Modeling and Forecasting the Co-Movement of International Yield Curve Drivers

Maria Sprincenatu

Dissertation
an der Fakultät für Mathematik, Informatik, und Statistik
der Ludwig–Maximilians–Universität
München

eingereicht von
Maria Sprincenatu

München 2019

Modeling and Forecasting the Co-Movement of International Yield Curve Drivers

Maria Sprincenatu

Dissertation
an der Fakultät für Mathematik, Informatik, und Statistik
der Ludwig–Maximilians–Universität
München

eingereicht von
Maria Sprincenatu

München 2019

1. Gutachter: Prof. Stefan Mittnik, PhD
2. Gutachter: Prof. Dr. Markus Haas, Christian-Albrecht-Universität zu Kiel
Tag der Einreichung: 25.06.2019
Tag der mündlichen Prüfung: 21.11.2019

"*The agony of breaking through personal limitations is the agony of spiritual growth. Art, literature, myth and cult, philosophy, and ascetic disciplines are instruments to help the individual past his limiting horizons into spheres of ever-expanding realization. As he crosses threshold after threshold, conquering dragon after dragon, the stature of the divinity that he summons to his highest wish increases, until it subsumes the cosmos. Finally, the mind breaks the bounding sphere of the cosmos to a realization transcending all experiences of form - all symbolizations, all divinities: a realization of the ineluctable void.*"

– Joseph Campbell, The Hero With a Thousand Faces

Acknowledgements

First and foremost, I would like to express my sincere gratitude to an outstanding person, researcher, and supervisor: Prof. Stefan Mittnik, PhD. Thank you for giving me the great opportunity to write both my Master and PhD theses under your supervision. Thank you for your extraordinary positivity, great spirit, and kindness. Most of all, thank you for helping me to find and prove the "null hypothesis" of my doctoral research, for sharing with me your unparalleled knowledge, ingenious ideas, and invaluable insights.

I would also like to extend my sincere thanks to Prof. Dr. Markus Haas, for joining my PhD committee as external reviewer, and Prof. Dr. Christian Heumann and Prof. Dr. Volker Schmid, for serving as internal members in the committee.

I also wish to express my deep appreciation to the entire Chair of Financial Econometrics. The Chair has been a source of genuine friendships as well as good advice and exchange of knowledge. Special thanks to my dear colleagues and friends, Christoph Berninger (aka Big B / Burnie) and Henry Port, for their support throughout my toughest PhD moments. Thank you, Burnie, for sharing your office with me, for your patience, and your illuminating thoughts that made you gain the title of "Big B". Thank you, Henry, for selling me your beautiful violin, the instrument that opened to me a great new world. I am also truly grateful to Martina Brunner for her unbounded kindness and great amount of assistance.

I also wish to thank my two great friends, Christina and Marco, for the delicious dinners and numerous moments of laughter.

Last and most importantly, a special thanks to my family. Words cannot express how grateful I am to my mother for all the sacrifices and hard work she has done for me, in order to allow me to pursue higher education. Thank you, Mother, for being my biggest supporter. Thank you so much for believing in me. You are the best mother in the world, and I owe my success to you and Babunea.

Zusammenfassung

Das Diebold-Li (2006) "Yields-Only" -Modell, erweitert von Diebold-Li-Yue (2008) auf den globalen Kontext, hat nach der Finanzkrise von 2008 an Popularität gewonnen, als die Aufsichtsbehörden der Marktbewertung und der Bilanzierung von Verbindlichkeiten einen höheren Stellenwert einräumten. Dank seiner Sparsamkeit, der genauen Parameterschätzung und der starken Vorhersagbarkeit auf lange Sicht gilt das Dieboli-Li-Modell als Stand der Technik für die Modellierung und Prognose von Zinskurven. Trotz seiner zahlreichen Vorteile ignoriert das Modell die dynamischen Eigenschaften der Zinskurvenfaktoren, die für die Prognose entscheidend sind.

Diese Dissertation zielt auf die Entwicklung neuer datengesteuerter Zustandsraummodelle, um die gleichgesichtete Entwicklung von Zinskurventreiber verschiedener Weltregionen vorherzusagen. Die Modelle sind so konzipiert, dass die dynamischen Eigenschaften der in den zugrunde liegenden Datenerzeugungsprozessen enthaltenen Zinskurventreiber erhalten bleiben. Im Sinne von Diebold-Li ermöglichen die Modelle die Prognose der Mitbewegung von Zinskurven verschiedener Weltregionen durch Vorhersage ihrer Treiber.

Unter der Verwendung von gehandelten Staatsanleihen für USA und Deutschland besteht der Modellierungsansatz darin, zunächst eine umfassende Untersuchung der dynamischen Eigenschaften von amerikanischen und deutschen Zinskurventreiber durchzuführen. Diese Studie liefert Hinweise auf die Stationarität der amerikanischen und deutschen Steigungen, die Nichtstationarität der Level und Krümmungen, einer Kointegrationsstruktur zwischen den Levels und den Krümmungen sowie das Vorhandensein einer Granger-Kausalität bei allen amerikanischen und deutschen Zinskurventreibern. Eine univariate und multivariate Zustandsraumanalyse zu Ausreißern und Strukturbrüchen zeigt Veränderungen in der Struktur der amerikanischen und deutschen Zinskurventreiber im Zeitraum der Finanzkrise von 2008. Diese vorübergehenden Änderungen scheinen über die Treiber hinweg synchron zu sein und ähneln Patches von Ausreißern und nicht strukturellen Brüchen. Eine Studie über die Vorhersagbarkeit der Geldpolitik der US-Fed und der EZB ermöglicht es, die Art der Ausreißer an einen Regimewechsel in der Geldpolitik der US-Fed und der EZB und die Fähigkeit der Marktteilnehmer, die geldpolitische Haltung nach der Finanzkrise vorherzusagen, zu verknüpfen. Die extremsten Ausreißer können in unseren Zustandsraummodellen unter Einbeziehung von Interventionsvariablen in die Ausgangsgleichung problemlos verarbeitet werden.

In einer rekursiven Out-of-Sample-Prognose mit dem Kalman-Filter und einer alle 12 Monate erfolgten Schätzung der Parameter untersuchen wir die Performance unserer neu entwickelten Zustandsraummodelle bei der gemeinsamen Prognose der amerikanischen und deutschen Zinskurven. Die Prognoseergebnisse sind vielversprechend und belegen, dass unser Full State Space Model (FSSM), das alle dynamischen Eigenschaften der Ertragsdaten berücksichtigt, das hochmoderne Diebold-Li-Modell übertrifft. Darüber hinaus überprüfen wir die Vorhersagegüte der Krümmungen, indem wir zwei zusätzliche Modelle nur für die amerikanischen und deutschen Level und Steigungen entwickeln und prognostizieren, nämlich das FSSM^{LS} und das MShock-FSSM^{LS}. Die schlechten Prognoseergebnisse für alle Zeiträume zeigen, dass die Krümmungen für unsere Zinsdaten eine Vorhersagegüte für die Zinskurven der USA und Deutschlands haben.

Summary

The Diebold-Li (2006) "Yields-Only" model, extended to the global context by Diebold-Li-Yue (2008), has gained significant popularity in the aftermath of the 2008 Financial Crisis, when regulators placed greater emphasis on the market valuation and accounting of liabilities. Thanks to its parsimony, accurate parameter estimation, and strong forecastability at long horizons, the Diebold-Li model is widely acknowledged as state-of-the-art for yield curve modeling and forecasting. Despite its numerous advantages, the model disregards the in-sample dynamic properties of the yield curve factors, which are crucial for forecasting the yield curve.

This thesis aims at developing new data-driven state space models to forecast the co-movement of yield curve drivers of different world regions. The models are designed to preserve the dynamic properties of the yield curve drivers embodied in their underlying data generation processes. In the spirit of Diebold-Li, the models allow forecasting the co-movement of yield curves of different world regions by forecasting their drivers.

Using actively traded government bond yields for US and Germany, the modeling approach consists in first conducting a comprehensive study of the dynamic properties of US and German yield curve drivers. This study provides evidence about the stationarity of the US and German slopes, nonstationarity of the levels and curvatures, cointegration structure between the levels and curvatures, and existence of Granger causality among all US and German yield curve drivers. A univariate and multivariate state-space study of outliers and structural breaks reveals alterations in the structure of US and German yield curve drivers in proximity of the 2008 Financial Crisis. These transient changes appear to be synchronized across the drivers and resemble of patches of outliers rather than of structural breaks. A study of the US Fed and the ECB monetary policy predictability allows linking the nature of the outliers to a regime change in the US Fed and ECB monetary policy and an increased ability of market participants in predicting the monetary policy stance after the Financial Crisis. The most blatant outliers can easily be handled in our state-space models with the inclusion of intervention variables in the measurement equation.

In a recursive out-of-sample forecasting exercise with the Kalman filter and re-estimation of the parameters every 12 months, we explore the performance of our newly developed state-space models in forecasting jointly the US and German yield curves. The forecasting results are promising, providing evidence that our Full State Space Model (FSSM), accounting for all the dynamic properties of the yield data, outperforms the state-of-the-art Diebold-Li model. In addition, we verify the forecasting power of the curvatures, by developing and forecasting with two additional models for the US and German levels and slopes only, i.e., the FSSM^{LS} and the MShock-FSSM^{LS}. The poor forecasting results at all horizons provide evidence that, for our sample of yields, the curvatures do have predictive power for the US and German yield curves.

Contents

List of Figures

List of Tables

List of Abbreviations

ABS PP	Asset-Backed Securities Purchase Programme
ADF	Augmented Dickey-Fuller (test)
AIC	Akaike Information Criterion
ARCH	Autoregressive Conditional Heteroskedastic
ARIMA	Autoregressive Integrated Moving Average
BIC	Bayesian Information Criterion
BofA	Bank of America
BVerfG	Bundesverfassungsgericht der Bundesrepublik Deutschland (German Federal Constitutional Court)
CBPP1	First Covered Bond Purchase Programme
CBPP2	Second Covered Bond Purchase Programme
CBPP3	Third Covered Bond Purchase Programme
CCF	Cross-Correlation Function
CLF	Concentrated Likelihood Function
CFPB	Consumer Financial Protection Bureau
CJEU	Court of Justice of the European Union
CRR/CRD IV	Capital Requirements Regulation and Directive
CSPP	Corporate Sector Purchase Programme
CT1	Core Tier 1 (Capital Ratio)
DF	(Rate on the) Deposit Facility
DGP	Data Generation Process
Dodd-Frank	Dodd-Frank Wall Street Reform and Consumer Protection Act
DJIA	Dow Jones Industrial Average
EBA	European Banking Authority
ECB	European Central Bank
ECM	Error Correction Model
EMU	European Monetary Union
FEIR	Forecast Error Impulse Response

FOMC	(United States) Federal Open Market Committee
FRBNY	(United States) Federal Reserve Bank of New York
FSSM	Full-State Space Model
FTOs	Fine-Tuning Operations
GLS	Generalized Least Squares
GC	Governing Council (of the ECB)
HICP	Harmonised Index of Consumer Prices
HMM	Hidden Markov Model
IORR rate	Interest Rate on Required Reserves
IOER rate	Interest Rate on Excess Reserves
IR	Impulse Response (Function)
IYCD	International Yield Curve Drivers
KPSS	Kwiatkowski–Phillips–Schmidt–Shin (test)
KS	(One-sample) Kolmogorov-Smirnov (test)
LF	Likelihood Function
LTCM	Long-Term Capital Management
LTROs	Long-Term Refinancing Operations
MA	Moving Average (Representation)
MBS	Mortgage-Backed Securities
MEP	Maturity Extension Program
MCMC	Markov Chain Monte Carlo (methods)
MLE	Maximum Likelihood Estimation
MLF	(Rate on the) Marginal Lending Facility
MRO	(Interest rate on the) Main Refinancing Operations
MSE	Mean Squared Error
NCBs	National Central Banks
OMOs	Open Market Operations
OMTs	Outright Monetary Transactions
ON RRP	Overnight Reverse Repurchase (Facility)
ON RRR rate	Overnight Reverse Repurchase Agreement Rate
OLS	Ordinary Least Squares
OPG	Outer Product of Gradients
PCA	Principal Component Analysis
PDF	Probability Density Function

PSPP	Public Sector Purchase Programme
QE	Quantitative Easing
QML	Quasi-Maximum Likelihood
RMSEs	Root Mean Square Errors
SFE	Squared Forecast Error
SOMA	System Open Market Account
SSECM	State-Space Error Correction Model
SPE	Standardize (One-Step) Prediction Errors (in SSM models)
SSM	Single Supervisory Mechanism
SSM	State-Space Model
SSOD	Standardized Smoothed Observation Disturbances (in SSM models)
SSSD	Standardized Smoothed State Disturbances (in SSM models)
SSVAR	State-Space Vector Autoregressive (Model)
SSVEC	State-Space Vector Error Correction (Model)
The Desk	(United States Federal Reserve Bank of New York) Open Market Desk
The Fed	(United States) Federal Reserve System
TARP	Troubled Asset Relief Program
TLTRO II	Targeted Longer-Term Refinancing Operations
TY	Toda-Yamamoto (approach to Granger noncausality)
VAR	Vector Autoregressive Model
VEC	Vector Error Correction Model

