nx \rfloor}{N} \frac{1}{\sqrt{N}} \sum_{n=1}^N (\mu(t) + \xi(t)n + \eta_n(t)) \\ & \quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right). \end{aligned}$$

Therefore,

$$\begin{aligned}
Z_N(x, t) &= \mu(t) \frac{1}{\sqrt{N}} \lfloor Nx \rfloor + \xi(t) \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} n + \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) \\
&\quad - \mu(t) \frac{1}{\sqrt{N}} N \frac{\lfloor Nx \rfloor}{N} - \frac{\lfloor Nx \rfloor}{N} \frac{1}{\sqrt{N}} \xi(t) \sum_{n=1}^N n - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N \eta_n(t) \\
&\quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N \eta_n(t) \\
&\quad + \mu(t) \frac{1}{\sqrt{N}} \lfloor Nx \rfloor - \mu(t) \frac{1}{\sqrt{N}} \lfloor Nx \rfloor \\
&\quad + \xi(t) \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} n - \frac{\lfloor Nx \rfloor}{N} \xi(t) \frac{1}{\sqrt{N}} \sum_{n=1}^N n \\
&\quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}$$

By Lemma 2,

$$\begin{aligned}
Z_N(x, t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N \eta_n(t) \\
&\quad + \xi(t) \frac{1}{\sqrt{N}} \frac{\lfloor Nx \rfloor (\lfloor Nx \rfloor + 1)}{2} - \frac{\lfloor Nx \rfloor}{N} \xi(t) \frac{1}{\sqrt{N}} \frac{N(N+1)}{2} \\
&\quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N \eta_n(t) \\
&\quad - \frac{1}{2} N^{3/2} (\hat{\xi}(t) - \xi(t)) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N \eta_n(t) \\
&\quad - \frac{1}{2} \left\{ \frac{N^{3/2}}{s_N} \sum_{n=1}^N \left(n - \left(\frac{N+1}{2} \right) \right) \eta_n(t) \right\} \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}$$

By Lemma 3,

$$Z_N(x, t) = V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) - \frac{1}{2} \frac{1}{N^{-3} s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\} \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right).$$

□

Lemma 5. *The following convergence holds*

$$\int \left\{ \frac{1}{N} \sum_{k=1}^N V_N\left(\frac{k}{N}, t\right) - \int_0^1 \Gamma_N(y, t) dy \right\}^2 dt \xrightarrow{P} 0,$$

where the Γ_N are the Gaussian processes in Theorem 15.

Proof. Since V_N is a step function with jumps at $y = \frac{k}{N}$,

$$\frac{1}{N} \sum_{k=1}^N V_N\left(\frac{k}{N}, t\right) = \int_0^1 V_N(y, t) dy.$$

We must thus show that

$$\left\| \int_0^1 V_N(y, \cdot) dy - \int_0^1 \Gamma_N(y, \cdot) dy \right\| \xrightarrow{P} 0.$$

Using the contractive property of integrals and relation (2.5.2), we have

$$\begin{aligned} \left\| \int_0^1 V_N(y, \cdot) dy - \int_0^1 \Gamma_N(y, \cdot) dy \right\| &\leq \int_0^1 \|V_N(y, \cdot) - \Gamma_N(y, \cdot)\| dy \\ &\leq \int_0^1 \sup_{0 \leq x \leq 1} \|V_N(y, \cdot) - \Gamma_N(y, \cdot)\| dy \\ &\leq \kappa_N = o_p(1) \end{aligned}$$

which proves Lemma 5. □

Lemma 6. *Consider the process $\Gamma(\cdot, \cdot)$ defined by (2.5.1) and set*

$$\Gamma^0(x, t) = \Gamma(x, t) + (2x - 3x^2)\Gamma(1, t) + (-6x + 6x^2) \int_0^1 \Gamma(y, t) dy. \quad (2.5.3)$$

Then

$$\int_0^1 \|\Gamma^0(x, \cdot)\|^2 dx = \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x) dx.$$

Proof. Expansion (2.5.1) implies that

$$\begin{aligned}
\Gamma^0(x, t) &= \Gamma(x, t) + (2x - 3x^2)\Gamma(x, t) + (-6x + 6x^2) \int_0^1 \Gamma(y, t) dy \\
&= \sum_{i=1}^{\infty} \sqrt{\lambda_i} W_i(x) \phi_i(t) + (2x - 3x^2) \sum_{i=1}^{\infty} \sqrt{\lambda_i} W_i(1) \phi_i(t) \\
&\quad + (-6x + 6x^2) \int_0^1 \sum_{i=1}^{\infty} \sqrt{\lambda_i} W_i(y) \phi_i(t) dy \\
&= \sum_{i=1}^{\infty} \sqrt{\lambda_i} \left\{ W_i(x) + (2x - 3x^2) W_i(1) \right. \\
&\quad \left. + (-6x + 6x^2) \int_0^1 W_i(y) dy \right\} \phi_i(t) \\
&= \sum_{i=1}^{\infty} \sqrt{\lambda_i} V_i(x) \phi_i(t),
\end{aligned}$$

where V_1, V_2, \dots are iid second-level Brownian bridges defined in (2.2.16). By the orthonormality of the eigenfunctions ϕ_i ,

$$\begin{aligned}
\int_0^1 \|\Gamma^0(x, \cdot)\|^2 dx &= \iint (\Gamma^0(x, t))^2 dt dx \\
&= \iint \left(\sum_{i=1}^{\infty} \sqrt{\lambda_i} V_i(x) \phi_i(t) \right)^2 dx dt \\
&= \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x) dx.
\end{aligned}$$

□

Lemma 7. For the processes $Z_N(\cdot, \cdot)$ and $\Gamma_N^0(\cdot, \cdot)$ defined, respectively, in (2.3.4) and (2.5.3),

$$\sup_{0 \leq x \leq 1} \|Z_N(x, \cdot) - \Gamma_N^0(x, \cdot)\| \xrightarrow{P} 0. \tag{2.5.4}$$

Proof. Using the decomposition given in Lemma 4, $Z_N(x, t)$ can be algebraically manipulated to be expressed as

$$Z_N(x, t) = V_N(x, t) + a_N(x) V_N(1, t) + b_N(x) \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right)$$

where

$$a_N(x) = \left(\frac{1}{2N^{-3}s_N} \left(\frac{N-1}{2N} \right) - 1 \right) \frac{\lfloor Nx \rfloor}{N} - \frac{1}{2N^{-3}s_N} \left(\frac{N-1}{2N} \right) \left(\frac{\lfloor Nx \rfloor}{N} \right)^2 \tag{2.5.5}$$

and

$$b_N(x) = \frac{1}{2N^{-3}s_N} \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right).$$

Notice that

$$\begin{aligned} \|Z_N(x, \cdot) - \Gamma_N^0(x, \cdot)\| &\leq \|V_N(x, t) - \Gamma_N(x, t)\| \\ &+ \left\| a_N(x)V_N(1, t) - (2x - 3x^2)\Gamma_N(1, t) \right\| \\ &+ \left\| b_N(x) \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) - (-6x + 6x^2) \int_0^1 \Gamma_N(y, t) dy \right\|. \end{aligned}$$

Thus, Lemma 7 will be proven once we have established the following relations:

$$\sup_{0 \leq x \leq 1} \|V_N(x, \cdot) - \Gamma_N(x, \cdot)\| \xrightarrow{P} 0; \quad (2.5.6)$$

$$\sup_{0 \leq x \leq 1} \|a_n(x)V_N(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| \xrightarrow{P} 0; \quad (2.5.7)$$

$$\sup_{0 \leq x \leq 1} \left\| b_n(x) \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, \cdot\right) - (-6x + 3x^2) \int_0^1 \Gamma_N(y, \cdot) dy \right\| \xrightarrow{P} 0. \quad (2.5.8)$$

Relation (2.5.6) is the conclusion of Theorem 15. The verification of relation (2.5.7) follows next.

Since

$$\begin{aligned} &\|a_n(x)V_N(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| \\ &\leq |a_n(x)| \|V_N(1, \cdot) - \Gamma_N(1, \cdot)\| + |a_N(x) - (2x - 3x^2)| \|\Gamma_N(1, \cdot)\|, \end{aligned}$$

relation (2.5.7) will hold once we have verified that

$$\sup_{N \geq 1} \left\{ \sup_{0 \leq x \leq 1} |a_N(x)| \right\} < \infty, \quad (2.5.9)$$

$$\|V_N(1, \cdot) - \Gamma_N(1, \cdot)\| \xrightarrow{P} 0, \quad (2.5.10)$$

$$\sup_{0 \leq x \leq 1} |a_N(x) - (2x - 3x^2)| \rightarrow 0, \quad (2.5.11)$$

$$\|\Gamma_N(1, \cdot)\| = O_P(1). \quad (2.5.12)$$

Relation (2.5.10) follows from Theorem 15. By (2.5.1),

$$\|\Gamma_N(1, \cdot)\|^2 \stackrel{\mathcal{D}}{=} \sum_{i=1}^{\infty} \lambda_i Z_i^2,$$

where the Z_i are independent standard normal and $\sum_{i=1}^{\infty} \lambda_i < \infty$. Thus relation (2.5.12) holds trivially because the distribution of the left-hand side does not depend on N . To show (2.5.9), set

$$d_N = \frac{1}{2N^{-3} s_N} \left(\frac{N-1}{2N} \right).$$

By Lemma 1, $d_N \rightarrow 3$, and

$$a_n(x) = (d_N - 1) \frac{\lfloor Nx \rfloor}{N} - d_N \left(\frac{\lfloor Nx \rfloor}{N} \right)^2.$$

Since $\lfloor Nx \rfloor \leq N$, $|a_N(x)| \leq |d_N - 1| + d_N$, and (2.5.9) follows. To show relation (2.5.11), first notice that

$$|a_N(x) - (2x - 3x^2)| \leq \left| (d_N - 1) \frac{\lfloor Nx \rfloor}{N} - 2x \right| + \left| d_N \left(\frac{\lfloor Nx \rfloor}{N} \right)^2 - 3x^2 \right|.$$

We must thus show that

$$\sup_{0 \leq x \leq 1} \left| (d_N - 1) \frac{\lfloor Nx \rfloor}{N} - 2x \right| \rightarrow 0, \quad (2.5.13)$$

and

$$\sup_{0 \leq x \leq 1} \left| d_N \left(\frac{\lfloor Nx \rfloor}{N} \right)^2 - 3x^2 \right| \rightarrow 0. \quad (2.5.14)$$

To show (2.5.13), first notice that $Nx - 1 \leq \lfloor Nx \rfloor \leq Nx$, which implies

$$(d_N - 1)x - 2x - \frac{1}{N}(d_N - 1) \leq (d_N - 1) \frac{\lfloor Nx \rfloor}{N} - 2x \leq (d_N - 1)x - 2x.$$

Both sides of the inequality are linear in x which implies the extrema occurs at the boundaries.

Thus for any $x \in [0, 1]$,

$$\left| (d_N - 1) \frac{\lfloor Nx \rfloor}{N} - 2x \right| \leq |d_N - 3| \rightarrow 0,$$

which proves relation (2.5.13). To show (2.5.14), notice that $(Nx - 1)^2 \leq \lfloor Nx \rfloor^2 \leq (Nx)^2$, implying

$$d_N x^2 - 3x^2 - \left(\frac{2x}{N} - \frac{1}{N^2} \right) d_N \leq d_N \left(\frac{\lfloor Nx \rfloor}{N} \right)^2 - 3x^2 \leq d_N x^2 - 3x^2.$$

For any $x \in [-1, 0]$, the quadratic functions on each side of the inequality are strictly decreasing.

Hence, the extrema occurs at the boundaries implying for any $x \in [0, 1]$,

$$\left| d_N \left(\frac{\lfloor Nx \rfloor}{N} \right)^2 - 3x^2 \right| \leq |d_N - 3| \rightarrow 0.$$

This proves (2.5.14) and hence relation (2.5.11) holds true. Thus relations (2.5.9), (2.5.10), (2.5.11) and (2.5.11) hold true which proves relation (2.5.7). That is

$$\sup_{0 \leq x \leq 1} \|a_n(x)V_N(1, \cdot) - (2x - 3x^2)\Gamma_N(1, \cdot)\| \xrightarrow{P} 0.$$

Next is the verification (2.5.8). As in the case of (2.5.7), it suffices to show that

$$\sup_{N \geq 1} \left\{ \sup_{0 \leq x \leq 1} |b_N(x)| \right\} < \infty, \quad (2.5.15)$$

$$\left\| \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, \cdot\right) - \int_0^1 \Gamma_N(y, \cdot) dy \right\| \xrightarrow{P} 0, \quad (2.5.16)$$

$$\sup_{0 \leq x \leq 1} |b_N(x) - (6x^2 - 6x)| \rightarrow 0, \quad (2.5.17)$$

$$\left\| \int_0^1 \Gamma_N(y, \cdot) dy \right\| = O_P(1). \quad (2.5.18)$$

The verification of (2.5.15) and (2.5.17) uses the same arguments as the verification of (2.5.9) and (2.5.11) so we do not need to present the details. Relation (2.5.16) coincides with Lemma 5, while relation (2.5.18) follows from Lemma 6. This completes the proof of Lemma 7. \square

Using the above lemmas, we can now present a compact proof of Theorem 10.

PROOF OF THEOREM 10: Recall that the test statistic R_N is defined by $R_N = \iint Z_N^2(x, t) dx dt$,

where

$$Z_N(x, t) = S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right)$$

with $S_N(x, t)$ and $\hat{\xi}(t)$ are respectively defined in equations (2.3.1) and (2.2.9). Recall that

$$\Gamma_N^0(x, t) = \Gamma_N(x, t) + (2x - 3x^2)\Gamma_N(1, t) + (-6x + 6x^2) \int_0^1 \Gamma_N(y, t) dy,$$

and

$$\Gamma^0(x, t) = \Gamma(x, t) + (2x - 3x^2)\Gamma(1, t) + (-6x + 6x^2) \int_0^1 \Gamma(y, t) dy.$$

From Lemma 7, we know that

$$\rho(Z_N(x, \cdot), \Gamma_N^0(x, \cdot)) = \sup_{0 \leq x \leq 1} \|Z_N(x, \cdot) - \Gamma_N^0(x, \cdot)\| \xrightarrow{P} 0.$$

By Theorem 15, $\Gamma_N^0(x, t) \xrightarrow{D} \Gamma^0(x, t)$. Thus, Theorem 14 implies that

$$Z_N(x, t) \xrightarrow{D} \Gamma^0(x, t).$$

By Lemma 6,

$$\iint (\Gamma^0(x, t))^2 dx dt \stackrel{d}{=} \sum_{i=1}^{\infty} \lambda_i \int_0^1 V_i^2(x) dx.$$

Thus, by the continuous mapping theorem,

$$R_N = \iint (Z_N(x, t))^2 dx dt \xrightarrow{D} \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx,$$

which proves the desired result. \square

2.5.3 Proof of Theorem 11

The key fact needed in the proof is the consistency of the sample eigenvalues $\hat{\lambda}_i$ and eigenfunctions $\hat{\phi}_i$. The required result, stated in (2.5.19), follows fairly directly from (2.2.4). However, the verification that (2.2.4) holds for the kernel estimator (2.2.14) is not trivial. The required result can be stated as follows.

Theorem 16. *Suppose Assumption 8 holds with $\delta = 0$ and $\kappa = 2$. If H_0 and Assumption 9 hold, then relation (2.2.4) holds.*

Observe that assuming that relation (2.2.3) in Assumption 8 holds with $\delta = 0$ and $\kappa = 2$ weakens the universal assumption that it holds with some $\delta > 0$ and $\kappa > 2 + \delta$.

We first present the proof of Theorem 11, which uses Theorem 16, and then turn to a rather technical proof of Theorem 16.

PROOF OF THEOREM 11: If Assumptions 8, 9, condition (2.2.7) and H_0 hold, then

$$\max_{1 \leq i \leq d} |\hat{\lambda}_i - \lambda_i| = o_p(1) \quad \text{and} \quad \max_{1 \leq i \leq d} \|\hat{\phi}_i - \hat{c}_i \phi_i\| = o_p(1), \quad (2.5.19)$$

where $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_d$ are unobservable random signs defined as $\hat{c}_i = \text{sign}(\langle \hat{\phi}_i, \phi_i \rangle)$. Indeed, Theorem 16 states that relation (2.2.4) holds under H_0 and Assumptions 8 and 9. Relations (2.5.19) follow from (2.2.4) and Lemmas 2.2. and 2.3 of Horváth and Kokoszka (2012) which state that the differences of the eigenvalues and eigenfunctions are bounded by the Hilbert–Schmidt norm of the difference of the corresponding operators.

Using (2.5.1), it is easy to see that for all N

$$\{\langle \Gamma_N^0(x, \cdot), \phi_i \rangle, 0 \leq x \leq 1, 1 \leq i \leq d\} \stackrel{\mathcal{D}}{=} \{\sqrt{\lambda_i} V_i(x), 0 \leq x \leq 1, 1 \leq i \leq d\}. \quad (2.5.20)$$

We first show that

$$\sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{c}_i \phi_i \rangle| \xrightarrow{P} 0. \quad (2.5.21)$$

By the Cauchy-Schwarz inequality and Lemma 7, we know

$$\sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot) - \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle| \leq \sup_{0 \leq x \leq 1} \|Z_N(x, \cdot) - \Gamma_N^0(x, \cdot)\| = o_p(1).$$

Again by the Cauchy-Schwarz inequality and (2.5.19), we have

$$\sup_{0 \leq x \leq 1} |\langle \Gamma_N^0(x, \cdot), \hat{\phi}_i - \hat{c}_i \phi_i \rangle| \leq \sup_{0 \leq x \leq 1} \|\Gamma_N^0(x, \cdot)\| \|\hat{\phi}_i - \hat{c}_i \phi_i\| = o_p(1).$$

Then using the triangle inequality and inner product properties,

$$\begin{aligned} & \sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{c}_i \phi_i \rangle| \\ &= \sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle + \langle \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{c}_i \phi_i \rangle| \\ &\leq \sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle| + \sup_{0 \leq x \leq 1} |\langle \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle - \langle \Gamma_N^0(x, \cdot), \hat{c}_i \phi_i \rangle| \\ &= \sup_{0 \leq x \leq 1} |\langle Z_N(x, \cdot) - \Gamma_N^0(x, \cdot), \hat{\phi}_i \rangle| + \sup_{0 \leq x \leq 1} |\langle \Gamma_N^0(x, \cdot), \hat{\phi}_i - \hat{c}_i \phi_i \rangle| \\ &= o_p(1), \end{aligned}$$

which proves relation (2.5.21). Thus by Theorem 14, (2.5.19), (2.5.21), (2.5.20) and the continuous mapping theorem,

$$R_N^0 = \sum_{i=1}^d \frac{1}{\hat{\lambda}_i} \int \langle Z_N(x, \cdot), \hat{\phi}_i \rangle^2 dx \xrightarrow{d} \sum_{i=1}^d \int V_i^2(x) dx.$$

□

PROOF OF THEOREM 16: Recall definitions of the kernels c and \hat{c} given, respectively, in (2.2.5) and (2.2.14). The claim will follow if we can show that

$$\iint \{\hat{\gamma}_0(t, s) - E[\eta_0(t)\eta_0(s)]\}^2 dt ds = o_P(1) \quad (2.5.22)$$

and

$$\iint \left\{ \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \hat{\gamma}_i(t, s) - \sum_{i \geq 1} E[\eta_0(s)\eta_i(t)] \right\}^2 dt ds = o_P(1). \quad (2.5.23)$$

These relations are established in a sequence of Lemmas which split the argument by isolating the terms related to the estimation of trend from those related to the autocovariances of the η_i . The latter terms were treated in Horváth et al. (2013), so the present proof focuses on the extra terms appearing in our context.

Lemma 8. *Under model (2.2.1), the following relation holds*

$$\begin{aligned} \hat{\gamma}_0(t, s) &= \frac{1}{N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t)) (\eta_i(s) - \bar{\eta}(s)) \\ &\quad - \frac{1}{Ns_N} \left\{ \sum_{i=1}^N \left(i - \frac{N+1}{2}\right) \eta_i(t) \right\} \left\{ \sum_{i=1}^N \left(i - \frac{N+1}{2}\right) \eta_i(s) \right\}. \end{aligned}$$

Proof. Observe that $\sum_{i=1}^N \left(i - \frac{N+1}{2}\right) = 0$, and so

$$\sum_{i=1}^N \left(i - \frac{N+1}{2}\right) i = \sum_{i=1}^N \left(i - \frac{N+1}{2}\right)^2 = s_N.$$

Also recall that

$$\hat{\xi}(t) - \xi(t) = \frac{1}{s_N} \sum_{i=1}^N \left(i - \frac{N+1}{2}\right) \eta_i(t) = \frac{1}{s_N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t)) \left(i - \frac{N+1}{2}\right).$$

We can express the residuals $e_n(t)$ as

$$\begin{aligned}
e_i(t) &= (X_i(t) - \bar{X}(t)) - \hat{\xi}(t) \left(i - \frac{N+1}{2} \right) \\
&= \mu(t) + i\xi(t) + \eta_i(t) - \frac{1}{N} \sum_{i=1}^N (\mu(t) + i\xi(t) + \eta_i(t)) - \hat{\xi}(t) \left(i - \frac{N+1}{2} \right) \\
&= (\eta_i(t) - \bar{\eta}(t)) - (\hat{\xi}(t) - \xi(t)) \left(i - \frac{N+1}{2} \right).
\end{aligned}$$

Then, by the above relations,

$$\begin{aligned}
\hat{\gamma}_0(t, s) &= \frac{1}{N} \sum_{i=1}^N e_i(t) e_i(s) \\
&= \frac{1}{N} \sum_{i=1}^N \left[(\eta_i(t) - \bar{\eta}(t)) - (\hat{\xi}(t) - \xi(t)) \left(i - \frac{N+1}{2} \right) \right] \left[(\eta_i(s) - \bar{\eta}(s)) - (\hat{\xi}(s) - \xi(s)) \left(i - \frac{N+1}{2} \right) \right] \\
&= \frac{1}{N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t)) (\eta_i(s) - \bar{\eta}(s)) \\
&\quad - (\hat{\xi}(s) - \xi(s)) \frac{1}{N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t)) \left(i - \frac{N+1}{2} \right) \\
&\quad - (\hat{\xi}(t) - \xi(t)) \frac{1}{N} \sum_{i=1}^N (\eta_i(s) - \bar{\eta}(s)) \left(i - \frac{N+1}{2} \right) \\
&\quad + (\hat{\xi}(t) - \xi(t)) (\hat{\xi}(s) - \xi(s)) \frac{1}{N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right)^2 \\
&= \frac{1}{N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t)) (\eta_i(s) - \bar{\eta}(s)) \\
&\quad - \frac{1}{Ns_N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(s) \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(t) \\
&\quad - \frac{1}{Ns_N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(t) \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(s) \\
&\quad + \frac{1}{N} \frac{1}{s_N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(t) \frac{1}{s_N} \sum_{i=1}^N \left(i - \frac{N+1}{2} \right) \eta_i(s) s_N.
\end{aligned}$$

The claim thus follows because the last two terms cancel. □

To lighten the notation, in the remainder of this section we set

$$k_i = i - \frac{N+1}{2}.$$

Lemma 9. *Relation (2.5.22) holds under the assumptions of Theorem 16.*

Proof. We must show that $\|\hat{\gamma}_0 - \gamma_0\| \xrightarrow{P} 0$, where $\gamma_0(t, s) = E[\eta_0(t)\eta_0(s)]$ and the norm is in $L^2([0, 1] \times [0, 1])$. We will use the decomposition of Lemma 8, i.e.

$$\hat{\gamma}_0(t, s) = \tilde{\gamma}_0(t, s) - v_N(t, s)$$

where

$$\tilde{\gamma}_0(t, s) = \frac{1}{N} \sum_{i=1}^N (\eta_i(t) - \bar{\eta}(t))(\eta_i(s) - \bar{\eta}(s))$$

and

$$v_N(t, s) = \frac{1}{Ns_N} \left\{ \sum_{i=1}^N k_i \eta_i(t) \right\} \left\{ \sum_{i=1}^N k_i \eta_i(s) \right\}.$$

It will be enough to show that

$$\|\tilde{\gamma}_0 - \gamma_0\| \xrightarrow{P} 0 \quad \text{and} \quad \|v_N\| \xrightarrow{P} 0.$$

The first convergence is the consistency of the sample covariance function which was proven by Horváth et al. (2013). The remainder of the proof is devoted to the verification that $\|v_N\| \xrightarrow{P} 0$.

Since

$$\|v_N\|^2 = \frac{1}{N^2 s_N^2} \int \left\{ \sum_{i=1}^N k_i \eta_i(t) \right\}^2 dt \int \left\{ \sum_{i=1}^N k_i \eta_i(s) \right\}^2 ds,$$

we must show that

$$\|v_N\| = \frac{1}{Ns_N} \left\| \sum_{i=1}^N k_i \eta_i \right\|^2 \xrightarrow{P} 0, \tag{2.5.24}$$

where the norm in (2.5.24) is in $L^2([0, 1])$. Using diagonal summation, we get

$$\begin{aligned} \left\| \sum_{i=1}^N k_i \eta_i \right\|^2 &= \sum_{i=1}^N \sum_{j=1}^N k_i k_j \langle \eta_i, \eta_j \rangle \\ &= \sum_{i=1}^N k_i^2 \|\eta_i\|^2 + 2 \sum_{l=1}^{N-1} \sum_{i=1}^{N-l} k_i k_{i+l} \langle \eta_i, \eta_{i+l} \rangle. \end{aligned}$$

Since $\sum_{i=1}^N k_i^2 = s_N$, we obtain

$$E\left[\sum_{i=1}^N k_i^2 \|\eta_i\|^2\right] = s_N E\|\eta_0\|^2. \quad (2.5.25)$$

We now turn to the expectation of the second term

$$E\left[\sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} \langle \eta_i, \eta_{i+\ell} \rangle\right] = \sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} E\langle \eta_i, \eta_{i+\ell} \rangle.$$

By stationarity, $E\langle \eta_0, \eta_\ell \rangle = E\langle \eta_{n-\ell}, \eta_n \rangle$. By Assumption 8, η_n can be approximated by $\eta_{n,\ell}$ which is independent of $\eta_{n-\ell}$. Therefore

$$E\langle \eta_{n-\ell}, \eta_n \rangle = E\langle \eta_{n-\ell}, \eta_{n,\ell} \rangle + E\langle \eta_{n-\ell}, \eta_n - \eta_{n,\ell} \rangle = E\langle \eta_{n-\ell}, \eta_n - \eta_{n,\ell} \rangle,$$

because $E\langle \eta_{n-\ell}, \eta_{n,\ell} \rangle = 0$. Observe that, using the Cauchy-Schwarz inequality twice,

$$\begin{aligned} |E\langle \eta_{n-\ell}, \eta_n - \eta_{n,\ell} \rangle| &\leq E\|\eta_{n-\ell}\| \|\eta_n - \eta_{n,\ell}\| \\ &\leq \{E\|\eta_{n-\ell}\|^2\}^{1/2} \{E\|\eta_n - \eta_{n,\ell}\|^2\}^{1/2}. \end{aligned}$$

It follows that

$$\begin{aligned} \left| E\left[\sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} \langle \eta_i, \eta_{i+\ell} \rangle\right] \right| &\leq \sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} |E\langle \eta_i, \eta_{i+\ell} \rangle| \\ &\leq \{E\|\eta_{n-\ell}\|^2\}^{1/2} \sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} \{E\|\eta_n - \eta_{n,\ell}\|^2\}^{1/2} \\ &\leq CN^3 \sum_{\ell=1}^{\infty} \{E\|\eta_n - \eta_{n,\ell}\|^2\}^{1/2}. \end{aligned}$$

Thus assuming

$$\sum_{\ell=1}^{\infty} \{E\|\eta_n - \eta_{n,\ell}\|^2\}^{1/2} < \infty,$$

we obtain

$$\left| E\left[\sum_{l=1}^{N-1} \sum_{i=1}^{N-\ell} k_i k_{i+\ell} \langle \eta_i, \eta_{i+\ell} \rangle\right] \right| \leq O(N^3) = O(s_N). \quad (2.5.26)$$

Combining (2.5.25) and (2.5.26), we see that

$$E\|v_N\| = \frac{1}{Ns_N} E\left\| \sum_{i=1}^N k_i \eta_i \right\|^2 = \frac{1}{Ns_N} O(s_N) = O(N^{-1}). \quad (2.5.27)$$

This proves relation (2.5.24). \square

Lemma 10. *Relation (2.5.23) holds under the assumptions of Theorem 16.*

Proof. Under H_0 ,

$$\begin{aligned} \hat{\gamma}_i(t, s) &= \frac{1}{N} \sum_{j=i+1}^N e_j(t) e_{j-i}(s) \\ &= \frac{1}{N} \sum_{j=i+1}^N \left[(\eta_j(t) - \bar{\eta}(t)) - (\hat{\xi}(t) - \xi(t)) k_j \right] \cdot \left[(\eta_{j-i}(s) - \bar{\eta}(s)) - (\hat{\xi}(s) - \xi(s)) k_{j-i} \right] \\ &= \frac{1}{N} \sum_{j=i+1}^N (\eta_j(t) - \bar{\eta}(t)) (\eta_{j-i}(s) - \bar{\eta}(s)) \\ &\quad - \frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(s) \sum_{j=i+1}^N k_{j-i} (\eta_j(t) - \bar{\eta}(t)) \\ &\quad - \frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(t) \sum_{j=i+1}^N k_j (\eta_{j-i}(s) - \bar{\eta}(s)) \\ &\quad + \frac{1}{Ns_N^2} \sum_{l=1}^N k_l \eta_l(s) \sum_{m=1}^N k_m \eta_m(t) \sum_{j=i+1}^N k_j k_{j-i}. \end{aligned}$$

Set

$$\bar{\gamma}_i(s, t) = \frac{1}{N} \sum_{j=i+1}^N (\eta_j(t) - \bar{\eta}(t)) (\eta_{j-i}(s) - \bar{\eta}(s)).$$

Then

$$\begin{aligned} \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \hat{\gamma}_i(t, s) &= \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \bar{\gamma}_i(t, s) \\ &\quad - \frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(s) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_{j-i} (\eta_j(t) - \bar{\eta}(t)) \\ &\quad - \frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(t) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_j (\eta_{j-i}(s) - \bar{\eta}(s)) \\ &\quad + \frac{1}{Ns_N^2} \sum_{l=1}^N k_l \eta_l(s) \sum_{m=1}^N k_m \eta_m(t) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_j k_{j-i}. \end{aligned}$$

Thus, in order to prove Lemma 10, we must establish all of the following relations:

$$\iint \left(\sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \bar{\gamma}_i(t, s) - \sum_{i \geq 1} E[\eta_0(s)\eta_i(t)] \right)^2 dt ds = o_P(1); \quad (2.5.28)$$

$$\iint \left(\frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(s) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_{j-i} (\eta_j(t) - \bar{\eta}(t)) \right)^2 ds dt = o_P(1); \quad (2.5.29)$$

$$\iint \left(\frac{1}{Ns_N} \sum_{l=1}^N k_l \eta_l(t) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_j (\eta_{j-i}(s) - \bar{\eta}(s)) \right)^2 ds dt = o_P(1); \quad (2.5.30)$$

$$\iint \left(\frac{1}{Ns_N^2} \sum_{l=1}^N k_l \eta_l(s) \sum_{m=1}^N k_m \eta_m(t) \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_j k_{j-i} \right)^2 ds dt = o_P(1). \quad (2.5.31)$$

Relation (2.5.28) has been established by Horváth et al. (2013), so it remains to deal with the remaining three relations which are due to the estimation of the trend.

Relations (2.5.29) and (2.5.30) follow by application of similar arguments, so we display only the verification of (2.5.29). Observe that the left-hand side of (2.5.29) is equal to

$$\left(\frac{1}{Ns_N} \left\| \sum_{l=1}^N k_l \eta_l \right\|^2 \right) \left(\frac{1}{Ns_N} \left\| \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_{j-i} (\eta_j - \bar{\eta}) \right\|^2 \right) \quad (2.5.32)$$

A bound for the expectation of the first factor is given in (2.5.27). In the second factor, the centering by $\bar{\eta}$ contributes asymptotically negligible terms, so this factor has the same order as

$$F_N = \frac{1}{Ns_N} \left\| \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_{j-i} \eta_j \right\|^2.$$

Observe that

$$\begin{aligned} E \left\| \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \sum_{j=i+1}^N k_{j-i} \eta_j \right\|^2 &= \sum_{i=1}^{N-1} \sum_{i'=1}^{N-1} K\left(\frac{i}{h}\right) K\left(\frac{i'}{h}\right) \sum_{l=1}^{N-1} \sum_{l'=1}^{N-1} k_l k_{l'} E\langle \eta_{i+l}, \eta_{i'+l'} \rangle \\ &= E \|\eta_0\|^2 \sum_{i=1}^{N-1} \sum_{i'=1}^{N-1} K\left(\frac{i}{h}\right) K\left(\frac{i'}{h}\right) \sum_{l=1}^{N-1} \sum_{l'=1}^{N-1} k_l k_{l'} \\ &= O(N^3) \left(\sum_{i=1}^{N-1} \left| K\left(\frac{i}{h}\right) \right| \right)^2 \end{aligned}$$

Notice that

$$\sum_{i=1}^{N-1} \left| K\left(\frac{i}{h}\right) \right| = h \sum_{i=1}^{N-1} \frac{1}{h} \left| K\left(\frac{i}{h}\right) \right| = O\left(h \int_0^1 |k(u)| du\right),$$

we see that $EF_N = O(N^{-4}N^3h^2) = O(N^{-1}h^2)$. Thus, (2.5.32) is of the order of (2.5.27),

$$O_P(N^{-1})O_P(N^{-1}h^2) = O_P\left(\frac{h^2}{N^2}\right) = o_P(1).$$

We now turn to the verification of (2.5.31) whose left-hand side can be written as

$$\begin{aligned} & \left\{ \int \left(\frac{1}{s_N} \sum_{l=1}^N k_l \eta(t) \right)^2 dt \right\}^2 \left\{ \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \frac{1}{N} \sum_{j=i+1}^N k_j k_{j-i} \right\}^2 \\ &= \frac{N^2}{S_N^2} \|v_N\|^2 \left\{ \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right) \frac{1}{N} \sum_{j=i+1}^N k_j k_{j-i} \right\}^2. \end{aligned}$$

Using (2.5.27), we see that the order of the above expression is

$$O\left(\frac{N^2}{N^6}\right)O_P(N^{-2})\{O_P(hN^2)\}^2 = O(N^{-4})O_P(N^{-2})O(h^2N^4) = O_P\left(\frac{h^2}{N^2}\right) = o_P(1).$$

This completes the proof of Lemma 10. □

2.5.4 Proof of Theorem 12

The proof of Theorem 12 is constructed from several lemmas.

