Three Essays on Structural Stability of Time Series Models

Inauguraldissertation

zur

Erlangung des Doktorgrades

der

Wirtschafts- und Sozialwissenschaftlichen Fakultät

der

Universität zu Köln

2019

vorgelegt

von

Sven Otto

aus

Hachenburg

Referent:	Prof. Dr. Jörg Breitung
Koreferent:	Prof. Dr. Dominik Wied
Tag der Promotion:	08.10.2019

Acknowledgements

This thesis would not have been possible without the support of many people, and I owe them a great amount of gratitude. First and foremost, I would like to thank my main supervisor Jörg Breitung for his guidance, encouragement, and steady support. During my time as a research assistant at his chair I learned a lot from his way of doing research, and his approaches of finding simple and elegant solutions to complex econometric problems inspired and motivated me. He was always willing and interested to discuss my projects and gave me invaluable comments and advice.

I am thankful to Dominik Wied for his support and feedback concerning technical issues and for accepting to be my second supervisor. Furthermore, I would like to thank Matei Demetrescu for his very helpful comments and support regarding the first chapter of this work, as well as Robinson Kruse-Becher for his support on my academic path. The joint project with Nazarii Salish was a great experience and I would like to thank him for many hours of fruitful discussions during my research visit in Madrid and via video chat.

The last five years wouldn't have been the same without my colleagues and fellow doctoral students at the Institute of Econometrics and Statistics at the University of Cologne, and I am thankful for many interesting academic and non-academic discussions, the pleasant working atmosphere, and the wonderful time.

Most importantly, I am deeply grateful to my family for their unconditional support over the years. Nora deserves more than gratitude for her support and belief in me.

Contents

Introduction **1**

1 Unit Root Testing with Slowly Varying Trends **4**
1.1 Introduction 4
1.2 The pooled estimator 6
1.3 Pseudo t-statistics for unit root testing 9
1.4 Testing under heteroskedasticity 11
1.5 Testing under short-run dynamics 14
1.6 Simulations 16
1.7 Empirical illustrations 25
1.8 Conclusion 25
Appendix to Chapter 1 27

2 Backward CUSUM for Testing and Monitoring Structural Change **47**
2.1 Introduction 47
2.2 The multivariate CUSUM process 50
2.3 CUSUM detectors 52
2.3.1 Forward CUSUM 53
2.3.2 Backward CUSUM 56
2.3.3 Stacked backward CUSUM 58
2.4 Simulations 60
2.4.1 Local asymptotic power and delay 60
2.4.2 Critical values and finite sample performance 64
2.5 Conclusion 67
Appendix to Chapter 2 68

3 A Dynamic Functional Factor Model for Yield Curves: Identification, Estimation, and Prediction **77**

3.1 Introduction 77
3.2 The dynamic functional factor model 79
3.3 The functional principal components estimator 85
3.4 Predictions and prediction bands 88
3.5 Application to yields for U.S. Treasuries 91
3.5.1 From discrete data to functional data 92
3.5.2 Functional principal component analysis 94
3.5.3 Yield curve prediction 95
3.6 Conclusion 100
Appendix to Chapter 3 101

References **119**

List of Figures

1.1 Plots of the trend functions 17

2.1 Retrospective testing and monitoring 48
2.2 Illustrative example for the backward CUSUM with a break in the mean . 57
2.3 Asymptotic local power curves for retrospective testing 61
2.4 Asymptotic local mean delay curves for monitoring with $m = 4$ 63
2.5 Size distributions of the retrospective detectors 63
2.6 Size distributions of the monitoring detectors with $m = 10$ 63

3.1 Nelson-Siegel loading functions 80
3.2 Yields of U.S. Treasury bonds 81
3.3 Fitted Nelson-Siegel curve and cubic B-spline representation 82
3.4 Explained variance of the factors and scree plot 91
3.5 Empirical functional principal components 92
3.6 Effects of the first six functional principal components 93
3.7 Empirical functional principal component scores 94
3.8 One-month ahead yield curve predictions and prediction bands 95

List of Tables

1.1 Asymptotic critical values for the fixed-b test 11
1.2 Trend functions 16
1.3 Size and size-adjusted powers under the zero-trend specification 19
1.4 Size and size-adjusted powers under different trends and i.i.d. errors (1/2) . 20
1.5 Size and size-adjusted powers under different trends and i.i.d. errors (2/2) . 21
1.6 Size and size-adjusted powers under different trends and AR(1) errors . . . 22
1.7 Size and size-adjusted powers of robust tests under constant trend and variance 23
1.8 Size and size-adjusted power of robust tests under breaks in trend and variance 24
1.9 Unit root tests applied to inflation rates 26

2.1 Asymptotic critical values for the retrospective tests 64
2.2 Empirical sizes of the retrospective tests 64
2.3 Size-adjusted powers of the retrospective tests 65
2.4 Asymptotic critical values for the stacked backward CUSUM monitoring . 66
2.5 Empirical sizes of the infinite horizon monitoring detectors 67
2.6 Empirical mean detection delays of the monitoring detectors 67

3.1 Root mean square forecast errors 96
3.2 Diebold-Mariano test statistics 97
3.3 Coverage rates and average widths of one-month-ahead interval forecasts . 98
3.4 Coverage rates of one-month-ahead simultaneous prediction bands 98

Introduction

The analysis of the inherent structure of time-dependent data is crucial for the selection of a suitable model. This thesis is comprised of three self-contained essays on time series models and the identification of their stability properties. Hypothesis tests are particularly useful in this context. While the first chapter presents a unit root test that is robust against an unknown nonparametric trend, the second chapter deals with testing for structural change in linear regression models. The third chapter is devoted to the analysis of the functional dependence structure of bond yields with different maturities, and discusses the identification and estimation of a functional factor model for yield curves from the perspective of functional data analysis. A more detailed description of each chapter is given in the remainder of the introduction.

While the first chapter is based on a single-author paper (see Otto 2019), the latter two chapters are joint works with Jörg Breitung (see Otto and Breitung 2019) and Nazarii Salish (see Otto and Salish 2019) respectively.

Chapter 1 The literature on unit root testing is large and comprehensive, beginning with the seminal works of Dickey and Fuller (1979), Said and Dickey (1984), Phillips (1987), and Phillips and Perron (1988). Elliott et al. (1996) presented a unit root test that exhibits optimality properties. These conventional unit root tests include the assumption that the deterministic component is either constant or linear. Since a misspecified trend model leads to power losses, many studies focused on unit root testing under a more flexible parametric structure for the trend component, such as structural break and smooth transition models with unknown breakpoint and magnitude, and approximations by Chebyshev polynomials and Fourier series. However, little attention has been devoted to the nonparametric treatment of deterministic trends.

The testing approach presented in this chapter is based on the idea that in a window every smooth trend can be approximated well by a constant if the window size is small enough. The time series is proposed to be divided into overlapping blocks, and the unit root hypothesis is tested based on a pooled regression across all these blocks. Since the

trend is estimated for each block separately, the pooled OLS estimator filters out the trend component, asymptotically. Both fixed-b and small-b block asymptotics are considered, and the limiting distributions of the t-statistics for the unit root hypothesis are derived. A nuisance parameter correction under heteroskedasticity provides heteroskedasticity-robust tests, and serial correlation is accounted for by a pre-whitening scheme. Furthermore, the limiting distributions under local alternatives are derived. An extensive Monte Carlo study shows that the proposed tests yield improved power results under both slowly varying trends and sharp breaks when compared to conventional unit root tests. Moreover, the tests are well sized and comparable to the conventional tests in terms of power if the trend is constant.

Chapter 2 The CUSUM test of Brown et al. (1975) for detecting changes in the coefficients of a linear regression and the corresponding monitoring procedure of Chu et al. (1996) suffer from low power and large detection delay. Therefore, two alternative detector statistics are proposed. The backward CUSUM detector sequentially cumulates the recursive residuals in reverse chronological order, whereas the stacked backward CUSUM detector considers a triangular array of backward cumulated residuals. While both the backward CUSUM detector and the stacked backward CUSUM detector are suitable for retrospective testing, only the stacked backward CUSUM detector can be monitored online. The limiting distributions of the maximum statistics under suitable sequences of alternatives are derived for retrospective testing and fixed endpoint monitoring. In the retrospective testing context, the local power of the tests is shown to be substantially higher than that for the conventional CUSUM test if a single break occurs after one third of the sample size. When applied to monitoring schemes, the detection delay of the stacked backward CUSUM is shown to be much shorter than that of the conventional monitoring CUSUM procedure. Moreover, an infinite horizon monitoring procedure and critical values are presented.

Chapter 3 The problem of yield curve modeling and forecasting from a functional time series perspective is discussed. In the fashion of the vector-valued factor models of Stock and Watson (2002) and Bai (2003), a functional factor model for yield curves is studied, in which the factors follow some linear autoregressive process. The model is identified by imposing suitable orthogonality conditions on the factors and the loading functions, while both factors and loadings are unknown. The model can be seen as an extension of the yield curve models by Nelson and Siegel (1987) and Diebold and Li (2006), in which the loadings are fixed and predefined. By applying the least squares principle, a functional principal

components based estimator is obtained, which is shown to be consistent. Furthermore, the minimum mean squared error forecast from the dynamic functional factor model is derived. By imposing normality of the factors and the errors, pointwise and simultaneous prediction bands are obtained from the forecast error curve distribution. The accuracy of the predictions and prediction bands is discussed in an out-of-sample experiment with monthly yield curves of U.S. Treasuries.

Chapter 1

Unit Root Testing with Slowly Varying Trends

1.1 Introduction

It is widely debated in the time series literature whether macroeconomic variables such as GDP, inflation, and interest rates are $I(1)$ or $I(0)$ around a deterministic trend. Dickey-Fuller-type unit root tests often fail to reject the null hypothesis for these time series. The trend component of a time series y_t is typically treated as known up to some parameter vector. The most commonly applied unit root tests, such as those developed by Dickey and Fuller (1979), Said and Dickey (1984), Phillips (1987), Phillips and Perron (1988), and Elliott et al. (1996), impose either a constant or a linear trend model. If, however, the deterministic trend component is nonlinear, highly persistent trend-stationary processes can be hardly distinguishable from unit root processes.

It is not only a misspecified trend model that may lead to high power losses, as an overparameterized model can also reduce the power of unit root tests. Therefore, many authors have suggested applying trend models that seem more suitable for macro data. Broken trend models with one-time changes in mean or slope with known breakpoint were first studied by Perron (1989) and Rappoport and Reichlin (1989). Christiano (1992) demonstrated that a broken trend model with an unknown breakpoint is more adequate, and Zivot and Andrews (1992), as well as Banerjee et al. (1992), proposed unit root tests for this framework. Structural changes in innovation variances were studied by Hamori and Tokihisa (1997), Kim et al. (2002), and Cavaliere (2005), while Cavaliere et al. (2011) considered unit root testing under broken trends together with nonstationary volatility. Leybourne et al. (1998), Kapetanios et al. (2003), and Kılıç (2011) allowed

for exponential smooth transitions from one trend regime to another. Bierens (1997) approximated a nonlinear mean function with Chebyshev polynomials, and Enders and Lee (2012) proposed a Fourier series approximation of the trend, which are approaches that can be used when the exact form and date of structural changes are unknown.

Dickey-Fuller-type tests are based on the t-statistic of the first-order autoregressive parameter. In case of a constant trend, the estimator is derived from a regression of Δy_t on $(y_{t-1} - \overline{y})$, where $\overline{y}$ is the sample mean. Schmidt and Phillips (1992) estimated the constant by the initial observation, which results in a regression of Δy_t on $(y_{t-1} - y_1)$. Whereas a constant is often not a good approximation, in a small block, a moderately varying trend can be approximated quite closely by a constant. To exploit this fact, we divide the series into $T - B$ overlapping blocks of length B. As the blocks can be considered as units of a panel, we follow the panel unit root tests proposed by Breitung (2000) and Levin et al. (2002) and consider a pooled regression of Δy_{j+t} on $(y_{j+t-1} - y_j)$ for $2 \leq t \leq T$ and $1 \leq j \leq T - B$. The deterministic function is approximated locally by a constant. Under a general class of piecewise continuous nonparametric trend functions, the resulting pooled estimator is consistent as $B, T \to \infty$. The limiting null distribution of the t-statistic is a functional of Brownian motions under fixed-b asymptotics. Under small-b asymptotics, a normal distribution is obtained.

The chapter is organized as follows: In Section 1.2 the autoregressive model with independent and homoskedastic errors is analyzed together with the asymptotic behavior of the pooled least squares estimator in the presence of a general nonparametric trend model. For both fixed-b and small-b block asymptotics, the limiting distributions are derived under both the unit root hypothesis and under local alternatives. Section 1.3 considers pseudo t-statistics for unit root testing, while Section 1.4 demonstrates that, under heteroskedasticity, nuisance parameters appear in the limiting distributions. The estimation of these parameters is discussed, and heteroskedasticity-robust test statistics are provided. In Section 1.5, a pre-whitening procedure is proposed in order to account for short-run dynamics, while Section 1.6 reports on Monte Carlo simulations. The tests are found to have only minor size distortions in small samples and are sized correctly in larger samples. It is shown that in the presence of slowly varying trends, pooled tests tend to yield higher power than conventional unit root tests. In Section 1.7, these tests are applied to the monthly inflation rates of 25 countries. The results provide some evidence in favor of inflation rates being trend-stationary around a slowly varying deterministic component. Finally, Section 1.8 presents the conclusion.

1.2 The pooled estimator

We are interested in inference concerning the autoregressive parameter ρ in the model

$$y_t = d_t + x_t, \quad x_t = \rho x_{t-1} + u_t, \quad t = 1, \ldots, T,$$

where ρ is close or equal to one. The deterministic trend component d_t is treated as nonstochastic and fixed in repeated samples, where its functional form is nonparametric and unknown.

Assumption 1.1 (trend component)**.** *The trend component is given by $d_t = d(t/T)$, where $d(r)$ is a piecewise Lipschitz continuous function.*

Note that any continuously differentiable function is Lipschitz continuous. Lipschitz functions are locally close to a constant value in the sense that there exists some $C < \infty$ such that $|d(r) - d(s)| \leq C|r - s|$ for all $r, s \in \mathbb{R}$. The piecewise Lipschitz condition allows for a partition with a finite number of intervals, such that $d(r)$ is Lipschitz continuous on each interval. This includes both smooth changes as well as abrupt breaks in the trend function. For the initial value, it is assumed that $E[x_0^2] < \infty$. We introduce the pooled estimator and the unit root test statistics under the following simplified assumptions on the error term, which are relaxed in the subsequent sections:

Assumption 1.2 (i.i.d. errors)**.** *The process $\{u_t\}_{t\in\mathbb{N}}$ is independently distributed, where $E[u_t] = 0$, $E[u_t^2] = \sigma^2$ and $E[u_t^4] < \infty$.*

The principal approach to dealing with a nonparametric, slowly varying trend is to approximate the unknown trend locally by a constant. Let B be some blocklength that satisfies $2 \leq B < T$. We divide the time series into $T - B$ overlapping blocks of length B and then block-wise estimate ρ via OLS under a constant trend specification. In the fashion of Schmidt and Phillips (1992), as well as Breitung and Meyer (1994), the constant trend is approximated by the first observation in each block. Thereafter, by pooling the $T - B$ individual block regressions, we obtain the following regression equation:

$$\Delta y_{t+j} = \phi(y_{t+j-1} - y_j) + u_{t+j}, \qquad t = 2, \ldots, B, \quad j = 1, \ldots, T - B,$$

where $\phi = \rho - 1$. The pooled OLS estimator is formulated as

$$\hat{\phi} = \hat{\rho} - 1 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta y_{t+j}(y_{t+j-1} - y_j)}{\sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2}.$$

In the following, we derive the asymptotic properties for the numerator and the denominator separately. The numerator and denominator statistics are defined as

$$\mathcal{Y}_{1,T} = \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta y_{t+j}(y_{t+j-1} - y_j), \quad \mathcal{Y}_{2,T} = \frac{1}{B^2 T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2,$$

such that $\sqrt{BT}(\hat{\rho} - 1) = \mathcal{Y}_{1,T}/\mathcal{Y}_{2,T}$. Their counterparts for a zero trend component are given by

$$\mathcal{X}_{1,T} = \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta x_{t+j}(x_{t+j-1} - x_j), \quad \mathcal{X}_{2,T} = \frac{1}{B^2 T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (x_{t+j-1} - x_j)^2.$$

In what follows, we show that, under the block procedure, the trend component can be ignored asymptotically. All asymptotic results are jointly derived for $B, T \to \infty$. While the statistics $\mathcal{X}_{1,T}$ and $\mathcal{X}_{2,T}$ are infeasible if d_t is unknown, they can be well approximated by $\mathcal{Y}_{1,T}$ and $\mathcal{Y}_{2,T}$ in the following sense:

Lemma 1.1. *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.2. Then, $|\mathcal{Y}_{1,T} - \mathcal{X}_{1,T}| = o_P(1)$, and $|\mathcal{Y}_{2,T} - \mathcal{X}_{2,T}| = o_P(1)$ as $B, T \to \infty$.*

The block procedure filters out the trend component in the numerator and the denominator asymptotically. Hence, applying Slutsky's theorem, we can write

$$\sqrt{BT}(\hat{\rho} - 1) = \frac{\mathcal{Y}_{1,T}}{\mathcal{Y}_{2,T}} = \frac{\mathcal{X}_{1,T}}{\mathcal{X}_{2,T}} + o_P(1).$$

In order to obtain the limiting distribution, we formulate the following properties for the numerator and denominator statistics:

Lemma 1.2. *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, and let u_t satisfy Assumption 1.2. Then, as $B, T \to \infty$, it follows that*

(a) $\mathcal{X}_{1,T} = \sum_{j=1}^{T} q_{j,T} - c \cdot (\sigma^2/2 + o_P(1))$, *where $\{q_{j,T}\}_{j \leq T, T \in \mathbb{N}}$ is a martingale difference array with*

$$Var\left[\sum_{j=1}^{T} q_{j,T}\right] = \sigma^4 \frac{(B-1)((T-B)(2B-1) - 2(B-2))}{6B^2T},$$

(b) $E[\mathcal{X}_{2,T}] = \sigma^2 \frac{(T-B)(B-1)}{2BT} + c \cdot O(B^{1/2}T^{-1/2})$, *and* $Var[\mathcal{X}_{2,T}] = O(BT^{-1})$.

Lemmas 1.1 and 1.2 imply that $Var[\mathcal{Y}_{2,T}] = O(1)$ if $B = O(T)$, whereas $Var[\mathcal{Y}_{2,T}] = o(1)$ if $B = o(T)$. This suggests distinguishing between these fundamentally different types of blocklength asymptotics. The fixed-b approach denotes the case where the relative blocklength B/T converges to some value b with $0 < b < 1$, such that B and T grow at the same rate. In the small-b approach, we consider a relative blocklength that converges to zero, while $B, T \to \infty$.[1] As the blocks are overlapping, the error terms in the pooled regression equation are correlated, but, fortunately, the correlation structure is known by construction. Lemmas 1.1 and 1.2 imply that $Var[\mathcal{Y}_{1,T}] \to \sigma^4/3$ and $E[\mathcal{Y}_{2,T}] \to \sigma^2/2$ as $B/T \to 0$ and $B, T \to \infty$. Together with the central limit theorem for martingale difference arrays, the following asymptotic result can be stated:

Theorem 1.1 (small-b asymptotics)**.** *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.2. Let $B/T \to 0$ as $B, T \to \infty$. Then,*

$$\mathcal{Y}_{1,T} \xrightarrow{\mathcal{D}} \mathcal{N}\bigg(-\frac{c\sigma^2}{2},\ \frac{\sigma^4}{3}\bigg), \quad \textit{and} \quad \mathcal{Y}_{2,T} \xrightarrow{p} \frac{\sigma^2}{2}.$$

As a direct consequence, the pooled estimator is asymptotically normally distributed under small-b asymptotics. Together with Slutsky's theorem, it follows that

$$\sqrt{BT}(\hat{\rho} - 1) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 4/3)$$

under the unit root hypothesis, which is given by $\rho = 1$ or equivalently by $c = 0$. Under fixed-b asymptotics, the numerator and denominator statistics can be represented as a partial sum process of the innovations. The functional central limit theorem then yields the following asymptotic result:

Theorem 1.2 (fixed-b asymptotics)**.** *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.2. Let $0 < b < 1$, and let $B/T \to b$ as $B, T \to \infty$. Then*

$$\begin{pmatrix} \mathcal{Y}_{1,T} \\ \mathcal{Y}_{2,T} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \frac{\sigma^2}{2b^{3/2}} \left(\int_0^{1-b} (J_{c/b}(b+r) - J_{c/b}(r))^2 - b(1-b) \right) \\ \frac{\sigma^2}{b^2} \int_0^{1-b} \int_r^{b+r} (J_{c/b}(s) - J_{c/b}(r))^2 \, \mathrm{d}s \, \mathrm{d}r \end{pmatrix},$$

where $J_c(r)$ is an Ornstein-Uhlenbeck process.

[1] Note that the terminology "fixed-b and small-b asymptotics" has also been used in the context of long-run variance estimation. Whereas Kiefer and Vogelsang (2005) use this wording for the asymptotics of the ratio of the truncation point to the sample size, we consider the ratio of the blocklength to the sample size.

Consequently, the pooled estimator is asymptotically represented as a functional of a standard Brownian motion $W(r)$ under the unit root hypothesis. If $\rho = 1$, then Theorem 1.2, together with the continuous mapping theorem, implies that

$$\sqrt{BT}(\hat{\rho} - 1) \xrightarrow{\mathcal{D}} \frac{b^{1/2} \int_0^{1-b} \left(W(b+r) - W(r)\right)^2 \, \mathrm{d}r + b^{3/2}(1-b)}{2 \int_0^{1-b} \int_r^{b+r} \left(W(s) - W(r)\right)^2 \, \mathrm{d}s \, \mathrm{d}r}$$

under fixed-b asymptotics. In comparison to the limiting distribution of the ρ-statistic in the Dickey-Fuller framework, the functional includes an additional integral, which results from pooling the block regressions.

1.3 Pseudo t-statistics for unit root testing

The principal concept of Dickey-Fuller-type unit root tests is to consider a t-test for the null hypothesis $H_0 : \rho = 1$. Following this approach in the pooled regression framework, the numerator of the t-statistic can be represented as $\hat{\rho} - 1 = \mathcal{Y}_{1,T}(\mathcal{Y}_{2,T}\sqrt{BT})^{-1}$. The standard error is obtained from the conditional variance of $\hat{\rho}$. Let

$$s_{\hat{\rho}}^2 = \hat{\sigma}^2 \Big(\sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2 \Big)^{-1} = \frac{\hat{\sigma}^2}{\mathcal{Y}_{2,T} B^2 T},$$

where $\hat{\sigma}^2$ denotes some consistent estimator for the error variance σ^2. The conventional t-statistic is then represented as $(\hat{\rho}-1)/s_{\hat{\rho}} = \sqrt{B}\mathcal{Y}_{1,T}/\sqrt{\hat{\sigma}^2 \mathcal{Y}_{2,T}}$ and diverges in probability under H_0. Accordingly, we consider a pseudo t-statistic of the form

$$\tau = \frac{\hat{\rho} - 1}{s_{\hat{\rho}}\sqrt{B}} = \frac{\mathcal{Y}_{1,T}}{\hat{\sigma}\sqrt{\mathcal{Y}_{2,T}}},$$

which is $O_P(1)$ under both small-b and fixed-b asymptotics. We consider the residuals $\hat{u}_t = y_t - \hat{\rho} y_{t-1}$ for $t = 2, \ldots, T$ and their sample mean $\overline{\hat{u}} = T^{-1} \sum_{j=1}^{T} \hat{u}_j$. For the error variance estimation, we distinguish between fixed-b and small-b block asymptotics and define

$$\hat{\sigma}^2_{\mathrm{sb}} = \frac{\sum_{j=1}^{T-B} \sum_{t=1}^{B} \Big(\hat{u}_{j+t} - \frac{1}{B} \sum_{k=1}^{B} \hat{u}_{j+k} \Big)^2}{(T-B)(B-1)}, \quad \hat{\sigma}^2_{\mathrm{fb}} = \frac{1}{T} \sum_{j=2}^{T} (\hat{u}_j - \overline{\hat{u}})^2.$$

The following consistency result can then be obtained:

Lemma 1.3. *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.2.*

(a) Let $B/T \to 0$ as $B, T \to \infty$. Then, $\hat{\sigma}^2_{sb} \xrightarrow{p} \sigma^2$.

(b) Let $0 < b < 1$, and let $B/T \to b$ as $B, T \to \infty$. Then, $\hat{\sigma}^2_{fb} \xrightarrow{p} \sigma^2$.

In what follows, the pseudo t-tests are defined. For the small-b pseudo t-statistic, we scale $\mathcal{Y}_{1,T}$ and $\mathcal{Y}_{2,T}$ by their finite sample variance and their expectation from Lemma 1.2 in order to avoid small-sample size distortions. Let

$$v_T^2 = \frac{(T-B)(2B-1) - 2(B-2)}{3B(T-B)},$$

which is equal to $\sigma^{-2}Var[\mathcal{X}_{1,T}]/E[\mathcal{X}_{2,T}]$ under the unit root hypothesis. The small-b pseudo t-statistic is then defined as

$$\tau\text{-SB} = \frac{\mathcal{Y}_{1,T}}{v_T \hat{\sigma}_{\text{sb}} \sqrt{\mathcal{Y}_{2,T}}} = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta y_{t+j}(y_{t+j-1} - y_j)}{\hat{\sigma}_{\text{sb}} \sqrt{\frac{(T-B)(2B-1)-2(B-2)}{3(T-B)} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2}}.$$

For the fixed-b statistic, we consider the unscaled versions and define

$$\tau\text{-FB} = \frac{\mathcal{Y}_{1,T}}{\hat{\sigma}_{\text{fb}} \sqrt{\mathcal{Y}_{2,T}}} = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta y_{t+j}(y_{t+j-1} - y_j)}{\hat{\sigma}_{\text{fb}} \sqrt{B \sum_{j=1}^{T-B} \sum_{t=2}^{B} (y_{t+j-1} - y_j)^2}}.$$

The unit root hypothesis is rejected in favor of stationarity if the test statistic is smaller than the α-quantile of the limiting distribution under H_0, where α is the significance level. From Theorems 1.1 and 1.2 and Lemma 1.3, together with the continuous mapping theorem and Slutsky's theorem, the following limiting result can be stated:

Corollary 1.1. *Let $\rho = 1$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.2.*

(a) Let $B/T \to 0$ as $B, T \to \infty$. Then, τ-SB $\xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

(b) Let $0 < b < 1$, and let $B/T \to b$ as $B, T \to \infty$. Then,

$$\tau\text{-FB} \xrightarrow{\mathcal{D}} \frac{\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 \, \mathrm{d}r - b(1-b)}{2\sqrt{b \int_0^{1-b} \int_r^{b+r} \left(W(s) - W(r)\right)^2 \, \mathrm{d}s \, \mathrm{d}r}},$$

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

Table 1.1: Asymptotic critical values for the fixed-b test

α	B/T								
	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9
0.2	-0.788	-0.812	-0.815	-0.799	-0.761	-0.701	-0.623	-0.520	-0.377
0.1	-1.126	-1.128	-1.104	-1.055	-0.987	-0.903	-0.798	-0.664	-0.486
0.05	-1.403	-1.375	-1.327	-1.257	-1.169	-1.067	-0.939	-0.781	-0.573
0.04	-1.486	-1.446	-1.391	-1.318	-1.222	-1.113	-0.978	-0.814	-0.600
0.03	-1.582	-1.534	-1.471	-1.394	-1.291	-1.169	-1.025	-0.855	-0.630
0.02	-1.709	-1.650	-1.579	-1.489	-1.374	-1.246	-1.094	-0.909	-0.669
0.01	-1.904	-1.830	-1.745	-1.639	-1.511	-1.361	-1.191	-0.995	-0.729
0.001	-2.431	-2.320	-2.203	-2.042	-1.882	-1.692	-1.480	-1.226	-0.905

Note: The sample paths of the standard Brownian motions contained in the asymptotic null distribution of τ-FB are simulated by a discretized version of $W(r)$ on a grid of 50,000 equidistant points. The empirical quantiles are obtained from 100,000 Monte Carlo repetitions.

For τ-SB we can rely on standard normal quantiles as critical values. However, the limiting distribution of τ-FB is nonstandard. Table 1.1 presents simulated left-tailed quantiles of the null distribution for various relative blocklengths B/T and significance levels.

From Theorems 1.1 and 1.2, it follows that the tests have power against alternatives of the form $\rho = 1 - c/\sqrt{BT}$, where $c > 0$.

1.4 Testing under heteroskedasticity

While stationary time-varying conditional variances such as ARCH and GARCH processes do not affect unit root testing, Hamori and Tokihisa (1997) showed that permanent changes in volatility, in contrast, dramatically alter the limiting distributions of unit root tests. Kim et al. (2002) reported that an abrupt break in the innovation variance can produce spurious rejections, while Cavaliere (2005) showed that nonstationary volatility induces a time-shift in the right-hand-side process of the functional central limit theorem. A variance-transformed Brownian process $W_\eta(r)$ appears in the limiting distributions of Dickey-Fuller-type unit root tests. Given the variance profile $\eta(s) = (\int_0^1 \sigma(r)^2 dr)^{-1} \int_0^s \sigma(r)^2 dr$, the transformed process is defined as $W_\eta(r) = W(\eta(r))$, where $0 \leq r \leq 1$. In what follows, we relax Assumption 1.2 and allow for heteroskedastic errors.

Assumption 1.3 (heteroskedastic errors)**.** *The process $\{u_t\}_{t\in\mathbb{N}}$ is independently distributed with $E[u_t] = 0$, $E[u_t^2] = \sigma_t^2$ and $E[u_t^4] < \infty$, where $\sigma_t = \sigma(t/T)$. The function $\sigma(r)$ is càdlàg, non-stochastic, strictly positive, and bounded.*

Notice that the approximation result of Lemma 1.1 is not affected by Assumption 1.3

and can be formulated under heteroskedasticity as follows:

Lemma 1.4. *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.3. Then, $|\mathcal{Y}_{1,T} - \mathcal{X}_{1,T}| = o_P(1)$, and $|\mathcal{Y}_{2,T} - \mathcal{X}_{2,T}| = o_P(1)$ as $B, T \to \infty$.*

However, nuisance parameters then appear in the limiting distributions of the numerator and denominator statistics.

Theorem 1.3. *Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.3.*

(a) Let $B/T \to 0$ as $B, T \to \infty$. Then,

$$\mathcal{Y}_{1,T} \xrightarrow{\mathcal{D}} \mathcal{N}\bigg(-\frac{c}{2}\int_0^1 \sigma^2(r)\,\mathrm{d}r,\ \frac{1}{3}\int_0^1 \sigma^4(r)\,\mathrm{d}r\bigg), \quad \text{and} \quad \mathcal{Y}_{2,T} \xrightarrow{p} \frac{1}{2}\int_0^1 \sigma^2(r)\,\mathrm{d}r.$$

(b) Let $0 < b < 1$, and let $B/T \to b$ as $B, T \to \infty$. Then,

$$\begin{pmatrix} \mathcal{Y}_{1,T} \\ \mathcal{Y}_{2,T} \end{pmatrix} \xrightarrow{\mathcal{D}} \begin{pmatrix} \frac{\int_0^1 \sigma^2(r)\,\mathrm{d}r}{2b^{3/2}}\big(\int_0^{1-b}(J_{c,b,\eta}(b+r) - J_{c,b,\eta}(r))^2 - b(1-b)\big) \\ \frac{\int_0^1 \sigma^2(r)\,\mathrm{d}r}{b^2}\int_0^{1-b}\int_r^{b+r}(J_{c,b,\eta}(s) - J_{c,b,\eta}(r))^2\,\mathrm{d}s\,\mathrm{d}r \end{pmatrix},$$

where $J_{c,b,\eta}(r)$ is a variance-transformed Ornstein-Uhlenbeck process, which is defined as $J_{c,b,\eta}(r) = \int_0^r e^{-(r-s)c/b} dW_\eta(s)$.

In order to correct for the additional nuisance terms, we consider the following estimators. Let

$$\hat{\kappa}^2 = \frac{\sum_{j=1}^{T-B}\sum_{t=1}^{B}\left(\hat{u}_j - \overline{\hat{u}}\right)^2\left(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^{B}\hat{u}_{j+k}\right)^2}{(T-B)(B-1)}$$

and let

$$\hat{\eta}(s) = \frac{\sum_{j=1}^{\lfloor sT\rfloor}\left(\hat{u}_j - \frac{1}{\lfloor sT\rfloor}\sum_{k=1}^{\lfloor sT\rfloor}\hat{u}_k\right)^2 + (sT - \lfloor sT\rfloor)\left(\hat{u}_{\lfloor sT\rfloor+1} - \frac{1}{\lfloor sT\rfloor+1}\sum_{k=1}^{\lfloor sT\rfloor+1}\hat{u}_k\right)^2}{\sum_{j=1}^{T}(\hat{u}_j - \overline{\hat{u}})^2},$$

where $s \in [0,1]$. The robust small-b statistic is then defined as

$$\tau\text{-SB}^{\mathrm{H}} = \frac{\mathcal{Y}_{1,T}}{v_T\hat{\kappa}\hat{\sigma}_{\mathrm{sb}}^{-1}\sqrt{\mathcal{Y}_{2,T}}} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Delta y_{t+j}(y_{t+j-1} - y_j)}{\hat{\kappa}\hat{\sigma}_{\mathrm{sb}}^{-1}\sqrt{\frac{(T-B)(2B-1)-2(B-2)}{3(T-B)}\sum_{j=1}^{T-B}\sum_{t=2}^{B}(y_{t+j-1} - y_j)^2}}.$$

Under fixed-b asymptotics, the nuisance term appears in the Gaussian process itself. By means of transforming the data with its inverse variance profile, Cavaliere and Taylor (2007) showed that the time-transformation in the Gaussian limiting processes can be inverted. Accordingly, we fix some auxiliary sample size $\widetilde{T} \geq T$ and consider the time-transformed series $\tilde{y}_t = y_{\lfloor \hat{\eta}^{-1}(t/\widetilde{T})\widetilde{T} \rfloor}$ for $t = 1, \ldots, \widetilde{T}$, where $\hat{\eta}^{-1}(s)$ is the inverse function of $\hat{\eta}(s)$. In practice, the observed time series is transformed in such a way that copies of adjacent observations between the sample points are inserted in highly volatilie periods. Since $\widetilde{T}$ can be set arbitrarily high, we do not need to discard any observations. We replace the original series in the test statistic by the time-transformed series and define

$$\widetilde{\mathcal{Y}}_{1,T} = \frac{1}{B^{3/2}\widetilde{T}^{1/2}} \sum_{j=1}^{\widetilde{T}-B} \sum_{t=2}^{B} \Delta\tilde{y}_{t+j}(\tilde{y}_{t+j-1} - \tilde{y}_j), \quad \widetilde{\mathcal{Y}}_{2,T} = \frac{1}{B^2\widetilde{T}} \sum_{j=1}^{\widetilde{T}-B} \sum_{t=2}^{B} (\tilde{y}_{t+j-1} - \tilde{y}_j)^2,$$

which yields the fixed-b heteroskedasticity-robust statistic

$$\tau\text{-FB}^{\text{H}} = \frac{\widetilde{\mathcal{Y}}_{1,T}}{\hat{\sigma}_{\text{fb}}\sqrt{\widetilde{\mathcal{Y}}_{2,T}}} = \frac{\sum_{j=1}^{\widetilde{T}-B}\sum_{t=2}^{B} \Delta\tilde{y}_{t+j}(\tilde{y}_{t+j-1} - \tilde{y}_j)}{\hat{\sigma}_{\text{fb}}\sqrt{B\sum_{j=1}^{\widetilde{T}-B}\sum_{t=2}^{B}(\tilde{y}_{t+j-1} - \tilde{y}_j)^2}}.$$

Theorem 1.4. *Let $\rho = 1$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.3.*

(a) Let $B/T \to 0$ as $B, T \to \infty$. Then, $\hat{\sigma}^2_{sb} \xrightarrow{p} \int_0^1 \sigma^2(r)\,\mathrm{d}r$, $\hat{\kappa}^2 \xrightarrow{p} \int_0^1 \sigma^4(r)\,\mathrm{d}r$, and $\tau\text{-}SB^H \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$.

(b) Let $0 < b < 1$, and let $B/\widetilde{T} \to b$ as $B, \widetilde{T} \to \infty$. Then, $\hat{\sigma}^2_{fb} \xrightarrow{p} \int_0^1 \sigma^2(r)\,\mathrm{d}r$, $\hat{\eta}(s) \xrightarrow{p} \eta(s)$ uniformly for all $s \in [0,1]$, and

$$\tau\text{-}FB^H \xrightarrow{\mathcal{D}} \frac{\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 \,\mathrm{d}r + b(1-b)}{2\sqrt{b\int_0^{1-b}\int_r^{b+r} \left(W(s) - W(r)\right)^2 \,\mathrm{d}s\,\mathrm{d}r}}.$$

The limiting distributions under the unit root hypothesis of the heteroskedasticity-robust test statistics coincide with those obtained in Section 1.3 under homoskedasticity. Hence, the critical values from those tests can be retained. For τ-SB$^{\text{H}}$, we consider standard normal quantiles, and, for τ-FB$^{\text{H}}$, we can apply the values from Table 1.1.

1.5 Testing under short-run dynamics

A more realistic scenario for macroeconomic variables is that error terms are serially correlated. We impose the following assumption on the error process:

Assumption 1.4 (serially correlated errors)**.** *The process $\{u_t\}_{t\in\mathbb{N}}$ possesses the stationary AR(p) representation $u_t = \sum_{i=1}^{p} \theta_i u_{t-i} + \epsilon_t$. The process $\{\epsilon_t\}_{t\in\mathbb{Z}}$ is independently distributed with $E[\epsilon_t] = 0$, $E[\epsilon_t^2] = \sigma_t^2$ and $E[\epsilon_t^4] < \infty$, where $\sigma_t = \sigma(t/T)$. The function $\sigma(r)$ is càdlàg, non-stochastic, strictly positive, and bounded. The lag order p satisfies $T^{-1/4}p \to 0$.*

In the fashion of Said and Dickey (1984), we allow the lag order to grow with the sample size. Asymptotically, this allows for fairly general forms of serially correlated errors, such as stationary and invertible ARMA processes. In order to correct for the effect of short-run dynamics, we follow Breitung and Das (2005) and consider the pre-whitened series $y_t^* = y_t - \sum_{i=1}^{p} \theta_i y_{t-i}$. The series decomposes into $y_t^* = d_t^* + x_t^*$, where the deterministic and the stochastic parts are given by $d_t^* = d_t - \sum_{i=1}^{p} \theta_i d_{t-i}$ and $x_t^* = x_t - \sum_{i=1}^{p} \theta_i x_{t-i}$ respectively. Note that $x_t^* - \rho x_{t-1}^* = \epsilon_t$, where ϵ_t satisfies the same conditions as u_t under Assumption 1.3. Consequently, if the unit root statistics are defined in terms of

$$\mathcal{X}_{1,T}^* = \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta x_{t+j}^* (x_{t+j-1}^* - x_j^*), \quad \mathcal{X}_{2,T}^* = \frac{1}{B^2 T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (x_{t+j-1}^* - x_j^*)^2$$

instead of $\mathcal{X}_{1,T}$ and $\mathcal{X}_{2,T}$, their limiting distributions then coincide with those presented in the previous sections.

Since the autoregressive parameters of the error process are unknown, they need to be estimated. We augment the regression equation with lagged values of the differenced series, such that

$$\Delta y_t = \varphi y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + e_t, \tag{1.1}$$

for $t = p+1, \ldots, T$, where e_t is a mean-zero error term. Let $\hat{\varphi}$ and $\hat{\beta}_1, \ldots, \hat{\beta}_p$ denote the OLS estimators of the parameters. In the following, we show that $(\hat{\beta}_1, \ldots, \hat{\beta}_p)'$ is consistent for $(\theta_1, \ldots, \theta_p)'$ under the unit root hypothesis:

Lemma 1.5. *Let $\rho = 1$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.4. Then $\max_{1\leq i\leq p} p|\hat{\beta}_i - \theta_i| = o_P(1)$ as $B, T \to \infty$.*

The estimated pre-whitened series is defined as $\hat{y}_t^* = y_t - \sum_{i=1}^p \hat{\beta}_i y_{t-i}$, and the corresponding numerator and denominator statistics are given by

$$\hat{\mathcal{Y}}_{1,T}^* = \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta\hat{y}_{t+j}^* (\hat{y}_{t+j-1}^* - \hat{y}_j^*), \quad \hat{\mathcal{Y}}_{2,T}^* = \frac{1}{B^2T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (\hat{y}_{t+j-1}^* - \hat{y}_j^*)^2.$$

Lemma 1.6. *Let $\rho = 1$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.4. Then, $|\hat{\mathcal{Y}}_{1,T}^* - \mathcal{X}_{1,T}^*| = o_P(1)$, and $|\hat{\mathcal{Y}}_{2,T}^* - \mathcal{X}_{2,T}^*| = o_P(1)$ as $B, T \to \infty$.*

The estimators $\hat{\sigma}_{\text{sb}}^{*2}$, $\hat{\sigma}_{\text{fb}}^{*2}$, $\hat{\kappa}^{*2}$, and $\hat{\eta}^*(s)$ are defined as their counterparts in Sections 1.3 and 1.4, except that the residuals are now defined as $\hat{u}_t = \hat{y}_t^* - \hat{\rho}^* \hat{y}_{t-1}^*$, where $\hat{\rho}^*$ is given by $\sqrt{BT}(\hat{\rho}^* - 1) = \hat{\mathcal{Y}}_{1,T}^* / \hat{\mathcal{Y}}_{2,T}^*$. Analogously to Section 1.4, we consider the time-transformed pre-whitened series $\tilde{y}_t^* = \hat{y}_{\lfloor \hat{\eta}^{-1}(t/\widetilde{T})\widetilde{T} \rfloor}^*$ for $t = 1, \ldots, \widetilde{T}$, where $\widetilde{T} \geq T$, and we define

$$\widetilde{\mathcal{Y}}_{1,T}^* = \frac{1}{B^{3/2}\widetilde{T}^{1/2}} \sum_{j=1}^{\widetilde{T}-B} \sum_{t=2}^{B} \Delta\tilde{y}_{t+j}^* (\tilde{y}_{t+j-1}^* - \tilde{y}_j^*), \quad \widetilde{\mathcal{Y}}_{2,T}^* = \frac{1}{B^2\widetilde{T}} \sum_{j=1}^{\widetilde{T}-B} \sum_{t=2}^{B} (\tilde{y}_{t+j-1}^* - \tilde{y}_j^*)^2.$$

The pre-whitened versions of the test statistics are then given by

$$\tau\text{-SB}^{\text{PW}} = \frac{\hat{\mathcal{Y}}_{1,T}^*}{v_T \hat{\sigma}_{\text{sb}}^* \sqrt{\hat{\mathcal{Y}}_{2,T}^*}}, \quad \tau\text{-SB}^{\text{H-PW}} = \frac{\hat{\mathcal{Y}}_{1,T}^*}{v_T \hat{\kappa}^* \hat{\sigma}_{\text{sb}}^{*-1} \sqrt{\hat{\mathcal{Y}}_{2,T}^*}},$$
$$\tau\text{-FB}^{\text{PW}} = \frac{\hat{\mathcal{Y}}_{1,T}^*}{\hat{\sigma}_{\text{fb}}^* \sqrt{\hat{\mathcal{Y}}_{2,T}^*}}, \quad \tau\text{-FB}^{\text{H-PW}} = \frac{\widetilde{\mathcal{Y}}_{1,T}^*}{\hat{\sigma}_{\text{fb}}^* \sqrt{\widetilde{\mathcal{Y}}_{2,T}^*}}.$$

Theorem 1.5. *Let $\rho = 1$, let d_t satisfy Assumption 1.1, and let u_t satisfy Assumption 1.4.*

(a) Let $B/T \to 0$ as $B, T \to \infty$. Then, $\hat{\sigma}_{sb}^{*2} \xrightarrow{p} \int_0^1 \sigma^2(r)\,\mathrm{d}r$, $\hat{\kappa}^{*2} \xrightarrow{p} \int_0^1 \sigma^4(r)\,\mathrm{d}r$, and $\tau\text{-}SB^{H\text{-}PW} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$. Furthermore, $\tau\text{-}SB^{PW} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1)$ if $\sigma_t^2 = \sigma^2$ for all t.

(b) Let $0 < b < 1$, and let $B/\widetilde{T} \to b$ as $B, \widetilde{T} \to \infty$. Then, $\hat{\sigma}_{fb}^{*2} \xrightarrow{p} \int_0^1 \sigma^2(r)\,\mathrm{d}r$, $\hat{\eta}(s) \xrightarrow{p} \eta(s)$ uniformly for all $s \in [0,1]$, and

$$\tau\text{-}FB^{H\text{-}PW} \xrightarrow{\mathcal{D}} \frac{\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 \,\mathrm{d}r + b(1-b)}{2\sqrt{b \int_0^{1-b} \int_r^{b+r} \left(W(s) - W(r)\right)^2 \,\mathrm{d}s\,\mathrm{d}r}}.$$

Furthermore, if $\sigma_t^2 = \sigma^2$ *for all* $t \in \mathbb{N}$*, then*

$$\tau\text{-}FB^{PW} \xrightarrow{\mathcal{D}} \frac{\int_0^{1-b} \left(W(b+r) - W(r)\right)^2 \mathrm{d}r + b(1-b)}{2\sqrt{b\int_0^{1-b}\int_r^{b+r} \left(W(s) - W(r)\right)^2 \mathrm{d}s\,\mathrm{d}r}}.$$

The lag order p is typically unknown in practice and can be chosen using conventional lag order selection methods, such as the Bayesian information criterion (BIC) or by the general-to-specific methodology in the fashion of Ng and Perron (1995). The maximum lag order is inspired by the rule of thumb provided by Schwert (1989) and is fixed as $p^* = \lfloor 4 \cdot (T/100)^{1/5} \rfloor$ or as $p^* = \lfloor 12 \cdot (T/100)^{1/5} \rfloor$.

1.6 Simulations

In this section, the finite sample performance of the unit root tests is evaluated by means of Monte Carlo simulations. The analysis includes several specifications for both the deterministic part d_t and the stochastic part x_t.

While the zero-trend $d_t = 0$ is the main benchmark, we consider several other trends including sharp breaks and smooth changes of different shapes. The trend specifications are presented in Table 1.2 and Figure 1.1. The parameter λ determines the size of the break. Similar trend functions are also considered in Jones and Enders (2014) in order to evaluate the performance of the unit root test by Enders and Lee (2012).

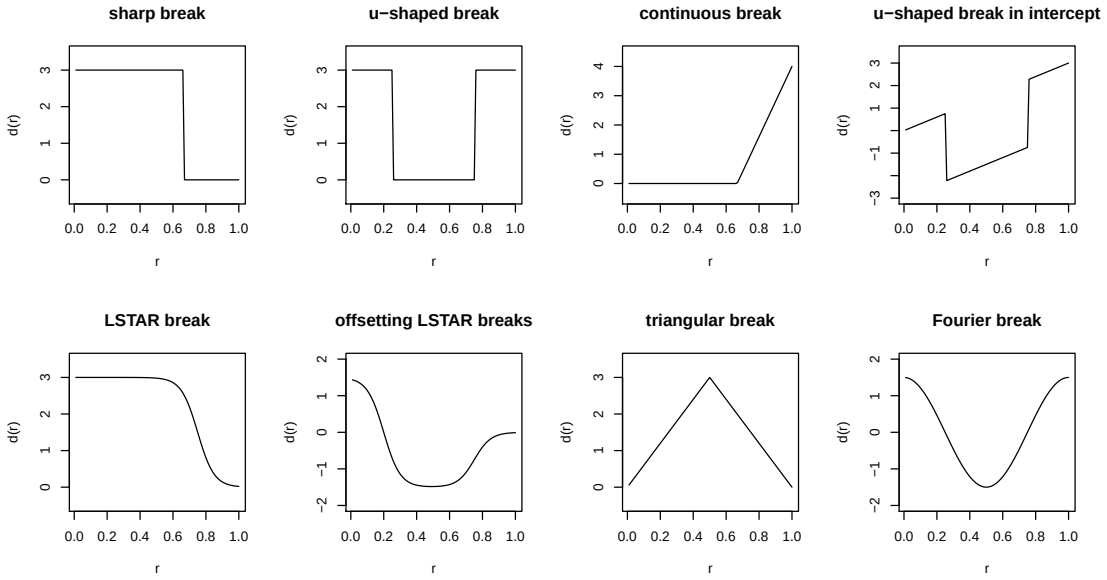
Table 1.2: Trend functions

	type of the trend	functional form
1	sharp break	$d(r) = \lambda \cdot 1_{\{r \leq 2/3\}}$
2	u-shaped break	$d(r) = \lambda \cdot 1_{\{r \leq 1/4\}} + \lambda \cdot 1_{\{r > 3/4\}}$
3	continuous break	$d(r) = \lambda \cdot (4r \cdot 1_{\{r > 2/3\}} - 8/3)$
4	u-shaped break in intercept	$d(r) = \lambda \cdot (r1_{\{r \leq 1/4\}} + (r-1)1_{\{1/4 < r \leq 3/4\}} + r1_{\{t > 3/4\}})$
5	LSTAR break	$d(r) = \lambda \cdot (1 + \exp(20(r - 0.75)))^{-1}$
6	offsetting LSTAR break	$d(r) = \lambda/(1 + \exp(20(r - 0.2))) - 0.5\lambda/(1 + \exp(20(r - 0.75)))$
7	triangular break	$d(r) = \lambda \cdot (2r1_{\{r \leq 1/2\}} + 2(1-r)1_{\{r > 1/2\}})$
8	Fourier break	$d(r) = \lambda \cdot 0.5\cos(2\pi r)$

Note: The functional form of the trend functions, which are considered in the Monte Carlo simulations, are presented. The parameter λ determines the size of the trend.