List of Symbols

$y_{it}(\tau)$	where $i = \{US, DE\}$ and t=[1999:01-2018:01]. Yields, in Levels	(%)
τ	Yield curve maturities, $\tau = \{6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y\}$	
$\Delta y_{it}(\tau)$	Yields, in First Differences	(%)
$s_{DE-US,t}(\tau)$	German-US yield curve spreads	(%)
l_{it}	(Nelson-Siegel) Country-Specific Level Factor	(%)
s_{it}	(Nelson-Siegel) Country-Specific Slope Factor	(%)
c_{it}	(Nelson-Siegel) Country-Specific Curvature Factor	(%)
$\boldsymbol{\alpha}$	Vector of adjustment speeds in VEC models	
$\boldsymbol{\beta}$	"Long-run" matrix in VEC models	
$\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$	Impact matrix VEC models	
$\boldsymbol{\beta}' x_{t-1}$	Common trends in VEC models	
$\boldsymbol{\alpha}\boldsymbol{\beta}' x_{t-1}$	Error correction term in VEC models	
ec_t^{ll}	Johansen estimated cointegrating relation in the $l_{US,t}/l_{DE,t}$ system	
ec_{it}^{llcc}	where $i = 1, 2, 3$. Johansen estimated cointegrating relations in the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system	
D_1	"All-Yields" dataset for PCA on all US and German yields	(%)
D_2	"DE-US Spread" dataset for PCA on German-US yield curve spreads	(%)
D_3	Refined "All-Factors" dataset for PCA stationary US and German yield curve factors and estimated cointegrating relations of I(1) country factors	(%)
$PC_{i,t}^{y}$	where $i = 1, 2, 3$. First 3 principal components of "All-Yields" dataset, D_1	(%)
$PC_{i,t}^{s}$	where $i = 1, 2, 3$. First 3 principal components of "DE-US Spread" dataset, D_2	(%)
$PC_{i,t}^{f}$	where $i = 1, 2, ..., 5$. First 5 principal components of "Refined All-Factors" dataset, D_3	(%)
$F_n(k)$	Wald, Lagrange multiplier	
k	Date of structural change	
supF_n	Quandt or "Sup" test statistic	

SupF(π_0)	Asymptotic distribution of the "Sup" test statistic
π_0	Single index on which the asymptotic distribution of the "Sup" test depends
m	Number of coefficient breaks in the structural breaks models
n	Number of variance breaks in the structural breaks models
$(T_1^c, \dots, T_m^c)$	Coefficient break points, in the Perron-Zhou (2008) workflow
$(T_1^v, \dots, T_n^v)$	Variance break points, in the Perron-Zhou (2008) workflow
K	Total number of break dates, in the Perron-Zhou (2008) workflow
(TP-1)	Testing problem 1 of $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, n = n_a\}$
(TP-2)	Testing problem 2 of $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$
(TP-3)	Testing problem 3 of $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a\}$
(TP-4)	Testing problem 4 of $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$
(TP-5)	Testing problem 5 of $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, 1 \leq n \leq N\}$
(TP-6)	Testing problem 6 of $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, 1 \leq n \leq N\}$
(TP-7)	Testing problem 7 of $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{1 \leq m \leq M, n = n_a\}$
(TP-8)	Testing problem 8 of $H_0 : \{m = n = 0\}$ versus $H_1 : \{1 \leq m \leq M, 1 \leq n \leq N\}$
(TP-9)	Testing problem 9 of $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a + 1, n = n_a\}$
(TP-10)	Testing problem 10 of $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a + 1\}$
sup$LR_{1,T}$	with sup$LR_{1,T}(n_a, \epsilon \vert m = n = 0)$, sup-Likelihood ratio test for testing (TP-1)
sup$LR_{2,T}$	with sup$LR_{2,T}(m_a, n_a, \epsilon \vert n = 0, m_a)$, sup-Likelihood ratio test for testing (TP-2)
sup$LR_{3,T}$	with sup$LR_{3,T}(m_a, n_a, \epsilon \vert m = 0, n_a)$, sup-Likelihood ratio test for testing (TP-3)
sup$LR_{4,T}$	with sup$LR_{4,T}$, sup-Likelihood ratio test for testing (TP-4)
Λ	Search set for possible values of the break fractions in coefficients $(\lambda_1^c, \dots, \lambda_m^c)$ and variance $(\lambda_1^v, \dots, \lambda_n^v)$
$W(\cdot)$	Wiener process
sup$LR_{1,T}^*$	Modified sup-Likelihood ratio test with asymptotic distribution free of nuisance parameters, for testing (TP-1)
sup$LR_{2,T}^*$	Modified sup-Likelihood ratio test with asymptotic distribution free of nuisance parameters, for testing (TP-2)

$\sup LR^*_{4,T}$	Modified sup-Likelihood ratio test with asymptotic distribution free of nuisance parameters, for testing (TP-4)
$UDmaxLR^*_{1,T}$	Equal-weight double maximum test, for testing (TP-5)
$UDmaxLR^*_{2,T}$	Equal-weight double maximum test, for testing (TP-6)
$UDmaxLR_{3,T}$	Equal-weight double maximum test, for testing (TP-7)
$UDmaxLR^*_{4,T}$	Equal-weight double maximum test, for testing (TP-8)
$UDmaxF_{3,T}$	Equal-weight double maximum test as maximum of the Wald-type test, for testing (TP-7) when errors are serially correlated
$UDmaxF_{4,T}$	Equal-weight double maximum test as maximum of the Wald-type test, for testing (TP-8) when errors are serially correlated
$(\widetilde{T}^c_1, \ldots, \widetilde{T}^c_m)$	Estimates of the break dates in the regression coefficient, assuming m breaks in coefficients
$(\widetilde{T}^v_1, \ldots, \widetilde{T}^v_n)$	Estimates of the break dates in the variance of errors, assuming n breaks in the variance
$\sup Seq_T(m+1, n \mid m, n)$	Sequential test for testing for an additional break in coefficients, assuming m breaks in the coefficients and n breaks in the variance of the errors
$\sup Seq_T(m, n+1 \mid m, n)$	Sequential test for testing for an additional break in variance, assuming m breaks in the coefficients and n breaks in the variance of the errors

SSM symbols

$\mathbf{y}_t$	$(n \times 1)$ vector of measurements
$\boldsymbol{\xi}_t$	$(r \times 1)$ vector of state variables
$\mathbf{x}_t$	$(k \times 1)$ vector of exogenous (predetermined) variables
$\mathbf{F}$	$(r \times r)$ state transition coefficient matrix
$\mathbf{B}$	$(r \times r)$ state disturbance loading coefficient matrix
$\mathbf{A}$	$(n \times k)$ coefficient matrix of the exogenous variables
$\mathbf{H}$	$(n \times r)$ measurement sensitivity coefficient matrix
$\mathbf{D}$	$(n \times r)$ observation innovation coefficient matrix
$\mathbf{Q}$	$(r \times r)$ state covariance matrix
$\mathbf{R}$	$(n \times n)$ observation covariance matrix
$\widehat{\boldsymbol{\xi}}_{t+1\mid t}$	Linear least squares forecasts of the state vector
$\mathbf{P}_{t+1\mid t}$	MSE of $\widehat{\boldsymbol{\xi}}_{t+1\mid t}$
$\widehat{\boldsymbol{\xi}}_{1\mid 0}$	Unconditional mean of $\boldsymbol{\xi}_1$
$\mathbf{P}_{1\mid 0}$	Unconditional variance of $\boldsymbol{\xi}_1$
$\widehat{\mathbf{y}}_{t+1\mid t}$	Measurements forecasts
$\widehat{\boldsymbol{\xi}}_{t\mid t}$	Updated projection of $\boldsymbol{\xi}$
$\mathbf{P}_{t\mid t}$	MSE of $\widehat{\boldsymbol{\xi}}_{t\mid t}$
$\mathbf{K}_t$	Kalman gain
$\boldsymbol{\xi}_{t+s}$	s-step-ahead state forecast
$\widehat{\boldsymbol{y}}_{t+s\mid t}$	s-period-ahead forecast of the measurements $\mathbf{y}$

$\widehat{\boldsymbol{\xi}}_{t\|T}$	Smoothed state
$\mathbf{P}_{t\|t-1}$	MSE of $\widehat{\boldsymbol{\xi}}_{t\|T}$
$\{\widehat{\boldsymbol{\xi}}_{t\|T}\}_{t=1}^{T}$	Full set of smoothed states
$\{\mathbf{P}_{t\|T}\}_{t=1}^{T}$	Full set of MSE
$\mathbf{m}_t$	One-step prediction errors
$\boldsymbol{e_t}$	Standardized one-step prediction errors
$\widehat{\boldsymbol{v}}_{\boldsymbol{t}}$	Smoothed state disturbances
$\widehat{\boldsymbol{w}}_{\boldsymbol{t}}$	Smoothed observation disturbances
$\boldsymbol{e}_{\boldsymbol{t}}^{*}$	Standardized smoothed observation disturbances
$\boldsymbol{r}_{\boldsymbol{t}}^{*}$	Standardized smoothed state disturbances
$Q(k)$	Box-Ljung statistic for independence
$H(h)$	Test statistic for homoskedasticity
N	Test statistic for normality
S	Skewness
K	Kurtosis
$\mathbf{x}_t$	2D vector holding $s_{US,t}$ and $s_{DE,t}$ in the 2D-VAR(5) model
$\mathbf{A}_i$	Autoregressive coefficient matrices in the 2D-VAR(5) model
$\mathbf{u}_t$	Vector of error terms in the 2D-VAR(5) model
$\Delta\mathbf{y}_t$	4D vector holding $\Delta l_{US,t}, \Delta l_{DE,t}, \Delta c_{US,t}$, and $\Delta c_{DE,t}$ in the 4D-VEC(3) model
$\mathbf{B}_i$	Short-run coefficient matrices in the 4D-VEC(3) model
$\boldsymbol{\eta}_t$	Vector of error terms in the 4D-VEC(3) model
$\boldsymbol{\Gamma}_t$	Matrix of state shocks
$\boldsymbol{\Lambda}_t$	Matrix of measurement shocks
$\boldsymbol{\delta}$	Shock magnitude

I dedicate this thesis to my family.
To my late grandmother, Antonina, for teaching me the principles of life.
I love you all dearly.

Chapter 1

Introduction

Modeling and forecasting the yield curve has been an evergreen topic of debate among monetary policy makers, academics, and bond market participants. Today the topic gains significant importance given the "global" feature of capital markets. Substantial increases of cross-border portfolio investments, asset ownership, and bank lending are the consequences of a worldwide process of financial integration. Financial integration brings with itself the benefits of smoother consumption, via cross-country asset diversification, and the challenges of spillover effects and macroeconomic shocks to the interest rate markets. These shocks transmit internationally via monetary policy and risk channels. Since the yield curves of different world regions co-move, giving rise to contemporaneous and non-contemporaneous dynamic interdependencies, central banks, international fixed income investors, and risk managers, all have a vital interest in developing term structure models that allow for a joint, global evolution of yield curves in multiple currency areas.

Despite significant scientific advances in the last decades, a closer look to the literature on yield curve modeling and forecasting reveals, however, a number of gaps and shortcomings.

The Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model has gained significant popularity in the aftermath of the 2008 Financial Crisis, when regulators placed greater emphasis on the market valuation and accounting of liabilities. Thanks to its parsimony, accurate parameter estimation, strong forecastability at long horizons, and convenient extension to the global context, the Diebold-Li model is widely acknowledged as state-of-the-art for yield curve modeling and forecasting (Diebold and Li, 2006; Diebold, Li, and Yue, 2008; Diebold and Rudebusch, 2013). Even though the model presents numerous advantages, it disregards the in-sample dynamic properties of the yield curve factors.

Yield curve variables are known to exhibit persistent, unit-root dynamics. This observation might suggest that yield curve variables are integrated of order one, [I(1)]. Nevertheless, yield curve variables are commonly modeled in levels, disregarding a potential cointegration structure. With respect to the correlation structure, yield curve factors are often assumed to be uncorrelated and, therefore, a diagonality constraint is imposed on the covariance matrix. Such a restrictive assumption excludes the possibility of lead-lag relationships arising from contemporaneous and non-contemporaneous dependencies of yield curves across different world regions (Chinn and Frankel, 2003; Belke and Gros, 2005; Anderton, Di Mauro, and Moneta, 2004, Stock and

Watson, 2005, Rey, 2016). The lead-lag relationships among yield curve variables are also not thoroughly analyzed with the objective of determining whether specific yield curve variables possess explanatory power for other yield curve variables.

Very often economic time series may exhibit changes in the serial correlation, mean, and volatility, and these changes might be due to sudden and unexpected external events in particular time periods (Lütkepohl, 2005). Unusual behavior or structural breaks in the relations between the yield curve drivers of different world regions has not yet been studied, despite the wide range of tools and procedures available for their detection. Popular global yield curve models, such as the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" Model, extended to the global context by Diebold, Li, and Yue, 2008, assume parameter stability and fit the global yield curve factors to a VAR(1) process. Disregarding the existence of potential structural breaks might undermine the model's forecasting accuracy and lead to unreliable inference.

The root causes or nature of structural breaks also remain a topic scarcely investigated in the context of interest rates. The 2008 Financial Crisis clearly induced a monetary policy regime change, given that central banks around the world transitioned from a traditional monetary policy to a more accommodative and nontraditional monetary policy, aiming at guiding market participants in understanding the future course of the policy stance (Hanspeter, 2004, ECB, 2011a, ECB, 2011b, ECB, 2011c, Wyplosz, 2013, De La Dehesa, 2013, Rodriguez and Carrasco, 2014, Verhelst, 2014 and Delivorias, 2015;Fed, 2018). Understanding whether changes in the ability of market participants in anticipating monetary policy decisions can cause structural breaks in the interest rates requires empirical evidence, especially for time periods incorporating the 2008 Financial Crisis. If there are good reasons to believe that a change in monetary policy regime did occur in a given sample period and that the regime change affected the ability of market participants in predicting monetary policy decisions, one can formulate the hypothesis that the structural breaks in the drivers of the yield curves stem from such a regime change. Another hypothesis requiring investigation is whether structural breaks in the univariate dynamics of yield curve drivers can actually be due to missing variables with explanatory power. Therefore, more accurate results with respect to the existence of structural breaks can be derived from testing for structural breaks in multivariate systems (Bai, Lumsdaine, and Stock, 1998; Bai, 2000; Hansen, 2003; Qu and Perron, 2007) and computational simplicity can be exploited in a state-space framework (Commandeur and Koopman, 2007; Commandeur, Koopman, and Ooms, 2011; Durbin and Koopman, 2012).