Lemma 11. *Under the alternative (2.2.2), for the functional slope estimate $\hat{\xi}$ defined by (2.2.9),*

$$\begin{aligned} N^{3/2}(\hat{\xi}(t) - \xi(t)) &= \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\} \\ &+ \frac{1}{N^{-3}s_N} \left\{ \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, t\right) - \left(\frac{N+1}{2N} \right) \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \right\}, \end{aligned}$$

where $V_N(x, t)$ is the partial sum process of the errors η_n defined in (2.3.2) and $Y_N(x, t)$ is the partial sum process of the random walk errors u_n defined by

$$Y_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} u_n(t). \quad (2.5.33)$$

Proof. Recall that $k_n = n - (N+1)/2$, $\sum_{n=1}^N k_n = 0$, $\sum_{n=1}^N nk_n = s_N$. Therefore, under alternative

(2.2.2),

$$\begin{aligned}
\hat{\xi}(t) - \xi(t) &= \frac{1}{s_N} \sum_{n=1}^N k_n X_n(t) - \xi(t) \\
&= \frac{1}{s_N} \sum_{n=1}^N k_n \left(\mu(t) + n\xi(t) + \sum_{i=1}^n u_i(t) + \eta_n(t) \right) - \xi(t) \\
&= \frac{1}{s_N} \left\{ \mu(t) \sum_{n=1}^N k_n + \xi(t) \sum_{n=1}^N nk_n \right\} \\
&\quad + \frac{1}{s_N} \sum_{n=1}^N k_n \eta_n(t) - \xi(t) + \frac{1}{s_N} \sum_{n=1}^N k_n \sum_{i=1}^n u_i(t) \\
&= \frac{1}{s_N} \xi(t) s_N + \frac{1}{s_N} \sum_{n=1}^N k_n \eta_n(t) - \xi(t) + \frac{1}{s_N} \sum_{n=1}^N k_n \sum_{i=1}^n u_i(t) \\
&= \frac{1}{s_N} \sum_{n=1}^N k_n \eta_n(t) + \frac{1}{s_N} \sum_{n=1}^N k_n \sum_{i=1}^n u_i(t).
\end{aligned}$$

Using definition (2.5.33),

$$\begin{aligned}
\frac{N^{-3/2}}{N^{-3}s_N} \sum_{n=1}^N k_n \sum_{i=1}^n u_i(t) &= \frac{N^{-3/2}}{N^{-3}s_N} \left\{ \sum_{n=1}^N n \sum_{i=1}^n u_i(t) - \left(\frac{N+1}{2} \right) \sum_{n=1}^N \sum_{i=1}^n u_i(t) \right\} \\
&= \frac{1}{N^{-3}s_N} \left\{ \sum_{n=1}^N \frac{n}{N} \frac{1}{\sqrt{N}} \sum_{i=1}^n u_i(t) - \left(\frac{N+1}{2N} \right) \sum_{n=1}^N \frac{1}{\sqrt{N}} \sum_{i=1}^n u_i(t) \right\} \\
&= \frac{1}{N^{-3}s_N} \left\{ \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, t\right) - \left(\frac{N+1}{2N} \right) \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \right\}.
\end{aligned}$$

The claim follows from the above relations and Lemma 3. □

Lemma 12. *Under the alternative, $Z_N(x, t)$ defined in (2.3.4) can be expressed as*

$$\begin{aligned}
Z_N(x, t) &= V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) \\
&\quad - \frac{1}{2} \frac{1}{N^{-3}s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\} \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&\quad + \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \\
&\quad - \frac{1}{2} \frac{1}{N^{-3}s_N} \left\{ \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, t\right) - \left(\frac{N+1}{2N} \right) \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \right\} \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}$$

Proof. Under H_A , the partial sum process $S_N(x, t)$ can be expressed as

$$\begin{aligned}
S_N(x, t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t) \\
&= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \left(\mu(t) + n\xi(t) + \sum_{i=1}^n u_i(t) + \eta_n(t) \right) \\
&= \left(\frac{1}{\sqrt{N}} \mu(t) \lfloor Nx \rfloor + \frac{1}{\sqrt{N}} \xi(t) \frac{\lfloor Nx \rfloor (\lfloor Nx \rfloor + 1)}{2} + \sum_{n=1}^{\lfloor Nx \rfloor} \frac{1}{\sqrt{N}} \sum_{i=1}^n u_i(t) + \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} \eta_n(t) \right) \\
&= \left(\frac{1}{\sqrt{N}} \mu(t) \lfloor Nx \rfloor + \frac{1}{\sqrt{N}} \xi(t) \frac{\lfloor Nx \rfloor (\lfloor Nx \rfloor + 1)}{2} + \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) + V_N(x, t) \right).
\end{aligned}$$

By Lemma 8,

$$\begin{aligned}
Z_N(x, t) &= S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t) - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&= \left(\frac{1}{\sqrt{N}} \mu(t) \lfloor Nx \rfloor + \frac{1}{\sqrt{N}} \xi(t) \frac{\lfloor Nx \rfloor (\lfloor Nx \rfloor + 1)}{2} + \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) + V_N(x, t) \right) \\
&\quad - \frac{\lfloor Nx \rfloor}{N} \left(\frac{1}{\sqrt{N}} \mu(t) N + \frac{1}{\sqrt{N}} \xi(t) \frac{N(N+1)}{2} + \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) + V_N(1, t) \right) \\
&\quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}$$

Therefore,

$$\begin{aligned}
Z_N(x, t) &= \frac{1}{\sqrt{N}} \mu(t) \lfloor Nx \rfloor - \frac{1}{\sqrt{N}} \mu(t) \lfloor Nx \rfloor + \frac{1}{2} N^{3/2} \xi(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&\quad + \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) + V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) \\
&\quad - \frac{1}{2} N^{3/2} \hat{\xi}(t) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \\
&= \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) + V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) \\
&\quad - \frac{1}{2} N^{3/2} (\hat{\xi}(t) - \xi(t)) \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right).
\end{aligned}$$

Then reexpressing $N^{3/2}(\hat{\xi}(t) - \xi(t))$, we get the desired expression.

□

Since the u_i satisfy Assumption 8, an analog of Theorem 15 holds, i.e. there exist Gaussian processes Λ_N equal in distribution to

$$\Lambda(x, t) = \sum_{i=1}^{\infty} \tau_i^{1/2} W_i(x) \psi_i(t), \quad (2.5.34)$$

where τ_i, ψ_i are, respectively, the eigenvalues and the eigenfunctions of the kernel (2.3.7). Moreover, for the partial sum process Y_N defined by (2.5.33),

$$l_n = \sup_{0 \leq x \leq 1} \|Y_N(x, \cdot) - \Lambda_N(x, \cdot)\| = o_p(1). \quad (2.5.35)$$

Lemma 13. *Under the alternative, the following convergence holds*

$$\sup_{0 \leq x \leq 1} \|N^{-1} Z_N^A(x, \cdot) - \Delta_N^0(x, \cdot)\| \xrightarrow{P} 0,$$

where the processes $Z_N^A(\cdot, \cdot)$ and $\Delta_N^0(\cdot, \cdot)$ are respectively defined by

$$\begin{aligned} Z_N^A(x, t) &= \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) - \frac{\lfloor Nx \rfloor}{N} \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \\ &\quad - \frac{1}{2} \frac{1}{N^{-3} s_N} \left\{ \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, t\right) - \left(\frac{N+1}{2N}\right) \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \right\} \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \end{aligned} \quad (2.5.36)$$

and

$$\Delta_N^0(x, t) = \int_0^x \Lambda_N(y, t) dy + (3x^2 - 4x) \int_0^1 \Lambda_N(y, t) dy + (-6x^2 + 6x) \int_0^1 y \Lambda_N(y, t) dy. \quad (2.5.37)$$

Proof. Notice that $N^{-1} Z_N^A(x, t)$ can be expressed as

$$\frac{1}{N} Z_N^A(x, t) = \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, t\right) \frac{1}{N} + f_N(x) \sum_{n=1}^N Y_N\left(\frac{n}{N}, t\right) \frac{1}{N} + g_N(x) \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, t\right) \frac{1}{N}$$

where

$$f_N(x) = \frac{1}{2N^{-3} s_N} \left(\frac{N+1}{2N} \right) \frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) - \frac{\lfloor Nx \rfloor}{N}$$

and

$$g_N(x) = \frac{1}{2N^{-3} s_N} \frac{\lfloor Nx \rfloor}{N} \left(1 - \frac{\lfloor Nx \rfloor}{N} \right).$$

By the triangle inequality,

$$\begin{aligned} \|N^{-1}Z_N(x, \cdot) - \Delta_N^0(x, \cdot)\| &\leq \left\| \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - \int_0^x \Lambda_N(y, \cdot) dy \right\| \\ &+ \left\| f_N(x) \sum_{n=1}^N Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - (3x^2 - 4x) \int_0^1 \Lambda_N(y, \cdot) dy \right\| \\ &+ \left\| g_N(x) \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - (-6x^2 + 6x) \int_0^1 y \Lambda_N(y, \cdot) dy \right\|. \end{aligned}$$

Thus, Lemma 1 will be proven once we have established the following relations:

$$\sup_{0 \leq x \leq 1} \left\| \sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - \int_0^x \Lambda_N(y, \cdot) dy \right\| \xrightarrow{P} 0; \quad (2.5.38)$$

$$\sup_{0 \leq x \leq 1} \left\| f_N(x) \sum_{n=1}^N Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - (3x^2 - 4x) \int_0^1 \Lambda_N(y, \cdot) dy \right\| \xrightarrow{P} 0; \quad (2.5.39)$$

$$\sup_{0 \leq x \leq 1} \left\| g_N(x) \sum_{n=1}^N \frac{n}{N} Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} - (-6x^2 + 6x) \int_0^1 y \Gamma_N(y, \cdot) dy \right\| \xrightarrow{P} 0. \quad (2.5.40)$$

Relations (2.5.39) and (2.5.40) follow from arguments fully analogous to those used in the proof of

Lemma 7. The verification of (2.5.38) is not difficult either. Observe that

$$\sum_{n=1}^{\lfloor Nx \rfloor} Y_N\left(\frac{n}{N}, \cdot\right) \frac{1}{N} = \int_0^x Y_N(y, \cdot) dy - r_N(x),$$

where

$$r_N(x) = \int_{\lfloor Nx \rfloor / N}^x Y_N(y, \cdot) dy.$$

Relation

$$\sup_{0 \leq x \leq 1} \left\| \int_0^x Y_N(y, \cdot) dy - \int_0^x \Lambda_N(y, \cdot) dy \right\| = o_p(1)$$

follows from (2.5.35) and the contractive property of the integral, which also implies that

$$\begin{aligned} \sup_{0 \leq x \leq 1} \|r_N(x)\| &\leq \sup_{0 \leq x \leq 1} \int_{\lfloor Nx \rfloor / N}^x \|Y_N(y, \cdot)\| dy \\ &\leq \frac{1}{N} \sup_{0 \leq x \leq 1} \|Y_N(y, \cdot)\| = O_p(N^{-1}). \end{aligned}$$

This completes the proof of Lemma 13. □

Lemma 14. *Under the alternative, the following convergence holds*

$$\sup_{0 \leq x \leq 1} \|N^{-1}Z_N(x, \cdot) - \Delta_N^0(x, \cdot)\| \xrightarrow{P} 0,$$

where the process $Z_N(\cdot, \cdot)$ is defined in (2.3.4) and the process $\Delta_N^0(\cdot, \cdot)$ in (2.5.37).

Proof. Under the alternative, Lemma 12 implies that, $Z_N(x, t)$ can be expressed as

$$Z_N(x, t) = Z_N^0(x, t) + Z_N^A(x, t),$$

where

$$\begin{aligned} Z_N^0(x, t) &= V_N(x, t) - \frac{\lfloor Nx \rfloor}{N} V_N(1, t) \\ &\quad - \frac{1}{2} \frac{1}{N^{-3} s_N} \left\{ \left(\frac{N-1}{2N} \right) V_N(1, t) - \frac{1}{N} \sum_{k=1}^{N-1} V_N\left(\frac{k}{N}, t\right) \right\} \left(\frac{\lfloor Nx \rfloor}{N} \left(\frac{\lfloor Nx \rfloor}{N} - 1 \right) \right) \end{aligned}$$

and $Z_N^A(x, t)$ is defined by (2.5.36). Notice under the null hypothesis, Lemma 4 implies that

$Z_N^0(x, t) = Z_N(x, t)$. Hence from Lemma 7, $\sup_{0 \leq x \leq 1} \|Z_N^0(x, \cdot) - \Gamma_N^0(x, \cdot)\| \xrightarrow{P} 0$, implying

$$\sup_{0 \leq x \leq 1} \left\| \frac{1}{N} Z_N^0(x, \cdot) \right\| = \frac{1}{N} O_p(1).$$

Thus the claim follows from Lemma 13 and the triangle inequality. \square

Lemma 15. *Consider the process $\Lambda(\cdot, \cdot)$ defined by (2.5.34) and set*

$$\Delta^0(x, t) = \int_0^x \Lambda(y, t) dy + (3x^2 - 4x) \int_0^1 \Lambda(y, t) dy + (-6x^2 + 6x) \int_0^1 y \Lambda(y, t) dy. \quad (2.5.41)$$

Then

$$\int_0^1 \|\Delta^0(x, \cdot)\|^2 dx = \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x) dx,$$

where τ_1, τ_2, \dots are eigenvalues of the long-run covariance function of the u_i , i.e. (2.3.7), and

$\Delta_1, \Delta_2, \dots$ are independent copies of the process Δ defined in (2.3.6).

Proof. Expansion (2.5.34) implies that

$$\begin{aligned}
\Delta^0(x, t) &= \int_0^x \Lambda(y, t) dy + (3x^2 - 4x) \int_0^1 \Lambda(y, t) dy + (-6x^2 + 6x) \int_0^1 y \Lambda(y, t) dy \\
&= \int_0^x \sum_{i=1}^{\infty} \sqrt{\tau_i} W_i(y) \psi_i(t) dy + (3x^2 - 4x) \int_0^1 \sum_{i=1}^{\infty} \sqrt{\tau_i} W_i(y) \psi_i(t) dy \\
&\quad + (-6x^2 + 6x) \int_0^1 y \sum_{i=1}^{\infty} \sqrt{\tau_i} W_i(y) \psi_i(t) dy \\
&= \sum_{i=1}^{\infty} \sqrt{\tau_i} \left\{ \int_0^x W_i(y) dy + (3x^2 - 4x) \int_0^1 W_i(y) dy \right. \\
&\quad \left. + (-6x^2 + 6x) \int_0^1 y W_i(y) dy \right\} \psi_i(t) \\
&= \sum_{i=1}^{\infty} \sqrt{\tau_i} \Delta_i(x) \psi_i(t).
\end{aligned}$$

The claim then follows from the orthonormality of the eigenfunctions ϕ_i . □

PROOF OF THEOREM 12: Recall that the test statistic R_N is defined by $R_N = \iint Z_N^2(x, t) dx dt$, with Z_N defined by (2.3.4). We want to show that under the alternative model (2.2.2),

$$\frac{1}{N^2} R_N \xrightarrow{\mathcal{D}} \sum_{i=1}^{\infty} \tau_i \int_0^1 \Delta_i^2(x) dx,$$

where $\Delta_1, \Delta_2, \dots$ are independent copies of the process defined by (2.3.6) and τ_1, τ_2, \dots are the eigenvalues of the long-run covariance kernel (2.3.7). By Lemma 14,

$$\rho(N^{-1} Z_N(x, \cdot), \Delta_N^0(x, \cdot)) = \sup_{0 \leq x \leq 1} \|N^{-1} Z_N(x, \cdot) - \Delta_N^0(x, \cdot)\| \xrightarrow{P} 0.$$

By construction, $\Delta_N^0 \stackrel{d}{=} \Delta^0$, so Theorem 14 implies that $N^{-1} Z_N \xrightarrow{d} \Delta^0$. By Lemma 15,

$$\iint (\Delta^0(x, t))^2 dx dt \stackrel{d}{=} \sum_{i=1}^{\infty} \lambda_i \int_0^1 \Delta_i^2(x) dx.$$

Thus by the continuous mapping theorem,

$$\frac{1}{N^2} R_N = \iint (N^{-1} Z_N(x, t))^2 dx dt \xrightarrow{d} \sum_{i=1}^{\infty} \lambda_i \int \Delta_i^2(x) dx.$$

□

Chapter 3

TESTING TREND STATIONARITY OF FUNCTIONAL TIME SERIES WITH APPLICATION TO YIELD AND DAILY PRICE CURVES

3.1 Introduction

Many econometric and financial data sets take the form of a time series of curves, or functions. The best known and most extensively studied data of this form are yield curves. Even though they are observed at discrete maturities, in financial theory they are viewed as continuous functions, one function per month or per day. The yield curves can thus be viewed as time series of curves, functional time series. Other examples include intraday price, volatility or volume curves. Intraday price curves are smooth, volatility and volume curves are noisy and must be smoothed before they can be effectively treated as curves. As with scalar and vector valued time series, it is important to describe the random structure of a functional time series. A fundamental question, which has received a great deal of attention in econometric research, is whether the time series has a random walk, or unit root, component. The present paper addresses this issue in the context of functional time series by proposing extensions of the KPSS test of Kwiatkowski et al. (1992) and applying them to several data sets.

The work of Kwiatkowski et al. (1992) was motivated by the fact that unit root tests developed by Dickey and Fuller (1979, 1981), and Said and Dickey (1984) indicated that most aggregate economic series had a unit root. In these tests, the null hypothesis is that the series has a unit root. Since such tests have low power in samples of sizes occurring in many applications, Kwiatkowski et al. (1992) proposed that trend stationarity should be considered as the null hypothesis, and the unit root should be the alternative. Rejection of the null could then be viewed as convincing evidence in favor of the unit root hypothesis. It was soon realized that the KPSS test of Kwiatkowski

et al. (1992) has a much broader utility. For example, Lee and Schmidt (1996) and Giraitis et al. (2003) used it to detect long memory, with short memory as the null hypothesis; de Jong et al. (1997) developed a robust version of the KPSS test. The work of Lo (1991) is crucial because he observed that under temporal dependence, to obtain parameter-free limit null distributions, statistics similar to the KPSS statistic must be normalized by the long run variance rather than by the sample variance.

In the functional setting, the null hypothesis of trend stationarity is stated as follows:

$$H_0 : X_n(t) = \mu(t) + n\xi(t) + \eta_n(t), \quad (3.1.1)$$

where n is the serial number of the day in our applications, and t refers to “time” for each function. For example, for the intraday price curves, t is the time within a trading day, measured in minutes or at an even finer resolution. For the yield curves, t does not correspond to physical time but to time until expiration, the maturity horizon of a bond. The functions μ and ξ correspond, respectively, to the intercept and slope. The errors η_n are also functions which model random departures of the observed functions X_n from a deterministic model. Under the alternative, the model contains a random walk component:

$$H_A : X_n(t) = \mu(t) + n\xi(t) + \sum_{i=1}^n u_i(t) + \eta_n(t), \quad (3.1.2)$$

where u_1, u_2, \dots are mean zero identically distributed random functions.

Our approach to testing exploits the ideas of functional data analysis (FDA), mostly those related to functional principal component expansions; several monographs, e.g. Ramsay and Silverman (2005) and Horváth and Kokoszka (2012), explain them in detail. Application of FDA methodology in an econometric context is not new. Among others, Kargin and Onatski (2008) studied prediction of yield curves, Müller et al. (2011) considered functional modeling of volatility, Kokoszka et al. (2014) used a regression type model to explain the shapes of price curves. A contribution most closely related to the present work is that of Horváth et al. (2014) who developed

a test of level stationarity. Incorporating a possible trend leads to different limit distributions and more complex numerical implementations.

The remainder of the paper is organized as follows. After introducing the required concepts and notation in Section 3.2, we present in Section 3.3 the large sample results needed to construct the tests. The resulting testing procedures are described in Section 3.4. Section 3.5 presents their applications to data representing bond, equity, forex and commodity markets. In this last section, we also examine and discuss finite sample properties of the tests.

3.2 Preliminaries

To understand the construction of the tests in the setting of functional time series, we must introduce some notation and definitions. This is the objective of the present section.

All random functions and deterministic functional parameters μ and ξ are assumed to be elements of the Hilbert space $L^2 = L^2([0, 1])$ with the inner product $\langle f, g \rangle = \int_0^1 f(t)g(t)dt$. This means that the domain of all functional observations, e.g. of the daily price or yield curves, has been normalized to be the unit interval. If the limits of integration are omitted, integration is over the interval $[0, 1]$. All random functions are assumed to be square integrable, i.e., $E \|\eta_n\|^2 < \infty$, $E \|u_n\|^2 < \infty$, where the norm is generated by the inner product, i.e. $\|f\|^2 = \int f^2(t)dt$.

Kwiatkowski et al. (1992) assumed that the errors η_n are iid, but subsequent research extended their work to errors which form a stationary time series, see, e.g., Giraitis et al. (2003) and the references therein. In the case of scalar observations, temporal dependence can be quantified in many ways, e.g., via structural, mixing or cumulant conditions, and a large number of asymptotic results established under such assumptions can be used. For functional time series, the corresponding results are much fewer and fall into two categories: 1) those derived assuming a linear, ARMA type, structure, see, e.g., Bosq (2000); 2) those assuming a nonlinear moving average representation (Bernoulli shifts) with the decay of dependence specified by a moment condition. We have established the asymptotic validity of our tests assuming very general conditions falling into the

second category. Detailed formulations of these conditions are presented in Kokoszka and Young (2015a). In essence, the error functions η_n and u_i need not be iid, but merely must form stationary and weakly dependent sequences.

Next we define the long-run covariance function of the errors η_n and its estimator. The long-run covariance function is defined as

$$c(t, s) = E\eta_0(t)\eta_0(s) + \sum_{i=1}^{\infty} (E\eta_0(t)\eta_i(s) + E\eta_0(s)\eta_i(t)). \quad (3.2.1)$$

The series defining the function $c(t, s)$ converges in $L^2([0, 1] \times [0, 1])$, see Horváth et al. (2013). The function $c(t, s)$ is positive definite. Therefore there exist eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, and orthonormal eigenfunctions $\phi_i(t)$, $0 \leq t \leq 1$, satisfying

$$\lambda_i \phi_i(t) = \int c(t, s) \phi_i(s) ds, \quad 0 \leq i \leq \infty. \quad (3.2.2)$$

The eigenvalues λ_i play a crucial role in our tests. They are estimated by the sample, or empirical, eigenvalues defined by

$$\hat{\lambda}_i \hat{\phi}_i(t) = \int \hat{c}(t, s) \hat{\phi}_i(s) ds, \quad 0 \leq i \leq N, \quad (3.2.3)$$

where $\hat{c}(\cdot, \cdot)$ is an estimator of (3.2.1), and N is the sample size of the functional time series. We use a kernel estimator similar to that introduced by Horváth et al. (2013), but with suitably defined residuals in place of the centered observations X_n . To define model residuals, consider the least squares estimators of the functional parameters $\xi(t)$ and $\mu(t)$ in model (3.1.1):

$$\hat{\xi}(t) = \frac{1}{s_N} \sum_{n=1}^N \left(n - \frac{N+1}{2} \right) X_n(t) \quad (3.2.4)$$

with

$$s_N = \sum_{n=1}^N \left(n - \frac{N+1}{2} \right)^2 \quad (3.2.5)$$

and

$$\hat{\mu}(t) = \bar{X}(t) - \hat{\xi}(t) \left(\frac{N+1}{2} \right). \quad (3.2.6)$$

The functional residuals are therefore

$$e_n(t) = (X_n(t) - \bar{X}(t)) - \hat{\xi}(t)\left(n - \frac{N+1}{2}\right), \quad (3.2.7)$$

where $1 \leq n \leq N$. Defining their empirical autocovariances by

$$\hat{\gamma}_i(t, s) = \frac{1}{N} \sum_{j=i+1}^N e_j(t)e_{j-i}(s), \quad 0 \leq i \leq N-1, \quad (3.2.8)$$

leads to the kernel estimator

$$\hat{c}(t, s) = \hat{\gamma}_0(t, s) + \sum_{i=1}^{N-1} K\left(\frac{i}{h}\right)(\hat{\gamma}_i(t, s) + \hat{\gamma}_i(s, t)). \quad (3.2.9)$$

It can be shown that under the usual assumptions on the kernel function K and the bandwidth h ($h \rightarrow \infty$, $h/N \rightarrow 0$),

$$\iint [\hat{c}(t, s) - c(t, s)]^2 dt ds \xrightarrow{P} 0, \quad \text{as } N \rightarrow \infty, \quad (3.2.10)$$

details are presented in Kokoszka and Young (2015a).

We conclude this section by stating the definitions of Gaussian stochastic processes which are needed to construct the limit distributions of the test statistics. Recall that if $\{W(x), 0 \leq x \leq 1\}$ is a standard Brownian motion (Wiener process), then the Brownian bridge is defined by $B(x) = W(x) - xW(1)$, $0 \leq x \leq 1$. The second-level Brownian bridge is defined by

$$\begin{aligned} V(x) = & W(x) + (2x - 3x^2)W(1) \\ & + (-6x + 6x^2) \int_0^1 W(y) dy, \quad 0 \leq x \leq 1. \end{aligned} \quad (3.2.11)$$

Both the Brownian bridge and the second-level Brownian bridge are special cases of the generalized Brownian bridge introduced by MacNeill (1978) who studied the asymptotic behavior of partial sums of polynomial regression residuals. Process (3.2.11) appears as the null limit of the KPSS statistic of Kwiatkowski et al. (1992). We will see in Section 3.3 that for functional data the limit involves an infinite sequence of independent and identically distributed second-level Brownian bridges $V_1(x), V_2(x), \dots$

3.3 Large sample limits

Horváth et al. (2014) developed tests of level-stationarity of a functional time series, i.e., of the null hypothesis $X_n(t) = \mu(t) + \eta_n(t)$, using the partial sum process

$$\begin{aligned} U_N(x, t) &= \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} (X_n(t) - \bar{X}(t)) \\ &= S_N(x, t) - \frac{\lfloor Nx \rfloor}{N} S_N(1, t), \end{aligned}$$

where $S_N(x, t)$ is the partial sum process of the curves $X_1(t), X_2(t), \dots, X_N(t)$ is defined by

$$S_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} X_n(t), \quad 0 \leq t, x \leq 1. \quad (3.3.1)$$

The process $U_N(x, t)$ has the form of a functional Brownian bridge. Their main statistic

$$\begin{aligned} T_N &= \iint U_N^2(x, t) dt dx \\ &= \int \|U_N(x, \cdot)\|^2 dx, \quad 0 \leq t, x \leq 1, \end{aligned}$$

is asymptotically distributed, under the null, as $\sum_{i=1}^{\infty} \lambda_i \int B_i^2(x) dx$, where $\lambda_1, \lambda_2, \dots$ are eigenvalues of the long-run covariance function of the observations X_n , and B_1, B_2, \dots are iid Brownian bridges. In the case of trend stationarity, a different distribution arises; the B_i must be replaced by second level Brownian bridges, and the λ_i are defined differently. The remainder of this section explains the details.

The test statistic for the trend-stationary case is based on the partial sum process of residuals (3.2.7), i.e., on the two-parameter process

$$Z_N(x, t) = \frac{1}{\sqrt{N}} \sum_{n=1}^{\lfloor Nx \rfloor} e_n(t). \quad (3.3.2)$$

A suitable test statistic is given by

$$\begin{aligned} R_N &= \iint Z_N^2(x, t) dt dx \\ &= \int \|Z_N(x, \cdot)\|^2 dx, \quad 0 \leq t, x \leq 1. \end{aligned} \quad (3.3.3)$$

It can be shown, Kokoszka and Young (2015a), that under the null hypothesis,

$$R_N \xrightarrow{d} \sum_{i=1}^{\infty} \lambda_i \int V_i^2(x) dx, \quad (3.3.4)$$

where $\lambda_1, \lambda_2, \dots$ are the eigenvalues of the long-run covariance function (3.2.1), and V_1, V_2, \dots are iid second-level Brownian bridges.

We now explain the issues arising in the functional case by comparing our result to that obtained by Kwiatkowski et al. (1992). If all curves are constant functions ($X_i(t) = X_i$ for $t \in [0, 1]$), the statistic R_N given by (3.3.3) is the numerator of the KPSS test statistic of Kwiatkowski et al. (1992), which is given by

$$\text{KPSS}_N = \frac{1}{N^2 \hat{\sigma}_N^2} \sum_{n=1}^N S_n^2 = \frac{R_N}{\hat{\sigma}_N^2},$$

where $\hat{\sigma}_N^2$ is a consistent estimator of the long-run variance σ^2 of the residuals. In the scalar case, (3.3.4) reduces to $R_N \xrightarrow{d} \sigma^2 \int_0^1 V^2(x) dx$, where $V(x)$ is a second-level Brownian bridge. If $\hat{\sigma}_N^2$ is a consistent estimator of σ^2 , the result of Kwiatkowski et al. (1992) is recovered, i.e. $\text{KPSS}_N \xrightarrow{d} \int_0^1 V^2(x) dx$. In the functional case, the eigenvalues λ_i can be viewed as long-run variances of the residual curves along the principal directions determined by the eigenfunctions of the kernel $c(\cdot, \cdot)$ defined by (3.2.1). To obtain a test analogous to the scalar KPSS test, with a parameter free limit null distribution, we must construct a statistic which involves a division by consistent estimators of the λ_i . We use only d largest eigenvalues in order not to increase the variability of the statistic caused by division by small empirical eigenvalues. A suitable statistic is

$$R_N^0 = \sum_{i=1}^d \frac{1}{\hat{\lambda}_i} \int_0^1 \langle Z_N(x, \cdot), \hat{\phi}_i \rangle^2 dx, \quad (3.3.5)$$

where the sample eigenvalues $\hat{\lambda}_i$ and eigenfunctions $\hat{\phi}_i$ are defined by (3.2.3). Statistic (3.3.5) extends the statistic KPSS_N . It can be shown that under suitable assumptions, Kokoszka and Young (2015a),

$$R_N^0 \xrightarrow{d} \sum_{i=1}^d \int_0^1 V_i^2(x) dx, \quad (3.3.6)$$

with the $V_i, 1 \leq i \leq d$, the same as in (3.3.4).

Section 3.4 describes how the tests based on relations (3.3.4) and (3.3.6) are implemented.

3.4 Algorithmic description of the test procedures

This section provides step-by-step descriptions of the test procedures based on limit relations (3.3.4) and (3.3.6).

Algorithm 17. [Monte Carlo test based on relation (3.3.4)]

1. Estimate the null model (3.1.1) and compute the residuals defined in equation (3.2.7).
2. Select kernel K and a bandwidth h in (3.2.9) and compute the eigenvalues $\hat{\lambda}_i \hat{\phi}_i$, $1 \leq i \leq N$, defined by (3.2.3).
3. Simulate a large number, say $G = 10,000$, of vectors $[V_1, V_2, \dots, V_N]$ consisting of independent second level Brownian bridge processes V_i defined in (3.2.11). Find the 95th percentile, $R_{critical}$, of the G replications of

$$R_N^* = \sum_{i=1}^N \hat{\lambda}_i \int_0^1 V_i^2(x) dx.$$

4. Compute the test statistic R_N defined in (3.3.3). If $R_N \geq R_{critical}$, reject H_0 at the 5% significance level.

In most applications, the $\hat{\lambda}_i$ decay very quickly to zero, so if N is large, it can be replaced in Algorithm 17 by a smaller number, e.g by $d = 20$, and the empirical distribution of the R_N^* can be replaced by that of the R_d^* . In Algorithm 17 the critical value must be obtained via Monte Carlo simulations for each data set. In Algorithm 18, tabulated critical values can be used. These depend on the number d of the functional principal components used to construct statistic R_N^0 . Typically d is a small, single digit, number. Table 3.1 lists selected critical values. They have been obtained by simulating $G = 10,000$ vectors $[V_1, V_2, \dots, V_d]$ and finding the percentiles of the G replications of

$$R^0(d) = \sum_{i=1}^d \int_0^1 V_i^2(x) dx. \tag{3.4.1}$$

Algorithm 18. [Asymptotic test based on relation (3.3.6)]

Table 3.1: Critical values of the distribution of the variable $R^0(d)$ given by (3.4.1).

	d	1	2	3	4	5
Size	10%	0.1201	0.2111	0.2965	0.3789	0.4576
	5%	0.1494	0.2454	0.3401	0.4186	0.5068
	1%	0.2138	0.3253	0.4257	0.5149	0.6131
	d	6	7	8	9	10
Size	10%	0.5347	0.6150	0.6892	0.7646	0.8416
	5%	0.5909	0.6687	0.7482	0.8252	0.9010
	1%	0.6960	0.7799	0.8574	0.9487	1.0326

1. Perform steps 1 and 2 of Algorithm 17.
2. Choose the smallest d such that $\sum_{i \leq d} \hat{\lambda}_i / \sum_{i \leq N} \hat{\lambda}_i > 0.85$.
3. Calculate the statistic R_N^0 given by (3.3.5) and reject H_0 if $R_N^0 > R_{critical}^0$, with the critical value given in Table 3.1.

The 85% rule in Step 2 is a rule of thumb; asking for 85% of the variance to be explained is based on good empirical results, leading to our choice above. In some applications, Step 2 may be replaced by a selection of d based on a visual fit of the truncated principal component expansion

$$X_n^{(d)}(t) = \hat{\mu}(t) + \sum_{j=1}^d \langle X_n, \hat{\phi}_j \rangle \hat{\phi}_j(t)$$

to the observed curves $X_n(t)$. In other applications, existing theory or experience may support certain choices of d . This is the case for the yield curves, which we use to illustrate the application of our tests (mean level plus $d = 2$ principal components). For financial data, d is generally small, with $d = 2, 3, 4$ being the typical values. However, for other types of data, e.g. for environmental data, d exceeding 10 may be needed. In such cases, caution is recommended in the application of Algorithm 18 as the resulting test may be numerically and statistically unstable due to the division by small $\hat{\lambda}_j$ which may exhibit large sampling variability.

An important step is the choice of h needed to estimate the long run covariance function. A great deal of research in this direction has been done for scalar and vector time series. For functional time series, the method proposed by Horváth et al. (2014b) often gives good results. It uses the flat top kernel

$$K(t) = \begin{cases} 1, & 0 \leq t < 0.1 \\ 1.1 - |t|, & 0.1 \leq t < 1.1 \\ 0, & |t| \geq 1.1 \end{cases} \quad (3.4.2)$$

advocated by Politis and Romano (Politis and Romano (1996), Politis and Romano (1999)) and Politis (2011), and a data-driven selection of h . This method performs well if the series length N is larger than several hundred, longer than the series we consider. In the simulations reported in Kokoszka and Young (2015a) a deterministic bandwidth $h = N^{2/5}$ (combined with the flat top kernel) produced good size and power. The optimal selection of h is not a focus of this paper, this complex issue must be investigated in a separate work. As in the scalar case, it is however unlikely that a selection procedure that is uniformly optimal for all dependence structures can be found. In testing problems, it is useful to use several values of h and trust results which do not depend on h in a reasonable range.