The stochastic part x_t is simulated both under the null hypothesis $\rho = 1$ and the alternative hypothesis $\rho = 0.9$. For the errors u_t, we consider an independent process as well as the AR(1) process $u_t = 0.5u_{t-1} + \epsilon_t$ with standard normal innovations. Furthermore, results with heteroskedastic innovations using the variance function $\sigma^2(r) = \lambda \cdot 1_{\{r \leq 2/3\}}$ are presented. As noted by Müller and Elliott (2003), the power of a unit root test depends

Figure 1.1: Plots of the trend functions



Note: The plots of the of the trend functions from Table 1.2, which are considered in the Monte Carlo simulations, are presented. The trend size is $\lambda = 3$.

on the initial condition. Thus, we consider both the zero initial condition $x_0 = 0$ as well as a random initial condition, where $x_0 = \sum_{k=1}^{T} \rho^{T-k} \tilde{\epsilon}_k$ is simulated from i.i.d. standard normal innovations $\tilde{\epsilon}_k$.

For the small-b tests, we consider blocklengths of the form $B = T^{\gamma}$ with parameters $\gamma \in \{0.5, 0.6, 0.7, 0.8\}$, and, for the fixed-$b$ versions, we implement $B = b \cdot T$ with relative blocklengths $b \in \{0.2, 0.3, 0.4, 0.5\}$. Size and power results are presented for τ-SB and τ-FB as well as for their pre-whitened and heteroskedasticity-robust versions. Both fixed lag augmentation as well as a flexible lag augmentation determined by the BIC are implemented. All empirical size levels are presented for a significance level of 5%, while the power results are size-adjusted. The models are simulated with 100,000 repetitions for sample sizes of $T = 100$ and $T = 300$.

In order to demonstrate the advantage of the fixed-b and small-b unit root tests, their finite sample results are compared to those obtained by conventional unit root tests. As the main benchmark, we consider the augmented Dickey-Fuller test by Said and Dickey (1984) with constant trend specification (ADF henceforth), which is the t-test for the hypothesis $\varphi = 0$ in the regression

$$\Delta y_t = \varphi y_{t-1} + \beta_0 + \sum_{i=1}^{p} \xi_i \Delta y_{t-i} + e_t.$$

Elliott et al. (1996) proposed detrending the series locally in the ADF regression. Let

the deterministic trend function be given by the vector z_t, and let $\bar{c} \in \mathbb{R}$. Furthermore, let $y_{\bar{c},t} = y_t - \bar{c}y_{t-1}$ and $Z_{\bar{c},t} = z_t - \bar{c}z_{t-1}$ for $t \geq 2$, and let $y_{\bar{c},1} = y_1$ and $Z_{\bar{c},1} = z_1$. The Dickey-Fuller GLS test is then the t-test for the hypothesis $\varphi = 0$ in the regression

$$\Delta y_t^d = \varphi y_{t-1}^d + \sum_{i=1}^{p} \xi_i \Delta y_{t-i}^d + e_t,$$

where $y_t^d = y_t - \hat{\beta}' z_t$ and where $\hat{\beta}$ is the OLS estimator from a regression of $y_{\bar{c},t}$ on $Z_{\bar{c},t}$. For the constant trend specification (DF-GLS henceforth), let $z_t = 1$ and $\bar{c} = 7$, and, for the linear trend specification (DF-GLS-trend henceforth), $z_t = (1, t)'$ and $\bar{c} = 13.5$ are considered. Elliott et al. (1996) showed that the Dickey-Fuller GLS test is optimal for the zero initial condition $x_0 = 0$.

An approach that does not assume a precise model for the trend component is that developed by Enders and Lee (2012) (EL henceforth). A flexible Fourier form is used to approximate smooth breaks in the trend function. Structural changes can be captured by the low frequency components of a series. In its simplest form, Enders and Lee (2012) considered the parametric trend model $d(r) = \alpha_0 + \gamma r + \alpha_1 \sin(2\pi r) + \beta_1 \cos(2\pi r)$. More frequencies could be included, but doing so could lead to an over-fitting problem. The test works as follows: First, the auxiliary regression

$$\Delta y_t = \delta_0 + \delta_1 \Delta \sin(2\pi t/T) + \delta_2 \Delta \cos(2\pi t/T) + v_t$$

is considered with OLS estimates $\widehat{\delta}_0$, $\widehat{\delta}_1$, and $\widehat{\delta}_2$. This yields the detrended series

$$\widetilde{S}_t = y_t - (y_1 - \widehat{\delta}_0 - \widehat{\delta}_1 \sin(\tfrac{2\pi}{T}) - \widehat{\delta}_2 \cos(\tfrac{2\pi}{T})) - \widehat{\delta}_0 t - (\widehat{\delta}_1 \sin(\tfrac{2\pi t}{T}) + \widehat{\delta}_2 \cos(\tfrac{2\pi t}{T})).$$

Finally, the test statistic is given by the t-statistic for the null hypothesis $\varphi = 0$ in the regression

$$\Delta y_t = \varphi \widetilde{S}_{t-1} + \beta_0 + \beta_1 \Delta \sin(2\pi t/T) + \beta_2 \Delta \cos(2\pi t/T) + \sum_{i=1}^{p} \xi_i \Delta \widetilde{S}_{t-i} + e_t.$$

For all tests, the lag augmentation order p is either fixed or flexibly determined by the BIC with a maximum lag order of $p^* = \lfloor 4 \cdot (T/100)^{1/5} \rfloor$.

The results presented in Tables 1.3–1.5 indicate that the pooled tests are slightly undersized for smaller sample sizes, where the size distortions become larger as the break gets larger. However, for larger sample sizes, the size distortions decline. Overall, the size levels are similar to those obtained from using the conventional unit root tests.

Table 1.3: Size and size-adjusted powers under the zero-trend specification

	zero initial condition				random initial condition			
	$T = 100$		$T = 300$		$T = 100$		$T = 300$	
	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$
i.i.d. errors – no lag augmentation (p=0)								
τ-SB, $B = T^{0.5}$	0.057	0.294	0.056	0.845	0.056	0.281	0.058	0.838
τ-SB, $B = T^{0.6}$	0.057	0.351	0.057	0.952	0.057	0.334	0.058	0.948
τ-SB, $B = T^{0.7}$	0.054	0.409	0.056	0.989	0.052	0.394	0.057	0.987
τ-SB, $B = T^{0.8}$	0.040	0.445	0.045	0.997	0.039	0.420	0.046	0.996
τ-FB, $B = 0.2T$	0.049	0.390	0.049	0.991	0.047	0.378	0.051	0.989
τ-FB, $B = 0.3T$	0.051	0.428	0.049	0.996	0.050	0.410	0.050	0.995
τ-FB, $B = 0.4T$	0.053	0.443	0.050	0.997	0.051	0.421	0.051	0.996
τ-FB, $B = 0.5T$	0.053	0.452	0.050	0.997	0.052	0.422	0.051	0.997
ADF	0.054	0.310	0.052	0.995	0.053	0.334	0.052	0.996
DF-GLS	0.078	0.661	0.058	1.000	0.076	0.490	0.059	0.931
DF-GLS-trend	0.069	0.294	0.053	0.993	0.069	0.255	0.054	0.945
EL	0.061	0.117	0.054	0.761	0.061	0.113	0.054	0.736
AR(1) errors – fixed lag augmentation (p=1)								
τ-SB$^{\text{PW}}$, $B = T^{0.5}$	0.010	0.309	0.020	0.805	0.008	0.331	0.018	0.819
τ-SB$^{\text{PW}}$, $B = T^{0.6}$	0.021	0.335	0.036	0.908	0.017	0.357	0.033	0.913
τ-SB$^{\text{PW}}$, $B = T^{0.7}$	0.031	0.368	0.044	0.963	0.027	0.387	0.043	0.964
τ-SB$^{\text{PW}}$, $B = T^{0.8}$	0.026	0.381	0.040	0.981	0.021	0.408	0.038	0.982
τ-FB$^{\text{PW}}$, $B = 0.2T$	0.024	0.355	0.040	0.968	0.020	0.374	0.038	0.969
τ-FB$^{\text{PW}}$, $B = 0.3T$	0.032	0.375	0.043	0.980	0.028	0.393	0.042	0.981
τ-FB$^{\text{PW}}$, $B = 0.4T$	0.037	0.379	0.045	0.983	0.031	0.403	0.044	0.985
τ-FB$^{\text{PW}}$, $B = 0.5T$	0.039	0.379	0.046	0.985	0.033	0.407	0.044	0.986
ADF	0.056	0.242	0.051	0.969	0.055	0.244	0.052	0.969
DF-GLS	0.077	0.589	0.058	1.000	0.079	0.522	0.059	0.990
DF-GLS-trend	0.071	0.239	0.052	0.968	0.069	0.229	0.053	0.951
EL	0.067	0.095	0.056	0.609	0.067	0.094	0.057	0.595
AR(1) errors – flexible lag augmentation where p is determined by BIC								
τ-SB$^{\text{PW}}$, $B = T^{0.5}$	0.004	0.357	0.015	0.837	0.003	0.397	0.013	0.850
τ-SB$^{\text{PW}}$, $B = T^{0.6}$	0.015	0.366	0.032	0.913	0.011	0.395	0.029	0.921
τ-SB$^{\text{PW}}$, $B = T^{0.7}$	0.026	0.387	0.042	0.960	0.020	0.414	0.040	0.962
τ-SB$^{\text{PW}}$, $B = T^{0.8}$	0.023	0.394	0.039	0.978	0.017	0.425	0.037	0.979
τ-FB$^{\text{PW}}$, $B = 0.2T$	0.019	0.377	0.039	0.964	0.014	0.406	0.036	0.967
τ-FB$^{\text{PW}}$, $B = 0.3T$	0.028	0.389	0.043	0.976	0.023	0.415	0.040	0.978
τ-FB$^{\text{PW}}$, $B = 0.4T$	0.034	0.389	0.044	0.980	0.027	0.423	0.043	0.982
τ-FB$^{\text{PW}}$, $B = 0.5T$	0.036	0.385	0.046	0.981	0.029	0.421	0.043	0.983
ADF	0.058	0.240	0.051	0.967	0.057	0.242	0.052	0.967
DF-GLS	0.085	0.542	0.060	0.999	0.086	0.480	0.061	0.986
DF-GLS-trend	0.082	0.219	0.054	0.955	0.080	0.206	0.056	0.932
EL	0.106	0.089	0.066	0.573	0.107	0.087	0.066	0.559

Note: Simulation results are reported for 100,000 replications. The zero-trend $d_t = 0$ is considered for all $t = 1, \ldots, T$. The AR(1) process is given by $u_t = 0.5u_{t-1} + \epsilon_t$. All innovations are simulated independently as standard normal random variables. For the small-b and fixed-b tests, the lag order p refers to the pre-whitening scheme, and, for the conventional tests, p is equal to the augmentation order. The random initial condition is simulated from T lagged innovations. For $\rho = 1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho = 0.9$, the values are size-adjusted.

Table 1.4: Size and size-adjusted powers under different trends and i.i.d. errors (1/2)

	$T=100,\ \rho=1$		$T=100,\ \rho=0.9$			$T=300,\ \rho=1$		$T=300,\ \rho=0.9$		
	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=6$	$\lambda=9$	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=6$	$\lambda=9$
sharp break										
τ-SB, $B=T^{0.5}$	0.056	0.039	0.248	0.162	0.104	0.056	0.053	0.816	0.725	0.586
τ-SB, $B=T^{0.6}$	0.056	0.039	0.281	0.159	0.085	0.058	0.054	0.928	0.834	0.656
τ-SB, $B=T^{0.7}$	0.053	0.038	0.296	0.124	0.043	0.056	0.054	0.970	0.856	0.566
τ-SB, $B=T^{0.8}$	0.044	0.048	0.281	0.075	0.014	0.047	0.052	0.969	0.688	0.173
τ-FB, $B=0.2T$	0.051	0.042	0.293	0.142	0.061	0.051	0.054	0.972	0.845	0.511
τ-FB, $B=0.3T$	0.056	0.056	0.284	0.093	0.024	0.053	0.063	0.970	0.720	0.224
τ-FB, $B=0.4T$	0.062	0.081	0.280	0.077	0.013	0.054	0.074	0.968	0.666	0.146
τ-FB, $B=0.5T$	0.062	0.085	0.292	0.088	0.016	0.053	0.073	0.972	0.690	0.162
ADF	0.050	0.023	0.158	0.027	0.002	0.051	0.040	0.887	0.255	0.006
DF-GLS	0.078	0.065	0.364	0.056	0.003	0.058	0.058	0.983	0.630	0.063
DF-GLS-trend	0.069	0.057	0.238	0.134	0.060	0.053	0.053	0.965	0.793	0.414
EL	0.060	0.044	0.110	0.088	0.071	0.053	0.050	0.710	0.569	0.397
u-shaped break										
τ-SB, $B=T^{0.5}$	0.055	0.022	0.213	0.113	0.069	0.057	0.046	0.785	0.615	0.413
τ-SB, $B=T^{0.6}$	0.056	0.022	0.231	0.098	0.049	0.058	0.047	0.898	0.694	0.397
τ-SB, $B=T^{0.7}$	0.057	0.025	0.209	0.052	0.016	0.055	0.048	0.945	0.630	0.197
τ-SB, $B=T^{0.8}$	0.040	0.018	0.252	0.065	0.017	0.046	0.044	0.941	0.443	0.046
τ-FB, $B=0.2T$	0.053	0.026	0.232	0.080	0.031	0.050	0.050	0.946	0.587	0.147
τ-FB, $B=0.3T$	0.056	0.033	0.231	0.059	0.017	0.051	0.054	0.940	0.442	0.048
τ-FB, $B=0.4T$	0.052	0.025	0.252	0.066	0.018	0.050	0.046	0.929	0.418	0.040
τ-FB, $B=0.5T$	0.045	0.010	0.243	0.061	0.018	0.047	0.031	0.922	0.377	0.032
ADF	0.046	0.011	0.178	0.056	0.019	0.049	0.030	0.911	0.381	0.045
DF-GLS	0.077	0.049	0.374	0.092	0.021	0.058	0.056	0.985	0.658	0.105
DF-GLS-trend	0.063	0.016	0.126	0.018	0.003	0.050	0.036	0.819	0.133	0.002
EL	0.064	0.053	0.108	0.090	0.076	0.055	0.057	0.703	0.547	0.377
continuous break										
τ-SB, $B=T^{0.5}$	0.049	0.015	0.264	0.188	0.115	0.053	0.037	0.839	0.818	0.782
τ-SB, $B=T^{0.6}$	0.049	0.014	0.299	0.185	0.093	0.054	0.036	0.944	0.916	0.853
τ-SB, $B=T^{0.7}$	0.046	0.012	0.327	0.165	0.063	0.053	0.035	0.981	0.937	0.783
τ-SB, $B=T^{0.8}$	0.035	0.011	0.355	0.175	0.059	0.043	0.029	0.989	0.919	0.640
τ-FB, $B=0.2T$	0.042	0.011	0.315	0.172	0.074	0.046	0.030	0.983	0.932	0.745
τ-FB, $B=0.3T$	0.043	0.012	0.331	0.160	0.055	0.046	0.031	0.988	0.914	0.631
τ-FB, $B=0.4T$	0.045	0.015	0.351	0.174	0.058	0.047	0.032	0.991	0.934	0.685
τ-FB, $B=0.5T$	0.046	0.016	0.360	0.184	0.062	0.047	0.033	0.992	0.941	0.704
ADF	0.046	0.011	0.149	0.019	0.001	0.047	0.029	0.891	0.273	0.007
DF-GLS	0.064	0.016	0.381	0.064	0.004	0.055	0.035	0.984	0.663	0.088
DF-GLS-trend	0.061	0.022	0.215	0.091	0.026	0.050	0.036	0.956	0.712	0.260
EL	0.059	0.048	0.116	0.111	0.104	0.053	0.049	0.754	0.725	0.678
u-shaped break in intercept										
τ-SB, $B=T^{0.5}$	0.053	0.018	0.209	0.108	0.063	0.056	0.043	0.784	0.607	0.403
τ-SB, $B=T^{0.6}$	0.054	0.018	0.221	0.088	0.042	0.057	0.043	0.897	0.680	0.373
τ-SB, $B=T^{0.7}$	0.055	0.019	0.194	0.042	0.012	0.054	0.043	0.940	0.592	0.158
τ-SB, $B=T^{0.8}$	0.039	0.015	0.238	0.054	0.013	0.045	0.040	0.933	0.381	0.027
τ-FB, $B=0.2T$	0.052	0.021	0.214	0.065	0.025	0.049	0.045	0.941	0.538	0.108
τ-FB, $B=0.3T$	0.054	0.026	0.214	0.046	0.012	0.051	0.050	0.931	0.382	0.027
τ-FB, $B=0.4T$	0.051	0.020	0.238	0.054	0.012	0.049	0.042	0.923	0.364	0.024
τ-FB, $B=0.5T$	0.043	0.008	0.237	0.057	0.014	0.047	0.029	0.919	0.340	0.022
ADF	0.043	0.007	0.118	0.012	0.001	0.047	0.026	0.785	0.075	0.000
DF-GLS	0.075	0.037	0.345	0.067	0.011	0.058	0.049	0.990	0.681	0.098
DF-GLS-trend	0.063	0.016	0.126	0.018	0.003	0.050	0.036	0.819	0.133	0.002
EL	0.064	0.053	0.108	0.090	0.076	0.055	0.057	0.703	0.547	0.377

Note: Simulation results are reported for 100,000 replications. The errors u_t are simulated independently as standard normal random variables. The series are not pre-whitened. For $\rho=1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho=0.9$, the values are size-adjusted.

Table 1.5: Size and size-adjusted powers under different trends and i.i.d. errors (2/2)

	$T=100,\ \rho=1$		$T=100,\ \rho=0.9$			$T=300,\ \rho=1$		$T=300,\ \rho=0.9$		
	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=6$	$\lambda=9$	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=6$	$\lambda=9$
LSTAR break										
τ-SB, $B=T^{0.5}$	0.052	0.022	0.269	0.211	0.145	0.055	0.043	0.840	0.826	0.800
τ-SB, $B=T^{0.6}$	0.051	0.020	0.308	0.209	0.115	0.056	0.042	0.945	0.926	0.883
τ-SB, $B=T^{0.7}$	0.047	0.017	0.332	0.171	0.063	0.053	0.038	0.983	0.949	0.835
τ-SB, $B=T^{0.8}$	0.036	0.014	0.348	0.161	0.048	0.044	0.032	0.988	0.907	0.598
τ-FB, $B=0.2T$	0.043	0.016	0.325	0.189	0.086	0.048	0.034	0.984	0.944	0.795
τ-FB, $B=0.3T$	0.045	0.016	0.334	0.160	0.053	0.047	0.034	0.987	0.908	0.612
τ-FB, $B=0.4T$	0.047	0.018	0.346	0.163	0.048	0.048	0.036	0.988	0.904	0.571
τ-FB, $B=0.5T$	0.048	0.020	0.355	0.171	0.051	0.049	0.037	0.989	0.907	0.574
ADF	0.049	0.019	0.179	0.037	0.004	0.050	0.037	0.926	0.416	0.030
DF-GLS	0.070	0.029	0.425	0.104	0.010	0.055	0.042	0.990	0.797	0.215
DF-GLS-trend	0.063	0.033	0.248	0.147	0.068	0.052	0.040	0.972	0.854	0.549
EL	0.059	0.046	0.115	0.112	0.106	0.053	0.050	0.751	0.716	0.666
offsetting LSTAR break										
τ-SB, $B=T^{0.5}$	0.050	0.018	0.267	0.196	0.125	0.053	0.040	0.841	0.822	0.786
τ-SB, $B=T^{0.6}$	0.050	0.016	0.297	0.188	0.097	0.055	0.038	0.945	0.919	0.859
τ-SB, $B=T^{0.7}$	0.046	0.014	0.324	0.165	0.064	0.053	0.035	0.981	0.935	0.770
τ-SB, $B=T^{0.8}$	0.035	0.011	0.338	0.153	0.053	0.043	0.029	0.985	0.877	0.506
τ-FB, $B=0.2T$	0.042	0.013	0.319	0.179	0.078	0.046	0.031	0.983	0.928	0.727
τ-FB, $B=0.3T$	0.044	0.014	0.334	0.159	0.059	0.047	0.031	0.985	0.882	0.530
τ-FB, $B=0.4T$	0.045	0.014	0.336	0.154	0.054	0.047	0.032	0.980	0.832	0.421
τ-FB, $B=0.5T$	0.046	0.014	0.326	0.137	0.043	0.047	0.032	0.979	0.798	0.336
ADF	0.052	0.048	0.239	0.121	0.046	0.050	0.044	0.978	0.838	0.476
DF-GLS	0.068	0.023	0.385	0.079	0.008	0.055	0.039	0.969	0.532	0.049
DF-GLS-trend	0.061	0.018	0.191	0.061	0.011	0.049	0.033	0.932	0.532	0.088
EL	0.060	0.050	0.116	0.114	0.109	0.053	0.050	0.755	0.736	0.704
triangular break										
τ-SB, $B=T^{0.5}$	0.051	0.020	0.267	0.204	0.136	0.055	0.042	0.840	0.822	0.793
τ-SB, $B=T^{0.6}$	0.050	0.019	0.308	0.205	0.114	0.056	0.041	0.945	0.924	0.879
τ-SB, $B=T^{0.7}$	0.047	0.016	0.339	0.191	0.083	0.054	0.039	0.983	0.951	0.829
τ-SB, $B=T^{0.8}$	0.036	0.012	0.346	0.172	0.065	0.045	0.032	0.983	0.871	0.526
τ-FB, $B=0.2T$	0.043	0.015	0.331	0.200	0.097	0.048	0.034	0.984	0.945	0.788
τ-FB, $B=0.3T$	0.046	0.016	0.343	0.184	0.075	0.048	0.035	0.983	0.880	0.555
τ-FB, $B=0.4T$	0.046	0.016	0.347	0.171	0.065	0.048	0.034	0.977	0.817	0.418
τ-FB, $B=0.5T$	0.047	0.016	0.349	0.170	0.063	0.048	0.034	0.982	0.838	0.441
ADF	0.052	0.043	0.231	0.098	0.025	0.051	0.046	0.971	0.768	0.331
DF-GLS	0.069	0.023	0.470	0.179	0.051	0.056	0.039	0.990	0.823	0.326
DF-GLS-trend	0.059	0.018	0.193	0.053	0.009	0.051	0.033	0.919	0.467	0.057
EL	0.060	0.055	0.116	0.113	0.113	0.053	0.050	0.760	0.755	0.746
Fourier break										
τ-SB, $B=T^{0.5}$	0.048	0.014	0.258	0.176	0.098	0.053	0.037	0.841	0.820	0.779
τ-SB, $B=T^{0.6}$	0.048	0.012	0.284	0.160	0.071	0.054	0.035	0.944	0.914	0.839
τ-SB, $B=T^{0.7}$	0.044	0.010	0.301	0.130	0.043	0.052	0.032	0.980	0.914	0.658
τ-SB, $B=T^{0.8}$	0.033	0.007	0.310	0.115	0.033	0.042	0.026	0.974	0.743	0.250
τ-FB, $B=0.2T$	0.041	0.009	0.299	0.144	0.054	0.046	0.028	0.981	0.895	0.578
τ-FB, $B=0.3T$	0.042	0.009	0.304	0.120	0.037	0.046	0.027	0.974	0.762	0.279
τ-FB, $B=0.4T$	0.043	0.009	0.309	0.114	0.033	0.046	0.028	0.963	0.673	0.185
τ-FB, $B=0.5T$	0.044	0.009	0.308	0.113	0.032	0.046	0.028	0.968	0.690	0.187
ADF	0.049	0.027	0.203	0.065	0.010	0.049	0.037	0.955	0.623	0.140
DF-GLS	0.064	0.015	0.430	0.140	0.041	0.054	0.034	0.989	0.784	0.235
DF-GLS-trend	0.058	0.011	0.160	0.032	0.004	0.049	0.028	0.891	0.318	0.016
EL	0.061	0.061	0.117	0.117	0.117	0.054	0.054	0.761	0.761	0.761

Note: Simulation results are reported for 100,000 replications. The errors u_t are simulated independently as standard normal random variables. The series are not pre-whitened. For $\rho=1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho=0.9$, the values are size-adjusted.

Table 1.6: Size and size-adjusted powers under different trends and AR(1) errors

	$T=100$, $\rho=1$		$T=100$, $\rho=0.9$		$T=300$, $\rho=1$		$T=300$, $\rho=0.9$	
	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=9$	$\lambda=3$	$\lambda=9$
sharp break								
τ-SB$^{\text{PW}}$, $B=T^{0.5}$	0.004	0.004	0.296	0.150	0.015	0.011	0.797	0.643
τ-SB$^{\text{PW}}$, $B=T^{0.6}$	0.015	0.013	0.310	0.152	0.031	0.028	0.888	0.759
τ-SB$^{\text{PW}}$, $B=T^{0.7}$	0.027	0.027	0.323	0.131	0.042	0.041	0.943	0.798
τ-SB$^{\text{PW}}$, $B=T^{0.8}$	0.026	0.038	0.320	0.092	0.040	0.046	0.958	0.674
τ-FB$^{\text{PW}}$, $B=0.2T$	0.021	0.024	0.320	0.143	0.039	0.043	0.947	0.787
τ-FB$^{\text{PW}}$, $B=0.3T$	0.035	0.051	0.322	0.109	0.045	0.059	0.956	0.692
τ-FB$^{\text{PW}}$, $B=0.4T$	0.042	0.075	0.319	0.094	0.047	0.064	0.958	0.657
τ-FB$^{\text{PW}}$, $B=0.5T$	0.043	0.075	0.323	0.109	0.048	0.061	0.960	0.672
u-shaped break								
τ-SB$^{\text{PW}}$, $B=T^{0.5}$	0.005	0.004	0.304	0.149	0.014	0.010	0.801	0.600
τ-SB$^{\text{PW}}$, $B=T^{0.6}$	0.015	0.013	0.310	0.134	0.031	0.026	0.886	0.676
τ-SB$^{\text{PW}}$, $B=T^{0.7}$	0.030	0.033	0.308	0.084	0.042	0.041	0.936	0.646
τ-SB$^{\text{PW}}$, $B=T^{0.8}$	0.022	0.017	0.331	0.112	0.039	0.038	0.948	0.521
τ-FB$^{\text{PW}}$, $B=0.2T$	0.023	0.026	0.313	0.116	0.039	0.046	0.939	0.616
τ-FB$^{\text{PW}}$, $B=0.3T$	0.032	0.036	0.321	0.099	0.043	0.050	0.947	0.517
τ-FB$^{\text{PW}}$, $B=0.4T$	0.032	0.025	0.329	0.113	0.043	0.039	0.946	0.504
τ-FB$^{\text{PW}}$, $B=0.5T$	0.029	0.011	0.326	0.105	0.042	0.031	0.947	0.474
continuous break								
τ-SB$^{\text{PW}}$, $B=T^{0.5}$	0.004	0.003	0.332	0.180	0.015	0.013	0.825	0.737
τ-SB$^{\text{PW}}$, $B=T^{0.6}$	0.014	0.010	0.340	0.182	0.031	0.029	0.904	0.830
τ-SB$^{\text{PW}}$, $B=T^{0.7}$	0.025	0.018	0.350	0.170	0.042	0.038	0.952	0.862
τ-SB$^{\text{PW}}$, $B=T^{0.8}$	0.022	0.016	0.359	0.174	0.038	0.036	0.970	0.847
τ-FB$^{\text{PW}}$, $B=0.2T$	0.018	0.013	0.348	0.176	0.038	0.034	0.956	0.858
τ-FB$^{\text{PW}}$, $B=0.3T$	0.027	0.020	0.355	0.164	0.042	0.039	0.968	0.841
τ-FB$^{\text{PW}}$, $B=0.4T$	0.032	0.025	0.357	0.174	0.044	0.041	0.972	0.859
τ-FB$^{\text{PW}}$, $B=0.5T$	0.034	0.027	0.357	0.181	0.046	0.042	0.973	0.868
LSTAR break								
τ-SB$^{\text{PW}}$, $B=T^{0.5}$	0.004	0.002	0.292	0.145	0.015	0.013	0.797	0.669
τ-SB$^{\text{PW}}$, $B=T^{0.6}$	0.013	0.007	0.305	0.157	0.031	0.028	0.893	0.803
τ-SB$^{\text{PW}}$, $B=T^{0.7}$	0.024	0.014	0.323	0.142	0.041	0.037	0.948	0.854
τ-SB$^{\text{PW}}$, $B=T^{0.8}$	0.020	0.012	0.333	0.136	0.039	0.034	0.965	0.816
τ-FB$^{\text{PW}}$, $B=0.2T$	0.017	0.010	0.320	0.153	0.038	0.033	0.952	0.847
τ-FB$^{\text{PW}}$, $B=0.3T$	0.026	0.015	0.328	0.136	0.042	0.037	0.964	0.817
τ-FB$^{\text{PW}}$, $B=0.4T$	0.031	0.019	0.331	0.138	0.044	0.040	0.966	0.813
τ-FB$^{\text{PW}}$, $B=0.5T$	0.034	0.021	0.331	0.144	0.045	0.041	0.968	0.810
Fourier break								
τ-SB$^{\text{PW}}$, $B=T^{0.5}$	0.004	0.003	0.335	0.217	0.015	0.013	0.826	0.755
τ-SB$^{\text{PW}}$, $B=T^{0.6}$	0.013	0.009	0.342	0.205	0.032	0.028	0.905	0.839
τ-SB$^{\text{PW}}$, $B=T^{0.7}$	0.025	0.015	0.349	0.180	0.041	0.036	0.951	0.846
τ-SB$^{\text{PW}}$, $B=T^{0.8}$	0.021	0.013	0.353	0.171	0.038	0.034	0.961	0.729
τ-FB$^{\text{PW}}$, $B=0.2T$	0.018	0.011	0.348	0.192	0.038	0.033	0.954	0.830
τ-FB$^{\text{PW}}$, $B=0.3T$	0.027	0.017	0.352	0.173	0.042	0.037	0.960	0.738
τ-FB$^{\text{PW}}$, $B=0.4T$	0.032	0.020	0.349	0.170	0.044	0.039	0.959	0.681
τ-FB$^{\text{PW}}$, $B=0.5T$	0.034	0.021	0.348	0.167	0.045	0.039	0.962	0.695

Note: Simulation results are reported for 100,000 replications. The errors u_t are simulated from $u_t = 0.5u_{t-1} + \epsilon_t$ with independent standard normal innovations, and the series are pre-whitened with a lag order that is determined from the BIC. For $\rho = 1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho = 0.9$, the values are size-adjusted.

Table 1.7: Size and size-adjusted powers of robust tests under constant trend and variance

	iid errors, $p = 0$				AR(1) errors, p is chosen by BIC			
	$T = 100$		$T = 300$		$T = 100$		$T = 300$	
	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$	$\rho = 1$	$\rho = 0.9$
τ-SB$^{\text{H}}$, $B = T^{0.5}$	0.063	0.298	0.057	0.849	0.006	0.356	0.016	0.838
τ-SB$^{\text{H}}$, $B = T^{0.6}$	0.064	0.355	0.059	0.953	0.018	0.365	0.033	0.913
τ-SB$^{\text{H}}$, $B = T^{0.7}$	0.062	0.408	0.058	0.988	0.032	0.381	0.044	0.960
τ-SB$^{\text{H}}$, $B = T^{0.8}$	0.049	0.434	0.048	0.996	0.032	0.380	0.042	0.976
τ-FB$^{\text{H}}$, $B = 0.2T$	0.044	0.348	0.046	0.979	0.020	0.315	0.037	0.942
τ-FB$^{\text{H}}$, $B = 0.3T$	0.046	0.386	0.047	0.989	0.028	0.331	0.042	0.960
τ-FB$^{\text{H}}$, $B = 0.4T$	0.047	0.400	0.048	0.992	0.033	0.337	0.043	0.966
τ-FB$^{\text{H}}$, $B = 0.5T$	0.049	0.402	0.048	0.993	0.034	0.337	0.044	0.970

Note: Simulation results are reported for 100,000 replications. All innovations are simulated independently as standard normal random variables, and the initial condition is $x_0 = 0$. The AR(1) process is given by $u_t = 0.5u_{t-1} + \epsilon_t$. For $\rho = 1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho = 0.9$, the values are size-adjusted.

The power of the pooled tests depends on the blocklength. In case of no break, a larger blocklength implies higher power results, which is in line with the theoretical findings that those tests have power in a $1/\sqrt{BT}$ neighborhood of the unit root hypothesis. For blocklengths of $B = T^{0.8}$ in the small-b case and $B = 0.5T$ in the fixed-b case, the power results are higher than for the ADF test and also larger than those obtained when performing the Dickey-Fuller GLS test under a random initial condition. Hence, we do not lose power under these small-sample specifications (although, asymptotically, those tests have power in a $1/T$ neighborhood of the unit root hypothesis). Furthermore, smaller sample sizes, such as $T^{0.6}$ in the small-b context and $0.3T$ in the fixed-b context, still yield reasonably high power. In particular, the EL test performs much worse in all cases. The size and power results obtained under the AR(1) error specification with both fixed and flexible lag augmentation for the pre-whitening scheme are similar to those produced by i.i.d. errors.

As the tests are designed to yield higher power in the presence of slowly varying trends and breaks, we compare the size-adjusted powers of the tests under the trend specifications presented in Table 1.2 and Figure 1.1. For large break sizes λ, it is shown that the smaller the blocklength, the greater the power results. In most cases, the pooled tests have greater power than the ADF, the DF-GLS, the DF-GLS-trend, and the EL test. Furthermore, the power results of the pooled tests are quite uniform across different trend specifications when compared to those of the conventional tests.

Table 1.6 shows that the pooled tests have reasonable size and power properties under the presence of AR(1) errors and different trend specifications. Furthermore, from Tables 1.7 and 1.8, we can conclude that the heteroskedasticity-robust tests are sized correctly and have good power properties in the presence of a break in the variance and in the trend function.

Table 1.8: Size and size-adjusted power of robust tests under breaks in trend and variance

	$T=100$				$T=300$			
	$\rho=1$		$\rho=0.9$		$\rho=1$		$\rho=0.9$	
	$\lambda=2$	$\lambda=3$	$\lambda=2$	$\lambda=3$	$\lambda=2$	$\lambda=3$	$\lambda=2$	$\lambda=3$
sharp break in variance								
τ-SB$^{\mathrm{H}}$, $B=T^{0.5}$	0.066	0.068	0.301	0.291	0.058	0.058	0.829	0.788
τ-SB$^{\mathrm{H}}$, $B=T^{0.6}$	0.072	0.075	0.358	0.347	0.062	0.062	0.941	0.916
τ-SB$^{\mathrm{H}}$, $B=T^{0.7}$	0.082	0.096	0.414	0.406	0.068	0.073	0.986	0.977
τ-SB$^{\mathrm{H}}$, $B=T^{0.8}$	0.085	0.124	0.443	0.433	0.082	0.116	0.997	0.996
τ-FB$^{\mathrm{H}}$, $B=0.2T$	0.043	0.042	0.346	0.318	0.045	0.045	0.958	0.900
τ-FB$^{\mathrm{H}}$, $B=0.3T$	0.044	0.043	0.384	0.350	0.047	0.045	0.977	0.930
τ-FB$^{\mathrm{H}}$, $B=0.4T$	0.044	0.043	0.408	0.366	0.048	0.046	0.987	0.944
τ-FB$^{\mathrm{H}}$, $B=0.5T$	0.043	0.042	0.417	0.379	0.046	0.047	0.993	0.968
sharp break in trend								
τ-SB$^{\mathrm{H}}$, $B=T^{0.5}$	0.063	0.063	0.275	0.251	0.058	0.058	0.837	0.820
τ-SB$^{\mathrm{H}}$, $B=T^{0.6}$	0.063	0.064	0.322	0.281	0.060	0.060	0.942	0.928
τ-SB$^{\mathrm{H}}$, $B=T^{0.7}$	0.061	0.061	0.353	0.294	0.058	0.058	0.981	0.969
τ-SB$^{\mathrm{H}}$, $B=T^{0.8}$	0.058	0.068	0.347	0.267	0.049	0.050	0.988	0.967
τ-FB$^{\mathrm{H}}$, $B=0.2T$	0.043	0.043	0.308	0.270	0.046	0.046	0.967	0.950
τ-FB$^{\mathrm{H}}$, $B=0.3T$	0.045	0.045	0.320	0.264	0.048	0.048	0.974	0.949
τ-FB$^{\mathrm{H}}$, $B=0.4T$	0.047	0.045	0.332	0.270	0.048	0.048	0.978	0.951
τ-FB$^{\mathrm{H}}$, $B=0.5T$	0.047	0.044	0.341	0.286	0.048	0.048	0.981	0.957
sharp break in trend and variance								
τ-SB$^{\mathrm{H}}$, $B=T^{0.5}$	0.067	0.068	0.283	0.256	0.059	0.058	0.816	0.767
τ-SB$^{\mathrm{H}}$, $B=T^{0.6}$	0.070	0.073	0.331	0.297	0.062	0.062	0.933	0.897
τ-SB$^{\mathrm{H}}$, $B=T^{0.7}$	0.080	0.091	0.372	0.334	0.068	0.072	0.979	0.959
τ-SB$^{\mathrm{H}}$, $B=T^{0.8}$	0.095	0.144	0.368	0.315	0.081	0.112	0.989	0.970
τ-FB$^{\mathrm{H}}$, $B=0.2T$	0.043	0.042	0.316	0.278	0.046	0.045	0.947	0.875
τ-FB$^{\mathrm{H}}$, $B=0.3T$	0.044	0.042	0.341	0.287	0.048	0.046	0.962	0.893
τ-FB$^{\mathrm{H}}$, $B=0.4T$	0.044	0.042	0.350	0.290	0.048	0.047	0.967	0.889
τ-FB$^{\mathrm{H}}$, $B=0.5T$	0.044	0.044	0.351	0.281	0.048	0.048	0.964	0.886

Note: Simulation results are reported for 100,000 replications. The errors u_t are simulated independently as standard normal random variables, and the series are not pre-whitened. The sharp break specification is defined by a break in the variance at 2/3 of the sample. For $\rho=1$, the rejection frequencies are based on the asymptotic critical values for a significance level of 5%, while, for $\rho=0.9$, the values are size-adjusted.

The blocklength B is a tuning parameter that needs to be chosen carefully, and any optimality result would depend on the actual trend model. In practice, however, the trend model is unknown, which makes it hard to derive an optimal blocklength. Although theoretical recommendations cannot be formulated based on the current analysis, the small-b tests with $B=T^{0.7}$ and the fixed-b tests with $T=0.2B$ yield very promising results for all trend functions studied in this paper and are therefore recommended as the default settings. Moreover, the empirical power of τ-SB does not exceed the empirical power of its heteroskedasticity-robust counterpart in the case of homoskedasticity. Hence, the τ-SB$^{\mathrm{H}}$ statistic can always be used in favor of τ-SB. The fixed-b statistic τ-FB has slightly better power results than τ-FB^{H} in the simulations.

1.7 Empirical illustrations

In order to illustrate the application of the test procedures, we apply the unit root tests to monthly annualized growth rates of the consumer price index. For monetary policy it is crucial to know whether shocks to the inflation rate will have a permanent or transitory effect. From an econometric point of view, the integrational properties of inflation rates affect the choice of an appropriate model. It has been widely debated in the literature whether inflation rates are $I(1)$ or $I(0)$. Early studies, such as that of MacDonald and Murphy (1989), have shown that conventional Dickey-Fuller tests are often unable to reject the unit root hypothesis for quarterly inflation rates. Evans and Lewis (1995), as well as Ng and Perron (2001), also find strong evidence that inflation rates are nonstationary. The work of Hassler and Wolters (1995), in which the authors used ARFIMA models for monthly data, produces mixed results. In contrast, Rose (1988) finds that for 18 countries, quarterly inflation rates are stationary. Using panel unit root tests, Lee and Wu (2001) provide evidence that the inflation rates of 13 OECD countries do not contain a unit root. Allowing for multiple breaks, Narayan and Narayan (2010) also find strong evidence for stationarity.

Our dataset includes 25 countries with 576 observations covering the period from 1971:1 to 2018:12.[2] The series is pre-whitened, where p is determined by the BIC using a maximal lag order of $p^* = \lfloor 4 \cdot (T/100)^{1/5} \rfloor$. The small-$b$ and fixed-b tests are then applied using different blocklengths. The results of both the pooled tests and some benchmark tests are presented in Table 1.9. For the conventional statistics, the augmented versions are applied with the same value for p.

At a 10% significance level, the τ-SB test with $B = T^{0.7}$ rejects the unit root hypothesis for 17 of 25 countries, and the τ-FB test with $B = 0.2T$ does so for 14 countries, whereas DF-GLS rejects H_0 for 12 countries, and ADF rejects H_0 only for five countries in the dataset. Tests that are more robust to changes in the deterministic component reject the hypothesis of a unit root more frequently than tests that assume a constant trend. Hence, inflation rates might be stationary around a slowly varying trend that reflects different regimes of monetary policy rather than a constant trend.

1.8 Conclusion

We have presented two variants of a unit root test under an unknown trend specification that are robust under both heteroskedasticity and autocorrelation. When applied to finite

[2]Source: https://data.oecd.org/

Table 1.9: Unit root tests applied to inflation rates

	τ-SB$^{\text{PW}}$ $B = T^{0.6}$	τ-SB$^{\text{PW}}$ $B = T^{0.7}$	τ-FB$^{\text{PW}}$ $B = 0.2T$	τ-FB$^{\text{PW}}$ $B = 0.3T$	ADF	DF-GLS	EL
AUT	−0.37	−1.36*	−1.12	−0.76	−1.81	−1.60	−2.84
BEL	−1.62*	−2.52***	−1.95***	−1.51**	−2.67*	−2.67***	−3.67
CAN	−0.13	−0.96	−0.58	−0.55	−1.72	−1.24	−2.41
CHE	−0.23	−1.48*	−1.37*	−0.96	−2.38	−1.14	−3.19
DEU	0.13	−0.93	−1.05	−0.82	−2.03	−1.55	−2.80
DNK	−2.31**	−2.12**	−1.15*	−0.73	−1.95	−1.56	−3.92*
ESP	−0.12	−0.37	−0.06	0.07	−1.19	−1.15	−2.45
FIN	0.13	−1.14	−0.57	−0.33	−1.46	−1.47	−2.44
FRA	0.04	−0.37	−0.16	−0.15	−1.24	−1.20	−2.31
GBR	−1.53*	−1.92**	−1.46**	−1.03	−2.32	−1.96**	−3.82*
GRC	−2.12**	−1.96**	−1.15*	−0.78	−2.13	−1.90*	−4.44**
IDN	−3.19***	−3.12***	−2.18***	−1.99***	−5.49***	−4.75***	−5.68***
IND	−3.11***	−3.47***	−2.73***	−2.46***	−6.03***	−4.76***	−6.05***
ITA	−0.95	−1.13	−0.60	−0.38	−1.56	−1.57	−3.24
JPN	−1.13	−1.31*	−0.73	−0.58	−2.13	−1.46	−3.20
KOR	−2.02**	−2.11**	−1.57**	−1.31*	−2.89**	−1.78*	−4.56**
LUX	1.42	−0.99	−0.71	−0.52	−1.85	−1.80*	−2.60
MEX	−2.05**	−1.93**	−1.76**	−1.71**	−3.14**	−2.75***	−4.21**
NLD	0.66	−0.26	−0.24	−0.02	−1.55	−0.76	−2.28
NOR	−1.39*	−1.99**	−1.27*	−0.96	−2.10	−1.53	−4.07**
PRT	−1.70**	−2.04**	−0.93	−0.47	−1.83	−1.82*	−3.75
SWE	−1.61*	−1.74**	−1.40**	−1.04	−2.04	−1.18	−4.38**
TUR	−1.76**	−1.45*	−1.29*	−1.11*	−2.34	−1.71*	−3.74
USA	−1.06	−1.57*	−1.38**	−1.24*	−2.37	−2.13**	−3.42
ZAF	−2.37***	−2.25**	−1.52**	−1.18*	−2.26	−2.11**	−5.03***

Note: Unit root test statistics for 25 countries using data from 1971:1 until 2018:12 are reported. ADF is the augmented Dickey-Fuller test with constant trend specification, DFGLS is the test developed by Elliott et al. (1996) with constant trend specification, and EL is the test by Enders and Lee (2012). The asterisks *, **, and *** denote the significance at the 10%, 5%, and 1% level, respectively. The lag order p is determined by the BIC.

samples, the tests show good size properties. The fixed-b pooled test statistic converges to a functional of a Brownian motion under the unit root hypothesis, while the small-b variant shows a standard normal distribution in the limit. Both heteroskedasticity- and autocorrelation-robust versions of the tests were introduced. Monte Carlo simulations indicate that, while under the zero-trend specification, the fixed-b and small-b tests perform similar to the conventional tests in terms of size and power, under sharp breaks as well as smooth changes in the trend, their power is much higher. In terms of power, the small-b tests with a blocklength of $B = T^{0.7}$ and the fixed-b tests with $B = 0.2 \cdot T$ perform well for moderately varying trends.

From the point of view of a practitioner, the τ-SB^H test has a number of advantages: First, the distribution is standard normal; thus, there is no need to resort to new tables, and p-values are easy to implement. Second, for the sample sizes used in the Monte Carlo simulation, the power tends to be higher than for conventional unit root tests under many trend specifications. Finally, the test is robust to heteroskedasticity.

Appendix to Chapter 1

Proof of Lemma 1.1

Since $d(r)$ is piecewise Lipschitz continuous on the unit interval, there are a finite number of points where $d(r)$ is not continuous. Let those points be given by $\{\pi_1, \ldots, \pi_L\}$, where $0 < \pi_1 < \ldots < \pi_L < 1$, and $L < \infty$. We can represent $d(r)$ by some function $\delta(r)$ that is Lipschitz continuous on the entire domain. Then, for any $r \in [0,1]$, we obtain $d(r) = \delta(r) + \sum_{l=1}^{L} \lambda_l 1_{\{r \geq \pi_l\}}$, with $\sum_{l=1}^{L} |\lambda_l| < \infty$. Let $t_l = \lfloor \pi_l T \rfloor$ for $l = 1, \ldots, L$, and let $\delta_t = \delta(t/T)$ for $t = 1, \ldots, T$. Then, $d_t = \delta_t + \sum_{l=1}^{L} \lambda_l 1_{\{t \leq t_l\}}$, and consequently,

$$\Delta d_{t+j} = \Delta\delta_t + \sum_{l=1}^{L} \lambda_l 1_{\{t+j=t_l\}}, \quad d_{t+j-1} - d_j = \delta_{t+j-1} - \delta_j + \sum_{k=1}^{t-1} \sum_{l=1}^{L} \lambda_l 1_{\{k+j=t_l\}}.$$

Due to the Lipschitz continuity of $\delta(r)$, there exists a constant $C_1 < \infty$, such that

$$|\Delta d_{t+j}| \leq C_1 T^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j=t_l\}}, \quad |d_{t+j-1} - d_j| \leq C_1 B T^{-1} + \sum_{k=1}^{t-1} \sum_{l=1}^{L} |\lambda_l| 1_{\{k+j=t_l\}}.$$

There also exists a constant $C_2 < \infty$, such that $\sum_{l=1}^{L} |\lambda_l| \leq C_2$. Furthermore, the error variance is bounded with $E[u_t^2] = \sigma^2 < \infty$ for all $t = 1, \ldots, T$. We then define the common constant $C = \max\{C_1, C_2, \sigma, 1\} < \infty$.