Shifting the focus from modeling to forecasting, although there are many studies on yield curve modeling, the literature on yield curve forecasting remains limited. The arbitrage-free (Hull and White, 1990; Heath, Jarrow, and Morton, 1992) and affine models (Vasicek, 1977; Cox, Ingersoll Jr, and Ross, 1977; Duffie and Kan, 1996), focusing primarily on the in-sample fit, are known to perform poorly out-of-sample (Duffee, 2002). When the goal is to forecast the yield curve out-of-sample, the domestic (Nelson and Siegel, 1987; Litzenberger, Squassi, and Weir, 1995; Balduzzi et al., 1996; Chen, 1996; Bliss, 1997a; Bliss, 1997b; Andersen and Lund, 1997; Dai and Singleton, 2000; De Jong and Santa-Clara, 1999; Jong, 2000; Brandt and Yaron, 2003; Duffee, 2002) and global term structure factor models (Diebold, Li, and Yue, 2008; Jotikasthira, Le, and Lundblad, 2015) are very often preferred. Of this group of models, the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model is state-of-the-art, given its strong forecastability at long

horizons, where the model performs noticeably better than standard benchmarks, such as the random walk, slope regression, Fama-Bliss forward rate regression, and other autoregressive models.

Very little research is available on models for forecasting jointly the yield curves of different world regions. The main contribution is the Dynamic Nelson-Siegel model of Diebold and Li, 2006, extended to the global context by Diebold, Li, and Yue, 2008. The idea behind the model is a hierarchy for global yields, in the sense that, country yield curves depend on country factors, which in turn depend on global factors. An empirical application of the model to the term structure of government bond yields for Germany, Japan, UK, and US, finds evidence supporting the existence, economic importance, and explanatory power of global yield factors.

In addition, it is worth recalling that the models mentioned so far rely on very restrictive assumptions concerning the dynamic properties of the yield data. Conventional modeling approaches in the literature assume stationarity of yields and refrain from exploring the dynamic properties from a forecasting perspective.

This thesis on *Modeling and Forecasting the Co-Movement of International Yield Curve Drivers* aims at fulfilling the literature gaps outlined above by developing new data-driven state-space models to forecast the co-movement of yield curve drivers of different world regions and, from the drivers, the yield curves. The thesis contributes with the following.

- A rigorous and comprehensive study of what drives yield curves in different world regions. This study aims at answering questions such as, what are the international yield curve drivers, what are their dynamic properties, and how do they co-move. Using an extended sample of US and German yields (including very recent observations), this study provides empirical evidence about the nonstationarity/stationarity properties of the US and German yield curve drivers, volatility clustering, correlation and cross-correlation structure, causality linkages and lead-lag relationships, cointegration structure, and impulse-response functions. These results are instrumental in developing new econometric models for forecasting the co-movement of international yield curve drivers.

- A univariate and multivariate study of structural breaks in the data generation processes of the US and German yield curve drivers. Starting in a univariate setting, we test for the presence of structural breaks in our sample period using the methods of Bai and Perron, 1998 and Perron and Zhou, 2008. The decision of which of the two methods to apply is based on whether the Chow test model assumptions of normal, serially uncorrelated, and homoskedastic errors are satisfied. As such, this study employs the most suitable tools for the detection of structural breaks and estimation of their timing by taking into account the dynamic properties of the data. Given the results of the univariate analysis, we investigate the root causes of structural breaks in the US and German yield curve drivers with a study of the US Fed and ECB monetary policy predictability. The novelty of this work is to assess monetary policy predictability in the context of the term structure of interest rates, in order to investigate and understand the root causes of structural breaks in the US and German yield curve drivers. Moreover, we assess monetary policy predictability on a significantly larger sample period compared to previous literature. From the beginning

of the European Monetary Union (EMU), 1999:01, and up to recent days, 2018:01, we analyze 18 years of daily data. Furthermore, with the aim of investigating whether the presence of structural breaks is due to missing variables, we switch from a univariate to a multivariate state-space setting to advance further empirical results about the existence of outliers and structural breaks. For this purpose, we develop new data-driven state-space models for the co-movement of the international drivers. The novelty of the models is that they are designed to preserve the dynamic properties of the US and German yield curve drivers embodied in the VAR model for the slopes and the VEC model for the levels and curvatures. We call the main version of the models the Full State-Space Model (FSSM). We estimate the FSSM via the Kalman filter and maximum likelihood and test for outliers and structural breaks in the FSSM using the standardized smoothed observation disturbances and standardized smoothed state disturbances, respectively. It turns out that both outlying values and structural breaks are present in the FSSM; however, the alterations in structure resemble more of patches of outliers rather than structural breaks. We explain how to adjust the FSSM for the most blatant outliers by including intervention variables in the measurement equation. We call this new version of the FSSM the MShock-FSSM.

- Finally, using our newly developed state-space models (i.e., the FSSM and the MShock-FSSM), this thesis contributes to the literature on yield curve forecasting by providing a study of joint US and German out-of-sample yield curve forecasting with models designed to preserve the dynamic properties of the yield data. In addition, we verify the predictive power of the yield curve curvatures, by developing and forecasting with two other models for the US and German yield curve levels and slopes only. These are the FSSM^{LS} and the MShock-FSSM^{LS} models. The ultimate goal is to understand how do models that account for all dynamic properties of the yield data perform compared to the state-of-the-art Diebold-Li model.

Structure of the Thesis

The thesis can roughly be divided into three parts.

Part I, entitled *Dynamic Properties*, consists of Chapters 2, 3, and 4 and lays the groundwork for the development of our econometric models.

Chapter 2 provides a revision of the known drivers of the yield curves and of the models available in the existing literature to describe their dynamic evolution. Adopting a forecasting perspective, this Chapter explains the evolution of the notorious Nelson and Siegel, 1987 exponential components framework from the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model to the global model and discusses how these models successfully forecast the yield curves in a parsimonious fashion by forecasting their underlying factors, known as level, slope, and curvature. Chapter 3 recalls the theoretical concepts and methods of time series analysis, which are employed in the subsequent study of the dynamic properties of the international yield curve drivers. Chapter 4 addresses empirically the research questions using actively traded US and German government bond yield curves. Screening out the most robust dynamic properties of

the US and German yield curves and yield curve drivers, this Chapter provides the fundamentals of the new econometric models developed in subsequent Chapters of this thesis.

Part II, entitled *Outliers and Structural Breaks*, consists of Chapters 5, 6, and 7 and is dedicated to a comprehensive econometric study of outliers and structural breaks in the dynamics of US and German yield curve drivers.

Chapter 5 starts with a univariate analysis of structural breaks. Using the methods of Bai and Perron, 1998 and Perron and Zhou, 2008, this Chapter reports and discusses the results of testing for multiple structural breaks in the data generation processes of the US and German levels, slopes, and curvatures, occurring at unknown timing. Chapter 6 questions the nature of structural breaks by conducting a study of the US Fed and ECB monetary policy predictability. The investigated assumption is whether the monetary policy regime change caused by the 2008 Financial Crisis induced a change in the ability of market participants in predicting monetary policy decisions and, ultimately, created breaks in the dynamics of the US and German yield curve drivers. Chapter 7 investigates whether the presence of structural breaks is due to variables with predictive power missing in the univariate dynamics of the US and German yield curve drivers. To verify this assumption, this Chapter adopts a multivariate state-space setting and develops a new data-driven state-space model, the FSSM, for the co-movement of the US and German yield curve drivers. The presence of outliers and structural breaks is tested in the FSSM, which is subsequently adjusted for the most blatant outliers with the inclusion of intervention variables in the measurement equation. The adjusted version of FSSM is the MShock-FSSM.

Part III, entitled *Forecasting*, consists of Chapter 8 and explores the performance of the FSSM and MShock-FSSM in out-of-sample yield curve forecasting.

To this regard, Chapter 8 performs a recursive out-of-sample forecasting exercise with re-estimation of the parameters every 12 months with the Kalman filter and maximum likelihood and produces term-structure forecasts at both short and long horizons, for US and Germany. The forecasting performance of our models is benchmarked to the Diebold-Li "Yields-Only" model. The aim is to understand how do models that account for all the dynamic properties of the yield data perform compared to the state-of-the-art Diebold-Li model. In addition, this Chapter verifies the predictive power of the US and German curvatures, by developing and forecasting with two additional models for the US and German levels and slopes only. These are the FSSM^{LS} and the MShock-FSSM^{LS} models.

Chapter 9 concludes the thesis, by highlighting the most important results and outlining perspectives for further research.

Supplementary materials to Part I and II are included in Appendices A and B, respectively.

Part I

Dynamic Properties

Dynamic Properties

Part I of this thesis lays the groundwork for the development of new econometric models for forecasting the co-movement of international yield curve drivers and, from the drivers, the international yield curves. The three main research questions we seek to answer here are: which are the international yield curve drivers? What are their dynamic properties? How do they co-move? Part I is structured as follows.

Chapter 2 reviews the most important scientific studies of domestic and global term structure modeling for forecasting purposes. The aim is to list the known drivers of yield curves and the models available to describe their dynamic evolution. Adopting a forecasting perspective, we discuss the evolution of the notorious Nelson and Siegel, 1987 exponential components framework from the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model to the global model, in which a potentially large set of country yield curves are modeled jointly. These models parsimoniously forecast the yield curves by forecasting their underlying factors, known as yield curve level, slope, and curvature.

Given that our research questions are best answered in a multivariate setting, Chapter 3 recalls the theoretical concepts and methods of multiple time series analysis, which are employed in the subsequent study.

Chapter 4 addresses empirically the research questions using actively traded US and German government bond yield curves, for the sample period running from ’1999:01’ to ’2018:01’ and for seven of the most liquid maturities. Overall, the results provide evidence about the stationarity of the US and German slopes, nonstationarity of the levels and curvatures, cointegration structure between the levels and curvatures, and existence of Granger causality among all US and German yield curve drivers. We find that the data generation processes most suitable to capture these dynamic properties are a 2D-VAR(5) model for the US and German slopes and a 4D-VEC(3) model for the levels and curvatures.

Chapter 2

International Yield Curve Variables

2.1 Introduction

The starting point of this thesis is a revision of the most relevant literature on domestic and global term structure modeling, in order to list the known drivers of the yield curves at a domestic and global level.

Yield curve models of the last decades pertain to four main groups: *arbitrage-free models*, *affine models*, *domestic term structure factor models*, and *global term structure models*.

The models belonging to the arbitrage-free approach (Hull and White, 1990; Heath, Jarrow, and Morton, 1992) are required to contribute to the accurate pricing of derivative products by ensuring the absence of arbitrage opportunities. This requirement is satisfied if the models can provide a perfect fit of the term structure of the interest rates at a given point in time. Focusing only on the in-sample fit, the arbitrage-free models neglect the term structure dynamics or the forecasting requirements.

The affine models (Vasicek, 1977; Cox, Ingersoll Jr, and Ross, 1977; Duffie and Kan, 1996) try to meet the requirement of modeling the dynamics of the instantaneous rate, from which yields at other maturities can be derived, upon assumptions about the risk premium. Similarly to the arbitrage-free models, the affine models focus primarily on the in-sample fit, as they are known to perform poorly out-of-sample (Duffee, 2002).

The domestic (Nelson and Siegel, 1987; Litzenberger, Squassi, and Weir, 1995; Balduzzi et al., 1996; Chen, 1996; Bliss, 1997a; Bliss, 1997b; Andersen and Lund, 1997; Dai and Singleton, 2000; De Jong and Santa-Clara, 1999; Jong, 2000; Brandt and Yaron, 2003; Duffee, 2002) and global (Diebold, Li, and Yue, 2008; Jotikasthira, Le, and Lundblad, 2015) term structure factor models, instead, focus on an accurate out-of-sample forecasting of the term structure of a given country. The pioneering work, from this perspective, is represented by the Nelson and Siegel, 1987 exponential components framework, with which we start our discussion in Section 2.2. Section 2.3 introduces the Diebold-Li model and discusses its novelty of interpreting the Nelson-Siegel parameters as time-varying yield curve level, slope, and curvature. Section 2.4 continues the exposition of Diebold-Li by introducing the Diebold-Li "Yields-Only" model and its extension to include macroeconomic variables. Along this line, Section 2.5 discusses the extension of the

Diebold-Li "Yields-Only" model to the global context by allowing for both global and country-specific yield curve drivers. Section 2.6 discusses Principal Component Analysis as a suitable tool to tackle the high dimensionality problem of yield curves and to capture the potential co-movement of yield curves. Finally, Section 2.7 discusses how Cointegration Analysis can be employed to obtain insights into the relations between yields of different maturities while effectively accounting for the potential presence of unit roots in yield data.

2.2 Nelson and Siegel, 1987

In the Nelson and Siegel, 1987 exponential components framework, the entire yield curve is modeled as a function of three parameters. Such a modeling framework represented a widely accepted solution to the problems associated with the high dimensionality of yield data and the need for a parsimonious model for the yield curves that would have the ability to represent the shapes generally associated with yield curves: monotonic, humped, and S-shaped.