3.5 Application to yield and daily price curves

In this section, we apply the test procedures of Section 3.4 to several financial data sets which can be viewed as time series of functions. The most extensively studied series of this type is the series of yield curves. In the past, the series of monthly yield curves have been typically studied, but in recent years high quality data at the daily frequency have become available. On a given day, a yield curve shows the yield (interest) earned on a fixed income instrument as a function of maturity. In most economic studies, these are yields on bills and bonds issued by central banks. The shape and level of these curves reflect the expectations of investors on the future direction of a specific economic area, see e.g., Chapter 10 of Campbell et al. (1997) or Diebold and Rudebusch (2013). Figures 3.1 (left panel) and 3.2 show, respectively, five consecutive yield curves and two sets

of 250 yield curves. The question we want to answer is whether the time series of yield curves can be treated as stationary time series with trend, or if they contain a random walk component. Visual examination and economic interpretation of these data leads to the conclusion that a pure trend model will not hold over very long periods of time which include periods of growth and recession and changes in central bank policies. Over shorter periods of time, the trend model may however hold, and may be useful to investors in fixed income securities.

The second type of functional time series we study are daily price curves like those shown in the right panel of Figure 3.1. As noted in the introduction, whether a time series of closing prices on a specific asset contains a random walk (is a unit root process) has been one of the most extensively studied topics in finance. In contrast to these studies, we consider the series of price curves. Out of a large number of assets that are of interest, we selected the S&P500 index, the US dollar index and light crude oil futures. These assets represent, respectively, the equity, currency and commodity markets. As for the bond market, trend stationarity will not hold over long periods of time, but our tests can identify periods for which it does hold.

The main objective of the empirical analysis presented in this section is to uncover commonalities and differences between the various classes of assets with respect to the trend behavior of specific daily functions. The analysis will also illustrate the statistical properties of the tests we propose.

3.5.1 Data description

As an example of the time series of yield curves we use the daily United States Federal Reserve yield curves defined for maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months. The available data covers all business days from January 2001 to December 2013.

The second data set is the Standard & Poor's 500 financial index (S&P500) in one minute resolution. The index is a weighted average of stock values of the largest 500 U.S. companies. At each trading day, we consider a price curve. The last value on day $n - 1$ is not the same as the first value on day n . An overnight jump of over half a percent is not unusual. Figure 3.3 shows

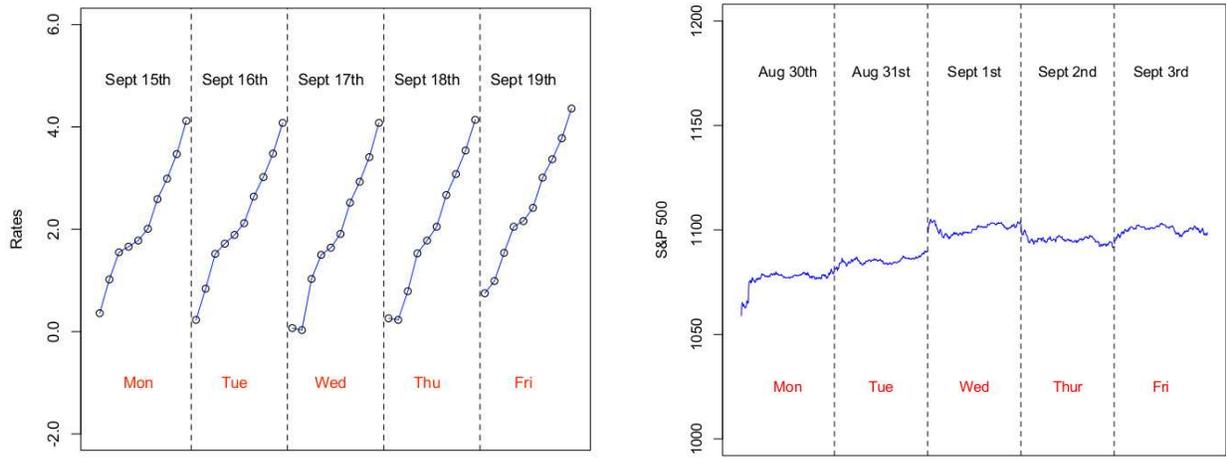


Figure 3.1: Left: five consecutive yield curves. Right: prices of the S&P 500 index over five consecutive days.

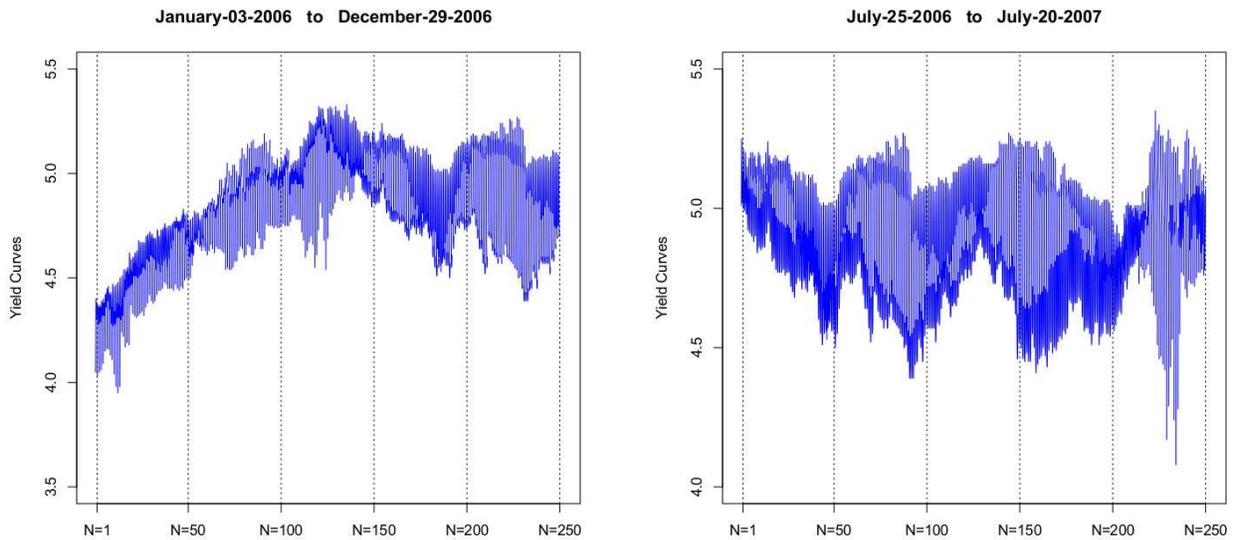


Figure 3.2: Consecutive yield curves over two time periods. Vertical lines show the location of sample sizes $N = 50, 100, 150, 200, 250$.

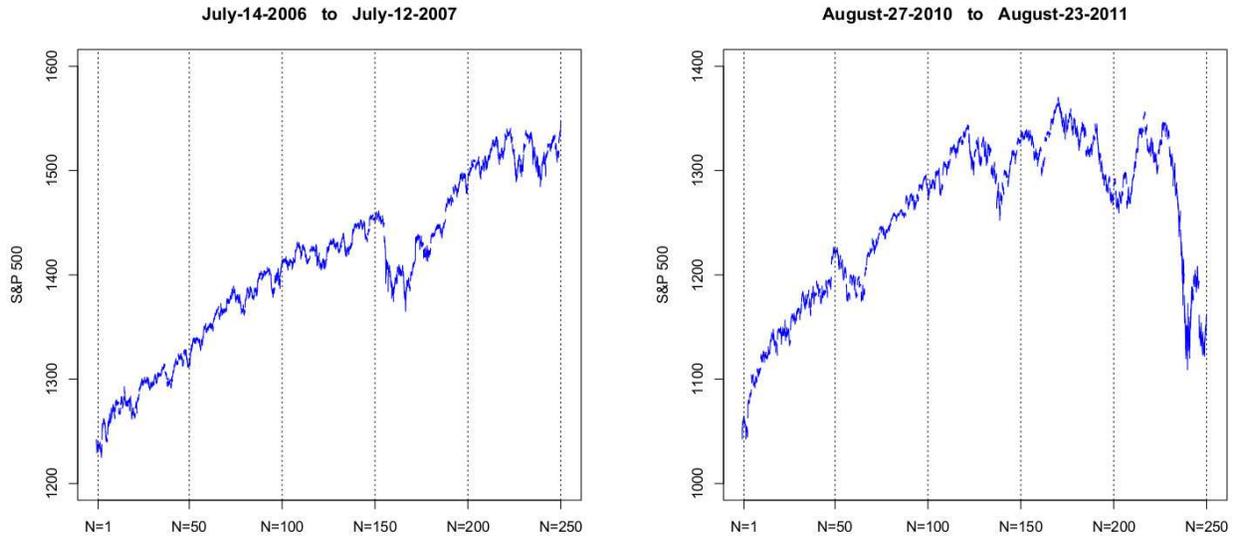


Figure 3.3: S&P 500 index for two different time periods.

the S&P500 index over two periods. The available data cover the period of 23 years from January 1989 to December 2012.

The third data set is the U.S. dollar index. As a weighted average of exchange rates against several major currencies, it measures the value of the U.S. dollar relative to a collection of other foreign currencies. A higher index indicates that the U.S. dollar is stronger compared to foreign currencies. The index is traded and used for the construction of derivative instruments. Similar to the S&P500 index, we use values in one minute resolution and consider one day as a single functional observation. However, instead of considering business days, we only exclude Saturdays for this data set. Figure 3.4 shows the U.S. dollar index over two different sampling periods. The available data cover the period of 23 years from January 1989 to December 2012.

The fourth data set consists of light crude oil futures. Light sweet crude oil futures and options are one of the worlds most highly traded energy products. Similar to the S&P500 and the U.S. dollar index, we use minute-by-minute prices, and consider one day as a single functional observation. We exclude only Saturdays for this data set. Figure 3.5 shows the light crude oil futures over two different sampling periods. The available data cover the period of 22 years from January 1989 to

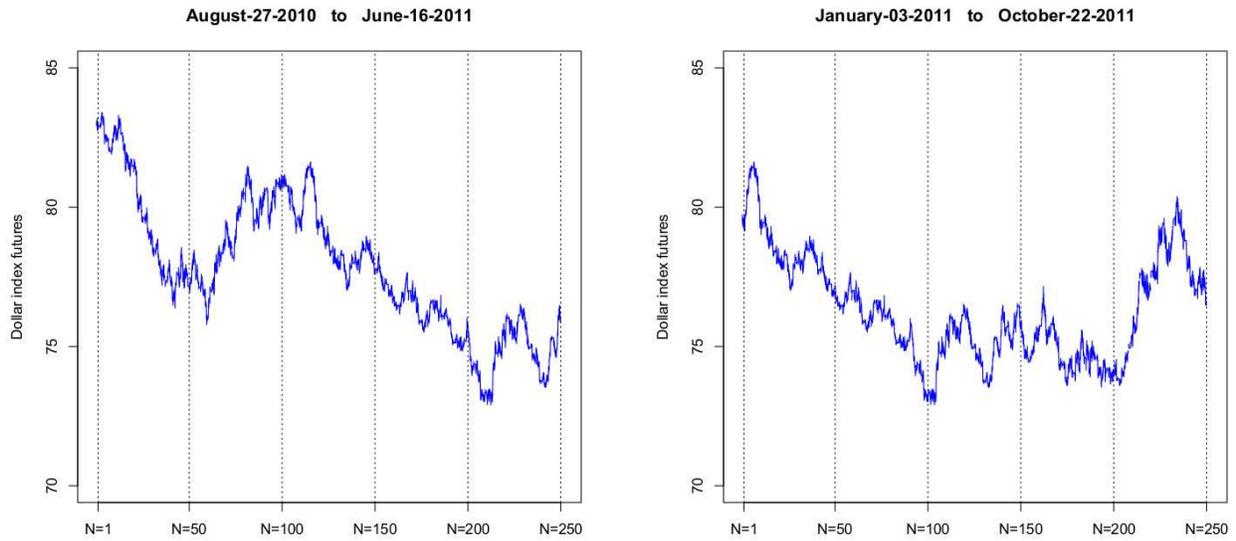


Figure 3.4: U.S. dollar index for two different time periods.

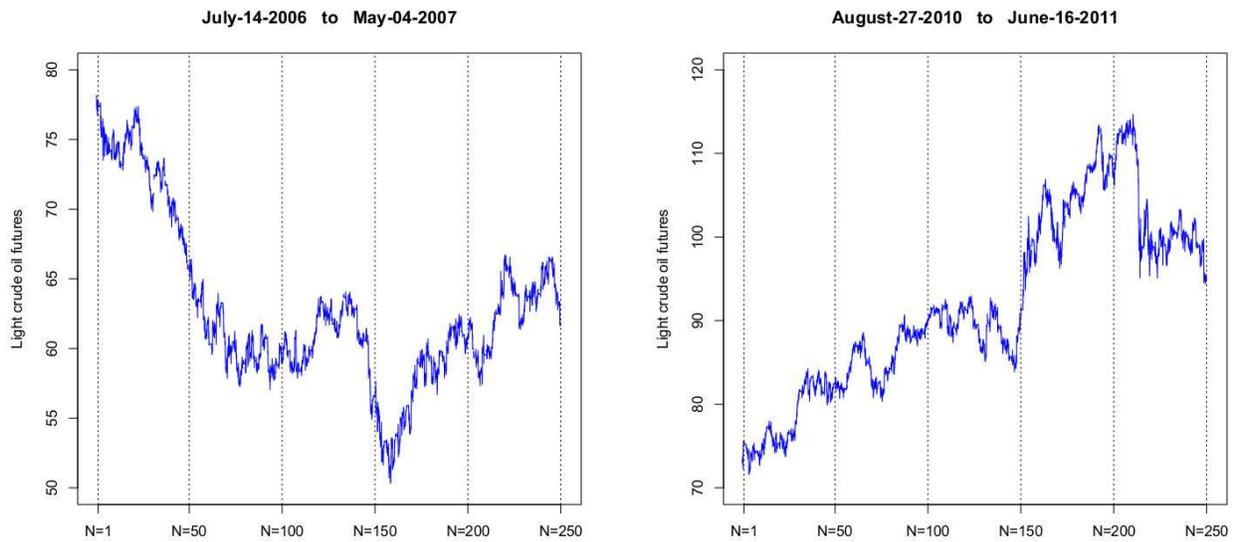


Figure 3.5: Light crude oil futures over two different time periods.

December 2011.

In Section 3.5.2, we show how the P-values broadly depend on the length of the series, N , and on the time period. We complement this analysis by focusing in Section 3.5.3 on selected time periods, those shown in Figures 3.2, 3.3, 3.4 and 3.5. This additional, more detailed analysis allows us to gain insights about the properties of the tests.

3.5.2 Long term trend characteristics of the curves

In this section, we display the P-values of the Monte Carlo test described in Algorithm 17 applied to the data described in Section 3.5.1. We also computed the P-values for the test based on Algorithm 18. While the P-values for the two tests are different, their general patterns are very similar, so to conserve space we focus on Algorithm 17 with the bandwidth $h = N^{2/5}$. We take a closer look at the differences between the two algorithms and the effect of the bandwidth in Section 3.5.3.

The main finding of our analysis is that for time periods of length $N = 100$ days, what corresponds roughly to the number of business days in four months, it is not uncommon that the null hypothesis of trend stationarity is not rejected. For periods covering the whole year, the null hypothesis is generally rejected. However, the proportion and temporal pattern of rejections are different for different assets. For example, for the yield curves there are hardly any period when H_0 can be accepted. This implies that this functional time series is not stationary even if a deterministic trend is allowed. This finding has implications for the prediction of yield curves; many methods assume a stationary model, some form of autoregression for factor coefficients. However, Chen and Niu (2014) obtained better prediction by assuming that the yield curves form only a locally stationary functional time series, i.e. stationary only on short subintervals. Our inferential procedures confirms the validity of such an approach. In the remainder of this section, we systematically present and discuss the results for all four data sets. For each asset, we consider all available consecutive, nonoverlapping periods of $N = 100$ and $N = 300$ days.

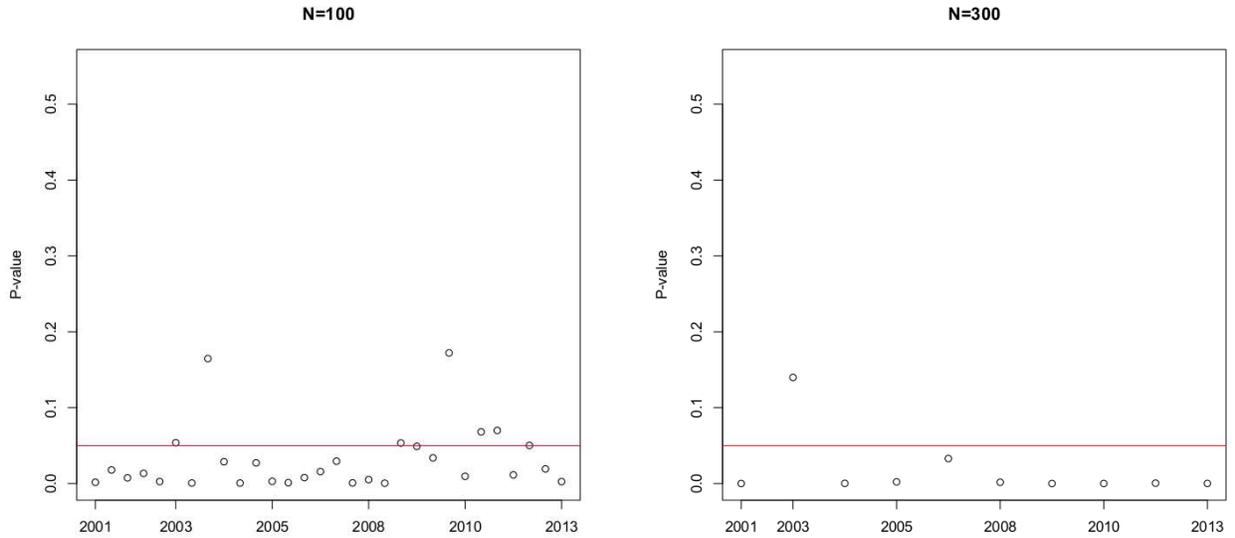


Figure 3.6: P-values of the test based on Algorithm 17 applied to the **Treasury Yield Curves**. The plot on the left shows thirty 100 day periods and the plot on the right shows ten 300 day periods.

Figure 3.6 exhibits the pattern of P-values for the Daily United States Federal Reserve yield curves. Focusing first on periods of length $N = 100$, we see that 23 out of the 30 periods show P-values below the significance level of 0.05. As the sample size increases to $N = 300$, we see that 9 out of the 10 longer periods have P-values below the significance level of 0.05. As noted above, an overriding conclusion is that the yield curves do not follow a stationary model even with a trend, and a presence of a random walk component or some other changes in the the stochastic structure must be taken into account.

Figure 3.7 shows the P-values for the S&P500 curves. For $N = 100$, 32 out of the 60 periods have P-values smaller than 0.05. In contrast to the yield curves, this shows that a stationary model with a trend can be suitable for many periods extending over several months; in most cases this corresponds to a persistent bull market, cf. Figure 3.3. However, as the sample size increases to $N = 300$, we see that 16 out of the 20 longer periods have P-values below 0.05; a bull market cannot last forever.

Figure 3.8 shows the P-values for the U.S. dollar index. For $N = 100$, 44 out of the 60 periods

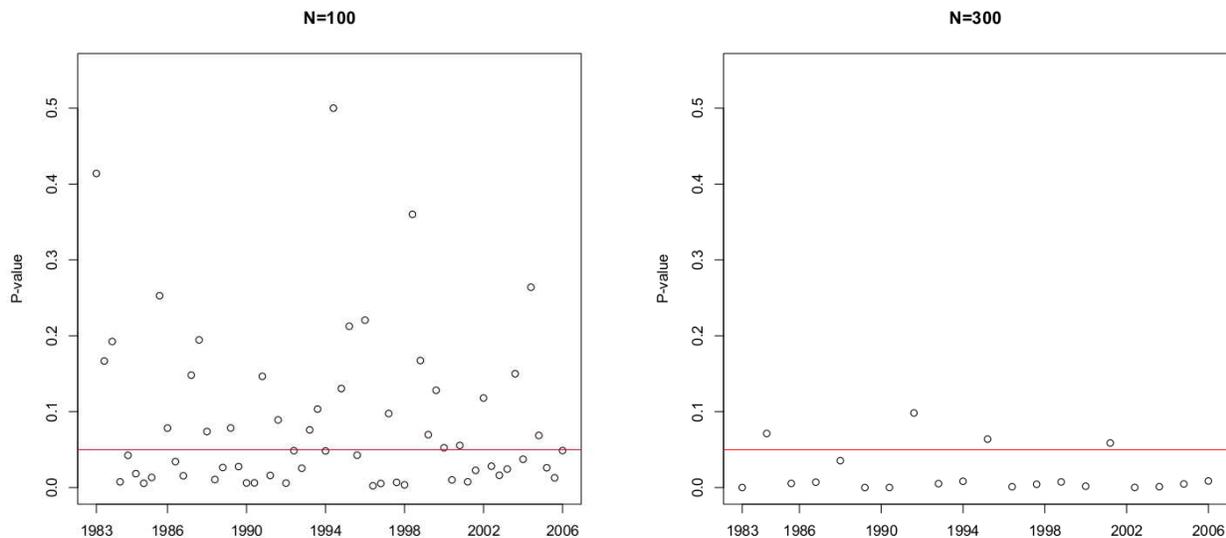


Figure 3.7: P-values of the test described in Algorithm 17 applied to the **S&P500 index**. The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods.

have P-values below 0.05. As the sample size increases to $N = 300$, we see that 17 out of the 20 longer periods have P-values smaller than 0.05. In terms of the trend behavior, the currency index is somewhere between the yield curves and the equity index. There are periods of trend stationarity but they are less frequent than for equities.

Finally, we turn to the light crude oil futures. Figure 3.9 shows the P-values. For $N = 100$, 42 out of the 60 periods have P-values smaller than 0.05. In this case, an interesting temporal pattern of these P-values can be seen. Starting from 1997, there are periods with increasing P-values, indicating that a trend model might often be suitable. This agrees with a persistent, almost linear, decline in prices from December 1996 to December 1998 followed by a long rise from January 1999 up to the summer of 2008, just before the financial crisis. These long periods were punctuated by short periods of reversals, so only in 3 out of 20 longer periods a trend model is accepted.

3.5.3 Properties of the tests

In this section, we elaborate on the findings of Section 3.5.2 in two ways: 1) we zoom in on specific time periods, those displayed in Figures 3.2, 3.3, 3.4 and 3.5, to establish a more direct

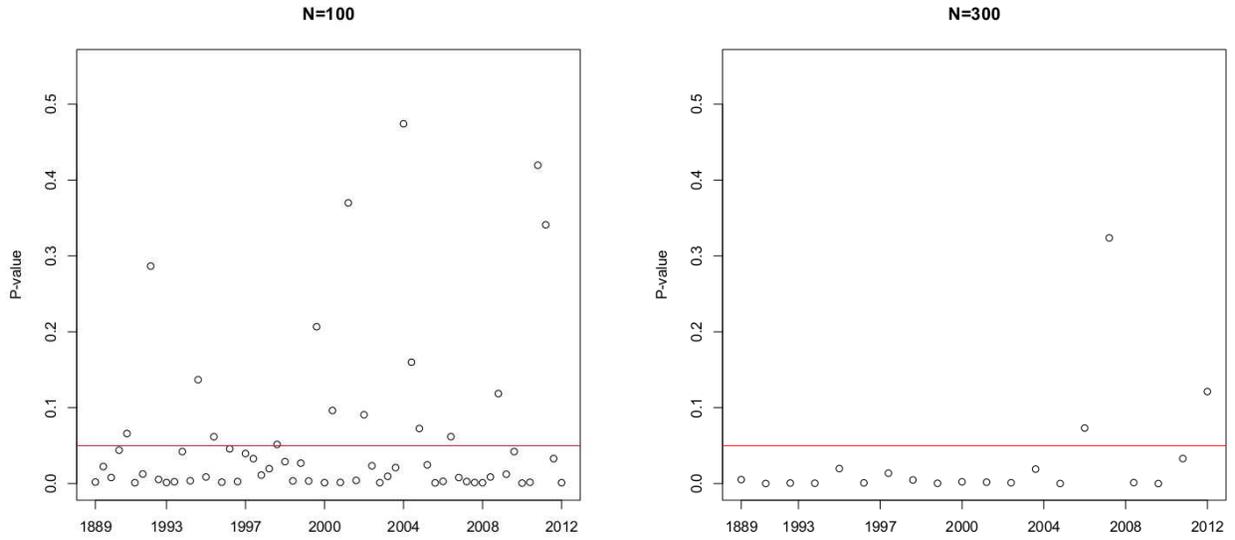


Figure 3.8: P-values of the test described Algorithm 17 applied to the **U.S. dollar index**. The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods.

connection between the data and the P-values, 2) we apply to these fewer periods both algorithms of Section 3.4 and use a selection of bandwidths h . The results are shown in Tables 3.2, 3.3, 3.4 and 3.5.

We begin by analyzing Table 3.2 which pertains to the yield curves shown in Figure 3.2. As for most other periods, the null hypothesis is rejected, except for a few cases corresponding to the bandwidth $h = N^{1/2}$. Simulations reported in Kokoszka and Young (2015a) show that for artificial data which resemble the yield curves, this bandwidth is too large. It makes the statistic too small and so the tests are too conservative. We also see that while the test based on the Monte Carlo distribution, statistic R_N , and the pivotal test based on R_N^0 generally give different P-values, the differences are small, and generally do not affect significance statements. Turning to the S&P500 index, for the two periods shown in Figure 3.3, Table 3.3 shows rejections, except for the first 100 days in the left panel of Figure 3.3. In some cases these rejections are weak if $h = N^{1/2}$; both tests again give the same conclusions in almost all cases. The conclusions for the U.S. Dollar index and Oil Futures are qualitatively the same as for the S&P500 data.

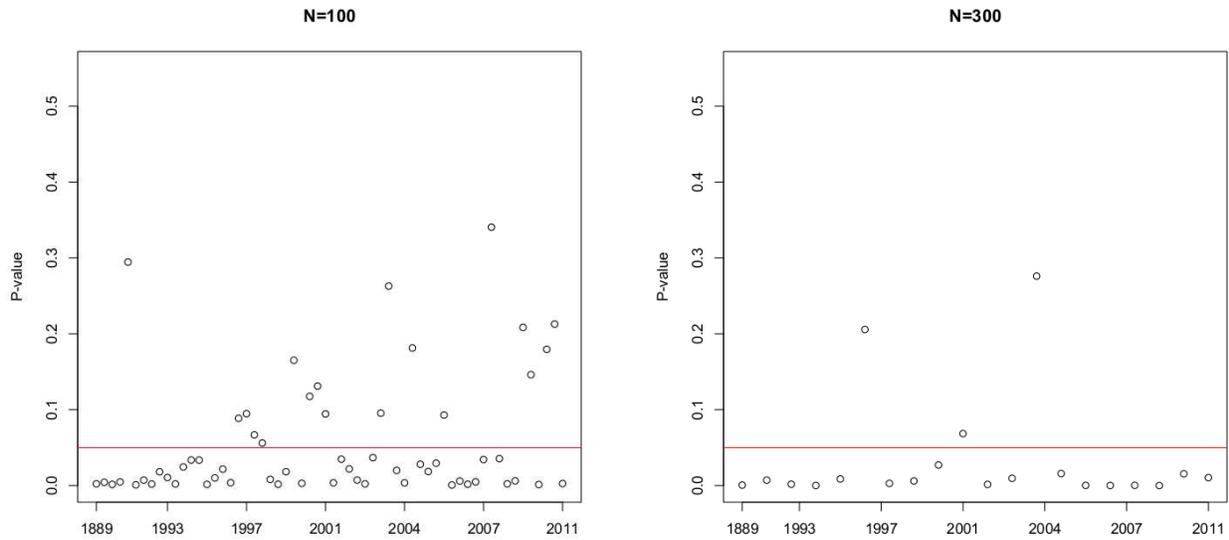


Figure 3.9: P-values of the test described Algorithm 17 applied to the **Oil Futures**. The plot on the left shows sixty 100 day periods and the plot on the right shows twenty 300 day periods.

The conclusion is that bandwidths $h = N^{1/3}$ or $h = N^{2/5}$ can be used for sample sizes in the range from 50 to 300. Both algorithms presented in Section 3.4 give practically the same results. We note that for the data we studied d was small, typically 2 or 3. If d is large, Algorithm 18 must be used with caution, as explained in Section 3.4.

Table 3.2: P-values of the tests of Section 3.4 applied to **Treasury Yield Curves**. The data are shown in Figure 3.2.

Time period	Sample size	Bandwidth	R_N	R_N^0
01/03/2006 – 03/15/2006	$N = 50$	$N^{1/3}$	0.0200	0.0206
		$N^{2/5}$	0.0432	0.0561
		$N^{1/2}$	0.0878	0.1361
01/03/2006 – 05/25/2006	$N = 100$	$N^{1/3}$	0.0022	0.0024
		$N^{2/5}$	0.0108	0.0116
		$N^{1/2}$	0.0410	0.0661
01/03/2006 – 08/07/2006	$N = 150$	$N^{1/3}$	0.0003	0.0005
		$N^{2/5}$	0.0013	0.0007
		$N^{1/2}$	0.0117	0.0086
01/03/2006 – 10/18/2006	$N = 200$	$N^{1/3}$	0.0000	0.0000
		$N^{2/5}$	0.0005	0.0001
		$N^{1/2}$	0.0015	0.0050
01/03/2006 – 12/29/2006	$N = 250$	$N^{1/3}$	0.0000	0.0000
		$N^{2/5}$	0.0000	0.0001
		$N^{1/2}$	0.0011	0.0051
Time period	Sample size	Bandwidth	R_N	R_N^0
07/25/2006 – 10/03/2006	$N = 50$	$N^{1/3}$	0.0065	0.0109
		$N^{2/5}$	0.0164	0.0290
		$N^{1/2}$	0.0404	0.0704
07/25/2006 – 12/14/2006	$N = 100$	$N^{1/3}$	0.0075	0.0169
		$N^{2/5}$	0.0272	0.0590
		$N^{1/2}$	0.0967	0.1885
07/25/2006 – 02/28/2007	$N = 150$	$N^{1/3}$	0.0002	0.0005
		$N^{2/5}$	0.0027	0.0044
		$N^{1/2}$	0.0230	0.0447
07/25/2006 – 05/09/2007	$N = 200$	$N^{1/3}$	0.0030	0.0081
		$N^{2/5}$	0.0166	0.0451
		$N^{1/2}$	0.1035	0.1965
07/25/2006 – 07/20/2007	$N = 250$	$N^{1/3}$	0.0000	0.0000
		$N^{2/5}$	0.0002	0.0010
		$N^{1/2}$	0.0090	0.0240

Table 3.3: P-values of the tests of Section 3.4 applied to **S&P500 index**. The data are shown in Figure 3.3.

Time period	Sample size	Bandwidth	R_N	R_N^0
07/14/2006 – 09/22/2006	$N = 50$	$N^{1/3}$	0.1364	0.2407
		$N^{2/5}$	0.1580	0.2584
		$N^{1/2}$	0.1443	0.1960
07/14/2006 – 12/04/2006	$N = 100$	$N^{1/3}$	0.2242	0.2620
		$N^{2/5}$	0.2669	0.2320
		$N^{1/2}$	0.2915	0.1707
07/14/2006 – 02/16/2007	$N = 150$	$N^{1/3}$	0.0001	0.0003
		$N^{2/5}$	0.0001	0.0013
		$N^{1/2}$	0.0057	0.0076
07/14/2006 – 05/01/2007	$N = 200$	$N^{1/3}$	0.0001	0.0001
		$N^{2/5}$	0.0001	0.0011
		$N^{1/2}$	0.0075	0.0120
07/14/2006 – 07/12/2007	$N = 250$	$N^{1/3}$	0.0013	0.0063
		$N^{2/5}$	0.0090	0.0254
		$N^{1/2}$	0.0519	0.1240
Time period	Sample size	Bandwidth	R_N	R_N^0
08/27/2010 – 11/05/2010	$N = 50$	$N^{1/3}$	0.0186	0.0457
		$N^{2/5}$	0.0324	0.0682
		$N^{1/2}$	0.0607	0.0967
08/27/2010 – 01/19/2011	$N = 100$	$N^{1/3}$	0.0329	0.0465
		$N^{2/5}$	0.0671	0.0916
		$N^{1/2}$	0.1365	0.1207
08/27/2010 – 03/31/2011	$N = 150$	$N^{1/3}$	0.0032	0.0088
		$N^{2/5}$	0.0112	0.0266
		$N^{1/2}$	0.0439	0.0762
08/27/2010 – 06/13/2011	$N = 200$	$N^{1/3}$	0.0000	0.0001
		$N^{2/5}$	0.0001	0.0009
		$N^{1/2}$	0.0026	0.0086
08/27/2010 – 08/23/2011	$N = 250$	$N^{1/3}$	0.0000	0.0000
		$N^{2/5}$	0.0001	0.0004
		$N^{1/2}$	0.0012	0.0055

Table 3.4: P-values of the tests of Section 3.4 applied to **U.S. dollar index**. The data are shown in Figure 3.4.

Time period	Sample size	Bandwidth	R_N	R_N^0
08/27/2010 – 10/24/2010	$N = 50$	$N^{1/3}$	0.0538	0.0350
		$N^{2/5}$	0.0980	0.0410
		$N^{1/2}$	0.1742	0.0562
08/27/2010 – 12/21/2010	$N = 100$	$N^{1/3}$	0.0001	0.0003
		$N^{2/5}$	0.0006	0.0014
		$N^{1/2}$	0.0079	0.0165
08/27/2010 – 02/18/2011	$N = 150$	$N^{1/3}$	0.0016	0.0012
		$N^{2/5}$	0.0093	0.0059
		$N^{1/2}$	0.0533	0.0243
08/27/2010 – 04/18/2011	$N = 200$	$N^{1/3}$	0.0022	0.0018
		$N^{2/5}$	0.0126	0.0087
		$N^{1/2}$	0.0711	0.0371
08/27/2010 – 06/16/2011	$N = 250$	$N^{1/3}$	0.0045	0.0089
		$N^{2/5}$	0.0219	0.0347
		$N^{1/2}$	0.1041	0.1146
Time period	Sample size	Bandwidth	R_N	R_N^0
01/03/2011 – 03/01/2011	$N = 50$	$N^{1/3}$	0.0039	0.0199
		$N^{2/5}$	0.0129	0.0455
		$N^{1/2}$	0.0350	0.1033
01/03/2011 – 04/30/2011	$N = 100$	$N^{1/3}$	0.0001	0.0006
		$N^{2/5}$	0.0003	0.0031
		$N^{1/2}$	0.0033	0.0178
01/03/2011 – 06/27/2011	$N = 150$	$N^{1/3}$	0.0001	0.0003
		$N^{2/5}$	0.0003	0.0021
		$N^{1/2}$	0.0029	0.0158
01/03/2011 – 08/24/2011	$N = 200$	$N^{1/3}$	0.0000	0.0001
		$N^{2/5}$	0.0001	0.0012
		$N^{1/2}$	0.0019	0.0107
01/03/2011 – 10/22/2011	$N = 250$	$N^{1/3}$	0.0038	0.0184
		$N^{2/5}$	0.0179	0.0609
		$N^{1/2}$	0.0691	0.1746

Table 3.5: P-values of the tests of Section 3.4 applied to **Oil Futures**. The data are shown in Figure 3.5.