Rearranging the model equation yields $\Delta x_t = (\rho - 1) x_{t-1} + u_t$. For the numerator statistic $\mathcal{Y}_{1,T}$, note that

$$\begin{aligned} &\Delta y_{t+j}(y_{t+j-1} - y_j) - \Delta x_{t+j}(x_{t+j-1} - x_j) \\ &= \Delta d_{t+j}(d_{t+j-1} - d_j) + \Delta d_{t+j}(x_{t+j-1} - x_j) + \Delta x_{t+j}(d_{t+j-1} - d_j), \end{aligned}$$

such that $\mathcal{Y}_{1,T} - \mathcal{X}_{1,T} = S_1 + S_2 + S_3$, where

$$S_1 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta d_{t+j}(d_{t+j-1} - d_j)}{B^{3/2} T^{1/2}}, \quad S_2 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta d_{t+j}(x_{t+j-1} - x_j)}{B^{3/2} T^{1/2}},$$
$$S_3 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta x_{t+j}(d_{t+j-1} - d_j)}{B^{3/2} T^{1/2}}.$$

In what follows, we show that S_1, S_2, and S_3 converge to zero in probability. For the first

term, we obtain

$$|S_1| \leq \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Big(CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j=t_l\}} \Big) \Big(CBT^{-1} + \sum_{k=1}^{t-1} \sum_{l=1}^{L} |\lambda_l| 1_{\{k+j=t_l\}} \Big)$$
$$\leq \frac{4C^2 B^2 T^{-1}}{B^{3/2}T^{1/2}} = o(1).$$

For the second term, note that $x_{t+j-1} - x_j = \sum_{k=1}^{t-1} u_{k+j} + (\rho - 1)x_{k+j-1}$. With the MA-representation $x_t = \sum_{m=0}^{t-1} \rho^m u_{t-m} + x_0$, we then decompose $S_2 = S_{2,1} + S_{2,2}$, where

$$S_{2,1} = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \Delta d_{t+j} u_{k+j}}{B^{3/2}T^{1/2}},$$
$$S_{2,2} = \frac{(\rho - 1) \sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \Big(\sum_{m=0}^{k+j-2} \rho^m u_{k+j-1-m} + x_0 \Big) \Delta d_{t+j}}{B^{3/2}T^{1/2}}.$$

Jensen's inequality implies that

$$E[|S_{2,1}|] \leq \frac{\sum_{t=2}^{B} \sum_{k=1}^{t-1} E[|\sum_{j=1}^{T-B} \Delta d_{t+j} u_{k+j}|]}{B^{3/2}T^{1/2}} \leq \frac{\sum_{t=2}^{B} \sum_{k=1}^{t-1} \sqrt{E[(\sum_{j=1}^{T-B} \Delta d_{t+j} u_{k+j})^2]}}{B^{3/2}T^{1/2}}$$
$$\leq \frac{C \sum_{t=2}^{B} \sum_{k=1}^{t-1} \sqrt{\sum_{j=1}^{T-B} |\Delta d_{t+j}|^2}}{B^{3/2}T^{1/2}} \leq \frac{C \sum_{t=2}^{B} \sqrt{\sum_{j=1}^{T-B} \Big(CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j=t_l\}} \Big)^2}}{B^{1/2}T^{1/2}}$$
$$\leq \frac{4C^2}{T^{1/2}} = o(1)$$

and

$$E[|S_{2,2}|] \leq \frac{(\rho - 1) \sum_{t=2}^{B} \sum_{k=1}^{t-1} E[|\sum_{j=1}^{T-B} \sum_{m=0}^{k+j-2} \rho^m \Delta d_{t+j} u_{k+j-1-m}|]}{B^{3/2}T^{1/2}} + \frac{cC}{T} E[|x_0|]$$
$$\leq \frac{C(\rho - 1) \sum_{t=2}^{B} \sum_{k=1}^{t-1} \sqrt{\sum_{j=1}^{T-B} (\frac{1}{1-\rho})^2 |\Delta d_{t+j}|^2}}{B^{3/2}T^{1/2}} + o(1)$$
$$\leq \frac{C \sum_{t=2}^{B} \sqrt{\sum_{j=1}^{T-B} |\Delta d_{t+j}|^2}}{B^{1/2}T^{1/2}} + o(1) \leq \frac{2C^2}{T^{1/2}} + o(1) = o(1).$$

Analogously, from $\Delta x_{t+j} = u_{t+j} + (\rho - 1)x_{t+j-1}$, we decompose $S_3 = S_{3,1} + S_{3,2}$, where

$$S_{3,1} = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} u_{t+j}(d_{t+j-1} - d_j)}{B^{3/2}T^{1/2}},$$
$$S_{3,2} = \frac{(\rho - 1) \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Big(\sum_{m=0}^{t+j-2} \rho^m u_{t+j-1-m} + x_0 \Big) (d_{t+j-1} - d_j)}{B^{3/2}T^{1/2}}.$$

Jensen's inequality yields

$$E[|S_{3,1}|] \leq \frac{\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{E[(\sum_{j=1}^{T-B}u_{t+j}\Delta d_{k+j})^2]}}{B^{3/2}T^{1/2}} \leq \frac{C\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{j=1}^{T-B}|\Delta d_{k+j}|^2}{B^{3/2}T^{1/2}}$$
$$\leq \frac{C\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{\sum_{j=1}^{T-B}(CT^{-1}+\sum_{l=1}^{L}|\lambda_l|1_{\{k+j=t_l\}})^2}}{B^{3/2}T^{1/2}} \leq \frac{4C^2}{T^{1/2}}$$

as well as

$$E[|S_{3,2}|] \leq \frac{(\rho-1)\sum_{t=2}^{B}E[|\sum_{j=1}^{T-B}\sum_{m=0}^{t+j-2}\rho^m u_{t+j-1-m}(d_{t+j-1}-d_j)|]}{B^{3/2}T^{1/2}} + \frac{4cC^2}{T}E[|x_0|]$$
$$\leq \frac{(\rho-1)\sum_{t=2}^{B}\sqrt{E[(\sum_{j=1}^{T-B}\sum_{m=0}^{t+j-2}\rho^m u_{t+j-1-m}(d_{t+j-1}-d_j))^2]}}{B^{3/2}T^{1/2}} + o(1)$$
$$\leq \frac{(\rho-1)\sum_{t=2}^{B}\sqrt{\frac{4C^4B}{(1-\rho)^2}}}{B^{3/2}T^{1/2}} + o(1) \leq \frac{2C^2}{T^{1/2}} + o(1) = o(1).$$

Consequently, $E[|\mathcal{Y}_{1,T}-\mathcal{X}_{1,T}|] \leq |S_1|+|S_2|+|S_3| = o(1)$, and, by Markov's inequality, it follows that $|\mathcal{Y}_{1,T}-\mathcal{X}_{1,T}| = o_P(1)$. To show the second result, we analogously rewrite

$$(y_{t+j-1}-y_j)^2-(x_{t+j-1}-x_j)^2 = (d_{t+j-1}-d_j)^2+2(x_{t+j-1}-x_j)(d_{t+j-1}-d_j)$$

and define

$$S_4 = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(d_{t+j-1}-d_j)^2}{B^2T}, \quad S_5 = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}2(x_{t+j-1}-x_j)(d_{t+j-1}-d_j)}{B^2T},$$

where $\mathcal{Y}_{2,T}-\mathcal{X}_{2,T} = S_4+S_5$. For the first term, note that

$$|S_4| \leq \frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Big(CBT^{-1}+\sum_{k=1}^{t-1}\sum_{l=1}^{L}|\lambda_l|1_{\{k+j=t_l\}}\Big)^2 \leq 4C^2T^{-1} = o(1).$$

From $(x_{t+j-1}-x_j) = \sum_{k=1}^{t-1}u_{k+j}+(\rho-1)x_{k+j-1}$ the second term is then decomposed into $S_5 = S_{5,1}+S_{5,2}$, where

$$S_{5,1} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}2u_{k+j}(d_{t+j-1}-d_j)}{B^2T},$$
$$S_{5,2} = \frac{2(\rho-1)\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}(\sum_{m=0}^{k+j-2}\rho^m u_{k+j-1-m}+x_0)(d_{t+j-1}-d_j)}{B^2T}.$$

Jensen's inequality yields

$$E[|S_{5,1}|] \leq \frac{2\sum_{t=2}^{B}\sum_{k=1}^{t-1} E[|\sum_{j=1}^{T-B} u_{k+j}(d_{t+j-1}-d_j)|]}{B^2T}$$
$$\leq \frac{2\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{E[(\sum_{j=1}^{T-B} u_{k+j}(d_{t+j-1}-d_j))^2]}}{B^2T} \leq \frac{2C\sum_{t=2}^{B}\sqrt{\sum_{j=1}^{T-B}|d_{t+j-1}-d_j|^2}}{BT}$$
$$\leq \frac{2C\sum_{t=2}^{B}\sqrt{\sum_{j=1}^{T-B}(CBT^{-1}+\sum_{k=1}^{t-1}\sum_{l=1}^{L}|\lambda_l|1_{\{k+j=t_l\}})^2}}{BT} \leq \frac{7C^2}{B^{1/2}} = o(1)$$

and

$$E[|S_{5,2}|] \leq \frac{2(\rho-1)}{B^2T}\sum_{t=2}^{B}\sum_{k=1}^{t-1} E\Big[\Big|\sum_{j=1}^{T-B}\sum_{m=0}^{k+j-2}\rho^m u_{k+j-1-m}(d_{t+j-1}-d_j)\Big|\Big] + \frac{4E[|x_0|]cC}{B}$$
$$\leq \frac{2(\rho-1)\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{E[(\sum_{j=1}^{T-B}\sum_{m=0}^{k+j-2}\rho^m u_{k+j-1-m}(d_{t+j-1}-d_j))^2]}}{B^2T}$$
$$\leq \frac{2(\rho-1)\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{\frac{4C^4B}{(1-\rho)^2}}}{B^2T} + o(1) \leq \frac{2C^2}{B^{1/2}} + o(1) = o(1).$$

Hence, $E[|\mathcal{Y}_{2,T}-\mathcal{X}_{2,T}|] \leq |S_4|+|S_5| = o(1)$, and, from Markov's inequality, it follows that $|\mathcal{Y}_{2,T}-\mathcal{X}_{2,T}| = o_P(1)$.

Proof of Lemma 1.2

From the representation $\Delta x_{t+j} = u_{t+j} + (\rho-1)x_{t+j-1}$, we decompose the numerator statistic into $\mathcal{X}_{1,T} = S_1+S_2+S_3+S_4$, where

$$S_1 = \sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\frac{u_{t+j}u_{k+j}}{B^{3/2}T^{1/2}}, \quad S_2 = \sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\frac{(\rho-1)u_{t+j}x_{k+j-1}}{B^{3/2}T^{1/2}},$$
$$S_3 = \sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\frac{(\rho-1)u_{k+j}x_{t+j-1}}{B^{3/2}T^{1/2}}, \quad S_4 = \sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\frac{(\rho-1)^2x_{t+j-1}x_{k+j-1}}{B^{3/2}T^{1/2}}.$$

Note that $S_2+S_3+S_4 = 0$ if $c=0$. First, we show that S_1 is the sum of the elements of a martingale difference array. We rearrange

$$S_1 = \sum_{t=1}^{B}\sum_{j=t+1}^{t+T-B}\sum_{k=1}^{t-1}\frac{u_ju_{k+j-t}}{B^{3/2}T^{1/2}} = \sum_{j=1}^{T}\sum_{t\in\mathcal{I}_j}\sum_{k=1}^{t-1}\frac{u_ju_{j-k}}{B^{3/2}T^{1/2}} = \sum_{j=1}^{T}q_{j,T},$$

where $\mathcal{I}_j = \{t \in \mathbb{N} : \ 1 \leq t \leq B, \ j + B - T \leq t \leq j - 1\}$. The elements of the sum $q_{j,T} = \sum_{t \in \mathcal{I}_j} \sum_{k=1}^{t-1} \frac{u_j u_{j-k}}{B^{3/2} T^{1/2}}$ form a martingale difference sequence for $j \leq T$ and $T \in \mathbb{N}$. Hence, we have $E[S_1] = 0$ and $Var[S_1] = \sum_{j=1}^{T} E[q_{j,T}^2] = \frac{1}{B^3 T} \sum_{j=1}^{T} E[\tilde{q}_{j,T}^2]$, where $\tilde{q}_{j,T} = B^{3/2} T^{1/2} q_{j,T}$. The index set can be expressed as

$$\mathcal{I}_j = \begin{cases} \{t \in \mathbb{N} : \ 2 \leq t \leq j - 1\} & \text{if } j \in [1, B], \\ \{t \in \mathbb{N} : \ 2 \leq t \leq B\} & \text{if } j \in [B + 1, T - B], \\ \{t \in \mathbb{N} : \ j + B - T \leq t \leq B\} & \text{if } j \in [T - B + 1, T]. \end{cases}$$

Note that by mathematical induction on n, the identity

$$\sum_{t=2}^{n} \sum_{k=1}^{t-1} a_k = \sum_{k=1}^{n-1} (n - k) a_k \tag{1.2}$$

holds true for any sequence $(a_t)_{t \in \mathbb{N}}$. For $j \in [1, B]$, it follows that

$$\tilde{q}_{j,T} = \sum_{t=2}^{j-1} \sum_{k=1}^{t-1} u_j u_{j-k} = \sum_{k=1}^{j-1} (j - 1 - k) u_{j-k} = \sum_{k=1}^{j-2} k u_j u_{k+1}, \quad E[\tilde{q}_{j,T}^2] = \sigma^4 \sum_{k=1}^{j-2} k^2.$$

Analogously, if $j \in [B + 1, T - B]$, we obtain

$$\tilde{q}_{j,T} = \sum_{t=2}^{B} \sum_{k=1}^{t-1} u_j u_{j-k} = \sum_{k=1}^{B} (B - k) u_{j-k} = \sum_{k=1}^{B-1} k u_j u_{j-B+k}, \quad E[\tilde{q}_{j,T}^2] = \sigma^4 \sum_{k=1}^{B-1} k^2. \tag{1.3}$$

If $j \in [T - B + 1, T]$, or, equivalently, if $i \in [1, B]$ for $i = j + B - T$, we have

$$\begin{aligned} \tilde{q}_{i,T}^2 &= \Big(\sum_{t=i}^{B} \sum_{k=1}^{t-1} u_j u_{j-k} \Big)^2 = \Big(\sum_{t=2}^{B} \sum_{k=1}^{t-1} u_j u_{j-k} - \sum_{t=2}^{i-1} \sum_{k=1}^{t-1} u_j u_{j-k} \Big)^2 \\ &= \Big(\sum_{k=1}^{B} (B - k) u_j u_{j-k} - \sum_{k=1}^{i-1} (i - 1 - k) u_j u_{j-k} \Big)^2 \\ &= \Big(\sum_{k=1}^{B-1} k u_j u_{j-B+k} \Big)^2 + \Big(\sum_{k=1}^{i-2} k u_j u_{T-B+k+1} \Big)^2 - 2 \sum_{k=1}^{i-1} \sum_{l=1}^{B} (i - 1 - k)(B - l) u_j^2 u_{j-k} u_{j-l} \end{aligned}$$

such that $E[\tilde{q}_{i,T}^2] = \sigma^4[\sum_{k=1}^{B-1} k^2 + \sum_{k=1}^{i-2}[k^2 - 2k(B-k)]]$. Combining all cases yields

$$\begin{aligned}\sum_{j=1}^{T} E[\tilde{q}_{j,T}^2] &= \sigma^4\Big[(T-B)\sum_{k=1}^{B-1} k^2 + \sum_{j=1}^{B}\sum_{k=1}^{j-2}[4k^2 - 2Bk]\Big] \\ &= \frac{\sigma^4 B(B-1)}{6}[(T-B)(2B-1) - 2(B-2)]\end{aligned}$$

by the Gaussian summation formulas, and, consequently,

$$Var[S_1] = \sigma^4 \frac{(T-B)(B-1)(2B-1) - 2(B-1)(B-2)}{6B^2T}.$$

It remains to show that $S_2 + S_3 + S_4 = c \cdot (\sigma^2/2 + o_P(1))$. Let C be the constant given by $C = \max\{E[u_j^4], \sigma^2, c, 1\} < \infty$. Note that

$$\sum_{m=0}^{T} |\rho|^m = \frac{1 - |\rho|^{T+1}}{1-\rho} = (1 - |\rho|^{T+1})\frac{\sqrt{BT}}{c} = o(\sqrt{BT}).$$

The second term then satisfies

$$\begin{aligned}E[|S_2|] &\leq \frac{(\rho-1)\sum_{t=2}^{B}\sum_{k=1}^{t-1} E[|\sum_{j=1}^{T-B}\sum_{m=0}^{k+j-2} \rho^m u_{k+j-1-m}u_{t+j}|]}{B^{3/2}T^{1/2}} \\ &\leq \frac{C^2\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sqrt{\sum_{j=1}^{T-B}\sum_{m=0}^{k+j-2}\rho^{2m}}}{B^2T} = o(1)\end{aligned}$$

and, for the fourth term, we obtain

$$\begin{aligned}E[|S_4|] &\leq \frac{(\rho-1)^2\sum_{t=2}^{B}\sum_{k=1}^{t-1} E[|\sum_{j=1}^{T-B}\sum_{m=0}^{t+j-2}\sum_{l=0}^{k+j-2}\rho^m\rho^l u_{t+j-1-m}u_{k+j-1-l}|]}{B^{3/2}T^{1/2}} \\ &\leq \frac{C^3B^2T\sqrt{(\sum_{m=0}^{T}|\rho|^m)^2}}{B^{5/2}T^{3/2}} = o(1).\end{aligned}$$

Hence, by Markov's inequality, $S_2 + S_4 = o_P(1)$. For the third term, we obtain

$$\begin{aligned}E[S_3] &= -\frac{c\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{m=0}^{t+j-2}\rho^m E[u_{k+j}u_{t+j-1-m}]}{B^2T} \\ &= -\frac{c\sigma^2(T-B)\sum_{t=2}^{B}\sum_{k=1}^{t-1}\rho^{k-1}}{B^2T} = -\frac{c\sigma^2}{2} + o(1)\end{aligned}$$

and

$$E[S_3^2] = \frac{c^2 E[(\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{m=0}^{t+j-2}\rho^m u_{k+j}u_{t+j-1-m})^2]}{B^4T^2}$$
$$= \frac{c^2\sigma^4(\sum_{t=2}^{B}\sum_{k=1}^{t-1}\rho^{k-1})^2}{B^4} + o(1) = \frac{c^2\sigma^4}{4} + o(1).$$

Hence $Var[S_3] = o(1)$, and, by Chebyshev's inequality, it follows that $S_3 = c\sigma^2/2 + o_P(1)$. To show (b), we decompose the denominator statistic into $\mathcal{X}_{2,T} = S_5 + S_6 + S_7$, where

$$S_5 = \frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Big(\sum_{k=1}^{t-1}u_{j+k}\Big)^2, \quad S_6 = \frac{2(\rho-1)}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}x_{j+k-1}u_{j+l},$$
$$S_7 = \frac{(\rho-1)^2}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\Big(\sum_{k=1}^{t-1}x_{j+k-1}\Big)^2.$$

The first term satisfies $E[S_5] = \frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sigma^2 = \sigma^2\frac{(T-B)(B-1)}{2BT}$ and

$$Var[S_5] = \frac{1}{B^4T^2}E\Big[\Big(\sum_{j=1}^{T-B}\sum_{t=2}^{B}\big(\sum_{k=1}^{t-1}u_{j+k}\big)^2\Big)^2\Big] = O(BT^{-1}).$$

Analogously to S_2 and S_4, we obtain

$$E[|S_6|] \leq \frac{2c\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sum_{l=1}^{t-1}E[|\sum_{j=1}^{T-B}\sum_{m=0}^{j+k-2}\rho^m u_{j+k-1-m}u_{j+l}|]}{B^{5/2}T^{3/2}}$$
$$\leq \frac{2cB^{1/2}}{T^{3/2}}\cdot O(T) = O(B^{1/2}T^{-1/2})$$

and

$$E[|S_7|] \leq \frac{c^2\sum_{j=2}^{B}\sum_{k,l=1}^{t-1}E[|\sum_{j=1}^{T-B}\sum_{m=0}^{j+k-2}\sum_{n=0}^{j+l-2}\rho^{m+n}u_{j+k-1-m}u_{j+k-1-n}|]}{B^3T^2}$$
$$\leq \frac{c^2}{T^2}o(T^{5/4}B^{1/4}) = o(1).$$

Since $S_6 = S_7 = 0$ if $c = 0$, the assertion follows.

Proof of Theorem 1.1

From the proof of Lemma 1.2, it follows that $\max_{1\leq j\leq T}E[q_{j,T}^2] = O(T^{-1})$. Then, by Jensen's inequality, $\max_{1\leq j\leq T}E[|q_{j,T}|] = O(T^{-1/2})$ and therefore $\max_{1\leq j\leq T}|q_{j,T}| \xrightarrow{p} 0$.

Furthermore, Lemma 1.2 yields $E[\mathcal{X}_{1,T}] = -\frac{c\sigma^2}{2} + o(1)$ and

$$V_T = Var\Big[\sum_{j=1}^{T} q_{j,T}\Big] = \frac{\sigma^4}{3} + o(1).$$

The first result then follows from Lemmas 1.1 and 1.8. Furthermore, Lemma 1.2 implies that $E[\mathcal{X}_{2,T}] = \sigma^2/2 + o(1)$ and that $Var[\mathcal{X}_{2,T}] = o(1)$. By Chebyshev's inequality together with Lemma 1.1, the second result follows.

Proof of Theorem 1.2

We rewrite

$$\begin{aligned}\Delta x_{t+j} x_{t+j-1} &= \frac{\Delta x_{t+j}(x_{t+j-1} + x_{t+j} - \Delta x_{t+j})}{2} \\ &= \frac{(x_{t+j} - x_{t+j-1})(x_{t+j} + x_{t+j-1}) - (\Delta x_{t+j})^2}{2} = \frac{x_{t+j}^2 - x_{t+j-1}^2 - (\Delta x_{t+j})^2}{2}\end{aligned}$$

such that

$$\begin{aligned}&\sum_{t=2}^{B} \Delta x_{t+j}(x_{t+j-1} - x_j) = \sum_{t=1}^{B} \frac{x_{t+j}^2 - x_{t+j-1}^2 - (\Delta x_{t+j})^2}{2} - \Delta x_{t+j} x_j \\ &= \frac{1}{2}(x_{j+B}^2 - x_j^2) - (x_{j+B} x_j - x_j^2) - \frac{1}{2}\sum_{t=1}^{B} (\Delta x_{t+j})^2 = \frac{(x_{j+B} - x_j)^2}{2} - \frac{1}{2}\sum_{t=1}^{B} (\Delta x_{t+j})^2.\end{aligned}$$

Let $Y_T(r) = T^{-1/2} x_{\lfloor rT \rfloor}$ for $r \geq 0$. Then, with Lemma 1.1,

$$\begin{aligned}\mathcal{Y}_{1,T} = \mathcal{X}_{1,T} + o_P(1) &= \frac{\sum_{j=1}^{T-B}(x_{B+j} - x_j)^2 - \sum_{j=1}^{T-B}\sum_{t=1}^{B}(\Delta x_{t+j})^2}{2B^{3/2}T^{1/2}} \\ &= \frac{\int_0^{1-b}(Y_T(b+r) - Y_T(r))^2 \,\mathrm{d}r - \frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=1}^{B}(\Delta x_{t+j})^2}{2b^{3/2}} + o_P(1).\end{aligned}$$

From $\Delta x_t = u_t$, it follows that

$$E\Big[\frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=1}^{B}(\Delta x_{t+j})^2\Big] = \frac{1}{T^2}\sum_{j=1}^{T-B}\sum_{t=1}^{B} E[u_{t+j}^2] = \frac{B(T-B)\sigma^2}{T^2} = b(1-b)\sigma^2 + o(1),$$

which implies that

$$\mathcal{Y}_{1,T} = \frac{\int_0^{1-b}(Y_T(b+r) - Y_T(r))^2 \,\mathrm{d}r - b(1-b)\sigma^2}{2b^{3/2}} + o_P(1). \tag{1.4}$$

Furthermore, Lemma 1.1 yields

$$\mathcal{Y}_{2,T} = \mathcal{X}_{2,T} + o_P(1) = \frac{1}{b^2}\int_0^{1-b}\int_r^{b+r}(Y_T(s) - Y_T(r))^2 \,\mathrm{d}s\,\mathrm{d}r + o_P(1). \tag{1.5}$$

The assertion follows from Lemma 1.7, together with the continuous mapping theorem.

Proof of Lemma 1.3

Since $(1-\hat{\rho}) = O_P(B^{-1/2}T^{-1/2})$ and $x_t = O_P(T^{1/2})$, the residuals satisfy

$$\begin{aligned}\hat{u}_t &= y_t - \hat{\rho}y_{t-1} = \Delta y_t + (1-\hat{\rho})y_{t-1} = \Delta d_t + u_t + (\rho-1)x_{t-1} + (1-\hat{\rho})y_{t-1} \\ &= u_t + \Delta d_t + \mathcal{O}_P(B^{-1/2}).\end{aligned}$$

Let $\overline{u}_j = \frac{1}{B}\sum_{k=1}^B u_{j+k}$ and $\overline{\Delta d}_j = \frac{1}{B}\sum_{k=1}^B \Delta d_{j+k}$. Then, for $t = 1, \ldots, B$,

$$\sum_{j=1}^{T-B}\Big(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^B \hat{u}_{j+k}\Big)^2 = \sum_{j=1}^{T-B}(u_{j+t} - \overline{u}_j + \Delta d_{t+j} - \overline{\Delta d}_j)^2 + O_P(TB^{-1/2}).$$

Then $\sum_{j=1}^{T-B}(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^B \hat{u}_{j+k})^2 = \sum_{j=1}^{T-B} u_{j+t}^2 + o_P(T)$, where $\sum_{j=1}^T (\Delta d_t)^2 = o(T)$ holds true due to the piecewise Lipschitz continuity of $d(r)$. Consequently,

$$\frac{1}{(T-B)(B-1)}\sum_{j=1}^{T-B}\sum_{t=1}^{B}\Big(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^B \hat{u}_{j+k}\Big)^2 = \frac{1}{T}\sum_{j=1}^{T-B} u_{j+t}^2 + o_P(1) = \sigma^2 + o_P(1)$$

as $B, T \to \infty$ and $B/T \to 0$. Under fixed-b asymptotics, we obtain

$$\frac{1}{T}\sum_{j=1}^T(\hat{u}_j - \overline{\hat{u}})^2 = \frac{1}{T}\sum_{j=1}^T u_j^2 + o_P(1) = \sigma^2 + o_P(1),$$

as $B/T \to b$, $0 < b < 1$ and $B, T \to \infty$. The assertion follows from Slutsky's theorem.

Proof of Lemma 1.4

We follow the proof of Lemma 1.1, except that the variance is time-varying and bounded with $E[u_t^2] = \sigma^2(t/T) < C_3^2$ for some constant $C_3 < \infty$. The common constant is given by $C = \max\{C_1, C_2, C_3, 1\} < \infty$, and the remaining steps follow analogously to the proof of Lemma 1.1.

Proof of Theorem 1.3

We follow the proof of Lemma 1.2 and obtain $S_2 + S_4 = o_P(1)$, as well as

$$E[S_3] = -\frac{c}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\rho^{t-k-1}E[u_{k+j}^2] = -\frac{c\int_0^1\sigma^2(r)\,\mathrm{d}r}{B^2}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\rho^{t-k-1} + o(1)$$
$$= -\frac{c}{2}\int_0^1\sigma^2(r)\,\mathrm{d}r + o(1)$$

and $Var[S_3] = o(1)$. Furthermore,

$$Var[S_1] = \frac{\sum_{j=1}^{T}E[\tilde{q}_{j,T}^2]}{B^3T} = \frac{\sum_{j=B+1}^{T-B}\sum_{k=1}^{B-1}k^2E[u_j^2]E[u_{j-B+k}^2]}{B^3T} + o(1)$$
$$= \int_{\frac{B}{T}}^{\frac{T-B}{T}}\int_0^1 s^2\sigma^2(r)\sigma^2(\tfrac{j-\lfloor(1-s)B\rfloor}{T})\,\mathrm{d}s\,\mathrm{d}r + o(1) = \int_0^1\int_0^1 s^2\sigma^4(r)\,\mathrm{d}s\,\mathrm{d}r + o(1)$$
$$= \frac{1}{3}\int_0^1\sigma^4(r)\,\mathrm{d}r + o(1).$$

Lemma 1.8 yields $\mathcal{X}_{1,T} \xrightarrow{d} \mathcal{N}(-\frac{c}{2}\int_0^1\sigma^2(r)\,\mathrm{d}r, \frac{1}{3}\int_0^1\sigma^4(r)\,\mathrm{d}r)$. For the denominator statistic, we obtain $S_6 + S_7 = o_P(1)$ and

$$E[S_5] = \frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\sigma^2(\tfrac{j+k}{T}) = \frac{1}{B^2T}\sum_{j=1}^{T-B}\sum_{k=1}^{B-1}(B-k)\sigma^2(\tfrac{j+k}{T})$$
$$= \int_0^{\frac{T-B}{T}}\int_0^1(1-s)\sigma^2(r+s\tfrac{B}{T})\,\mathrm{d}s\,\mathrm{d}r + o(1) = \int_0^1\int_0^1(1-s)\sigma^2(r)\,\mathrm{d}s\,\mathrm{d}r + o(1)$$
$$= \frac{1}{2}\int_0^1\sigma^2(r)\,\mathrm{d}r + o(1),$$

while $Var[S_5] = o(1)$. Then, following the proof of Theorem 1.1, result (a) follows with Lemma 1.4.

To show (b), we consider equation (1.4) under heteroskedasticity, which is given by

$$\mathcal{Y}_{1,T} = \frac{\int_0^{1-b}(Y_T(b+r) - Y_T(r))^2 \,\mathrm{d}r - b(1-b)\overline{\sigma}^2}{2b^{3/2}} + o_P(1),$$

where $\overline{\sigma}^2 = \int_0^1 \sigma^2(r)\,\mathrm{d}r$. Together with equation (1.5) and Lemma 1.7, the assertion follows according to Slutsky's theorem and the continuous mapping theorem.

Proof of Theorem 1.4

From the proof of Lemma 1.3, we have $\sum_{j=1}^{T-B}(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^{B}\hat{u}_{j+k})^2 = \sum_{j=1}^{T-B}u_{j+t}^2 + o_P(T)$ and $\sum_{j=1}^{\lfloor sT\rfloor}(\hat{u}_j - \overline{\hat{u}})^2 = \sum_{j=1}^{\lfloor sT\rfloor}u_j^2 + o_P(T)$, where $s \in [0,1]$. Then, as $B, T \to \infty$ and $B/T \to 0$, we obtain $\hat{\sigma}^2_{\mathrm{sb}} = \int_0^1 \sigma^2(r)\,\mathrm{d}r + o_P(1)$ and $\hat{\kappa}^2 = \int_0^1 \sigma^4(r)\,\mathrm{d}r + o_P(1)$. Then, (a) follows from the proof of Theorem 1.1. For (b), note that fixed-b asymptotics yield

$$\frac{1}{T}\sum_{j=1}^{\lfloor sT\rfloor}\Big(\hat{u}_j - \frac{1}{\lfloor sT\rfloor}\sum_{k=1}^{\lfloor sT\rfloor}\hat{u}_k\Big)^2 = \int_0^s \sigma^2(r)\,\mathrm{d}r + o_P(1), \quad s \in [0,1],$$

as $B/T \to b$, $0 < b < 1$ and $B, T \to \infty$. Then, $\hat{\sigma}^2_{\mathrm{fb}} = \int_0^1 \sigma^2(r)\,\mathrm{d}r + o_P(1)$ follows with $s = 1$. Furthermore, Slutsky's theorem implies that $\hat{\eta}(s) = \eta(s) + o_P(1)$ holds pointwise for all $s \in [0,1]$. Uniform convergence then follows by Dini's theorem since both $\hat{\eta}(s)$ and $\eta(s)$ are continuous, monotone, and bounded. Following the notation of Lemma 1.7, the statistic τ-FB$^{\mathrm{H}}$ is a continuous functional of $\tilde{Y}_T(r)$, and, analogously to the proof of Theorem 1.2, the assertion follows with the continuous mapping theorem and Slutsky's theorem.

Proof of Lemma 1.5

Equation (1.1) can be rewritten as $\Delta y_t = z_t'\beta + e_t$, where $z_t = (\Delta y_{t-1}, \ldots, \Delta y_{t-p}, y_{t-1})'$ and $\beta = (\beta_1, \ldots, \beta_p, \varphi)'$ for $t = p+1, \ldots, T$. The least squares estimator for β is then given by $\hat{\beta} = (\sum_{t=p+1}^{T} z_t z_t')^{-1}\sum_{t=p+1}^{T} z_t \Delta y_t$. We derive a consistent estimator for the coefficients $(\theta_1, \ldots, \theta_p)'$ that is asymptotically equal to $(\hat{\beta}_1, \ldots, \hat{\beta}_p)'$. From Assumption 1.4, it follows that $\Delta y_t = \Delta d_t + \Delta x_t = \Delta d_t + \phi x_{t-1} + \sum_{i=1}^{p}\theta_i u_{t-i} + \epsilon_t$. Then, from $\rho = 1$, we obtain $\Delta y_t - \Delta d_t^* = \phi y_{t-1} + \sum_{i=1}^{p}\theta_i \Delta y_{t-i} + \epsilon_t$, which can be equivalently rewritten as $\Delta y_t - \delta_t = z_t'\vartheta + \epsilon_t$ for $t = p+1, \ldots, T$, where $\vartheta = (\theta_1, \ldots, \theta_p, \phi)'$. The OLS estimator is given by $\hat{\vartheta} = (\sum_{t=p+1}^{T} z_t z_t')^{-1}\sum_{t=p+1}^{T} z_t(\Delta y_t - \Delta d_t^*)$. Note that the estimator satisfies $p\|\hat{\vartheta} - \vartheta\|_V = o_P(1)$, where $\|\cdot\|_V$ is an arbitrary vector norm on $\mathbb{R}^{p+1}$. In the following, we show that $p\|\hat{\beta} - \hat{\vartheta}\|_V = o_P(1)$. Note that $\hat{\beta} - \hat{\vartheta} = (\sum_{t=p+1}^{T} z_t z_t')^{-1}\sum_{t=p+1}^{T} z_t \Delta d_t^*$. For

notational convenience, let $\theta_0 = -1$. Then, $\Delta d_t^* = -\sum_{i=1}^p \theta_i \Delta d_{t-i}$. Following the proof of Lemma 1.1, there exists a constant $C < \infty$ such that the deterministic part satisfies $|\Delta d_t^*| \leq \sum_{i=1}^p |\theta_i|(CT^{-1} + \sum_{l=1}^L |\lambda_l| 1_{\{t-i=t_l\}})$, where $\sum_{l=1}^L |\lambda_l| < C$, $\sum_{i=0}^p |\theta_i| < pC$ and $0 \leq t_l \leq 1$ for $l = 1, \ldots, L$. Then,

$$
\begin{aligned}
&\Big\| \frac{p}{T} \sum_{t=p+1}^{T} z_t \Delta d_t^* \Big\|_V \leq \frac{p}{T} \sum_{t=p+1}^{T} |\Delta d_t^*| \|z_t\|_V \\
&\leq \frac{p}{T} \sum_{t=p+1}^{T} \sum_{i=0}^{p} |\theta_i| \Big(CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t-i=t_l\}} \Big) \|z_t\|_V \\
&\leq \frac{p}{T^{1/2}} \Big(\frac{C^2}{T^{3/2}} \sum_{t=p+1}^{T} \|z_t\|_V + \frac{1}{pT^{1/2}} \sum_{i=0}^{p} \sum_{l=1}^{L} |\theta_i||\lambda_l| \|z_{t_l+i}\|_V \Big) \\
&= O_P(p^2 T^{-1/2}) = o_P(1).
\end{aligned}
$$

Let $\|\cdot\|_M$ be the matrix norm induced by $\|\cdot\|_V$. Then, $\|(\frac{1}{T}\sum_{t=p+1}^T z_t z_t')^{-1}\|_M = O_P(1)$ and consequently $p\|\hat{\beta} - \hat{\vartheta}\|_V \leq \|(\frac{1}{T}\sum_{t=p+1}^T z_t z_t')^{-1}\|_M \, \|\frac{p}{T}\sum_{t=p+1}^T z_t \Delta d_t^*\|_V = o_P(1)$. The triangle inequality then yields $p\|\hat{\beta} - \vartheta\| \leq p\|\hat{\beta} - \hat{\vartheta}\|_V + \|\hat{\vartheta} - \vartheta\|_V = o_P(1)$, and the assertion follows by setting $\|\cdot\|_V$ equal to the maximum norm.

Proof of Lemma 1.6

For notational convenience, we define $\hat{\theta}_i = \hat{\beta}_i$ for $i = 1, \ldots, p$ and $\theta_0 = \hat{\theta}_0 = -1$. The pre-whitened series satisfy $y_t^* = -\sum_{i=0}^p \theta_i y_{t-i}$, $d_t^* = -\sum_{i=0}^p \theta_i d_{t-i}$, $x_t^* = -\sum_{i=0}^p \theta_i x_{t-i}$, and the estimated pre-whitened series are given by $\hat{y}_t^* = -\sum_{i=0}^p \hat{\theta}_i y_{t-i}$, $\hat{d}_t^* = -\sum_{i=0}^p \hat{\theta}_i d_{t-i}$, and $\hat{x}_t^* = -\sum_{i=0}^p \hat{\theta}_i x_{t-i}$. From Lemma 1.5, it follows that $(\hat{\theta}_0, \ldots, \hat{\theta}_p)' \xrightarrow{p} (\theta_0, \ldots, \theta_p)'$. Furthermore, the estimated pre-whitened series satisfies $\hat{x}_t^* = x_t^* + \sum_{i=0}^p (\theta_i - \hat{\theta}_i) x_{t-i}$, and $\rho = 1$ yields $\Delta x_t^* = \epsilon_t$, as well as $\Delta \hat{x}_t^* = -\sum_{i=0}^p \hat{\theta}_i u_{t-i} = \epsilon_t + \sum_{i=0}^p (\theta_i - \hat{\theta}_i) u_{t-i}$. Following the proof of Lemma 1.1, there exists a common constant $C \in [1, \infty)$ such that

$$
\begin{aligned}
|\Delta d_{t+j}^*| &\leq \sum_{i=0}^{p} |\theta_i| |\Delta d_{t+j-i}| \leq \sum_{i=0}^{p} |\theta_i| \Big(CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j-i=t_l\}} \Big), \\
|d_{t+j-1}^* - d_j^*| &\leq \sum_{i=0}^{p} |\theta_i| |d_{t+j-1-i} - d_{j-i}| \leq \sum_{i=0}^{p} |\theta_i| \Big(CBT^{-1} + \sum_{k=1}^{t-1} \sum_{l=1}^{L} |\lambda_l| 1_{\{k+j-i=t_l\}} \Big),
\end{aligned}
$$

where the discontinuity points are given by $t_l \in [0, 1]$ for $l = 1, \ldots, L$. Furthermore, the constant satisfies $\sum_{l=1}^L |\lambda_l| < C$ and $\sum_{i=0}^p |\theta_i| < pC$. The stationary error term has an MA representation $u_t = \sum_{m=0}^\infty \psi_m \epsilon_{t-m}$, where $\sum_{m=0}^\infty |\psi_m| \leq C$. The variance of ϵ_t is bounded

with $\sigma^2(r) \leq C^2$ for all $r \leq 1$. Let

$$\hat{\mathcal{X}}^*_{1,T} = \frac{1}{B^{3/2}T^{1/2}} \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta \hat{x}^*_{t+j}(\hat{x}^*_{t+j-1} - \hat{x}^*_j), \quad \hat{\mathcal{X}}^*_{2,T} = \frac{1}{B^2 T} \sum_{j=1}^{T-B} \sum_{t=2}^{B} (\hat{x}^*_{t+j-1} - \hat{x}^*_j)^2.$$

In the first part of the proof, we show that $|\hat{\mathcal{Y}}^*_{1,T} - \hat{\mathcal{X}}^*_{1,T}| = o_P(1)$ and $|\hat{\mathcal{Y}}^*_{2,T} - \hat{\mathcal{X}}^*_{2,T}| = o_P(1)$. We decompose $\hat{\mathcal{Y}}^*_{1,T} - \hat{\mathcal{X}}^*_{1,T} = S_1 + S_2 + S_3$ and $\hat{\mathcal{Y}}^*_{2,T} - \hat{\mathcal{X}}^*_{2,T} = S_4 + S_5$, where

$$S_1 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta \hat{d}^*_{t+j}(\hat{d}^*_{t+j-1} - \hat{d}^*_j)}{B^{3/2}T^{1/2}}, \quad S_2 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta \hat{d}^*_{t+j}(\hat{x}^*_{t+j-1} - \hat{x}^*_j)}{B^{3/2}T^{1/2}},$$
$$S_3 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta \hat{x}^*_{t+j}(\hat{d}^*_{t+j-1} - \hat{d}^*_j)}{B^{3/2}T^{1/2}}$$

and

$$S_4 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} (\hat{d}^*_{t+j-1} - \hat{d}^*_j)^2}{B^2 T}, \quad S_5 = \frac{\sum_{j=1}^{T-B} \sum_{t=2}^{B} 2(\hat{x}^*_{t+j-1} - \hat{x}^*_j)(\hat{d}^*_{t+j-1} - \hat{d}^*_j)}{B^2 T}.$$

For S_1, note that

$$\sum_{j=1}^{T-B} \sum_{t=2}^{B} |\Delta \hat{d}^*_{t+j}(\hat{d}^*_{t+j-1} - \hat{d}^*_j)| \leq \sum_{i_1,i_2=0}^{p} \sum_{j=1}^{T-B} \sum_{t=2}^{B} |\hat{\theta}_{i_1}||\hat{\theta}_{i_2}||A_{i_1,i_2,j,t}|,$$

where $A_{i_1,i_2,j,t} = (CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j-i=t_l\}})(CBT^{-1} + \sum_{k=1}^{t-1} \sum_{l=1}^{L} |\lambda_l| 1_{\{k+j-i=t_l\}})$. From $\sum_{j=1}^{T-B} \sum_{t=2}^{B} |A_{i_1,i_2,j,t}| \leq 4C^2 B$, it follows that $|S_1| \leq \frac{4C^2}{B^{1/2}T^{1/2}} (\sum_{i=0}^{p} |\hat{\theta}_i|)^2$ and that

$$E[|S_1|] \leq \frac{4C^2}{B^{1/2}T^{1/2}} \Big(\sum_{i=0}^{p} |\theta_i| \Big)^2 + o(1) = O(p^2 B^{-1/2} T^{-1/2}) = o(1).$$

For S_2, note that

$$\begin{aligned}
&\Big| \sum_{j=1}^{T-B} \sum_{t=2}^{B} \Delta \hat{d}^*_{t+j}(\hat{x}^*_{t+j-1} - \hat{x}^*_j) \Big| = \Big| \sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \Delta \hat{d}^*_{t+j} \Delta \hat{x}^*_{k+j} \Big| \\
&= \Big| \sum_{j=1}^{T-B} \sum_{t=2}^{B} \sum_{k=1}^{t-1} \sum_{i_1,i_2=0}^{p} \hat{\theta}_{i_1} \hat{\theta}_{i_2} \Delta d_{t+j-i_1} u_{k+j-i_2} \Big| \\
&\leq \sum_{t=2}^{B} \sum_{k=1}^{t-1} \sum_{i_1,i_2=0}^{p} \sum_{m=0}^{\infty} \Big| \psi_m \hat{\theta}_{i_1} \hat{\theta}_{i_2} \sum_{j=1}^{T-B} \Delta d_{t+j-i_1} \epsilon_{k+j-i_2-m} \Big|.
\end{aligned}$$

Then,

$$E[|S_2|] \leq \frac{1}{B^{3/2}T^{1/2}} \sum_{m=0}^{\infty} \sum_{i_1,i_2=0}^{p} \sum_{t=2}^{B} \sum_{k=1}^{t-1} |\psi_m| E\Big[\Big|\hat{\theta}_{i_1}\hat{\theta}_{i_2} \sum_{j=1}^{T-B} \Delta d_{t+j-i_1}\epsilon_{k+j-i_2-m}\Big|\Big].$$

From the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned}
&\sum_{t=2}^{B}\sum_{k=1}^{t-1} E\Big[\Big|\hat{\theta}_{i_1}\hat{\theta}_{i_2} \sum_{j=1}^{T-B} \Delta d_{t+j-i_1}\epsilon_{k+j-i_2-m}\Big|\Big] \\
&\leq \sqrt{E[\hat{\theta}_{i_1}^2\hat{\theta}_{i_2}^2]} \sum_{t=2}^{B}\sum_{k=1}^{t-1} \sqrt{E\Big[\Big(\sum_{j=1}^{T-B} \Delta d_{t+j-i_1}\epsilon_{k+j-i_2-m}\Big)^2\Big]} \\
&\leq |\theta_{i_1}\theta_{i_2} + o(1)| \sum_{t=2}^{B}\sum_{k=1}^{t-1} \sqrt{\sum_{j=1}^{T-B} |\Delta d_{t+j-i_1}|^2\sigma_{k+j-i_2-m}^2} \\
&\leq CB|\theta_{i_1}\theta_{i_2} + o(1)| \sum_{t=2}^{B} \sqrt{\sum_{j=1}^{T-B} \Big(CT^{-1} + \sum_{l=1}^{L} |\lambda_l| 1_{\{t+j-i_1=t_l\}}\Big)^2} \leq 2C^2B^{3/2}|\theta_{i_1}\theta_{i_2} + o(1)|.
\end{aligned}$$

Consequently, $E[|S_2|] \leq \frac{2C^5p^2}{T^{1/2}} + o(1) = O(p^2T^{-1/2}) = o(1)$. For S_3, note that

$$\begin{aligned}
&\Big|\sum_{j=1}^{T-B} \Delta\hat{x}_{t+j}^*(\hat{d}_{t+j-1}^* - \hat{d}_j^*)\Big| = \Big|\sum_{i_1,i_2=0}^{p} \sum_{j=1}^{T-B} \hat{\theta}_{i_1}\hat{\theta}_{i_2} u_{t+j-i_1}(d_{t+j-1-i_2} - d_{j-i_2})\Big| \\
&\leq \sum_{i_1,i_2=0}^{p} \sum_{m=0}^{\infty} \Big|\psi_m\hat{\theta}_{i_1}\hat{\theta}_{i_2} \sum_{j=1}^{T-B} \epsilon_{t+j-i_1-m}(d_{t+j-1-i_2} - d_{j-i_2})\Big|.
\end{aligned}$$

Furthermore,

$$\begin{aligned}
&E\Big[\Big(\sum_{j=1}^{T-B} \epsilon_{t+j-i_1-m}(d_{t+j-1-i_2} - d_{j-i_2})\Big)^2\Big] = \sum_{j=1}^{T-B} |d_{t+j-1-i_2} - d_{j-i_2}|^2\sigma_{t+j-i_1-m}^2 \\
&\leq C\sum_{j=1}^{T-B} \Big(CBT^{-1} + \sum_{k=1}^{t-1}\sum_{l=1}^{L} |\lambda_l| 1_{\{k+j-i_2=t_l\}}\Big)^2 \leq 4C^3B.
\end{aligned}$$

From the Cauchy-Schwarz inequality, it follows that

$$E\Big[\Big|\hat{\theta}_{i_1}\hat{\theta}_{i_2}\sum_{j=1}^{T-B}\epsilon_{t+j-i_1-m}(d_{t+j-1-i_2}-d_{j-i_2})\Big|\Big]$$
$$\leq\sqrt{E[\hat{\theta}_{i_1}^2\hat{\theta}_{i_2}^2]}\sqrt{E\Big[\Big(\sum_{j=1}^{T-B}\epsilon_{t+j-i_1-m}(d_{t+j-1-i_2}-d_{j-i_2})\Big)^2\Big]}\leq 2C^{3/2}B|\theta_{i_1}\theta_{i_2}+o(1)|.$$

Consequently,

$$E[|S_3|]\leq\frac{1}{B^{3/2}T^{1/2}}\sum_{t=2}^{B}\sum_{i_1,i_2=0}^{p}\sum_{m=0}^{\infty}|\psi_m|E\Big[\Big|\hat{\theta}_{i_1}\hat{\theta}_{i_2}\sum_{j=1}^{T-B}\epsilon_{t+j-i_1-m}(d_{t+j-1-i_2}-d_{j-i_2})\Big|\Big]$$
$$\leq\frac{2C^{3/2}B}{B^{3/2}T^{1/2}}\Big(\sum_{i=0}^{p}|\theta_i+o(1)|\Big)^2\sum_{m=0}^{\infty}|\psi_m|=O(p^2B^{-1/2}T^{-1/2})=o(1).$$