Nelson and Siegel derive their parsimonious model by noticing that a class of functions that readily generates the typical yield curve shapes is that associated with solutions to differential or difference equations. This observation is supported by the expectations theory of the term structure of interest rates, since, if spot rates are generated by a differential equation, then forward rates, being the object of the forecasting exercise, will be the solution to the equations. Reporting the example of Nelson and Siegel, if the instantaneous forward rate at maturity m, denoted $r(m)$, is given by the solution to a second-order differential equation with real and unequal roots, we would have:

$$r(m) = \beta_0 + \beta_1 e^{-m/\tau_1} + \beta_2 e^{-m/\tau_2}, \tag{2.1}$$

where τ_1 and τ_2 are time constants associated with the equation, and β_0, β_1, and β_2 are determined by initial conditions. Nelson and Siegel argue that this equation generates a family of forward rate curves that take on monotonic, humped, or S shapes depending on the values of β_1 and β_2 and that also have asymptote β_0. The yield to maturity, denoted $R(m)$, is the average of the forward rates:

$$R(m) = \frac{1}{m} \int_0^m r(x)dx, \tag{2.2}$$

and the yield curve implied by the model displays the same range of shapes. Since the model in 2.1 turns out to be overparameterized, the more parsimonious model proposed by Nelson and Siegel (that can generate the same range of shapes) is given by the solution equation for the case of equal roots:

$$r(m) = \beta_0 + \beta_1 e^{-m/\tau} + \beta_2 [(m/\tau)e^{-m/\tau}]. \tag{2.3}$$

In this model, one can evaluate the shape flexibility by interpreting the coefficients as measuring the strengths of the short-, medium-, and long-term components of the forward rate curve, and hence, of the yield curve. The long-term component, β_0, is a constant that does not

decay to zero in the limit. The medium-term component, β_2, is the only function in the model that starts out at zero (and is therefore not short term) and decays to zero (and is therefore not long term). The short-term component, β_1, has the fastest decay of all functions within the model that decay monotonically to zero. The novelty of the Nelson and Siegel model lies, therefore, in the fact that with the appropriate choices of weights for the three components, one can generate a variety of yield curves based on forward rate curves with monotonic and humped shapes.

The weakness of the Nelson-Siegel model, however, is that it is a static model. The model's parameters, β_0, β_1, β_2 and m, do not have a time subscript. This weakness was approached by Diebold and Li, 2006, who also provide a different interpretation of the model's parameters.

2.3 Diebold and Li, 2006

Diebold and Li, 2006 take the Nelson-Siegel yield curve, dynamize it by making the three model's parameters varying over time and invert the interpretation of the model. Recalling the notation of Diebold and Li, if $y(\tau)$ denotes the continuously compounded zero-coupon nominal yield to maturity τ, the Nelson-Siegel yield curve [1] reads as follows:

$$y_t(\tau) = \beta_{1t} + \beta_{2t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) - \beta_{3t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \nu(\tau) \tag{2.4}$$

where $\beta_{1t}, \beta_{2t}, \beta_{3t}$, and τ are the model's parameters and the exponential terms are the model's variables. The inverted interpretation of the Nelson-Siegel model is obtained by considering the three parameters as three *latent dynamic factors* (and, thus, as the model's variables) and the exponential terms as the *factor loadings* (and, thus, as the model's parameters). Moreover, the three latent dynamic factors can be interpreted in terms of level, slope, and curvature of the yield curve[2].

As Diebold-Li explain, the long-term factor, β_{1t}, governs the yield curve level, as an increase in β_{1t} increases all yields equally. This behavior happens because the loading is identical at all maturities, thereby β_{1t} changes the level of the yield curve. The short-term factor, β_{2t} governs the yield curve slope, as an increase in β_{2t} increases short yields more than long yields, because the short rates load on β_{2t} more heavily, thereby changing the slope of the yield curve. Finally, the medium-term factor, β_{3t}, governs the yield curve curvature. An increase in β_{3t} will have little effect on very short or very long yields, which load minimally on it, but will increase medium-term yields, which load more heavily on it, thereby increasing yield curve curvature.

Let L_t, S_t and C_t denote the level, slope, and curvature. The dynamic Nelson-Siegel yield

[1] The Nelson-Siegel yield curve is obtained by integrating the Nelson-Siegel forward rate curve in 2.3 over all maturities m (or τ, following the notation of Diebold-Li).

[2] Diebold and Li, 2006 define empirically the yield curve level as $y_t(\infty) = \beta_{1t}$, the yield curve slope as the ten-year yield minus the three-month yield, and the yield curve curvature as twice the two-year yield minus the sum of the ten-year and the three-month yields.

curve, as re-interpreted by Diebold and Li, reads as follows:

$$y_t(\tau) = L_t + S_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} \right) - C_t \left(\frac{1 - e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau} \right) + \nu_t(\tau) \tag{2.5}$$

Diebold and Li propose and estimate autoregressive models for the factors to show that the models are consistent with most of the stylized facts regarding the yield curve[3] and that the three time-varying factors can be estimated with high efficiency. The Diebold-Li model (also known as the "yields-only" model) gained significant popularity as a successful domestic factor model. Even today it is widely used to fit intra-country bond yields, facilitating extraction of latent level, slope, and curvature components.

Subsequent financial research has strengthen the evidence that changes in the yield curve are attributable to a few unobservable factors. In the context of asset pricing, financial economists and bond traders have developed and estimated numerous models to characterize the movement of these latent factors and, thereby, that of the yield curve. Few of these models, however, investigate the nature of the factors, identify the underlying forces that drive their movements or study the response of these factors to macroeconomic variables (Wu, 2003).

2.4 Diebold, Rudebusch, and Aruoba, 2006

Since the interest rates represent the tool monetary policy makers use to determine the economic environment, the domestic factor models were quickly evolved to include observable macroeconomic variables, such as, the real activity, inflation, and the monetary policy instrument (Ang and Piazzesi, 2003,Diebold, Rudebusch, and Aruoba, 2006). These models succeed at improving the out-of-sample forecasting performance by accounting for the dynamic interactions between the macroeconomy and the yield curve.

Diebold, Rudebusch, and Aruoba, 2006 exploit the state-space representation of the dynamic Nelson-Siegel model in 2.5, to provide a model that characterizes the dynamic interactions between the macroeconomy and the yield curve.

The dynamic Nelson-Siegel model is naturally cast in state-space form if one assumes an autoregressive structure for the factor dynamics. The transition equation governing the dynamics of the state vector is:

$$\begin{pmatrix} L_t - \mu_L \\ S_t - \mu_S \\ C_t - \mu_C \end{pmatrix} = \begin{pmatrix} \alpha_{11} & \alpha_{12} & \alpha_{13} \\ \alpha_{21} & \alpha_{22} & \alpha_{23} \\ \alpha_{31} & \alpha_{32} & \alpha_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} - \mu_L \\ S_{t-1} - \mu_S \\ C_{t-1} - \mu_C \end{pmatrix} + \begin{pmatrix} \eta_t(L) \\ \eta_t(S) \\ \eta_t(C) \end{pmatrix} \tag{2.6}$$

$t = 1, ..., T$. The measurement equation relating the set of N yields to the three unobservable

[3] For a comprehensive list of these stylized facts, we refer the reader to Diebold and Li, 2006, p. 343.

factors is:

$$\begin{pmatrix} y_t(\tau_1) \\ y_t(\tau_2) \\ \dots \\ y_t(\tau_N) \end{pmatrix} = \begin{pmatrix} 1 & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} & \frac{1-e^{-\tau_1\lambda}}{\tau_1\lambda} - e^{-\tau_1\lambda} \\ 1 & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} & \frac{1-e^{-\tau_2\lambda}}{\tau_2\lambda} - e^{-\tau_2\lambda} \\ \vdots & \vdots & \vdots \\ 1 & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} & \frac{1-e^{-\tau_N\lambda}}{\tau_N\lambda} - e^{-\tau_N\lambda} \end{pmatrix} \begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} + \begin{pmatrix} \epsilon_t(\tau_1) \\ \epsilon_t(\tau_2) \\ \vdots \\ \epsilon_t(\tau_N) \end{pmatrix} \tag{2.7}$$

$t = 1, \dots, T$. The orthogonal, Gaussian white noise processes η_t and ϵ_t are defined such that:

$$\begin{pmatrix} \eta_t \\ \epsilon_t \end{pmatrix} \sim WN \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} Q & 0 \\ 0 & H \end{pmatrix} \right) \tag{2.8}$$

under the assumptions that the covariance matrix H is diagonal and that the covariance matrix Q is non-diagonal. Diagonality of the H matrix implies that deviations of yields of various maturities from the yield curve are uncorrelated and non-diagonality of the Q matrix implies that shocks to the three term structure factors can be correlated.

Within this setup, Diebold, Rudebusch, and Aruoba find strong evidence of macroeconomic effects on the future yield curve and weak evidence of yield curve effects on future macroeconomic developments.

2.5 Diebold, Li, and Yue, 2008

With the financial integration reaching a global magnitude, the evolution of the term structure models to the global environment becomes inevitable. The academia, policy makers, and market participants start to ask the question of why do term structures in different currencies co-move? The existence of global factors, the nature of dynamic cross-country bond yield interactions, the international spillovers and the transmission channels of the international macroeconomic shocks represent economic phenomena requiring high attention. Al Awad and Goodwin, 1998 examine the dynamic linkages between short-run and long-run weekly real interest rates, for a set of multiple countries, to provide evidence of well-integrated international asset markets. The integration of financial markets is particularly strong in the long run, when full transmission of international interest rates to local interest rates takes place (Frankel, Schmukler, and Serven, 2004). Yield curve fluctuations are transmitted across different currencies through the monetary policy channel and through the risk compensation channel. Taking the example of the United States and the European Union, a reciprocal leader-follower relationship does seem to exist, in the long run, between the ECB and the Fed (Chinn and Frankel, 2003; Belke and Gros, 2005). The ECB follows the Fed in setting its monetary policy; the Fed is also increasingly influenced by the ECB, although the relationship is asymmetric.

The international financial integration implies that the monetary policy shocks of the leading countries are transmitted internationally to the follower countries, affecting their financial conditions and giving rise to co-movement of business cycles across countries (Anderton, Di Mauro, and Moneta, 2004, Stock and Watson, 2005, Rey, 2016).

The prominent work of Diebold, Li, and Yue, 2008 finds strong evidence for the existence of global yield factors, their high economic importance, and their direct linkage to global macroeconomic fundamentals such as inflation and real activity. These findings stem from the extension of the Diebold and Li, 2006 model to a global context, in which a potentially large set of country yield curves is modeled by allowing for both global and country-specific factors. The so-called "Generalized Nelson-Siegel Model" of Diebold, Li, and Yue is a hierarchical dynamic model for the global yields. The hierarchy is built by allowing the country yield curves to depend on country factors and the country factors to depend on global factors. More specifically, the global yields are allowed to depend on global factors as follows:

$$Y_t(\tau) = L_t + S_t\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau}\right) - C_t\left(\frac{1-e^{-\lambda_t\tau}}{\lambda_t\tau} - e^{-\lambda_t\tau}\right) + V_t(\tau). \tag{2.9}$$

The global factors follow a first-order vector autoregression[4]:

$$\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix} = \begin{pmatrix} \Phi_{11} & \Phi_{12} & \Phi_{13} \\ \Phi_{21} & \Phi_{22} & \Phi_{23} \\ \Phi_{31} & \Phi_{32} & \Phi_{33} \end{pmatrix} \begin{pmatrix} L_{t-1} \\ S_{t-1} \\ C_{t-1} \end{pmatrix} + \begin{pmatrix} u_t^l \\ u_t^s \\ u_t^c \end{pmatrix} \tag{2.10}$$

where u_{it} are disturbances such that $Eu_{it}^n u_{i't'}^{n'} = (\sigma_i^n)^2$ if $i = i'$, $t = t'$ and $n = n'$, and 0 otherwise, $n = l, s, c$. The index i denotes the country in the set being modeled. Each country's yield curve remains characterized by the dynamic Nelson-Siegel functional form:

$$y_{it}(\tau) = l_{it} + s_{it}\left(\frac{1-e^{-\lambda_{it}\tau}}{\lambda_{it}\tau}\right) - c_{it}\left(\frac{1-e^{-\lambda_{it}\tau}}{\lambda_{it}\tau} - e^{-\lambda_{it}\tau}\right) + \nu_{it}(\tau). \tag{2.11}$$

The country specific factors, l_{it}, s_{it}, c_{it}, are allowed to load on global common factors, L_{it}, S_{it}, C_{it}, as well as on country idiosyncratic factors[5]:

$$\begin{aligned} l_{it} &= \alpha_i^l + \beta_i^l L_t + \epsilon_{it}^l \\ s_{it} &= \alpha_i^s + \beta_i^s S_t + \epsilon_{it}^s \\ c_{it} &= \alpha_i^c + \beta_i^c C_t + \epsilon_{it}^c \end{aligned} \tag{2.12}$$

where $\left\{\alpha_i^l, \alpha_i^s, \alpha_i^c\right\}$ are constant terms, $\left\{\beta_i^l, \beta_i^s, \beta_i^c\right\}$ are the loadings on global common factors, $\left\{\epsilon_i^l, \epsilon_i^s, \epsilon_i^c\right\}$ are country idiosyncratic factors, $i = 1, ..., N$.

The country idiosyncratic factors follow a first-order vector autoregression, representing the dynamic equation for the country idiosyncratic factors:

$$\begin{pmatrix} \epsilon_{it}^l \\ \epsilon_{it}^s \\ \epsilon_{it}^c \end{pmatrix} = \begin{pmatrix} \phi_{i,11} & \phi_{i,12} & \phi_{i,13} \\ \phi_{i,21} & \phi_{i,22} & \phi_{i,23} \\ \phi_{i,31} & \phi_{i,32} & \phi_{i,33} \end{pmatrix} \begin{pmatrix} \epsilon_{i,t-1}^l \\ \epsilon_{i,t-1}^s \\ \epsilon_{i,t-1}^c \end{pmatrix} + \begin{pmatrix} u_{it}^l \\ u_{it}^s \\ u_{it}^c \end{pmatrix} \tag{2.13}$$

[4]This dynamic equation for the global factors will subsequently represent the transition equation in the state-space model.