Time period	Sample size	Bandwidth	R_N	R_N^0
07/14/2006 – 09/10/2006	$N = 50$	$N^{1/3}$	0.0161	0.0185
		$N^{2/5}$	0.0333	0.0340
		$N^{1/2}$	0.0691	0.0781
07/14/2006 – 11/07/2006	$N = 100$	$N^{1/3}$	0.0115	0.0243
		$N^{2/5}$	0.0318	0.0477
		$N^{1/2}$	0.0963	0.1116
07/14/2006 – 01/07/2007	$N = 150$	$N^{1/3}$	0.0000	0.0001
		$N^{2/5}$	0.0004	0.0007
		$N^{1/2}$	0.0055	0.0084
07/14/2006 – 03/06/2007	$N = 200$	$N^{1/3}$	0.0001	0.0004
		$N^{2/5}$	0.0001	0.0037
		$N^{1/2}$	0.0128	0.0342
07/14/2006 – 05/04/2007	$N = 250$	$N^{1/3}$	0.0000	0.0000
		$N^{2/5}$	0.0001	0.0006
		$N^{1/2}$	0.0028	0.0110
Time period	Sample size	Bandwidth	R_N	R_N^0
08/27/2010 – 10/24/2010	$N = 50$	$N^{1/3}$	0.1117	0.1159
		$N^{2/5}$	0.1771	0.1931
		$N^{1/2}$	0.2752	0.2694
08/27/2010 – 12/21/2010	$N = 100$	$N^{1/3}$	0.0882	0.2514
		$N^{2/5}$	0.1567	0.3883
		$N^{1/2}$	0.2343	0.4744
08/27/2010 – 02/19/2011	$N = 150$	$N^{1/3}$	0.0007	0.0079
		$N^{2/5}$	0.0044	0.0267
		$N^{1/2}$	0.0176	0.0620
08/27/2010 – 04/18/2011	$N = 200$	$N^{1/3}$	0.0030	0.0174
		$N^{2/5}$	0.0144	0.0589
		$N^{1/2}$	0.0551	0.1189
08/27/2010 – 06/15/2011	$N = 250$	$N^{1/3}$	0.0164	0.0704
		$N^{2/5}$	0.0530	0.1682
		$N^{1/2}$	0.1707	0.2828

Chapter 4

CHANGE POINT TESTS IN FUNCTIONAL FACTOR MODELS WITH APPLICATION TO YIELD CURVES

4.1 Introduction

Recent advances in yield curve modeling have brought to the fore a class of functional data models which we call functional factor models. This paper proposes several methods for testing the hypothesis of the existence of change points in such models. While there has been extensive research on yield curve prediction, modeling via regime switching processes and even change point estimation, we are not aware of any previous research concerned with a change point testing problem.

Yield curve modeling has been an important direction of economic research over many decades. A solid account of the classical theory related to the so called affine models is presented in Chapter 8 of Campbell et al. (1997); Filipović (2009) presents a continuous finance theory perspective. An approach that has gained wide acceptance in recent years is the Nelson–Siegel model and its dynamic modification, Diebold and Rudebusch (2013). This paper is concerned with the detection of change points in models that generalize the dynamic Nelson–Siegel model. The most direct motivation for our research comes from the recent work of Chen and Niu (2014) who show that accounting for possible change points in the term structure improves yield curve predictions. The purpose of this work is to develop formal tests of significance for the existence of change points in a general class of functional models for yield curves. We now elaborate on our contribution.

Denote by $t_j, 1 \leq j \leq J$, the maturities ordered from the shortest (1 month) to the longest (10 years). The general form of the dynamic Nelson–Siegel model can be written as

$$X_i(t_j) = \sum_{k=1}^K \beta_{i,k} f_k(t_j) + \varepsilon_i(t_j). \quad (4.1.1)$$

The index i refers to time periods at which the curves are available, it typically indexes days or months. The functions f_k are postulated to have a specific parametric form; in the standard model, $K = 3$ functions are used. The function f_1 is equal to one, its weight $\beta_{i,1}$ represents the general level of yields in period i . The function f_2 is decreasing, and for negative $\beta_{i,2}$ models the increase of yields with maturity. The function f_3 has a hump at maturities of 2–3 years, and models the curvature of the yield curve over such maturities, detailed formulas are presented Section 4.5. The attribute “dynamic” stems from the fact that the weights $\beta_{i,k}$ are time series; in a static model $\beta_{i,k} = \beta_k$ does not depend on period i . The objective of this work is to develop significance tests whose null hypothesis is that the mean structure of the K series $\{\beta_{i,k}\}$ is constant over a time period under consideration against the alternative that it changes at unknown change points. Our approach allows the error structure to change at prespecified points. Precise definitions of the mean and error structures are given in Section 4.2. Before proceeding further, we give an illustrative example. Figure 4.1 shows 100 yield curves centered at the height of the financial crisis of 2008. Visual inspection shows that the typical shape of the curves, which we will quantify as the mean structure, changed in mid–September. If we apply our test assuming that the error structure is the same over the sample period, then the test does not reject the null hypothesis that the mean structure has not changed. If, however, the error structure is allowed to change on Sep 16, then the test rejects the null hypothesis with a very high significance.

The above example illustrates the need for flexibility in modeling the error structure when testing for change points in the mean structure. The prespecified break points in the error structure can reflect exogenous information or can be based on exploratory analysis of the data. In Figure 4.1, the mean structure is reflected by the general level and range of the curves, the error structure by the “wiggleness” in the top, middle and bottom parts of the curves. The errors, volatility, of the curves increased after the middle of the sample. The change in mean structure is fairly obvious in Figure 4.1; this period is used to emphasize a main message of this paper that without taking

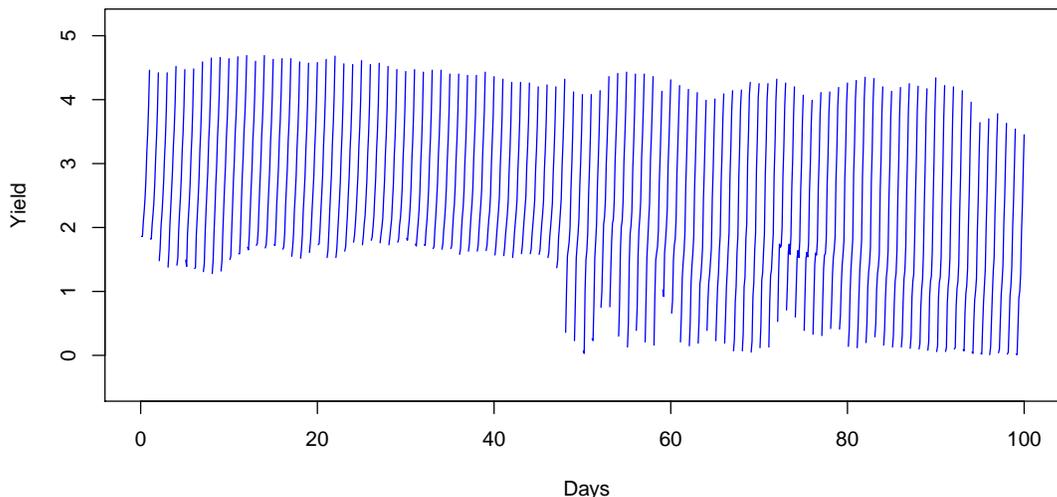


Figure 4.1: US yield curves over $N = 100$ business days from July-08-2008 to November-28-2008. The central time point corresponds to the Lehman Brothers collapse.

instability in the errors structure into account, decision on the existence of change points in the mean structure can be incorrect. In the context of scalar time series, this point has recently been emphasized by Dalla et al. (2015).

The fact that the stochastic structure of yield curve changes has been recognized for many decades. An established approach is to use hidden Markov chains to estimate the structural changes together with the parameters of the affine structure, Nieh et al. (2010). In addition to the aforementioned paper of Chen and Niu (2014), the only other paper concerned with change point estimation is Chib and Kang (2013). The main difference between hidden Markov models and change point models is that in the former only a few, typically two, states are assumed and the system moves between them. In a change point model, no such assumption is made, the parameters can take any values between the change points. This is the paradigm advocated by Chib and Kang (2013) who use a Bayesian approach to estimate the change points. This paper is concerned with testing for the presence of change points. We propose frequentist procedures based on asymptotic distributions of test statistics.

The methodology and theory we develop does not depend on the specific form of the curves f_k , which we will call factor curves, extending the usual terminology of multivariate statistics. Lenglwiler and Lenz (2010) and Hays et al. (2012), among others, argue that the standard Nelson–Siegel factor curves are not optimal in some respects. In our methodological and theoretical exposition, we merely assume that the f_k are square integrable and linearly independent. We then go a step further, and assume that the yield curves follow the model $X_i(t_j) = \tau_i(t_j) + \eta_i(t_j)$, where the curves τ_i describe the mean structure and the η_i are error curves. We will derive tests that allow to test for the presence of change points in the form of the functions τ_i without any parametric assumptions.

The remainder of the paper is organized as follows. In Section 4.2, we specify the general functional factor model. Testing for the presence of change point in this model is addressed in Section 4.3. In Section 4.4, we extend our approach to the nonparametric setting. Finite sample performance of the proposed methods is studied by a simulation study in Section 4.5. Main conclusions are summarized in Section 4.6. Section 4.7 contains proofs of the asymptotic results on which the tests are based. Their numerical implementation is explained in Section 4.8.

4.2 Functional Factor Model

In this section, we introduce the general functional factor model. In Section 4.4, we introduce the fully functional model which does not assume a factor structure.

We consider the functional term structure model

$$X_i(t) = \sum_{k=1}^K \beta_{i,k} f_k(t) + \varepsilon_i(t), \quad 1 \leq i \leq N. \quad (4.2.1)$$

This is the same model as (4.1.1), but the maturity t is modeled as a continuous variable, following the Nelson–Siegel paradigm. An empirical justification is that fractions of bonds with standard maturities can be traded at any time. The functions f_k are known, and we merely assume that

$$\int f_k^2(t) dt < \infty, \quad 1 \leq k \leq K. \quad (4.2.2)$$

The random coefficients $\beta_{i,k}$ can be decomposed as

$$\beta_{i,k} = \mu_{i,k} + b_{i,k}, \quad Eb_{i,k} = 0.$$

We want to test the change point hypothesis

$$H_0 : \boldsymbol{\mu}_1 = \boldsymbol{\mu}_2 = \dots = \boldsymbol{\mu}_N, \quad (4.2.3)$$

where

$$\boldsymbol{\mu}_i = [\mu_{i,1}, \mu_{i,2}, \dots, \mu_{i,K}]^\top$$

We thus want to test if the first order structure of the process (4.2.1) changes at some unknown points. Under the alternative, there are at most R possible changes in the mean structure at times

$$1 = r_0 < r_1 < r_2 < \dots < r_R < r_{R+1} = N.$$

When testing for a change in the mean, it is generally assumed in change point analysis, e.g. Csörgő and Horváth (1997), Horváth and Rice (2014), that the second order properties do not change. Even in the simplest case of independent normal observations, allowing for a change in both mean and variance leads to quite complex asymptotic theory, Horváth (1993). In the setting of model (4.2.1), the error structure is captured by the terms $\sum_{k=1}^K b_{i,k}f_k(t) + \varepsilon_i(t)$. We allow their stochastic structure to change at specific points

$$1 = i_0 < i_1 < i_2 < \dots < i_M < i_{M+1} = N.$$

In application to yield curves, the i_m can be determined as dates of substantial central bank intervention, times of events of economic impact, or by exploratory analysis of the variability of the curves. The dates i_m reflect available exogenous information and are treated as known. We refer to them as *break points*. By contrast, the *change points* r_ℓ are unknown, and their existence is to be tested. In Figure 4.1, the break point visually practically coincides with the change point. The break point is reflected by the higher variability of the yield curves in the second half of the sample.

The top and bottom parts of the graph are more variable, have larger errors. The change point is reflected by the wider range of the curves. In the Nelson–Siegel model, the decreasing curve f_2 describes the spread of yields between short and long maturities. For $\mu_2^* < \mu_2 < 0$, the yield curves containing $\mu_2^* f_2$ will have a larger spread than the curves containing $\mu_2 f_2$.

We now formulate model assumptions. Introduce the error vectors

$$\mathbf{a}_i(t) = [b_{i,1}, \dots, b_{i,K}, \varepsilon_i(t)]^\top, \quad 1 \leq i \leq N. \quad (4.2.4)$$

The vectors \mathbf{a}_i are stationary on each interval $(i_m, i_{m+1}]$, and have mean zero. Their dependence structure is described by the following assumptions.

Assumption 19. *Assume that the vectors \mathbf{a}_i defined by (4.2.4) admit the representation*

$$\mathbf{a}_i = g_m(\delta_i, \delta_{i-1}, \dots), \quad i_m < i \leq i_{m+1}, \quad m = 0, 1, \dots, M,$$

where $g_0, g_1, \dots, g_M : S^\infty \mapsto L^2$ are unknown deterministic measurable functions. The random elements $\{\delta_i, -\infty < i < \infty\}$ are iid with values in a measurable space S .

Broadly speaking, Assumption 19 requires merely that on the segments of stationarity are some abstract (non-linear) moving averages of abstract errors. Representations of this type impose a very flexible dependence structure and have become popular over the last decade, e.g. Wu (2005), Shao and Wu (2007), Aue et al. (2009), Hörmann and Kokoszka (2010), Hörmann et al. (2013) and Kokoszka and Reimherr (2013b). Observe that the same sequence $\{\delta_i\}$ is used in all functions g_m . This reflects the intuition that even though the stochastic structure can change from segment to segment, there is a dependence between the segments; Kokoszka and Leipus (2000), among others, used this paradigm in a change point problem for scalar ARCH models.

Next, we formalize the assumption that on each subinterval $(i_m, i_{m+1}]$ the sequence $\{\mathbf{a}_i, 1 \leq i \leq N\}$ is weakly dependent. We use the notion of approximability which has been recently used in the analysis of time series of functions. Chapter 16 of Horváth and Kokoszka (2012) puts this notion

in a historical context and provides a number of its applications. To formulate it in our context of segment-stationarity, we must extend the part-sequences $\{\mathbf{a}_i, i_m < i \leq i_{m+1}\}$ to the full domain of all integers. We thus denote by $\{\mathbf{a}_i^{(m)}\}$ the above sequence extended to all integers, i.e.

$$\mathbf{a}_i^{(m)} = g_m(\delta_i, \delta_{i-1}, \dots), \quad -\infty < i < \infty. \quad (4.2.5)$$

Assumption 20. For some $\delta > 0$ and $\kappa > 2 + \delta$,

$$\max_{1 \leq m \leq M+1} \sum_{\ell=1}^{\infty} (E \|\mathbf{a}_{i,\ell}^{(m)} - \mathbf{a}_i^{(m)}\|^{2+\delta})^{1/\kappa} < \infty, \quad (4.2.6)$$

where $\mathbf{a}_{i,\ell}^{(m)}$ is defined by $\mathbf{a}_{i,\ell}^{(m)} = g_m(\delta_i, \delta_{i-1}, \dots, \delta_{i-\ell+1}, \delta_{i-\ell}^*, \delta_{i-\ell-1}^*, \dots)$, and δ_k^* are independent copies of δ_i , independent of $\{\delta_i, -\infty < i < \infty\}$.

The essence of Assumption 20 is that the impact of innovations δ_i far back in the past becomes negligible; replacing them by independent copies does not affect the distribution of the $\mathbf{a}_i^{(m)}$ much. Condition (4.2.6) quantifies the magnitude of the effect of such a replacement. It allows us to control remainder terms arising by replacing the sequence $\mathbf{a}_i^{(m)}$ by sequences consisting of variables that are independent for sufficiently large lags (m -dependent sequences). The arbitrary constants δ and κ in (4.2.6) are needed to guarantee that a weak approximation theorem of Berkes et al. (2013b), which we use in our proofs, holds for every segment.

Our last assumption states that the segments of stationarity have asymptotically comparable lengths.

Assumption 21. We assume that $i_m = i_m(N)$ and

$$\lim_{N \rightarrow \infty} N^{-1} i_m(N) = \theta_m, \quad 1 \leq m \leq M, \quad (4.2.7)$$

with $0 = \theta_0 < \theta_1 < \theta_2 < \dots < \theta_M < \theta_{M+1} = 1$.

4.3 Detection Through Projections onto Factors

The method presented in this section directly exploits representation (4.2.1). A method that does not use the factor structure is presented in Section 4.4.

Test statistics can be derived either using the vector of projections

$$\mathbf{z}_i = [\langle X_i, f_1 \rangle, \dots, \langle X_i, f_K \rangle]^\top \quad (4.3.1)$$

or cumulative estimates $\hat{\boldsymbol{\mu}}_k$ computed using the first k observations. These two approaches are equivalent, as we now explain. We begin with the CUSUM process of the vectors (4.3.1). Introduce the matrix

$$\mathbf{C} = [\langle f_k, f_j \rangle, 1 \leq k, j \leq K].$$

The matrix \mathbf{C} is deterministic and known. We assume that the f_k are linearly independent. That is, $\sum_{k=1}^K a_k f_k = 0$ in L^2 implies $a_1 = a_2 = \dots = a_K = 0$. It is easy to check that this condition implies that the columns of \mathbf{C} are linearly independent, so \mathbf{C}^{-1} exists. By (4.2.1),

$$\mathbf{z}_i = \mathbf{C}\boldsymbol{\mu}_i + \boldsymbol{\gamma}_i, \quad (4.3.2)$$

where

$$\boldsymbol{\gamma}_i = \mathbf{C}\mathbf{b}_i + \boldsymbol{\varepsilon}_i,$$

and where

$$\mathbf{b}_i = [b_{i,1}, b_{i,2}, \dots, b_{i,K}]^\top, \quad \boldsymbol{\varepsilon}_i = [\langle \varepsilon_i, f_1 \rangle, \langle \varepsilon_i, f_2 \rangle, \dots, \langle \varepsilon_i, f_K \rangle]^\top.$$

Since the matrix \mathbf{C} is deterministic and invertible, relation (4.3.2) implies that a change in the vectors $\boldsymbol{\mu}_i$ is equivalent to a change in the \mathbf{z}_i at the same change points. Test statistics will thus be based on the CUSUM process

$$\boldsymbol{\alpha}_N(x) = N^{-1/2} \left(\sum_{i=1}^{\lfloor Nx \rfloor} \mathbf{z}_i - \frac{\lfloor Nx \rfloor}{N} \sum_{i=1}^N \mathbf{z}_i \right), \quad 0 \leq x \leq 1. \quad (4.3.3)$$

Another route that leads to the process (4.3.3) is through cumulative least squares estimators of the vector $\boldsymbol{\mu} = [\mu_1, \mu_2, \dots, \mu_K]^\top$, which does not depend on i under H_0 . The estimator based on the whole sample minimizes the least-squares criterion

$$U_N(\boldsymbol{\mu}) = U_N(\mu_1, \mu_2, \dots, \mu_K) = \sum_{i=1}^N \left\| X_i - \sum_{k=1}^K \mu_k f_k \right\|^2.$$

It is given by $\hat{\boldsymbol{\mu}}_N = N^{-1} \mathbf{C}^{-1} \sum_{i=1}^N \mathbf{z}_i$. Denote by $\hat{\boldsymbol{\mu}}_k$ the estimator based on the first k functions, i.e. $\hat{\boldsymbol{\mu}}_k = k^{-1} \mathbf{C}^{-1} \sum_{i=1}^k \mathbf{z}_i$. Then

$$N^{-1/2} k (\hat{\boldsymbol{\mu}}_k - \hat{\boldsymbol{\mu}}_N) = \mathbf{C}^{-1} \boldsymbol{\alpha}_N (kN^{-1}).$$

Thus, functionals of the process $N^{1/2} x (\hat{\boldsymbol{\mu}}_{[Nx]} - \hat{\boldsymbol{\mu}}_N)$, $0 \leq x \leq 1$, are the same as those of the process (4.3.3), up to the multiplication by a known deterministic matrix.

Our next goal is to specify the limit distribution of the process $\boldsymbol{\alpha}_N$. Notice that $\boldsymbol{\gamma}_i = \mathbf{w}(\mathbf{a}_i)$, where \mathbf{w} is a known function and \mathbf{a}_i is defined in (4.2.4). This allows us to introduce the $M + 1$ infinite domain stationary sequences

$$\boldsymbol{\gamma}_i^{(m)} = \mathbf{w}(\mathbf{a}_i^{(m)}), \quad -\infty < i < \infty,$$

and define their long-run covariance matrices

$$\mathbf{V}_m = \sum_{\ell=-\infty}^{\infty} \text{Cov}(\boldsymbol{\gamma}_i^{(m)}, \boldsymbol{\gamma}_{i+\ell}^{(m)}).$$

Theorem 22. *If H_0 (4.2.3) and Assumptions 19, 20, 21 hold, and the factors f_k are linearly independent, then*

$$\boldsymbol{\alpha}_N \xrightarrow{d} \mathbf{G}^0, \quad \text{in } D^K([0, 1]),$$

where the process \mathbf{G}^0 is defined by

$$\mathbf{G}^0(x) = \mathbf{G}(x) - x\mathbf{G}(1), \tag{4.3.4}$$

and $\mathbf{G}(x)$, $x \in [0, 1]$, is a mean zero \mathbb{R}^K -valued Gaussian process with covariances

$$E[\mathbf{G}(x)\mathbf{G}^\top(y)] = \sum_{j=1}^m (\theta_j - \theta_{j-1}) \mathbf{V}_j + (x - \theta_m) \mathbf{V}_{m+1}, \quad \theta_m \leq x \leq \theta_{m+1}, y \geq x.$$

Theorem 22 is proven in Section 4.7.

The covariances of the process $\mathbf{G}^0(x)$ can be computed explicitly. They are given by

$$\begin{aligned} \mathbf{Q}(x, y) &:= E[\mathbf{G}^0(x)\mathbf{G}^0(y)^\top] \\ &= (1-y) \left[\sum_{j=1}^m (\theta_j - \theta_{j-1}) \mathbf{V}_j + (x - \theta_m) \mathbf{V}_{m+1} \right] \\ &\quad - x \left[\sum_{j=1}^{m'} (\theta_j - \theta_{j-1}) \mathbf{V}_j + (y - \theta_{m'}) \mathbf{V}_{m'+1}(t, s) \right] \\ &\quad + xy \sum_{j=1}^{M+1} (\theta_j - \theta_{j-1}) \mathbf{V}_j, \end{aligned} \tag{4.3.5}$$

where $0 \leq x \leq y \leq 1$, $\theta_m \leq x \leq \theta_{m+1}$ and $\theta_{m'} \leq y \leq \theta_{m'+1}$.

Tests can be based on the Cramér–von–Mises functional

$$\mathcal{C}_N = \int_0^1 \|\alpha_N(x)\|^2 dx, \tag{4.3.6}$$

where $\|\cdot\|$ is the Euclidean norm in \mathbb{R}^K , or the Kolmogorov–Smirnov functional

$$\mathcal{K}_N = \sup_{0 \leq x \leq 1} \|\alpha_N(x)\|, \tag{4.3.7}$$

or their weighted versions. In finite samples, Cramér–von–Mises tests generally perform better, and we therefore focus on the statistic \mathcal{C}_N . By Theorem 22,

$$\mathcal{C}_N \xrightarrow{d} \int_0^1 \|\mathbf{G}^0(x)\|^2 dx. \tag{4.3.8}$$

To perform the test, we must simulate the distribution of the right–hand side of (4.3.8). We propose two methods. The first one is based on the following general result.

Proposition 23. *Let $\mathbf{\Gamma}(x)$, $x \in [0, 1]$, be a mean zero \mathbb{R}^K -valued Gaussian processes with covariances $\mathbf{R}(x, y) = E[\mathbf{\Gamma}(x)\mathbf{\Gamma}(y)^\top]$. Then*

$$\int_0^1 \|\mathbf{\Gamma}(x)\|^2 dx \stackrel{d}{=} \sum_{j=1}^{\infty} \lambda_j Z_j^2, \tag{4.3.9}$$

where the Z_j are independent standard normal random variables and the λ_j are the eigenvalues of covariance kernel $\mathbf{R}(\cdot, \cdot)$, i.e.

$$\int_0^1 \mathbf{R}(x, y) \phi_j(y) dy = \lambda_j \phi_j(x), \quad (4.3.10)$$

where the $\phi_j(x)$ are orthonormal eigenfunctions defined on the unit interval and taking values in \mathbb{R}^K .

Proposition 23 follows from the Karhunen–Loève decomposition of a Gaussian element in a separable Hilbert space. To enhance the understanding of this result, we present a proof in Section 4.7. It shows that to approximate the distribution of $\int_0^1 \|\mathbf{G}^0(x)\|^2 dx$, it is enough to compute the eigenvalues λ_j in (4.3.10), with \mathbf{Q} given by (4.3.5) in place of \mathbf{R} . To this end, we must estimate the long-run covariance matrices \mathbf{V}_m . These estimates are also needed in the second method of approximating the limit in (4.3.8), which we now discuss.

The second method is based on generating replications of the process \mathbf{G}^0 . By (4.3.4), this reduces to generating replications of the process \mathbf{G} . In the course of the proof of Theorem 22, it is shown that

$$\mathbf{G}(x) = \sum_{j=1}^m (\theta_j - \theta_{j-1})^{1/2} \mathbf{G}_j(1) + (\theta_{m+1} - \theta_m)^{1/2} \mathbf{G}_{m+1} \left(\frac{x - \theta_m}{\theta_{m+1} - \theta_m} \right), \quad x \in (\theta_m, \theta_{m+1}]. \quad (4.3.11)$$

Each process \mathbf{G}_j is a mean zero Gaussian process with $E[\mathbf{G}_j(x) \mathbf{G}_j(y)^\top] = \min(x, y) \mathbf{V}_j$. It can be simulated as $\mathbf{G}_j(x) = \mathbf{L}_j \mathbf{W}(x)$, where $\mathbf{L}_j \mathbf{L}_j^\top = \mathbf{V}_j$ and $\mathbf{W} = [W_1, W_2, \dots, W_K]^\top$ consists of K independent standard Wiener processes. The decomposition of the long-run variance matrix uses either the upper or lower Cholesky decomposition. This representation always exists because \mathbf{V}_j is non-negative definite. In order to simulate the Gaussian processes \mathbf{G}_j , we must compute the estimated long-run covariance matrices $\widehat{\mathbf{V}}_j$, for which computationally efficient R implementations exist, details are described in Section 4.8.

We conclude this section by stating the consistency of the test based on convergence (4.3.8). To keep the statement simple, we consider only one change point, but it can be shown by the same

technique, merely with a more complex notation, that the statistic \mathcal{C}_N diverges if there are more than one change points.

Theorem 24. *Suppose there is a change point $r \in (0, 1)$ such that $\boldsymbol{\mu} \neq \boldsymbol{\mu}^*$, where $\boldsymbol{\mu} = \boldsymbol{\mu}_1 = \dots = \boldsymbol{\mu}_{[Nr]}$, $\boldsymbol{\mu}^* = \boldsymbol{\mu}_{[Nr]+1} = \dots = \boldsymbol{\mu}_N$. Then,*

$$N^{-1}\mathcal{C}_N \xrightarrow{P} \|\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|^2 \int_0^1 \{g^*(x, r)\}^2 dx, \quad (4.3.12)$$

with the function g^* defined in (4.7.5).

Theorem 24 is proven in Section 4.7.1. It implies that $\mathcal{C}_N \xrightarrow{P} \infty$, provided $\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)$ is not a zero vector. Since the limit distribution in (4.3.8) does not depend on the vectors $\boldsymbol{\mu}_i$, it shows that the probability of rejection approaches one, as $N \rightarrow \infty$.

4.4 A Nonparametric Functional Approach

In this section, we consider a testing procedure that does not assume model (4.2.1). To motivate it, we rewrite model (4.2.1) as

$$X_i(t) = \tau_i(t) + \eta_i(t), \quad 1 \leq i \leq N, \quad (4.4.1)$$

where

$$\tau_i(t) = \sum_{k=1}^K \mu_{i,k} f_k(t), \quad \eta_i(t) = \sum_{k=1}^K b_{ik} f_k(t) + \varepsilon_i(t). \quad (4.4.2)$$

In (4.4.2), the mean functions τ_i and the error functions η_i are expressed in terms of the components of model (4.2.1). However, such a specific form is not assumed in (4.4.1). Functional factor models, and earlier affine models, postulate some form of parametric dependence of the yield curve on a small number of parameters. Formulation (4.4.1) can be viewed as a nonparametric model emphasizing the main first order structure described by the functions τ_i , which are all equal to a function τ under H_0 . The form of the function τ can be arbitrary. It can be estimated by nonparametric

methods, but our focus here is on testing if it does not change. In this setting, the null hypothesis becomes

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_N, \quad (4.4.3)$$

with an alternative formulated analogously as in Section 4.2. Observe that under model (4.2.1), in light of (4.4.2), the null hypothesis (4.4.3) is equivalent to the null hypothesis (4.2.3).

The problem of testing for a change point in the functional means τ_i , without any reference to yield curves, was addressed by Berkes et al. (2009) who assumed that the error curves η_i are iid. Hörmann and Kokoszka (2010) extended that test by allowing to the η_i to be stationary and weakly dependent. In both cases, the tests are based on projections of the data on the estimated functional principal components of the errors η_i . (Both tests are described, respectively, in Chapters 6 and 16 of Horváth and Kokoszka (2012).) In our setting, the η_i are not stationary; the functional principal components, and hence the corresponding projections cannot be defined. It is however possible to derive tests based directly on the functional CUSUM process

$$\alpha_N(x, t) = \frac{1}{\sqrt{N}} \left(\sum_{i=1}^{\lfloor Nx \rfloor} X_i(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{i=1}^N X_i(t) \right). \quad (4.4.4)$$

In the remainder of this section, we develop the required theory and derive the tests.

We begin by stating assumptions analogous to Assumptions 19 and 20. The error functions η_i are mean zero and form stationary sequences on the intervals $(i_m, i_{m+1}]$. By $\eta_i^{(m)}$ we denote their extensions to the infinite domain consisting of all integers.

Assumption 25. *The functions η_i in (4.4.1) admit the representation*

$$\eta_i = g_m(\delta_i, \delta_{i-1}, \dots), \quad i_m < i \leq i_{m+1}, \quad m = 0, 1, \dots, M,$$

where the functions g_m and the errors δ_i satisfy conditions of Assumption 19

Assumption 26. *For some $\delta > 0$ and $\kappa > 2 + \delta$,*

$$\max_{1 \leq m \leq M+1} \sum_{\ell=1}^{\infty} (E \|\eta_{i,\ell}^{(m)} - \eta_i^{(m)}\|^{2+\delta})^{1/\kappa} < \infty, \quad (4.4.5)$$

where $\eta_{i,\ell}^{(m)}$ is defined by $\eta_{i,\ell}^{(m)} = g_m(\delta_i, \delta_{i-1}, \dots, \delta_{i-\ell+1}, \delta_{i-\ell}^*, \delta_{i-\ell-1}^*, \dots)$, and δ_k^* are independent copies of δ_i , independent of $\{\delta_i, -\infty < i < \infty\}$.

Consider the long-run covariance kernels defined by

$$D_m(t, s) = \sum_{\ell=-\infty}^{\infty} E[\eta_0^{(m)}(t)\eta_\ell^{(m)}(s)], \quad 0 \leq t, s \leq 1. \quad (4.4.6)$$

The existence of the L^2 -limit $D_m(\cdot, \cdot)$ was established by Horváth et al. (2013). The following theorem, proven in Section 4.7, is an analog of Theorem 22.

Theorem 27. *Under H_0 (4.4.3) and Assumptions 25, 26, 21, we can define Gaussian processes $\Gamma_N^0(x, t)$, $0 \leq x, t \leq 1$, such that*

$$\sup_{0 \leq x \leq 1} \|\alpha_N(x, t) - \Gamma_N^0(x, t)\| \xrightarrow{P} 0.$$

Each process Γ_N^0 is defined by

$$\Gamma_N^0(x, t) = \Gamma_N(x, t) - x\Gamma_N(1, t),$$

where Γ_N is mean zero Gaussian with covariances

$$E[\Gamma_N(x, t)\Gamma_N(y, s)] = \sum_{j=1}^m (\theta_j - \theta_{j-1})D_j(t, s) + (x - \theta_m)D_{m+1}(t, s), \quad \theta_m \leq x \leq \theta_{m+1}, x \leq y.$$

The covariances $E[\Gamma_N^0(x, t)\Gamma_N^0(y, s)]$ can be computed explicitly. Assuming

$$0 \leq x \leq y \leq 1, \quad \theta_m \leq x \leq \theta_{m+1}, \quad \theta_{m'} \leq y \leq \theta_{m'+1}, \quad (4.4.7)$$

$$\begin{aligned} E[\Gamma_N(x, t)\Gamma_N(1, s)] &= \sum_{j=1}^m (\theta_j - \theta_{j-1})D_j(t, s) + (x - \theta_m)D_{m+1}(t, s), \\ E[\Gamma_N(y, s)\Gamma_N(1, t)] &= \sum_{j=1}^{m'} (\theta_j - \theta_{j-1})D_j(t, s) + (y - \theta_{m'})D_{m'+1}(t, s), \\ E[\Gamma_N(1, s)\Gamma_N(1, t)] &= \sum_{j=1}^{M+1} (\theta_j - \theta_{j-1})D_j(t, s). \end{aligned}$$

Assuming (4.4.7), we thus obtain

$$\begin{aligned}
U^0(x, y; t, s) &:= E[\Gamma_N^0(x, t)\Gamma_N^0(y, s)] \\
&= E[(\Gamma_N(x, t) - x\Gamma_N(1, t))(\Gamma_N(y, s) - y\Gamma_N(1, s))] \\
&= (1 - y) \left[\sum_{j=1}^m (\theta_j - \theta_{j-1}) D_j(t, s) + (x - \theta_m) D_{m+1}(t, s) \right] \\
&\quad - x \left[\sum_{j=1}^{m'} (\theta_j - \theta_{j-1}) D_j(t, s) + (y - \theta_{m'}) D_{m'+1}(t, s) \right] \\
&\quad + xy \sum_{j=1}^{M+1} (\theta_j - \theta_{j-1}) D_j(t, s).
\end{aligned} \tag{4.4.8}$$

The complex structure of U^0 is what distinguishes the present research from the methods developed by Berkes et al. (2009) and Hörmann and Kokoszka (2010). The later work can be considered as a special case, in which $D_1 = D_2 = \dots = D_{M+1} (= D)$. In that case, $U^0(x, y; t, s) = (\min(x, y) - xy)D(t, s)$. Consequently, new approaches are needed to implement tests based on Theorem 27, even though they use the usual functionals. Denote by $\{\Gamma^0(x, t), 0 \leq x, t \leq 1\}$ a process with the same distribution as each Γ_N^0 . As in Section 4.3, we focus on the Cramér-von-Mises functional

$$\mathcal{V}_N = \iint \alpha_N^2(x, t) dt dx, \tag{4.4.9}$$

and the convergence

$$\mathcal{V}_N \xrightarrow{d} \iint \{\Gamma^0(x, t)\}^2 dt dx. \tag{4.4.10}$$

There is no explicit formula for the distribution of the limit in (4.4.10). It depends on the unknown long-run covariance kernels D_j , $1 \leq j \leq M + 1$, whose estimation is discussed below. Once these estimates are computed, the approximations of right-hand side of (4.4.10) is based on the relation

$$\|\Gamma^0\|^2 = \sum_{j=1}^{\infty} \lambda_j Z_j^2, \tag{4.4.11}$$

where the Z_j are iid standard normal and the λ_j are the eigenvalues of U^0 given by (4.4.8). If \widehat{U}^0 is an estimator of U^0 , then the λ_j are approximated by the eigenvalues $\widehat{\lambda}_j$ defined by

$$\iint \widehat{U}^0(x, y; t, s) \widehat{\varphi}_j(y, s) dy ds = \widehat{\lambda}_j \widehat{\varphi}_j(x, t).$$

and $\sum_{j=1}^{\infty} \lambda_j Z_j^2$ by $\sum_{j=1}^N \hat{\lambda}_j Z_j^2$. The numerical computation of the $\hat{\lambda}_j$ is not trivial, details are described in Section 4.8. The simulation of the limit process Γ^0 is even more challenging, and is not pursued.