For S_4, we obtain

$$|S_4|\leq\frac{1}{B^2T}\sum_{t=2}^{B}\sum_{j=1}^{T-B}\Big(\sum_{i=0}^{p}|\hat{\theta}_i|\Big(C^2BT^{-1}+\sum_{k=1}^{t-1}\sum_{l=1}^{L}|\lambda_l|1_{\{k+j-i=t_l\}}\Big)\Big)^2\leq\frac{4C^4}{T}\Big(\sum_{i=0}^{p}|\hat{\theta}_i|\Big)^2,$$

which implies that $E[|S_4|]\leq\frac{4C^4}{T}\Big(\sum_{i=0}^{p}|\theta_i|\Big)^2+o(1)=O(p^2T^{-1})=o(1)$. For the fifth term, we obtain

$$|S_5|\leq\frac{2}{B^2T}\sum_{i=0}^{p}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Big|\sum_{j=1}^{T-B}\hat{\theta}_i u_{k+j-i}(\hat{d}^*_{t+j-1}-\hat{d}^*_j)\Big|$$
$$\leq\frac{2}{B^2T}\sum_{i_1,i_2=0}^{p}\sum_{m=0}^{\infty}\sum_{t=2}^{B}\sum_{k=1}^{t-1}\Big|\sum_{j=1}^{T-B}\psi_m\hat{\theta}_{i_1}\hat{\theta}_{i_2}\epsilon_{k+j-m-i_1}(d_{t+j-1-i_2}-d_{j-i_2})\Big|.$$

Note that $E[\hat{\theta}_{i_1}^2]=\theta_{i_1}^2+o(1)$ and $E[\hat{\theta}_{i_2}^2]=\theta_{i_2}^2+o(1)$ and that

$$E\Big[\Big(\sum_{j=1}^{T-B}\epsilon_{k+j-m-i_1}(d_{t+j-1-i_2}-d_{j-i_2})\Big)^2\Big]=\sum_{j=1}^{T-B}\sigma^2_{k+j-m-i_1}|d_{t+j-1-i_2}-d_{j-i_2}|^2$$
$$\leq C^2\sum_{j=1}^{T-B}\Big(CBT^{-1}+\sum_{k=1}^{t-1}\sum_{l=1}^{L}|\lambda_l|1_{\{k+j-i_2=t_l\}}\Big)^2\leq 4C^4B.$$

Then, by the Cauchy-Schwarz inequality,

$$E[|S_5|] \leq \frac{4C^2B^{1/2}}{B^2T} \sum_{i_1,i_2=0}^{p} \sum_{m=0}^{\infty} \sum_{t=2}^{B} \sum_{k=1}^{t-1} |\psi_m||\theta_{i_1}\theta_{i_2} + o(1)| \leq \frac{4C^5}{B^{1/2}} + o(1) = o(1).$$

By Markov's inequality, it follows that $S_1 + S_2 + S_3 = o_P(1)$ and that $S_4 + S_5 = o_P(1)$. In the second part of the proof, we show that $|\hat{\mathcal{X}}^*_{1,T} - \mathcal{X}^*_{1,T}| = o_P(1)$ and $|\hat{\mathcal{X}}^*_{2,T} - \mathcal{X}^*_{2,T}| = o_P(1)$. Note that the estimated pre-whitened series satisfies $\hat{x}^*_t = x^*_t + \sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)x_{t-i}$ such that $\Delta\hat{x}^*_{t+j} = \Delta x^*_{t+j} + \sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)\Delta x_{t+j-i}$ as well as

$$\hat{x}^*_{t+j-1} - \hat{x}^*_j = x^*_{t+j-1} - x^*_j + \sum_{i=1}^{p}(\theta_i - \hat{\theta}_i)(x_{t+j-1-i} - x_{j-1}).$$

Then,

$$\begin{aligned}
&\Delta\hat{x}^*_{t+j}(\hat{x}^*_{t+j-1} - \hat{x}^*_j) - \Delta x^*_{t+j}(x^*_{t+j-1} - x^*_j) \\
&= (\Delta\hat{x}^*_{t+j} - \Delta x^*_{t+j})(\hat{x}^*_{t+j-1} - \hat{x}^*_j) + \Delta x^*_{t+j}[(\hat{x}^*_{t+j-1} - \hat{x}^*_j) - (x^*_{t+j-1} - x^*_j)] \\
&= \sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)\Delta x_{t+j-i}(\hat{x}^*_{t+j-1} - \hat{x}^*_j) + \sum_{i=1}^{p}(\theta_i - \hat{\theta}_i)\Delta x^*_{t+j}(x_{t+j-1-i} - x_{j-1})
\end{aligned}$$

and

$$\begin{aligned}
&(\hat{x}^*_{t+j-1} - \hat{x}^*_j)^2 - (x^*_{t+j-1} - x^*_j)^2 \\
&= \Big(\sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)x_{t+j-1-i} - x_{j-i}\Big)^2 - 2\sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)(x^*_{t+j-1} - x^*_j)(x_{t+j-1-i} - x_{j-i}).
\end{aligned}$$

Hence, we can decompose $\hat{\mathcal{X}}^*_{1,T} - \mathcal{X}^*_{1,T} = S_6 + S_7$, where

$$\begin{aligned}
S_6 &= \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)\Delta x_{t+j-i}(\hat{x}^*_{t+j-1} - \hat{x}^*_j)}{B^{3/2}T^{1/2}}, \\
S_7 &= \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{i=1}^{p}(\theta_i - \hat{\theta}_i)\Delta x^*_{t+j}(x_{t+j-1-i} - x_{j-1})}{B^{3/2}T^{1/2}}
\end{aligned}$$

and $\hat{\mathcal{X}}^*_{2,T} - \mathcal{X}^*_{2,T} = S_8 + S_9$, where

$$S_8 = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(\sum_{i=0}^{p}(\theta_i - \hat{\theta}_i)(x_{t+j-1-i} - x_{j-i})^2}{B^2T},$$
$$S_9 = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{i=0}^{p} 2(\theta_i - \hat{\theta}_i)(x^*_{t+j-1} - x^*_j)(x_{t+j-1-i} - x_{j-i})}{B^2T}.$$

Let $A_i = p(\theta_i - \hat{\theta}_i)$. From Lemma 1.5, it follows that $\max_{1\le i\le p} A_i = o_P(1)$. We then rearrange the terms such that

$$S_6 = \frac{1}{p}\sum_{i=1}^{p} A_i S_{6,i}, \quad S_7 = \frac{1}{p}\sum_{i=1}^{p} A_i S_{7,i}, \quad S_8 = \frac{1}{p^2}\sum_{i_1,i_2=1}^{p} A_{i_1}A_{i_2}S_{8,i_1,i_2}, \quad S_9 = \frac{1}{p}\sum_{i=1}^{p} A_i S_{9,i},$$

where

$$S_{6,i} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(\hat{x}^*_{t+j-1} - \hat{x}^*_j)u_{t+j-i}}{B^{3/2}T^{1/2}}, \quad S_{7,i} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}(x_{t+j-1} - x_j)\epsilon_{t+j}}{B^{3/2}T^{1/2}},$$
$$S_{8,i_1,i_2} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k_1,k_2=1}^{t-1} u_{k_1+j-i_1}u_{k_2+j-i_2}}{B^2T},$$
$$S_{9,i} = \frac{\sum_{j=1}^{T-B}\sum_{t=2}^{B}\sum_{k_1,k_2=1}^{t-1} \epsilon_{k_1+j}u_{k_2+j-i}}{B^2T}$$

with $0 \le i, i_1, i_2 \le p$. From the Cauchy-Schwarz inequality, it follows that

$$\begin{aligned} E[|S_{6,i}|] &\le \sum_{i_2=0}^{p}\sum_{m_1,m_2=0}^{\infty}\sum_{t=2}^{B}\psi_{m_1}\psi_{m_2}\frac{|\theta_{i_2} + o(1)|}{B^{3/2}T^{1/2}}\sqrt{\Big(\sum_{j=1}^{T-B}\sum_{k=1}^{t-1}\epsilon_{t+j-i_1-m_1}\epsilon_{k+j-i_2-m_2}\Big)^2} \\ &\le \frac{1}{B^{3/2}T^{1/2}}C^3B\sqrt{TBC^4} = O(1) \end{aligned}$$

and Jensen's inequality yields

$$\begin{aligned} E[|S_{7,i}|] &\le \frac{1}{B^{3/2}T^{1/2}}\sum_{m=0}^{\infty}\sum_{t=2}^{B}\psi_m\sqrt{E\Big[\Big(\sum_{j=1}^{T-B}\sum_{k=1}^{t-1}\epsilon_{t+j}\epsilon_{k+j-m}\Big)^2\Big]} \\ &\le \frac{1}{B^{3/2}T^{1/2}}CB\sqrt{TBC^4} = O(1). \end{aligned}$$

From $E[(\sum_{j=1}^{T-B}\sum_{k_2=1}^{t-1}\epsilon_{k_1+j-i_1-m_1}\epsilon_{k_2+j-i_2-m_2})^2] \leq TBC^4$ and from Jensen's inequality, it follows that

$$\begin{aligned}E[|S_{8,i_1,i_2}|] &\leq \frac{1}{B^2T}\sum_{t=2}^{B}\sum_{k_1=1}^{t-1}\sum_{m_1,m_2=0}^{\infty} E\Big[\Big|\sum_{j=1}^{T-B}\sum_{k_2=1}^{t-1}\psi_{m_1}\psi_{m_2}\epsilon_{k_1+j-i_1-m_1}\epsilon_{k_2+j-i_2-m_2}\Big|\Big] \\ &\leq C^4 = O(1).\end{aligned}$$

Finally, for the last term, we obtain

$$\begin{aligned}E[|S_{9,i}|] &\leq \frac{2}{B^2T}\sum_{t=2}^{B}\sum_{k_1=1}^{t-1}\sum_{m=0}^{\infty} E\Big[\Big|\sum_{j=1}^{T-B}\sum_{k_2=1}^{t-1}\psi_m\epsilon_{k_1+j}\epsilon_{k_2+j-i-m}\Big|\Big] \\ &\leq \frac{2}{B^2T}\sum_{t=2}^{B}\sum_{k_1=1}^{t-1}\sum_{m=0}^{\infty}|\psi_m|\sqrt{\sum_{j=1}^{T-B}\sum_{k_2=1}^{t-1}\sigma^2_{k_1+j}\sigma^2_{k_2+j-i-m}} \leq 2C^3 = O(1).\end{aligned}$$

Then, for any $0 \leq i, i_1, i_2 \leq p$, we have $S_{6,i} = O_P(1)$, $S_{7,i} = O_P(1)$, $S_{8,i_1,i_2} = O_P(1)$, and $S_{9,i} = O_P(1)$, which follows from Markov's inequality. By the law of large numbers, it then follows that $S_6 + S_7 = o_P(1)$ and $S_8 + S_9 = o_P(1)$ Consequently, the triangle inequality implies that $|\hat{\mathcal{Y}}^*_{1,T} - \mathcal{X}^*_{1,T}| \leq |\hat{\mathcal{Y}}^*_{1,T} - \hat{\mathcal{X}}^*_{1,T}| + |\hat{\mathcal{X}}^*_{1,T} - \mathcal{X}^*_{1,T}| = o_P(1)$ as well as $|\hat{\mathcal{Y}}^*_{2,T} - \mathcal{X}^*_{2,T}| \leq |\hat{\mathcal{Y}}^*_{2,T} - \hat{\mathcal{X}}^*_{2,T}| + |\hat{\mathcal{X}}^*_{2,T} - \mathcal{X}^*_{2,T}| = o_P(1)$.

Proof of Theorem 1.5

First, note that

$$\hat{u}_t = \hat{y}^*_t - \hat{\rho}^*\hat{y}_{t-1} = \Delta\hat{y}^*_t + (1-\hat{\rho}^*)\hat{y}^*_{t-1} = \Delta\hat{d}^*_t + \Delta\hat{x}^*_t + (1-\hat{\rho}^*)\hat{y}^*_{t-1},$$

where $\Delta\hat{x}^*_t = \epsilon_t + (\rho-1)x^*_{t-1} + \sum_{i=1}^{p}(\theta_i - \hat{\beta}_i)x_{t-i}$. Then, for all $s \in [0,1]$,

$$\begin{aligned}\hat{u}_{\lfloor sT\rfloor} &= \epsilon_{\lfloor sT\rfloor} + \Delta\hat{d}^*_{\lfloor sT\rfloor} + (\rho-1)x^*_{\lfloor sT\rfloor-1} + \sum_{i=1}^{p}(\theta_i - \hat{\beta}_i)x_{\lfloor sT\rfloor-i} + (1-\hat{\rho}^*)\hat{y}^*_{\lfloor sT\rfloor-1} \\ &= \epsilon_{\lfloor sT\rfloor} + o_P(1),\end{aligned}$$

which follows from the fact that $p = o(T^{1/4})$, $\Delta d_{\lfloor sT\rfloor} = \mathcal{O}(T)$, $(\rho-1) = O(B^{-1/2}T^{-1/2})$ as well as $\Delta x_{\lfloor sT\rfloor} = O_P(1)$, $\max_{1\leq i\leq p} p|\hat{\beta}_i - \theta_i| = o_P(1)$, $(\hat{\rho}^* - 1) = O_P(B^{-1/2}T^{-1/2})$, and

$\hat{y}_{\lfloor sT \rfloor} = O(T^{1/2})$. Let $\overline{\epsilon}_j = \frac{1}{B}\sum_{k=1}^{B} \epsilon_{j+k}$. Then, for any $t = 1, \ldots, B$, we have

$$\sum_{j=1}^{B} \Big(\hat{u}_{t+j} - \frac{1}{B}\sum_{k=1}^{B} \hat{u}_{j+k} \Big)^2 = \sum_{j=1}^{T-B} (\epsilon_{t+j} - \overline{\epsilon}_j)^2 + o_P(T) = \sum_{j=1}^{T-B} \epsilon_{t+j}^2 + o_P(T).$$

Analogously, $\sum_{j=1}^{\lfloor sT \rfloor} (\hat{u}_j - \overline{\hat{u}})^2 = \sum_{j=1}^{\lfloor sT \rfloor} \epsilon_j^2 + o_P(T)$ for all $s \in [0,1]$. Then, as $B, T \to \infty$ and $B/T \to 0$, we obtain

$$\frac{1}{(T-B)(B-1)} \sum_{j=1}^{T-B} \sum_{t=1}^{B} \Big(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^{B} \hat{u}_{j+k} \Big)^2 = \int_0^1 \sigma^2(r)\,\mathrm{d}r + o_P(1),$$
$$\frac{1}{(T-B)(B-1)} \sum_{j=1}^{T-B} \sum_{t=1}^{B} (\hat{u}_j - \overline{\hat{u}})^2 \Big(\hat{u}_{j+t} - \frac{1}{B}\sum_{k=1}^{B} \hat{u}_{j+k} \Big)^2 = \int_0^1 \sigma^4(r)\,\mathrm{d}r + o_P(1)$$

and the consistency of $\hat{\sigma}^{*2}_{\mathrm{sb}}$ and $\hat{\kappa}^{*2}$ follows by Slutsky's theorem. For (b), note that fixed-b asymptotics yields

$$\frac{1}{T} \sum_{j=1}^{\lfloor sT \rfloor} \Big(\hat{u}_j - \frac{1}{\lfloor sT \rfloor} \sum_{k=1}^{\lfloor sT \rfloor} \hat{u}_k \Big)^2 = \int_0^s \sigma^2(r)\,\mathrm{d}r + o_P(1), \quad s \in [0,1],$$

as $B/T \to b$, $0 < b < 1$ and $B, T \to \infty$. The consistency of $\hat{\sigma}^{*2}_{\mathrm{fb}}$ then follows with $s = 1$. Furthermore, Slutsky's theorem implies that (ii) holds pointwise. The uniform convergence then follows by Dini's theorem since both $\hat{\eta}(s)$ and $\eta(s)$ are continuous, monotone, and bounded.

Finally, since the pre-whitened numerator and denominator statistics $(\mathcal{X}^*_{1,T}, \mathcal{X}^*_{2,T})$ under Assumption 1.4 have the same properties as $(\mathcal{X}_{1,T}, \mathcal{X}_{2,T})$ under Assumption 1.3, the assertion follows with Lemma 1.6.

Central Limit Theorems

Lemma 1.7 (FCLTs). *Let $\{u_t\}_{t\in\mathbb{N}}$ be independently distributed with $E[u_t] = 0$, $E[u_t^2] = \sigma_t^2$ and $E[u_t^4] < \infty$, and let $\sigma_t = \sigma(t/T)$, where the function $\sigma(r)$ is càdlàg, non-stochastic, strictly positive, and bounded. Let $\rho = 1 - c/\sqrt{BT}$ with $c \geq 0$. Let $\hat{\eta}(r)$ be a consistent estimator for the variance profile $\eta(r)$, and let $\tilde{x}_{\lfloor rT \rfloor} = x_{\lfloor \hat{\eta}^{-1}(r)T \rfloor}$ and $\tilde{u}_{\lfloor rT \rfloor} = u_{\lfloor \hat{\eta}^{-1}(r)T \rfloor}$. Let $W(r)$ be a standard Brownian motion, and let $W_\eta(r) = W(\eta(r))$ be its variance-transformed counterpart (see Davidson 1994, p.486). An Ornstein-Uhlenbeck process is defined by $J_c(r) = \int_0^r e^{-(r-s)c} dW(s)$, and the variance-transformed Ornstein-Uhlenbeck*

process is defined by $J_{c,\eta}(r) = \int_0^r e^{-(r-s)c} dW_\eta(s)$. *Furthermore, let*

$$X_T(r) = \sum_{k=1}^{\lfloor rT \rfloor} \frac{u_k}{\sqrt{T}}, \quad Y_T(r) = \frac{x_{\lfloor rT \rfloor}}{\sqrt{T}}, \quad \tilde{X}_T(r) = \sum_{k=1}^{\lfloor rT \rfloor} \frac{\tilde{u}_k}{\sqrt{T}}, \quad \tilde{Y}_T(r) = \frac{\tilde{x}_{\lfloor rT \rfloor}}{\sqrt{T}}.$$

Let $B/T \to b$, *where* $0 < b < 1$. *Then,* (*a*) $X_T \Rightarrow \overline{\sigma} W_\eta$, (*b*) $Y_T \Rightarrow \overline{\sigma} J_{c/b,\eta}$, (*c*) $\tilde{X}_T \Rightarrow \overline{\sigma} W$, *and* (*d*) $\tilde{Y}_T \Rightarrow \overline{\sigma} J_{c/b}$, *as* $B, T \to \infty$, *where* $\overline{\sigma}^2 = \int_0^1 \sigma^2(r)\,\mathrm{d}r$ *is the average variance, and where "* $\Rightarrow$*" denotes weak convergence on the càdlàg space* $D[0,1]$ *together with a suitable norm.*

Proof. Result (a) follows from Lemmas 1 and 2 in Cavaliere (2005). To show (b), we set $u_0 = x_0$ for convenience. Note that a Taylor expansion around 0 yields $e^{-x} = 1 - x + o(x)$, which implies that $\rho = 1 - c/\sqrt{BT} = \exp(-c/\sqrt{BT}) + o(1/\sqrt{BT})$. Then, with the continuous mapping theorem, we obtain

$$\begin{aligned} \frac{1}{\overline{\sigma}\sqrt{T}} x_{\lfloor rT \rfloor} &= \sum_{k=0}^{\lfloor rT \rfloor} \rho^{\lfloor rT \rfloor - k} \frac{u_k}{\overline{\sigma}\sqrt{T}} = \sum_{k=0}^{\lfloor rT \rfloor} e^{-(\lfloor rT \rfloor - k)c/\sqrt{BT}} \frac{u_k}{\overline{\sigma}\sqrt{T}} + o_P(1) \\ &= \int_0^r e^{-(r-s)c/b} dX_T(s) + o_P(1) \Rightarrow \int_0^r e^{-(r-s)c/b} dW_\eta(s) = J_{c/b,\eta}(r). \end{aligned}$$

Result (c) follows by Theorem 1 in Cavaliere and Taylor (2008), and (d) follows analogously to the proof of (b). □

Lemma 1.8 (CLT for md-arrays)**.** *Let* $\{\{q_{j,T}\}_{1 \le j \le T}\}_{T \in \mathbb{N}}$ *be a martingale difference array with* $V_T = Var[\sum_{j=1}^T q_{j,T}] < \infty$ *and* $\max_{1 \le j \le T} |q_{j,T}| \xrightarrow{p} 0$. *Then, as* $T \to \infty$,

$$\frac{1}{\sqrt{V_T}} \sum_{j=1}^T q_{j,T} \xrightarrow{d} \mathcal{N}(0,1).$$

Proof. The result follows from Theorem 24.3 in Davidson (1994). □

Chapter 2

Backward CUSUM for Testing and Monitoring Structural Change

2.1 Introduction

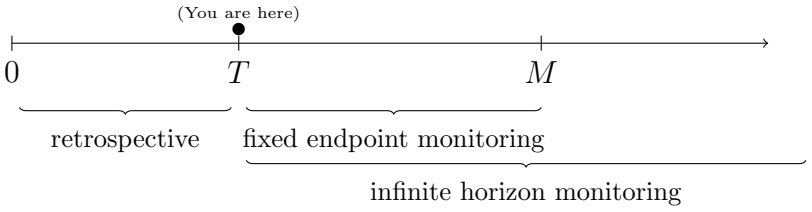
Cumulative sums have become a standard statistical tool for testing and monitoring structural changes in time series models. The CUSUM test was introduced by Brown et al. (1975) as a structural break test for the coefficient vector in the linear regression model $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + u_t$ with time index t. Under the null hypothesis, there is no structural change, such that $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t = 1, \ldots, T$, while, under the alternative hypothesis, the coefficient vector changes at unknown time T^*, where $1 < T^* \leq T$.

Sequential tests, such as the CUSUM test, consist of a detector statistic and a critical boundary function. The CUSUM detector sequentially cumulates standardized one-step ahead forecast errors, which are also referred to as recursive residuals. The detector is evaluated for each time point within the testing period, and, if its path crosses the boundary function at least once, the null hypothesis is rejected.

A variety of retrospective structural break tests have been proposed in the literature. Krämer et al. (1988) investigated the CUSUM test of Brown et al. (1975) under a more general setting. The MOSUM tests by Bauer and Hackl (1978) and Chu et al. (1995) are based on a moving time window of fixed length. A CUSUM test statistic that cumulates OLS residuals was proposed by Ploberger and Krämer (1992), and Ploberger et al. (1989) presented a fluctuation test based on a sequence of OLS estimates. Kuan and Hornik (1995) studied generalized fluctuation tests. Andrews (1993) and Andrews and Ploberger (1994) proposed a sup-Wald test, and the tests by Nyblom (1989) and Hansen (1992) consider maximum likelihood scores instead of residuals.

Since the seminal work of Chu et al. (1996), increasing interest has been focused on monitoring structural stability in real time. Sequential monitoring procedures consist of a detector statistic and a boundary function that are evaluated for periods beyond some historical time span $\{1, 2, \ldots, T\}$. It is assumed that there is no structural change within the historical time period. The monitoring time span with $t > T$ can either have a fixed endpoint $M < \infty$ or an infinite horizon (see Figure 2.1). In the fixed endpoint setting, the monitoring period starts at $T+1$ and ends at M, while the boundary function depends on the ratio $m = M/T$. This setting is suitable if the length of the monitoring period is known in advance. In case of an infinite horizon, the monitoring time span does not need to be specified before the monitoring procedure starts. The null hypothesis of no structural change is rejected whenever the path of the detector crosses some critical boundary function for the first time. Leisch et al. (2000), Zeileis et al. (2005), and Wied and Galeano (2013) proposed CUSUM-based monitoring procedures for a fixed endpoint, whereas Chu et al. (1996), Horváth et al. (2004), and Aue et al. (2006) considered an infinite monitoring horizon.

Figure 2.1: Retrospective testing and monitoring



A drawback of the conventional retrospective CUSUM test is its low power, whereas the conventional monitoring CUSUM procedure exhibits large detection delays. This is due to the fact that the pre-break recursive residuals are uninformative, as their expectation is equal to zero up to the break date, while the recursive residuals have a non-zero expectation after the break. Hence, the cumulative sums of the recursive residuals typically contain a large number of uninformative residuals that only add noise to the statistic. In contrast, if one cumulates the recursive residuals backwards from the end of the sample to the beginning, the cumulative sum collects the informative residuals first, and the likelihood of exceeding the critical boundary will typically be larger than when cumulating residuals from the beginning onwards. In this paper, we show that such backward CUSUM tests may indeed have a much higher power and lower detection delay than the conventional

forward CUSUM tests.

Another way of motivating the backward CUSUM testing approach is to consider the simplest possible situation, where, under the null hypothesis, it is assumed that the process is generated as $y_t = \beta + u_t$, with β and $\sigma^2 = Var(u_t)$ assumed to be known. We are interested in testing the hypothesis, that at some time period T^*, the mean changes to some unknown value $\beta^* > 0$. To test this hypothesis, we introduce the dummy variable D_t^*, which is unity for $t \geq T^*$ and zero elsewhere. For this one-sided testing problem, there exists a uniform most powerful (UMP) test statistic, which is the t-statistic of the hypothesis $\delta = 0$ in the regression $(y_t - \beta) = \delta D_t^* + u_t$:

$$\tau_{T^*} = \frac{1}{\sigma\sqrt{T - T^* - 1}} \sum_{t=T^*}^{T} (y_t - \beta).$$

If β is unknown, we may replace it by the full sample mean $\overline{y}$, resulting in the backward cumulative sum of the OLS residuals from period T through T^*. Note that if T^* is unknown, the test statistic is computed for all possible values of T^*, whereas the starting point of the backward cumulative sum T remains constant. Since the sum of the OLS residuals is zero, it follows that the test is equivalent to a test based on the forward cumulative sum of the OLS residuals. In contrast, if we replace β with the recursive mean $\overline{\mu}_{t-1} = (t-1)^{-1} \sum_{i=1}^{t-1} y_t$, we obtain a test statistic based on the backward cumulative sum of the recursive residuals (henceforth, backward CUSUM). In this case, however, the test is different from a test based on the forward cumulative sum of the recursive residuals (henceforth, forward CUSUM). This is due to the fact that the sum of the recursive residuals is an unrestricted random variable. Accordingly, the two versions of the test may have quite different properties. In particular, it turns out that the backward CUSUM is much more powerful than the standard forward CUSUM at the end of the sample. Accordingly, this version of the CUSUM test procedure is better suited for the purpose of real-time monitoring, where it is crucial to be powerful at the end of the sample.

Furthermore, the conventional CUSUM test has no power against alternatives that do not affect the unconditional mean of y_t. In order to obtain tests that have power against breaks of this kind, we extend the existing invariance principle for recursive residuals to a multivariate version and consider a vector-valued CUSUM process instead of the univariate CUSUM detector. For both retrospective testing and monitoring, we propose a vector-valued sequential statistic in the fashion of the score-based cumulative sum statistic of Nyblom (1989) and Hansen (1992). The application of a vector norm then yields a detector and a sequential test, that has power against a much larger class of structural

breaks.

In Section 2.2, the limiting distribution of the multivariate CUSUM process is derived under both the null hypothesis and local alternatives. Section 2.3 introduces the forward CUSUM, the backward CUSUM, and the stacked backward CUSUM tests for both retrospective testing and monitoring. While the backward CUSUM is only defined for $t \leq T$ and can thus be implemented only for retrospective testing, the stacked backward CUSUM cumulates recursive residuals backwardly in a triangular scheme and is therefore suitable for real-time monitoring. In Section 2.4, the local powers of the tests are compared. In the retrospective setting, the powers of the backward CUSUM and the stacked backward CUSUM tests are substantially higher than that of the the conventional forward CUSUM test if a single break occurs after one third of the sample size. In the case of monitoring, the detection delay of the stacked backward CUSUM under local alternatives is shown to be much lower than that of the monitoring CUSUM detector by Chu et al. (1996). Furthermore, simulated critical values as well as Monte Carlo simulation results are presented. Finally, Section 2.5 concludes.

2.2 The multivariate CUSUM process

We consider the multiple linear regression model

$$y_t = \boldsymbol{x}_t' \boldsymbol{\beta}_t + u_t, \quad t \in \mathbb{N},$$

where $\boldsymbol{x}_t = (1, x_{t2}, \ldots, x_{tk})'$ is the vector of regressor variables, and y_t is the dependent variable. The $k \times 1$ vector of regression coefficients $\boldsymbol{\beta}_t$ depends on the time index t, and u_t is an error term. Let $\{(y_t, \boldsymbol{x}_t')',\ 1 \leq t \leq T\}$ be the set of historical observations, such that the time point T divides the time horizon into the retrospective time period $1 \leq t \leq T$ and the monitoring period $t > T$. We impose the following assumptions on the regressors and the error term, which are common in the literature on CUSUM statistics and also include the case of lagged dependent variables (see e.g. Krämer et al. 1988):

Assumption 2.1. *(a) Let* $\boldsymbol{C}_T = T^{-1} \sum_{t=1}^{T} \boldsymbol{x}_t \boldsymbol{x}_t'$ *be the empirical covariance matrix, and let* $\|\cdot\|_M$ *denote some matrix norm. Then,* $\operatorname{plim}_{T\to\infty} \|\boldsymbol{C}_T - \boldsymbol{C}\|_M = 0$*, where* $\boldsymbol{C}$ *is a positive definite* $k \times k$ *matrix. Furthermore, there exists some* $\delta > 0$ *such that* $\lim_{T\to\infty} T^{-1} \sum_{t=1}^{T} E|x_{tj}|^{2+\delta} < \infty$ *for all* $j = 2, \ldots, k$.

(b) Let $\mathcal{F}_t$ *be the* σ*-algebra generated by* $\{(\boldsymbol{x}_{i+1}', u_i)',\ i \leq t\}$*. The error process* $\{u_t\}$ *is a martingale difference sequence with respect to* $\mathcal{F}_t$*, where* $E[u_t|\mathcal{F}_{t-1}] = 0$ *and*

$$E[u_t^2|\mathcal{F}_{t-1}] = \sigma^2 \ \textit{with}\ 0 < \sigma^2 < \infty.$$

Recursive residuals for linear regression models were introduced by Brown et al. (1975) as standardized one-step ahead forecast errors. Let $\widehat{\boldsymbol{\beta}}_{t-1} = \big(\sum_{i=1}^{t-1} \boldsymbol{x}_i \boldsymbol{x}_i'\big)^{-1} \big(\sum_{i=1}^{t-1} \boldsymbol{x}_i y_i\big)$ be the OLS estimator at time $t-1$. The recursive residuals are given by

$$w_t = \frac{y_t - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_{t-1}}{\sqrt{1 + \boldsymbol{x}_t' (\sum_{i=1}^{t-1} \boldsymbol{x}_i \boldsymbol{x}_i')^{-1} \boldsymbol{x}_t}}, \quad t \geq k+1,$$

and $w_t = 0$ for $t = 1, \ldots, k$.

For testing against structural changes in the regression coefficient vector, Brown et al. (1975) introduced the sequential statistic $Q_t = (\widehat{\sigma}^2 T)^{-1/2} \sum_{j=1}^{t} w_j$ for $t = 1, \ldots, T$, where $\widehat{\sigma}^2$ is a consistent estimator for σ^2. In the monitoring context, Chu et al. (1996) considered the detector statistic $Q_t - Q_T$ for $t > T$. The limiting behavior of the underlying empirical process has been thoroughly analyzed in the literature. Under $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, Sen (1982) showed that $Q_{\lfloor rT \rfloor} = (\widehat{\sigma}^2 T)^{-1/2} \sum_{j=1}^{\lfloor rT \rfloor} w_j$ converges weakly and uniformly to a standard Brownian motion $W(r)$ for $r \in [0,1]$, while Horváth et al. (2004) proved the same result for $r \geq 1$. Ploberger and Krämer (1990) studied local alternatives of the form $H_1 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0 + T^{-1/2} \mathbf{g}(t/T)$, where $\mathbf{g}(r)$ is piecewise constant and bounded. Let $\boldsymbol{\mu} = \lim_{T\to\infty} (\overline{\boldsymbol{x}}_1, \ldots, \overline{\boldsymbol{x}}_k)'$ be the mean regressor, where $\overline{\boldsymbol{x}}_j$ is the sample mean of the j-th component of the regressors, and let

$$\mathbf{h}(r) = \frac{1}{\sigma} \int_0^r \mathbf{g}(z)\, \mathrm{d}z - \frac{1}{\sigma} \int_0^r \int_0^z \frac{1}{z} \mathbf{g}(v)\, \mathrm{d}v\, \mathrm{d}z. \tag{2.1}$$

The authors showed that $Q_{\lfloor rT \rfloor}$ converges weakly and uniformly to $W(r) + \boldsymbol{\mu}' \mathbf{h}(r)$ for $r \in [0,1]$. As noted by Krämer et al. (1988), if the break vector $\mathbf{g}(r)$ is orthogonal to $\boldsymbol{\mu}$, the limiting distributions under H_0 and H_1 coincide. Hence, if a break in the coefficient vector does not affect the unconditional mean of y_t, then the CUSUM tests of Brown et al. (1975) and Chu et al. (1996) have no power against such an alternative.

Accordingly, for any fixed relative endpoint $m < \infty$, we consider a multivariate cumulative sum process of recursive residuals, which is defined as

$$\mathbf{Q}_T(r) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t, \qquad r \in [0, m].$$

Following Krämer et al. (1988), the consistent estimator $\widehat{\sigma}^2 = (T-k-1)^{-1} \sum_{j=1}^{T} (w_j - \overline{w})^2$ is considered. Note that $\mathbf{Q}_T(r)$ is a vector of piecewise constant processes, where its

domain can be divided into the retrospective time period $r \in [0, 1]$ and the monitoring period $r \in (1, m]$. Hence, each component of $\mathbf{Q}_T(r)$ is in the space $D([0, m])$ of càdlàg functions on the domain $[0, m]$, and $\mathbf{Q}_T(r)$ is an element of the k-fold product space $D([0, m])^k = D([0, m]) \times \ldots \times D([0, m])$. The space is equipped with the Skorokhod metric (see Billingsley 1999, p.166 and p.244), and the symbol "$\Rightarrow$" denotes weak convergence with respect to this metric. The result presented below summarizes the limiting behavior of $\mathbf{Q}_T(r)$ for both the retrospective and the monitoring time period under both H_0 and H_1:

Theorem 2.1. *Let $\{(\boldsymbol{x}_t, u_t)\}_{t\in\mathbb{N}}$ satisfy Assumption 2.1, let $\mathbf{g}(r)$ be piecewise constant and bounded, and let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0 + T^{-1/2}\mathbf{g}(t/T)$ for all $t \in \mathbb{N}$. Then, for any fixed and positive $m < \infty$,*

$$\mathbf{Q}_T(r) \Rightarrow \mathbf{W}(r) + \boldsymbol{C}^{1/2}\mathbf{h}(r), \quad r \in [0, m],$$

as $T \to \infty$, where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions and $\mathbf{h}(r)$ is defined as in (2.1).

Note that the function $\mathbf{g}(r)$ is constant if and only if $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$. Under H_0, we then obtain $\boldsymbol{C}^{1/2}\mathbf{h}(r) = \mathbf{0}$, and thus $\mathbf{Q}_T(r) \Rightarrow \mathbf{W}(r)$. By contrast, under a local alternative with a non-constant break function $\mathbf{g}(r)$, it follows that $\mathbf{h}(r)$ is non-zero, and, consequently, $\boldsymbol{C}^{1/2}\mathbf{h}(r)$ is non-zero, since $\boldsymbol{C}^{1/2}$ is positive definite. The limiting distributions of $\mathbf{Q}_T(r)$ under both H_0 and H_1 thus coincide only for the trivial case where $\mathbf{g}(r)$ is constant. Therefore, tests that are based on $\mathbf{Q}_T(r)$ have power against a larger class of alternatives than the tests of Brown et al. (1975) and Chu et al. (1996).

2.3 CUSUM detectors

In this section, we consider sequential tests for both retrospective testing and monitoring that are based on the multivariate CUSUM processes $\mathbf{Q}_T(r)$. The null hypothesis of no structural change in the regression coefficient vector is formulated as $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathcal{T}$, where the testing period is given by

$$\mathcal{T} = \begin{cases} \{t \in \mathbb{N} : \ 1 \leq t \leq T\} & \text{in the retrospective context,} \\ \{t \in \mathbb{N} : \ T < t \leq \lfloor mT \rfloor\} & \text{in the monitoring context.} \end{cases}$$

The sequential tests consist of a detector statistic and a critical boundary function, in which the detector is evaluated for each time point within the testing period, and, if its

path crosses the boundary function at least once, the null hypothesis is rejected. While $m < \infty$ is fixed in advance when using the fixed endpoint monitoring scheme, in the infinite horizon monitoring context, m can be set arbitrarily high and can be even increased during the monitoring process. Furthermore, the non-contamination assumption $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for the historical time period $t = 1, \ldots, T$ is imposed in the monitoring context. While the forward CUSUM detectors for retrospective testing and monitoring are discussed in Section 2.3.1, we introduce the backward CUSUM detector in Section 2.3.2 and the stacked backward CUSUM detectors in Section 2.3.3. We make the following assumption on the boundary function:

Assumption 2.2. *The boundary function is of the form $b(r) = \lambda_\alpha \cdot d(r)$, where λ_α denotes the critical value, which depends on the significance level α, and $d(r)$ is a continuous function that satisfies $d(r) > C > 0$ for some constant C for all $r \geq 0$. Furthermore, $\lim_{r\to\infty} \sqrt{r \ln(\ln(r))}/d(r) = 0$.*

While continuity and positivity of the boundary ensures the applicability of the continuous mapping theorem, the latter condition is motivated by the law of the iterated logarithm and ensures, that the test statistics and its limits are bounded in probability.

2.3.1 Forward CUSUM

As an extension of the univariate CUSUM detector by Brown et al. (1975) we consider the multivariate retrospective CUSUM detector

$$\mathbf{Q}_{t,T} = \mathbf{Q}_T\left(\tfrac{t}{T}\right) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=1}^{t} \boldsymbol{x}_t w_j, \quad 1 \leq t \leq T.$$

The vector-valued detector is inspired by Hansen (1992)'s score-based cumulative sum statistic. While Hansen (1992) considered OLS residuals and proposed averaging all entries of the vector-valued cumulative sum, we consider recursive residuals and the maximum vector entry. Let $\|\boldsymbol{a}\| = \max_{i=1,\ldots,k} |a_i|$, $\boldsymbol{a} = (a_1, \ldots, a_k)' \in \mathbb{R}^k$, be the maximum norm. The null hypothesis is rejected if the path of $\|\mathbf{Q}_{t,T}\|$ exceeds the critical boundary function $b_t = \lambda_\alpha \cdot d\big(t/T\big)$ for at least some time index within the retrospective testing period. The critical value λ_α determines the significance level α such that

$$\lim_{T\to\infty} P\Big(\|\mathbf{Q}_{t,T}\| \geq \lambda_\alpha \cdot d\big(\tfrac{t}{T}\big) \text{ for at least one index } t = 1, \ldots, T \,\Big|H_0\Big) = \alpha.$$

Let $\boldsymbol{\mathcal{M}}_Q^{\text{ret}} = \max_{1\leq t\leq T} \|\mathbf{Q}_{t,T}\|/d\big(t/T\big)$ denote the maximum statistic representation of the CUSUM detector. The above condition can be equivalently expressed as

$$\lim_{T\to\infty} P(\boldsymbol{\mathcal{M}}_Q^{\text{ret}} \geq \lambda_\alpha | H_0) = \alpha.$$

Hence, λ_α is the $(1-\alpha)$ quantile of the limiting null distribution of $\boldsymbol{\mathcal{M}}_Q^{\text{ret}}$. Note that $\boldsymbol{\mathcal{M}}_Q^{\text{ret}}$ together with the critical value λ_α defines a one-shot test that is equivalent to the sequential CUSUM test.

For real-time monitoring, we follow Chu et al. (1996) and define the multivariate retrospective CUSUM detector as

$$\mathbf{Q}_{t,T}^{\text{mon}} = \mathbf{Q}_T\big(\tfrac{t}{T}\big) - \mathbf{Q}_T(1) = \frac{1}{\widehat{\sigma}\sqrt{T}} \sum_{j=T+1}^{t} \boldsymbol{x}_t w_j, \quad t > T,$$

and H_0 is rejected if its maximum norm $\|\mathbf{Q}_{t,T}^{\text{mon}}\|$ exceeds the boundary $b_t = \lambda_\alpha \cdot d\big((t-T)/T\big)$ at least once for some $t > T$. For a fixed endpoint $M = \lfloor mT \rfloor$, where $1 < m < \infty$, let $\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} = \max_{T<t\leq mT} \|\mathbf{Q}_{t,T}^{\text{mon}}\|/d\big((t-T)/T\big)$ be the corresponding maximum statistic. Under the null hypothesis of no structural change, the following limiting results can be stated:

Theorem 2.2. *Under Assumptions 2.1 and 2.2, and under $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, it follows that*

(a) $\boldsymbol{\mathcal{M}}_Q^{ret} \xrightarrow{\mathcal{D}} \sup\limits_{r\in(0,1)} \dfrac{\|\mathbf{W}(r)\|}{d(r)}$,

(b) $\boldsymbol{\mathcal{M}}_{Q,m}^{mon} \xrightarrow{\mathcal{D}} \sup\limits_{r\in(0,m-1)} \dfrac{\|\mathbf{W}(r)\|}{d(r)} \overset{\mathcal{D}}{=} \sup\limits_{r\in(0,\frac{m-1}{m})} \dfrac{\|\boldsymbol{B}(r)\|}{(1-r)d\big(\frac{r}{1-r}\big)}, \quad 1 < m < \infty,$

(c) $\lim\limits_{m\to\infty} \lim\limits_{T\to\infty} P\big(\boldsymbol{\mathcal{M}}_{Q,m}^{mon} \leq \lambda\big) = P\Big(\sup\limits_{r\in(0,1)} \dfrac{\|\boldsymbol{B}(r)\|}{(1-r)d\big(\frac{r}{1-r}\big)} \leq \lambda \Big), \quad \lambda \in \mathbb{R},$

as $T \to \infty$, where $\boldsymbol{W}(r)$ is a vector of k independent standard Brownian motions and $\boldsymbol{B}(r)$ is a vector of k independent standard Brownian bridges.

Remark 2.1. Note that Theorem 2.2(c) does not provide a uniform convergence result for $m, T \to \infty$. Since the typical proof strategy using a functional central limit theorem together with the continuous mapping theorem cannot be applied in this case due to the unboundedness of the test statistic for $m = \infty$, Aue et al. (2006), Fremdt (2015), and Gösmann et al. (2019) imposed additional stochastic approximation conditions. Horváth

et al. (2004) showed that under the assumption that the errors and the regressors are independent across all leads and lags, the univariate CUSUM statistic of recursive residuals satisfies these stochastic approximation conditions. However, in practice, Theorem 2.2(c) is sufficient to define a conservative infinite horizon monitoring scheme for any arbitrary large but fixed $m < \infty$, so that its type I error does not exceed the significance level α asymptotically as $T \to \infty$. Let λ_α be the asymptotic critical value defined by

$$P\left(\sup_{r\in(0,1)} \frac{\|\boldsymbol{B}(r)\|}{(1-r)d\left(\frac{r}{1-r}\right)} \geq \lambda_\alpha \right) = \alpha.$$

Then, under the null hypothesis, it follows that $\lim_{T\to\infty} P(\boldsymbol{\mathcal{M}}^{\text{mon}}_{Q,m} \geq \lambda_\alpha) \leq \alpha$, for any $m \in (1, \infty)$, and $\lim_{m\to\infty} \lim_{T\to\infty} P(\boldsymbol{\mathcal{M}}^{\text{mon}}_{Q,m} \geq \lambda_\alpha) = \alpha$. △

While, for one-shot tests, the critical value determines the type I error, for sequential tests, the critical boundary involves two degrees of freedom. Besides the test size, which is controlled asymptotically by an appropriately chosen value for λ_α, the shape of the boundary determines the distribution of the first boundary crossing under the null hypothesis, which is also referred to as the "distribution of the size" (see Anatolyev and Kosenok 2018). Brown et al. (1975) suggested the linear boundary function

$$b(r) = \lambda_\alpha(1 + 2r), \tag{2.2}$$

which is our main benchmark. In this case, the retrospective maximum statistic satisfies

$$\max_{1\leq t\leq T} \frac{\|\mathbf{Q}_{t,T}\|}{1 + 2\left(\frac{t}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\mathbf{W}(r)\|}{1 + 2r}$$

under H_0, as $T \to \infty$, whereas, for the monitoring maximum statistic, we obtain

$$\max_{T+1\leq t\leq mT} \frac{\|\mathbf{Q}_{t,T}\|}{1 + 2\left(\frac{t}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,\frac{m-1}{m})} \frac{\|\boldsymbol{B}(r)\|}{1 + r} \leq \sup_{r\in(0,1)} \frac{\|\boldsymbol{B}(r)\|}{1 + r}. \tag{2.3}$$

The linear boundary is widely applied in practice, but, as already noted by Brown et al. (1975), the crossing probabilities cannot be constant for all potential relative crossing time points r. The authors argued that it is more natural to consider a boundary that is proportional to the standard deviation of the limiting process. Such a boundary is given by the radical function $b(r) = \lambda_\alpha\sqrt{r}$. As noted by Zeileis (2004), if there is a single break in the middle or at the end of the retrospective sample, there is no power gain using the radical boundary when compared to the linear boundary. Only in cases where a break occurs at the beginning of the sample, some increased power may be observed.

Another problem associated with the radical boundary is that it is not bounded away from zero. In order to obtain critical values and avoid size distortions, some trimming at the beginning of the sample in the fashion of the sup-Wald test by Andrews (1993) is necessary. For infinite horizon monitoring, Chu et al. (1996) also considered a boundary function of radical type, which is given by

$$b(r) = \sqrt{(r+1)\ln\left(\tfrac{r+1}{\alpha^2}\right)}. \tag{2.4}$$

The boundary is based on a result on boundary crossing probabilities for the path of Brownian motions. Robbins and Siegmund (1970) showed that

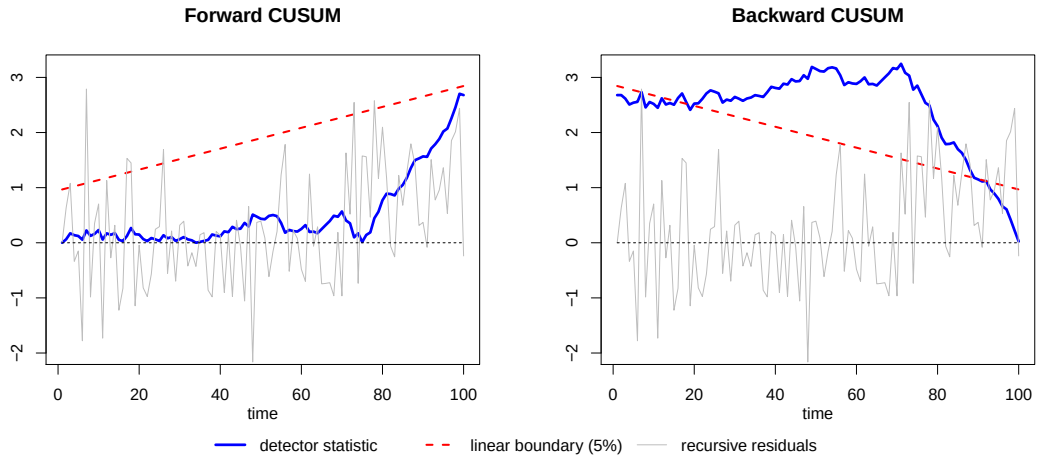
$$P\Big(|W(r)| \geq \sqrt{(r+1)\ln\left(\tfrac{r+1}{\alpha^2}\right)} \text{ for some } r \geq 0\Big) = \alpha,$$

and the univariate monitoring CUSUM detector together with the radical boundary by Chu et al. (1996) thus yields a sequential test that has size α, as $m \to \infty$. Anatolyev and Kosenok (2018) derived a theoretical boundary that yields a uniformly distributed size. However, their boundary has no closed form solution and is only valid for the univariate retrospective and fixed endpoint monitoring cases. Furthermore, simulations, which are omitted here, indicate that their approximative boundary does indeed yield a uniform size distribution, but that the CUSUM test performs uniformly worse in terms of power compared to the test when using the linear boundary of Brown et al. (1975). Note that in the context of infinite horizon monitoring the size cannot be uniformly distributed.

2.3.2 Backward CUSUM

An alternative approach is to cumulate the recursive residuals in reversed order. Suppose there is a single break point at time T^*. Then, $\{w_t,\ t < T^*\}$ are the residuals from the pre-break period, and $\{w_t,\ t \geq T^*\}$ are those from the post-break period. The pre-break residuals do not contain any information about the break and have mean zero. The partial sum process $T^{-1/2}\sum_{j=1}^{t} w_j$ has a random walk behavior for the pre-break period $t < T^*$, and cumulating those residuals brings nothing but noise to the detector statistic. In contrast, under a structural break, the post-break residuals have nonzero mean and reveal relevant information about a possible break. In order to focus on the post-break residuals, we consider backwardly cumulated partial sums of the form $T^{-1/2}\sum_{j=0}^{t-1} w_{T-j}$.