[5]These equations represent the country factor decomposition equation and, as well as, the measurement equation in the state-space model.

where u_{it}^n are disturbances such that $Eu_{it}^n u_{i't'}^{n'} = (\sigma_i^n)^2$ if $i = i'$, $t = t'$ and $n = n'$, and 0 otherwise, $n = l, s, c$.

An important feature of the Diebold, Li and Yue approach is that it does not require that the global yields or the global yield factors be observed. As the authors explain, the "global yields" are substituted via the factor structure in 2.9. Doing so eliminates the need to observe the global yields or even define them directly. The underlying factors, L, S, and C are treated as latent factors in a state-space framework, in which the measurement equation is given by:

$$\underbrace{\begin{pmatrix} y_{1t}(\tau_1) \\ y_{1t}(\tau_2) \\ \cdots \\ y_{Nt}(\tau_J) \end{pmatrix}}_{\substack{\text{Global} \\ \text{Yields}}} = A \underbrace{\begin{pmatrix} \alpha_1^l \\ \alpha_1^s \\ \cdots \\ \alpha_N^c \end{pmatrix}}_{\substack{\text{constant} \\ \text{terms for} \\ \text{N-countries}}} + B \underbrace{\begin{pmatrix} L_t \\ S_t \\ C_t \end{pmatrix}}_{\substack{\text{Global} \\ \text{Yield} \\ \text{Factors}}} + A \underbrace{\begin{pmatrix} \epsilon_{1t}^l \\ \epsilon_{1t}^s \\ \vdots \\ \epsilon_{Nt}^c \end{pmatrix}}_{\substack{\text{Country} \\ \text{Idiosyncratic} \\ \text{Factors}}} + \underbrace{\begin{pmatrix} \nu_{1t}(\tau_1) \\ \nu_{1t}(\tau_2) \\ \vdots \\ \nu_{Nt}(\tau_J) \end{pmatrix}}_{\substack{\text{disturbances for} \\ \text{the Country} \\ \text{Specific Yields}}} \tag{2.14}$$

where

$$A = \begin{pmatrix} 1 & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1}\right) & \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1}\right) & 0 & \cdots & 0 \\ 1 & \left(\frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2}\right) & \left(\frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2}\right) & 0 & \cdots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 1 & \left(\frac{1-e^{-\lambda\tau_J}}{\lambda\tau_J}\right) & \left(\frac{1-e^{-\lambda\tau_J}}{\lambda\tau_J} - e^{-\lambda\tau_J}\right) \end{pmatrix}$$

and

$$B = \begin{pmatrix} \beta_1^l & \beta_1^s \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1}\right) & \beta_1^c \left(\frac{1-e^{-\lambda\tau_1}}{\lambda\tau_1} - e^{-\lambda\tau_1}\right) \\ \beta_1^l & \beta_1^s \left(\frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2}\right) & \beta_1^c \left(\frac{1-e^{-\lambda\tau_2}}{\lambda\tau_2} - e^{-\lambda\tau_2}\right) \\ \cdots & \cdots & \cdots \\ \beta_N^l & \beta_N^s \left(\frac{1-e^{-\lambda\tau_J}}{\lambda\tau_J}\right) & \beta_N^c \left(\frac{1-e^{-\lambda\tau_J}}{\lambda\tau_J} - e^{-\lambda\tau_J}\right) \end{pmatrix}$$

and the transition equations are the union of 2.10 and 2.13, i.e., the transition equations are the union of the dynamic equation for the global factors and of the dynamic equation for the country idiosyncratic factors.

The estimation results of Diebold, Li and Yue indicate that global yield factors do indeed exist and are economically important. The global level (relating to global inflation) and the global slope (relating to real economic activity) are found to explain significant fractions of country yield curve dynamics.

In terms of global macroeconomic fundamentals, other works support the results of Diebold, Li and Yue. Global inflation (Borio and Filardo, 2007; Ciccarelli and Mojon, 2010; Byrne, Fazio, and Fiess, 2012) and international business cycles (Lumsdaine and Prasad, 2003; Kose, Otrok, and Whiteman, 2003; Hellerstein, 2011; Dahlquist and Hasseltoft, 2013) explain large portions of the variance of country-specific inflation and global bond risk premia.

In the tradition of Diebold, Li, and Yue, 2008, Spencer and Liu, 2010; Bauer and Rios, 2012; Abbritti et al., 2013; Jotikasthira, Le, and Lundblad, 2015; Byrne, Cao, and Korobilis, 2017, among others, extend the country-specific term structure models to the multi-country setting to incorporate the international dynamics of the term structure and find that yield curve fluctuations across different currencies are highly correlated, that macroeconomic variables are important drivers of international term and foreign exchange risk premia as well as expected exchange rate changes, and that global factors explain long-term dynamics in yield curves, as opposed to domestic factors, which are instead accountable for short-run movements.

2.6 Principal Component Analysis for Yield Curves

One of the main challenges posed by the yield data is the high dimensionality. Expanding from a domestic to an international environment, this challenge becomes even more critical. Principal Component Analysis (PCA) is the model reduction technique most widely applied to the interest rate markets to describe yield curve dynamics in a parsimonious manner. Malava, 1999 performs direct PCA of international yields to find that 14 principal components are needed to explain 99% of variability in the joint term structure of LIBOR USD, JPY, EUR, and GBP yield curves.

In the context of yield curves, the PCA technique is used not only to tackle the high-dimensionality issue but also to capture potential co-movement of yield curves. Phoa, 2000 uses PCA to look at intra-country yield data and identify patterns of co-movement between yields at different maturities. Using U.S. Treasury market data, the author finds that two major kinds of co-movement explain most variation in bond yields. For an inter-country analysis, the author uses PCA to decompose international 10-year bond yields and conclude that the global shift factor explains more movement in the intra-country models than it does in the inter-country ones.

For a more novel study of the term structure of sovereign yield spreads, Wellmann and Trück, 2018 apply PCA to five sovereign spread data sets of advanced economies (Australia, Canada, Switzerland, Japan, UK and the US) to show that the term structure of all sovereign spreads is driven by three latent factors that one can interpret as spread level, spread slope, and spread curvature. Estimation results show that the three spread factors explain approximately 99% of the entire variation in the term structure of spreads between US interest rates and yields in Australia, Canada, Japan, Switzerland, and UK. In-sample prediction results confirm that the predictive power of the extracted latent spread factors for the exchange rate movements and excess return.

Wide uses of PCA are found in fixed income risk measurement and management for risk estimation, risk reporting, and scenario analysis (Golub and Tilman, 2000). For risk estimation and interest rate risk measurement purposes, the benefit of PCA relates to its ability to parsimoniously describe complex structures. The entire distribution of interest rates can be described by the more compact distribution of principal components. Working with the distribution of fewer variables (which retain most of the variability of the initial structure) comes at reduced simulation costs and increased accuracy, in the cases where, for example, risk systems employ Monte Carlo simulation methods to estimate the distribution of portfolio returns. Risk reporting

is simplified as contributions to portfolio risk can be analyzed via factors derived directly from actual market data, with no need of a priori postulation. PCA of yield curves provides helpful insights when performing scenario analysis, since, before applying yield curve shocks, one can understand the shape and dynamics of yield curve movements. Finally, PCA allows to describe the joint distribution of interest rates, thus, the probability of any particular risk scenario can be calculated.

2.7 Cointegration Analysis for Yield Curves

Most of the studies mentioned so far build their models upon principles of Economic Theory; principles which appear to be violated, however, given evidence flowing in from today's markets environment. Such principles state that nominal bond yields can not assume negative values, but they would eventually do so if they contained unit roots. Economic Theory excludes the existence of unit roots in the nominal yield series and, hence, the existence of integrated [I(1)] nominal yields, upon justification that nominal yields have a lower bound support at zero and an upper bound support lower than infinity. The works of Nelson and Siegel, 1987; Dai and Singleton, 2000; Duffee, 2002; Ang and Piazzesi, 2003; Diebold and Li, 2006; Diebold, Rudebusch, and Aruoba, 2006; Diebold, Li, and Yue, 2008, among others, disregard the in-sample properties of nominal yield data and model yields in levels.

The theory of cointegrated vector autoregressive models, on the other side, is abundant in applications to the term structure of interest rates in order to deal effectively with the unit root and, hence, with the nonstationarity property of these time series and, more generally, to investigate the relations between yields of different maturities.

The work of Hall, Anderson, and Granger, 1992, which extended the bivariate cointegration approach of Engle and Granger, 1987 and Campbell and Shiller, 1987 to the multivariate case, strengthens the expectations theory of the term structure, which states that the long-term bond rate is determined purely by current and future expected short-term rates. According to this expectation hypothesis, the spreads between different maturities make up the cointegrating vectors. Considering only one common trend, the term premia exhibits a mean-reverting or even a constant tendency.

The subsequent studies of Shea, 1992; Zhang, 1993; and Carstensen, 2003; among others, support the results of Hall, Anderson, and Granger, 1992.

With regards to the more technical aspects of the cointegrating relationships estimation methods, the likelihood-based approach proposed by Johansen, 1992 is shown to exhibit better properties compared to the ordinary least squares methods, nonlinear least squares, principal components, and canonical correlations (Gonzalo, 1994). Johansen's cointegration testing approach (Johansen, 1995) considers a nonstationary vector autoregressive process, integrated of order 1 and generated by i.i.d. Gaussian errors. If the innovations demostrate, instead, leptokurtic behavior, the non-Gaussian pseudolikelihood ratio test proposed by Lucas, 1997 is shown to have higher power than the Gaussian test of Johansen.

2.8 Conclusion

In the present Chapter, we reviewed the most important scientific works on domestic and global term structure modeling. The aim was to list the known drivers of yield curves and the state-of-the-art models available to describe their dynamic evolution.

Adopting a forecasting perspective, we started our discussion with the notorious Nelson and Siegel, 1987 exponential components framework, in which the entire yield curve is modeled as a function of three parameters. Despite being a parsimonious model that allows for a wide variety of yield curves, the Nelson-Siegel model remains a static model. We discussed how this weakness was approached by Diebold and Li, 2006, who dynamized the model by allowing the parameters to vary over time. The three parameters drive the entire yield curve by governing the yield curve level, slope, and curvature. Within the Diebold-Li framework, the three drivers can be estimated with high efficiency and because the yield curve depends entirely on these three drivers, forecasting the yield curve is equivalent to forecasting the yield curve drivers. The Diebold-Li "Yields-Only" model gained significant popularity as a successful domestic factor model, extensively used to fit intra-country bond yields, easily allowing for extensions to incorporate macroeconomic variables (Diebold, Rudebusch, and Aruoba, 2006) and extensions to the global context (Diebold, Li, and Yue, 2008), by modeling a potentially large set of country yield curves. In addition, we discussed how PCA can be a suitable tool to tackle the high dimensionality problem of yield curves and to capture the potential co-movement of yield curves. Most of the models we reviewed rely upon principles of Economic Theory, stating that nominal bond yields cannot assume negative values but they would eventually do so if they contained unit roots. To this regard, we discussed how the theory of cointegrated vector autoregressive models can provide helpful insights into the relations between yields of different maturities while effectively dealing with the unit root and, hence, with the nonstationarity property of yield data.

Chapter 3

Methods for Multiple Time Series Analysis

3.1 Introduction

In this chapter, we review and discuss the methods we employ to obtain insights into the dynamic structure of systems of international yield curve drivers. Given a system of international yield curve drivers, it is of interest to us to learn the dynamic interrelationships between these variables. The questions we seek to answer are the following. What are the dynamic properties of the single yield curve variables? How do the variables co-move? Are there contemporaneous dynamic interdependencies? Is there commonality in the movements? Are there non-contemporaneous dependency patterns like causality linkages or lead-lag relationships? Do the variables have common trends so that they move together to some extent?

As these questions are best answered in a multivariate setting, in the sequel, we recall (mainly from Hamilton, 1994; Lütkepohl and Krätzig, 2004; Shumway and Stoffer, 2000; Lütkepohl, 2005; Tsay, 2005; Rachev et al., 2007; Brockwell and Davis, 2013; and Box et al., 2015;) and discuss basic to advanced techniques of multiple time series analysis.

The Chapter proceeds as follows. Section 3.2 defines the Vector Autoregressive (VAR) models and discusses their underlying assumptions, properties, and estimation methods. VAR models are suitable methods for exploiting *lead-lag relationships* among variables. Along these lines, Section 3.3 introduces the concept of Granger Causality and discusses how this method can be employed to obtain useful information about the nature of the interactions among variables and the way they influence each other. Section 3.4 explains how the concept of Granger causality can be studied further with the Impulse Response (IR) function, in order to quantify the influence and the temporal profile of a change in one variable on the other variable in the system. From Section 3.5 onwards, we discuss the equivalent methods for Vector Error Correction (VEC) processes. More specifically, Section 3.5 introduces the VEC models, which are suitable for modeling and analyzing data that exhibit nonstationary behavior. Section 3.6 explains how to test for Granger noncausality in a cointegrating system using the Toda and Yamamoto, 1995 approach and Section 3.7 explains how to derive the IR function. Section 3.8 reviews the theory

of PCA, as a technique commonly employed when dealing with high dimensional data. Section 3.9 concludes the chapter.

3.2 Vector Autoregressive (VAR) Models: Assumptions, Properties, and Estimation Methods

VAR Models Defined

Vector autoregressive (VAR) models are models of vector of variables as autoregressive processes, where each variable depends linearly on its own lagged values and those of the other variables in the vector. The future values of the process are a weighted sum of past and present values plus some noise. Because of this variable dependency, VAR models are suitable for exploiting *lead-lag relationships*, i.e., relationships where the values of the "leader" variables anticipate values of the "laggard" variables.