We now turn to the estimation of the covariance kernels D_j . We describe the method proposed by Horváth et al. (2013). Set $N_j = i_j - i_{j-1}$. Before defining the sample covariance kernel \hat{D}_{j,N_j} , we need to introduce more notation. Let $X_{i_{j-1}+1}, X_{i_{j-1}+2}, \dots, X_{i_j}$ denote the j th subset of observations. Define the n th residual of the j th subset by

$$e_{j,n}(t) = X_{i_{j-1}+n}(t) - \bar{X}_{N_j}(t), \quad 1 \leq n \leq N_j, \quad 1 \leq j \leq M,$$

where $\bar{X}_{N_j}(t)$ is the subset's sample mean defined by

$$\bar{X}_{N_j}(t) = \frac{1}{N_j} (X_{i_{j-1}+1} + X_{i_{j-1}+2} + \dots + X_{i_j}).$$

The j th subset's autocovariances are defined by

$$\hat{\gamma}_{j,\ell,N_j}(t,s) = \frac{1}{N_j} \sum_{i=\ell+1}^{N_j} e_{j,i}(t) e_{j,i-\ell}(s), \quad 1 \leq j \leq M,$$

We then define the j th subset's long-run kernel estimator by

$$\hat{D}_{j,N_j}(t,s) = \hat{\gamma}_{j,0,N_j}(t,s) + \sum_{\ell=1}^{N_j-1} K\left(\frac{\ell}{h}\right) \{\hat{\gamma}_{j,\ell,N_j}(t,s) + \hat{\gamma}_{j,\ell,N_j}(s,t)\}. \quad (4.4.12)$$

Using Theorem 2 of Horváth et al. (2013), it is easy to verify that $\|\hat{U}_N^0 - U^0\| \xrightarrow{P} 0$, where \hat{U}_N^0 is defined analogously to U^0 (4.4.8) with the D_j replaced by the \hat{D}_{j,N_j} and θ_j by $N^{-1}i_j$. The kernel K must satisfy the following assumption.

Assumption 28. *The function K is continuous, bounded, $K(0) = 1$ and $K(u) = 0$ if $|u| > c$, for some $c > 0$. The smoothing bandwidth $h = h(N)$ satisfies $h(N) \rightarrow \infty$, $N^{-1}h(N) \rightarrow 0$, as $N \rightarrow \infty$.*

We conclude this section with the following consistency result proven in Section 4.7.

Theorem 29. *Suppose Assumptions 25, 26, 21 hold and there is a change point $r \in (0, 1)$ such that $\tau \neq \tau^*$, where $\tau = \tau_1 = \dots = \tau_{[Nr]}$, $\tau^* = \tau_{[Nr]+1} = \dots = \tau_N$. Then,*

$$N^{-1}\mathcal{V}_N \xrightarrow{P} \|\tau - \tau^*\|^2 \int_0^1 \{g^*(x, r)\}^2 dx, \quad (4.4.13)$$

where $\|\cdot\|$ is the norm defined in $L^2[0, 1]$ and the function g^* defined in (4.7.5).

4.5 Finite Sample Performance

In this section, we assess the empirical size and power of the procedures introduced in Sections 4.3 and 4.4. We emphasize the importance of incorporating a break point in the error structure. This section is not meant to be an extensive empirical study of yield curves, but rather a study of the proposed statistical methods. We however take care to use simulated data that closely resemble actual yield curves. For ease of reference, we begin by listing in Table 4.1 the procedures we study. We work with zero coupon US yield curves defined at maturities of 1, 3, 6, 12, 24, 36, 60, 84, 120 and 360 months. Figure 4.5 shows five consecutive yield curves.

Table 4.1: Testing procedures

Method	Description
ProjSim	Projections onto factors approach of Section 4.3; limit approximated by simulating process (4.3.11).
ProjEigen	Projections onto factors approach of Section 4.3; limit approximated by simulating the RHS of (4.3.9).
NFEigen	Nonparametric functional approach of Section 4.4; limit approximated by simulating the RHS of (4.4.11).

We begin by illustrating the behavior of the methods using three sampling periods listed in the first column of Table 4.2. In the first two periods, whose central 100 days are shown, respectively in Figures 4.1 and 4.5, we expect a change point in the mean structure. In the third period, see Figure 4.5, we do not expect such a change point. The periods were chosen in such a way that the

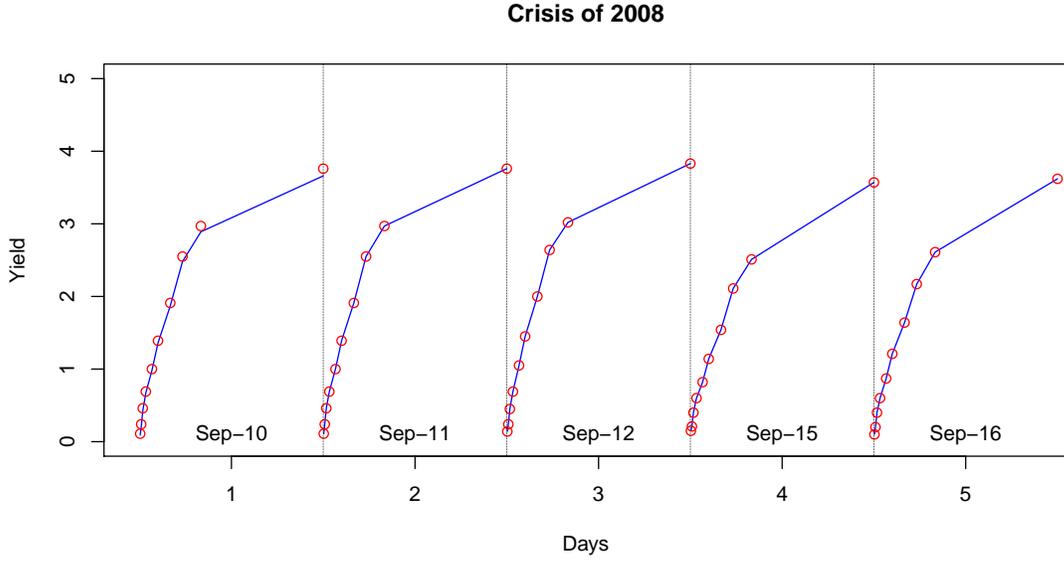


Figure 4.2: US yield curves at five business days around the Lehman Brothers collapse

potential break point in the error structure is in the middle of the sampling period. To perform the tests, we assume either no break point, designation “no” in the “Break Point” column of Table 4.2, or one break point in the middle of the sample, $\theta_1 = 1/2$, designation “yes”. We emphasize that a break point in the error structure should be viewed as an option in the application of the tests; we test for change points in the “main” mean structure. Table 4.2 shows that if a change in the mean structure exists, but a break point in the error structure is not taken into account, then the change point may not be detected. This finding is confirmed by the simulation study that we now describe.

As simulated data we use realizations of the dynamic Nelson–Siegel model

$$X_n(t) = \beta_{i,1}f_1(t, \lambda) + \beta_{i,2}f_2(t, \lambda) + \beta_{i,3}f_3(t, \lambda) + \varepsilon_i(t), \quad (4.5.1)$$

where

$$f_1(t, \lambda) = 1, \quad f_2(t, \lambda) = \frac{1 - e^{-\lambda t}}{\lambda t}, \quad f_3(t, \lambda) = \frac{1 - e^{-\lambda t}}{\lambda t} - e^{-\lambda t}. \quad (4.5.2)$$

Curves (4.5.2) are shown in Figure 5.5. The value of the parameter λ is chosen to maximize $f_3(t, \lambda)$ at the maturity of 30 months, Diebold and Li (2003). Since we simulate data on the rescaled

Table 4.2: Application of the test procedures to yield curves over three sampling periods. We expect small P-values in periods (1)–(2), large in period (3).

Sampling Period	Method	Break Point	P-value
(1) 03/20/2008 – 03/19/2009	ProjSim	yes	1.5%
	ProjEigen	yes	1.7%
	NFEigen	yes	0.1%
	ProjSim	no	87.9%
	ProjEigen	no	85.2%
	NFEigen	no	26.2%
(2) 06/30/2005 - 06/29/2006	ProjSim	yes	0.1%
	ProjEigen	yes	0.0%
	NFEigen	yes	0.2%
	ProjSim	no	56.7%
	ProjEigen	no	50.5%
	NFEigen	no	57.2%
(3) 02/16/2012 – 02/14/2013	ProjSim	yes	68.1%
	ProjEigen	yes	66.9%
	NFEigen	yes	55.8%
	ProjSim	no	80.4%
	ProjEigen	no	77.7%
	NFEigen	no	76.3%

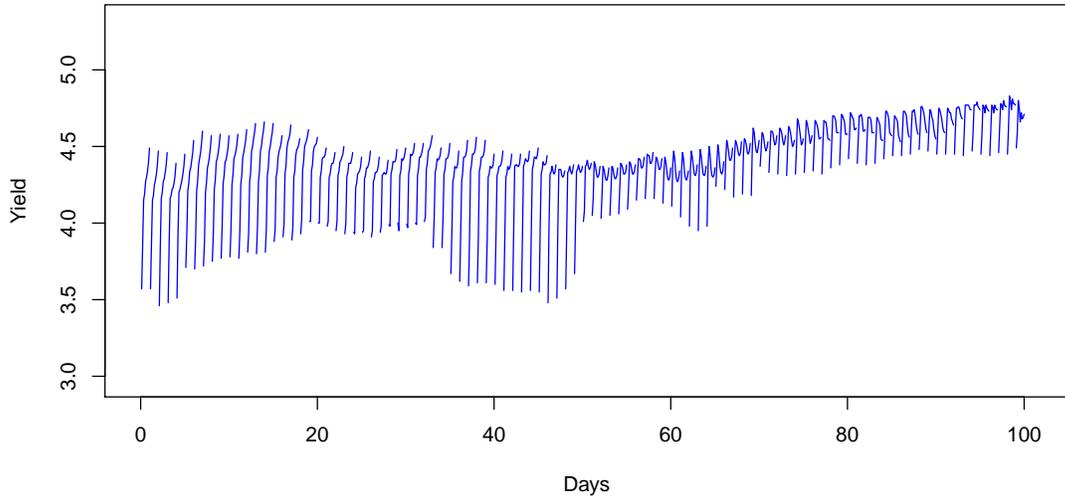


Figure 4.3: Yield curves on $N = 100$ business days form October-18-2005 to March-14-2006.

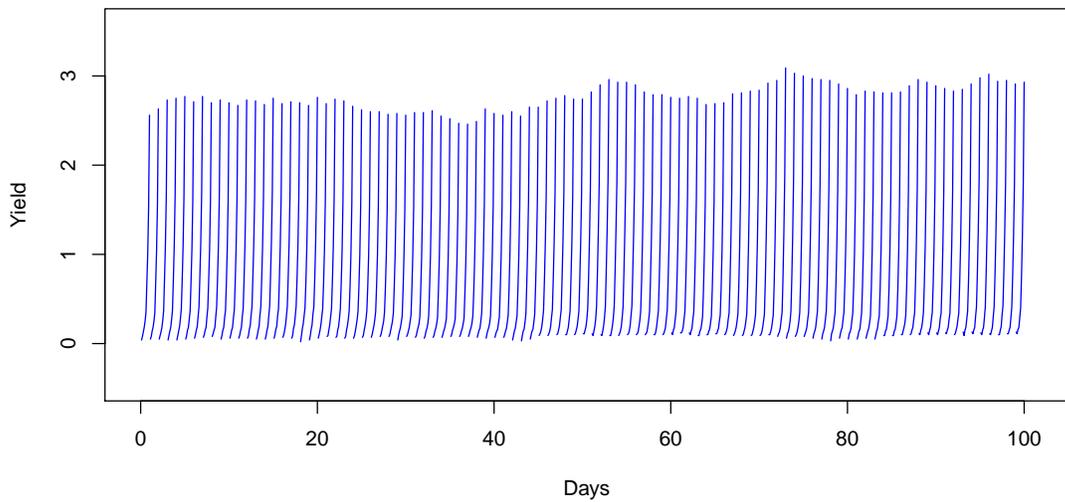


Figure 4.4: Yield curves on $N = 100$ business days from June-4-2012 to October-24-2012.

interval $[0, 1]$, we use $\lambda = 21.5194$ which maximizes $f_3(t, \lambda)$ at $t = 30/360 = 0.0833$. To assess the sensitivity of the results to the specific form of the factors, we also performed simulations using factor curves $f_1(t) = 1, f_2(t) = t, f_3(t) = t(1 - t)$. The properties of the methods and the empirical rejection rates are very similar, so we do not include the additional tables.

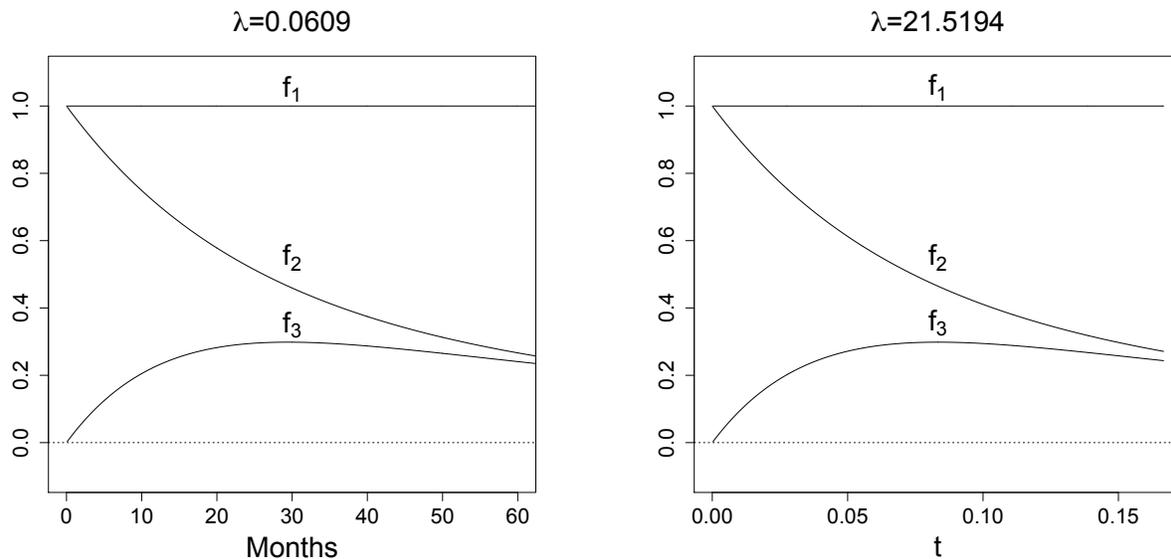


Figure 4.5: Nelson–Siegel factors $f_1(t, \lambda), f_2(t, \lambda), f_3(t, \lambda)$. Left panel: curves with $\lambda = 0.0609$ corresponding to the domain of real yield curves. Right panel: curves with $\lambda = 21.52$ corresponding to the unit interval.

The coefficients $\beta_{i,k}$ are generated as AR(1) processes

$$\beta_{i,k} = \mu_k(1 - \varphi_k) + \varphi_k\beta_{i-1,k} + u_{i,k}, \quad u_{i,k} \sim N(0, \sigma_k^2), \quad i = 1, \dots, N, \quad k = 1, 2, 3.$$

To resemble real data, the values of model parameters used in simulations are obtained as follows. For a specified sampling period, we compute least squares estimates of $\beta_{i,1}, \beta_{i,2}, \beta_{i,3}$ for each day i . This is the approach recommended by Diebold and Rudebusch (2013). Treating these estimates as a realization of an AR(1) processes, we estimate μ_k, φ_k and σ_k^2 by maximum likelihood. Following Bech and Lengwiler (2012), the presence of a break point in the error structure is simulated by using different AR(1) error variances σ_k^2 before and after the break point. Finally, the error curves

ε_i are simulated as random curves

$$\varepsilon_i(t) = \frac{2}{25}\zeta_{i1} + \frac{1}{25}\zeta_{i2} \sin(2\pi t), \quad t \in [0, 1], \quad (4.5.3)$$

chosen so that they are of the same size as the actual residuals, Figure 4.6. The series $\zeta_{i,j}$ are autoregressions defined by $\zeta_{i,j} = .9\zeta_{i-1,j} + Z_{i,j}$, $Z_{i,j} \sim N(0, 1)$, $j = 1, 2$, $i = 1, \dots, N$. The parameters chosen above represent well the error curves over periods where no break point in the error structure is visible.

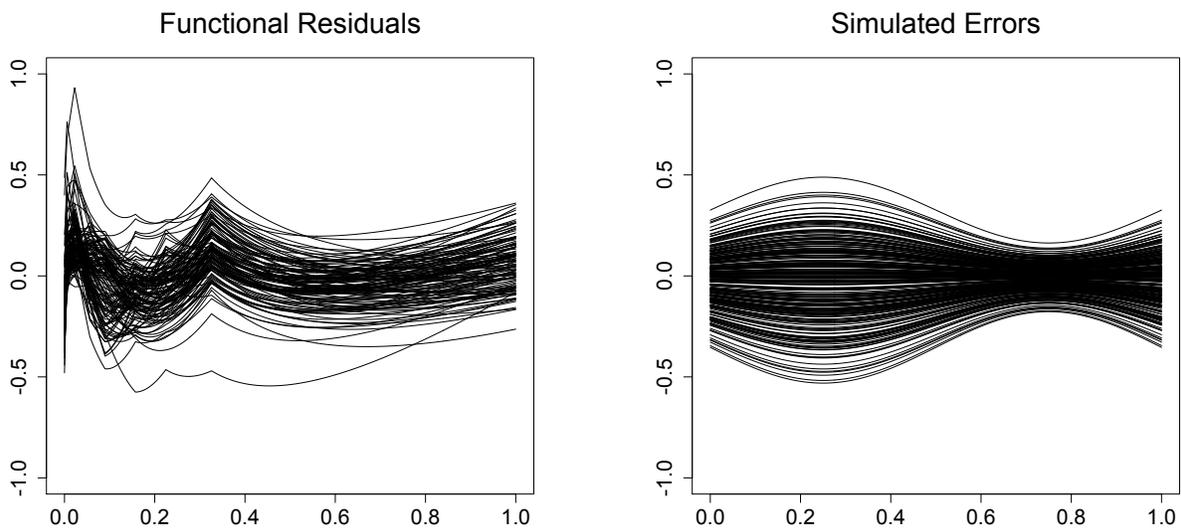


Figure 4.6: The left panel shows the residuals of the dynamic Nelson-Siegel model estimated over $N = 250$ business days from March-20-2008 to March-19-2008. The right panel shows $N = 100$ error curves simulated using Equation (4.5.3).

Using the approach described above, one can specify a large number of realistic data generating processes. We estimated several segments consisting of 250 and 125 consecutive yield curves. The following values are fairly representative, and we use them to simulate data.

AR(1) coefficients: $\varphi_1 = 0.90$, $\varphi_2 = .90$, $\varphi_3 = .90$.

AR(1) error variances:

$$\text{Var}[u_{i,1}] = 0.003, \quad i \leq i_1, \quad \text{Var}[u_{i,1}] = 0.012, \quad i > i_1,$$

$$\text{Var}[u_{i,2}] = 0.006, \quad i \leq i_1, \quad \text{Var}[u_{i,2}] = 0.026, \quad i > i_1,$$

$$\text{Var}[u_{i,3}] = 0.063, \quad i \leq i_1, \quad \text{Var}[u_{i,3}] = 0.095, \quad i > i_1.$$

We use two different locations of the break point θ_1 , 1/2 and 2/3.

Means under H_0 :

$$\boldsymbol{\mu} = \begin{pmatrix} 4.54 & -2.82 & -3.03 \end{pmatrix}^\top. \quad (4.5.4)$$

Means under $H_{A(1)}$:

$$\boldsymbol{\mu}_i = \begin{pmatrix} 4.54 \\ -2.82 \\ -3.03 \end{pmatrix}, \quad i \leq N/2; \quad \boldsymbol{\mu}_i = \begin{pmatrix} 4.20 \\ -3.00 \\ -3.20 \end{pmatrix}, \quad i > N/2. \quad (4.5.5)$$

Means under $H_{A(2)}$:

$$\boldsymbol{\mu}_i = \begin{pmatrix} 4.54 \\ -2.82 \\ -3.03 \end{pmatrix}, \quad i \leq N/2; \quad \boldsymbol{\mu}_i = \begin{pmatrix} 3.89 \\ -3.32 \\ -3.32 \end{pmatrix}, \quad i > N/2. \quad (4.5.6)$$

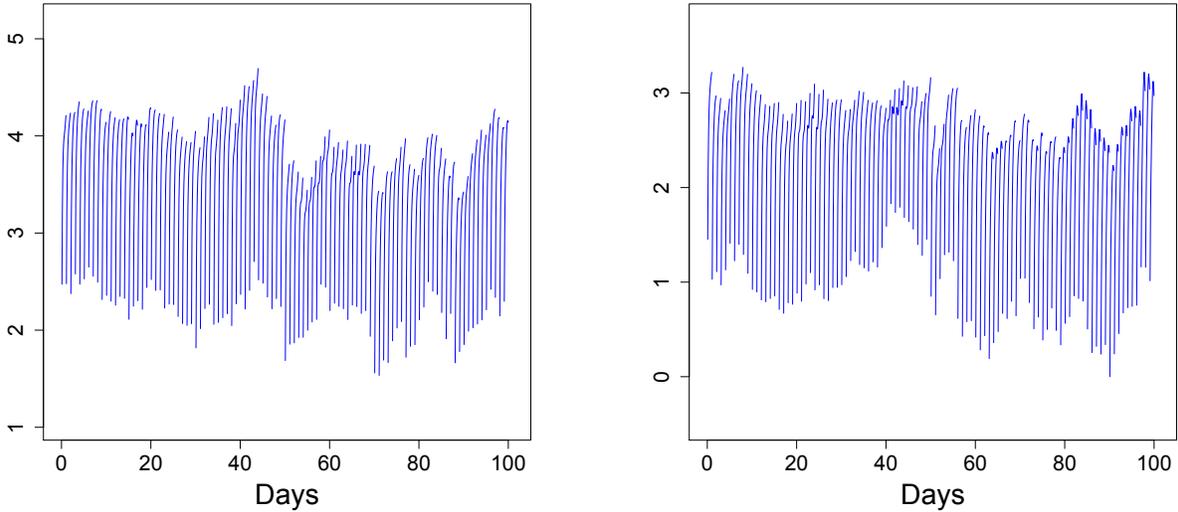


Figure 4.7: Both panels show simulated yield curves under the alternative hypothesis (4.5.5). In the left panel, including the break point leads to P-value = 0.5%, without it, P-value = 38.1%. In the right panel, when the break point is included, P-value = 0.0%, otherwise P-value = 27.8%.

The alternative $H_{A(2)}$ is a larger departure from H_0 than $H_{A(1)}$, and somewhat exaggerates the change points that can be expected in real data; $H_{A(1)}$ is very realistic. Examples of simulated processes which satisfy $H_{A(1)}$ are shown in Figure 4.7. It is often difficult to tell by eye if a change point is present or not. For the simulation study, we investigate how the power behaves when the change point coincides with the breakpoint ($CP = BP$), and when the change point is different from the breakpoint ($CP \neq BP$). In the first scenario, we use $CP = BP = N/2$, in the second $CP = N/2$ and $BP = 2N/3$.

The empirical rejection rates, based on one thousand replications are displayed in Tables 4.3, 4.4 and 4.5. For $N = 500$, if a break point is taken into account, all methods have correct size, within the chance error. For $N = 250$, the size is overinflated. As preliminary examples indicated, if the break point is not taken into account, all methods can fail to detect an existing change point with a large probability. In that case, all procedures also suffer from nonmonotonic power, i.e. power is smaller for the larger departure from H_0 . Taking the break point into account preserves monotonicity, in addition to leading to tests of practically useable power and correct size for sufficiently large sample size.

In the setting described so far, the projection methods ProjSim and ProjEigen have an automatic advantage because they use projections on the same factor curves that are used to generate the data. To further investigate the performance of these two approaches, we will impose a different factor structure on the data generating process than the structure used to compute test statistic (4.3.6). We will use the same data generating process as before, which imposes the factor structure in dynamic Nelson-Siegel model (4.5.1). However, the tests will be applied using the *intelligible*

factors introduced by Lengwiler and Lenz (2010). Their factor curves have the form

$$f_1(t, \alpha_1, \alpha_2, b_1, b_2, b_3) = 1 + \frac{\frac{1}{t}(b_3 - \log(\alpha_3)b_1)(1 - \alpha_2^t) - \frac{1}{m}(b_2 - \log(\alpha_2)b_1)(1 - \alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4},$$

$$f_2(t, \alpha_1, \alpha_2, b_1, b_2, b_3) = \frac{-\frac{1}{t}b_3(1 - \alpha_2^t) - \frac{1}{m}b_2(1 - \alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4},$$

$$f_3(t, \alpha_1, \alpha_2, b_1, b_2, b_3) = \frac{\frac{1}{t}\log(\alpha_3)(1 - \alpha_2^t) - \frac{1}{t}\log(\alpha_2)(1 - \alpha_3^t)}{\log(\alpha_2)b_3 - \log(\alpha_3)b_4}.$$

Estimating the parameters $\alpha_2, \alpha_3, b_1, b_2, b_3$ requires a nested optimization. We use the values obtained in Lengwiler and Lenz (2010), i.e. $\alpha_2 = 0.1133, \alpha_3 = 0.6798, b_1 = 0.2674, b_2 = -.4343$ and $b_3 = -.2584$. The resulting factor curves are displayed in Figure 4.8. To make the intelligible factors conform to the simulated yield curves generated on the unit interval, we transformed them in a similar manner as the Nelson–Siegel factors. The parameters of the transformed intelligible factors are $\alpha_2 = 0.1133^{30}, \alpha_3 = 0.6798^{30}, b_1 = 0.2674, b_2 = -.4343 * 30$ and $b_3 = -.2584 * 30$. These transformed intelligible factors are displayed in the right panel of Figure 4.8.

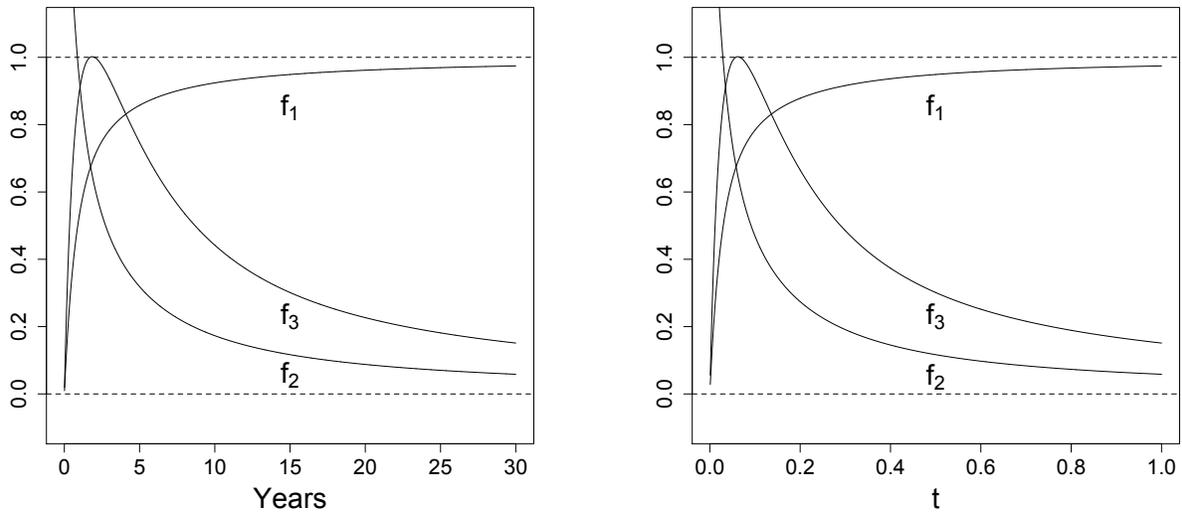


Figure 4.8: Intelligible factor curves. Left: corresponding to real yield curves; Right: transformed to the unit interval.

Tables 4.6 and 4.7 show the rejection rates in the case of a misspecified factor structure. Compared to the correctly specified structure, the sizes become somewhat overinflated, especially in the case the projEigen method. However, employing the intelligible factors in the ProjSim test improves the empirical size for $N = 250$.

Table 4.3: Empirical size and power for the **ProjSim** approach.

$(CP = BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	7.8%	61.1%	97.3%
	Yes	5%	$N = 500$	5.2%	84.6%	100.0%
	Yes	10%	$N = 250$	11.6%	70.7%	98.1%
	Yes	10%	$N = 500$	8.5%	90.3%	100.0%
	No	5%	$N = 250$	1.6%	9.8 %	2.8%
	No	5%	$N = 500$	2.8%	50.8%	39.9%
	No	10%	$N = 250$	5.7%	25.6%	12.9%
	No	10%	$N = 500$	6.3%	73.7%	77.1%
$(CP \neq BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	6.8%	65.8%	98.0%
	Yes	5%	$N = 500$	5.1%	87.5%	100.0%
	Yes	10%	$N = 250$	12.4%	75.9%	98.9%
	Yes	10%	$N = 500$	8.7%	91.5%	100%
	No	5%	$N = 250$	0.9%	7.8%	0.8%
	No	5%	$N = 500$	2.6%	48.7%	25.2%
	No	10%	$N = 250$	4.3%	23.3%	7.3%
	No	10%	$N = 500$	5.3%	72.5%	61.2%

Table 4.4: Empirical size and power for the **ProjEigen** approach.

$(CP = BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	8.5%	68.7%	99.2%
	Yes	5%	$N = 500$	5.1%	88.6%	100%
	Yes	10%	$N = 250$	13.7%	76.7%	99.5%
	Yes	10%	$N = 500$	9.4%	92.3%	100%
	No	5%	$N = 250$	2.6%	19.3%	12.6%
	No	5%	$N = 500$	3.4%	68.5%	81.5%
	No	10%	$N = 250$	8%	43.3%	41.7%
	No	10%	$N = 500$	7.3%	84.9%	97.6%
$(CP \neq BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	8.7%	71.9%	97.2%
	Yes	5%	$N = 500$	5.8%	89.2%	100%
	Yes	10%	$N = 250$	14.9%	79.2%	98.9%
	Yes	10%	$N = 500$	11.9%	94.7%	100%
	No	5%	$N = 250$	3.2%	19.4%	10.5%
	No	5%	$N = 500$	4.8%	70.2%	76%
	No	10%	$N = 250$	8.9%	43.3%	36.6%
	No	10%	$N = 500$	9.6%	86.6%	97.9%

Table 4.5: Empirical size and power for the **NFEigen** approach.

$(CP = BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	7.9%	67.7%	98.3%
	Yes	5%	$N = 500$	5.7%	90.7%	100%
	Yes	10%	$N = 250$	12.3%	76.5%	99.7%
	Yes	10%	$N = 500$	10.9%	94.2%	100%
	No	5%	$N = 250$	2.2%	13.8%	3.8%
	No	5%	$N = 500$	4.2%	64.8%	54.6%
	No	10%	$N = 250$	7.4%	37.2%	22.4%
	No	10%	$N = 500$	7.8%	83.6%	88.5%
$(CP \neq BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	9.7%	71.8%	99.1%
	Yes	5%	$N = 500$	6.1%	92.5%	100%
	Yes	10%	$N = 250$	14.6%	79.4%	99.6%
	Yes	10%	$N = 500$	12.6%	96%	100%
	No	5%	$N = 250$	3%	13.1%	2.4%
	No	5%	$N = 500$	5.1%	63.5%	34.1%
	No	10%	$N = 250$	7.7%	38.3%	15.6%
	No	10%	$N = 500$	10.1%	83.8%	80.1%

Table 4.6: Empirical size and power for the **ProjSim** approach with misspecified factors.

$(CP = BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	6.3%	62.3%	97.7%
	Yes	5%	$N = 500$	6.2%	85.7%	100.0%
	Yes	10%	$N = 250$	11.4%	71.6%	99.2%
	Yes	10%	$N = 500$	10.8%	91.2%	100.0%
	No	5%	$N = 250$	0.7%	10.3%	0.3%
	No	5%	$N = 500$	3.1%	53.6%	41.1%
	No	10%	$N = 250$	3.1%	26.6%	13.6%
	No	10%	$N = 500$	7.2%	73.6%	77.5%
$(CP \neq BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	6.3%	62.1%	97.45%
	Yes	5%	$N = 500$	4.5%	87.8%	100.0%
	Yes	10%	$N = 250$	11.3%	72.4%	98.9%
	Yes	10%	$N = 500$	8.2%	92.4%	100.0%
	No	5%	$N = 250$	1.2%	8.7%	0.9%
	No	5%	$N = 500$	2.2%	51.9%	23.0%
	No	10%	$N = 250$	4.4%	24.7%	6.5%
	No	10%	$N = 500$	5.4%	73.9%	65.7%

Table 4.7: Empirical size and power for the **ProjEigen** approach with misspecified factors.

$(CP = BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	8.4%	66.7%	98.1%
	Yes	5%	$N = 500$	6.9%	89.1%	100%
	Yes	10%	$N = 250$	13.5%	75.1%	99.1%
	Yes	10%	$N = 500$	10.5%	93.1%	100%
	No	5%	$N = 250$	1.9%	21.2%	13.2%
	No	5%	$N = 500$	3.3%	65.7%	82%
	No	10%	$N = 250$	7.1%	43.4%	39.4%
	No	10%	$N = 500$	9.5%	84.9%	97.3%
$(CP \neq BP)$	Break point	Significance level	Sample size	H_0	$H_{A(1)}$	$H_{A(2)}$
	Yes	5%	$N = 250$	8.6%	69.9%	98.9%
	Yes	5%	$N = 500$	7.4%	91.9%	100%
	Yes	10%	$N = 250$	14.5%	77.6%	99.5%
	Yes	10%	$N = 500$	11.7%	95.5%	100%
	No	5%	$N = 250$	3.5%	20%	8.8%
	No	5%	$N = 500$	6.2%	70.9%	77.6%
	No	10%	$N = 250$	9.6%	42.9%	34.6%
	No	10%	$N = 500$	9.6%	86.2%	97.2%

4.6 Summary

We introduced several asymptotic methods to test the null hypothesis that the mean structure of the a sequence of curves does not change. The tests are motivated by application to yield curves. In this context, the mean structure does not refer merely to the level of the curves, but also to their range and other aspects of their shape, most prominently concavity. We observed the importance of the error structure, which refers to the random variability in the aspects of the curves listed above.