Figure 2.2: Illustrative example for the backward CUSUM with a break in the mean



Note: The process $y_t = \mu_t + u_t$, $t = 1, \dots, T$, is simulated for $T = 100$ with $\mu_t = 0$ for $t < 75$, $\mu_t = 1$ for $t \geq 75$, and i.i.d. standard normal innovations u_t. Since $k = 1$, the detectors are univariate, and the vector norm is simply the absolute value. The bold solid line paths are the trajectories of $|\mathbf{Q}_{t,T}|$ and $|\mathbf{BQ}_{t,T}|$. In the background, the recursive residuals are plotted. The dotted lines shows the linear boundary (2.2) with $\alpha = 5\%$ and $\lambda_\alpha = 0.948$.

We define the retrospective backward CUSUM detector as

$$\mathbf{BQ}_{t,T} = \mathbf{Q}_T(1) - \mathbf{Q}_T\left(\tfrac{t-1}{T}\right) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=t}^{T} \boldsymbol{x}_t w_j,$$

where $1 \leq t \leq T$. The null hypothesis is rejected if the path of $\|\mathbf{BQ}_{t,T}\|$ exceeds the boundary $b_t = \lambda_\alpha \cdot d\big((T-t-1)/T\big)$ for at least some time index t.

Theorem 2.3. *Under Assumptions 2.1 and 2.2, and under $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, it follows that*

$$\boldsymbol{\mathcal{M}}_{BQ}^{ret} = \max_{1 \leq t \leq T} \frac{\|\mathbf{BQ}_{t,T}\|}{d\left(\frac{T-t+1}{T}\right)} \xrightarrow{\mathcal{D}} \sup_{r \in (0,1)} \frac{\|\mathbf{W}(r)\|}{d(r)}$$

as $T \to \infty$, where $\boldsymbol{W}(r)$ is a vector of k independent standard Brownian motions.

Using the same boundary as for the retrospective CUSUM, the limiting null distributions of their maximum statistics coincide. A simple illustrative example of the detector paths together with the linear boundary of Brown et al. (1975) are depicted in Figure 2.2, in which a process with $k = 1$ and a single break in the mean at 3/4 of the sample is simulated.

Unlike the forward CUSUM detector, the backward CUSUM detector is not measurable with respect to the filtration of available information at time t and is therefore not suitable for a monitoring procedure. The path of $\|\mathbf{BQ}_{t,T}\|$ is only defined for $t \leq T$, as its endpoint

T is fixed.

2.3.3 Stacked backward CUSUM

In order to combine the advantages of the backward CUSUM with the measurability properties of the forward CUSUM for monitoring, we propose the stacked backward CUSUM detector. Let

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t) = \max_{1\leq s\leq t} \frac{\|\mathbf{Q}_T\big(\frac{t}{T}\big) - \mathbf{Q}_T\big(\frac{s-1}{T}\big)\|}{d\big(\frac{t-s+1}{T}\big)}$$

be the backward CUSUM statistic with endpoint t. The idea is to compute this statistic sequentially for each time point $t = 1, \ldots, T$, yielding $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(1)$, $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(2), \ldots, \boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(T)$. The stacked backward CUSUM statistic is the maximum among this sequence of backward CUSUM statistics. An important feature of this sequence is that it is measurable with respect to the filtration of information at time t and $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t)$ can thus be adapted for real-time monitoring.

The stacked backward CUSUM detector can be defined as

$$\mathbf{SBQ}_{s,t,T} = \mathbf{Q}_T\big(\tfrac{t}{T}\big) - \mathbf{Q}_T\big(\tfrac{s-1}{T}\big) = \frac{1}{\widehat{\sigma}\sqrt{T}} \boldsymbol{C}_T^{-1/2} \sum_{j=s}^{t} \boldsymbol{x}_t w_j, \quad 1 \leq s \leq t < \infty.$$

Since the upper and the lower summation index of $\mathbf{SBQ}_{s,t,T}$ are both flexible with $s \leq t$, this induces a triangular scheme. H_0 is rejected if $\|\mathbf{SBQ}_{s,t,T}\|$ exceeds the two-dimensional boundary $b_{s,t} = \lambda_\alpha \cdot d\big((t-s+1)/T\big)$ for some s and t with $1 \leq s \leq t \leq T$, or, equivalently, if the double maximum statistic

$$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}} = \max_{1\leq t\leq T} \boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t) = \max_{1\leq t\leq T} \max_{1\leq s\leq t} \frac{\|\mathbf{SBQ}_{s,t,T}\|}{d\big(\frac{t-s+1}{T}\big)}$$

exceeds λ_α.

The backward CUSUM maximum statistic $\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}(t)$ is itself a sequential statistic. Stacking all those maximum statistics on one another leads to an additional maximum and a double supremum in the limiting distribution. The stacked backward CUSUM uses the recursive residuals in a multiple way such that the set over which the maximum is taken has many more elements than the forward CUSUM and the backward CUSUM. For $t = 1$ only w_1 is cumulated, for $t = 2$ the residuals w_2 and w_1 are cumulated, for $t = 3$ we consider w_3, w_2, and w_1, and so forth. A similar procedure was proposed by Dette and Gösmann (2019) in the context of likelihood ratio (LR) tests for change point detection.

Their detector is given by the maximum of a triangular array of LR statistics, which also leads to a double maximum statistic.

The triangular detector can also be monitored on-line across all the time points $t > T$. The null hypothesis is rejected if $\|\mathbf{SBQ}_{s,t,T}\|$ exceeds $b_{s,t} = \lambda_\alpha \cdot d\big((t-s+1)/T\big)$ at least once for some s and t with $T < s \leq t$. Analogously to the retrospective case, let

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{mon}}(t) = \max_{T<s\leq t} \frac{\|\mathbf{SBQ}_{s,t,T}\|}{d\big(\frac{t-s+1}{T}\big)}$$

be the sequence of backward CUSUM maximum statistics for $t > T$, and let

$$\boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}} = \max_{T<t\leq\lfloor mT\rfloor} \boldsymbol{\mathcal{M}}_{BQ}^{\text{mon}}(t) = \max_{T<t\leq\lfloor mT\rfloor} \max_{T<s\leq t} \frac{\|\mathbf{SBQ}_{s,t,T}\|}{d\big(\frac{t-s+1}{T}\big)}$$

be its maximum statistic for some fixed endpoint $m \in (1,\infty)$.

Theorem 2.4. *Under Assumptions 2.1 and 2.2, and under $H_0 : \boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, it follows that*

(a) $$\boldsymbol{\mathcal{M}}_{SBQ}^{ret} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r)-\mathbf{W}(s)\|}{d(r-s)},$$

(b) $$\begin{aligned}\boldsymbol{\mathcal{M}}_{SBQ,m}^{mon} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,m-1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r)-\mathbf{W}(s)\|}{d(r-s)} \\ &\overset{\mathcal{D}}{=} \sup_{r\in(0,\frac{m-1}{m})} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)d\big(\frac{r-s}{(1-r)(1-s)}\big)}, \quad 1<m<\infty,\end{aligned}$$

(c) $$\lim_{m\to\infty}\lim_{T\to\infty} P\big(\boldsymbol{\mathcal{M}}_{SBQ,m}^{mon} \leq \lambda\big) = P\bigg(\sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)d\big(\frac{r-s}{(1-r)(1-s)}\big)} \leq \lambda\bigg),$$
$\lambda \in \mathbb{R}$,

as $T \to \infty$, where $\boldsymbol{W}(r)$ is a vector of k independent standard Brownian motions and $\boldsymbol{B}(r)$ is a vector of k independent standard Brownian bridges.

Analogously to the forward CUSUM, for the linear boundary of Brown et al. (1975), it follows that, for any $m \in (1,\infty)$,

$$\begin{aligned}\max_{T+1\leq t<mT} \max_{T\leq s\leq t-1} \frac{\|\mathbf{SBQ}_{s,t,T}^{\text{mon}}\|}{1+2(\frac{t-s}{T})} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,\frac{m-1}{m})} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)+2(r-s)} \\ &\leq \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)+2(r-s)} \qquad (2.5)\end{aligned}$$

under H_0, as $T \to \infty$.

2.4 Simulations

In this section, we compare both the asymptotic and finite sample properties of the tests. While in Section 2.4.1 local asymptotic power an local asymptotic mean delay curves are simulated, we present simulation results on the finite sample size and power in Section 2.4.2. Furthermore, asymptotic critical values for the tests are provided.

2.4.1 Local asymptotic power and delay

In order to illustrate the advantages of the backward CUSUM and the stacked backward CUSUM tests, we consider the simple model $y_t = \beta_t + u_t$ with a local break in the mean. Let the mean be given by $\beta_t = \beta^0 + T^{-1/2} g(t/T)$, where $g(r)$ is a piecewise constant and bounded function. Note that in this case the multivariate CUSUM process coincides with the univariate CUSUM process $Q_{\lfloor rT \rfloor}$. Furthermore, note that the covariance matrix $\boldsymbol{C}$ is equal to unity, and the vector norm for $k = 1$ is simply the absolute value. Theorem 2.1 yields $Q_T(r) \Rightarrow W(r) + h(r)$, where

$$h(r) = \frac{1}{\sigma} \int_0^r g(z)\,\mathrm{d}z - \frac{1}{\sigma} \int_0^r \int_0^z \frac{1}{z} g(v)\,\mathrm{d}v\,\mathrm{d}z,$$

and together with the continuous mapping theorem, it follows that

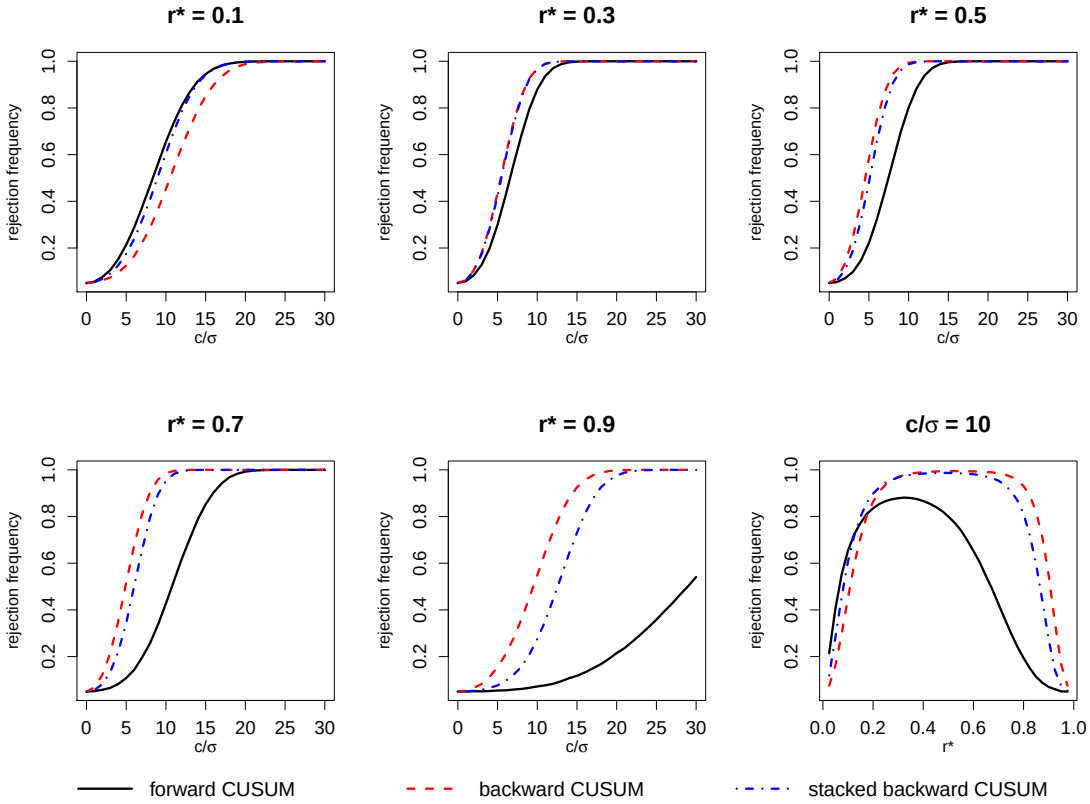
$$\begin{aligned}
\boldsymbol{\mathcal{M}}_{Q}^{\mathrm{ret}} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{|W(r) + h(r)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{BQ}^{\mathrm{ret}} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{|W(r) + h(1) - h(1-r)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{SBQ}^{\mathrm{ret}} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{|W(r) - W(s) + h(r) - h(s)|}{d(r-s)},
\end{aligned}$$

as $T \to \infty$. While, under H_0, the limiting distributions for the retrospective forward CUSUM and the retrospective backward CUSUM coincide, they differ from each other under the alternative. The maximum statistics in the fixed endpoint monitoring case satisfy

$$\begin{aligned}
\boldsymbol{\mathcal{M}}_{Q,m}^{\mathrm{mon}} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,m-1)} \frac{|W(r) + h(r+1) - h(1)|}{d(r)}, \\
\boldsymbol{\mathcal{M}}_{SBQ,m}^{\mathrm{mon}} &\xrightarrow{\mathcal{D}} \sup_{r\in(0,m-1)} \sup_{s\in(0,r)} \frac{|W(r) - W(s) + h(r+1) - h(s+1)|}{d(r-s)},
\end{aligned}$$

as $T \to \infty$.

Figure 2.3: Asymptotic local power curves for retrospective testing



Note: The plots show simulated local power curves. While, for the plots at the top and the first two plots at the bottom, the break location is fixed with $r^* \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ and local break sizes c/σ are shown on the x-axis, for the last plot, the local break size is fixed with $c/\sigma = 10$, and the breakpoint locations r^* are given on the x-axis. The linear boundary (2.2) is implemented for a significance level of $\alpha = 5\%$.

Generally, none of the tests can be shown to be uniformly more powerful in comparison to the other tests. However, we can compare the tests under particular alternatives. We consider a single break in the mean, where the break function is given by $g(r) = c \cdot 1_{\{r \geq r^*\}}$ and r^* denotes the break location. Then,

$$h(r) = \frac{c}{\sigma} \int_{r^*}^{r} \mathrm{d}z - \frac{c}{\sigma} \int_0^r \int_{r^*}^{z} \frac{1}{z} \,\mathrm{d}v \,\mathrm{d}z = \frac{cr^*}{\sigma} \int_{r^*}^{r} \frac{1}{z} \,\mathrm{d}z = \frac{cr^*(\ln(r) - \ln(r^*))1_{\{r \geq r^*\}}}{\sigma}.$$

Simulated asymptotic local power curves under the limiting distribution at a 5% significance level are presented in Figure 2.3 for the retrospective case. The Brownian motions are approximated on a grid of 1,000 equidistant points, and the linear boundary $d(r) = 1 + 2r$ is implemented. The size-adjusted rejection rates are obtained from 100,000 Monte Carlo repetitions for different break locations. The plots show that for a single break that is located after 15% of the sample size, the backward CUSUM and the stacked backward CUSUM clearly outperform the forward CUSUM in terms of power.

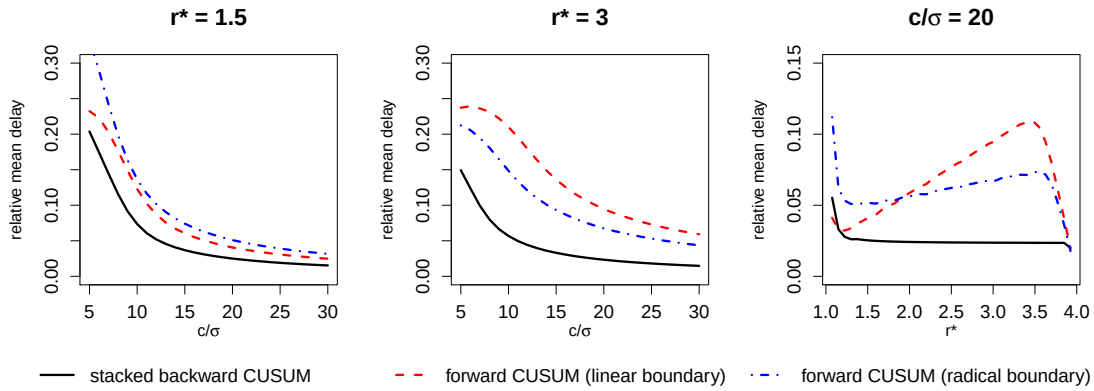
The backward CUSUM performs best for $r^* > 0.3$, while the stacked backward CUSUM outperforms the other two tests if the break is located at around 1/5 of the sample size.

For the monitoring case with fixed endpoint $m = 2$, the local power curves of the forward CUSUM test and the stacked backward CUSUM test have exactly the same shape as in the retrospective case. The monitoring local power curve for a break at $r^* \in (1, 2)$ then coincides with the corresponding retrospective curve in Figure 2.3 with a single break at $r^* - 1$. Hence, the power of the stacked backward CUSUM is always higher than that of the forward CUSUM if $r^* \geq 1.15$. However, the delay between the actual break and the detection time point is a much more important performance measure for monitoring detectors than the power itself, since every fixed nontrivial alternative will be detected if the monitoring horizon is long enough. Let τ be the stopping time of the time point of the first boundary crossing, and let the mean local relative delay be given by $E\big[\tau/T | r^* \leq \tau/T \leq m\big] - r^*$.

Figure 2.4 presents the simulated mean local relative delay curves for the fixed endpoint $m = 4$ for $\boldsymbol{\mathcal{M}}^{\text{mon}}_{SBQ,4}$ with the linear boundary, for $\boldsymbol{\mathcal{M}}^{\text{mon}}_{Q,4}$ with the linear boundary, and for $\boldsymbol{\mathcal{M}}^{\text{mon}}_{Q,4}$ with the radical boundary by Chu et al. (1996). The mean local relative delay of the stacked backward CUSUM is much lower than that of the forward CUSUM. Furthermore, the mean local relative delay is constant across different break locations, with the exception of breaks that are located at $r^* < 1.15$.

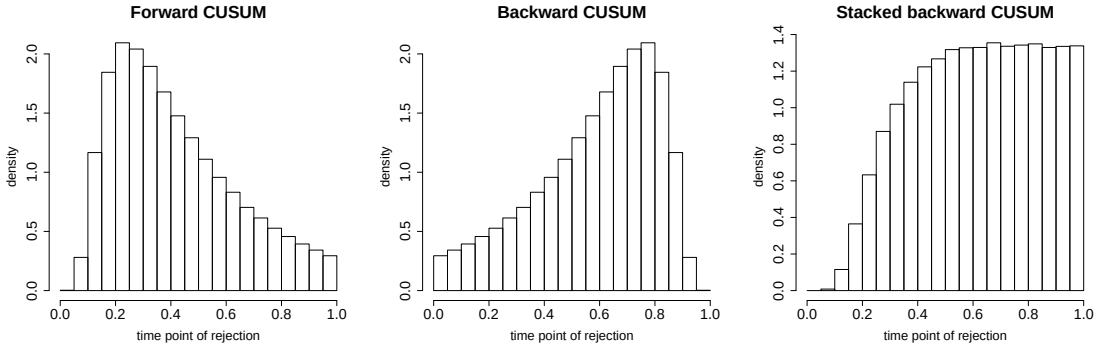
Moreover, we compare the asymptotic distributions of the size, which is the distribution of the time point of the first boundary crossing under H_0. Figure 2.5 presents histograms of the asymptotic size distributions for retrospective testing under the linear boundary. For the forward CUSUM, the highest rejection rates under H_0 are obtained at relative locations between 0.15 and 0.4 of the sample. For the backward CUSUM, the picture is mirror-inverted, such that most weight is put on rejections at relative locations between 0.6 and 0.85. The distribution for the forward CUSUM is right-skewed, whereas, for the backward CUSUM, it is left-skewed. For the stacked backward CUSUM, the distribution is much closer to a uniform distribution, although it is slightly left-skewed. Note that the size distributions provide information about the location of false rejections, but, when comparing Figure 2.3 with Figure 2.5, it is reasonable to assume that this is also related to the distribution of the power across different time points. There is no consensus on which distribution should be preferred, as whether one wishes to put more weight on particular regions of time points of rejection depends on the particular application. However, Zeileis et al. (2005) and Anatolyev and Kosenok (2018) argue that if no further information is available, one might prefer a uniform distribution to a skewed one. Figure 2.6 presents the distributions of the size for the fixed monitoring horizon with $m = 10$. The distribution

Figure 2.4: Asymptotic local mean delay curves for monitoring with $m = 4$



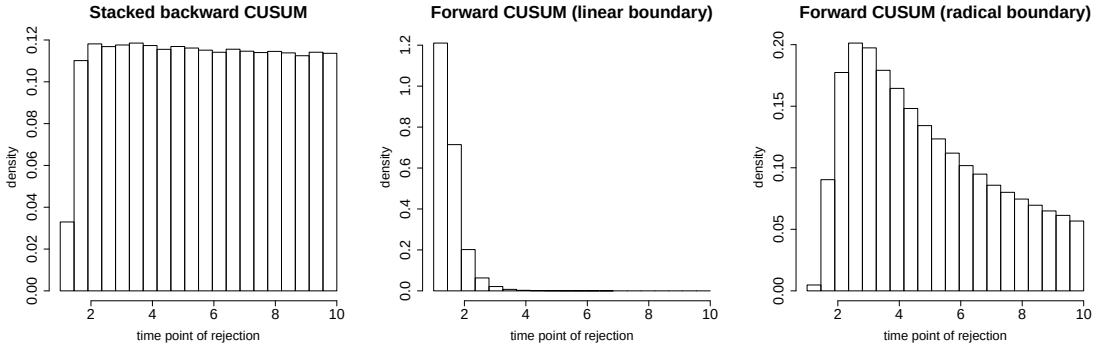
Note: The plots show simulated local mean delay curves, where the relative mean delays are given on the y-axis. While, for the first two plots, the break locations are fixed with $r^* \in \{1.5, 3\}$ and local break sizes c/σ are given on the x-axis, for the last plot, the local break size is fixed with $c/\sigma = 20$, and the breakpoint locations r^* are given on the x-axis. The linear boundary (2.2) is considered for $\alpha = 5\%$.

Figure 2.5: Size distributions of the retrospective detectors



Note: The plots show the frequencies of the location of the first boundary exceedance under the null hypothesis. The frequencies are based on random draws under the limiting distribution of the maximum statistics of the forward CUSUM, the backward CUSUM, and the stacked backward CUSUM detector using the linear boundary in (2.2) with a significance level of 5% under a model with $k = 1$.

Figure 2.6: Size distributions of the monitoring detectors with $m = 10$



Note: The plots show the frequencies of the location of the first boundary exceedance under the null hypothesis. The frequencies are based on random draws under the limiting distribution of the monitoring maximum statistics with $m = 10$. The stacked backward CUSUM detector using the linear boundary, the forward CUSUM detector using the linear boundary, and the forward CUSUM detector using the radical boundary by Chu et al. (1996) are considered at a significance level of 5% under a model with $k = 1$.

Table 2.1: Asymptotic critical values for the retrospective tests

	$\mathcal{M}_Q^{\text{ret}}$ and $\mathcal{M}_{BQ}^{\text{ret}}$					$\mathcal{M}_{SBQ}^{\text{ret}}$				
k	20%	10%	5%	2.5%	1%	20%	10%	5%	2.5%	1%
1	0.734	0.847	0.945	1.034	1.143	1.018	1.113	1.198	1.278	1.374
2	0.839	0.941	1.032	1.115	1.219	1.107	1.196	1.277	1.352	1.442
3	0.895	0.993	1.081	1.163	1.260	1.156	1.244	1.321	1.392	1.481
4	0.933	1.029	1.114	1.192	1.287	1.190	1.275	1.350	1.419	1.506
5	0.962	1.056	1.139	1.216	1.307	1.216	1.299	1.372	1.441	1.526
6	0.985	1.077	1.160	1.235	1.323	1.237	1.317	1.388	1.457	1.541
7	1.005	1.095	1.176	1.249	1.338	1.253	1.333	1.404	1.471	1.556
8	1.021	1.110	1.189	1.261	1.349	1.268	1.347	1.418	1.483	1.566

Note: Critical values λ_α are reported for the linear boundary in (2.2) from 100,000 Monte Carlo repetitions. The Gaussian processes in the limiting distributions are simulated on a grid of 10,000 equidistant points.

Table 2.2: Empirical sizes of the retrospective tests

	$k=1$			$k=2$			$k=3$			$k=4$		
T	100	200	500	100	200	500	100	200	500	100	200	500
$\mathcal{M}_Q^{\text{ret}}$	3.8	4.2	4.6	4.0	4.4	4.5	4.0	4.4	4.5	4.1	4.3	4.5
$\mathcal{M}_{BQ}^{\text{ret}}$	4.1	4.2	4.6	4.8	4.7	4.6	5.4	4.9	4.6	6.0	5.3	4.7
$\mathcal{M}_{SBQ}^{\text{ret}}$	2.8	3.5	4.2	3.9	4.0	4.2	4.7	4.5	4.2	5.7	4.9	4.4

Note: Simulated rejection rates under H_0 are presented in percentage points. The values are obtained from 100,000 Monte Carlo repetitions using the critical values from Table 2.1 at a significance level of 5% for the linear boundary (2.2). The cases $k = 1, \ldots, 4$ represent the models $y_t = \beta_1 + u_t$, $y_t = \beta_1 + \beta_2 x_{t2} + u_t$, $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, and $y_t = \beta_1 + \beta_2 x_{t2} + \beta_3 x_{t3} + u_t$, respectively, where x_{t2}, x_{t3}, x_{t4}, and u_t are simulated independently as standard normal random variables for all $t = 1, \ldots, T$.

for the stacked backward CUSUM is much closer to a uniform distribution compared to those of the forward CUSUM variants.

2.4.2 Critical values and finite sample performance

Table 2.1 presents critical values for the retrospective case using the linear boundary, while the empirical size results for a significance level of 5% are shown in Table 2.2. The tests have only minor size distortions in finite samples.

The empirical powers of the retrospective tests are compared with that of the sup-Wald test of Andrews (1993). The sup-Wald statistic is given by

$$\max_{r \in [r_0, 1-r_0]} T \cdot \frac{S_0 - S_1(r) - S_2(r)}{r(1-r)},$$

where S_0 is the OLS residual sum of squares using observations $\{1, \ldots, T\}$, $S_1(r)$ is the OLS residual sum of squares using observations $\{1, \ldots, \lfloor rT \rfloor\}$, and $S_2(r)$ is the OLS residual sum of squares using observations $\{\lfloor rT \rfloor + 1, \ldots, T\}$. The parameter r_0 defines the lower and upper trimming parameters. In the subsequent simulations, we consider $r_0 = 0.15$,

Table 2.3: Size-adjusted powers of the retrospective tests

	Model (2.6) ($k=1$)				Model (2.7) ($k=2$)			
	$\boldsymbol{\mathcal{M}}_{Q}^{\text{ret}}$	$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}$	$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}}$	supW	$\boldsymbol{\mathcal{M}}_{Q}^{\text{ret}}$	$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}}$	$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}}$	supW
$r^* = 0.1$	46.9	28.3	40.7	26.3	32.5	19.0	25.9	21.5
$r^* = 0.2$	63.5	65.0	71.2	73.9	47.2	47.4	51.7	59.3
$r^* = 0.3$	67.1	84.0	83.9	86.8	50.8	70.3	68.1	75.3
$r^* = 0.4$	63.5	91.5	88.7	91.4	47.1	81.9	75.9	82.3
$r^* = 0.5$	54.0	93.8	89.4	92.5	38.2	85.7	77.0	84.3
$r^* = 0.6$	39.4	93.3	86.6	91.4	26.6	84.1	72.0	82.2
$r^* = 0.7$	23.4	89.0	77.0	86.9	15.6	75.5	58.9	75.3
$r^* = 0.8$	11.0	74.2	51.6	74.1	8.2	56.0	37.0	59.5
$r^* = 0.9$	5.5	31.4	12.9	26.2	5.1	24.6	13.3	21.4

Note: Simulated size-adjusted rejection rates under models (2.6) and (2.7) are presented in percentage points for a significance level of 5% and a sample size of $T = 100$, where supW denotes the sup-Wald test with $r_0 = 0.15$. The values are obtained from 100,000 Monte Carlo repetitions for a sample size of $T = 100$, while the linear boundary (2.2) is implemented.

which is the default setting suggested by Andrews (1993). The limiting distribution is given by $\sup_{r\in[r_0,1-r_0]} \boldsymbol{B}(r)'\boldsymbol{B}(r)/(r(1-r))$, and critical values for different values of r_0 and k are tabulated in Andrews (1993). The author showed that the sup-Wald test has weak optimality properties in the sense that, in the case of a single structural break, its local power curve approaches the power curve from the infeasible point optimal maximum likelihood test asymptotically, as the significance level tends to zero. Note that the sup-Wald statistic is not suitable for monitoring, since its numerator statistic $T(S_0 - S_1(t/T) - S_2(t/T))$ is not measurable with respect to the filtration of information at time t.

We illustrate the finite sample performance for a simple model with $k = 1$ and a break in the mean, which is given by

$$y_t = \mu_t + u_t, \quad \mu_t = 2 + 0.8 \cdot 1_{\{\frac{t}{T} \geq r^*\}}, \quad u_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{2.6}$$

and for a univariate linear regression model with a break in the slope coefficient, which is given by

$$y_t = \mu_t + \beta_t x_t + u_t, \quad \mu_t = 2, \quad \beta_t = 1 + 0.8 \cdot 1_{\{\frac{t}{T} \geq r^*\}}, \quad x_t, u_t \overset{iid}{\sim} \mathcal{N}(0,1), \tag{2.7}$$

where $t = 1, \ldots, T$. Table 2.3 presents the size-adjusted power results.

First, we observe that the backward CUSUM and the stacked backward CUSUM outperform the forward CUSUM, except for the case $r^* = 0.1$. Second, while the forward CUSUM test has much lower power than the sup-Wald test, the reversed order cumulation structure in the backward CUSUM seems to compensate for this weakness of the forward CUSUM test. The backward CUSUM performs equally well than the sup-Wald test, which is remarkable since, as discussed previously, the latter test has weak optimality proper-

Table 2.4: Asymptotic critical values for the stacked backward CUSUM monitoring

	$k=1$			$k=2$			$k=3$			$k=4$		
m	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.2	0.782	0.859	1.024	0.859	0.935	1.092	0.902	0.975	1.129	0.932	1.003	1.152
1.4	0.941	1.030	1.208	1.028	1.111	1.277	1.076	1.156	1.320	1.108	1.185	1.345
1.6	1.026	1.113	1.292	1.111	1.192	1.365	1.158	1.238	1.406	1.189	1.269	1.432
1.8	1.077	1.162	1.344	1.161	1.244	1.411	1.208	1.286	1.452	1.240	1.317	1.476
2	1.113	1.198	1.374	1.196	1.277	1.442	1.244	1.321	1.481	1.275	1.350	1.506
3	1.211	1.293	1.462	1.291	1.366	1.524	1.334	1.407	1.558	1.363	1.436	1.582
4	1.262	1.339	1.500	1.336	1.410	1.564	1.378	1.450	1.599	1.407	1.478	1.621
6	1.316	1.390	1.544	1.387	1.460	1.606	1.428	1.496	1.638	1.456	1.522	1.660
8	1.346	1.419	1.569	1.417	1.486	1.629	1.456	1.522	1.661	1.483	1.548	1.686
10	1.367	1.440	1.588	1.437	1.503	1.644	1.475	1.540	1.677	1.500	1.565	1.703
∞	1.450	1.514	1.648	1.512	1.573	1.703	1.547	1.606	1.736	1.570	1.629	1.760
	$k=5$			$k=6$			$k=7$			$k=8$		
m	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
1.2	0.954	1.023	1.170	0.972	1.041	1.186	0.987	1.054	1.198	1.000	1.065	1.206
1.4	1.133	1.208	1.366	1.152	1.225	1.381	1.167	1.241	1.396	1.181	1.253	1.409
1.6	1.214	1.293	1.452	1.235	1.311	1.466	1.251	1.325	1.477	1.265	1.339	1.488
1.8	1.265	1.340	1.496	1.283	1.357	1.511	1.300	1.372	1.525	1.315	1.385	1.537
2	1.299	1.372	1.526	1.317	1.388	1.541	1.333	1.404	1.556	1.347	1.418	1.566
3	1.386	1.457	1.601	1.404	1.472	1.615	1.420	1.487	1.629	1.433	1.500	1.640
4	1.429	1.497	1.638	1.446	1.513	1.651	1.461	1.527	1.665	1.473	1.539	1.679
6	1.476	1.541	1.680	1.492	1.557	1.696	1.507	1.571	1.709	1.519	1.583	1.718
8	1.503	1.567	1.706	1.519	1.582	1.718	1.533	1.596	1.728	1.545	1.607	1.739
10	1.520	1.584	1.718	1.536	1.599	1.732	1.551	1.612	1.744	1.562	1.623	1.752
∞	1.589	1.647	1.775	1.604	1.661	1.788	1.617	1.673	1.799	1.627	1.683	1.807

Note: Critical values λ_α for $\boldsymbol{\mathcal{M}}^{\text{mon}}_{SBQ,m}$ are reported using the linear boundary (2.2). The Gaussian processes in the limiting distributions are simulated on a grid of 10,000 equidistant points with 100,000 Monte Carlo repetitions. The case $m = \infty$ corresponds to the right-hand side of equation (2.5).

ties. Finally, while the sup-Wald statistic and the backward CUSUM detector are not suitable for monitoring, the stacked backward CUSUM test is much more powerful than the forward CUSUM test, and its detector statistic is therefore well suited for real-time monitoring.

For the monitoring case, the critical values for the stacked backward CUSUM are shown in Table 2.4. For the forward CUSUM with the linear boundary (2.2), the simulated 5% critical values for $m = \infty$ using the representation of the right hand side of equation (2.3) are given by 0.957 for $k = 1$ and 1.044 for $k = 2$.

In order to evaluate the finite sample performances of the monitoring detectors, we consider models (2.6) and (2.7) for the time points $t = T + 1, \ldots, \lfloor mT \rfloor$. We simulate the series up to the fixed endpoints $m \in \{1.5, 2, 4, 10\}$, while the critical values for the case $m = \infty$ are implemented. Table 2.5 presents the size results. Note, that the tests are undersized by construction, as not all of the size is used up to the time point $\lfloor mT \rfloor$. For $k \geq 2$, we observe some size distortions for small sample sizes. The results in Table 2.6

Table 2.5: Empirical sizes of the infinite horizon monitoring detectors

	$k=1$						$k=2$					
	$T=100$			$T=500$			$T=100$		$T=200$		$T=500$	
horizon	SBQ	Q	CSW	SBQ	Q	CSW	SBQ	Q	SBQ	Q	SBQ	Q
$m=1.5$	0.1	2.8	0.0	0.1	3.0	0.0	0.5	4.5	0.2	3.7	0.1	3.2
$m=2$	0.2	4.2	0.1	0.2	4.4	0.1	1.4	6.6	0.7	5.5	0.4	4.8
$m=4$	1.0	4.7	0.9	0.9	4.8	0.8	4.8	7.3	2.5	6.0	1.4	5.2
$m=6$	1.7	4.7	1.6	1.4	4.8	1.4	7.7	7.4	4.1	6.0	2.3	5.2
$m=8$	2.4	4.7	2.0	2.0	4.8	1.8	10.3	7.4	5.7	6.0	3.3	5.2
$m=10$	3.1	4.7	2.3	2.7	4.8	2.0	12.7	7.4	7.2	6.0	4.3	5.2

Note: Simulated rejection rates under H_0 are presented in percentage points. The linear boundary (2.2) is implemented, while critical values for $\alpha = 5\%$ and $m = \infty$ are considered. The values are obtained from 100,000 random draws of the models $y_t = \beta_1 + u_t$ and $y_t = \beta_1 + \beta_2 x_{t2} + u_t$ for $t = 1, \ldots, \lfloor mT \rfloor$, where x_{t2} and u_t are i.i.d. and standard normal. While SBQ and Q correspond to the stacked backward CUSUM and the forward CUSUM with critical values for the case $m = \infty$, the univariate test by Chu et al. (1996) using the radical boundary (2.4) is denoted by CSW.

Table 2.6: Empirical mean detection delays of the monitoring detectors

	Model (2.6)			Model (2.7)	
	SBQ	Q	CSW	SBQ	Q
$r^* = 1.5$	41.4	39.4	53.6	62.2	50.4
$r^* = 2$	38.4	59.4	60.1	57.7	77.0
$r^* = 2.5$	36.9	79.2	65.8	54.6	103.4
$r^* = 3$	36.0	99.1	71.1	52.4	129.6
$r^* = 5$	34.5	178.0	89.4	48.1	233.6
$r^* = 10$	33.5	374.6	124.2	45.7	487.8

Note: The empirical mean detection delays are obtained from 100,000 Monte Carlo repetitions using size-adjusted critical values for a significance level of 5%, where models (2.6) and (2.7) are simulated for $t = 1, \ldots, \lfloor mT \rfloor$ with $T = 100$ and $m = 20$. While SBQ and Q correspond to the stacked backward CUSUM and the forward CUSUM with the linear boundary (2.2) and with critical values for the case $m = \infty$, the univariate test by Chu et al. (1996) with the radical boundary (2.4) is denoted by CSW.

show that the mean delay for the stacked backward CUSUM is much lower than that of the forward CUSUM and is almost constant across the breakpoint locations.

2.5 Conclusion

Two alternatives to the conventional CUSUM detectors by Brown et al. (1975) and Chu et al. (1996) have been proposed. It has been demonstrated that a detector that backwardly cumulates recursive residuals yields much higher power than when using forwardly cumulated recursive residuals when the break is located in the middle or at the end of the sample. Furthermore, the stacked triangular array of backwardly cumulated recursive residuals can be applied for monitoring and yields a much lower detection delay than that of the monitoring procedure by Chu et al. (1996). Due to the multivariate nature of the tests, we also have power against structural breaks that do not affect the unconditional mean of the dependent variable.

Appendix to Chapter 2

We present some auxiliary lemmas that are required to prove Theorem 2.1.

Lemma 2.1. *Let $\{(\boldsymbol{x}_t, u_t)\}_{t\in\mathbb{N}}$ satisfy Assumption 2.1, let $\mathcal{F}_t$ be the σ-algebra generated by $\{(\boldsymbol{x}'_{i+1}, u_i)',\ i \le t\}$, and let $\boldsymbol{z}_t$ be some k-dimensional and $\mathcal{F}_{t-1}$-measurable vector sequence that satisfies $\|T^{-1}\sum_{t=1}^{T}\boldsymbol{z}_t\boldsymbol{z}'_t - \boldsymbol{R}\|_M = o_P(1)$ for some positive definite matrix $\boldsymbol{R}$, where $\|\cdot\|_M$ denotes some matrix norm on $\mathbb{R}^{k\times k}$. Then, for any fixed and positive $m < \infty$,*

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{z}_t\epsilon_t \Rightarrow \sigma\boldsymbol{R}^{-1/2}\mathbf{W}(r), \quad r\in[0,m],$$

as $T\to\infty$, where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions.

Proof. A direct consequence of the functional central limit theorem for multiple time series on the space $D([0,1])^k$ given by Theorem 2.1 in Phillips and Durlauf (1986) is that $M^{-1/2}\sum_{t=1}^{\lfloor sM\rfloor}\boldsymbol{z}_t u_t \Rightarrow \sigma\boldsymbol{R}^{-1/2}\mathbf{W}(s)$, $s\in[0,1]$, as $M\to\infty$. A similar result is also stated in Lemma 3 in Krämer et al. (1988). Then, on the space $D([0,m])^k$,

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{z}_t\epsilon_t = \frac{\sqrt{m}}{\sqrt{M}}\sum_{t=1}^{\lfloor (r/m)M\rfloor}\boldsymbol{z}_t\epsilon_t \Rightarrow \sqrt{m}\mathbf{W}(r/m) \overset{\mathcal{D}}{=} \mathbf{W}(r), \quad r\in[0,m].$$

□

Lemma 2.2. *Let $\{(\boldsymbol{x}_t, u_t)\}_{t\in\mathbb{N}}$ satisfy Assumption 2.1, let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t\in\mathbb{N}$, and let $m\in(0,\infty)$. Then, as $T\to\infty$,*

$$\sup_{r\in[0,m]}\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{x}_t w_t - \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{x}_t u_t + \int_0^r\frac{1}{s}\Big(\frac{1}{\sqrt{T}}\sum_{j=1}^{\lfloor sT\rfloor}\boldsymbol{x}_j u_j\Big)\,\mathrm{d}s\right\|_V = o_P(1),$$

where $\|\cdot\|_V$ denotes some vector norm on $\mathbb{R}^k$.

Proof. Let $t > k$, and let $f_t = (1+(t-1)^{-1}\boldsymbol{x}'_t\boldsymbol{C}_{t-1}^{-1}\boldsymbol{x}_t)^{1/2}$ be the denominator of w_t.

Furthermore, let

$$\mathbf{Y}_t(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t, \quad \mathbf{X}_T(r) = \frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t u_t, \quad \mathbf{Z}_T(r) = \int_0^r \frac{1}{s} \mathbf{X}_T(s)\, \mathrm{d}s,$$

$$\mathbf{Z}_{1,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \frac{1}{f_t} \boldsymbol{x}_t u_t, \quad \mathbf{Z}_{2,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \frac{1}{f_t(t-1)} \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}_{t-1}^{-1} \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j,$$

$$\mathbf{Z}_{3,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \frac{1}{t-1} \boldsymbol{x}_t \boldsymbol{x}_t' \boldsymbol{C}^{-1} \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j, \quad \mathbf{Z}_{4,T}(r) = \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor - 1} \frac{1}{t-1} \boldsymbol{C}_t \boldsymbol{C}^{-1} \sum_{j=1}^{t} \boldsymbol{x}_j u_j.$$

From the fact that

$$f_t w_t = y_t - \boldsymbol{x}_t' \widehat{\boldsymbol{\beta}}_{t-1} = u_t - \boldsymbol{x}_t' \Big(\sum_{j=1}^{t-1} \boldsymbol{x}_j \boldsymbol{x}_j' \Big)^{-1} \Big(\sum_{j=1}^{t-1} \boldsymbol{x}_j u_j \Big) = u_t - \boldsymbol{x}_t' \boldsymbol{C}_{t-1}^{-1} \Big(\frac{1}{t-1} \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j \Big)$$

it follows that $\mathbf{Y}_t(r) = \mathbf{Z}_{1,T}(r) - \mathbf{Z}_{2,T}(r)$. Thus, it remains to show

$$\sup_{r \in [0,m]} \|\mathbf{Z}_{1,T}(r) - \mathbf{X}_T(r)\|_V = o_P(1), \tag{2.8}$$

$$\sup_{r \in [0,m]} \|\mathbf{Z}_{2,T}(r) - \mathbf{Z}_{3,T}(r)\|_V = o_P(1), \tag{2.9}$$

$$\sup_{r \in [0,m]} \|\mathbf{Z}_{3,T}(r) - \mathbf{Z}_{4,T}(r)\|_V = o_P(1), \tag{2.10}$$

$$\sup_{r \in [0,m]} \|\mathbf{Z}_{4,T}(r) - \mathbf{Z}_T(r)\|_V = o_P(1), \tag{2.11}$$

as $T \to \infty$, which implies that $\sup_{r \in [0,m]} \|\mathbf{Y}_t(r) - \mathbf{X}_T(r) + \mathbf{Z}_T(r)\|_V = o_P(1)$. For equation (2.8), note that $\sqrt{t}(f_t^{-1} - 1) = O_P(1)$ for all $t > k$, and let $\boldsymbol{z}_t = \sqrt{T}(f_t^{-1} - 1)\boldsymbol{x}_t$, which satisfies the conditions of Lemma 2.1 for $t > k$. Then,

$$\sup_{r \in [0,m]} \|\mathbf{Z}_{1,T}(r) - \mathbf{X}_T(r)\|_V \leq \sup_{r \in [0,m]} \left(\Big\| \frac{1}{T} \sum_{t=k+1}^{\lfloor rT \rfloor} \boldsymbol{z}_t u_t \Big\|_V + \Big\| \frac{1}{\sqrt{T}} \sum_{t=1}^{k} \boldsymbol{x}_t u_t \Big\|_V \right) = O_P(T^{-1/2}).$$

For equation (2.9), we have

$$\begin{aligned} \sup_{r \in [0,m]} \|\mathbf{Z}_{2,T}(r) - \mathbf{Z}_{3,T}(r)\|_V &= \sup_{r \in [0,m]} \Big\| \frac{1}{\sqrt{T}} \sum_{t=k+1}^{\lfloor rT \rfloor} \frac{1}{t-1} \boldsymbol{x}_t \boldsymbol{x}_t' \big(f_t^{-1} \boldsymbol{C}_{t-1}^{-1} - \boldsymbol{C}^{-1} \big) \sum_{j=1}^{t-1} \boldsymbol{x}_j u_j \Big\|_V \\ &= \sup_{r \in [0,m]} \Big\| \int_0^r \frac{1}{s} \boldsymbol{x}_{\lfloor sT \rfloor} \boldsymbol{x}_{\lfloor sT \rfloor}' \big(f_{\lfloor sT \rfloor}^{-1} \boldsymbol{C}_{\lfloor sT \rfloor}^{-1} - \boldsymbol{C}^{-1} \big) \mathbf{X}_T(s)\, \mathrm{d}s \Big\|_V + O_P(T^{-1/2}) = o_P(1), \end{aligned}$$

since $\sup_{r\in[0,m]} \|f^{-1}_{\lfloor sT\rfloor}\boldsymbol{C}^{-1}_{\lfloor sT\rfloor} - \boldsymbol{C}^{-1}\|_M = o_P(1)$ due to Assumption 2.1(a), where $\|\cdot\|_M$ denotes the induced matrix norm of $\|\cdot\|_V$.