A vector autoregressive model of order p, VAR(p), has the following general form[1]:

$$\mathbf{x}_t = \mathbf{A}_1\mathbf{x}_{t-1} + \mathbf{A}_2\mathbf{x}_{t-2} + \cdots + \mathbf{A}_p\mathbf{x}_{t-p} + \mathbf{s}_t + \varepsilon_t \tag{3.1}$$

where $\mathbf{x}_t = (x_{1,t}, \ldots, x_{n,t})'$ is a multivariate stochastic series in vector notation; $\mathbf{A}_i$, $i = 1, 2, \ldots, p$ are deterministic $n \times n$ matrices; $\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{n,t})'$ is a multivariate noise with variance-covariance matrix $\mathbf{\Omega}$; and $\mathbf{s}_t = (s_{1,t}, \ldots, s_{n,t})'$ is a vector of deterministic terms[2].

Stationarity, Stability, and Invertibility

The theory of VAR models assumes the *stationarity, stability, and invertibility* conditions[3]. A stochastic process is called *weakly stationary* or covariance-stationary if the expectation of x_t, $E(x_t)$, and the autocovariances, $Cov(x_t, x_{t-k})$, do not vary with time and are finite. A process is called *strictly stationary* if all finite-dimensional distributions are time-invariant. Rachev et al., 2007 call *weakly asymptotically stationary* a process that starts at a time origin and is such that its first and second moments (i.e., expectations and variances-covariances) converge to finite limits. The *Wold decomposition theorem*[4], states that any zero-mean, covariance stationary process $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ can be represented in a unique way as the sum of stochastic process and linearly predictable deterministic process, where the stochastic part is represented as an infinite moving average.

The *stability* conditions of a VAR process require that the roots of the reverse characteristic equation be strictly outside of the unit circle[5]. If the stability conditions are satisfied, then the relative VAR process is stationary if it extends on the entire time axis and is asymptotically

[1] In the sequel, we use the notation of Rachev et al., 2007.

[2] In our applications, the deterministic term will consist of constant intercept terms, i.e., $\mathbf{s}_t = \nu$.

[3] For a complete explanation of these conditions, we refer the reader to Rachev et al., 2007, p. 322

[4] Wold, 1938.

[5] If the roots of the reverse characteristic equation are outside the unit circle, all past shocks (i.e., noise terms) decay exponentially over time.

stationary if it starts from initial conditions. Stability, therefore, implies stationarity. The converse, however, is not true. There are stationary processes that are not stable.

If a VAR process satisfies the stability conditions and is stationary, then the process is *invertible* and can be written in an *infinite moving average* representation.

Solutions of VAR(p) Models

With respect to the solutions for VAR(p) models, which are given by the sum of a deterministic part and a stochastic part, we recall that the deterministic part depends on the initial conditions and deterministic terms and that stochastic part depends on random shocks. More specifically, if the process is stable, the stochastic part is a weighted sum of the most recent shocks, as the shocks in the distant past have only a negligible effect. If the process is integrated, the stochastic part is the cumulation of all past shocks, as the effects of shocks never decay over time. If the process is explosive, then shocks are amplified as time passes.

Equivalence of VAR(p) and VAR(1) Models

One key fact about VAR(p) models is that they can be simplified to VAR(1) models by adding appropriate variables. In particular, an n-dimensional VAR(p) model of the form

$$\mathbf{x}_t = (\mathbf{A}_1 L + \mathbf{A}_2 L^2 + \cdots + \mathbf{A}_p L^p)\mathbf{x}_t + \mathbf{s}_t + \varepsilon_t \tag{3.2}$$

is transformed into the following np-dimensional VAR(1) model

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_t + \mathbf{S}_t + \mathbf{W}_t \tag{3.3}$$

where

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_n & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I}_n & \dots & 0 & 0 \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{I}_n & 0 \end{bmatrix}, \mathbf{S}_t = \begin{bmatrix} \mathbf{s}_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{W}_t \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{3.4}$$

where $\mathbf{X}_t$, $\mathbf{S}_t$, and $\mathbf{W}_t$ are $np \times 1$ vectors and $\mathbf{A}$ is a $np \times np$ square matrix. The explicit solutions to higher-order VAR processes can be obtained by considering the equivalent VAR(1) process. It can be demonstrated that the reverse characteristic equation of the VAR(1) system has the same roots as those of the original VAR(p) system.

Forecasting with VAR Models

When forecasting with VAR models, a widely used criterion is the minimization of the *mean square error* (MSE). If a process $\mathbf{y}_t$ is generated by a VAR(p) process, the optimal h-step ahead

forecast according to the MSE criterion is the conditional expectation:

$$E_t(\mathbf{y}_{t+h}) \equiv E(\mathbf{y}_{t+h}|\mathbf{y}_s, s \leq t). \tag{3.5}$$

If the error terms are strict white noise, then the optimal forecast of a VAR model reads as follows:

$$E_t(\mathbf{y}_{t+h}) = \mathbf{v} + \mathbf{A}_1 E_t(\mathbf{y}_{t+h-1}) + \cdots + \mathbf{A}_p E_t(\mathbf{y}_{t+h-p}). \tag{3.6}$$

Least Squares Estimation of Stable VAR Models

Stable unrestricted[6] VAR models can be conveniently estimated with multivariate *least squares* (LS) methods.

Considering the np-dimensional VAR(1) representation of the stable unrestricted N-dimensional VAR(p) process:

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{V} + \mathbf{U}_t \tag{3.7}$$

where

$$\mathbf{X}_t = \begin{bmatrix} x_t \\ x_{t-1} \\ \vdots \\ x_{t-p+1} \end{bmatrix}, \mathbf{A} = \begin{bmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \dots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ \mathbf{I}_N & 0 & \dots & 0 & 0 \\ 0 & \mathbf{I}_N & \dots & 0 & 0 \\ 0 & 0 & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & \mathbf{I}_N & 0 \end{bmatrix},$$

$$\mathbf{V} = \begin{bmatrix} \mathbf{v} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \mathbf{U}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}.$$

Matrix **A** is called the *companion matrix* of the VAR(p) system. In the case of a stable unrestricted VAR(p) process, the multivariate GLS estimator coincides with the OLS estimator computed equation by equation.

To display the estimation workflow, following Rachev et al., 2007, we represent the autoregressive process in 3.7 as a single-matrix equation.

Suppose that a sample of T observations of the N-variate variable $\mathbf{x}_t$, $t = 1, \dots, T$ and a presample of p initial conditions $\mathbf{x}_{-p+1}, \dots, \mathbf{x}_0$ are given. The first step consists in stacking all observations $\mathbf{x}_t$, $t = 1, \dots, T$ and noise terms in two separate $NT \times 1$ vectors, to obtain the following compact notations. For all observations $\mathbf{x}_t$:

$$\mathbf{x} = \text{vec}(\mathbf{X}),$$

[6]The models are called "unrestricted" if the estimation process is allowed to determine any possible outcome, and "restricted" if the estimation process restricts parameters in some way.

$$\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_T) = \begin{pmatrix} x_{1,1} & \cdots & x_{1,T} \\ \vdots & \ddots & \vdots \\ x_{N,1} & \cdots & x_{N,T} \end{pmatrix}.$$

$\mathbf{x}$ is a $(NT \times 1)$ vector where all observations are stacked, while $\mathbf{X}$ is a $(N \times T)$ matrix where each column represents an N-variate observation. For the noise terms:

$$\mathbf{u} = \text{vec}(\mathbf{U}),$$

$$\mathbf{U} = \begin{pmatrix} \varepsilon_{1,1} & \cdots & \varepsilon_{1,T} \\ \vdots & \ddots & \vdots \\ \varepsilon_{N,1} & \cdots & \varepsilon_{N,T} \end{pmatrix},$$

where $\mathbf{U}$ is a $(N \times T)$ matrix such that each column represents an n-variate innovation term.

The noise terms are assumed to have a nonsingular covariance matrix, Σ, and the covariance matrix of $\mathbf{u}$, Σ_u, is a block-diagonal matrix where all diagonal blocks are equal to Σ:

$$\Sigma_u = \mathbf{I}_T \otimes \Sigma = \begin{pmatrix} \Sigma & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \Sigma \end{pmatrix}. \tag{3.8}$$

The covariance structure in 3.8 reflects the assumption of white-noise innovations that excludes the possibility of autocorrelations and cross correlations in the innovation terms. The VAR(p) model can now be written compactly in two equivalent ways as follows[7]:

$$\mathbf{X} = \mathbf{AW} + \mathbf{U}, \tag{3.9}$$

$$\mathbf{x} = \mathbf{w}\beta + \mathbf{u}. \tag{3.10}$$

The next step in the estimation workflow consists in writing the weighted sum of squared residuals as:

$$S = \mathbf{u}'\Sigma_u^{-1}u = \sum_{t=1}^{T} \varepsilon_t'\Sigma^{-1}\varepsilon_t. \tag{3.11}$$

For a given set of observations, the quantity S is a function of the model parameters, $S = S(\beta)$, and the least squares estimate of the model parameters, $\hat{\beta}$, are obtained by minimizing $S = S(\beta)$ with respect to beta requiring equating the vector of partial derivatives to zero:

$$\frac{\partial S(\beta)}{\partial \beta} = 0.$$

[7]The long definition of regressor matrix **w** and of the matrices **W, A** and vector **B** are provided in Rachev et al., 2007, p. 347.

Using the expression in 3.9, the matrix $\mathbf{A}$ is estimated as:

$$\hat{\mathbf{A}} = \mathbf{X}\mathbf{W}'(\mathbf{W}\mathbf{W}')^{-1}. \tag{3.12}$$

The relationship between the estimated matrix $\hat{\mathbf{A}}$ and the parameter estimates, $\hat{\beta}$, is as follows:

$$\hat{\beta} = ((\mathbf{W}\mathbf{W}')^{-1}\mathbf{W} \otimes \mathbf{I}_N)\mathbf{x}, \tag{3.13}$$

$$\begin{aligned} \text{vec}\hat{\mathbf{A}} &= ((\mathbf{W}\mathbf{W}')^{-1}\mathbf{W} \otimes \mathbf{I}_N)\text{vec}(\mathbf{X}) \\ &= \text{vec}(\mathbf{X}\mathbf{W}'(\mathbf{W}\mathbf{W}')^{-1}). \end{aligned} \tag{3.14}$$

Estimating the Number of Lags

Assuming that the type of model is correctly specified and that it is a VAR(p) model, it is important to determine correctly the order p of the model (i.e., the number of lags of the model). This importance stems from the fact that, increasing the model order reduces the size of residuals but tends also to reduce the forecasting ability of the model. By increasing the number of parameters, the in-sample accuracy increases at the cost of worse out-of-sample forecasting ability[8].

We determine the correct number of lags in the VAR models in a systematic manner, as follows. We start with a maximum lag length $m = 12$ (because of the monthly frequency of the data). We then run the VAR model in level for lag length (1:12) and calculate the Lütkepohl's version of Akaike Information Criterion (AIC) given by:

$$AIC = ln(|\Sigma_u|) + \frac{2pK^2}{T}. \tag{3.15}$$

where Σ_u is the white-noise covariance matrix, p is the order of the VAR process fitted to the data, K is the dimension of the time series and T is the sample size. We chose the lag length that minimizes the Lütkepohl's version of AIC. Next, we confirm that for the selected lag length the residuals of the VAR model are not correlated. The decision is made based on the *multivariate Portmanteau* and *Breusch-Godfrey test for serially correlated errors*[9]. We may have to modify the lag length, if there is autocorrelation.

3.3 Structural Analysis with VAR Models: Granger Causality

Multivariate time series models capture the co-movement and dependencies between several time series variables over time. Employing VAR models, useful information can be obtained about the nature of interactions among the variables and the way they influence each other.

[8]For a comprehensive explanation of this intuition, we refer the reader to Lütkepohl, 2005, p. 146 and Rachev et al., 2007, p. 357.

[9]Box and Pierce, 1970; Ljung and Box, 1978; Castle and Hendry, 2010.

Granger, 1969 has defined a concept of causality, which can be employed to draw conclusions about the causal direction with which the variables influence each other. The idea behind the concept of Granger causality is that a cause cannot come after the effect. Thus, if a variable X affects a variable Y, the former should help improving the predictions of the latter variable.

Underlying Idea of Granger Causality

The underlying idea of the concept of Granger Causality can be formalized as follows. Let Z_t be a weakly stationary stochastic process and define the set of Z_t's up to and including period t by

$$Z_t^- = \{Z_{t-i} | i = 0, 1, \dots\}.$$

The expression $A_t^- \ B_t^-$ refers to the set-theoretic subtraction. Let

$$P(A_t | B_t^-)$$

denote the conditional best, unbiased least-squares predictor of A_t utilizing all the information contained in B_t^-; and let $\sigma^2(A_t^- | B_t^-)$ denote the corresponding prediction-error variance, i.e.,

$$\sigma^2(A_t | B_t^-) = Var\left[A_t - P(A_t | B_t^-)\right]$$

Assuming that we want to predict A_t using the information set $\mathcal{I}_{t-1}$[10] then:

$$P(A_t | \mathcal{I}_{t-1})$$

denotes the best one-step-ahead predictor for A_t in the sense defined above, and

$$\sigma^2(A_t | \mathcal{I}_{t-1}) \tag{3.16}$$

is the corresponding one-step-ahead predictor-error variance[11]. With this notation, the following definitions are established:

Definition 1. *X Granger-causes Y, if*

$$\sigma^2(Y_t | \mathcal{I}_{t-1}) < \sigma^2(Y_t | \mathcal{I}_{t-1} \backslash X_{t-1}^-)$$

Definition 2. *There exists instantaneous Granger-causation between X and Y, if*

$$\sigma^2(Y_t | \mathcal{I}_t \backslash Y_t) < \sigma^2(Y_t | \mathcal{I}_t \backslash Y_t \backslash X_t^-)$$

Definition 3. *There exists feedback between X and Y in the Granger sense, if X Granger-causes Y and Y Granger-causes X.*

Definition 4. *X and Y are independent in the Granger-sense if neither Granger-causes the other.*

[10] $\mathcal{I}_{t-1}$ denotes a set containing information accumulated up to and including time $t-1$.