Two tests, called projSim and projEigen, are based on projections of the factor curves, for example on the Nelson–Siegel curves or the intelligible factors of Lengwiler and Lenz (2010). The difference between them is that projSim is based on simulating data that approximately satisfy H_0 (no change point), while projEigen is based on approximating suitable eigenvalues (it also required generating an MC distribution). These two tests require a specification of a factor structure. The third approach, NFEigen, does not require any factor structure; it is a nonparametric version of the method projEigen.

Based on our data analysis and simulation study, the following conclusions can be drawn.

1. If a possible break point in the error structure is not taken into account in any of the testing procedures, an existing change point in the mean structure can fail to be detected with a large probability.
2. The test are generally well calibrated if $N = 500$.
3. If $N = 250$, all tests have a tendency to overreject at the 5% level, i.e. the empirical type I error tends to be larger than 5%.
4. If the intelligible factors are used, the empirical size of projSim test improves at the 5% nominal level, but the size of the ProjEigen test deteriorates.

The objective of this paper has been to develop the methodology and theory for change point testing in a setting of factor models commonly used in the analysis of yield curves. Our simulation study, while representative, is limited. Keeping this in mind, we recommend method ProjSim with intelligible factors and method NFEigen.

4.7 Proofs of the Asymptotic Results

4.7.1 Proofs of the results of Section 4.3

The main result of Section 4.3 is Theorem 22. It follows from Lemmas 16 and 17. We will use the following notation:

$$N_m = i_m - i_{m-1} \quad \text{and} \quad J_m(x) = \{j : i_{m-1} < j \leq [i_{m-1} + N_m x]\}.$$

Lemma 16. *Suppose Assumptions 19, 20 and 21 are satisfied. Then, for each N , we can define $M + 1$ independent Gaussian \mathbb{R}^K -valued processes $\mathbf{G}_{N,m}$, $1 \leq m \leq M + 1$, such that*

$$E\mathbf{G}_{N,m}(x) = 0 \quad \text{and} \quad E[\mathbf{G}_{N,m}(x)\mathbf{G}_{N,m}(y)^\top] = \min(x, y)\mathbf{V}_m,$$

and for all $1 \leq m \leq M + 1$,

$$\max_{0 \leq x \leq 1} \left\| N_m^{-1/2} \sum_{j \in J_m(x)} \gamma_j - \mathbf{G}_{N,m} \right\| \xrightarrow{P} 0. \quad (4.7.1)$$

Proof. Lemma 16 is a consequence of Theorem 1.1 of Berkes et al. (2013b). Their approximation principle is applied to each segment of stationarity of the γ_i . The only difference is that Berkes et al. (2013b) consider L^2 -valued processes, whereas (4.7.1) involves \mathbb{R}^K -valued processes. All arguments used by Berkes et al. (2013b) remain valid, the inner product must be interpreted as the inner product in \mathbb{R}^K rather than in L^2 .

The processes $N_m^{-1/2} \sum_{j \in J_m(x)} \gamma_j$, $1 \leq M \leq M + 1$, are not independent under our assumptions. The proof of Berkes et al. (2013b) shows that it is enough to consider ℓ -dependent sequences, cf. Assumption 20. These ℓ -dependent sequences are asymptotically independent for any $\ell \geq 1$. Therefore, we obtain the independence of the approximating sequences $\mathbf{G}_{N,m}$. \square

Lemma 17. *Suppose Assumptions 19, 20 and 21 are satisfied. Then*

$$\max_{0 \leq x \leq 1} \left\| \left\| N^{-1/2} \sum_{1 \leq j \leq Nx} \gamma_j - \mathbf{G}_N(x) \right\| \right\| \xrightarrow{P} 0,$$

where \mathbf{G}_N is a mean zero Gaussian process with covariances

$$E[\mathbf{G}_N(x)\mathbf{G}_N^\top(y)] = \sum_{j=1}^m (\theta_j - \theta_{j-1})\mathbf{V}_j + (x - \theta_m)\mathbf{V}_{m+1}, \quad \theta_m < x \leq \theta_{m+1}, \quad x \leq y \leq 1. \quad (4.7.2)$$

Proof. For $\theta_m < x \leq \theta_{m+1}$, we can write the partial sum process as

$$\sum_{1 \leq j \leq Nx} \gamma_j = \sum_{j=1}^m \sum_{i_{j-1} < i \leq i_j} \gamma_i + \sum_{i_m < i \leq Nx} \gamma_i. \quad (4.7.3)$$

Therefore, by Lemma 16,

$$N^{-1/2} \sum_{1 \leq j \leq Nx} \gamma_j = \sum_{j=1}^m \left(\frac{N_j}{N} \right)^{1/2} N_j^{-1/2} \sum_{i_{j-1} < i \leq i_j} \gamma_i + \left(\frac{N_{m+1}}{N} \right)^{1/2} N_{m+1}^{-1/2} \sum_{i_m < i \leq Nx} \gamma_i$$

can be approximated by

$$\mathbf{G}_N(x) = \sum_{j=1}^m (\theta_j - \theta_{j-1})^{1/2} \mathbf{G}_{N,j}(1) + (\theta_{m+1} - \theta_m)^{1/2} \mathbf{G}_{N,m+1} \left(\frac{x - \theta_m}{\theta_{m+1} - \theta_m} \right).$$

The process $\mathbf{G}_N(x)$ is a linear combination of Gaussian processes and hence is Gaussian. The covariance structure (4.7.2) follows from the independence of $\mathbf{G}_{N,1}, \mathbf{G}_{N,2}, \dots, \mathbf{G}_{N,M+1}$ and there covariances established in Lemma 16. \square

PROOF OF THEOREM 22: By the triangle inequality,

$$\begin{aligned} \|\alpha_N(x) - \mathbf{G}_N^0(x)\| &= \left\| \frac{1}{\sqrt{N}} \left(\sum_{1 \leq i \leq Nx} \gamma_j - \frac{[Nx]}{N} \sum_{i=1}^N \gamma_j \right) - (\mathbf{G}_N(x) - x\mathbf{G}_N(1)) \right\| \\ &\leq \left\| \frac{1}{\sqrt{N}} \sum_{1 \leq i \leq Nx} \gamma_j - \mathbf{G}_N(x) \right\| + \left\| \frac{[Nx]}{N} \frac{1}{\sqrt{N}} \sum_{i=1}^N \gamma_j - x\mathbf{G}_N(1) \right\| \end{aligned}$$

Thus, by Lemma 17, $\max_{0 \leq x \leq 1} \|\alpha_N(x) - \mathbf{G}_N^0(x)\| \xrightarrow{P} 0$, and the claim follows.

PROOF OF PROPOSITION 23: Suppose Γ is a zero mean random element in a separable Hilbert space, which satisfies $E\|\Gamma\|^2 < \infty$. Then, Γ admits the Karhunen–Loève decomposition, $\Gamma = \sum_{j=1}^{\infty} \xi_j \varphi_j$,

where the φ_j are (deterministic) orthonormal eigenvectors of the covariance operator of Γ , and ξ_j are random variables, $\xi_j = \langle X, \varphi_j \rangle$. The covariance operator of Γ is defined by $f \mapsto E[\langle \Gamma, f \rangle \Gamma]$, so the φ_j satisfy $E[\langle \Gamma, \varphi_j \rangle \Gamma] = \lambda_j \varphi_j$. If Γ is Gaussian, then the ξ_j are independent and normal with means zero and variances λ_j .

In the setting of Proposition 23, we consider the Hilbert space of R^K -valued functions

$$\mathbf{f}(x) = [f_1(x), f_2(x), \dots, f_K(x)]^\top, \quad x \in [0, 1],$$

with the inner product

$$\langle \mathbf{f}, \mathbf{g} \rangle = \sum_{k=1}^K \int_0^1 f_k(x) g_k(x) dx.$$

Direct verification shows that

$$E[\langle \Gamma, \mathbf{f} \rangle \Gamma](x) = \int_0^1 \mathbf{R}(x, y) \mathbf{f}(y) dy.$$

Therefore, the Karhunen–Loève decomposition with the eigenelements in (4.3.10) is

$$\Gamma(x) = \sum_{j=1}^{\infty} \sqrt{\lambda_j} Z_j \phi_j(x).$$

By the orthonormality of the ϕ_j , we obtain

$$\int_0^1 \|\Gamma(x)\|^2 dx = \langle \Gamma, \Gamma \rangle = \sum_{j=1}^{\infty} \lambda_j Z_j^2.$$

PROOF OF THEOREM 24: Under the change point alternative,

$$\mathbf{z}_i = \begin{cases} \mathbf{C}\boldsymbol{\mu} + \gamma_i, & 1 \leq i \leq \ell^*, \\ \mathbf{C}\boldsymbol{\mu}^* + \gamma_i, & \ell^* < i \leq N, \end{cases}$$

with $\ell^* = [Nr]$. Therefore, the CUSUM process can be expressed as

$$\boldsymbol{\alpha}_N(x) = \boldsymbol{\beta}_N(x) + N^{-1/2} g_N(x, r) \mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*),$$

where

$$\boldsymbol{\beta}_N(x) = N^{-1/2} \left(\sum_{1 \leq i \leq Nx} \gamma_i - \frac{[Nx]}{N} \sum_{i=1}^N \gamma_i \right),$$

$$g_N(x, r) = \frac{[Nx]}{N}(N - [Nr])I(\{x \leq r\}) + \frac{[Nr]}{N}(N - [Nx])I(\{x > r\}), \quad (4.7.4)$$

and $I(A)$ is the indicator function of set A . We also define the function

$$g^*(x, r) = x(1 - r)I(\{x \leq r\}) + r(1 - x)I(\{x > r\}), \quad (4.7.5)$$

and notice that for fixed $x, r \in [0, 1]$,

$$N^{-1}g_N(x, r) \rightarrow g^*(x, r), \quad \text{as } N \rightarrow \infty. \quad (4.7.6)$$

Under H_A , the Cramér–von–Mises test statistic can be expressed

$$\begin{aligned} \mathcal{C}_N &= \int_0^1 \|\boldsymbol{\alpha}_N(x)\|^2 dx \\ &= \int_0^1 \|\boldsymbol{\beta}_N(x)\|^2 dx + \int_0^1 \left\| N^{-1/2}g_N(x, r)\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*) \right\|^2 dx \\ &\quad + 2 \int_0^1 N^{-1/2}g_N(x, r)\boldsymbol{\beta}_N^\top(x)\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)dx. \end{aligned}$$

Recall the Gaussian limit process defined in Theorem 22. Then,

$$\int_0^1 \|\boldsymbol{\beta}_N(x)\|^2 dx \xrightarrow{d} \int_0^1 \|\mathbf{G}^0(x)\|^2 dx = O_P(1).$$

By (4.7.6) and the continuous mapping theorem,

$$N^{-1/2} \int_0^1 N^{-1/2}g_N(x, r)\boldsymbol{\beta}_N^\top(x)\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)dx \xrightarrow{d} \int_0^1 g^*(x, r)[\mathbf{G}^0(x)]^\top \mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)dx = O_P(1),$$

The second term is deterministic and dominates the other two terms. By (4.7.6),

$$N^{-1} \int_0^1 \left\| N^{-1/2}g_N(x, r)\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*) \right\|^2 dx \rightarrow \|\mathbf{C}(\boldsymbol{\mu} - \boldsymbol{\mu}^*)\|^2 \int_0^1 \{g^*(x, r)\}^2 dx.$$

Combining the above limits, we obtain relation (4.3.12).

4.7.2 Proofs of the results of Section 4.4

The proof of Theorem 27 is analogous to the proof of Theorem 22. Recall the notation N_j and $J_m(x)$ introduced at the beginning of Section 4.7.1. Throughout this section, we assume that the Assumptions of Theorem 27 hold.

Lemma 18 follows from Theorem 1.1 of Berkes et al. (2013b); the argument for the independence of the $M + 1$ processes $\Gamma_{N,m}$ is presented in the proof of Lemma 16.

Lemma 18. *For each N , we can define we can define $M + 1$ independent Gaussian processes $\Gamma_{N,1}, \Gamma_{N,2}, \dots, \Gamma_{N,M+1}$ such that for all $1 \leq m \leq M + 1$,*

$$E\Gamma_{N,m}(x, t) = 0, \quad E\Gamma_{N,m}(x, t)\Gamma_{N,m}(y, s) = \min(x, y)D_m(t, s)$$

and

$$\max_{0 \leq x \leq 1} \left\| N_m^{-1/2} \sum_{j \in J_m(x)} \eta_j(t) - \Gamma_{N,m}(x, t) \right\| \xrightarrow{P} 0.$$

The proof of Lemma 19 is analogous to the proof of Lemma 17, so it is omitted.

Lemma 19. *Define the process*

$$\Gamma_N(x, t) = \sum_{l=1}^m (\theta_j - \theta_{j-1})^{1/2} \Gamma_{N,j}(1, t) + (\theta_{m+1} - \theta_m)^{1/2} \Gamma_{N,m+1} \left(\frac{x - \theta_m}{\theta_{m+1} - \theta_m}, t \right),$$

$\theta_m \leq x \leq \theta_{m+1}$, $0 \leq m \leq M + 1$. Then

$$\max_{0 \leq x \leq 1} \left\| N^{-1/2} \sum_{j=1}^{\lfloor Nx \rfloor} \eta_j(t) - \Gamma_N(x, t) \right\| \xrightarrow{P} 0.$$

THE PROOF OF THEOREM 27: Under $H_0 : \tau_1 = \tau_2 = \dots = \tau_N = \tau$, so

$$\sum_{1 \leq i \leq Nx} X_i(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{i=1}^N X_i(t) = \sum_{1 \leq i \leq Nx} \eta_i(t) - \frac{\lfloor Nx \rfloor}{N} \sum_{i=1}^N \eta_i(t).$$

It remains to apply Lemma 19 and the triangle inequality, analogously as in the proof of Theorem 22.

PROOF OF THEOREM 29: The following proof is similar to the proof of Theorem 24. Under the change point alternative,

$$X_i(t) = \begin{cases} \tau(t) + \eta_i(t), & 1 \leq i \leq \ell^*, \\ \tau^*(t) + \eta_i(t), & \ell^* < i \leq N, \end{cases}$$

with $\ell^* = \lfloor Nr \rfloor$. The CUSUM process can then be expressed as

$$\alpha_N(x, t) = \beta_N(x, t) + N^{-1/2} g_N(x, r)(\tau(t) - \tau^*(t)),$$

where

$$\beta_N(x, t) = N^{-1/2} \left(\sum_{1 \leq i \leq Nx} \eta_i(t) - \frac{[Nx]}{N} \sum_{i=1}^N \eta_i(t) \right),$$

and $g_N(x, r)$ is the function defined in Equation (4.7.4). Under H_A , the Cramér–von–Mises test statistic can be expressed as

$$\begin{aligned} \mathcal{V}_N &= \iint \alpha_N^2(x, t) dt dx \\ &= \iint \beta_N^2(x, t) dt dx + \|\tau - \tau^*\|^2 \int_0^1 N^{-1} g_N^2(x, r) dx \\ &\quad + 2 \iint N^{-1/2} g_N(x, r) \beta_N(x, t) (\tau(t) - \tau^*(t)) dt dx, \end{aligned}$$

where $\|\cdot\|$ is the norm defined in $L^2[0, 1]$. By the Gaussian limit process defined in Theorem 27,

$$\iint \beta_N^2(x, t) dx dt \xrightarrow{d} \iint (\Gamma^0(x, t))^2 dx dt = O_P(1).$$

By (4.7.6) and the continuous mapping theorem,

$$N^{-1/2} \iint N^{-1/2} g_N(x, r) \beta_N(x, t) (\tau(t) - \tau^*(t)) dt dx \xrightarrow{d} \int_0^1 g^*(x, r) \Gamma^0(x, t) (\tau(t) - \tau^*(t)) dt dx = O_P(1),$$

The second term is deterministic and dominates the other two terms. By (4.7.6),

$$N^{-1} \|\tau - \tau^*\|^2 \int_0^1 N^{-1} g_N^2(x, r) dx \rightarrow \|\tau - \tau^*\|^2 \int_0^1 \{g^*(x, r)\}^2 dx.$$

Combining the above limits, we obtain relation (4.4.13).

4.8 Details of the Numerical Implementation of the Tests

This section provides details of the numerical implementation of the methods proposed in this paper. We use abbreviations introduced in Table 4.1.

4.8.0.0.1 Method ProjSim To lighten the notation, we describe the method in case of a single break point, $M = 1$. Given functional observations X_1, X_2, \dots, X_N , we proceed as follows. We estimate the long-run covariance matrices $\widehat{\mathbf{V}}_1$ and $\widehat{\mathbf{V}}_2$, respectively, of the series $\mathbf{data}_1 = \{\mathbf{z}_i, i \leq i_1\}$

and $\mathbf{data}_2 = \{\mathbf{z}_i, i > i_1\}$. This allows us to generate a large number R , say $R = 10E4$ of processes $\{\widehat{G}_r^0(x), x \in [0, 1]\}$, as described following (4.3.11). For each replication r , we compute $\mathcal{C}_N^{(r)} = \int_0^1 \|G_r^0(x)\|^2 dx$. The P-value is computed as the fraction of the $\mathcal{C}_N^{(r)}$ which exceed \mathcal{C}_N computed from the data. Critical values are the empirical quantiles of the $\mathcal{C}_N^{(r)}, 1 \leq r \leq R$.

To compute the long-run variance matrices, we use the Newey-West estimator implemented in the R package `sandwich`. The main function used is `lrvar()`. One important point is that the function `lrvar()` gives the long-run variance of the sample mean and hence is scaled by the sample size. To adjust for this, we need to multiply the function by the j th subset's sample size. More specifically, we estimate \mathbf{V}_j by

$$\widehat{\mathbf{V}}_j = N_j * \text{lrvar}(\mathbf{data}_j, \text{type}="Newey-West", \text{prewhite}=\text{TRUE}, \text{adjust}=\text{FALSE}), \quad j = 1, 2,$$

where N_j is the sample size of the j th subset. Another important point is that for each estimated long-run covariance matrix, we apply prewhiting.

4.8.0.0.2 Method ProjEigen The limit distribution of the statistic \mathcal{C}_N is approximated as

$$\int_0^1 \|\mathbf{G}^0(x)\|^2 dx = \sum_{j=1}^{\infty} \lambda_j Z_j^2 \approx \sum_{j=1}^J \widehat{\lambda}_j Z_j^2 = \widehat{\mathcal{C}}_J,$$

where the Z_j are i.i.d. standard normal and the $\widehat{\lambda}_j$ are eigenvalues defined by

$$\int_0^1 \widehat{\mathbf{R}}^0(x, y) \widehat{\phi}_j(y) dy = \widehat{\lambda}_j \widehat{\phi}_j(x).$$

For $0 \leq x \leq y \leq 1, i_m \leq Nx \leq i_{m+1}$ and $i_{m'} \leq Ny \leq i_{m'+1}$, the kernel $\widehat{\mathbf{R}}^0$ is defined by

$$\begin{aligned} \widehat{\mathbf{R}}_N^0(x, y) = & (1 - y) \left[\sum_{j=1}^m (\widehat{\theta}_j - \widehat{\theta}_{j-1}) \widehat{\mathbf{V}}_j + (x - \widehat{\theta}_m) \widehat{\mathbf{V}}_{m+1} \right] \\ & - x \left[\sum_{j=1}^{m'} (\widehat{\theta}_j - \widehat{\theta}_{j-1}) \widehat{\mathbf{V}}_j + (y - \widehat{\theta}_{m'}) \widehat{\mathbf{V}}_{m'+1} \right] \\ & + xy \sum_{j=1}^{M+1} (\widehat{\theta}_j - \widehat{\theta}_{j-1}) \widehat{\mathbf{V}}_j, \end{aligned}$$

where $\hat{\theta}_j = N^{-1}i_j$ and where $\hat{\mathbf{V}}_j$ are the estimators defined above.

Using orthonormal basis functions $\psi_i(y), 0 \leq y \leq 1$, and the standard orthonormal basis $\{\mathbf{e}_i\}_{i=1}^K$ of \mathbb{R}^K , we express $\hat{\mathbf{R}}_N^0(x, y)$ as

$$\hat{\mathbf{R}}_0(x, y) = \sum_{i,j=1}^K \sum_{k,l=1}^{\infty} a_{ijkl} \mathbf{e}_i \mathbf{e}_j^\top \psi_k(x) \psi_l(y),$$

with

$$a_{ijkl} = \int \mathbf{e}_i^\top \hat{\mathbf{R}}_0(x, y) \mathbf{e}_j \psi_k(x) \psi_l(y) dx dy. \quad (4.8.1)$$

Using the first p functions ψ_i , the eigenvalues $\hat{\lambda}_n$ are computed by solving the equation

$$\int_0^1 \sum_{i,j=1}^K \sum_{k,l=1}^p a_{ijkl} \psi_k(x) \psi_l(y) \mathbf{e}_i \mathbf{e}_j^\top \hat{\phi}_n(y) dy = \hat{\lambda}_n \hat{\phi}_n(y),$$

which is reduced to finding the eigenvalues of a matrix \mathbf{A} whose components are given by

$$A_{\mu\nu} = a_{(\lfloor \frac{\mu-1}{p} \rfloor + 1), (\lfloor \frac{\nu-1}{p} \rfloor + 1), (\mu - \lfloor \frac{\mu-1}{p} \rfloor p), (\nu - \lfloor \frac{\nu-1}{p} \rfloor p)} \quad \text{for } \mu, \nu = 1, \dots, K \times p,$$

with a_{ijkl} given by (4.8.1). We used $\psi_j(x) = \sqrt{2} \sin(j\pi x), j = 1, \dots, p = 64$ ($K = 3$).

Using $J = N/5$, we compute the replications

$$\hat{\mathcal{C}}^{(r)} = \sum_{j=1}^{N/5} \hat{\lambda}_j Z_j^2, \quad r = 1, \dots, 10E4.$$

The resulting empirical distribution is used to compute P-values or critical values.

4.8.0.0.3 Method NFEigen The key difficulty is to compute the eigenvalues $\hat{\lambda}_j$ defined by

$$\int_0^1 \int_0^1 \hat{U}^0(x, y, t, s) \hat{\phi}_j(y, s) dy ds = \hat{\lambda}_j \hat{\phi}_j(x, t).$$

In contrast to method ProjEigen, the readily available long-run covariance matrix estimators cannot be used.

To estimate the \hat{D}_{j, N_j} defined in (4.4.12), we use the flat-top kernel

$$K(t) = \begin{cases} 1, & |t| < 0.1, \\ 1.1 - |t|, & 0.1 \leq |t| < 1.1, \\ 0, & 1.1 \leq |t|, \end{cases}$$

with bandwidth parameter $h = 16$, which works well for sample size $N = 250$ and $N = 500$.

Using a complete orthonormal set $\{\psi_i\}_{i=1}^{\infty} \in L^2([0, 1])$, the function $\hat{U}^0(x, y, t, s)$ is expressed as

$$\hat{U}^0(x, y, t, s) = \sum_{i,j,k,l=1}^{\infty} a_{ijkl} \psi_i(x) \psi_j(y) \psi_k(t) \psi_l(s),$$

with

$$a_{ijkl} = \int \hat{U}^0(x, y, t, s) \psi_i(x) \psi_j(y) \psi_k(t) \psi_l(s) dx dy dt ds. \quad (4.8.2)$$

We set $a_{ijkl} = 0$ for $i, j, k, l > p$. The eigenvalue $\hat{\lambda}_n$ is then computed by solving the equation

$$\iint \sum_{i,j,k,l=1}^p a_{ijkl} \psi_i(x) \psi_j(y) \psi_k(t) \psi_l(s) \hat{\phi}_n(y, s) dy ds = \hat{\lambda}_n \hat{\phi}_n(x, t). \quad (4.8.3)$$

Expanding $\hat{\phi}_n$ as

$$\hat{\phi}_n(x, t) = \sum_{i,j=1}^{\infty} d_{ij}^{(n)} \psi_i(x) \psi_j(t)$$

and recalling that the functions ψ_j are orthonormal, equation (4.8.3) can be further simplified to a $p^2 \times p^2$ system of linear equations

$$\sum_{j,l=1}^p a_{ijkl} d_{jl}^{(n)} = \hat{\lambda}_n d_{ik}^{(n)} \quad \text{for } i, k = 1, \dots, p.$$

Thus the eigenvalues $\hat{\lambda}_n$ can be estimated by finding the eigenvalues of a matrix \mathbf{A} whose components are given by the projection coefficients a_{ijkl} , e.g.

$$A_{\mu\nu} = a_{\left(\lfloor \frac{\mu-1}{p} \rfloor + 1\right), \left(\lfloor \frac{\nu-1}{p} \rfloor + 1\right), \left(\mu - \lfloor \frac{\mu-1}{p} \rfloor p\right), \left(\nu - \lfloor \frac{\nu-1}{p} \rfloor p\right)} \quad \text{for } \mu, \nu = 1, \dots, p^2,$$

with a_{ijkl} given by (4.8.2).

We used $p = 64$, $\psi_j(x) = \sqrt{2} \sin(j\pi x)$, and $J = 200$ first eigenvalues to compute the replications of $\sum_{j=1}^J \hat{\lambda}_j Z_j^2$.

Chapter 5

DETERMINATION OF THE INTERVAL OF INCREASING CUMULATIVE RETURNS PRIOR TO MACROECONOMIC ANNOUNCEMENTS

5.1 Introduction

Inspired by Lucca and Moench (2015), we develop a data-driven approach for detecting intervals of statistically increasing cumulative returns in the presence of macroeconomic announcements. We examine the intraday cumulative return curve in the context of Functional Data Analysis (FDA). This is done by finding an interval with a positive derivative. To construct the confidence band, we develop two methods: asymptotic and bootstrap. Our approach can precisely determine the turning points of the derivatives and thus detect the exact interval of the “pre-FOMC.” We then apply our approach to re-examine the pre-FOMC drift phenomenon. Our results reveal more information about how the return curves change previous to the FOMC meeting. This methodology can be applied to investigate the impact of events on high frequency asset returns.

Lucca and Moench (2015) document an interesting phenomenon that occurs on the day of the scheduled Federal Open Market Committee (FOMC) announcement. The authors find large average excess returns on U.S. equities in anticipation of monetary policy decisions made at scheduled meetings of the Federal Open Market Committee (FOMC). This phenomenon, which was observed over the last few decades, is named as the “pre-FOMC drift.” These pre-FOMC returns have increased over time and account for sizable fractions of total annual realized stock returns. To infer upon this phenomenon, Lucca and Moench (2015) formally assesses the magnitudes of excess stock market returns prior to scheduled FOMC announcements by running a simple dummy variable regression model:

$$rx_t = \beta_0 + \beta_1 \mathbf{1}_t(\text{FOMC}) + \beta_x X_t + \epsilon_t. \quad (5.1.1)$$

The response rx_t denotes the cum-dividend log excess return on the SPX over the risk-free rate in percentage points. The explanatory variable is a dummy variable equal to one on scheduled pre-FOMC announcement windows and zero otherwise. X_t is a vector of additional controls.

Although the “dummy” approach is generally used in the macroeconomic announcement literature, the method lacks novelty and the ability to detect precisely where the cumulative returns are statistically increasing and decreasing. In order to assess the impact of macroeconomic announcements, most if not all literature artificially separate data into pre- and post- announcement and run similar regression analysis or simple statistics to show the difference (or similarity) pre- and post- announcement. Ideally, the data can tell us where to cut it instead of using brute force.

In this paper we want to answer the following questions:

1. Do the cumulative returns change direction at exactly the announcement time (i.e. 14:15 for FOMC meeting)?
2. If not, can we detect the change points statistically?

To answer these injuries, we develop a data-driven approach for detecting intervals of statistically increasing cumulative returns in the presence of macroeconomic announcements. This is done by finding an interval with a positive derivative and extending the ideas developed by Liu and Müller (2009).

Liu and Müller (2009) develop a method of recovering the derivative of sparsely observed observations. The authors accomplish this task by expanding on traditional local polynomial smoothing, generalizing it to a FDA setting. Their approach in conjunction with functional principal component scores for sparse data leads to a practical solution of recovering derivatives for sparsely observed functions. The authors apply their developed methods to online auction price dynamics.

Investigated by Lucca and Moench (2015), the pre-FOMC drift is a phenomenon that deserves more attention. To study this drift in more detail, we expand on the ideas of Liu and Müller

(2009) and develop a method of recovering the derivative of intraday cumulative returns. Our proposed method accounts for the uncertainty related to our estimation, which produces $(1-p)100\%$ confidence bands of the derivative. To construct the $(1-p)100\%$ confidence band, we develop two methods: asymptotic and bootstrap. These new approaches can then be used to estimate time intervals of statistically increasing intraday cumulative returns under the presence of scheduled macroeconomic announcements.

We apply the asymptotic procedure to the same sampling period presented in Lucca and Moench (2015). We construct an interval of statistically increasing cumulative returns of the S&P500 index using 132 trading days in which the FOMC announcement occurred. This sampling period begins September-1994 and ends March-2011. For this motivating example, we only look at a single day, where Lucca and Moench (2015) investigated a three day period centered about the announcement day. When using this sampling period, the asymptotic confidence band crosses zero 47 minutes before the FOMC announcement. The announcement occurs at 14:15 and the interval of statistically increasing returns begins 09:30 and ends at 13:28. This procedure is illustrated in Figure 5.1.

Then we apply the approach to several other combinations: WTI Sweet Light Crude Oil Futures returns before the Weekly Petroleum Status Report, i.e., the crude oil inventory announcements of the U.S. Energy Information Administration (EIA).

In addition to the work cited above, this paper is related to different streams of literature. High-frequency, and intraday financial data in general, have been an important focus of research in finance, econometrics and statistics for over two decades. The literature is enormous; an introduction is given in Chapters 5 and 6 of Tsay (2005), and to list a few influential publications, which contain literature overview, we cite Engle and Russel (2004), Barndorff-Nielsen and Shephard (2004), Hayashi and Yoshida (2005), Wang and Zou (2010), and Andersen et al. (2012). The impact of scheduled macroeconomic news on assets has been an important research topic over the last few years. There have been several studies that established the impact or lack thereof scheduled

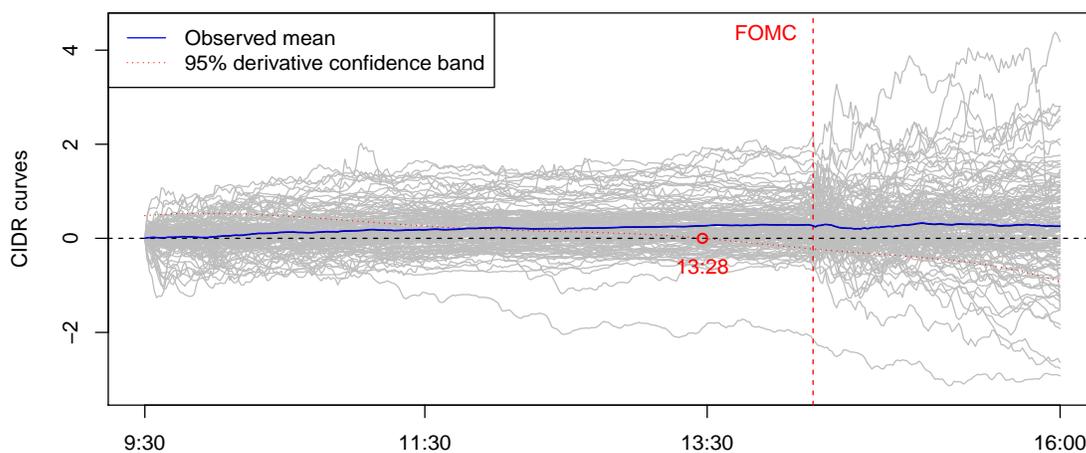


Figure 5.1: Intraday Cumulative Returns of S&P500 on the FOMC Days

The S&P500 intraday cumulative returns on the scheduled FOMC announcement days from September 1994 through March 2011. The domain of each curve consists of a single trading day (390 minutes). The FOMC announcement is located at $t = 285$ minutes which corresponds to 14:15. The asymptotic procedure is used to estimate when the true maximum occurs with 95% confidence. The 95% asymptotic band crosses zero at $t = 238$ minutes or 13:28.

announcements on various types of assets, see e.g. Elder et al. (2011) for an inquiry of this type which also contains references to earlier work.

The paper is organized as follows. Section 5.2 introduces the statistical model and relevant definitions of functional data analysis. Section 5.3 gives details on how the derivative is estimated, which incorporates a functional analogue of local polynomial regression. Section 5.4 describes the asymptotic and bootstrap procedures used to construct the confidence bands, which includes numerical implementations and large sample approximations. Section 5.5 discuss the data used in the empirical study. Section 5.6 is an extensive empirical study. We apply the two developed procedures to selected financial indexes in the presence of scheduled macroeconomic announcements. Section 5.7 is a simulation study assessing the performance of the asymptotic and bootstrap procedures. Section 5.8 summarizes the statistical and financial findings. Section 5.9 includes proofs of the propositions.

5.2 Statistical Model and Assumptions

We consider Cumulative Intraday Returns (CIDRs) of financial assets. The CIDR's are defined by

$$R_i(t) = 100\{\log P_i(t) - \log P_i(0)\}, \quad (5.2.1)$$

where $P_i(t)$ is the price of the asset at time t on day i . Since $R_i(t) \approx 100(P_i(t) - P_i(0))/P_i(0)$, the CIDR's reflect the shape of the daily price curves. In contrast to the price curves $P_i(t)$, the $R_i(t)$ form an approximately stationary functional time series.

To answer the questions discussed in the Introduction with statistical significance, a suitable statistical model is needed. We postulate that the observed CIDR curves, $R_1(t), \dots, R_N(t)$, follow the model

$$R_i(t) = \mu(t) + \varepsilon_i(t), \quad i = 1, \dots, N, \quad t \in [0, 1], \quad (5.2.2)$$

where μ is the unknown mean function, $\mu(t) = ER_i(t)$, and the curves ε_i are independent identically distributed random functions with zero mean, that is, $E\varepsilon_i(t) = 0$.