To show (2.10), we apply Abel's formula of summation by parts, which is given by

$$\sum_{t=1}^{n} \boldsymbol{A}_t\boldsymbol{b}_t = \sum_{t=1}^{n} \boldsymbol{A}_t\boldsymbol{b}_n + \sum_{t=1}^{n-1}\sum_{j=1}^{t} \boldsymbol{A}_j(\boldsymbol{b}_t - \boldsymbol{b}_{t+1}), \qquad \boldsymbol{A}_t \in \mathbb{R}^{k\times k}, \quad \boldsymbol{b}_t \in \mathbb{R}^k, \quad n \in \mathbb{N}. \tag{2.12}$$

With $\boldsymbol{A}_t = \boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{C}^{-1}$ and $\boldsymbol{b}_t = (t-1)^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j u_j$, it follows that

$$\begin{aligned}\mathbf{Z}_{3,T}(r) &= \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor} \boldsymbol{A}_t\boldsymbol{b}_t = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor} \boldsymbol{A}_t\boldsymbol{b}_{\lfloor rT\rfloor} + \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t} \boldsymbol{A}_j(\boldsymbol{b}_t - \boldsymbol{b}_{t+1}) \\ &= \mathbf{Z}_{5,T}(r) - \mathbf{Z}_{6,T}(r) + \mathbf{Z}_{4,T}(r),\end{aligned}$$

where

$$\begin{aligned}\mathbf{Z}_{5,T}(r) &= \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor} \boldsymbol{A}_t\boldsymbol{b}_{\lfloor rT\rfloor} = \widetilde{\boldsymbol{C}}_{\lfloor rT\rfloor}\boldsymbol{C}^{-1}\mathbf{X}_T(\tfrac{\lfloor rT\rfloor-1}{T}), \qquad \widetilde{\boldsymbol{C}}_t = \frac{1}{t-1}\sum_{j=k+1}^{t}\boldsymbol{x}_j\boldsymbol{x}_j', \\ \mathbf{Z}_{6,T}(r) &= \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t} \boldsymbol{A}_j(\boldsymbol{b}_{t+1} - \tfrac{t-1}{t}\boldsymbol{b}_t) = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor-1} \boldsymbol{C}_t\boldsymbol{C}^{-1}\boldsymbol{x}_t u_t, \\ \mathbf{Z}_{4,T}(r) &= \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t} \boldsymbol{A}_j(\boldsymbol{b}_t - \tfrac{t-1}{t}\boldsymbol{b}_t) = \frac{1}{\sqrt{T}}\sum_{t=k+1}^{\lfloor rT\rfloor-1} \frac{1}{t-1}\boldsymbol{C}_t\boldsymbol{C}^{-1}\sum_{j=1}^{t}\boldsymbol{x}_j u_j.\end{aligned}$$

With Assumption 2.1(a) and Lemma 2.1, it follows that

$$\begin{aligned}\sup_{r\in[0,m]} \big\|\mathbf{Z}_{5,T}(r) - \mathbf{X}_T(r)\big\|_V &\le \sup_{r\in[0,m]} \big\|(\widetilde{\boldsymbol{C}}_{\lfloor rT\rfloor}\boldsymbol{C}^{-1} - \mathbf{I}_k)\mathbf{X}_T(r)\big\|_V + o_P(1) = o_P(1), \\ \sup_{r\in[0,m]} \big\|\mathbf{Z}_{6,T}(r) - \mathbf{X}_T(r)\big\|_V &\le \sup_{r\in[0,m]} \Big\|\frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{C}_t\boldsymbol{C}^{-1} - \boldsymbol{I}_k)\boldsymbol{x}_t u_t\Big\|_V + o_P(1) = o_P(1).\end{aligned}$$

Then, the triangle inequality yields $\sup_{r\in[0,m]} \|\mathbf{Z}_{3,T}(r) - \mathbf{Z}_{4,T}(r)\| = o_P(1)$. Finally, (2.11) follows from Assumption 2.1(a) together with Lemma 2.1 and the continuous mapping

theorem, which imply that

$$\begin{aligned}
&\sup_{r\in[0,m]} \left\| \mathbf{Z}_{4,T}(r) - \int_0^r \frac{1}{s}\Big(\frac{1}{\sqrt{T}} \sum_{j=1}^{\lfloor sT \rfloor} \boldsymbol{x}_j u_j\Big)\,\mathrm{d}s \right\|_V \\
&\leq \sup_{r\in[0,m]} \left\| \int_0^r \frac{1}{s}(\boldsymbol{C}_{\lfloor rT \rfloor}\boldsymbol{C}^{-1} - \boldsymbol{I}_k)\mathbf{X}_T(s)\,\mathrm{d}s \right\|_V + o_P(1) = o_P(1).
\end{aligned}$$

□

Lemma 2.3. *Let $W(r)$ be a standard Brownian motion. Furthermore, for $r \geq 0$, let $F(W(r)) = W(r) - \int_0^r z^{-1}W(z)\,\mathrm{d}z$. Then $F(W(r)) \overset{\mathcal{D}}{=} W(r)$.*

Proof. Note that, by the Cauchy-Schwarz inequality and Jensen's inequality, we obtain $\int_0^r z^{-1}E[|W(z)|]\,\mathrm{d}z < \infty$ as well as $\int_0^r z^{-1}E[|W(r)W(z)|]\,\mathrm{d}z < \infty$, which justifies the application of Fubini's theorem in the subsequent steps. Since both $W(r)$ and $F(W(r))$ are Gaussian, it remains to show that their covariance functions coincide. First, note that $E[F(W(r))] = E[W(r)] = 0$. Furthermore, let w.l.o.g. $r \leq s$. Then, the assertion follows from

$$\begin{aligned}
&E[F(W(r))F(W(s))] - E[W(r)W(s)] \\
&= \int_0^r \int_0^s \frac{E[W(z_1)W(z_2)]}{z_1 z_2}\,\mathrm{d}z_2\,\mathrm{d}z_1 - \int_0^s \frac{E[W(r)W(z_2)]}{z_2}\,\mathrm{d}z_2 - \int_0^r \frac{E[W(s)W(z_1)]}{z_1}\,\mathrm{d}z_1 \\
&= (2r + r\ln(s) - r\ln(r)) - (r + r\ln(s) - r\ln(r)) - r = 0.
\end{aligned}$$

□

Lemma 2.4. *Let $\{(\boldsymbol{x}_t, u_t)\}_{t\in\mathbb{N}}$ satisfy Assumption 2.1, let $\boldsymbol{\beta}_t = \boldsymbol{\beta}^0$ for all $t \in \mathbb{N}$, and let $m \in (0, \infty)$. Then, as $T \to \infty$,*

$$\frac{1}{\sqrt{T}} \sum_{t=1}^{\lfloor rT \rfloor} \boldsymbol{x}_t w_t \Rightarrow \sigma \boldsymbol{C}^{-1/2}\mathbf{W}(r), \quad r \in [0, m],$$

where $\mathbf{W}(r)$ is a vector of k independent standard Brownian motions.

Proof. Let $\mathbf{X}_T(r) = T^{-1/2}\sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j u_j$, and let $\mathbf{Y}_T(r) = T^{-1/2}\sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j w_j$. Lemma 2.2 states that $\sup_{r\in[0,m]} \|\mathbf{Y}_T(r) - F(\mathbf{X}_T(r))\|_V = o_P(1)$, where $F(\cdot)$ is defined as in Lemma 2.3. Therefore, the Skorokhod metric of $F(\mathbf{X}_T(r))$ and $\mathbf{Y}_T(r)$ tends to zero in probability. Lemma 2.1 implies that $F(\mathbf{X}_T(r)) \Rightarrow F(\sigma\boldsymbol{C}^{-1/2}\mathbf{W}(r)) = \sigma\boldsymbol{C}^{-1/2}F(\mathbf{W}(r))$. Furthermore, from Lemma 2.3, it follows that $F(\mathbf{W}(r)) \overset{\mathcal{D}}{=} \mathbf{W}(r)$. Consequently, $\mathbf{Y}_T(r) \Rightarrow \sigma\boldsymbol{C}^{-1/2}\mathbf{W}(r)$.

□

Lemma 2.5. *Let $\|\cdot\|_V$ be some vector norm on $\mathbb{R}^k$, and let $\|\cdot\|_M$ be the induced matrix norm. Let $\mathbf{h}$ be a $\mathbb{R}^k$-valued function of bounded variation, and let $\{\boldsymbol{A}_t\}_{t\in\mathbb{N}}$ be a sequence of random $(k\times k)$ matrices with $\sup_{r\in[0,m]}\|T^{-1}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\|_M = o_P(1)$, where $m\in(0,\infty)$. Then, as $T\to\infty$,*

$$\sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\mathbf{h}(\tfrac{t}{T})\Big\|_V = o_P(1).$$

Proof. By the application of Abel's formula of summation by parts, which is given in (2.12), it follows that

$$\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\mathbf{h}(\tfrac{t}{T}) = \sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\mathbf{h}(\tfrac{\lfloor rT\rfloor}{T}) + \sum_{t=1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t}(\boldsymbol{A}_j-\boldsymbol{A})(\mathbf{h}(\tfrac{t}{T})-\mathbf{h}(\tfrac{t+1}{T})).$$

The fact that $\mathbf{h}(r)$ is of bounded variation yields

$$\sup_{r\in[0,m]}\|\mathbf{h}(r)\|_V = O(1), \qquad \sup_{r\in[0,m]}\Big\|\sum_{t=1}^{\lfloor rT\rfloor-1}\frac{t}{T}(\mathbf{h}(\tfrac{t}{T})-\mathbf{h}(\tfrac{t+1}{T}))\Big\|_V = O(1).$$

Consequently,

$$\sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\mathbf{h}(\tfrac{\lfloor rT\rfloor}{T})\Big\|_V \le \sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(\boldsymbol{A}_t-\boldsymbol{A})\Big\|_M\Big\|\mathbf{h}(\tfrac{\lfloor rT\rfloor}{T})\Big\|_V = o_P(1)$$

and

$$\begin{aligned}&\sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor-1}\sum_{j=1}^{t}(\boldsymbol{A}_j-\boldsymbol{A})(\mathbf{h}(\tfrac{t}{T})-\mathbf{h}(\tfrac{t+1}{T}))\Big\|_V\\ &\le \sup_{r\in[0,m]}\sum_{t=1}^{\lfloor rT\rfloor-1}\frac{t}{T}\Big\|\frac{1}{t}\sum_{j=1}^{t}(\boldsymbol{A}_j-\boldsymbol{A})\Big\|_M\Big\|\mathbf{h}(\tfrac{t}{T})-\mathbf{h}(\tfrac{t+1}{T})\Big\|_V = o_P(1).\end{aligned}$$

Then, by the triangle inequality, the assertion follows. □

Proof of Theorem 2.1

Let $w_t^* = f_t^{-1}(y_t^* - \boldsymbol{x}_t'\widehat{\boldsymbol{\beta}}_{t-1}^*)$, which are recursive residuals from a regression without any structural break, where $f_t = (1 + (t-1)^{-1}\boldsymbol{x}_t'\boldsymbol{C}_{t-1}^{-1}\boldsymbol{x}_t)^{1/2}$,

$$y_t^* = \boldsymbol{x}_t'\boldsymbol{\beta}^0 + u_t, \quad \text{and} \quad \widehat{\boldsymbol{\beta}}_{t-1}^* = \Big(\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\Big)^{-1}\Big(\sum_{j=1}^{t-1}\boldsymbol{x}_j y_j^*\Big).$$

Then, $y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_t + u_t = y_t^* + T^{-1/2}\boldsymbol{x}_t'\mathbf{g}(t/T)$, and

$$\widehat{\boldsymbol{\beta}}_{t-1} = \widehat{\boldsymbol{\beta}}_{t-1}^* + \frac{1}{\sqrt{T}(t-1)}\boldsymbol{C}_{t-1}^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\mathbf{g}(j/T).$$

Furthermore, $w_t = w_t^* + f_t^{-1}T^{-1/2}\boldsymbol{x}_t'\mathbf{g}(t/T) - f_t^{-1}T^{-1/2}(t-1)^{-1}\boldsymbol{C}_{t-1}^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\mathbf{g}(j/T)$. We can decompose the partial sum process as $\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{x}_t w_t = \mathbf{S}_{1,T}(r) + \mathbf{S}_{2,T}(r) + \mathbf{S}_{3,T}(r)$, where

$$\mathbf{S}_{1,T}(r) = \frac{1}{\sqrt{T}}\sum_{t=1}^{\lfloor rT\rfloor}\boldsymbol{x}_t w_t^*, \quad \mathbf{S}_{2,T}(r) = \frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor} f_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t'\mathbf{g}(\tfrac{t}{T}),$$
$$\mathbf{S}_{3,T}(r) = -\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}\frac{1}{f_t(t-1)}\boldsymbol{x}_t\boldsymbol{x}_t'\boldsymbol{C}_{t-1}^{-1}\sum_{j=1}^{t-1}\boldsymbol{x}_j\boldsymbol{x}_j'\mathbf{g}(\tfrac{j}{T}).$$

Let $\|\cdot\|_V$ be some vector norm on $\mathbb{R}^k$, and let $\|\cdot\|_M$ be the induced matrix norm. Lemma 2.4 yields $\mathbf{S}_{1,T}(r) \Rightarrow \sigma\boldsymbol{C}^{1/2}\mathbf{W}(r)$. For the second term, note that, from Assumption 2.1(a) and the fact that $\sqrt{T}(f_T^{-1} - 1) = O_P(1)$, it follows that

$$\sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(f_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{C})\Big\|_M = o_P(1). \tag{2.13}$$

Since $\mathbf{g}(r)$ is piecewise constant and therefore of bounded variation, Lemma 2.5 yields

$$\sup_{r\in[0,m]}\Big\|\mathbf{S}_2(r) - \int_0^r \boldsymbol{C}\mathbf{g}(s)\,\mathrm{d}s\Big\|_V = \sup_{r\in[0,m]}\Big\|\frac{1}{T}\sum_{t=1}^{\lfloor rT\rfloor}(f_t^{-1}\boldsymbol{x}_t\boldsymbol{x}_t' - \boldsymbol{C})\mathbf{g}(\tfrac{t}{T})\Big\|_V = o_P(1).$$

For the third term, let

$$\mathbf{p}_1(r) = \frac{1}{\lfloor rT \rfloor} \boldsymbol{C}_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{x}_j \boldsymbol{x}_j' \mathbf{g}(\tfrac{j}{T}), \quad \mathbf{p}_2(r) = \frac{1}{\lfloor rT \rfloor} \boldsymbol{C}_{\lfloor rT \rfloor}^{-1} \sum_{j=1}^{\lfloor rT \rfloor} \boldsymbol{C} \mathbf{g}(\tfrac{j}{T}),$$
$$\mathbf{p}_3(r) = \frac{1}{\lfloor rT \rfloor} \sum_{j=1}^{\lfloor rT \rfloor} \mathbf{g}(\tfrac{j}{T}).$$

From Assumption 2.1(a), it follows that $\sup_{r\in[0,m]} \|\mathbf{p}_2(r) - \mathbf{p}_3(r)\|_M = o_P(1)$. Furthermore, from Lemma 2.5 and from the fact that $\sup_{r\in[0,m]} \|\frac{1}{\lfloor rT \rfloor} \sum_{t=1}^{\lfloor rT \rfloor} (\boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C})\|_M = o_P(1)$, it follows that $\sup_{r\in[0,m]} \|\mathbf{p}_1(r) - \mathbf{p}_2(r)\|_V = o_P(1)$. Thus, $\sup_{r\in[0,m]} \|\mathbf{p}_1(r) - \mathbf{p}_3(r)\|_V = o_P(1)$. Consequently,

$$\begin{aligned}
&\sup_{r\in[0,m]} \Big\| \mathbf{S}_{3,T}(r) + \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' \mathbf{h}_3(\tfrac{t-1}{T}) \Big\|_V \\
&\leq \sup_{r\in[0,m]} \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} \|f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t'\|_M \|\mathbf{p}_1(\tfrac{t-1}{T}) - \mathbf{p}_3(\tfrac{t-1}{T})\|_V,
\end{aligned}$$

which is $o_P(1)$. Since $\mathbf{p}_3$ is a partial sum of a piecewise constant function, it is of bounded variation, and, together with (2.13), we can apply Lemma 2.5. Then,

$$\sup_{r\in[0,m]} \Big\| \frac{1}{T} \sum_{t=1}^{\lfloor rT \rfloor} (f_t^{-1} \boldsymbol{x}_t \boldsymbol{x}_t' - \boldsymbol{C}) \mathbf{p}_3(\tfrac{t-1}{T}) \Big\| = o_P(1),$$

which yields

$$\begin{aligned}
&\sup_{r\in[0,m]} \Big\| \mathbf{S}_{3,T}(r) + \int_0^r \int_0^s \frac{1}{s} \boldsymbol{C} \mathbf{g}(v) \, \mathrm{d}v \, \mathrm{d}s \Big\|_V \\
&= \sup_{r\in[0,m]} \Big\| \mathbf{S}_{3,T}(r) + \frac{1}{T} \boldsymbol{C} \sum_{t=1}^{\lfloor rT \rfloor} \mathbf{p}_3(\tfrac{t-1}{T}) \Big\|_V + o_P(1) = o_P(1).
\end{aligned}$$

Finally, Slutsky's theorem implies that $\mathbf{S}_{1,T}(r) + \mathbf{S}_{2,T}(r) + \mathbf{S}_{3,T}(r) \Rightarrow \sigma \boldsymbol{C}^{1/2} \mathbf{W}(r) + \sigma \boldsymbol{C} \mathbf{h}(r)$, which yields

$$\mathbf{Q}_T(r) = \widehat{\sigma}^{-1} \boldsymbol{C}_T^{-1/2} (\mathbf{S}_{1,T}(r) + \mathbf{S}_{2,T}(r) + \mathbf{S}_{3,T}(r)) \Rightarrow \mathbf{W}(r) + \boldsymbol{C}^{1/2} \mathbf{h}(r),$$

since $\widehat{\sigma}^2$ is consistent for σ^2 (see Krämer et al. 1988).

Proof of Theorem 2.2

First, note that Assumption 2.2 and the law of the iterated logarithm ensure that all random variables considered in this proof are bounded in probability and are well defined. Let $m \in (1, \infty)$ be fixed. Under H_0, Theorem 2.1 yields $\mathbf{Q}_T(r) \Rightarrow \mathbf{W}(r)$, $r \in [0, m]$. Then, (a) follows with the continuous mapping theorem. For result (b), the continuous mapping theorem implies that

$$\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} = \sup_{r\in(0,m]} \frac{\|\mathbf{Q}_T(r) - \mathbf{Q}_T(1)\|}{d(r-1)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,m]} \frac{\|\mathbf{W}(r) - \mathbf{W}(1)\|}{d(r-1)} \overset{\mathcal{D}}{=} \sup_{r\in(0,m-1)} \frac{\|\mathbf{W}(r)\|}{d(r)}.$$

To obtain a supremum over a subset of the unit interval, we consider the bijective function $g : (0, (m-1)/m) \to (0, m-1)$ that is given by $g(\eta) = \eta/(1-\eta)$. Furthermore, note that $\mathbf{W}(g(\eta)) \overset{\mathcal{D}}{=} \mathbf{B}(\eta)/(1-\eta)$, where $\mathbf{B}(r)$ is a vector of k independent standard Brownian bridges. This follows from the fact that both $\mathbf{W}(g(\eta))$ and $\mathbf{B}(\eta)/(1-\eta)$ are Gaussian with mean zero and have the same covariance function. Consequently,

$$\sup_{r\in(0,m-1)} \frac{\|\mathbf{W}(r)\|}{d(r)} = \sup_{\eta\in(0,\frac{m-1}{m})} \frac{\|\mathbf{W}(g(\eta))\|}{d(g(\eta))} \overset{\mathcal{D}}{=} \sup_{\eta\in(0,\frac{m-1}{m})} \frac{\|\mathbf{B}(\eta)\|}{(1-\eta)d\big(\frac{\eta}{1-\eta}\big)}.$$

Finally, for (c), we obtain

$$\begin{aligned}\lim_{m\to\infty}\lim_{T\to\infty} P\big(\boldsymbol{\mathcal{M}}_{Q,m}^{\text{mon}} \leq \lambda\big) &= \lim_{m\to\infty} P\bigg(\sup_{r\in(0,\frac{m-1}{m})} \frac{\|\boldsymbol{B}(r)\|}{(1-r)d\big(\frac{r}{1-r}\big)} \leq \lambda\bigg) \\ &= P\bigg(\sup_{r\in(0,1)} \frac{\|\boldsymbol{B}(r)\|}{(1-r)d\big(\frac{r}{1-r}\big)} \leq \lambda\bigg),\end{aligned}$$

which follows from the monotonicity of $\sup_{r\in(0,\frac{m-1}{m})} \|\boldsymbol{B}(r)\|/((1-r)d(\frac{r}{1-r}))$ in m and the fact that the probability measure is σ-additive.

Proof of Theorem 2.3

Theorem 2.1 and the continuous mapping theorem imply that

$$\boldsymbol{\mathcal{M}}_{BQ}^{\text{ret}} = \sup_{r\in(0,1)} \frac{\|\mathbf{Q}_T(1) - \mathbf{Q}_T(r)\|}{d(1-r)} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \frac{\|\mathbf{W}(1) - \mathbf{W}(r)\|}{d(1-r)} \overset{\mathcal{D}}{=} \sup_{r\in(0,1)} \frac{\|\mathbf{W}(r)\|}{d(r)}.$$

Proof of Theorem 2.4

Analogously to the proof of Theorem 2.2,

$$\boldsymbol{\mathcal{M}}_{SBQ}^{\text{ret}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r)-\mathbf{W}(s)\|}{d(r-s)}, \quad \boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}} \xrightarrow{\mathcal{D}} \sup_{r\in(0,m-1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r)-\mathbf{W}(s)\|}{d(r-s)}$$

follow with Theorem 2.1 and the continuous mapping theorem. Furthermore, let the function $g : (0,(m-1)/m) \to (0,m-1)$ be given by $g(\eta) = \eta/(1-\eta)$. Then,

$$\begin{aligned}
&\sup_{r\in(0,m-1)} \sup_{s\in(0,r)} \frac{\|\mathbf{W}(r)-\mathbf{W}(s)\|}{d(r-s)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{s\in(0,g(\eta))} \frac{\|\mathbf{W}(g(\eta))-\mathbf{W}(s)\|}{d(g(\eta)-s)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,\eta)} \frac{\|\mathbf{W}(g(\eta))-\mathbf{W}(g(\zeta))\|}{d(g(\eta)-g(\zeta))} \\
&\overset{\mathcal{D}}{=} \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,\eta)} \frac{\|\mathbf{B}(\eta)/(1-\eta)-\mathbf{W}(\zeta)/(1-\zeta)\|}{d\big(\frac{\eta}{1-\eta}-\frac{\zeta}{1-\zeta}\big)} \\
&= \sup_{\eta\in(0,\frac{m-1}{m})} \sup_{\zeta\in(0,r)} \frac{\|(1-\zeta)\mathbf{B}(\eta)-(1-\eta)\mathbf{B}(\zeta)\|}{(1-\eta)(1-\zeta)d\big(\frac{\eta-\zeta}{(1-\eta)(1-\zeta)}\big)}.
\end{aligned}$$

Finally, for (c), we obtain

$$\begin{aligned}
&\lim_{m\to\infty}\lim_{T\to\infty} P\big(\boldsymbol{\mathcal{M}}_{SBQ,m}^{\text{mon}} \le \lambda\big) \\
&= \lim_{m\to\infty} P\bigg(\sup_{r\in(0,\frac{m-1}{m})} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)d\big(\frac{r-s}{(1-r)(1-s)}\big)} \le \lambda\bigg) \\
&= P\bigg(\sup_{r\in(0,1)} \sup_{s\in(0,r)} \frac{\|(1-s)\mathbf{B}(r)-(1-r)\mathbf{B}(s)\|}{(1-r)(1-s)d\big(\frac{r-s}{(1-r)(1-s)}\big)} \le \lambda\bigg),
\end{aligned}$$

which follows from the monotonicity in m and the fact that the probability measure is σ-additive.

Chapter 3

A Dynamic Functional Factor Model for Yield Curves: Identification, Estimation, and Prediction

3.1 Introduction

The number of factors that explain the shape of the term structure of bond yields is typically much smaller than the number of available maturities. According to Diebold et al. (2005), yields of the same bond with different maturities are driven by only a few sources of risk. Factor models thus received much attention in the literature on bond yield modeling, as they allow a high dimensional problem to be reduced to a lower dimensional one.

In the classical factor analysis, both the latent factors and the factor loadings are estimated using principal components, while the model is identified by imposing orthogonality conditions (see, e.g., Anderson 2003). Litterman and Scheinkman (1991), Ang et al. (2006), and Joslin et al. (2014) conducted a factor analysis on the yields of U.S. Treasuries with different maturities and demonstrated that the first three principal components explain more than 98% of the variation in the yields. Bliss (1997) proposed a dynamic version of the classical factor model for bond yields, where the first three principal components follow an autoregressive process. For surveys of classical and more recent results in factor analysis, see the review articles by Stock and Watson (2012) and Breitung and Choi (2013).

Another approach that has gained wide acceptance in practice is based on the Nelson-Siegel term structure model (see Nelson and Siegel 1987), where the underlying yield curve

is assumed to follow a certain parametric functional structure. The standard version of this model considers three predefined loading functions, which are denoted as the level, the slope, and the curvature function. The corresponding factors are interpreted as the long-term, short-term, and medium-term factors. Extensions and alternative factor loadings were proposed in Svensson (1995), Gürkaynak et al. (2007), Christensen et al. (2009), and Park et al. (2009). The dynamic Nelson-Siegel (DNS) model, which was introduced by Diebold and Li (2006), considers common factor dynamics in the form of a vector autoregressive process. In a two-step procedure, the factors are first estimated by ordinary least squares, and the dynamic process is then estimated given the fitted factor values from the first step. For a review on the DNS approach, see Diebold and Rudebusch (2013).

The dynamic versions of both the classical factor model and the Nelson-Siegel model are particularly useful for predicting the term structure of bond yields, since a predictor for the bond yields can be obtained from the minimum mean squared error (MSE) forecast of the factors. However, both approaches have their advantages and disadvantages. While the classical factor analysis does not depend on a specific form of the loadings and provides an optimal basis system, it considers yield curves as vector-valued objects and does not take into account the functional nature of the underlying curve. On the other hand, while the Nelson-Siegel model considers smooth loading functions and generates a fully functional representation of the yield curve, its predefined loadings are not optimal in some sense, as argued, for instance, in Lengwiler and Lenz (2010) and Hays et al. (2012).

In this paper, we follow recent advances in functional data analysis (for reviews, see, e.g., Ramsay and Silverman 2005, Hörmann and Kokoszka 2012, and Hsing and Eubank 2015). We consider a dynamic functional factor model, which is akin to the models proposed in Hays et al. (2012) and Bardsley et al. (2017). In contrast to Hays et al. (2012), where the estimation is achieved by applying an expectation maximization algorithm, we suggest a different estimation procedure and identify the model by imposing orthogonality conditions on the loading functions and suitable assumptions on the factors and the error term. We propose a functional principal component (FPC) estimator, which results from applying the least squares principle. It is shown that the model parameters are estimated consistently. Following Aue et al. (2015), the factors are modeled as an autoregressive process. Moreover, in the fashion of Stock and Watson (2002), forecasts are conducted in a two-step procedure, where the predictive model is estimated given the estimated factors. Analogously to Diebold and Li (2006), a forecast for the entire yield curve is obtained from the estimated functional factor model and the optimal MSE forecast for the factors. Furthermore, we derive both pointwise and simultaneous prediction bands for these forecasts. While the pointwise bands are asymptotically exact, the simultaneous

bands are conservative.

The paper is organized as follows: In Section 3.2, the functional factor model and its identification conditions are defined, while in Section 3.3, the FPC estimator is derived, and consistency is shown. Predictions for the entire yield curve as well as pointwise and simultaneous prediction bands are provided in Section 3.4. Then, in Section 3.5, the predictor is applied to a dataset of monthly U.S. Treasury yields, and Section 3.6 concludes.

3.2 The dynamic functional factor model

We consider a time series of yield curves $Y_t(r)$ with time to maturity $r \in [a, b]$ at time $t = 1, \ldots, T$, where a denotes the lowest time to maturity, and b denotes the longest one. The curves are assumed to be already given as square-integrable functions on the domain $[a, b]$. Following Hays et al. (2012), a general dynamic functional factor model with K factors can be formulated as

$$Y_t(r) = \mu(r) + \sum_{l=1}^{K} F_{l,t}\psi_l(r) + \epsilon_t(r), \quad t = 1, \ldots, T, \quad r \in [a, b],$$

where $\mu(r)$ is an intercept function, $F_{l,t}$ denotes the l-th factor at time t, and $\psi_l(r)$ is the l-th loading for time to maturity r. While $\mu(r)$ and $\psi_l(r)$ are deterministic terms, the vector of factors $F_t = (F_{1,t}, \ldots, F_{K,t})'$ is assumed to be a time series that follows some linear autoregressive process, which stems the attribute "dynamic" in the name of the model, and $\epsilon_t(r)$ is an idiosyncratic error term. In vector notation, the model can be equivalently expressed as

$$Y_t(r) = \mu(r) + \Psi'(r)F_t + \epsilon_t(r), \quad t = 1, \ldots, T, \quad r \in [a, b], \tag{3.1}$$

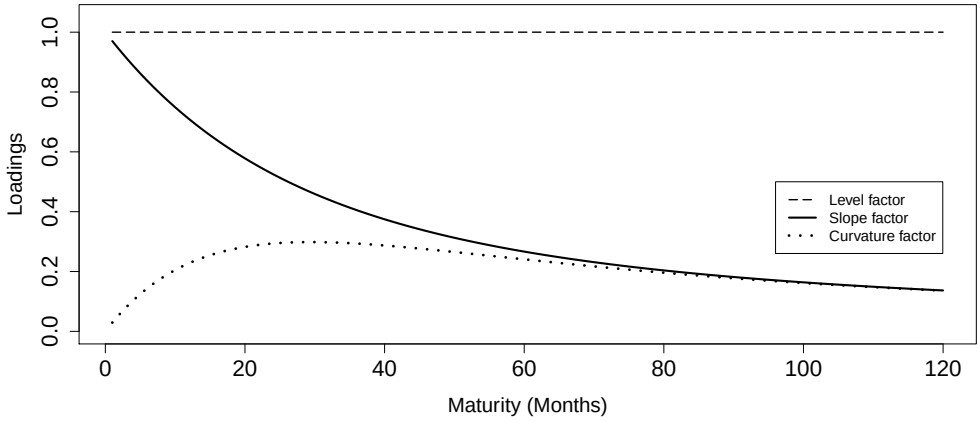
where the loadings at time to maturity r are given by $\Psi(r) = (\psi_1(r), \ldots, \psi_K(r))'$.

While model (3.1) can be viewed as an extension of the vector-valued dynamic factor model for bond yields by Bliss (1997) to a continuous domain, it also nests the widely used DNS framework, in which the loadings $\Psi(r)$ are predefined. Diebold and Li (2006) specified the Nelson-Siegel model with three fixed loading functions given by

$$\psi_1(r) = 1, \quad \psi_2(r) = \frac{1 - e^{-\lambda r}}{\lambda r}, \quad \psi_3(r) = \frac{1 - e^{-\lambda r}}{\lambda r} - e^{-\lambda r}, \tag{3.2}$$

where the decay parameter is fixed, and the intercept is zero. The Nelson-Siegel loadings

Figure 3.1: Nelson-Siegel loading functions



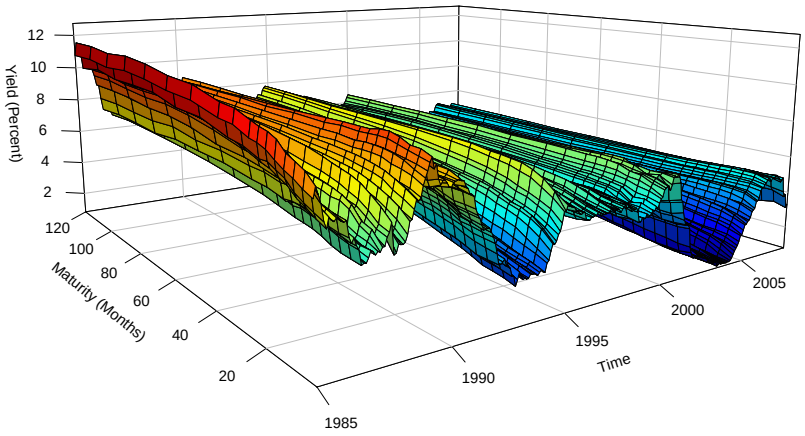
Note: The figure presents a plot of the loading functions considered in Diebold and Li (2006). The decay parameter in equation (3.2) is set to $\lambda = 0.0609$, which maximizes the curvature factor at a maturity of $r = 30$ months.

are illustrated in Figure 3.1, where the first loading $\psi_1(r)$ represents the level of the yield curve at time t, the function $\psi_2(r)$ corresponds to the slope and the third function $\psi_3(r)$ has a hump at mid-term maturities and represents the curvature of the yield curve. A plot of the yield curve data from Jungbacker et al. (2014) is depicted in Figure 3.2, and Figure 3.3 presents an example of a fitted Nelson-Siegel yield curve.

In this paper, we neither impose restrictive assumptions on the shape of the loading functions, as in the DNS framework, nor do we treat yield curves as vectors, as in the classical factor analysis. We consider model (3.1) to be a fully functional model and assume that both the factors F_t and loadings $\Psi(r)$ are unknown. Analogously to the vector-valued factor model, this implies an ambiguity, since $\Psi'(r)F_t = \Psi'(r)QQ^{-1}F_t$ for any non-singular $K \times K$ matrix Q. Thus, K^2 restrictions are necessary to make $\Psi(r)$ and F_t separately identifiable. Assuming that the loadings are pairwise orthogonal imposes $K(K-1)/2$ restrictions, and another $K(K-1)/2$ restrictions follow by assuming that the factors are pairwise uncorrelated. The remaining K restrictions can be obtained by normalizing either the loadings or the factors. Since the consideration of orthonormal loading functions is common in functional data analysis, we normalize the loadings. Another ambiguity follows from the fact that $\mu(r) + \Psi'(r)F_t = \mu(r) + \Psi'(r)F_0 + \Psi'(r)(F_t - F_0)$ for any $F_0 \in \mathbb{R}^K$, and K restrictions are obtained by setting the mean of the factors to zero.

To formalize these restrictions, we define a Hilbert space structure for the underlying function space $H = L^2([a,b])$, which denotes the space of real-valued square integrable functions on the domain $[a,b]$. The inner product for the function space is defined as $\langle x, y\rangle = \int_a^b x(r)y(r)\,\mathrm{d}r$, where $x, y \in H$, and the norm is given by $\|x\|^2 = \langle x, x\rangle$. The following identification conditions are imposed:

Figure 3.2: Yields of U.S. Treasury bonds



Note: The figure depicts a plot of the monthly yield curves of U.S. Treasuries from January 1985 until December 2007. The dataset is taken from Jungbacker et al. (2014) and is an extension of the dataset considered in Diebold and Li (2006).

Assumption 3.1. *(a) The loading functions are pairwise orthogonal; i.e.,* $\langle \psi_k, \psi_l \rangle = 0$, *for all* $k, l = 1, \ldots, K$ *with* $k \neq l$.

(b) The loading functions are normalized, so that $\|\psi_l\| = 1$ *for all* $l = 1, \ldots, K$.

(c) The factors satisfy $F_t \sim \mathcal{N}(0, \Sigma_F)$ *for all* $t = 1, \ldots, T$, *where* $\Sigma_F = diag(\lambda_1, \ldots, \lambda_K)$ *with distinct entries* $\lambda_1 > \lambda_2 > \ldots > \lambda_K > 0$.

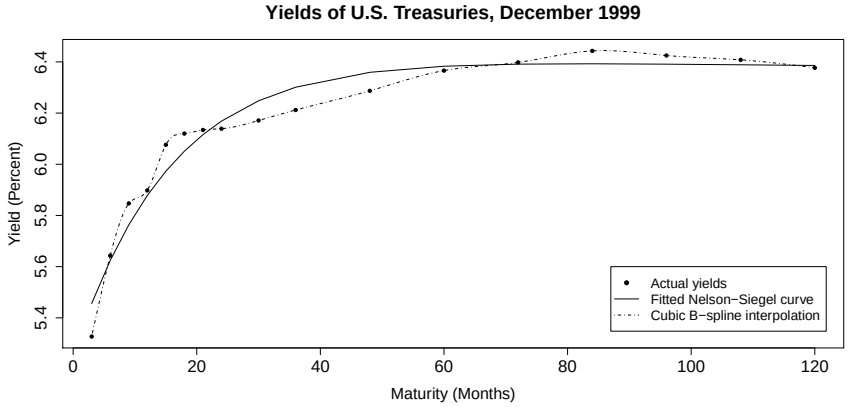
The assumptions are similar to the identification conditions for the classical factor model (see Stock and Watson 2002 and Bai and Ng 2013). The loadings are assumed to be pairwise orthonormal, while the factors are uncorrelated. Distinct variances of the factors serve to identify the individual entries of F_t, which could otherwise be consistently estimated only up to an orthonormal transformation. Note that Assumption 3.1 is a local identification condition, since $\Psi(r)$ and F_t are identified separately only up to a sign change. Changing the sign of both the loadings and the factors will leave the common component $\Psi'(r)F_t$ unchanged. A global identification condition can be obtained by fixing the sign for either the loadings or the factors.

Assumption 3.2. *(a) The functions* $\mu(r)$ *and* $\psi_l(r)$, $l = 1, \ldots, K$, *are bounded.*

(b) The errors $\{\epsilon_t\}_{t \in \mathbb{N}}$ *are an i.i.d. sequence of zero mean Gaussian random functions on the domain* $[a, b]$ *with covariance kernel* $c_\epsilon(r, s)$.

For each $r, s \in [a, b]$, the error term is normally distributed with $E[\epsilon_t(r)] = 0$ and $Cov[\epsilon_t(r), \epsilon_t(s)] = c_\epsilon(r, s)$. Note that Hays et al. (2012) proposed normality conditions on the factors and errors similar to those made in Assumptions 3.1 and 3.2. In contrast

Figure 3.3: Fitted Nelson-Siegel curve and cubic B-spline representation



Note: The figure illustrates the yield curve from the dataset of Jungbacker et al. (2014) for December 1999. While the solid line presents the fitted Nelson-Siegel curve, the dash-dotted line is a functional representation of the yields from the B-spline basis expansion (see Remark 3.1 and Section 3.5.1).

to their estimation procedure, we impose normality only to derive the prediction bands. For the estimation of the model and the forecast of the yield curve itself, normality of the errors and the factors is not necessary.

Moreover, Assumptions 3.1(c), 3.2(a), and 3.2(b) imply that $\{Y_t\}_{t\in\mathbb{N}}$ is a sequence of H-valued random functions with $E\|Y_t\|^2 \leq \|\mu\|^2 + E\|\epsilon_t\|^2 < \infty$. It follows that $Y_t(r)$ is covariance stationary with mean $E[Y_t(r)] = \mu(r) < \infty$, and covariance kernel

$$c_Y(r,s) = Cov[Y_t(r), Y_t(s)] = \sum_{l=1}^{K} \lambda_l \psi_l(r)\psi_l(s) + c_\epsilon(r,s) < \infty$$

for all t. The covariance operator of $Y_t(r)$ is given by

$$C_Y(x)(r) = \int_a^b c_Y(r,s)x(s)\,\mathrm{d}s, \quad x \in H,\ r \in [a,b],$$

and the eigenequation is defined as

$$\int_a^b c_Y(r,s)\psi(s)\,\mathrm{d}s = \lambda\psi(s), \quad \lambda \in \mathbb{R},\ \psi \in H,\ r \in [a,b],$$

where $(\lambda, \psi(s))$ is called an eigenvalue and eigenfunction pair of C_Y. The covariance operator of $\epsilon_t(r)$ is analogously defined as

$$C_\epsilon(x)(r) = \int_a^b c_\epsilon(r,s)x(s)\,\mathrm{d}s, \quad x \in H,\ r \in [a,b].$$

Since C_Y and C_ϵ are positive semi-definite and symmetric (see Hörmann and Kokoszka 2012), all eigenvalues are nonnegative, and the eigenfunctions are orthogonal. An orthonormal eigenfunction that corresponds to the l-th largest eigenvalue of C_Y is called the l-th FPC. We next establish a link between the functional factor model (3.1) and the FPCs.

Assumption 3.3. *(a)* $\langle \epsilon_t, \psi_l \rangle = 0$ *for all* $l = 1, \ldots, K$, $t = 1, \ldots, T$.

(b) $E\|\epsilon_t\|^2 < \lambda_K$.

(c) C_ϵ *has* $L - K$ *nonzero eigenvalues, which are denoted as* $\lambda_{K+1} > \lambda_{K+2} > \ldots > \lambda_L$.

Assumption 3.3 is similar to the conditions made in Forni et al. (2000) for the dynamic vector-valued factor model. The first assumption implies that the idiosyncratic component $\epsilon_t(r)$ is orthogonal to the common component $\Psi'(r)F_t$, which separates H into a factor space and its orthogonal complement, while the second assumption guarantees some limited amount of cross-correlation in the idiosyncratic component. The covariances of $\epsilon_t(r)$ are bounded in the sense that there is no direction in H along which the idiosyncratic component has a variance that exceeds $Var[F_{K,t}] = \lambda_K$. The third assumption implies that the error term is an element of an $(L - K)$-dimensional subspace of H, and the yield curve series consequently takes values in an L-dimensional subspace of H, where $K \leq L < \infty$.

Next, we show that under Assumptions 3.1–3.3, the variances of the factors $\lambda_1, \ldots, \lambda_K$ coincide with the K largest eigenvalues of C_Y, while the factor loadings $\psi_1(r), \ldots, \psi_K(r)$ are the first K FPCs. Furthermore, the factors are equal to the projections of the demeaned yield curve onto the corresponding FPC. This relation is summarized as follows:

Lemma 3.1. *Let Assumptions 3.1–3.3 hold true.*

(a) The largest K *eigenvalues of* C_Y *are given by* $\lambda_1, \ldots, \lambda_K$, *while* $\psi_1(r), \ldots, \psi_K(r)$ *are corresponding eigenfunctions.*

(b) The factors satisfy $F_{l,t} = \langle Y_t - \mu, \psi_l \rangle$ *for all* $l = 1, \ldots, K$ *and* $t = 1, \ldots, T$.

The eigenvalues of C_Y are given by $\lambda_1 > \lambda_2 > \ldots > \lambda_L > 0$, and the FPCs are the corresponding orthonormal eigenfunctions $\psi_1(r), \ldots, \psi_L(r)$, which are identified up to a sign change. The term $\langle Y_t - \mu, \psi_l \rangle$ is called the l-th FPC score of $Y_t(r)$ and describes the contribution of the l-th FPC to the curve $Y_t(r)$. While Aue et al. (2015) imposed a vector autoregressive structure on the empirical counterpart of the FPC scores, we incorporate

this assumption into the model (3.1) and assume that the factors follow a VAR process. This dynamic structure is exploited in Section 3.4 to obtain predictions of the yield curve and to construct prediction bands.

Assumption 3.4. *The factors follow the VAR(p) process* $F_t = A(L)F_{t-1} + \eta_t$ *with* $K \times K$ *lag polynomial matrices* $A(L) = A_1 + A_2 L + \ldots + A_p L^{p-1}$. *The VAR process is stable; i.e., all roots of the characteristic polynomial lie outside the unit circle. The innovation vectors* $\{\eta_t\}_{t\in\mathbb{N}}$ *are i.i.d. with* $\eta_t = (\eta_{1,t}, \ldots, \eta_{K,t})' \sim \mathcal{N}(0, \Sigma_\eta)$, *where* $\Sigma_\eta = diag(\sigma_1^2, \ldots, \sigma_K^2)$. *Furthermore,* $E[\eta_{l,t}\epsilon_{t+j}(r)] = 0$ *for all* $t \in \mathbb{N}$, $j \geq 0$, $l = 1, \ldots, K$, *and* $r \in [a, b]$.

Since the factors have zero mean and are uncorrelated, the VAR model does not include a constant, and the entries of the innovation vector are uncorrelated. Although the contemporaneous correlations of the factors are zero by Assumption 3.1(c), the cross-correlations of the factors might be nonzero, and the lag polynomial matrices are therefore not necessarily diagonal. However, Diebold and Li (2006) argued that VAR models for bond yield factors tend to produce poor forecasts compared to those from univariate autoregressive models, since a model with a large number of parameters is prone to overfitting. Hyndman and Ullah (2007) also argued that a simple univariate autoregressive model for each FPC score is more adequate in the forecasting context, since contemporaneous correlations of the FPC scores are zero, and the cross-correlations at non-zero lags are often quite small. Therefore, we discuss both multivariate and univariate predictive time series models for the factors in Section 3.4.

Remark 3.1. Throughout the theoretical part of the paper, we assume that the yield curves are already given in functional form as elements of H. In practice, however, only vectors of bond yields with a finite number N of times to maturity are observed. The problem of transforming the vector observations into functions has been extensively studied in the literature on functional data analysis and is well understood (see, e.g., Ramsay and Silverman 2005 for a review of the available techniques). The data is interpolated or approximated by means of a basis expansion, which is typically performed using a spline basis in the case of non-periodic data. In the empirical part of the paper, we employ B-spline techniques to reconstruct yield curves (see Figure 3.3). Since only a finite number of basis functions are necessary to interpolate N observations, the reconstructed yield curve observations are in practice elements of a finite-dimensional subspace of H. More details on this topic are provided in Section 3.5.1.

3.3 The functional principal components estimator

While in the classical factor analysis the principal components estimator is obtained as the solution to some least squares problem, an analogous least squares problem arises for the functional factor model (3.1). Let

$$\widehat{\mu}(r) = \frac{1}{T}\sum_{t=1}^{T} Y_t(r), \quad r \in [a,b],$$

denote the empirical mean function of $Y_t(r)$. We consider the demeaned curve $Y_t(r) - \widehat{\mu}(r)$ and minimize the objective function

$$\sum_{t=1}^{T} \left\| Y_t - \widehat{\mu} - \sum_{l=1}^{K} F_{l,t}\psi_l \right\|^2 \tag{3.3}$$

subject to the orthonormality conditions given by Assumptions 3.1(a) and 3.1(b). Taking the loading functions as given, we minimize (3.3) with respect to the factors $F_{l,t}$ for all $l = 1, \ldots, K$ and $t = 1, \ldots, T$. Note that addition and multiplication for functional objects are defined pointwise, so that $(Y_t - \widehat{\mu})(r) = Y_T(r) - \widehat{\mu}(r)$ and $(F_{l,t}\psi_l)(r) = F_{l,t} \cdot \psi_l(r)$ for all $r \in [a,b]$.

Lemma 3.2. *Under Assumptions 3.1(a) and 3.1(b),*

$$\min_{F_{l,t}} \sum_{s=1}^{T} \left\| Y_s - \widehat{\mu} - \sum_{k=1}^{K} F_{k,s}\psi_k \right\|^2 \tag{3.4}$$

is reached, if $F_{l,t} = \langle Y_t - \widehat{\mu}, \psi_l \rangle$ *for any* $l = 1, \ldots, K$ *and* $t = 1, \ldots, T$.

The objective function is minimized by setting $F_{l,t}$ equal to the projection of the demeaned yield curve at time t onto the l-th loading function. The least squares problem is consequently equivalent to minimizing

$$\sum_{t=1}^{T} \left\| Y_t - \widehat{\mu} - \sum_{l=1}^{K} \langle Y_t - \widehat{\mu}, \psi_l \rangle \psi_l \right\|^2 \tag{3.5}$$

with respect to $\psi_1(r), \ldots, \psi_K(r)$. Hence, we have to find a K-dimensional orthonormal system in H that minimizes (3.5), which is a well known problem in the literature on

functional data analysis and extensively studied in Hörmann and Kokoszka (2012). Let

$$\widehat{c}_Y(r,s) = \frac{1}{T}\sum_{t=1}^{T}\big(Y_t(r) - \widehat{\mu}(r)\big)\big(Y_t(s) - \widehat{\mu}(s)\big), \quad r,s \in [a,b],$$

be the sample covariance operator. The eigenequation is given by

$$\int_a^b \widehat{c}_Y(r,s)\widehat{\psi}(s)\,\mathrm{d}s = \widehat{\lambda}\widehat{\psi}(r), \quad r \in [a,b], \tag{3.6}$$

where $\widehat{\lambda} \in \mathbb{R}$ is an eigenvalue and $\widehat{\psi}(r)$ a corresponding eigenfunction of the sample covariance operator, which is defined as $\widehat{C}_Y(x)(r) = \int_a^b \widehat{c}_Y(r,s)x(s)\,\mathrm{d}s$, $x \in H$, $r \in [a,b]$. Note that all eigenvalues are nonnegative, since $\widehat{C}_Y$ is positive semi-definite, and the eigenspaces are pairwise orthogonal, since $\widehat{C}_Y$ is symmetric. Furthermore, the underlying function space H is countably infinite-dimensional, which yields the existence of a sequence of eigenpairs $\{(\widehat{\lambda}_j, \widehat{\psi}_j(r))\}_{j\in\mathbb{N}}$ with orthonormal eigenfunctions $\{\widehat{\psi}_j(r)\}_{j\in\mathbb{N}}$ and with eigenvalues $\{\widehat{\lambda}_j\}_{j\in\mathbb{N}}$ that are arranged in decreasing order $\widehat{\lambda}_1 \geq \widehat{\lambda}_2 \geq \ldots \geq 0$. Following Hörmann and Kokoszka (2012), the eigenfunction $\widehat{\psi}_l(r)$ is called the l-th empirical FPC, and the term $\widehat{F}_{l,t} = \langle Y_t - \widehat{\mu}, \widehat{\psi}_l\rangle$ is called the l-th empirical FPC score of $Y_t(r)$. The vector of empirical FPCs is given by $\widehat{\Psi}(r) = (\widehat{\psi}_1(r), \ldots, \widehat{\psi}_K(r))'$, and the vector of empirical scores is defined as $\widehat{F}_t = (\widehat{F}_{1,t}, \ldots, \widehat{F}_{K,t})'$.

Theorem 3.1. *Under Assumptions 3.1(a) and 3.1(b), the solution to the least squares problem*

$$\min_{\substack{\Psi(r)\\ F_1,\ldots,F_T}} \sum_{t=1}^{T}\Big\|Y_t - \widehat{\mu} - \Psi'(r)F_t\Big\|^2$$

is reached if $F_t = \widehat{F}_t$ for all $t = 1,\ldots,T$ and $\Psi(r) = \widehat{\Psi}(r)$ for all $r \in [a,b]$.

The l-th empirical FPC score has mean zero, since $T^{-1}\sum_{t=1}^{T}\widehat{F}_{l,t} = \langle \widehat{\mu} - \widehat{\mu}, \widehat{\psi}_l\rangle = 0$, and is uncorrelated with the k-th empirical score for $k \neq l$, while its sample variance is equal to $\widehat{\lambda}_l$, since

$$\frac{1}{T}\sum_{t=1}^{T}\widehat{F}_{k,t}\widehat{F}_{l,t} = \frac{1}{T}\sum_{t=1}^{T}\int_a^b\int_a^b \widehat{c}_Y(r,s)\widehat{\psi}_l(s)\widehat{\psi}_k(r)\,\mathrm{d}s\,\mathrm{d}r = \int_a^b \hat{\lambda}_l\widehat{\psi}_l(r)\widehat{\psi}_k(r)\,\mathrm{d}r = \widehat{\lambda}_l \cdot 1_{\{k=l\}},$$

for any $k, l \in \mathbb{N}$.

Note that the sequence $\{\widehat{\psi}_j(r)\}_{j\in\mathbb{N}}$ of empirical FPCs cannot be unique. If $\widehat{\psi}_l(r)$ is an

orthonormal eigenfunction of $\widehat{C}_Y$ with respect to the eigenvalue $\widehat{\lambda}_j$, then $-\widehat{\psi}_l(r)$ possesses the same orthonormality properties. However, if the first K eigenvalues of $\widehat{C}_Y$ are distinct, with $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \ldots > \widehat{\lambda}_K > \widehat{\lambda}_{K+1} \geq 0$, then the orthonormal eigenvectors $\widehat{\psi}_1(r), \ldots, \widehat{\psi}_K(r)$ are unique up to a sign change, and the common components $\widehat{F}_{1,t}\widehat{\psi}_1(r), \ldots, \widehat{F}_{K,t}\widehat{\psi}_K(r)$ are uniquely determined.