[11] The predictor $P(A_t | B_t^- \ A_t)$ uses contemporaneous information for predicting A_t, while $P(A_t | B_{t-1}^-)$ does not.

First Suggestions of Granger Causality: Sample Cross Correlation Function

The *cross-correlation function* (CCF) is the data analysis tool, which can be employed to derive first suggestions on the existence of Granger causality and lead-lag relationships among given variables.

Given two time series, x_t and y_t, the sample cross-correlation function can be used to determine whether the series x_t may be related to past lags of the y_t-series. Therefore, the sample cross correlation function is helpful for identifying lags of the y_t-variable that might be useful predictors of x_t and, hence, for identifying potential Granger causality between *yield curve drivers*. Following Box et al., 2015, p. 474, the sample cross covariance function is an estimate of the covariance between two time series, x_t and y_t, at lags $k = 0, \pm 1, \pm 2, ...$
For data pairs $(x_1, y_1), (x_2, y_2), ..., (x_T, y_T)$, an estimate of the lag k cross-covariance is:

$$c_{xy}(k) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \bar{x})(y_{t+k} - \bar{y}); & \text{if } k = 0, 1, 2, ... \\ \frac{1}{T}\sum_{t=1}^{T+k}(y_t - \bar{y})(x_{t-k} - \bar{x}); & \text{if } k = 0, -1, -2, ... \end{cases}$$

where $\bar{x}$ and $\bar{y}$ are the sample means of the series. Let the *sample standard deviations* of the series be denoted as

$$s_x = \sqrt{c_{xx}(0)}, \qquad \text{where} \qquad c_{xx}(0) = Var(x),$$

$$s_y = \sqrt{c_{yy}(0)}, \qquad \text{where} \qquad c_{yy}(0) = Var(y).$$

Then an estimate of the *cross correlation* is:

$$r_{xy}(k) = \frac{c_{xy}(k)}{s_x s_y} \qquad \text{with} \qquad k = 0, \pm 1, \pm 2, ...$$

Granger Test and Types of Granger Causality

The *Granger test* is a test based on a truncated version of the infinite AR representation of the bivariate process. The order at which the autoregression is truncated can be determined by an information criterion, such as AIC or BIC, and should be sufficiently large, so that the residuals resemble white noise. We determine the truncation order as specified in Section 3.2.

The null hypothesis of the Granger test is that x_t does not Granger-cause y_t. Testing this hypothesis requires the estimation of the regression

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \epsilon_t \tag{3.17}$$

A test of the hypothesis that x_t does not Granger-cause y_t amounts to testing the null hypothesis

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0,$$

which can be done by an F-test. The F-test will compare the unrestricted regression of y_t against the restricted regression. Depending on what kind of realizations of x_t are included in the equation of y_t, one can distinguish 3 types of Granger causality:

1. No delayed Granger causality;
2. No causality at all, i.e., no instantaneous and no delayed Granger causality;
3. No delayed but instantaneous Granger causality.

Type 1: No delayed Granger causality

A test for *no delayed Granger causality* from x_t to y_t requires the inclusion of lagged realizations of x_t in the unrestricted regression of y_t (3.17), i.e.,

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \epsilon_t$$

A test of the hypothesis that there is no delayed causality from x_t to y_t amounts to testing the null hypothesis

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_p = 0$$

Thus, the no delayed Granger causality test consists of the following steps:

1. Estimate the unrestricted regression

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=1}^{p} \beta_i x_{t-i} + \epsilon_t$$

   using ordinary least squares, where lag length p is sufficiently large, so that the estimated residuals, $\hat{\epsilon}_t$, $t = 1, \ldots, T$, resemble white noise.

2. Using again ordinary least squares, estimate the restricted regression

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \epsilon_{R,t} \tag{3.18}$$

3. Compute the $F-$statistic

$$F = \frac{T-k-1}{q} \frac{\sum_{t=1}^{T} (\epsilon_{R,t}^2 - \epsilon_t^2)}{\sum_{t=1}^{T} \epsilon_t^2} \sim F_{q,T-k-1} \tag{3.19}$$

4. Reject the null hypothesis of no delayed Granger causality from x_t to y_t, if the computed value of the $F-$statistic[12] exceeds the critical value of the $F-$distribution with q and $T-k-1$ degrees of freedom.

Type 2: No causality at all: no instantaneous and no delayed Granger causality

A test for *no causality at all, i.e., no instantaneous and no delayed Granger causality* from x_t to y_t requires the inclusion of contemporaneous realizations of x_t in the unrestricted regression of y_t (3.17), i.e.,

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=0}^{p} \beta_i x_{t-i} + \epsilon_t \tag{3.20}$$

while the restricted regression (3.18) remains as in Type 1. The null hypothesis now becomes

$$H_0 : \beta_0 = \beta_1 = \cdots = \beta_p = 0;$$

and the critical value for the F-statistic (3.19) has to be taken from the $F-$distribution with $p+1$ and $T-2p-2$ degrees of freedom.

Type 3: No delayed but instantaneous Granger-causation

A test for *no delayed but instantaneous Granger causality* from x_t to y_t requires the inclusion of contemporaneous realizations of x_t in the unrestricted regression of y_t (3.17), i.e.,

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \sum_{i=0}^{p} \beta_i x_{t-i} + \epsilon_t \tag{3.21}$$

while in the restricted regression (3.18) we include also the *contemporaneous realization* of x_t

$$y_t = c + \sum_{i=1}^{p} \alpha_i y_{t-i} + \beta_0 x_t + \epsilon_{R,t} \tag{3.22}$$

The null hypothesis now becomes

$$H_0 : \beta_0 \neq 0, \beta_1 = \beta_2 = \cdots = \beta_p = 0;$$

and the critical value for the F-statistic (3.19) has to be taken from the $F-$distribution with p and $T-2p+1$ degrees of freedom.

[12]The F-test for multiple linear restrictions is a test of whether or not a group of variables has an effect on y, meaning that we are testing whether these variables are jointly significant. In equation 3.19, q is the number of restrictions (i.e., the number of independent variables that are dropped), T is the number of observations, and k is the number of independent variables.

3.4 Structural Analysis with VAR Models: Impulse Response Analysis

The concept of causality can be studied further by quantifying the influence and the temporal profile of a change of one variable on the other variable in the system. It is of our interest to know the response of one yield curve driver to an impulse in another yield curve driver, in order to gain deeper understanding of the dynamic interrelationships within a system and determine whether one yield curve driver is causal for another one.

Responses to Forecast Errors

Impulse response functions and variance decompositions indicate how the endogenous variables respond to external influences. In a VAR model, where all variables are endogenous, the only external inputs are the disturbances, which amount to one-step prediction errors. They are "surprises", "shocks", or "innovations" which cannot be explained by the model and past data. Considering the VAR(1) representation

$$Y_t = \nu + \mathbf{A}Y_{t-1} + U_t$$

of the higher order VAR(p) process. Under the stability assumption, the process $\mathbf{Y}_t$ has a *moving average (MA) representation*, where $\mathbf{Y}_t$ is expressed in terms of past and present error or innovation vectors $\mathbf{U}_t$ and the mean term $\boldsymbol{\mu}$:

$$Y_t = \boldsymbol{\mu} + \sum_{i=0}^{\infty} \mathbf{A}^i U_{t-i}. \tag{3.23}$$

The moving average representation captures the responses of the Y-variables with respect to the prediction errors and the impulse responses are the elements of the upper left-hand $(K \times K)$ block of $\mathbf{A}^i$. Furthermore, given that the MA representation can be found by premultiplying 3.23 by the $(K \times Kp)$ matrix $J := [I_K : 0 : \cdots : 0]$[13]:

$$\begin{aligned} y_t = JY_t &= J\boldsymbol{\mu} + \sum_{i=0}^{\infty} J\mathbf{A}^i J' J U_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i} \end{aligned} \tag{3.24}$$

where $\mu := J\boldsymbol{\mu}$, $\Phi_i := J\mathbf{A}^i J'$ and, due to the special structure of the white noise process U_t, we have $U_t = J'JU_t$ and $JU_t = u_t$. The matrix $\mathbf{A}^i$ can be shown to be the i-th coefficient matrix Φ_i of the MA representation in 3.24. The jk-th element of Φ_i, $\phi_{jk,i}$, represents the reaction of

[13] Lütkepohl, 2005, p. 18

the j-th variable of the system to a unit shock in variable k, i periods ago, provided the effect is not contaminated by other shocks to the system.[14]

If the variables have different scales, it is sometimes convenient to consider innovations of one standard deviation rather than unit shocks. Since the average size of the innovations occurring in a system depends on their standard deviation, a rescaling of the impulse responses may sometimes give a better picture of the dynamic relationships.

Following Proposition 2.2 of Lütkepohl, 2005, the impulse responses are zero if one of the variables does not Granger-cause the other variables taken as a group.

Accumulated Responses

If we are interested in quantifying the accumulated effect over several or more periods of a shock in one variable, we need to sum up the MA coefficient matrices[15]. The $k-$th column of $\Psi_n := \sum_{i=0}^{n} \Phi_i$ contains the *accumulated responses* over n periods to a unit shock in the $k-$th variable of the system and these quantities are sometimes called $n-$th *interim multipliers*. The total accumulated effects for all future periods are obtained by summing up all the MA coefficient matrices. $\Psi_\infty := \sum_{i=0}^{\infty} \Phi_i$ is sometimes called the matrix of *long-run effects* or *total multipliers*. Because the MA operator $\Phi(z)$ is the inverse of the VAR operator $A(z) = I_k - A_1 z - \cdots - A_p z^p$, the long-run effects are easily obtained as

$$\Psi_\infty = \Phi(1) = (I_K - A_1 - \cdots - A_p)^{-1}. \tag{3.25}$$

Responses to Orthogonal Impulses

The impulse responses to forecast errors assume that a shock occurs only in one variable at a time. This assumption is violated in cases where shocks in different variables are not independent and/or the error terms are correlated. In these cases, a shock in one variable is likely to be accompanied by a shock in another variable, hence, it is reasonable to perform the impulse response analysis in terms of the MA representation:

$$y_t = \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \tag{3.26}$$

where the components of $w_t = (w_{1t}, \ldots, w_{Kt})'$ are uncorrelated and have unit variance, $\Sigma_w = I_K$. Recalling from Lütkepohl, 2005, the representation in 3.26 is obtained by decomposing Σ_u as $\Sigma_u = PP'$, where P is a lower triangular matrix, and defining $\Theta_i = P^{-1}u_t$. Within this representation, it is reasonable to assume that a change in one component of w_t has no effect on the other components because the components are orthogonal (uncorrelated). The jk-th element of Θ_i is assumed to represent the effect on variable j of a unit innovation in the k-th

[14] Lütkepohl, 2005, p. 52. Because the u_t are just the one-step ahead forecast errors of the VAR process, the shocks may be regarded as forecast errors and the impulse responses are sometimes referred to as *forecast error impulse responses*.

[15] Lütkepohl, 2005, p. 55

variable that has occurred i periods ago. Further on, Lütkepohl, 2005 shows how to relate these impulse responses to a VAR model.

3.5 Vector Error Correction (VEC) Models: Assumptions, Properties, and Estimation Methods

Cointegration: Definition and Key Features

Many economic and financial time series tend to exhibit nonstationary behavior and very often it is necessary to have models that accommodate the nonstationary features of the data, especially when the modeler is interested in analyzing the original variables rather than the rates of change. The idea behind cointegration is that there are feedback mechanisms that force nonstationary processes to stay close together. In the sequel, we review the cointegration theory. The exposition follows Rachev et al., 2007.

The concept of cointegration was introduced by Granger, 1981 and can be intuitively characterized in terms of its three key features:

- *Reduction of order of cointegration*, in the sense that, cointegration is a property of processes integrated of order one that admit linear combinations integrated of order zero (stationary). Formally, suppose that n time series $x_{i,t}$, integrated of the same order d are given. If there is a linear combination of the series

$$\delta_t = \sum_{i=1}^{n} \beta_i x_{i,t} \tag{3.27}$$

  that is integrated of order $e < d$, then the series are said to be cointegrated and such a linear combination is called a *cointegrating relationship*. Cointegrated processes are characterized by a short-term dynamics and a long-run equilibrium, which is the relationship between the processes after eliminating the short-term dynamics[16]. Generally, there can be many linearly independent cointegrating relationships. Given n processes integrated of order one, there can be a maximum of $n - 1$ cointegrating relationships. The cointgeration vectors $[\beta_i]$ are not unique. In fact, given two cointegrating vectors $[\alpha_i]$ and $[\beta_i]$ such that

$$\sum_{i=1}^{n} \alpha_i X_i, \qquad \sum_{i=1}^{n} \beta_i X_i \tag{3.28}$$

  are integrated of order e, any linear combination of the cointegrating vectors is another cointegrating vector as the linear combination

$$A \sum_{i=1}^{n} \alpha_i X_i + B \sum_{i=1}^{n} \beta_i X_i \tag{3.29}$$

[16] In this sense, the long-run equilibrium denotes the static regression function and thus, it does not mean that the cointegrated processes tend to a long-run equilibrium.

is integrated of order e.

- *Linear regression*, in the sense that, two or more processes integrated of order one are said to be cointegrated if it is possible to make meaningful linear regression of one process on the other(s).
- *Common trends*: given n time series $x_{i,t}, i = 1, \dots, n$ with $k < n$ cointegrating relationships, it is possible to determine $n - k$ integrated time series $u_{j,t}, j = 1, \dots, n - k$, called *common trends*, such that any of the n original processes can be expressed as a linear regression on the common trends[17] plus a stationary disturbance:

$$x_{i,t} = \sum_{j=1}^{n-k} \gamma_j u_{j,t} + \eta_{i,t}. \tag{3.30}$$

Following the original work of Stock and Watson, 1988, Rachev et al., 2007 show how, in a set of cointegrated processes, each process can be expressed in terms of a reduced number of common stochastic trends.