Without loss of generality, we assume that all functions are defined on the unit interval $[0, 1]$. This can always be achieved by an affine transformation of time. From now on, integration without indicating the limits, always refers to integration over $[0, 1]$. We assume that μ belongs to the space L^2 of square integrable functions, i.e. $\int \mu^2(t)dt < \infty$. The error functions are also assumed to be in L^2 and to be square integrable, i.e. $E \int \varepsilon_i^2(t)dt < \infty$. The covariance operator C of random functions $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ is defined by

$$C(y)(t) = \int_0^1 c(s, t)y(s)ds, \quad y \in L^2[0, 1], \quad (5.2.3)$$

where the kernel $c(\cdot, \cdot)$ is the covariance function defined by

$$c(s, t) = E[\varepsilon(s)\varepsilon(t)] = E[(R(t) - \mu(t))(R(s) - \mu(s))], \quad 0 \leq s, t \leq 1. \quad (5.2.4)$$

Model (5.2.2) can intuitively be thought of as a single path of a Brownian motion. A more detailed background can be found in several recent monographs, see e.g. Chapters 2 and 3 of Horváth and Kokoszka (2012).

5.3 Estimation of the Derivative

In model (5.2.2), the mean function μ is a parameter to which inference applies. We are specifically interested in estimating the derivative $\mu'(t)$ at any given time point t and finding a confidence interval for this derivative. We use a framework akin to the local polynomial smoothing, see Fan and Gijbels (1996), but suitably modified to apply to model (5.2.2). A local polynomial estimator is calculated by minimizing the objective function

$$S_N(t; \alpha) = \sum_{i=1}^N \int_0^1 \{R_i(s) - (\alpha_0 + \alpha_1(s-t) + \dots + \alpha_{p-1}(s-t)^{p-1})\}^2 K\left(\frac{s-t}{h}\right) ds, \quad (5.3.1)$$

with respect to $\alpha_0, \alpha_1, \dots, \alpha_{p-1}$. We denote the estimated derivative by $\hat{\mu}'(t) = \hat{\alpha}_1$. Minimizing (5.3.1) is a different procedure than the standard local polynomial smoothing of a single function observed with noise. It makes sense only in the context of model (5.2.2), and is similar to the

approach of Liu and Müller (2009), who used a sum rather than the integral, as they focused on sparsely observed curves. A review of the local polynomial smoothing is presented in Section 5.9.1. The CIDR curves are smooth and densely observed, so the integral notation is appropriate.

5.3.1 Estimation of model (5.2.2)

In the framework of model (5.2.2), minimizing (5.3.1) for a fixed t leads to an estimator of the parameter vector

$$\boldsymbol{\alpha}(t) = \boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_{p-1}]'. \quad (5.3.2)$$

It is possible to compute the minimum using the Newton's method; it can be shown that the Hessian does not depend on $\boldsymbol{\alpha}$ and can be calculated in closed form for any specific kernel K . As a matter of practical implementation, it is however easier to relate the minimization of (5.3.1) to the minimization of (5.9.1) for which computationally efficient R implementations exist. We now explain how this is done.

Partition the unit interval into $J + 1$ intervals using equispaced points $0 = s_0 < s_1 < \dots < s_J =$

1. For fixed time t , define the functional response vector and design matrix by

$$\mathbf{R}_i = \begin{bmatrix} R_i(s_1) \\ R_i(s_2) \\ \vdots \\ R_i(s_J) \end{bmatrix}, \quad \mathbf{U}_t = \begin{bmatrix} 1 & (s_1 - t) & (s_1 - t)^2 & \cdots & (s_1 - t)^{p-1} \\ 1 & (s_2 - t) & (s_2 - t)^2 & \cdots & (s_2 - t)^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (s_J - t) & (s_J - t)^2 & \cdots & (s_J - t)^{p-1} \end{bmatrix}. \quad (5.3.3)$$

Define the diagonal weight matrix by

$$\boldsymbol{\Omega}_{t,h} = \text{diag}\left\{K\left(\frac{s_1 - t}{h}\right), K\left(\frac{s_2 - t}{h}\right), \dots, K\left(\frac{s_J - t}{h}\right)\right\}. \quad (5.3.4)$$

Notice that the weight matrices (5.9.2) and (5.3.4) are basically the same except that in (5.9.2) N is the number of points at which a single curve is observed and in (5.3.4), J is the number of partition points which can be arbitrarily chosen. The connection between the two settings is established by the following proposition.

Proposition 30. *For every fixed t , the least squares objective function $S_N(t; \boldsymbol{\alpha})$ defined in (5.3.1) achieves its minimum at*

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_{p-1} \end{bmatrix} = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\alpha}}_i(t),$$

where

$$\hat{\boldsymbol{\alpha}}_i(t) = \hat{\boldsymbol{\alpha}}_i = \begin{bmatrix} \hat{\alpha}_{0,i} \\ \hat{\alpha}_{1,i} \\ \vdots \\ \hat{\alpha}_{p-1,i} \end{bmatrix} = (\mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{R}_i. \quad (5.3.5)$$

The proof of Proposition 30 is given in Section 5.9.2. Comparing equation (5.3.5) to (5.9.3), we see that each $\hat{\boldsymbol{\alpha}}_i$ can be computed using available software and treating the vector \mathbf{R}_i defined in (5.3.3) as a curve observed at points s_j . For the empirical study in Section 5.6 and simulation study in Section 5.7, the R function `locpoly(\cdot)` is implemented to compute (5.3.5). This function can be found in the R package `KernSmooth`. In all calculations, we use $p = 3$ (local quadratic fit) because it is known to give well-behaved estimates of the local slope.

5.3.2 Asymptotic approximation to the distribution of $\hat{\mu}'(t)$

To lighten the notation, set

$$\mathbf{Q}_{t,h} = (\mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_{t,h}, \quad (5.3.6)$$

so that $\hat{\boldsymbol{\alpha}}_i(t) = \mathbf{Q}_{t,h} \mathbf{R}_i$. Notice that $\mathbf{Q}_{t,h}$ is a function of both the fixed time $t \in [0, 1]$ and the bandwidth h . In contrast to the usual asymptotic theory for local smoothing, we do not consider $h \rightarrow 0$ with the sample size. This is because we want to obtain an approximation for an actual h we use. In such a setting, $\hat{\boldsymbol{\alpha}}(t)$ is in general a biased estimator of $\boldsymbol{\alpha}(t)$, but under weak assumptions on the local structure of the function μ in (5.2.2), the asymptotic approximation described in Section 5.4 can be justified. Proposition 31 gives the details of the limiting distribution with fixed h . Before stating it, we present some preliminary notation and arguments.

The expected value of the response vector \mathbf{R}_i defined in (5.3.3), is

$$\mathbb{E}[\mathbf{R}_i] = \begin{bmatrix} \mathbb{E}R_i(s_1) \\ \mathbb{E}R_i(s_2) \\ \vdots \\ \mathbb{E}R_i(s_J) \end{bmatrix} = \begin{bmatrix} \mu(s_1) \\ \mu(s_2) \\ \vdots \\ \mu(s_J) \end{bmatrix} =: \boldsymbol{\mu},$$

which implies $E[\hat{\boldsymbol{\alpha}}(t)] = \mathbf{Q}_{t,h}\boldsymbol{\mu}$. In Proposition 30, we discretize the unit interval into J points. The mean vector $\boldsymbol{\mu}$ is the discretized representation of mean function μ in model (5.2.2). The covariance function c defined in (5.2.4) can also be represented in a discrete form

$$\boldsymbol{\Sigma}_\varepsilon = \begin{bmatrix} c(s_1, s_1) & c(s_1, s_2) & \cdots & c(s_1, s_J) \\ c(s_2, s_1) & c(s_2, s_2) & \cdots & c(s_2, s_J) \\ \vdots & \vdots & \ddots & \vdots \\ c(s_J, s_1) & c(s_J, s_2) & \cdots & c(s_J, s_J) \end{bmatrix}, \quad (5.3.7)$$

where $c(\cdot, \cdot)$ is the covariance function defined in Equation (5.2.4). Denote the covariance matrix of estimator $\hat{\boldsymbol{\alpha}}(t)$ by $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}$. This matrix can be expressed as

$$\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}} = \text{Var}[\hat{\boldsymbol{\alpha}}_i(t)] = \text{Var}[\mathbf{Q}_{t,h}\mathbf{R}_i] = \mathbf{Q}_{t,h} \text{Var}[\mathbf{R}_i] \mathbf{Q}'_{t,h} = \mathbf{Q}_{t,h} \boldsymbol{\Sigma}_\varepsilon \mathbf{Q}'_{t,h}. \quad (5.3.8)$$

Proposition 31. *Under model (5.2.2), for fixed $t \in [0, 1]$ and $h > 0$,*

$$\sqrt{N}(\hat{\boldsymbol{\alpha}}(t) - \mathbf{Q}_{t,h}\boldsymbol{\mu}) \xrightarrow{\mathcal{D}} N(0, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}),$$

where $\mathbf{Q}_{t,h}$ is defined in Equation (5.3.6) and $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}$ is defined in Equation (5.3.8).

Proof. The $\mathbf{R}_1, \mathbf{R}_2, \dots, \mathbf{R}_N$ are independent and identically distributed random vectors with common mean $\mathbf{Q}_{t,h}\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}$. The result follows immediately from the multivariate central limit theorem, e.g. Lütkepohl (2005) p. 691. \square

Proposition 31 implies that

$$\sqrt{N}(\hat{\alpha}_1(t) - \alpha_1^*(t)) \xrightarrow{\mathcal{D}} N(0, \sigma_\varepsilon^2),$$

where $\alpha_1^*(t)$ is the second component of vector $\mathbf{Q}_{t,h}\boldsymbol{\mu}$ and σ_ε^2 is the second diagonal element of covariance matrix $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\alpha}}}$. To construct the confidence band of interest, we need to estimate σ_ε^2 and address the bias $\alpha_1^*(t) - \alpha_1(t)$.

Define the empirical covariance function \hat{c} by

$$\hat{c}(s, t) = \frac{1}{N} \sum_{i=1}^N (R_i(t) - \bar{R}(t))(R_i(s) - \bar{R}(s)), \quad 0 \leq s, t \leq 1,$$

where the sample mean is defined by

$$\bar{R}(s) = \frac{1}{N} \sum_{i=1}^N R_i(s), \quad 0 \leq s \leq 1.$$

Discretizing the unit square and evaluating the respective points in the empirical covariance function $\hat{c}(s, t)$ gives a natural estimator of $\boldsymbol{\Sigma}_\varepsilon$. More specifically,

$$\hat{\boldsymbol{\Sigma}}_\varepsilon = \begin{bmatrix} \hat{c}(s_1, s_1) & \hat{c}(s_1, s_2) & \cdots & \hat{c}(s_1, s_J) \\ \hat{c}(s_2, s_1) & \hat{c}(s_2, s_2) & \cdots & \hat{c}(s_2, s_J) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{c}(s_J, s_1) & \hat{c}(s_J, s_2) & \cdots & \hat{c}(s_J, s_J) \end{bmatrix}. \quad (5.3.9)$$

The bias $\alpha_1^*(t) - \alpha_1(t)$ cannot, in general, be eliminated for a fixed h . It however vanishes at every t for which the mean function μ admits the representation

$$\mu(s_j) = \sum_{l=1}^{p-1} \alpha_l (s_j - t)^l, \quad i = 1, 2, \dots, N, \quad j = 1, 2, \dots, J, \quad (5.3.10)$$

where the coefficients α_l depend on t . The model can then be expressed using matrix representation

$$\mathbf{R}_i = \mathbf{U}_t \boldsymbol{\alpha} + \boldsymbol{\varepsilon}_i, \quad i = 1, 2, \dots, N.$$

Hence, the estimator $\hat{\boldsymbol{\alpha}}(t)$ is unbiased because

$$\mathbb{E}[\hat{\boldsymbol{\alpha}}(t)] = \mathbf{Q}_{t,h}\boldsymbol{\mu} = (\mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t \boldsymbol{\alpha} = \boldsymbol{\alpha}.$$

Representation (5.3.10) will not hold exactly at every $t = t_j$, but is a reasonable approximation underlying the local polynomial method. This leads to the approximation in distribution

$$\sqrt{N}(\hat{\alpha}_1(t) - \mu'(t)) \approx N(0, \hat{\sigma}_\varepsilon^2), \quad (5.3.11)$$

where $\hat{\sigma}_\varepsilon^2$ is the second diagonal element of the matrix $\mathbf{Q}_{t,h}\widehat{\Sigma}_\varepsilon\mathbf{Q}'_{t,h}$ with the matrix $\mathbf{Q}_{t,h}$ defined in (5.3.6) and the matrix Σ_ε in (5.3.9). The matrices \mathbf{U}_t and $\Omega_{t,h}$ in (5.3.6) are defined, respectively, in (5.3.3) and (5.3.4).

5.4 Asymptotic and Bootstrap Procedures

Our objective is to determine the t -region over which the function μ is increasing. Equivalently, we want to determine the values of t for which $\mu'(t) > 0$. To achieve this objective with a specified statistical significance p , we have to find for each t , a random lower bound $\mu_L(t)$ such that $P(\mu'(t) > \mu'_L(t)) = 1 - p$. Note that using $p = 0.05$ defines a one-sided 95 percent confidence interval. If $\mu'_L(t) > 0$, we can be 95 percent confident that the expected cumulative return is increasing at time t . In the following, we often suppress the time t , as it is fixed when constructing the confidence band.

Computation of the exact value of μ'_L is not possible; we would need to know the exact distribution of $\hat{\mu}' - \mu'$. In that infeasible case, if q_U is the $(1 - p)$ th quantile of the distribution of $\hat{\mu}' - \mu'$, we would set $\mu'_L = \hat{\mu}' - q_U$. The distribution of $\hat{\mu}' - \mu'$ is however unknown and must be approximated in some way. We consider asymptotic and bootstrap approximations.

The asymptotic approximation we use is stated as relation (5.3.11), which leads to approximation:

$$q_U \approx N^{-1/2}\hat{\sigma}_\varepsilon\Phi_{1-p}, \quad (5.4.1)$$

where the standard deviation $\hat{\sigma}_\varepsilon$ is defined immediately below relation (5.3.11). The symbol Φ_{1-p} is the $(1 - p)$ th quantile of the standard normal distribution. For the empirical study, we denote the asymptotic approximation by **AS**.

In the bootstrap approximation, we replace the distribution of $\hat{\mu}' - \mu'$ by the empirical distribution of $\hat{\alpha}_1^*$. The bootstrap replications are obtained according to Algorithm 32. For convenience, label the bootstrap procedure as **BS**. Denote the empirical p th (e.g. 5th) percentile of the B values

of $\hat{\alpha}_1^*$ by q_L^* . We then approximate μ'_L by q_L^* . Note the implemented bootstrap procedure follows from Section 6.3.4 of Ruppert (2011). The percentile bootstrap method is known to work well if the distribution of $\hat{\alpha}_1^*$ is symmetric about $\hat{\alpha}_1$, which is reasonable in our setting. We emphasize that the bootstrap samples are selected from the set of all CIDR curves R_1, R_2, \dots, R_N , not separately for each t .

Algorithm 32. *[Bootstrap] For $b = 1, 2, \dots, B$, repeat the following steps:*

1. *Select a random sample with replacement of N curves $R^* = \{R_1^*, R_2^*, \dots, R_N^*\}$ from the original data set.*
2. *Using the bootstrap sample R^* , compute $\hat{\alpha}_1^*$ which minimizes (5.3.1). (We use Equation (5.3.5).)*
3. *Use the lower quantile q_L^* of the empirical distribution of $\hat{\alpha}_1^*$ to construct the lower bound $\mu'_L = q_L^*$.*

In practice, we partition the interval $[0, 1]$ using points $0 = t_0 < t_1 < t_2 < \dots < t_J = 1$, and calculate the approximations to $\mu'_L(t_j)$ defined above.

Both the asymptotic and bootstrap procedures can be transformed to find intervals of statistically decreasing cumulative returns. To accomplish this, we have to find for each t , a random upper bound $\mu_U(t)$ such that $P(\mu'(t) < \mu_U(t)) = 1 - p$. In this paper, we only focus on intervals of increasing cumulative returns.

5.5 Data

In this empirical study, we apply the asymptotic procedure and the bootstrap procedure to selected financial data sets under the presence of scheduled macroeconomic announcements. There are two financial data sets under consideration. The first is intraday cumulative returns of S&P500 index. This is a slight variation from the work of Lucca and Moench (2015), who use 2pm-to-2pm SPX excess cumulative returns. A typical trading day consists of $J = 390$ minutes ranging

from 09:30 to 16:00. The second financial index is NYMEX light crude oil futures for which, we consider the intraday cumulative returns from 10:00 to 15:00, which gives refinement $J = 300$ minutes.¹ In the empirical study, analyzing a larger window about the announcement is often appropriate. We consider either a one day or a three day window. The one day analysis refers to only looking at functional observations that come from the actual announcement day. A three day window corresponds to a single functional observation covering a three day span. For the S&P500 cumulative returns, a one day window has refinement $J = 390$ minutes while a three day window has refinement $J = 1170$ minutes.

The first macroeconomic announcement we consider is the scheduled Federal Open Market Committee meetings. To investigate the pre-FOMC drift, we run a one day analysis using the sampling period from Lucca and Moench (2015) on the S&P500 cumulative returns. This constitutes 132 scheduled FOMC meetings ranging from September-1994 and ending March-2011. A longer sampling period of 145 days is also considered, which span February-1994 through March-2012. To assess the impact of the FOMC meetings on crude oil cumulative futures, we look at all scheduled meeting days from February-1994 through ending March-2011, which results in 166 days. We run a one day window on oil futures in the presence of the FOMC meetings.

The next macroeconomic announcement is the US Change in Nonfarm Payrolls, which constitutes 157 days starting from January-1999 through March-2012. For this case, we assess its influence on the S&P500 cumulative returns. The announcement occurs at 8:30, which is before the start of the trading day. Consequently, we chose a three day window for this setting. Many researches consider the Change in Nonfarm Payrolls announcement to be one of the most impactful announcements, see e.g., Elder et al. (2011). To gain further insight, we consider both ex-post

¹Our sample period for CL is June 2003 to December 2014. During this period, the futures trading hours changed. From June 2003 to 6/11/2006, Crude Oil Futures cease trading between 9:30 and 10:00 for maintenance. After 6/12/2006, the maintenance period is between 16:15 and 17:00. Thus, for the period between 2003 and 2006, there is no trading between 9:30 and 10:00. Therefore, For simplification, we only consider the timer period between 10:00 and 15:00 for CL.

positive and negative shocks on the market. Namely, we consider when the market exceeds expectations and when the market is less than expected. The expectations come from the Bloomberg survey median.

The third announcement considered is the Energy Information Administration's announcement of crude oil inventory. This is known to be a killer or buster of short run crude oil prices. We test its impact on crude oil futures cumulative returns with announcement days ranging from June-2003 through December-2014. This announcement occurs mostly on Wednesday at 10:30 am which results in a total of 601 days. Since the announcement occurs during the business day, we only consider a one day analysis as opposed to a three day window. For the crude oil inventory announcement, we also consider ex-post positive and negative shocks related to market expectations measured by the Bloomberg Survey median. For ease of reference, Tables 5.1 and 5.2 summarize the above data sets, announcements, sampling periods and testing procedures.

The intraday index and futures price data is obtained from TickData. The market expectations of the corresponding macroeconomic news are from Bloomberg.

5.6 Empirical Applications

5.6.1 Pre-FOMC Drift

For both procedures, the confidence bands are computed using least squares optimization method described in Section 5.3.1. For the **BS** procedure, use $B = 1000$ bootstrap replications. The confidence bands are estimated using levels 90%, 95% and 99%. The kernel function $K(\cdot)$ is chosen to be gaussian with seven bandwidths $h = .01, .05, .10, .15, .20, .30, .50$. Using too big of a bandwidth will result in under-fitting while too small of a bandwidth will result in over-fitting. Visual inspection supports the use of bandwidth $h = .10$, which is also chosen in all figures, chosen to demonstrate the relevant procedures.

After running the **AS** and **BS** procedures to construct the confidence bands, the intervals of statistically increasing returns are determined by where the confidence bands are greater than zero.

These time periods can be a single interval or a union of disjoint intervals. In the summary tables, “NA” reflects the case when the confidence band erratically crosses zero due to over-fitting. The NA is displayed when the confidence band produces at least four intervals. In the summary tables, “none” reflects the case when the confidence band is under zero for the entire trading day. In this case, no intervals of statistically increasing returns have been discovered. To determine intervals of statistically decreasing returns, upper confidence bands must be constructed, which is not a focus in this paper.

Table 5.4 shows the one day window **SP** intervals of increasing cumulative returns in the presence of the **FOMC** scheduled meetings. This is the same sampling period used in Lucca and Moench (2015). The results show that with different bandwidths and different confidence levels, both the **AS** and **BS** approaches are able to detect the intervals of increasing cumulative returns within the one day window. Different bandwidths and different confidence levels do impact the results. In general, results of approaches with most of the combinations of bandwidth and confidence levels show that the cumulative returns do increase before the announcement at 14:15. However, the results also answer our first question. That is, the cumulative returns do not change direction exactly at 14:15. Instead, the increasing return intervals end before 14:00 for all combinations. The pre-FOMC drift does exist, but it ends earlier than the announcement. The results also clearly show that increasing bandwidth and confidence levels make the intervals tighter. The **AS** procedure produces similar results as the **BS** procedure. The different bandwidths and confidence levels do not change the fundamental pattern.

Figure 5.1 displays all the cumulative return curves, the mean and the 95% derivative confidence band with $h = 0.1$. The figure shows that the average cumulative returns increase statistically until approximately 47 minutes (at 13:28) before the scheduled **FOMC** announcements.

Table 5.5 shows one day window **SP** intervals of increasing cumulative returns in the presence of the **FOMC** scheduled meetings using the extended sample size. The pattern is similar to Table 5.4.

When using the increased sample size, the intervals of statistically increasing returns is roughly 10 minutes shorter. For example, when using $h = .10$ and 95% confidence, the **AS** procedure produces the interval (09:30,13:28) for the smaller sample size and (09:30,13:17) for the larger sample size.

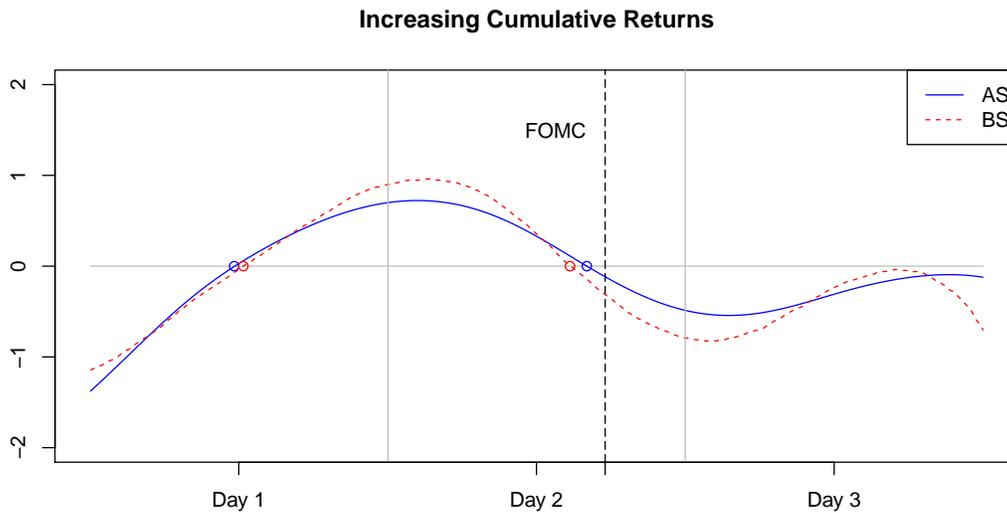


Figure 5.2: Three Day Window **SP** 95% Confidence Bands: FOMC Days

Tables 5.6 and 5.7 show three day window **SP** intervals of increasing cumulative returns in the presence of the **FOMC** scheduled meetings. For this output, the larger sampling period is used. The pattern is in-line with Table 5.5 and adds more information. As illustrated in Figure 5.2, the cumulative returns tend to start statistically increasing on the business day before the announcement and stop increasing before the **FOMC** announcement. When using $h = .10$ and 95% confidence, the **AS** procedure produces the interval (12:39,16:00) for the day preceding the announcement and the interval (09:30,13:51) on the announcement day. This pattern is roughly uniform for the different sample sizes and announcements. The **AS** procedure and **BS** procedures produce similar conclusions. In summary, the results show that the pre-FOMC drift starts at roughly noon of the day before the announcement day and continues to about half hour before 14:15.

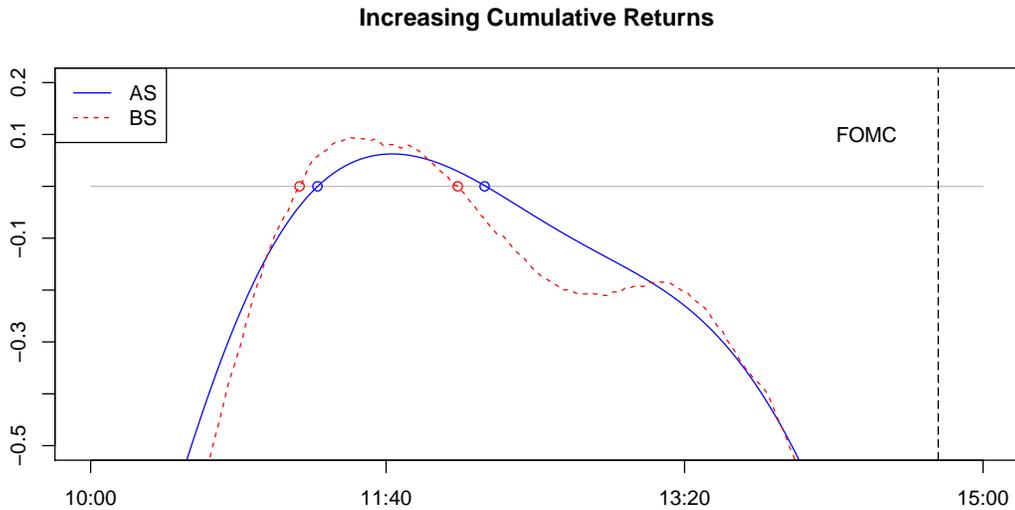


Figure 5.3: One Day Window **CL** 95% Confidence Bands: FOMC Days

5.6.2 Change in Crude Oil Inventory

In much macroeconomic announcement literature, the EIA Change in Crude Oil Inventory announcements have been shown to impact the price movement of crude oil futures, at least at the intraday level. This impact is known to vanish quickly. In this section, we apply our approaches to detect whether there is a pre-announcement drift around the EIA Change in Crude Oil Inventory announcements.

Tables 5.8, 5.9, and 5.10 show one day window **CL** intervals of increasing cumulative returns in the presence of the **CI** scheduled meetings. Table 5.8 reports the results for all announcement days. The results basically show an increasing interval after 12:00 for all combinations. It does not contain much information at all since the announcement is at 10:30. Thus, to further investigate the impact of the change in inventory, we divide the data into two sub-samples: days with ex-post negative shocks – less than expected change of inventory and days with positive shocks – more than expected change of inventory. The expectation is measured as the Bloomberg Survey median. In reality, one does not observe the shocks until the announcements, but as an ex-post examination,

most literature adopt this approach.

Table 5.8 reports the results for the days with ex-post negative shocks. Intuitively, negative shocks should lead to price increase or increasing returns of **CL**. The results are consistent with intuition. Although different combinations of bandwidth and confidence levels give different increasing return intervals, in general, the intervals contains 10:30. For instance, the **AS** approach with 0.10 bandwidth and 95% confidence levels detect the increasing return interval (10:23:11:25). This is consistent with the literature: negative shocks results in up price movements of CL and the impact vanish fairly quickly. The interval seems to indicate that there is some pre-announcement drift, but only several minutes before the announcements on average.

Table 5.10 presents the results for the days with ex-post positive shocks. Not surprisingly, we do not observe any increasing intervals containing 10:30. Intuitively, positive shocks upsets CL prices and thus, we are more likely to detect a decreasing interval. However, the after announcement intervals (mostly start after 12:00) are consistent with the “correcting of over reaction” phenomenon observed in the literature. Most likely, after a positive shock, the market overreacts to the shock for more than an hour and then corrects the over reaction quickly after.

5.6.3 The Change in Nonfarm Payrolls

The Change in Nonfarm Payrolls which is included in the monthly Employment Situation Summary from the Bureau of Labor Statistics is considered the “King of announcement.” In this section, we investigate the impact of the Nonfarm Payrolls on the cumulative return of the S&P500 index. Table 5.11 shows three day window **CL** intervals of increasing cumulative returns in the presence of the **CM** scheduled meetings. The single table shows all relevant intervals for positive shocks, negative shocks and all announcement days. Validated by Figure 5.5, the confidence bands all exist below zero, which indicates the S&P500 cumulative returns are not shown to be statistically increasing under the change in conform payroll announcement. This conclusion is consistent with

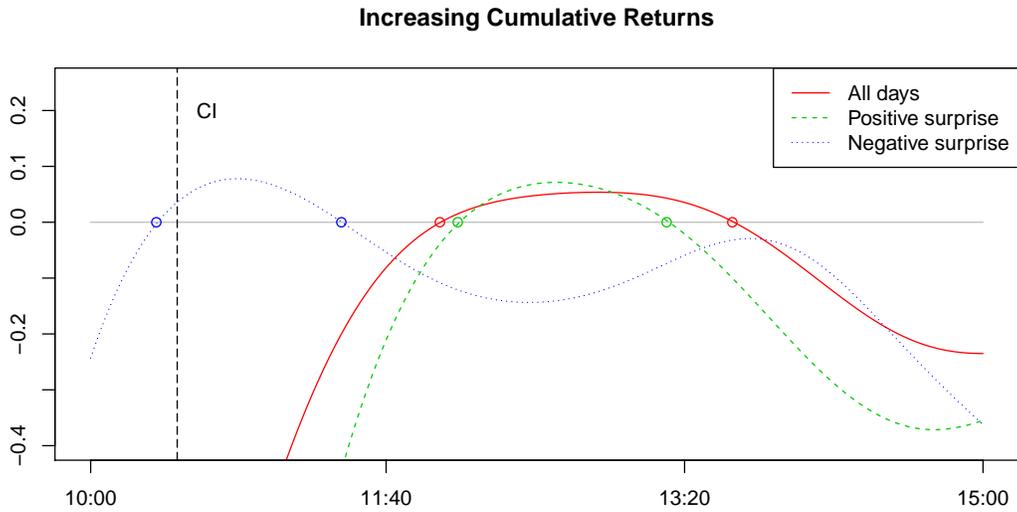


Figure 5.4: One Day Window **CL** Asymptotic 95% Confidence Bands: CI Days

Lucca and Moench (2015). The few intervals that do exist are likely a result of overfitting, since they occur with small bandwidths $h = .01, .05$.

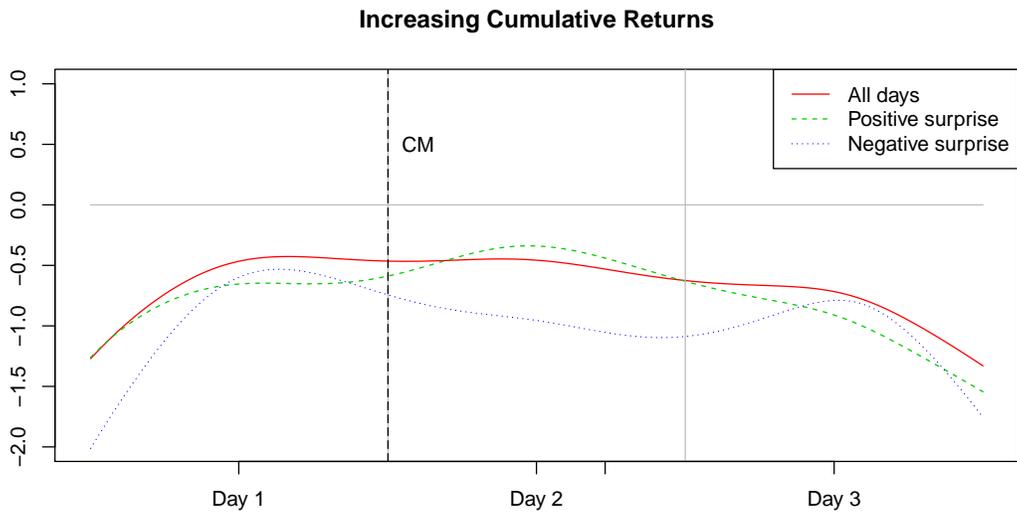


Figure 5.5: Three Day Window **SP** Asymptotic 95% Confidence Bands: CM Days

5.7 Simulation study

To assess the performance of both asymptotic approach described by approximation (5.3.11) and bootstrap approach described by Algorithm 32, we perform a small simulation study using curves that resemble actual intraday cumulative returns. With regards to model (5.2.2), assume two different functional means:

$$\mu(t) = -169/128t(t - 16/13), \quad 0 \leq t \leq 1, \quad (5.7.1)$$

and

$$\mu(t) = -169/64t(t - 16/13), \quad 0 \leq t \leq 1. \quad (5.7.2)$$

Rescaled to 390 minutes to represent a typical trading day, these curves resembles the average pattern of cumulative returns. The two different functions have the same vertex with different amplitudes. The true maximum of both quadratics is located at $t = 240/390$, which corresponds to 240 minutes when $\mu(t)$ is not scaled to the unit interval. The maximum of (5.7.1) is .5 and the maximum of (5.7.2) is 1. The random functional errors $\varepsilon_i(t)$ are assumed to be an AR(1) process simulated at discrete time points $t_j, j = 1, 2, \dots, J$. More specifically,

$$\varepsilon_i(t_j) = 0.99 * \varepsilon_i(t_{j-1}) + u_j, \quad j = 2, \dots, J, \quad i = 1, 2, \dots, N,$$

where the u_j 's are iid mean zero random variables with variance $\sigma^2 = (1/25)^2$. Collectively, the cumulative returns are simulated using the two models

$$R_i(t_j) = -169/128t_j(t_j - 16/13) + \varepsilon_i(t_j), \quad j = 2, \dots, J, \quad i = 1, 2, \dots, N, \quad (5.7.3)$$

and

$$R_i(t_j) = -169/64t_j(t_j - 16/13) + \varepsilon_i(t_j), \quad j = 2, \dots, J, \quad i = 1, 2, \dots, N. \quad (5.7.4)$$

Consequently, the true interval of increasing returns is expressed as:

$$\text{Interval of increasing futures} = [0, 240) \text{ minutes.}$$

For convenience, we call $t = 240$ minutes the *true cut-off* value. In our study, we simulate N curves and estimate the $(1-p)\%$ confidence band using bandwidth 0.10. Denote the location where the confidence band crosses zero by t_c . Using bootstrap Algorithm 32, define t_c as the value that satisfies

$$\mu_L(t_c) = 0.$$

Using approximation (5.3.11), define t_c as the value that satisfies

$$\hat{\alpha}_1(t_c) - N^{-1/2} \hat{\sigma}_\epsilon \Phi_{1-p} = 0.$$

With $(1-p)\%$ confidence, this quantity defines our large sample interval of increasing returns, i.e.,

$$(1-p)\% \text{ interval of increasing cumulative returns} = (0, t_c) \text{ minutes.}$$

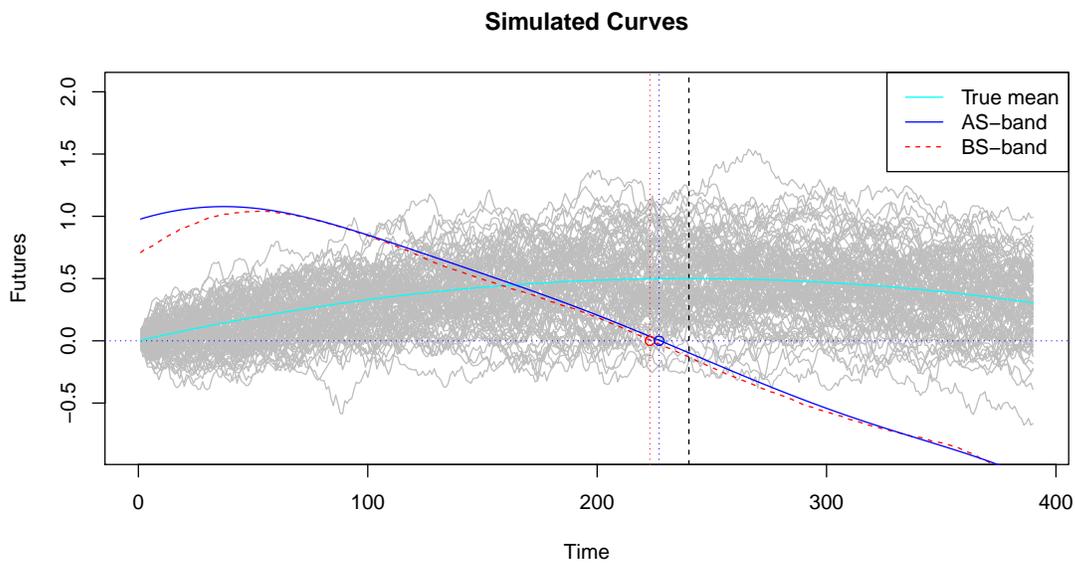


Figure 5.6: $N = 100$ simulated curves using refinement $J = 390$ minutes. The **BS** and **AS** procedures are used to estimate when the true maximum occurs with 95% confidence. The true maximum is located at 240 minutes and the true amplitude of the mean $\mu(t)$ is $1/2$. For this realization, the **BS** confidence band crosses at 223 minutes and the **AS** confidence band crosses at 227 minutes.