Consistency results for the empirical mean function, the empirical FPCs, the empirical FPC scores, and the eigenvalues of the empirical covariance operator are obtained under fairly general conditions. Hörmann and Kokoszka (2010) established consistency for weakly dependent data. Salish and Gleim (2019) expanded the consistency results to strongly dependent data. The results in Salish and Gleim (2019) can be used to establish uniform consistency under the functional factor model (3.1), as described in the following lemma.

Lemma 3.3. *Let Assumptions 3.1–3.4 hold true. Then, as* $T \to \infty$,

(a) $\sup_{r\in[a,b]} \left|\widehat{\mu}(r) - \mu(r)\right| = O_P(T^{-1/2})$,

(b) $\sup_{r,s\in[a,b]} \left|\widehat{c}_Y(r,s) - c_Y(r,s)\right| = O_P(T^{-1/2})$,

(c) $\left|\widehat{\lambda}_l - \lambda_l\right| = O_P(T^{-1/2})$,

(d) $\sup_{r\in[a,b]} \left|s_l\widehat{\psi}_l(r) - \psi_l(r)\right| = O_P(T^{-1/2})$,

(e) $\max_{1\leq t\leq T} \left|s_l\langle Y_t - \widehat{\mu}, \widehat{\psi}_l\rangle - \langle Y_t - \mu, \psi_l\rangle\right| = O_P(T^{-1/2})$,

for all $l = 1, \ldots, L$, *where* $s_l = sign(\langle\widehat{\psi}_l, \psi_l\rangle)$.

Accordingly, it follows that $\sup_{r\in[a,b]} |\widehat{Y}_t(r) - \mu(r) - \Psi'(r)F_t| = O_P(T^{-1/2})$, as $T \to \infty$, where the fitted yield curve is formulated as

$$\widehat{Y}_t(r) = \widehat{\mu}(r) + \sum_{l=1}^{K} \widehat{F}_{l,t}\widehat{\psi}_l(r), \quad r \in [a,b].$$

In practice, equation (3.6) can be solved by using, for instance, the `fda`-package, which is available for R and MATLAB (see Ramsay et al. 2009). If the yield curves are transformed from vector-valued data, as discussed in Remark 3.1, then they are elements of a finite-dimensional subspace of H, which implies that, in practice, only a finite number of nonzero eigenvalues are obtained.

3.4 Predictions and prediction bands

The optimal one-step ahead forecast of the factors with respect to MSE loss is given by

$$F_{T+1|T} = E[F_{T+1}|I_T] = \sum_{i=1}^{p} A_i F_{T-i+1},$$

where $I_T = \{Y_t,\ t \leq T\}$. For longer horizons $h \in \mathbb{N}$, the h-step ahead forecast is obtained by the chain rule of forecasting as

$$F_{T+h|T} = E[F_{T+h}|I_T] = \sum_{i=1}^{p} A_i F_{T+h-i|T},$$

where $F_{T+j|T} = F_{T+j}$ for $j \leq 0$. A forecast for the entire yield curve can be formulated with the help of the following theorem:

Theorem 3.2. *Let $g(I_T) \in H$ be a forecast function for Y_{T+h} given the information up to time T. Then, $E\|Y_{T+h} - g(I_T)\|^2$ is minimized if $g(I_T)(r) = E[Y_{T+h}(r)|I_T]$ for any $r \in [a, b]$.*

The minimum MSE h-step ahead forecast for the yield curve is thus given by

$$Y_{T+h|T}(r) = E[Y_{T+h}(r)|I_T] = \mu(r) + \Psi'(r)E[F_{T+1}|I_T] = \mu(r) + \Psi'(r)F_{T+h|T}, \tag{3.7}$$

where $r \in [a, b]$, and the forecast error curve is obtained as

$$\begin{aligned} e_{T+h|T}(r) = Y_{T+h}(r) - Y_{T+h|T}(r) &= \Psi'(r)(F_{T+h} - F_{T+h|T}) + \epsilon_{T+h}(r) \\ &= \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \eta_{T+h-i}\Big) + \epsilon_{T+h}(r), \quad r \in [a, b], \end{aligned}$$

where the matrices Φ_i are defined by the recursion $\Phi_i = \sum_{j=1}^{p} \Phi_{i-j} A_j$ with $\Phi_0 = I_K$ and $\Phi_j = 0$ for $j < 0$. From Assumptions 3.2 and 3.4, we conclude that $e_{T+h}(r)$ is a Gaussian random function with $E[e_{T+h}(r)] = 0$ and covariance kernel

$$c_{e,h}(r, s) = Cov[e_{T+h}(r), e_{T+h}(s)] = \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i'\Big)\Psi(s) + c_\epsilon(r, s), \quad r, s \in [a, b].$$

For any fixed $r \in [a,b]$, it follows that

$$\frac{Y_{T+h}(r) - Y_{T+h|T}(r)}{\sqrt{c_{e,h}(r,r)}} \sim \mathcal{N}(0,1). \tag{3.8}$$

The above formulas are infeasible, since $\mu(r)$, $\Psi(r)$, F_T, $A(L)$, and Σ_η are unknown and must be replaced by consistent estimators. For the intercept, the loadings, and the factors, consistent estimators are provided by Lemma 3.3. Following the two-step procedures of Stock and Watson (2002), Diebold and Li (2006), and Aue et al. (2015), the estimated factors can be used to estimate the VAR model parameters. If the factors are known, then the least squares estimator for the $(p \times Kp)$-matrix $B = (A_1, \ldots, A_p)$ is equal to $\widetilde{B} = FZ'(ZZ')^{-1}$, where $F = (F_{p+1}, \ldots, F_T)$, $Z = (Z_p, \ldots, Z_{T-1})$, and $Z_t = (F_t', \ldots, F_{t-p+1}')'$ (see, e.g., Lütkepohl 2005, Section 3). We thus consider its feasible counterpart which is given by $\widehat{B} = \widehat{F}\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}$, where $\widehat{F} = (\widehat{F}_{p+1}, \ldots, \widehat{F}_T)$, $\widehat{Z} = (\widehat{Z}_p, \ldots, \widehat{Z}_{T-1})$, and $\widehat{Z}_t = (\widehat{F}_t', \ldots, \widehat{F}_{t-p+1}')'$. The residuals are obtained as $\widehat{\eta}_t = \widehat{F}_t - \sum_{i=1}^p \widehat{A}_i \widehat{F}_{t-i}$, where $(\widehat{A}_1, \ldots, \widehat{A}_p) = \widehat{B}$, and the covariance matrix of η_t is estimated by $\widehat{\Sigma}_\eta = \mathrm{diag}(\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_K^2)$, where $\widehat{\sigma}_l^2 = (T-p-1)^{-1} \sum_{t=p+1}^T \widehat{\eta}_{lt}^2$.

Lemma 3.4. *Let $S = diag(s_1, \ldots, s_K)$, where $s_l = sign\langle \widehat{\psi}_l, \psi_l \rangle$. Then, under Assumptions 3.1–3.4, it follows that*

(a) $\|S\widehat{A}_i S - A_i\|_M = O_P(T^{-1/2})$, *for all* $i = 1, \ldots, p$,

(b) $\|\widehat{\Sigma}_\eta - \Sigma_\eta\|_M = O_P(T^{-1/2})$,

as $T \to \infty$, where $\|\cdot\|_M$ denotes some matrix norm.

The feasible h-step ahead factor forecast is defined by the recursion

$$\widehat{F}_{T+h|T} = \sum_{i=1}^p \widehat{A}_i \widehat{F}_{T+h-i|T},$$

where $\widehat{F}_{T+j|T} = \widehat{F}_{T+j}$ for $j \leq 0$, and the feasible h-step ahead forecast for the yield curve is obtained as

$$\widehat{Y}_{T+h|T}(r) = \widehat{\mu}(r) + \widehat{\Psi}'(r)\widehat{F}_{T+h|T}. \tag{3.9}$$

Moreover, the estimated covariance kernel of the forecast error function is given by

$$\widehat{c}_{e,h}(r,s) = \widehat{\Psi}'(r)\Big(\sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_\eta \widehat{\Phi}_i'\Big)\widehat{\Psi}(s) + \sum_{l=K+1}^{L} \widehat{\lambda}_l \widehat{\psi}_l(r)\widehat{\psi}_l(s), \quad r,s \in [a,b].$$

Note that the h-step ahead predictor for stationary functional time series by Aue et al. (2015) coincides with (3.9). The authors assumed a VAR(p) model for the first K empirical FPC scores, and they discussed its relation to the optimal forecast from the functional autoregressive model (FAR) of order p. The consistent estimator by Bosq (2000) for the FAR(p) model is based on a truncated FPC decomposition. Aue et al. (2015) demonstrated that the predictor (3.9) and the h-step ahead predictor by Bosq (2000) are asymptotically equivalent, as $T \to \infty$, if the truncation parameter for the estimator by Bosq (2000) is set to K.

The following result indicates that (3.9) is an asymptotically optimal minimum MSE forecast for model (3.1), as $T \to \infty$.

Lemma 3.5. *Let Assumptions 3.1–3.4 hold true. Then, as $T \to \infty$,*

(a) $\sup_{r\in[a,b]} \left|\widehat{Y}_{T+h|T}(r) - Y_{T+h|T}(r)\right| = O_P(T^{-1/2})$,

(b) $\sup_{r,s\in[a,b]} \left|\widehat{c}_{e,h}(r,s) - c_{e,h}(r,s)\right| = O_P(T^{-1/2})$.

Together with equation (3.8), we can formulate both pointwise and simultaneous $(1-\alpha)$ prediction bands. The first part of the following theorem provides an interval forecast for each fixed time to maturity $r \in [a,b]$ and some given significance level α, while the second part presents a conservative simultaneous prediction band for the entire yield curve.

Theorem 3.3. *Under Assumptions 3.1–3.4, it follows that*

(a) $\lim_{T\to\infty} P\left(\frac{|Y_{T+h}(r) - \widehat{Y}_{T+h|T}(r)|}{\omega(r)} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha, \quad \text{for all } r \in [a,b],$

(b) $\lim_{T\to\infty} P\left(\frac{|Y_{T+1}(r) - \widehat{Y}_{T+1|T}(r)|}{\omega(r)} \leq \sqrt{\chi^2_{L,1-\alpha}}, \quad \text{for all } r \in [a,b]\right) \geq 1-\alpha,$

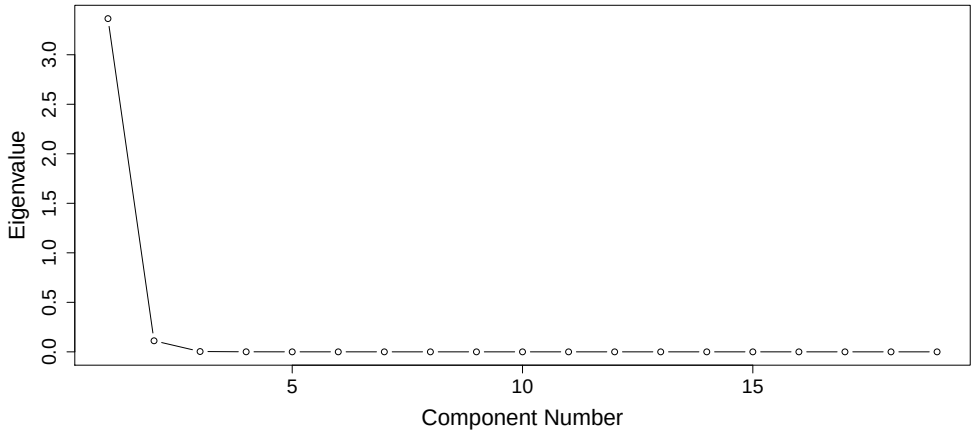
where u_ν is the ν-quantile of the standard normal distribution, $\chi^2_{L,\nu}$ is the ν-quantile of the chi-squared distribution with L degrees of freedom, and

$$\omega^2(r) = \widehat{\Psi}'(r)\left(\sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_\eta \widehat{\Phi}_i'\right)\widehat{\Psi}(r) + \sum_{l=K+1}^{L} \widehat{\lambda}_l \widehat{\psi}_l^2(r).$$

Estimating the functional factor model via FPCs yields a model that asymptotically explains $(\sum_{l=1}^{K}\lambda_l)/(\sum_{l=1}^{\infty}\lambda_l)$ of the variability in Y_t, while $\sum_{l=1}^{\infty}\lambda_l = E\|Y_t-\mu\|^2 < \infty$. The proportion of the variance explained by the l-th FPC is given by $\lambda_l/\sum_{k=1}^{\infty}\lambda_k$. Hence, a scree plot provides a useful selection criterion for the number of relevant factors K. The lag

Figure 3.4: Explained variance of the factors and scree plot

FPC	1	2	3	4	5	6	7	8	9	10
eigenvalue	3.3649	0.1119	0.0033	0.0006	0.0002	0.0001	0.0001	0.0001	0.0001	0.0000
expl.var.	96.65%	3.21%	0.10%	0.02%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
cumulative	96.65%	99.87%	99.96%	99.98%	99.99%	99.99%	99.99%	99.99%	99.99%	100.00%



Note: The table at the top presents the proportion of the explained variance by the l-th FPC score and the cumulative proportions, which are given by $\widehat{\lambda}_l / \sum_{k=1}^{\infty} \widehat{\lambda}_k$ and $(\sum_{j=1}^{l} \widehat{\lambda}_j) / \sum_{k=1}^{\infty} \widehat{\lambda}_k$. The figure at the bottom presents a scree plot of the eigenvalues of the empirical covariance operator of the yield curves.

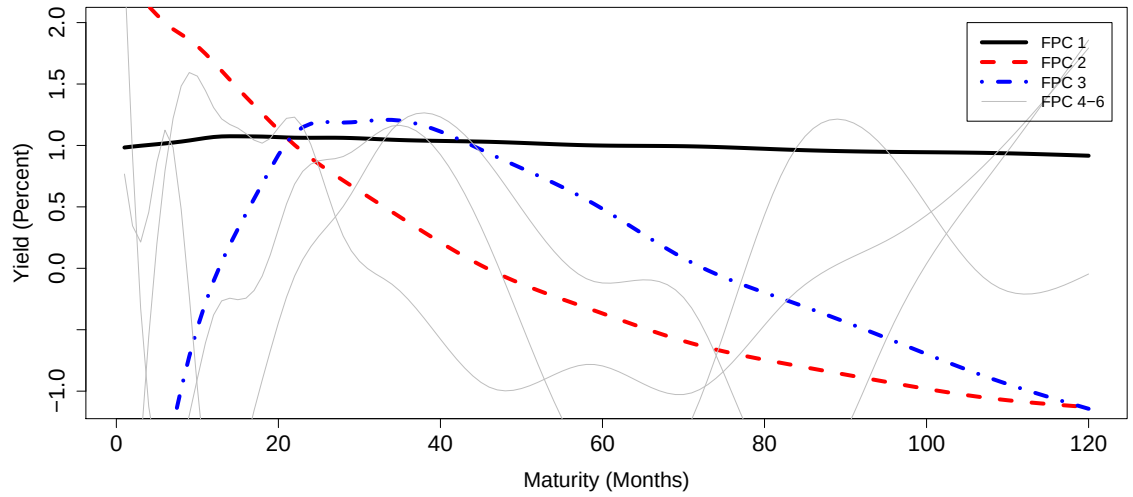
order p can also be identified using an information criterion from the VAR literature. Given the relation of the VAR(p) model for the factors and the FAR(p) model for the curves, as discussed above, the testing procedure for selecting the lag order in the FAR(p) model of Kokoszka and Reimherr (2013) also provides a useful criterion to select p. Furthermore, the functional FPE criterion of Aue et al. (2015) allows for a simultaneous identification of the lag order p and the number of factors K.

3.5 Application to yields for U.S. Treasuries

To apply the prediction and prediction band methodology, we consider a panel of monthly unsmoothed Fama-Bliss zero-coupon yields of U.S. Treasuries with fixed maturities of 3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months, from January 1985 until December 2007 (see Figure 3.2). The dataset is taken from Jungbacker et al. (2014), which is available in the Journal of Applied Econometrics Data Archive.[1] It extends the dataset of Diebold and Li (2006). The time span ranges from the period after the Volcker disinflation until the 2008 financial crisis, which can be treated as a consistent monetary policy regime (see Mönch 2012).

[1] see `http://qed.econ.queensu.ca/jae/`.

Figure 3.5: Empirical functional principal components



Note: The plot presents the first six empirical functional principal components (FPCs) of the empirical covariance operator of the yield curves from the full dataset of Figure 3.2.

3.5.1 From discrete data to functional data

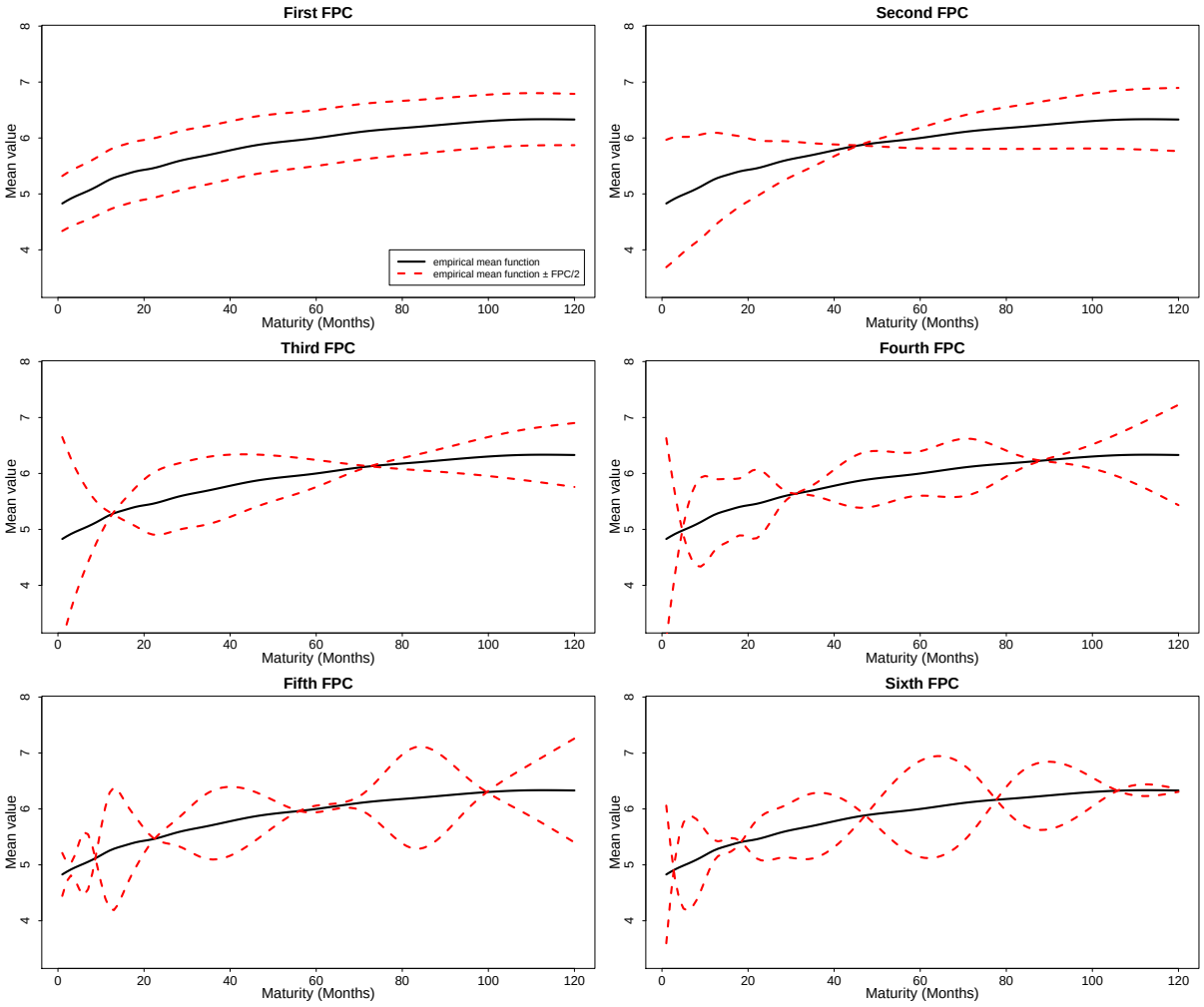
The data is given in the form of a panel of bond yields y_{it} at time $t = 1, \ldots, T$ and times to maturity r_i, $i = 1, \ldots, N$, with $r_1 < r_2 < \ldots < r_N$. We convert the panel to a functional time series $Y_t(r)$, $t = 1, \ldots, T$, $r \in [a, b]$, by means of a cubic B-spline expansion, where $a = r_1$ and $b = r_N$. The curve is represented as $Y_t(r) = \sum_{j=1}^{J} c_{tj}\phi_j(r)$, $r \in [a, b]$, where $\{\phi_j\}_{j=1,\ldots,J}$ are basis functions on the domain $[a, b]$, and $c_{t1}, \ldots, c_{tJ}$ are basis coefficients. The B-spline basis is defined with respect to a sequence of knots. At each interior observation point $r_2, \ldots, r_{N-1}$, we set a single interior knot, which yields $N-2$ basis functions. For cubic B-splines, another four basis functions are necessary to ensure twice continuous differentiability. The number of basis functions is hence $J = N + 2$. For the definition of and a comprehensive discussion on the B-spline basis, see de Boor (2001).

Minimizing the sum of squared errors $\sum_{i=1}^{N}(Y_t(r_i) - y_{it})^2$ is not feasible, since $J > N$. Following Ramsay and Silverman (2005), we introduce the roughness penalty $\lambda > 0$ and minimize the criterion

$$\min_{c_{t1},\ldots,c_{tJ}} \sum_{i=1}^{N}(Y_t(r_i) - y_{it})^2 + \lambda \int_a^b \left(D^2 Y_t(r)\right)^2 \mathrm{d}r$$

for each $t = 1, \ldots, T$, where $D^2 Y_t(r)$ is the second derivative of $Y_t(r)$. By choosing a roughness penalty λ that is close to zero, the approximation errors at the observed points $r_1, \ldots, r_N$ are almost zero, such that $Y_t(r)$ interpolates the observed bond yields

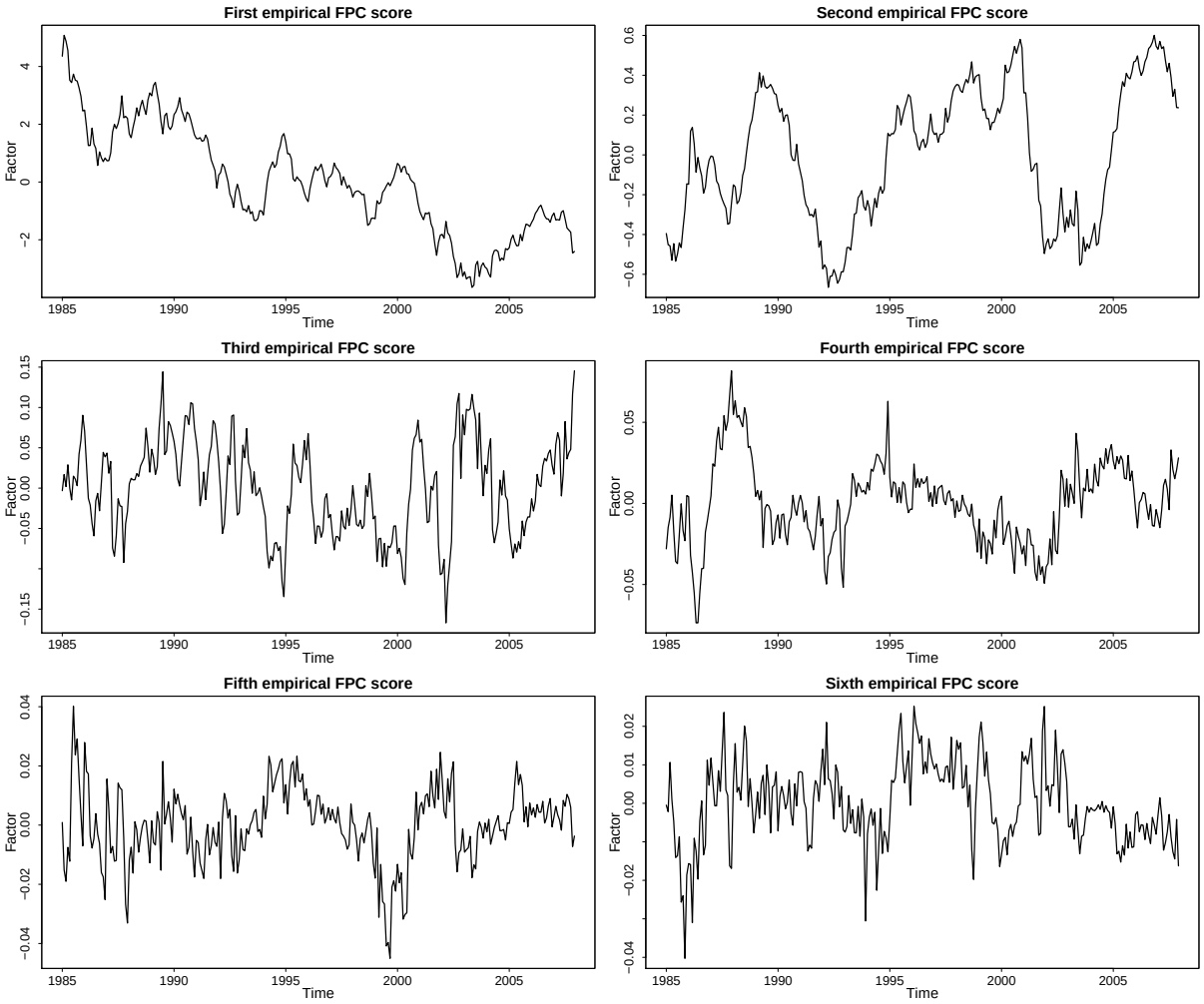
Figure 3.6: Effects of the first six functional principal components



Note: The figure presents the plot of $\widehat{\mu}(r) \pm 0.5 \cdot \widehat{\psi}_l(r)$ for $l = 1, \ldots, 6$, where the empirical functional principal components (FPCs) are identified up to a sign change.

$y_{1t}, \ldots, y_{Nt}$. For our application, we set $\lambda = 10^{-8}$. A comprehensive overview of the problem of transforming discrete data to functional data is provided in Ramsay and Silverman (2005) and Ramsay et al. (2009). The basis representation not only guarantees a smooth yield curve and but also produces yields at maturities that we do not observe. Furthermore, it handles missing values across maturities in a natural way. Note that the B-Spline functions for a basis of an $(N + 2)$-dimensional subspace of H. Therefore, at most $N + 2$ eigenvalues of the empirical covariance function are nonzero.

Figure 3.7: Empirical functional principal component scores



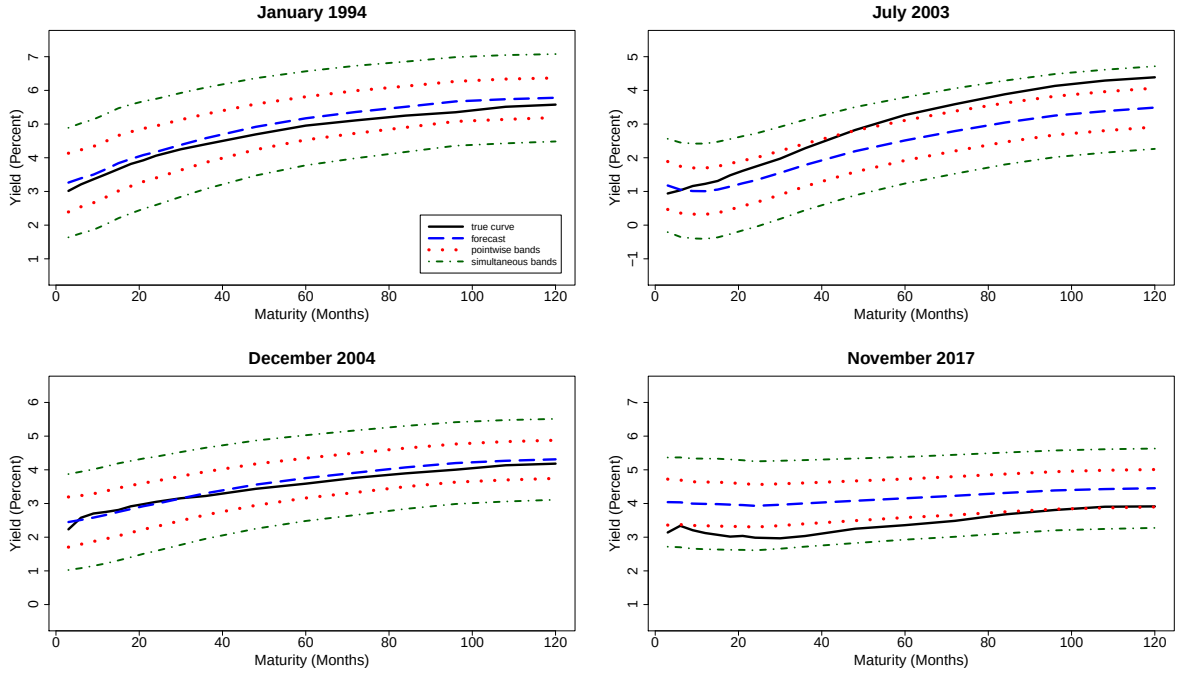
Note: The figure presents the plot of the empirical functional principal component (FPC) scores $\widehat{F}_{l,t}$ for $l = 1, \ldots, 6$, which are estimated using the full dataset of Figure 3.2.

3.5.2 Functional principal component analysis

Since the data consists of 17 times to maturity, the transformed functional objects are elements of a 19-dimensional subspace of H, which implies that $\widehat{C}_Y$ has at most 19 nonzero eigenvalues. Figure 3.4 depicts a scree plot of the eigenvalues. The estimated explained variance of the l-th FPC is given by $\widehat{\lambda}_l / \sum_{k=1}^{\infty} \widehat{\lambda}_k$, while $\sum_{k=1}^{\infty} \widehat{\lambda}_k = \sum_{k=1}^{19} \widehat{\lambda}_k = 3.4814$. The first empirical FPC explains more than 96% of the variation of the curve.

Figure 3.5 presents orthonormal eigenfunctions for the empirical covariance operator's six largest eigenvalues, which are identified up to a sign change. In the foreground, the eigenfunctions belonging to the largest three eigenvectors are illustrated, while the next five eigenfunctions are plotted in the background. The shape of the first three loading

Figure 3.8: One-month ahead yield curve predictions and prediction bands



Note: Four exemplary one-step ahead forecasts, pointwise prediction bands and simultaneous prediction bands ($\alpha = 5\%$) are illustrated for the model with $K = 3$ factors and AR(1) factor dynamics. While the predictions for January 1994 and December 2014 are usual cases, the ones for July 2003 and November 2017 are exceptional, since the true curve crosses the pointwise prediction band.

functions are similar to the level, shape, and curvature function of the Nelson-Siegel model (see Figure 3.1). Figure 3.6 illustrates a plot of $\widehat{\mu}(r) \pm 0.5\widehat{\psi}_l(r)$ for $l = 1, \ldots, 6$, which illustrates the effect of each factor on the mean function across all times to maturity. The series of the first six empirical FPC scores are presented in Figure 3.7.

3.5.3 Yield curve prediction

To evaluate and compare the forecasting performances in an out-of-sample experiment, we follow the sequential setting in Diebold and Li (2006) and forecast the yield curve for each month from January 1994 until the end of the sample. Let P denote the set of prediction time points, so that the h-step ahead yield curve forecast for time $t \in P$ is estimated using all observations from the beginning of the sample until the time point $t - h$.

For the predictor from the functional dynamic factor model in equation (3.9), we consider the fixed numbers of $K = 3$, $K = 4$, and $K = 6$ factors and three different

Table 3.1: Root mean square forecast errors

	RW	Nelson-Siegel		Functional principal components estimator								
K		3	3	3	4	6	3	4	6	3	4	6
dynamics	–	AR1	VAR1		AR1			AR2			VAR1	
1-month-ahead forecast												
3m	0.215	0.216	0.194	0.206	0.202	0.202	0.205	0.199	0.199	0.187	0.185	0.190
6m	0.213	0.229	0.206	0.205	0.209	0.210	0.197	0.201	0.201	0.199	0.194	0.196
9m	0.228	0.250	0.236	0.231	0.227	0.227	0.221	0.217	0.217	0.233	0.216	0.218
12m	0.246	0.264	0.249	0.250	0.244	0.243	0.242	0.236	0.236	0.255	0.240	0.242
15m	0.258	0.277	0.261	0.267	0.263	0.259	0.259	0.255	0.252	0.271	0.259	0.258
18m	0.266	0.280	0.269	0.275	0.271	0.269	0.267	0.264	0.262	0.279	0.268	0.268
21m	0.273	0.286	0.279	0.285	0.282	0.280	0.278	0.275	0.273	0.288	0.278	0.280
24m	0.279	0.297	0.293	0.292	0.288	0.288	0.285	0.282	0.282	0.296	0.286	0.288
30m	0.288	0.297	0.298	0.294	0.294	0.294	0.289	0.288	0.288	0.298	0.292	0.295
36m	0.293	0.297	0.300	0.294	0.294	0.295	0.290	0.290	0.290	0.298	0.295	0.297
48m	0.299	0.299	0.304	0.295	0.298	0.299	0.293	0.295	0.296	0.300	0.298	0.301
60m	0.289	0.298	0.303	0.289	0.291	0.291	0.287	0.289	0.288	0.295	0.294	0.296
72m	0.284	0.290	0.294	0.282	0.285	0.285	0.281	0.283	0.283	0.288	0.288	0.290
84m	0.276	0.282	0.286	0.278	0.279	0.277	0.277	0.278	0.276	0.285	0.283	0.282
96m	0.270	0.272	0.275	0.274	0.272	0.272	0.274	0.273	0.272	0.280	0.278	0.279
108m	0.262	0.266	0.268	0.267	0.264	0.265	0.268	0.265	0.265	0.273	0.267	0.270
120m	0.263	0.271	0.273	0.270	0.262	0.263	0.271	0.263	0.264	0.274	0.264	0.268
TRMSFE	0.266	0.276	0.272	0.269	0.268	0.267	0.265	0.264	0.263	0.273	0.266	0.268
6-month-ahead forecast												
3m	0.814	0.820	0.723	0.725	0.734	0.734	0.744	0.752	0.753	0.724	0.667	0.669
6m	0.829	0.867	0.806	0.794	0.798	0.798	0.804	0.808	0.808	0.820	0.745	0.750
9m	0.837	0.879	0.853	0.829	0.826	0.826	0.834	0.832	0.832	0.871	0.785	0.792
12m	0.852	0.889	0.874	0.856	0.851	0.851	0.863	0.859	0.859	0.904	0.823	0.832
15m	0.864	0.895	0.887	0.874	0.870	0.870	0.884	0.881	0.880	0.924	0.851	0.861
18m	0.859	0.888	0.897	0.873	0.870	0.870	0.883	0.880	0.880	0.926	0.861	0.872
21m	0.860	0.884	0.906	0.873	0.871	0.870	0.883	0.881	0.880	0.928	0.869	0.881
24m	0.861	0.885	0.920	0.872	0.869	0.869	0.881	0.879	0.878	0.928	0.874	0.887
30m	0.849	0.858	0.902	0.847	0.846	0.846	0.858	0.858	0.858	0.906	0.863	0.877
36m	0.835	0.838	0.889	0.825	0.825	0.826	0.839	0.839	0.839	0.886	0.852	0.867
48m	0.799	0.801	0.859	0.790	0.792	0.793	0.805	0.808	0.808	0.852	0.831	0.847
60m	0.776	0.785	0.848	0.768	0.770	0.770	0.784	0.786	0.786	0.834	0.820	0.837
72m	0.739	0.749	0.809	0.734	0.736	0.736	0.752	0.754	0.754	0.799	0.790	0.809
84m	0.716	0.717	0.779	0.712	0.712	0.712	0.731	0.732	0.731	0.776	0.770	0.791
96m	0.694	0.688	0.749	0.694	0.694	0.694	0.718	0.717	0.717	0.759	0.754	0.776
108m	0.674	0.671	0.732	0.683	0.681	0.681	0.705	0.703	0.704	0.744	0.737	0.760
120m	0.659	0.672	0.732	0.675	0.671	0.671	0.698	0.695	0.695	0.732	0.723	0.749
TRMSFE	0.798	0.815	0.836	0.793	0.792	0.792	0.807	0.806	0.806	0.845	0.803	0.817
12-month-ahead forecast												
3m	1.392	1.348	1.355	1.272	1.279	1.279	1.294	1.301	1.301	1.369	1.187	1.192
6m	1.418	1.382	1.437	1.344	1.346	1.347	1.350	1.352	1.352	1.462	1.265	1.273
9m	1.402	1.372	1.463	1.359	1.358	1.358	1.357	1.355	1.355	1.490	1.282	1.291
12m	1.408	1.370	1.469	1.378	1.375	1.375	1.376	1.373	1.373	1.509	1.308	1.317
15m	1.400	1.357	1.461	1.379	1.377	1.377	1.383	1.381	1.380	1.509	1.322	1.332
18m	1.374	1.330	1.446	1.354	1.352	1.352	1.359	1.357	1.357	1.484	1.310	1.321
21m	1.348	1.305	1.432	1.328	1.326	1.326	1.334	1.333	1.332	1.458	1.297	1.307
24m	1.326	1.289	1.424	1.304	1.302	1.302	1.310	1.308	1.307	1.432	1.282	1.293
30m	1.280	1.237	1.373	1.248	1.248	1.248	1.260	1.260	1.260	1.375	1.247	1.259
36m	1.233	1.193	1.328	1.196	1.196	1.196	1.213	1.213	1.214	1.320	1.213	1.226
48m	1.146	1.119	1.249	1.115	1.117	1.116	1.142	1.143	1.143	1.235	1.163	1.179
60m	1.092	1.081	1.207	1.061	1.062	1.063	1.093	1.094	1.095	1.184	1.133	1.150
72m	1.026	1.030	1.148	1.011	1.013	1.013	1.050	1.052	1.052	1.132	1.097	1.117
84m	0.983	0.986	1.101	0.972	0.973	0.972	1.014	1.014	1.014	1.091	1.067	1.091
96m	0.946	0.947	1.060	0.946	0.945	0.945	0.994	0.993	0.993	1.064	1.047	1.073
108m	0.913	0.924	1.037	0.928	0.927	0.927	0.975	0.974	0.975	1.044	1.029	1.057
120m	0.895	0.929	1.040	0.918	0.916	0.916	0.965	0.963	0.964	1.030	1.015	1.046
TRMSFE	1.226	1.200	1.306	1.195	1.195	1.195	1.214	1.214	1.214	1.317	1.197	1.211

Note: Root mean square forecast errors (RMSFEs) and trace root mean square forecast errors (TRMSFEs) are reported. Furthermore, RW is the random walk forecast,

Table 3.2: Diebold-Mariano test statistics

	Functional principal components estimator								
K	3	4	6	3	4	6	3	4	6
dynamics		AR1			AR2			VAR1	
Test against the RW forecast									
1-month-ahead forecast									
3m	-0.96	-2.19*	-2.31*	-0.96	-2.11*	-2.25*	-2.80*	-3.22*	-2.60*
12m	0.77	-0.63	-0.87	-0.55	-1.50	-1.66*	1.25	-0.82	-0.47
60m	-0.02	0.72	0.65	-0.38	0.02	-0.09	1.31	0.82	1.12
6-month-ahead forecast									
3m	-4.83*	-4.48*	-4.46*	-2.64*	-2.35*	-2.32*	-3.29*	-3.92*	-3.77*
12m	0.26	-0.08	-0.08	0.45	0.27	0.27	2.27	-0.79	-0.54
60m	-0.66	-0.50	-0.50	0.38	0.45	0.45	2.37	1.41	1.92
12-month-ahead forecast									
3m	-4.21*	-4.04*	-4.05*	-2.23*	-2.09*	-2.10*	-0.62	-3.49*	-3.45*
12m	-1.21	-1.33	-1.33	-0.69	-0.75	-0.75	3.51	-1.81*	-1.69*
60m	-1.40	-1.34	-1.34	0.04	0.07	0.08	2.58	0.83	1.20
Test against the DNS forecast with AR(1) factor dynamics									
1-month-ahead forecast									
3m	-2.18*	-2.71*	-2.35*	-1.70*	-2.70*	-2.53*	-3.70*	-4.03*	-3.32*
12m	-2.98*	-4.06*	-4.12*	-3.13*	-3.84*	-3.91*	-0.97	-2.12*	-1.80*
60m	-1.98*	-1.31	-1.44	-2.19*	-1.63	-1.80*	-0.58	-1.14	-0.35
6-month-ahead forecast									
3m	-6.60*	-5.92*	-5.90*	-4.28*	-3.77*	-3.72*	-3.25*	-4.01*	-3.87*
12m	-2.25*	-2.60*	-2.61*	-2.28*	-2.72*	-2.73*	0.49	-1.59	-1.36
60m	-1.16	-1.03	-1.03	-0.10	0.05	0.06	2.37	1.29	1.92
12-month-ahead forecast									
3m	-3.14*	-2.82*	-2.83*	-2.13*	-1.86*	-1.87*	0.46	-2.64*	-2.56*
12m	0.27	0.17	0.17	0.37	0.20	0.21	3.08	-1.07	-0.93
60m	-0.81	-0.76	-0.75	0.84	0.91	0.94	3.56	1.25	1.70

Note: The benchmark statistics for the Diebold-Mariano tests are the random-walk (RW) forecast and the dynamic Nelson-Siegel (DNS) forecast with AR(1) factor dynamics. A negative value indicates the superiority of the functional principal components (FPCs) based forecast, while the asterisks denote significance at the 5% level.

specifications for the factor dynamics, which are given by

$$F_{l,t} = a_l F_{l,t-1} + \eta_{l,t}, \quad l = 1, \dots, K, \tag{AR1}$$

$$F_{l,t} = a_l F_{l,t-1} + b_l F_{t,t-2} + \eta_{l,t}, \quad l = 1, \dots, K, \tag{AR2}$$

$$F_t = A F_{t-1} + \eta_t. \tag{VAR1}$$

As reported in many studies, bond yields are close to being nonstationary, which makes it difficult to outperform the simple random walk (RW) no-change forecast (see Duffee 2002, Ang and Piazzesi 2003, Diebold and Li 2006, Mönch 2008, Caldeira and Torrent 2017). This is also reflected by the fact that the first eigenvalue in Figure 3.4 is quite large compared to the others and that augmented Dickey-Fuller tests are not able to reject the unit root hypothesis for the first and the second empirical FPC score series in Figure 3.7 at the usual significance levels. We thus consider the RW forecast as the main benchmark.

Another benchmark is the forecast obtained from the DNS model. The factors are estimated by regressing the available yields onto the Nelson-Siegel loadings given by equa-

Table 3.3: Coverage rates and average widths of one-month-ahead interval forecasts

	Functional principal components estimator with $K = 3$ components								
nominal coverage	85			90			95		
dynamics	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)	AR1	AR2	VAR1
3m	97.02	96.43	95.24	97.02	97.02	96.43	**97.62**	**97.62**	**97.02**
	(1.09)	(1.06)	(0.73)	(1.25)	(1.21)	(0.84)	(1.48)	(1.44)	(1.00)
6m	96.43	97.02	95.24	97.02	97.02	96.43	98.21	98.21	**97.62**
	(1.05)	(1.02)	(0.77)	(1.20)	(1.16)	(0.88)	(1.43)	(1.39)	(1.05)
9m	96.43	96.43	**89.88**	97.62	97.62	96.43	98.21	98.21	98.21
	(1.03)	(1.00)	(0.82)	(1.18)	(1.15)	(0.94)	(1.41)	(1.36)	(1.12)
12m	94.64	95.24	**89.88**	97.02	97.02	**92.86**	98.21	**97.62**	**97.02**
	(1.03)	(1.00)	(0.87)	(1.18)	(1.14)	(1.00)	(1.41)	(1.36)	(1.19)
15m	92.26	92.86	**89.29**	95.24	95.83	**91.67**	**97.62**	**97.62**	**97.02**
	(1.03)	(1.00)	(0.91)	(1.18)	(1.14)	(1.04)	(1.41)	(1.36)	(1.24)
18m	92.26	91.67	90.48	95.24	94.64	**91.67**	98.21	**97.62**	**95.83**
	(1.02)	(0.99)	(0.93)	(1.17)	(1.13)	(1.06)	(1.39)	(1.35)	(1.27)
21m	90.48	91.07	**89.88**	**92.86**	**92.26**	**91.67**	**97.62**	**97.02**	**96.43**
	(1.01)	(0.97)	(0.95)	(1.15)	(1.11)	(1.08)	(1.37)	(1.32)	(1.29)
24m	**89.88**	90.48	**89.88**	**92.86**	**91.07**	**92.26**	**97.02**	**95.24**	**96.43**
	(0.99)	(0.96)	(0.97)	(1.13)	(1.10)	(1.10)	(1.35)	(1.31)	(1.31)
30m	**89.29**	90.48	**89.29**	**92.26**	**92.26**	**91.67**	**96.43**	**94.64**	**96.43**
	(0.97)	(0.94)	(0.97)	(1.11)	(1.08)	(1.11)	(1.33)	(1.28)	(1.32)
36m	**87.50**	**89.29**	**87.50**	**92.26**	**91.07**	**92.26**	**94.64**	**93.45**	**97.02**
	(0.95)	(0.91)	(0.97)	(1.08)	(1.05)	(1.11)	(1.29)	(1.25)	(1.32)
48m	**86.31**	**87.50**	**86.90**	**91.67**	**89.88**	**91.67**	**93.45**	**92.86**	**94.64**
	(0.92)	(0.89)	(0.97)	(1.05)	(1.01)	(1.10)	(1.25)	(1.21)	(1.32)
60m	**86.90**	**86.31**	**88.69**	**91.07**	**90.48**	**92.26**	**94.05**	**93.45**	**94.64**
	(0.89)	(0.86)	(0.94)	(1.01)	(0.98)	(1.08)	(1.21)	(1.17)	(1.28)
72m	**86.90**	**87.50**	**88.10**	**92.26**	**90.48**	**92.26**	**95.24**	**94.05**	**95.83**
	(0.89)	(0.86)	(0.94)	(1.01)	(0.98)	(1.07)	(1.21)	(1.17)	(1.28)
84m	**87.50**	**86.90**	**88.69**	**91.07**	**89.88**	**92.26**	**95.83**	**93.45**	**96.43**
	(0.87)	(0.84)	(0.92)	(0.99)	(0.96)	(1.05)	(1.18)	(1.14)	(1.25)
96m	**87.50**	**87.50**	**89.29**	**92.86**	**90.48**	**92.86**	**95.24**	**94.05**	**95.83**
	(0.86)	(0.83)	(0.90)	(0.98)	(0.95)	(1.03)	(1.17)	(1.13)	(1.23)
108m	**88.69**	**88.10**	**89.29**	**92.26**	**91.67**	**93.45**	**96.43**	**96.43**	**97.62**
	(0.86)	(0.84)	(0.91)	(0.99)	(0.95)	(1.04)	(1.18)	(1.14)	(1.24)
120m	90.48	**87.50**	**89.29**	**92.86**	**92.26**	**93.45**	**97.62**	**96.43**	**97.62**
	(0.86)	(0.83)	(0.90)	(0.98)	(0.95)	(1.02)	(1.17)	(1.13)	(1.22)

Note: The out-of-sample performance of the pointwise $(1-\alpha)$-prediction bands from Theorem 3.3 is presented for the levels $\alpha = 15\%$, $\alpha = 10\%$, and $\alpha = 5\%$. The true coverage rates are presented in percentage points, while the average widths of the interval forecasts are indicated in the brackets . Bold numbers indicate that the coverage test by Christoffersen (1998) does not reject the hypothesis that the expected coverage coincides with the nominal coverage at the 5% level.

Table 3.4: Coverage rates of one-month-ahead simultaneous prediction bands

	Functional principal components estimator with $K = 3$ components								
nominal coverage	85			90			95		
dynamics	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)	AR(1)	AR(2)	VAR(1)
true coverage	100	100	100	100	100	100	100	100	100
average width	(2.43)	(2.35)	(2.44)	(2.55)	(2.46)	(2.56)	(2.73)	(2.64)	(2.74)

Note: The out-of-sample performance of the simultaneous $(1-\alpha)$-prediction bands from Theorem 3.3 is presented for the levels $\alpha = 15\%$, $\alpha = 10\%$, and $\alpha = 5\%$, and $L = 10$. The true coverage rates are presented in percentage points, while the average widths of the interval forecasts are indicated in the brackets.

tion (3.2) for a fixed value of $\lambda = 0.0609$. In a second step, a linear autoregressive model is fitted to the estimated factors from the first step, which, analogously to equation (3.7), gives rise to a forecast of the yield curve. Diebold and Li (2006) considered VAR(1) and univariate AR(1) factor dynamics and demonstrated that the model has a good forecasting performance for larger forecasting horizons.

The accuracies of the curve forecasts are evaluated by computing root mean square forecast errors

$$RMSFE(r_i) = \sqrt{\frac{1}{|P|}\sum_{t\in P}\left(\widehat{Y}_{t|t-h}(r_i) - Y_t(r_i)\right)^2}$$

and trace root mean square forecast errors

$$TRMSFE = \sqrt{\frac{1}{N}\frac{1}{|P|}\sum_{i=1}^{N}\sum_{t\in P}\left(\widehat{Y}_{t|t-h}(r_i) - Y_t(r_i)\right)^2},$$

where $r_1, \ldots, r_N$ are the available times to maturity in the data, and $|P|$ denotes the number of prediction time points. The results are presented in Table 3.1. The forecasts based on the FPC estimator with AR(1) factor dynamics have a lower TRMSFE than the DNS forecasts for all forecasting horizons, and a lower TRMSFE than the RW forecast for the six-month and 12-month ahead forecasts. The Diebold-Mariano test results in Table 3.2 indicate that the FPC-based forecasts outperform the RW and DNS forecasts significantly for bonds with a short time to maturity.

To evaluate the prediction bands, we consider the same out-of-sample setting as that for the curve forecasts. The coverage rates of the pointwise and simultaneous bands from Theorem 3.3 are listed in Tables 3.3 and 3.4, where we set $L = 10$ for the simultaneous bands (see Figure 3.4). Figure 3.8 presents four exemplary forecasts, pointwise prediction bands, and simultaneous prediction bands. The accuracies of the interval forecasts from the pointwise prediction bands are evaluated using the test by Christoffersen (1998), which compares the nominal coverage of the interval forecast with the true coverage. The results are provided in Table 3.3, and the null hypothesis that the expected coverage coincides with the nominal coverage is not rejected for the medium and longer maturity bonds. While the true coverage rates for the pointwise prediction bands do not deviate significantly from the nominal coverage for many times to maturity, the simultaneous bands are conservative and have a 100% coverage at a nominal coverage level of 85%, so that there is no yield curve in our sample that exceeds the simultaneous prediction bands.

3.6 Conclusion

We have introduced an identification strategy for a functional factor model by imposing orthogonality conditions for the loading functions that are elements of the Hilbert space of square integrable functions. The conditions are similar to those from the vector-valued factor models by Stock and Watson (2002) and Bai (2003).

Using results from functional data analysis, an FPC estimator is derived, and consistency results are presented. The minimum MSE h-step ahead forecast coincides with the predictor proposed in Aue et al. (2015) and thus provides a model-based justification of the prediction strategy in Aue et al. (2015). Furthermore, pointwise and simultaneous prediction bands are derived from the forecast error curve distribution. In an out-of-sample experiment with yield curve data, the forecasting procedure provides a higher accuracy when compared to the conventional DNS model.

Since the number of factors K and the lag order p are assumed to be known or selected by heuristic arguments, further research is required to answer the question of how these numbers can be identified.

Appendix to Chapter 3

We first show the following auxiliary result:

Lemma 3.6. *Let $\{Y_t\}_{t\in\mathbb{N}}$ be a functional time series that satisfies* (3.1) *and Assumptions 3.1, 3.2, 3.3, and 3.4. Then, the demeaned series $X_t(r) = Y_t(r) - \mu(r)$, $r \in [a, b]$, admits the Karhunen-Loève representation, $X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t}\psi_l(r)$, and, for its FPC scores $\theta_{l,t}$, there exists some constants $R < \infty$ and $\beta \in (1, \infty)$, such that*

$$\left|E\left[\theta_{l_1,t}\theta_{l_2,t-h}\right]\right| \leq R\, h^{-\beta}\sqrt{\lambda_{l_1}\lambda_{l_2}}, \tag{3.10}$$

$$\sum_{\tau_1,\tau_2,\tau_3=-\infty}^{\infty} |\kappa_{l_1,l_2,l_3,l_4}(\tau_1,\tau_2,\tau_3)| \leq R\sqrt{\lambda_{l_1}\lambda_{l_2}\lambda_{l_3}\lambda_{l_4}}, \tag{3.11}$$

for all $t, h, l_1, l_2, l_3, l_4, \in \mathbb{N}$, where $\kappa_{l_1,l_2,l_3,l_4}(\tau_1,\tau_2,\tau_3)$ denotes the 4-th order cumulant of $(\theta_{l_1,t}, \theta_{l_2,t+\tau_1}, \theta_{l_3,t+\tau_2}, \theta_{l_4,t+\tau_3})$, and $\lambda_l = E[\theta_{l,t}^2]$.

Proof. Since $E\|X_t\|^2 < \infty$, the demeaned series admits the Karhunen-Loève representation; that is,

$$X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t}\psi_l(r),$$

where $\{\psi_l\}_{l\in\mathbb{N}}$ represents the corresponding sequences of eigenfunctions, and $\theta_{l,t} = \langle X_t, \psi_l\rangle$ denotes the l-th FPC score of $X_t(r)$ (see Hörmann and Kokoszka 2012) . From (3.1), we have that

$$X_t(r) = \sum_{l=1}^{K} F_{l,t}\psi_l(r) + \sum_{l=1}^{\infty} \varepsilon_{l,t}\varphi_l(r),$$

where the second term on the right-hand side is the Karhunen-Loève representation of the error term $\epsilon_t(r)$. Following the identification results of Theorem 3.1, it follows that

$$\theta_{l,t} = \begin{cases} F_{l,t} & \text{for } 1 \leq l \leq K, \\ \varepsilon_{l-K,t} & \text{for } l > K. \end{cases} \tag{3.12}$$

Finally, since $F_{l,t}$ satisfies Assumption 3.4, the first K FPC scores can be written as a linear process based on the innovation sequence $\{\eta_t\}_{t\in\mathbb{N}}$. Moreover, $\varepsilon_{k,t}$ are uncorrelated and Gaussian for all $k, t \in \mathbb{N}$ by Assumption 3.2 and the properties of the Karhunen-Loève representation. Then, following the discussion in Salish and Gleim (2019) in Section 2.2, the scores $\theta_{l,t}$ satisfy restrictions (3.10) and (3.11). □

Proof of Lemma 3.1

Since the covariance kernel of $\epsilon_t(r)$ is bounded and $\langle \epsilon_t, \psi_l \rangle = 0$ from Assumption 3.3, it follows that $\int_a^b c_\epsilon(r,s)\psi_l(s)\,\mathrm{d}s = E[\epsilon_t(r)\langle \epsilon_t, \psi_l\rangle] = 0$ for all $l = 1, \ldots, K$ and $t = 1, \ldots, T$, and

$$\int_a^b c_Y(r,s)\psi_l(s)\,\mathrm{d}s = \int_a^b \Big(\sum_{k=1}^K \lambda_k \psi_k(r)\psi_k(s)\Big)\psi_l(s)\,\mathrm{d}s + \int_a^b c_\epsilon(r,s)\psi_l(s)\,\mathrm{d}s = \lambda_l \psi_l(r)$$

for all $r \in [a,b]$. As a consequence, λ_l is an eigenvalue of C_Y, and $\psi_l(r)$ is a corresponding eigenfunction. Let $\{\xi_j\}_{j\in\mathbb{N}}$ be the sequence of eigenvalues of C_ϵ, and let $\{v_j(r)\}_{j\in\mathbb{N}}$ be a sequence of corresponding orthonormal eigenfunctions. Then, $\{\xi_j, v_j(r)\}_{j\in\mathbb{N}}$ are also eigenpairs of C_Y. Furthermore, $\{v_j\}_{j\in\mathbb{N}}$ forms a basis of H, and, by Parseval's identity, $E\|\epsilon_t\|^2 = \sum_{j=1}^\infty E\langle \epsilon_t, \psi_j\rangle^2 = \sum_{j=1}^\infty \langle C_\epsilon(\psi_j), \psi_j\rangle = \sum_{j=1}^\infty \xi_j$, and $E\|\epsilon_t\|^2 < \lambda_K < \ldots < \lambda_1$ by Assumption 3.2(c). As a result, $\lambda_1, \ldots, \lambda_K$ are the K largest eigenvalues of C_Y. The second result follows from

$$\langle Y_t - \mu, \psi_l\rangle = \sum_{k=1}^K F_{k,t}\langle \psi_k, \psi_l\rangle + \langle \epsilon_t, \psi_l\rangle = F_{l,t},$$

since the eigenfunctions are orthonormal, and $\langle \epsilon_t, \psi_l\rangle = 0$.

Proof of Lemma 3.2

From Assumptions 3.1(a) and 3.1(b), it follows that $\sum_{t=1}^T \|\sum_{l=1}^K F_{lt}\psi_l\|^2 = \sum_{t=1}^T \sum_{l=1}^K F_{lt}^2$, which yields

$$\sum_{t=1}^T \Big\|Y_t - \widehat{\mu} - \sum_{l=1}^K F_{lt}\psi_l\Big\|^2 = \sum_{t=1}^T \Big(\|Y_t - \widehat{\mu}\|^2 - 2\sum_{l=1}^K F_{lt}\langle Y_t - \widehat{\mu}, \psi_l\rangle + \sum_{l=1}^K F_{lt}^2\Big).$$

The minimization problem (3.4) is then equivalent to minimizing $F_{lt}^2 - 2F_{lt}\langle Y_t - \widehat{\mu}, \psi_l\rangle$ for all $l = 1, \ldots, K$ and $t = 1, \ldots, T$, and the unique minimum is attained if $F_{lt} = \langle Y_t - \widehat{\mu}, \psi_l\rangle$.

Proof of Theorem 3.1

Let $X_t(r) = Y_t(r) - \widehat{\mu}(r)$. The goal is to find an orthonormal basis $\psi_1(r), \ldots, \psi_K(r)$, such that $\sum_{t=1}^T \|X_t - \sum_{l=1}^K \langle X_t, \psi_l\rangle\psi_l\|^2$ is minimized. We follow Hörmann and Kokoszka (2012) and Horváth and Kokoszka (2012), where this problem is treated in detail. From

Assumptions 3.1(a) and 3.1(b), it follows that

$$\sum_{t=1}^{T} \|Y_t - \mu - \sum_{l=1}^{K} \langle Y_t - \mu, \psi_l \rangle \psi_l\|^2 = \sum_{t=1}^{T} \|X_t\|^2 - \sum_{t=1}^{T} \sum_{l=1}^{K} \langle X_t, \psi_l \rangle^2.$$

Note that $\sum_{t=1}^{T} \sum_{l=1}^{K} \langle X_t, \psi_l \rangle^2 = T \sum_{l=1}^{K} \langle \widehat{C}_Y(\psi_l), \psi_l \rangle$. Then, the spectral decomposition yields $\langle \widehat{C}_Y(\psi_l), \psi_l \rangle = \sum_{j=0}^{\infty} \widehat{\lambda}_j \langle \psi_l, \widehat{\psi}_j \rangle^2$. Following Theorem 3.2 in Horváth and Kokoszka (2012), $\sum_{l=1}^{K} \sum_{j=0}^{\infty} \widehat{\lambda}_j \langle \psi_l, \widehat{\psi}_j \rangle^2$ is maximized if $\psi_l(r) = \widehat{\psi}_l(r)$ for all $l = 1, \ldots, K$. Finally, the assertion follows with Lemma 3.2.

Proof of Lemma 3.3

Since $E\|Y_t(r) - \mu(r)\|^2 < \infty$, the demeaned series $X_t(r) = Y_t(r) - \mu(r)$ admits the Karhunen-Loève representation,

$$X_t(r) = \sum_{l=1}^{\infty} \theta_{l,t} \psi_l(r),$$

where $\{\lambda_l\}_{l \in \mathbb{N}}$ is the sequence of eigenvalues of the covariance operator of $X_t(r)$ in decreasing order, and $\{\psi_l(r)\}_{l \in \mathbb{N}}$ is an orthonormal sequence of corresponding eigenfunctions. The FPC scores $\theta_{l,t} = \int_a^b X_t(r) \psi_l(r) \, \mathrm{d}r$ satisfy $E[\theta_{l,t} \theta_{m,t}] = \lambda_l \cdot 1_{\{l=m\}}$. Furthermore, since $\int_a^b (\psi_l(r))^2 \, \mathrm{d}r = 1$, Lebesgue's criterion for Riemann integrability implies that, for any $l \geq 1$, the eigenfunction $\psi_l(r)$ is bounded. That is, a constant $M < \infty$ exists, such that

$$\sup_{r \in [a,b]} |\psi_\ell(r)| \leq M \quad \text{for all } l \in \mathbb{N}. \tag{3.13}$$

Proof of (a): From equation (3.13) we have

$$\begin{aligned}
&\Big(\sup_{r \in [a,b]} \big| \widehat{\mu}(r) - \mu(r) \big| \Big)^2 = \sup_{r \in [a,b]} \big(\widehat{\mu}(r) - \mu(r) \big)^2 = \sup_{r \in [a,b]} \left(\frac{1}{T} \sum_{t=1}^{T} X_t(r) \right)^2 \\
&= \sup_{r \in [a,b]} \frac{1}{T^2} \sum_{t_1, t_2 = 1}^{T} \sum_{l_1, l_2 = 1}^{\infty} \theta_{l_1,t_1} \theta_{l_2,t_2} \psi_{l_1}(r) \psi_{l_2}(r) \leq \frac{M^2}{T^2} \sum_{t_1, t_2 = 1}^{T} \sum_{l_1, l_2 = 1}^{\infty} \theta_{l_1,t_1} \theta_{l_2,t_2} \\
&= \frac{M^2}{T^2} \Bigg(\sum_{t=1}^{T} \sum_{l_1, l_2 = 1}^{\infty} \theta_{l_1,t} \theta_{l_2,t} + 2 \sum_{h=1}^{T-1} \sum_{t=h+1}^{T} \sum_{l_1, l_2 = 1}^{\infty} \theta_{l_1,t} \theta_{l_2,t-h} \Bigg).
\end{aligned}$$

With the identification (3.12), it follows that $E[\sum_{l_1,l_2=1}^{\infty}\theta_{l_1,t}\theta_{l_2,t}]=\sum_{l=1}^{\infty}\lambda_l$, and

$$E\Big[\sum_{l_1,l_2=1}^{\infty}\theta_{l_1,t}\theta_{l_2,t-h}\Big]=\sum_{l_1,l_2=1}^{K}E[\theta_{l_1,t}\theta_{l_2,t-h}]\leq\frac{R}{h}\sum_{l_1,l_2=1}^{K}\sqrt{\lambda_{l_1}\lambda_{l_2}},$$

where the last inequality follows from Lemma 3.6. Then,

$$E\Big[\Big(\sup_{r\in[a,b]}\big|\widehat{\mu}(r)-\mu(r)\big|\Big)^2\Big]\leq\frac{M^2}{T}\sum_{l=1}^{\infty}\lambda_l+\frac{2M^2R}{T^2}\sum_{h=1}^{T-1}\frac{T-h}{h}\sum_{l_1,l_2=1}^{K}\sqrt{\lambda_{l_1}\lambda_{l_2}}=O(T^{-1})$$

Chebyshev's inequality thus concludes the first part of the lemma.
Proof of (b): From (a), it follows that $\sup_{r,s\in[a,b]}|\widehat{c}_Y(r,s)-\widehat{c}_X(r,s)|=O_P(T^{-1/2})$, where $\widehat{c}_X(r,s)=T^{-1}\sum_{l=1}^{T}X_t(r)X_t(s)$. It hence suffices to show consistency for $\widehat{c}_X(r,s)$. From equation (3.13), it follows that

$$\begin{aligned}
&\sup_{r,s\in[a,b]}\big|\widehat{c}_X(r,s)-c_X(r,s)\big|=\sup_{r,s\in[a,b]}\Big|\frac{1}{T}\sum_{t=1}^{T}\big(X_t(r)X_t(s)-E[X_t(r)X_t(s)]\big)\Big|\\
&=\sup_{r,s\in[a,b]}\Big|\frac{1}{T}\sum_{l_1,l_2=1}^{\infty}\sum_{t=1}^{T}\big(\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\big)\psi_{l_1}(r)\psi_{l_2}(s)\Big|\\
&\leq\frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\Big|\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\Big|.
\end{aligned}$$

Then, by Jensen's inequality,

$$\begin{aligned}
&E\Big[\Big|\sup_{r,s\in[a,b]}\big|\widehat{c}_X(r,s)-c_X(r,s)\big|\Big|\Big]\leq\frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}E\Big[\Big|\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\Big|\Big]\\
&\leq\frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\sqrt{E\Big[\Big(\sum_{t=1}^{T}\theta_{l_1,t}\theta_{l_2,t}-E[\theta_{l_1,t}\theta_{l_2,t}]\Big)^2\Big]}\\
&=\frac{M^2}{T}\sum_{l_1,l_2=1}^{\infty}\sqrt{\sum_{t_1,t_2=1}^{T}E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\theta_{l_1,t_2}\theta_{l_2,t_2}\big]-E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\big]E\big[\theta_{l_1,t_2}\theta_{l_2,t_2}\big]}.
\end{aligned}$$

The 4-th order cumulants have the following property:

$$\kappa_{(x_1,x_2,x_3,x_4)}=E[x_1x_2x_3x_4]-E[x_1x_2]E[x_3x_4]-E[x_1x_3]E[x_2x_4]-E[x_1x_4]E[x_2x_3],$$

which yields

$$\sum_{t_1,t_2=1}^{T} E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\theta_{l_1,t_2}\theta_{l_2,t_2}\big] - E\big[\theta_{l_1,t_1}\theta_{l_2,t_1}\big]E\big[\theta_{l_1,t_2}\theta_{l_2,t_2}\big]$$
$$= \sum_{t_1,t_2=1}^{T} \kappa_{l_1,l_2,l_1,l_2}(0,|t_1-t_2|,|t_1-t_2|) + E\big[\theta_{l_1,t_1}\theta_{l_1,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_2,t_2}\big] + E\big[\theta_{l_1,t_1}\theta_{l_2,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_1,t_2}\big].$$

From Lemma 3.6, it follows that

$$\sum_{t_1,t_2=1}^{T} \kappa_{l_1,l_2,l_1,l_2}(0,|t_1-t_2|,|t_1-t_2|) \leq R\lambda_{l_1}\lambda_{l_2},$$

and

$$\begin{aligned}
&\sum_{t_1,t_2=1}^{T} E\big[\theta_{l_1,t_1}\theta_{l_1,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_2,t_2}\big] + E\big[\theta_{l_1,t_1}\theta_{l_2,t_2}\big]E\big[\theta_{l_2,t_1}\theta_{l_1,t_2}\big] \\
&= \sum_{t=1}^{T} \big(E\big[\theta_{l_1,t}^2\big]E\big[\theta_{l_2,t}^2\big] + E\big[\theta_{l_1,t}\theta_{l_2,t}\big]^2\big) \\
&\quad + 2\sum_{h=1}^{T-1}\sum_{t=h+1}^{T} \big(E\big[\theta_{l_1,t}\theta_{l_1,t+h}\big]E\big[\theta_{l_2,t}\theta_{l_2,t+h}\big] + E\big[\theta_{l_1,t}\theta_{l_2,t+h}\big]E\big[\theta_{l_2,t}\theta_{l_1,t+h}\big]\big) \\
&\leq T\lambda_{l_1}\lambda_{l_2}\Big(2 + \frac{4R}{T}\sum_{h=1}^{T-1}\frac{T-h}{h^{\beta}}\Big).
\end{aligned}$$

Hence, a constant $C < \infty$ exists, such that

$$E\Big[\Big|\sup_{r,s\in[a,b]}\big|\widehat{c}_X(r,s) - c_X(r,s)\big|\Big|\Big] \leq \frac{C}{\sqrt{T}}\sum_{l_1,l_2=1}^{\infty}\sqrt{\lambda_{l_1}\lambda_{l_2}} = \frac{C}{\sqrt{T}}\sum_{l_1,l_2=1}^{L}\sqrt{\lambda_{l_1}\lambda_{l_2}} = O(T^{-1/2}),$$

since $Y_t(r)$ takes values in an L-dimensional subspace of H, and the assertion follows with Markov's inequality.

Proof of (c): The result follows from Lemma 3.6 and Corollary 2 in Salish and Gleim (2019).

Proof of (d): From Lemma 3.6 and Corollary 2 in Salish and Gleim (2019) ir follows that

$\|s_l\widehat{\psi}_l - \psi_l\| = O_P(T^{-1/2})$. Then, the Cauchy-Schwarz inequality yields

$$\begin{aligned}
&\sup_{r\in[a,b]} \Big|\widehat{\lambda}_l s_l \widehat{\psi}_l(r) - \lambda_l \psi_l(r)\Big| = \sup_{r\in[a,b]} \left| \int_a^b \widehat{c}_Y(r,s) s_l \widehat{\psi}_l(s)\,\mathrm{d}s - \int_a^b c_Y(r,s)\psi_l(s)\,\mathrm{d}s \right| \\
&= \sup_{r\in[a,b]} \left| \int_a^b \big(\widehat{c}_Y(r,s) - c_Y(r,s)\big) s_l \widehat{\psi}_l(s)\,\mathrm{d}s + \int_a^b c_Y(r,s)\big(s_l\widehat{\psi}_l(s) - \psi_l(s)\big)\,\mathrm{d}s \right| \\
&\le \|\psi_l\| \left(\sup_{r\in[a,b]} \sqrt{\int_a^b \big(\widehat{c}_Y(r,s) - c_Y(r,s)\big)^2\,\mathrm{d}s} \right) + \big\|s_l\widehat{\psi}_l - \psi_l\big\| \left(\sup_{r\in[a,b]} \sqrt{\int_a^b c_Y^2(r,s)\,\mathrm{d}s} \right) \\
&\le \sup_{r,s\in[a,b]} \big|\widehat{c}_Y(r,s) - c_Y(r,s)\big| + O_P(T^{-1/2}) = O_P(T^{-1/2}),
\end{aligned}$$

which follows from (b). Finally, with (c), and Slutsky's theorem, we obtain

$$\big|s_l\widehat{\psi}_l(r) - \psi_l(r)\big| = \frac{1}{\lambda_l}\Big|\widehat{\lambda}_l s_l \widehat{\psi}_l(r) - \lambda_l\psi_l(r)\Big| + O_P(T^{-1/2}) = O_P(T^{-1/2}).$$

Proof of (e): The triangle inequality and the Cauchy-Schwarz inequality yield

$$\begin{aligned}
&\max_{1\le t\le T} \big|s_l\langle Y_t - \widehat{\mu}, \widehat{\psi}_l\rangle - \langle Y_t - \mu, \psi_l\rangle\big| \\
&= \max_{1\le t\le T} \big|\langle Y_t - \widehat{\mu}, s_l\widehat{\psi}_l\rangle - \langle Y_t - \widehat{\mu}, \psi_l\rangle + \langle Y_t - \widehat{\mu}, \psi_l\rangle - \langle Y_t - \mu, \psi_l\rangle\big| \\
&\le \max_{1\le t\le T} \big(\big|\langle Y_t - \widehat{\mu}, s_l\widehat{\psi}_l - \psi_l\rangle\big| + \big|\langle \mu - \widehat{\mu}, \psi_l\rangle\big|\big) \\
&\le \|s_l\widehat{\psi}_l - \psi_l\|\big(\max_{1\le t\le T}\|Y_t - \widehat{\mu}\|\big) + \|\widehat{\mu} - \mu\| = O_P(T^{-1/2}),
\end{aligned}$$

which follows from (a) and (d).

Proof of Theorem 3.2

Note that

$$\begin{aligned}
&E\|Y_{T+h} - g(I_T)\|^2 = E\|Y_{T+h} - E[Y_{T+h}|I_T] + E[Y_{T+h}|I_T] - g(I_T)\|^2 \\
&= E\|Y_{T+h} - E[Y_{T+h}|I_T]\|^2 - 2E[\langle Y_{T+h} - E[Y_{T+h}|I_T], E[Y_{T+h}|I_T] - g(I_T)\rangle] \\
&\quad + E\|E[Y_{T+h}|I_T] - g(I_T)\|^2.
\end{aligned}$$

While $E\|Y_{T+h} - E[Y_{T+h}|I_T]\|^2$ does not depend on $g(I_T)$, the second term satisfies

$$\begin{aligned}
&E\big[\big\langle Y_{T+h} - E[Y_{T+h}|I_T],\ E[Y_{T+h}|I_T] - g(I_T)\big\rangle\big] \\
&= E\big[E\big[\big\langle Y_{T+h} - E[Y_{T+h}|I_T],\ E[Y_{T+h}|I_T] - g(I_T)\big\rangle\big|I_T\big]\big] \\
&= E\big[\big\langle E\big[Y_{T+h} - E[Y_{T+h}|I_T]\big|I_T\big],\ E[Y_{T+h}|I_T] - g(I_T)\big\rangle\big] \\
&= E\big[\big\langle 0,\ E[Y_{T+h}|I_T] - g(I_T)\big\rangle\big] = 0
\end{aligned}$$

because of the law of iterated expectation. Finally, $E\|E[Y_{T+h}|I_T] - g(I_T)\|^2$ takes the smallest value when $g(I_T) = E[Y_{T+h}|I_T]$.

Proof of Lemma 3.4

From Proposition 3.1 in Lütkepohl (2005), it follows that $\|\widetilde{B} - B\|_M = O_P(T^{-1/2})$, and Lemma 3.3(e) yields $\|S\widehat{F} - F\|_M = O_P(T^{-1/2})$ and $\|(S \otimes I_p)\widehat{Z} - Z\|_M = O_P(T^{-1/2})$, where $\otimes$ denotes the Kronecker product. The continuous mapping theorem implies that $\|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S \otimes I_p) - Z'(ZZ')^{-1}\|_M = O_P(T^{-1/2})$. Then,

$$\begin{aligned}
\|S\widehat{B}(S \otimes I_p) - \widetilde{B}\|_M &\leq \|S\widehat{F} - F\|_M \cdot \|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S \otimes I_p)\|_M \\
&+ \|F\|_M \cdot \|\widehat{Z}'(\widehat{Z}\widehat{Z}')^{-1}(S \otimes I_p) - Z'(ZZ')^{-1}\|_M = O_P(T^{-1/2}).
\end{aligned}$$

Therefore, $\|S\widehat{B}(S \otimes I_p) - B\|_M = O_P(T^{-1/2})$, which follows by the triangle inequality, and $\|S\widehat{A}_i S - A_i\|_M = O_P(T^{-1/2})$ for all $i = 1, \ldots, p$. The second result follows from Proposition 3.2 in Lütkepohl (2005) and the fact that $\widehat{\eta}_{lt}^2 = (s_l\widehat{\eta}_{lt})^2$ for all t and l.

Proof of Lemma 3.5

For the first result, let S be defined as in Lemma 3.4. Then,

$$\begin{aligned}
&\sup_{r\in[a,b]} \big|\widehat{\Psi}'(r)\widehat{F}_{T+h|T} - \Psi'(r)F_{T+h|T}\big| \\
&= \sup_{r\in[a,b]} \Big|\sum_{i=1}^{p} (S\widehat{\Psi}(r))'(S\widehat{A}_i S)(S\widehat{F}_{T+h-i|T}) - \Psi'(r)A_i F_{T+h-i|T}\Big| = O_P(T^{-1/2}),
\end{aligned}$$

which follows from Lemmas 3.3(d) and 3.3(e), and 3.4. Finally, by the triangle inequality,

$$\begin{aligned}
&\sup_{r\in[a,b]} \left|\widehat{Y}_{T+h|T}(r) - Y_{T+h|T}(r)\right| \\
&\leq \sup_{r\in[a,b]} \left|\widehat{\mu}(r) - \mu(r)\right| + \sup_{r\in[a,b]} \left|\widehat{\Psi}'(r)\widehat{F}_{T+h|T} - \Psi'(r)F_{T+h|T}\right| = O_P(T^{-1/2})
\end{aligned}$$

because of Lemma 3.3(a). For the second result, note that

$$c_{e,h}(r,s) = \Psi'(r)\Big(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i'\Big)\Psi(s) + \sum_{l=K+1}^{L} \lambda_l \psi_l(r)\psi_l(s), \quad r,s \in [a,b],$$

since $c_\epsilon(r,s) = \sum_{l=K+1}^{L} \lambda_l \psi_l(r)\psi_l(s)$ by Mercer's theorem. Then, analogously to the proof for the first result,

$$\begin{aligned}
&\sup_{r,s\in[a,b]} \left|\widehat{c}_{e,h}(r,s) - c_{e,h}(r,s)\right| \\
&\leq \sum_{i=0}^{h} \sup_{r,s\in[a,b]} \left|(S\widehat{\Psi}(r))' S\widehat{\Phi}_i S\widehat{\Sigma}_\eta (S\widehat{\Phi}_i S)' S\widehat{\Psi}(s) - \Psi'(r)\Phi_i\Sigma_\eta\Phi_i\Psi(s)\right| \\
&\quad + \sum_{l=K+1}^{L} \sup_{r,s\in[a,b]} \left|\widehat{\lambda}_l s_l \widehat{\psi}_l(r) s_l \widehat{\psi}_l(s) - \lambda_l \psi_l(r)\psi_l(s)\right| = O_P(T^{-1/2}),
\end{aligned}$$

by Lemmas 3.3 and 3.4.

Proof of Theorem 3.3

From Equation (3.8), it follows that for any fixed $r \in [a,b]$,

$$P\left(\frac{|Y_{T+h}(r) - Y_{T+h|T}(r)|}{\sqrt{\Psi'(r)(\sum_{i=0}^{h-1} \Phi_i\Sigma_\eta\Phi_i')\Psi(s) + \sum_{l=K+1}^{L} \lambda_l\psi_l(r)\psi_l(s)}} \leq u_{1-\frac{\alpha}{2}}\right) = 1-\alpha.$$

Then, (a) follows by Lemma 3.5 and Slutsky's theorem. For (b), let $\theta_{l,T} = \langle Y_t - \mu, \psi_l\rangle$ for $l > K$, and consider the $(L\times 1)$-vector $\delta = ((\sum_{i=0}^{h-1} \Phi_i\eta_{T+h-i})', \theta_{K+1,T+h}, \ldots, \theta_{L,T+h})'$. Furthermore, let $V(r) = (\psi_1(r), \ldots, \psi_L(r))'$, Then, $e_{T+h|T}(r) = V'(r)\delta$, where $\delta \sim \mathcal{N}(0, \Sigma_\delta)$

with

$$\Sigma_\delta = \begin{pmatrix} \sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i' & & 0 & \\ & \lambda_{K+1} & & \\ 0 & & \ddots & \\ & & & \lambda_L \end{pmatrix},$$

and $\delta' \Sigma_\delta^{-1} \delta \sim \chi^2(K)$. Therefore,

$$1 - \alpha = P\big(\delta' \Sigma_\delta^{-1} \delta \leq \chi^2_{L,1-\alpha}\big) = P\big((\Sigma_\delta^{-1/2}\delta)'(\Sigma_\delta^{-1/2}\delta) \leq \chi^2_{L,1-\alpha}\big).$$

A result from linear algebra states that for any fixed vector $x \in \mathbb{R}^L$ and any constant $c > 0$, $x'x \leq c^2$ if and only if $|a'x| \leq c\sqrt{a'a}$ for all $a \in \mathbb{R}^L$ (see Lemma A.4 in Yao et al. 2005). Then,

$$P\big((\Sigma_\delta^{-1/2}\delta)'(\Sigma_\delta^{-1/2}\delta) \leq \chi^2_{L,1-\alpha}\big) = P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathbb{R}^k\Big).$$

Let $\mathcal{E} = \{a \in \mathbb{R}^L :\ a = \Sigma_\delta^{1/2} V(r),\ r \in [a,b]\}$, which is a subset of $\mathbb{R}^L$. Therefore,

$$P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathbb{R}^k\Big) \leq P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathcal{E}\Big).$$

Furthermore,

$$P\Big(\big|a'(\Sigma_\delta^{-1/2}\delta)\big| \leq \sqrt{\chi^2_{L,1-\alpha} a'a},\ \forall a \in \mathcal{E}\Big) = P\left(\frac{\big|V'(r)\delta\big|}{\sqrt{V'(r)\Sigma_\delta V(r)}} \leq \sqrt{\chi^2_{L,1-\alpha}},\ \forall r \in [a,b]\right).$$

As a consequence,

$$\begin{aligned} &P\left(\frac{|e_{T+h|T}(r)|}{\sqrt{\sum_{i=0}^{h-1} \Psi'(r)\Phi_i\Sigma_\eta\Phi_i'\Psi(r) + \sum_{l=K+1}^{L} \lambda_l \psi_l(r)}} \leq \sqrt{\chi^2_{L,1-\alpha}} \quad \forall r \in [a,b]\right) \\ &= P\left(\frac{\big|V'(r)\delta\big|}{\sqrt{V'(r)\Sigma_\delta V(r)}} \leq \sqrt{\chi^2_{L,1-\alpha}},\ \forall r \in [a,b]\right) \geq 1 - \alpha. \end{aligned}$$

Finally, by Lemmas 3.3–3.5 and Slutsky's theorem,

$$\begin{aligned}
&\frac{|Y_{T+h}(r) - \widehat{Y}_{T+h|T}(r)|}{\sqrt{\widehat{\Psi}'(r)(\sum_{i=0}^{h-1} \widehat{\Phi}_i \widehat{\Sigma}_\eta \widehat{\Phi}_i')\widehat{\Psi}(r) + \sum_{l=K+1}^{L} \widehat{\lambda}_l \widehat{\psi}_l(r)}} \\
&= \frac{|e_{T+h|T}(r)|}{\sqrt{\Psi'(r)(\sum_{i=0}^{h-1} \Phi_i \Sigma_\eta \Phi_i')\Psi(r) + \sum_{l=K+1}^{L} \lambda_l \psi_l(r)}} + o_P(1),
\end{aligned}$$

and the assertion follows.

Bibliography

Anatolyev, S. and Kosenok, G. (2018). Sequential testing with uniformly distributed size. *Journal of Time Series Econometrics*, 10:1941–1928.

Anderson, T. W. (2003). *An introduction to multivariate statistical analysis, third edition.* John Wiley & Sons.

Andrews, D. W. (1993). Tests for parameter instability and structural change with unknown change point. *Econometrica*, 61:821–856.

Andrews, D. W. and Ploberger, W. (1994). Optimal tests when a nuisance parameter is present only under the alternative. *Econometrica*, 62:1383–1414.

Ang, A. and Piazzesi, M. (2003). A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables. *Journal of Monetary Economics*, 50:745–787.

Ang, A., Piazzesi, M., and Wei, M. (2006). What does the yield curve tell us about GDP growth? *Journal of Econometrics*, 131:359–403.

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

Aue, A., Norinho, D. D., and Hörmann, S. (2015). On the prediction of stationary functional time series. *Journal of the American Statistical Association*, 110:378–392.

Bai, J. (2003). Inferential theory for factor models of large dimensions. *Econometrica*, 71:135–171.

Bai, J. and Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176:18–29.

Banerjee, A., Lumsdaine, R. L., and Stock, J. H. (1992). Recursive and sequential tests of the unit-root and trend-break hypotheses: theory and international evidence. *Journal of Business & Economic Statistics*, 10:271–287.

Bardsley, P., Horváth, L., Kokoszka, P., and Young, G. (2017). Change point tests in functional factor models with application to yield curves. *The Econometrics Journal*, 20:86–117.

Bauer, P. and Hackl, P. (1978). The use of MOSUMS for quality control. *Technometrics*, 20:431–436.

Bierens, H. J. (1997). Testing the unit root with drift hypothesis against nonlinear trend stationarity, with an application to the us price level and interest rate. *Journal of Econometrics*, 81:29–64.

Billingsley, P. (1999). *Convergence of probability measures, second edition*. John Wiley & Sons.

Bliss, R. R. (1997). Movements in the term structure of interest rates. *Economic Review (Federal Reserve Bank of Atlanta)*, 82:16–34.

Bosq, D. (2000). *Linear processes in function spaces: theory and applications*. Springer.

Breitung, J. (2000). The local power of some unit root tests for panel data. In *Advances in Econometrics, Vol. 15: Nonstationary Panels, Panel Cointegration, and Dynamic Panels*, pages 161–177. Emerald Group Publishing.

Breitung, J. and Choi, I. (2013). Factor models. In *Handbook of Research Methods and Applications in Empirical Macroeconomics*, pages 249–265. Edward Elgar Publishing.

Breitung, J. and Das, S. (2005). Panel unit root tests under cross-sectional dependence. *Statistica Neerlandica*, 59:414–433.

Breitung, J. and Meyer, W. (1994). Testing for unit roots in panel data: are wages on different bargaining levels cointegrated? *Applied Economics*, 26:353–361.

Brown, R. L., Durbin, J., and Evans, J. M. (1975). Techniques for testing the constancy of regression relationships over time. *Journal of the Royal Statistical Society. Series B*, 37:149–192.

Caldeira, J. and Torrent, H. (2017). Forecasting the US term structure of interest rates using nonparametric functional data analysis. *Journal of Forecasting*, 36:56–73.

Cavaliere, G. (2005). Unit root tests under time-varying variances. *Econometric Reviews*, 23:259–292.

Cavaliere, G., Harvey, D. I., Leybourne, S. J., and Taylor, A. R. (2011). Testing for unit roots in the presence of a possible break in trend and nonstationary volatility. *Econometric Theory*, 27:957–991.

Cavaliere, G. and Taylor, A. R. (2007). Testing for unit roots in time series models with non-stationary volatility. *Journal of Econometrics*, 140:919–947.

Cavaliere, G. and Taylor, A. R. (2008). Time-transformed unit root tests for models with non-stationary volatility. *Journal of Time Series Analysis*, 29:300–330.

Christensen, J. H., Diebold, F. X., and Rudebusch, G. D. (2009). An arbitrage-free generalized Nelson-Siegel term structure model. *The Econometrics Journal*, 12:C33–C64.

Christiano, L. J. (1992). Searching for a break in GNP. *Journal of Business & Economic Statistics*, 10:237–250.

Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39:841–862.

Chu, C.-S. J., Hornik, K., and Kaun, C.-M. (1995). Mosum tests for parameter constancy. *Biometrika*, 82:603–617.

Chu, C.-S. J., Stinchcombe, M., and White, H. (1996). Monitoring structural change. *Econometrica*, 64:1045–65.

Davidson, J. (1994). *Stochastic limit theory: an introduction for econometricians*. Oxford University Press.

de Boor, C. (2001). *A practical guide to splines, revised edition*. Springer.

Dette, H. and Gösmann, J. (2019). A likelihood ratio approach to sequential change point detection for a general class of parameters. *Journal of the American Statistical Association*, forthcoming.

Dickey, D. A. and Fuller, W. A. (1979). Distribution of the estimators for autoregressive time series with a unit root. *Journal of the American Statistical Association*, 74:427–431.

Diebold, F. X. and Li, C. (2006). Forecasting the term structure of government bond yields. *Journal of Econometrics*, 130:337–364.

Diebold, F. X., Piazzesi, M., and Rudebusch, G. D. (2005). Modeling bond yields in finance and macroeconomics. *American Economic Review*, 95:415–420.

Diebold, F. X. and Rudebusch, G. D. (2013). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *The Journal of Finance*, 57:405–443.

Elliott, G., Rothenberg, T. J., and Stock, J. H. (1996). Efficient tests for an autoregressive unit root. *Econometrica*, 64:813–836.

Enders, W. and Lee, J. (2012). A unit root test using a fourier series to approximate smooth breaks. *Oxford Bulletin of Economics and Statistics*, 74:574–599.

Evans, M. D. and Lewis, K. K. (1995). Do expected shifts in inflation affect estimates of the long-run fisher relation? *The Journal of Finance*, 50:225–253.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82:540–554.

Fremdt, S. (2015). Page's sequential procedure for change-point detection in time series regression. *Statistics*, 49(1):128–155.

Gösmann, J., Kley, T., and Dette, H. (2019). A new approach for open-end sequential change point monitoring. *https://arxiv.org/abs/1906.03225*.

Gürkaynak, R. S., Sack, B., and Wright, J. H. (2007). The US treasury yield curve: 1961 to the present. *Journal of Monetary Economics*, 54:2291–2304.

Hamori, S. and Tokihisa, A. (1997). Testing for a unit root in the presence of a variance shift. *Economics Letters*, 57:245–253.

Hansen, B. E. (1992). Testing for parameter instability in linear models. *Journal of Policy Modeling*, 14:517–533.

Hassler, U. and Wolters, J. (1995). Long memory in inflation rates: international evidence. *Journal of Business & Economic Statistics*, 13:37–45.

Hays, S., Shen, H., Huang, J. Z., et al. (2012). Functional dynamic factor models with application to yield curve forecasting. *The Annals of Applied Statistics*, 6:870–894.

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 *Handbook of statistics*, volume 30, pages 157–186. Elsevier.

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

Horváth, L. and Kokoszka, P. (2012). *Inference for functional data with applications*. Springer.

Hsing, T. and Eubank, R. (2015). *Theoretical foundations of functional data analysis, with an introduction to linear operators*. John Wiley & Sons.

Hyndman, R. J. and Ullah, M. S. (2007). Robust forecasting of mortality and fertility rates: a functional data approach. *Computational Statistics & Data Analysis*, 51:4942–4956.

Jones, P. M. and Enders, W. (2014). On the use of the flexible fourier form in unit root tests, endogenous breaks, and parameter instability. In *Recent Advances in Estimating Nonlinear Models*, pages 59–83. Springer.

Joslin, S., Priebsch, M., and Singleton, K. J. (2014). Risk premiums in dynamic term structure models with unspanned macro risks. *The Journal of Finance*, 69:1197–1233.

Jungbacker, B., Koopman, S. J., and Van der Wel, M. (2014). Smooth dynamic factor analysis with application to the US term structure of interest rates. *Journal of Applied Econometrics*, 29:65–90.

Kapetanios, G., Shin, Y., and Snell, A. (2003). Testing for a unit root in the nonlinear STAR framework. *Journal of Econometrics*, 112:359–379.

Kiefer, N. M. and Vogelsang, T. J. (2005). A new asymptotic theory for heteroskedasticity-autocorrelation robust tests. *Econometric Theory*, 21:1130–1164.

Kılıç, R. (2011). Testing for a unit root in a stationary ESTAR process. *Econometric Reviews*, 30:274–302.

Kim, T.-H., Leybourne, S., and Newbold, P. (2002). Unit root tests with a break in innovation variance. *Journal of Econometrics*, 109:365–387.

Kokoszka, P. and Reimherr, M. (2013). Determining the order of the functional autoregressive model. *Journal of Time Series Analysis*, 34:116–129.

Krämer, W., Ploberger, W., and Alt, R. (1988). Testing for structural change in dynamic models. *Econometrica*, 56:1355–1369.

Kuan, C.-M. and Hornik, K. (1995). The generalized fluctuation test: A unifying view. *Econometric Reviews*, 14:135–161.

Lee, H.-Y. and Wu, J.-L. (2001). Mean reversion of inflation rates: evidence from 13 OECD countries. *Journal of Macroeconomics*, 23:477–487.

Leisch, F., Hornik, K., and Kuan, C.-M. (2000). Monitoring structural changes with the generalized fluctuation test. *Econometric Theory*, 16:835–854.

Lengwiler, Y. and Lenz, C. (2010). Intelligible factors for the yield curve. *Journal of Econometrics*, 157:481–491.

Levin, A., Lin, C.-F., and Chu, C.-S. J. (2002). Unit root tests in panel data: asymptotic and finite-sample properties. *Journal of Econometrics*, 108:1–24.

Leybourne, S., Newbold, P., and Vougas, D. (1998). Unit roots and smooth transitions. *Journal of Time Series Analysis*, 19:83–97.

Litterman, R. and Scheinkman, J. (1991). Common factors affecting bond returns. *The Journal of Fixed Income*, 1:54–61.

Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer.

MacDonald, R. and Murphy, P. D. (1989). Testing for the long run relationship between nominal interest rates and inflation using cointegration techniques. *Applied Economics*, 21:439–447.

Mönch, E. (2008). Forecasting the yield curve in a data-rich environment: A no-arbitrage factor-augmented var approach. *Journal of Econometrics*, 146:26–43.

Mönch, E. (2012). Term structure surprises: the predictive content of curvature, level, and slope. *Journal of Applied Econometrics*, 27:574–602.

Müller, U. K. and Elliott, G. (2003). Tests for unit roots and the initial condition. *Econometrica*, 71:1269–1286.

Narayan, P. K. and Narayan, S. (2010). Is there a unit root in the inflation rate? new evidence from panel data models with multiple structural breaks. *Applied Economics*, 42:1661–1670.

Nelson, C. R. and Siegel, A. F. (1987). Parsimonious modeling of yield curves. *The Journal of Business*, 60:473–489.

Ng, S. and Perron, P. (1995). Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag. *Journal of the American Statistical Association*, 90:268–281.

Ng, S. and Perron, P. (2001). Lag length selection and the construction of unit root tests with good size and power. *Econometrica*, 69:1519–1554.

Nyblom, J. (1989). Testing for the constancy of parameters over time. *Journal of the American Statistical Association*, 84:223–230.

Otto, S. (2019). Unit root testing with slowly varying trends. Working paper, `https://wisostat.uni-koeln.de/sites/statistik/pdf/otto2019.pdf`.

Otto, S. and Breitung, J. (2019). Backward CUSUM for testing and monitoring structural change. Working paper, `https://wisostat.uni-koeln.de/sites/statistik/pdf/ottobreitung2019.pdf`.

Otto, S. and Salish, N. (2019). A dynamic functional factor model for yield curves: identification, estimation, and prediction. Working paper, `https://wisostat.uni-koeln.de/sites/statistik/pdf/ottosalish2019.pdf`.

Park, B. U., Mammen, E., Härdle, W., and Borak, S. (2009). Time series modelling with semiparametric factor dynamics. *Journal of the American Statistical Association*, 104:284–298.

Perron, P. (1989). The great crash, the oil price shock, and the unit root hypothesis. *Econometrica*, 57:1361–1401.

Phillips, P. C. (1987). Time series regression with a unit root. *Econometrica*, 55:277–301.

Phillips, P. C. and Durlauf, S. N. (1986). Multiple time series regression with integrated processes. *The Review of Economic Studies*, 53:473–495.

Phillips, P. C. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75:335–346.

Ploberger, W. and Krämer, W. (1990). The local power of the cusum and cusum of squares tests. *Econometric Theory*, 6:335–347.

Ploberger, W. and Krämer, W. (1992). The cusum test with ols residuals. *Econometrica*, 60:271–285.

Ploberger, W., Krämer, W., and Kontrus, K. (1989). A new test for structural stability in the linear regression model. *Journal of Econometrics*, 40:307–318.

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

Ramsay, J. and Silverman, B. (2005). *Functional data analysis*. Springer.

Rappoport, P. and Reichlin, L. (1989). Segmented trends and non-stationary time series. *The Economic Journal*, 99:168–177.

Robbins, H. and Siegmund, D. (1970). Boundary crossing probabilities for the wiener process and sample sums. *The Annals of Mathematical Statistics*, 41:1410–1429.

Rose, A. K. (1988). Is the real interest rate stable? *The Journal of Finance*, 43:1095–1112.

Said, S. E. and Dickey, D. A. (1984). Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71:599–607.

Salish, N. and Gleim, A. (2019). A moment-based notion of time dependence for functional time series. *Journal of Econometrics*, `https://doi.org/10.1016/j.jeconom.2019.03.007`.

Schmidt, P. and Phillips, P. C. (1992). Lm tests for a unit root in the presence of deterministic trends. *Oxford Bulletin of Economics and Statistics*, 54:257–287.

Schwert, G. W. (1989). Tests for unit roots: a monte carlo investigation. *Journal of Business & Economic Statistics*, 7:147–159.

Sen, P. K. (1982). Invariance principles for recursive residuals. *The Annals of Statistics*, 10:307–312.

Stock, J. H. and Watson, M. W. (2002). Forecasting using principal components from a large number of predictors. *Journal of the American Statistical Association*, 97:1167–1179.

Stock, J. H. and Watson, M. W. (2012). Dynamic factor models. In *The Oxford Handbook of Economic Forecasting*, pages 35–39. Oxford University Press.

Svensson, L. E. (1995). Estimating forward interest rates with the extended nelson & siegel method. *Quarterly Review (Sveriges Riksbank)*, 3:13–26.

Wied, D. and Galeano, P. (2013). Monitoring correlation change in a sequence of random variables. *Journal of Statistical Planning and Inference*, 143:186–196.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, 100:577–590.

Zeileis, A. (2004). Alternative boundaries for cusum tests. *Statistical Papers*, 45:123–131.

Zeileis, A., Leisch, F., Kleiber, C., and Hornik, K. (2005). Monitoring structural change in dynamic econometric models. *Journal of Applied Econometrics*, 20:99–121.

Zivot, E. and Andrews, D. W. (1992). Further evidence on the great crash, the oil-price shock, and the unit-root hypothesis. *Journal of Business & Economic Statistics*, 10:251–270.