To illustrate the procedure, Figure 5.6 shows $N = 100$ generated curves with true mean assumed as Equation (5.7.1). The 95% lower confidence bands are computed using both procedures. The simulation algorithm is described below:

Algorithm 33. *[Simulation]*

1. Simulate N curves based on equations (5.7.3) and (5.7.4). Here we use sample sizes $N = 100, 250$ and refinement $J = 390$.
2. Estimate 95% lower confidence band using **AS** approximation (5.3.11) and **BS** Algorithm 32. Using the confidence bands, find cut-off values t_c for both procedures.
3. With $R = 200$, compute the MSE using both procedures where

$$MSE = \left\{ \frac{1}{R} \sum_{r=1}^R \left(t_c^{(r)} - t \right)^2 \right\}^{1/2}.$$

Table 5.3 displays the simulation results of Algorithm 33. For all cases, the **AS** procedure did a better job at identifying the true cut-off compared to the **BS** procedure. Simulated curves that exhibit a larger amplitude in $\mu(t)$ also produce a smaller MSE . This is intuitive because the trend is easier to identify which in turn, makes it easier to estimate the true cut-off $t = 240$. The increased sample size also produces a smaller MSE for all cases.

5.8 Conclusion

There has been extensive literature that studied the impact of macroeconomic news on returns of financial assets. In a recent paper, Lucca and Moench (2015) documents large average excess returns on U.S. equities in anticipation of monetary policy decisions made at scheduled meetings of the Federal Open Market Committee (FOMC). However, their conclusion is based on dividing the data artificially at the announcement point of time and run simple regression to show the existence of the pre-FOMC drift.

In this paper, we develop a data-driven approach for detecting intervals of statistically increasing cumulative returns in the presence of macroeconomic announcements. We examine the intraday cumulative return curve in the context of Functional Data Analysis (FDA). This is done by finding an interval with a positive derivative. To construct the confidence band, we develop two methods: asymptotic and bootstrap. Our approach can precisely determine the turning points of the

derivatives and thus detect the exact interval of the “pre-FOMC.” The approaches can be applied to study the impacts of any “events” on the returns of financial assets.

We then apply our approaches to re-examine the pre-FOMC drift phenomenon. Our results reveal more information about how the return curves change previous to the FOMC meeting. First, the pre-FOMC drift does exist. Second, the drift ends earlier than the announcement time of 14:15, not exactly at the announcement time. Therefore, artificially cutting the data at the time of the announcement is questionable or at least not very accurate.

We also apply our approaches to examine the impacts of Change in Crude Oil Inventory on the cumulative returns of WTI Light Sweet Crude oil futures. When the announcement days are divided into sub-samples of positive and negative ex-post inventory shocks, the approaches are able to detect the increasing return intervals for both sub-samples. The intervals detected are consistent with the literature and contain richer information. The result of applying the approaches to the impact of Change in Nonfarm payroll on the S&P500 index is consistent with Lucca and Moench (2015).

Table 5.1: Tickers and Abbreviation

This table shows the Tickers (symbols) and abbreviations used in the paper.

Assets, Events, etc.	Symbol (Abbreviation)
S & P 500 E-mini futures	SP
Light crude oil futures NYMEX	CL
Federal open market committee meetings (14:15 or 2:15 pm)	FOMC
Change in nonfarm payroll (8:30 am)	CM
Inventory change in crude oil (10:30 am)	CI
Asymptotic Procedure	AS
Bootstrap Procedure	BS
Positive Shock (Actual > Expected)	POS
Negative Shock (Actual < Expected)	NEG

Table 5.2: Sampling Periods and Sizes

This table reports the sample periods in this study.

Financial index	Announcement	Sampling period	Sample size
SP	FOMC	09/94 - 03/11	132
	FOMC	09/94 - 03/12	147
	CM-FULL	01/99 - 03/12	157
	CM-POS	01/99 - 03/12	86
	CM-NEG	01/99 - 03/12	71
CL	FOMC	02/94 - 12/14	166
	CI-FULL	06/03 - 12/14	601
	CI-POS	06/03 - 12/14	378
	CI-NEG	06/03 - 12/14	223

Table 5.3: MSE for Simulations

This table reports the simulated MSE values of the bootstrap procedure (**BS**) and large sample procedure (**AS**). Each method uses confidence level 95% and bandwidth $h = .10$. The MSE values are calculated using the full 390 minute trading day. The units of the empirical MSE values are in minutes.

Procedure	Cut-off	Amplitude of $\mu(t)$	$N = 100$	$N = 250$
BS	t_c	Small	26.79	16.78
	t_c	Big	12.55	8.21
AS	t_c	Small	20.03	12.42
	t_c	Big	9.45	6.23

Table 5.4: FOMC and **S&P** One Day Increasing Return Intervals: 1994 to 2011

This table presents the one day **SP** intervals of increasing returns in the presence of the **FOMC** scheduled meetings for the sample period September 1994 to March 2011. This is the same sampling period as pre-FOMC drift paper.

h	Confidence Level (%)		
	90	95	99
Panel A: Asymptotic Procedure			
0.01	NA	NA	NA
0.05	(09:37,13:40)	(09:41,13:37)	(09:48,11:49) (12:30,13:32)
0.10	(09:30,13:37)	(09:30,13:28)	(09:30,13:13)
0.15	(09:30,13:36)	(09:30,13:24)	(09:30,13:07)
0.20	(09:30,13:35)	(09:30,13:23)	(09:30,13:05)
0.30	(09:30,13:35)	(09:30,13:23)	(09:30,13:04)
0.50	(09:30,13:35)	(09:30,13:22)	(09:30,13:04)
Panel B: Bootstrap Procedure			
0.01	NA	NA	NA
0.05	(09:45,12:01) (12:25,13:44)	(09:48,11:56) (12:29,13:42)	(09:52,11:47) (12:34,13:38)
0.10	(09:30,13:38)	(09:30,13:32)	(09:35,13:27)
0.15	(09:30,13:38)	(09:30,13:28)	(09:30,13:12)
0.20	(09:30,13:38)	(09:30,13:24)	(09:30,13:06)
0.30	(09:30,13:34)	(09:30,13:24)	(09:30,13:06)
0.50	(09:30,13:35)	(09:30,13:21)	(09:30,13:05)

Table 5.5: FOMC and **S&P** One Day Increasing Return Intervals: 1994 to 2012

This table presents the intervals of increasing returns in the presence of the **FOMC** scheduled meetings for the sample period February 1994 to March 2012. This is a different sampling period as the pre-FOMC drift paper.

h	Confidence Level (%)		
	90	95	99
Panel A: Asymptotic Procedure			
0.01	NA	NA	NA
0.05	(09:35,13:36)	(09:39,12:06) (12:18,13:33)	(09:46,11:46) (12:35,13:27)
0.10	(09:30,13:27)	(09:30,13:17)	(09:30,13:02)
0.15	(09:30,13:27)	(09:30,13:15)	(09:30,12:58)
0.20	(09:30,13:28)	(09:30,13:16)	(09:30,12:58)
0.30	(09:30,13:28)	(09:30,13:16)	(09:30,12:58)
0.50	(09:30,13:29)	(09:30,13:16)	(09:30,12:58)
Panel B: Bootstrap Procedure			
0.01	NA	NA	NA
0.05	NA	(09:46,11:53) (12:30,13:38)	(09:51,11:45) (12:38,13:31)
0.10	(09:30,13:30)	(09:30,13:26)	(09:34,13:14)
0.15	(09:30,13:27)	(09:30,13:18)	(09:30,13:00)
0.20	(09:30,13:26)	(09:30,13:20)	(09:30,13:06)
0.30	(09:30,13:26)	(09:30,13:16)	(09:30,12:56)
0.50	(09:30,13:26)	(09:30,13:16)	(09:30,12:56)

Table 5.6: FOMC and **S&P** Three Days Increasing Return Intervals: Asymptotic Procedure

This table presents the three day **SP** asymptotic intervals of increasing returns in the presence of the **FOMC** scheduled meetings for the sample period February 1994 to March 2012 (**different as pre-FOMC drift paper**).

h	CL(%)	Day 1	Day 2 (FOMC)	Day 3
0.01	90	NA	(09:30,11:56) (12:30,13:42) (14:44,15:05)	(12:34,13:09) (13:36,14:37)
0.01	95	NA	(09:30,11:53) (12:33,13:39) (14:48,15:01)	(12:39,13:05) (13:44,13:56)
0.01	99	(12:26,12:41) (15:46,16:00)	(09:30,11:13) (11:25,11:47) (12:38,13:33)	none
0.05	90	(12:48,16:00)	(09:30,13:34)	(13:06,14:23)
0.05	95	(13:06,16:00)	(09:30,13:21)	none
0.05	99	(13:50,16:00)	(09:30,13:03)	none
0.10	90	(12:24,16:00)	(09:30,14:11)	(14:24,16:00)
0.10	95	(12:39,16:00)	(09:30,13:51)	none
0.10	99	(13:13,16:00)	(09:30,13:18)	none
0.15	90	(11:11,16:00)	(09:30,14:47)	none
0.15	95	(11:33,16:00)	(09:30,14:18)	none
0.15	99	(12:20,16:00)	(09:30,13:28)	none
0.20	90	(09:30,16:00)	(09:30,15:06)	none
0.20	95	(09:30,16:00)	(09:30,14:32)	none
0.20	99	(10:39,16:00)	(09:30,13:33)	none
0.30	90	(09:30,16:00)	(09:30,15:23)	none
0.30	95	(09:30,16:00)	(09:30,14:45)	none
0.30	99	(09:30,16:00)	(09:30,13:37)	none
0.50	90	(09:30,16:00)	(09:30,15:35)	none
0.50	95	(09:30,16:00)	(09:30,14:54)	none
0.50	99	(09:30,16:00)	(09:30,13:40)	none

Table 5.7: FOMC and **S&P** Three Days Increasing Return Intervals: Bootstrap Procedure

This table presents the three day **SP** Bootstrap intervals of increasing returns in the presence of the **FOMC** scheduled meetings for the sample period February 1994 to March 2012.

h	CL(%)	Day 1	Day 2 (FOMC)	Day 3
0.01	90	NA	NA	NA
0.01	95	NA	NA	(12:40,13:04) (13:39,13:55)
0.01	99	(12:24,12:42) (15:51,15:55)	NA	(12:47,13:01)
0.05	90	(09:30,09:44) (12:54,16:00)	(09:30,13:29)	(12:51,14:27)
0.05	95	(13:19,16:00)	(09:30,13:22)	(13:06,14:06)
0.05	99	(14:16,16:00)	(09:30,13:02)	none
0.10	90	(12:40,16:00)	(09:30,13:46)	(13:26,15:03)
0.10	95	(12:51,16:00)	(09:30,13:29)	none
0.10	99	(13:23,16:00)	(09:30,13:00)	none
0.15	90	(12:19,16:00)	(09:30,14:15)	(14:32,16:00)
0.15	95	(12:37,16:00)	(09:30,13:52)	none
0.15	99	(13:09,16:00)	(09:30,13:24)	none
0.20	90	(11:29,16:00)	(09:30,14:47)	none
0.20	95	(11:47,16:00)	(09:30,14:13)	none
0.20	99	(12:23,16:00)	(09:30,13:25)	none
0.30	90	(09:30,16:00)	(09:30,15:04)	none
0.30	95	(09:30,16:00)	(09:30,14:34)	none
0.30	99	(10:03,16:00)	(09:30,13:49)	none
0.50	90	(09:30,16:00)	(09:30,15:32)	none
0.50	95	(09:30,16:00)	(09:30,14:38)	none
0.50	99	(09:30,16:00)	(09:30,13:21)	none

Table 5.8: EIA and **CL** One Day Increasing Return Intervals

This table presents the one day **CL** intervals of increasing returns in the presence of Crude Oil Inventory Change releases for the sample period June 2003 to December 2014.

h	Confidence Level (%)		
	90	95	99
Panel A: Asymptotic Procedure			
0.01	NA	NA	NA
0.05	NA	NA	(11:44,11:57) (13:05,13:25)
0.10	(11:48,13:50)	(11:58,13:36)	none
0.15	(12:01,13:57)	(12:12,13:37)	none
0.20	(12:09,14:12)	(12:20,13:46)	none
0.30	(12:15,15:00)	(12:29,14:18)	none
0.50	(12:19,15:00)	(12:33,15:00)	none
Panel B: Bootstrap Procedure			
0.01	NA	NA	NA
0.05	NA	NA	NA
0.10	(11:34,13:46) (14:51,15:00)	(11:40,12:22) (12:43,13:40)	(13:06,13:08)
0.15	(11:49,13:50)	(12:03,13:32)	none
0.20	(12:01,13:52)	(12:08,13:34)	(12:28,13:06)
0.30	(12:09,14:19)	(12:19,13:53)	none
0.50	(12:15,15:00)	(12:29,14:31)	(13:08,13:22)

Table 5.9: EIA and **CL** One Day Increasing Return Intervals: Negative Inventory Shock

This table presents the one day **CL** intervals of increasing returns in the presence of negative shock of Crude Oil Inventory Change scheduled meetings for the sample period June 2003 to December 2014.

h	Confidence Level (%)		
	90	95	99
Panel A: Asymptotic Procedure			
0.01	NA	NA	(10:25,10:31) (13:29,13:38)
0.05	(10:23,11:11) (13:17,13:55)	(10:30,11:00) (13:28,13:46)	none
0.10	(10:04,11:42) (13:19,14:05)	(10:23,11:25)	none
0.15	(10:00,12:14) (13:12,14:24)	(10:00,11:31)	none
0.20	(10:00,13:27)	(10:00,11:29)	none
0.30	(10:00,13:14)	none	none
0.50	(10:00,13:12)	(11:45,12:30)	none
Panel B: Bootstrap Procedure			
0.01	NA	NA	(10:26,10:31) (13:30,13:37)
0.05	NA	(10:27,10:53) (13:31,13:44)	none
0.10	(10:15,11:31) (13:17,14:00)	(10:36,11:10) (13:44,13:46)	none
0.15	(10:02,11:44) (13:18,14:09)	(10:29,11:24)	none
0.20	(10:00,12:15) (13:10,14:22)	(10:02,11:39)	none
0.30	(10:00,13:29)	(10:04,10:09) (10:11,10:14) (11:00,15:00)	none
0.50	(10:00,13:17)	none	none

Table 5.10: EIA and **CL** One Day Increasing Return Intervals: Positive Inventory Shock

This table presents the one day **CL** intervals of increasing returns in the presence of positive shock of Crude Oil Inventory Change releases for the sample period June 2003 to December 2014.

h	Confidence Level (%)		
	90	95	99
Panel A: Asymptotic Procedure			
0.01	NA	(11:36,12:02) (12:50,13:04)	(11:44,11:55) (11:59,12:54) (12:57,15:00)
0.05	NA	(11:37,12:14) (12:48,13:27)	(11:48,12:00) (13:03,13:11)
0.10	(11:56,13:27)	(12:04,13:14)	none
0.15	(12:16,13:27)	(12:28,13:09)	none
0.20	(12:31,13:36)	none	none
0.30	(12:43,14:24)	(13:11,13:33)	none
0.50	(12:48,15:00)	(13:06,15:00)	none
Panel B: Bootstrap Procedure			
0.01	NA	NA	(11:45,11:51) (11:58,12:00) (12:54,15:00)
0.05	NA	(11:34,12:09) (12:50,13:28)	(11:40,12:03) (13:01,13:13)
0.10	(11:40,13:30) (14:58,15:00)	(11:46,13:22)	none
0.15	(12:00,13:24)	(12:08,13:15)	none
0.20	(12:13,13:25)	(12:20,13:16)	none
0.30	(12:35,13:44)	none	none
0.50	(12:45,14:59)	(13:06,14:13)	none

Table 5.11: CM and **S&P** Three Days Increasing Return Intervals: Bootstrap Procedure

This table presents the three day **SP** Bootstrap intervals of increasing returns in the presence of the **FOMC** scheduled meetings for the sample period February 1994 to March 2012.

Shock	h	CL(%)	Day 1	Day 2 (CM)	Day 3
Panel A: Asymptotic Procedure					
Pos	0.01	90	(11:47,11:56)	(15:20,15:54)	none
	0.01	95	none	(15:23,15:46)	none
Neg	0.01	90	none	none	(15:47,16:00)
	0.01	95	none	(15:23,15:44)	none
All	0.01	90	(15:39,16:00)	(09:30,09:32) (15:19,15:52)	none
	0.01	95	none	(15:23,15:45)	none
Panel B: Bootstrap Procedure					
Pos	0.01	90	(11:46,12:02)	(11:33,11:40) (15:22,15:53)	none
	0.01	95	(11:49,11:58)	(15:24,15:50)	none
	0.01	99	none	(15:30,15:43)	none
Neg	0.01	90	none	none	(15:42,15:52)
	0.05	90	(14:20,14:22) (14:25,16:00)	none	none
All	0.01	90	(10:36,10:43) (15:42,16:00)	(09:30,09:38) (15:21,15:51)	none
	0.01	95	none	(15:25,15:47)	none
	0.01	99	none	(15:32,15:39)	none

5.9 Proofs

5.9.1 Review of local polynomial smoothing

Local polynomial smoothing and derivative estimation have been developed in the context of the model

$$x_i = f(s_i) + \varepsilon_i, \quad i = 1, 2, \dots, N,$$

in which f is an unknown smooth function observed at time points s_i with a measurement error or noise ε_i . An introduction to this technique is presented in Chapter 21 of Ruppert (2011), a detailed and comprehensive treatment is given in Fan and Gijbels (1996). We merely present a few formulas we needed in the following.

To estimate the function f and its derivatives, for every t , we minimize

$$\sum_{i=1}^N \{x_i - (\alpha_0 + \alpha_1(s_i - t) + \dots + \alpha_{p-1}(s_i - t)^{p-1})\}^2 K\left(\frac{s_i - t}{h}\right), \quad (5.9.1)$$

with respect to $\alpha_0, \alpha_1, \dots, \alpha_{p-1}$. The smoothed curve at time t is then given by the estimated intercept $\hat{f}(t) = \hat{\alpha}_0$, and the estimated derivative by $\hat{f}'(t) = \hat{\alpha}_1$. In conjunction with the kernel $K(\cdot)$, the bandwidth h governs the level of smoothness. A large bandwidth results in oversmoothing, while a small bandwidth results in over-fitting the curve.

Closed form formula for the estimator of the parameter vector $\boldsymbol{\alpha} = [\alpha_0, \alpha_1, \dots, \alpha_{p-1}]'$ follows easily using weighted least squares. For fixed t , define the response vector and design matrix respectively by

$$\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, \quad \mathbf{U}_t = \begin{bmatrix} 1 & (s_1 - t) & (s_1 - t)^2 & \dots & (s_1 - t)^{p-1} \\ 1 & (s_2 - t) & (s_2 - t)^2 & \dots & (s_2 - t)^{p-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & (s_N - t) & (s_N - t)^2 & \dots & (s_N - t)^{p-1} \end{bmatrix}.$$

Define the diagonal weight matrix by

$$\boldsymbol{\Omega}_{t,h} = \text{diag}\left\{K\left(\frac{s_1 - t}{h}\right), K\left(\frac{s_2 - t}{h}\right), \dots, K\left(\frac{s_N - t}{h}\right)\right\}. \quad (5.9.2)$$

The estimated parameter vector is given by the weighted least squares solution

$$\hat{\boldsymbol{\alpha}}(t) = \hat{\boldsymbol{\alpha}} = \begin{bmatrix} \hat{\alpha}_0 \\ \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_{p-1} \end{bmatrix} = (\mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_{t,h} \mathbf{x}. \quad (5.9.3)$$

5.9.2 Proof of Proposition 30

Proof. Under the partition $P = \{0 = s_0 < s_1 < \cdots < s_J = 1\}$, the least squares criterion $S_N(t; \boldsymbol{\alpha})$ defined in (5.3.1) can be expressed

$$\begin{aligned} S_N(t; \boldsymbol{\alpha}) &= \sum_{i=1}^N \int_0^1 \{R_i(s) - (\alpha_0 + \alpha_1(s-t) + \cdots + \alpha_{p-1}(s-t)^{p-1})\}^2 K\left(\frac{s-t}{h}\right) ds \\ &= \sum_{i=1}^N \frac{1}{J} \sum_{j=1}^J \{R_i(s_j) - (\alpha_0 + \alpha_1(s_j-t) + \cdots + \alpha_{p-1}(s_j-t)^{p-1})\}^2 K\left(\frac{s_j-t}{h}\right) \\ &= \frac{1}{J} \sum_{i=1}^N SSE_i(t; \boldsymbol{\alpha}) \\ &= \frac{1}{J} \sum_{i=1}^N (\mathbf{R}_i - \mathbf{U}_t \boldsymbol{\alpha})' \boldsymbol{\Omega}_t (\mathbf{R}_i - \mathbf{U}_t \boldsymbol{\alpha}) \\ &= \frac{1}{J} \sum_{i=1}^N \{\mathbf{R}'_i \boldsymbol{\Omega}_t \mathbf{R}_i - \mathbf{R}'_i \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha} - \boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i + \boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha}\} \end{aligned}$$

The quantities $\mathbf{R}'_i \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha}$ and $\boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i$ are scalars which implies they have the same transpose.

This gives

$$S_N(t; \boldsymbol{\alpha}) = \frac{1}{J} \sum_{i=1}^N \{\mathbf{R}'_i \boldsymbol{\Omega}_t \mathbf{R}_i - 2\boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i + \boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha}\}.$$

Using matrix calculus,

$$\begin{aligned} \frac{\partial}{\partial \boldsymbol{\alpha}} S_N(t; \boldsymbol{\alpha}) &= \frac{1}{J} \sum_{i=1}^N \frac{\partial}{\partial \boldsymbol{\alpha}} \{\mathbf{R}'_i \boldsymbol{\Omega}_t \mathbf{R}_i - 2\boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i + \boldsymbol{\alpha}' \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha}\} \\ &= -\frac{2}{J} \sum_{i=1}^N \{\mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i - \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha}\} \\ &= -\frac{2}{J} \left\{ \sum_{i=1}^N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i - N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha} \right\}. \end{aligned}$$

Setting $\frac{\partial}{\partial \boldsymbol{\alpha}} S_N(t; \boldsymbol{\alpha})$ equal to zero gives relation

$$\sum_{i=1}^N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i - N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha} = 0.$$

By the definition of \mathbf{U}_t and $\boldsymbol{\Omega}_t$, the matrix $\mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t$ is nonnegative definite and hence invertible.

Thus the least squares solution that minimizes $S_N(t; \boldsymbol{\alpha})$ is

$$\hat{\boldsymbol{\alpha}} = \frac{1}{N} \sum_{i=1}^N \hat{\boldsymbol{\alpha}}_i = \frac{1}{N} \sum_{i=1}^N (\mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t)^{-1} \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i.$$

To show that $\hat{\boldsymbol{\mu}}$ is indeed a minimum, notice that

$$\begin{aligned} \frac{\partial^2 S_N(t; \boldsymbol{\alpha})}{\partial \boldsymbol{\alpha}' \partial \boldsymbol{\alpha}} &= -\frac{2}{J} \frac{\partial}{\partial \boldsymbol{\alpha}'} \left\{ \sum_{i=1}^N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{R}_i - N \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t \boldsymbol{\alpha} \right\} \\ &= \frac{2N}{J} \mathbf{U}'_t \boldsymbol{\Omega}_t \mathbf{U}_t. \end{aligned}$$

The above matrix is nonnegative definite which guarantees $\hat{\boldsymbol{\alpha}}$ is a minimum. This completes the proof of Proposition 30.

□

Bibliography

- Andersen, T. G., Fusari, N., and Todorov, V. (2012). Parametric inference and dynamic state recovery from option panels. Technical report, NBER.
- Antoniadis, A., Paparoditis, E., and Sapatinas, T. (2006). A functional wavelet–kernel approach for time series prediction. *Journal of the Royal Statistical Society, Series B*, 68:837–857.
- Aue, A., Dubard-Norinho, D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392.
- Aue, A., Hörmann, S., Horváth, L., and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *The Annals of Statistics*, 37:4046–4087.
- Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariance: High frequency based covariance, regression and correlation in financial economics. *Econometrica*, 72:885–925.
- Bech, M. and Lengwiler, Y. (2012). The financial crisis and the changing dynamics of the yield curve. Technical report, Bank of International Settlements.
- Berkes, I., Gabrys, R., Horváth, L., and Kokoszka, P. (2009). Detecting changes in the mean of functional observations. *Journal of the Royal Statistical Society (B)*, 71:927–946.
- Berkes, I., Hörmann, S., and Schauer, J. (2011). Split invariance principles for stationary processes. *The Annals of Probability*, 39:2441–2473.
- Berkes, I., Horváth, L., and Rice, G. (2013a). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, 123:385–403.
- Berkes, I., Horváth, L., and Rice, G. (2013b). Weak invariance principles for sums of dependent random functions. *Stochastic Processes and their Applications*, 123:385–403.

- Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley, New York.
- Bosq, D. (2000). *Linear Processes in Function Spaces*. Springer, New York.
- Brockwell, P. J. and Davis, R. A. (1991). *Time Series: Theory and Methods*. Springer, New York.
- Campbell, J. Y., Lo, A. W., and MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, New Jersey.
- Chen, Y. and Niu, L. (2014). Adaptive dynamic Nelson–Siegel term structure model with applications. *Journal of Econometrics*, 180:98–115.
- Chib, S. and Kang, K. H. (2013). Change points in affine arbitrage-free term structure models. *Journal of Financial Econometrics*, 14:302–334.
- Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. Wiley.
- Dalla, V., Giraitis, L., and Phillips, P. C. B. (2015). Testing mean stability of heteroskedastic time series. Technical report, Yale University.
- de Jong, R. M., Amsler, C., and Schmidt, P. (1997). A robust version of the KPSS test based on indicators. *Journal of Econometrics*, 137:311–333.
- Dickey, D. A. and Fuller, W. A. (1979). Distributions of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.
- Dickey, D. A. and Fuller, W. A. (1981). Likelihood ratio statistics for autoregressive time series with unit root. *Econometrica*, 49:1057–1074.
- Diebold, F. and Rudebusch, G. (2013). *Yield Curve Modeling and Forecasting: The Dynamic Nelson–Siegel Approach*. Princeton University Press.
- Diebold, F. X. and Li, C. (2003). Forecasting the term structure of government bond yields. Working Paper 10048, National Bureau of Economic Research.

- Elder, J., Miao, H., and Ramchander, S. (2011). Impact of macroeconomics news on metal futures. *Journal of Banking and Finance*, 36:51–65.
- Engle, R. F. and Russel, J. R. (2004). Analysis of high frequency financial data. In Ait-Shahlia, Y., editor, *Handbook of Financial Econometrics*. North Holland.
- Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and its Applications*. Chapman & Hall/CRC.
- Filipović, D. (2009). *Term-Structure Models*. Springer.
- Giraitis, L., Kokoszka, P. S., Leipus, R., and Teyssière, G. (2003). Rescaled variance and related tests for long memory in volatility and levels. *Journal of Econometrics*, 112:265–294.
- Hayashi, T. and Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11:359–379.
- Hays, S., Shen, H., and Huang, J. Z. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6:870–894.
- Hörmann, S., Horváth, L., and Reeder, R. (2013). A functional version of the ARCH model. *Econometric Theory*, 29:267–288.
- Hörmann, S., Kidzinski, L., and Hallin, M. (2015). Dynamic functional principal components. *Journal of the Royal Statistical Society (B)*, 77:319–348.
- Hörmann, S. and Kokoszka, P. (2010). Weakly dependent functional data. *The Annals of Statistics*, 38:1845–1884.
- Hörmann, S. and Kokoszka, P. (2012). Functional time series. In Rao, C. R. and Rao, T. S., editors, *Time Series*, volume 30 of *Handbook of Statistics*. Elsevier.

- Horváth, L. (1993). The maximum likelihood method for testing changes in the parameters of normal observations. *The Annals of Statistics*, 21:671–680.
- Horváth, L., Hušková, M., and Kokoszka, P. (2010). Testing the stability of the functional autoregressive process. *Journal of Multivariate Analysis*, 101:352–367.
- Horváth, L. and Kokoszka, P. (2012). *Inference for Functional Data with Applications*. Springer.
- Horváth, L., Kokoszka, P., and Reeder, R. (2013). Estimation of the mean of functional time series and a two sample problem. *Journal of the Royal Statistical Society (B)*, 75:103–122.
- Horváth, L., Kokoszka, P., and Rice, G. (2014). Testing stationarity of functional time series. *Journal of Econometrics*, 179:66–82.
- Horváth, L. and Rice, G. (2014). Extensions and some classical methods in change point analysis (with discussions). *TEST*, 23:219–290.
- Horváth, L., Rice, G., and Whipple, S. (2014b). Adaptive bandwidth selection in the long run covariance estimator of functional time series. *Computational Statistics and Data Analysis*, 00:000–000. Forthcoming.
- Hsing, T. and Eubank, R. (2015). *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley.
- Jentsch, C. and Subba Rao, S. (2015). A test for second order stationarity of a multivariate time series. *Journal of Econometrics*, 185:124–161.
- Kargin, V. and Onatski, A. (2008). Curve forecasting by functional autoregression. *Journal of Multivariate Analysis*, 99:2508–2526.
- Kokoszka, P. and Leipus, R. (2000). Change-point estimation in ARCH models. *Bernoulli*, 6:513–539.

- Kokoszka, P., Miao, H., and Zhang, X. (2014). Functional dynamic factor model for intraday price curves. *Journal of Financial Econometrics*, 00:000–000. Forthcoming.
- Kokoszka, P. and Reimherr, M. (2013a). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, 34:116–129.
- Kokoszka, P. and Reimherr, M. (2013b). Predictability of shapes of intraday price curves. *Econometrics Journal*, 16:285–308.
- Kokoszka, P. and Young, G. (2015a). Kpss test for functional time series. Technical report, Colorado State University.
- Kokoszka, P. and Young, G. (2015b). Testing trend stationarity of functional time series with application to yield and daily price curves. Technical report, Colorado State University.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root? *Journal of Econometrics*, 54:159–178.
- Lee, D. and Schmidt, P. (1996). On the power of the KPSS test of stationarity against fractionally integrated alternatives. *Journal of Econometrics*, 73:285–302.
- Lengwiler, Y. and Lenz, C. (2010). Intelligible factors for the yield curve. *Journal of Econometrics*, 157:481–491.
- Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on–line auction dynamics. *Journal of the American Statistical Association*, 104:704–717.
- Lo, A. W. (1991). Long-term memory in stock market prices. *Econometrica*, 59:1279–1313.

- Lucca, D. O. and Moench, E. (2015). The pre-FOMC announcement drift. *The Journal of Finance*, 70:329–371.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis*. Springer.
- MacNeill, I. B. (1978). Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *The Annals of Statistics*, 6:422–433.
- Müller, H.-G., Sen, R., and Stadtmüller, U. (2011). Functional data analysis for volatility. *Journal of Econometrics*, 165:233–245.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica*, 55:703–08.
- Nieh, C., Wu, S., and Zeng, Y. (2010). Regime shifts and the term structure of interest rates. In Lee, A. C. and Lee, C., editors, *Handbook of Quantitative Finance and Risk Management*, pages 1121–1133. Springer.
- Panaretos, V. M. and Tavakoli, S. (2013). Fourier analysis of stationary time series in function space. *The Annals of Statistics*, 41:568–603.
- Politis, D. N. (2011). Higher-order accurate, positive semidefinite estimation of large sample covariance and spectral density matrices. *Econometric Theory*, 27:1469–4360.
- Politis, D. N. and Romano, J. P. (1996). On flat-top spectral density estimators for homogeneous random fields. *Journal of Statistical Planning and Inference*, 51:41–53.
- Politis, D. N. and Romano, J. P. (1999). Multivariate density estimation with general fat-top kernels of infinite order. *Journal of Multivariate Analysis*, 68:1–25.

- Pötscher, B. and Prucha, I. (1997). *Dynamic Non-linear Econometric Models. Asymptotic Theory*. Springer.
- Ramsay, J. O. and Silverman, B. W. (2005). *Functional Data Analysis*. Springer.
- Ruppert, D. (2011). *Statistics and Data Analysis for Financial Engineering*. Springer.
- Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive–moving average models of unknown order. *Biometrika*, 71:599–608.
- Shao, X. and Wu, W. B. (2007). Asymptotic spectral theory for nonlinear time series. *The Annals of Statistics*, 35:1773–1801.
- Tsay, R. S. (2005). *Analysis of Financial Time Series*. Wiley.
- Wang, Y. and Zou, J. (2010). Vast volatility matrix estimation for high–frequency financial data. *The Annals of Statistics*, 38:953–978.
- Wu, W. (2005). *Nonlinear System Theory: Another Look at Dependence*, volume 102 of *Proceedings of The National Academy of Sciences of the United States*. National Academy of Sciences.