Error Correction Models

A multivariate integrated process is cointegrated if and only if it can be represented in the *error correction model* (ECM) form (or *vector error correction* (VEC) model) with appropriate restrictions. Adding the error-correction term to a VAR model in differences produces the error-correction form:

$$\Delta \mathbf{x}_t = (\mathbf{\Phi}_1 L + \mathbf{\Phi}_2 L^2 + \cdots + \mathbf{\Phi}_{p-1} L^{p-1}) \Delta \mathbf{x}_t + \mathbf{\Pi} L^p \mathbf{x}_t + D\mathbf{s}_t + \boldsymbol{\varepsilon}_t \tag{3.31}$$

where the $p - 1$ terms are in first differences and the last term is in levels. The term in levels can be placed at any lag. Cointegration is then expressed as restrictions on the matrix Π, as follows:

$$\Delta \mathbf{x}_t T = \left(\sum_{i=1}^{P-1} \mathbf{A} L^i \right) \Delta \mathbf{x}_{t-1} + \boldsymbol{\alpha} \boldsymbol{\beta}' \mathbf{x}_{t-1} + \boldsymbol{\varepsilon}_{\boldsymbol{t}} \tag{3.32}$$

where $\boldsymbol{\alpha}$ is an $n \times r$ matrix, $\boldsymbol{\beta}$ is an $n \times r$ matrix with $\boldsymbol{\alpha}\boldsymbol{\beta}' = \mathbf{\Pi}$. In the ECM representation in 3.32, $\boldsymbol{\beta}$ is a cointegrating vector, the combination $\boldsymbol{\beta}' x_{t-1}$ reflects common trends and measures the "error" in the data (i.e., the deviation from the stationary mean) at time $t - 1$, while $\boldsymbol{\alpha}$ is the vector containing the loading factors of the common trends. $\boldsymbol{\alpha}$ can be viewed as the vector of adjustment speeds (i.e., the rate at which the series "correct" from disequilibrium). The combination $\boldsymbol{\alpha}\boldsymbol{\beta}' x_{t-1}$ is the error-correction term. If $r = 0$, there is no common trend and no cointegration exists between the processes; if $r = n$, the processes are stationary; if $n > r > 0$, processes are integrated and there are cointegrating relationships.

[17] Cointegration classifies as a dimensionality reduction technique, as the common trends are the common drivers of a set of processes. Common trends as integrated processes were first discussed by Stock and Watson, 1989 and Stock and Watson, 1998.

Maximum Likelihood Estimation of Cointegrated VAR Models

When it comes to estimating nonstationary and nonstable processes (i.e., processes, in which the averages, variances, or covariances may vary with time), the maximum likelihood (ML) procedure represents the state-of-the-art estimation method. In the sequel, we recall the ML methodology following the exposition of Rachev et al., 2007. Rachev et al. describe the ML estimation methodology for cointegrated processes as introduced by Banerjee and Hendry, 1992 to then connect with the original *reduced rank regression* method of Johansen, 1991. Writing the cointegrated VAR using the ECM formulation as follows:

$$\Delta \mathbf{x}_t = -\mathbf{\Pi} \mathbf{x}_{t-1} + \mathbf{F}_1 \Delta \mathbf{x}_{t-1} + \mathbf{F}_2 \Delta \mathbf{x}_{t-2} + \cdots + \mathbf{F}_{p-1} \Delta \mathbf{x}_{t-p+1} + \boldsymbol{\varepsilon}_t, \tag{3.33}$$

where the innovations are assumed to be independent identically distributed (IID) multivariate, correlated, Gaussian variables. The method of Banerjee and Hendry uses the mathematical technique of *concentrated likelihood* to transform the original likelihood function (LF) into a function of a smaller number of variables, called the *concentrated likelihood function* (CLF). With respect to the process in 3.33, define

$$\mathbf{X} = (\mathbf{x}_0, \ldots, \mathbf{x}_{T-1})$$

$$\Delta \mathbf{x}_t = \begin{pmatrix} \Delta x_{1,t} \\ \vdots \\ \Delta x_{n,t} \end{pmatrix}$$

$$\Delta \mathbf{X} = (\Delta \mathbf{x}_1, \ldots, \Delta \mathbf{x}_T) = \begin{pmatrix} \Delta x_{1,1} & \cdots & \Delta x_{1,T} \\ \vdots & \ddots & \vdots \\ \Delta x_{n,1} & \cdots & \Delta x_{n,T} \end{pmatrix}$$

$$\Delta \mathbf{Z} = \begin{pmatrix} \Delta \mathbf{x}_t \\ \vdots \\ \Delta \mathbf{x}_{t-p+2} \end{pmatrix},$$

$$\Delta \mathbf{Z} = \begin{pmatrix} \Delta \mathbf{x}_0 & \cdots & \Delta \mathbf{x}_{T-1} \\ \vdots & \ddots & \vdots \\ \Delta \mathbf{x}_{-p+2} & \cdots & \Delta \mathbf{x}_{T-p+1} \end{pmatrix} = \begin{pmatrix} \Delta x_{1,0} & \cdots & \Delta x_{1,T-1} \\ \vdots & \ddots & \vdots \\ \Delta x_{n,0} & \cdots & \Delta x_{n,T-1} \\ \vdots & \ddots & \vdots \\ \Delta x_{1,-p+2} & \cdots & \Delta x_{1,T} \\ \vdots & \ddots & \vdots \\ \Delta x_{n,-p+2} & \cdots & \Delta x_{n,T} \end{pmatrix}$$

$$\mathbf{F} = (\mathbf{F}_1, \mathbf{F}_2, \ldots, \mathbf{F}_{p-1}).$$

Assuming $\mathbf{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ and using matrix notation, the model can be written in compact form as follows:

$$\Delta\mathbf{X} = \mathbf{F}\Delta\mathbf{Z} - \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X} + \mathbf{U}. \tag{3.34}$$

Rachev et al. show that the log likelihood function is then given by:

$$\begin{aligned}\log(l) = &-\frac{nT}{2}\log(2\pi) - \frac{T}{2}\log(|\mathbf{\Sigma}_u|) \\ &-\frac{1}{2}\text{trace}((\Delta\mathbf{X} - \mathbf{F}\mathbf{Z} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})'\mathbf{\Sigma}_u^{-1}(\Delta\mathbf{X} - \mathbf{F}\mathbf{Z} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})\end{aligned} \tag{3.35}$$

and that the concentrated likelihood after removing $\mathbf{\Sigma}$ is:

$$\begin{aligned}l^{CL} &= K - \frac{T}{2}\log|\mathbf{U}\mathbf{U}'| \\ &= K - \frac{T}{2}\log|(\Delta\mathbf{X} - \mathbf{F}\mathbf{Z} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})(\Delta\mathbf{X} - \mathbf{F}\mathbf{Z} + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})'|\end{aligned} \tag{3.36}$$

where K is a constant that includes all the constant terms left after concentrating. The next step consists in eliminating the $\mathbf{F}$ terms to reach the following log-likelihood function:

$$l^{CP} = K - \frac{T}{2}\log|\Delta\mathbf{X}\mathbf{M}\Delta\mathbf{X}' + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}\mathbf{M}\Delta\mathbf{X}' + \Delta\mathbf{X}\mathbf{M}(\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})' + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X}\mathbf{M}(\boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{X})'| \tag{3.37}$$

where $\mathbf{M} = \mathbf{I}_T - \Delta\mathbf{Z}'(\Delta\mathbf{Z}\Delta\mathbf{Z}')^{-1}\Delta\mathbf{Z}$. Defining $\mathbf{R}_0 = \mathbf{\Delta}\mathbf{X}\mathbf{M}, \mathbf{R}_1 = \mathbf{X}\mathbf{M}$ and

$$\mathbf{S}_{ij} = \frac{\mathbf{R}_i\mathbf{R}_j}{T}, i, j = 1, 2$$

The CLF can be rewritten as follows:

$$l^{C\Pi}(\boldsymbol{\alpha}\boldsymbol{\beta}') = K - \frac{T}{2}\log|\mathbf{S}_{00} - \mathbf{S}_{10}\boldsymbol{\alpha}\boldsymbol{\beta}' - \mathbf{S}_{01}(\boldsymbol{\alpha}\boldsymbol{\beta}')' + \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{S}_{11}(\boldsymbol{\alpha}\boldsymbol{\beta}')'|. \tag{3.38}$$

The results following the Johansen method can be obtained by applying the method of *reduced rank regression*. The Johansen method eliminates the terms $\mathbf{F}$ by regressing $\Delta\mathbf{x}_t$ and $\Delta\mathbf{x}_{t-1}$ on $(\Delta\mathbf{x}_{t-1}, \Delta\mathbf{x}_{t-2}, \ldots, \Delta\mathbf{x}_{t-p+1})$ to obtain the following residuals:

$$\mathbf{R}_{0t} = \Delta\mathbf{x}_t + \mathbf{D}_1\Delta\mathbf{x}_{t-1} + \mathbf{D}_2\Delta\mathbf{x}_{t-2} + \cdots + \mathbf{D}_{p-1}\Delta\mathbf{x}_{t-p+1} \tag{3.39}$$

$$\mathbf{R}_{1t} = \Delta\mathbf{x}_{t-1} + \mathbf{E}_1\Delta\mathbf{x}_{t-1} + \mathbf{E}_2\Delta\mathbf{x}_{t-2} + \cdots + \mathbf{E}_{p-1}\Delta\mathbf{x}_{t-p+1} \tag{3.40}$$

where

$$\mathbf{D} = (\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_{p-1}) = \mathbf{\Delta}\mathbf{X}\mathbf{\Delta}\mathbf{Z}'(\mathbf{\Delta}\mathbf{Z}\Delta\mathbf{Z}')^{-1}$$

and

$$\mathbf{E} = (\mathbf{E}_1, \mathbf{E}_2, \ldots, \mathbf{E}_{p-1}) = \mathbf{X}\boldsymbol{\Delta}\mathbf{Z}'(\boldsymbol{\Delta}\mathbf{Z}\boldsymbol{\Delta}\mathbf{Z}')^{-1}.$$

The original model is reduced to a simpler model:

$$\mathbf{R}_{0t} = \boldsymbol{\alpha}\boldsymbol{\beta}'\mathbf{R}_{1t} + \mathbf{u}_t \tag{3.41}$$

with the following likelihood function:

$$l(\boldsymbol{\alpha}\boldsymbol{\beta}') = K_1 - \frac{T}{2}\log|(\mathbf{R}_0 + \mathbf{R}_1(\boldsymbol{\alpha}\boldsymbol{\beta}'))'(\mathbf{R}_0 + \mathbf{R}_1(\boldsymbol{\alpha}\boldsymbol{\beta}'))|. \tag{3.42}$$

The CLF of the likelihood function in 3.42 is the same as the one in 3.38. The maximum of the log-likelihood function of the Johansen method is:

$$l_{max} = K - \frac{T}{2}\log|\mathbf{S}_{00} - \frac{T}{2}\sum_{i=1}^{r}\log(1-\lambda_i)| \tag{3.43}$$

where the eigenvalues λ_i can be interpreted as the canonical correlations between $\Delta\mathbf{x}_t$ and $\Delta\mathbf{x}_{t-1}$.

Estimating the Number of Cointegrating Relationships: The Johansen Test For Cointegration

The Johansen ML estimation method depends on correctly estimating the cointegration rank, i.e., the number r of cointegrating relationships. At the core of the Johansen method is the relationship between the *impact matrix*, $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, and the size of its eigenvalues. The eigenvalues depend on the composition of the deterministic terms of the VEC model. The Johansen method incorporates the testing procedure into the process of model estimation and, in doing so, it avoids conditional estimates. The method, therefore, first infers the cointegration rank by testing the number of eigenvalues that are statistically different from zero, then conducts model estimation under the rank constraints.

The cointegration rank can be determined using the *trace test*, which assesses which eigenvalues correspond to stationary and which to non-stationary relations. A small eigenvalue indicates a unit root and thus a very persistent and possibly non-stationary process. Formally, the trace test assesses the null hypothesis of $H(r)$ of cointegration rank less than or equal to r against the alternative hypothesis $H(k)$, where k is the dimension of the data. The test reads as follows:

$$\lambda_{trace} = -T[\log(1-\lambda_{T+1}) + \cdots + \log(1-\lambda_K)]. \tag{3.44}$$

Estimating the Number of Lags

The weakness of the Johansen approach is that it is sensitive to the lag length. Therefore, before testing for the cointegration rank, the optimal lag structure for the VEC(q) model has to be

selected. We do so by exploiting the fact that, by collecting the first differences, a VEC(q) model can be converted to a VAR(p) model in levels, with $p = q+1$. We then follow the same systematic approach described in 3.2, with the additional step that the lag length for the VEC(q) model will equal the lag length chosen for the equivalent VAR(p) model minus one (p=q+1), since, for cointegration testing, we are running the model in first difference, and, hence, lose one lag.

3.6 Structural Analysis with VEC Models: Granger Causality

The restrictions characterizing Granger noncausality in cointegrated systems are the same as in the stable case (Lütkepohl, 2005). More specifically, if we consider a VAR(p) model in levels as the representation of the data generation process,

$$\mathbf{y}_t = \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{u}_t$$

and the vector $\mathbf{y}_t$ is partitioned in M- and $(K-M)$-dimensional subvectors $\mathbf{z}_t$ and $\mathbf{x}_t$,

$$\mathbf{y}_t = \begin{bmatrix} \mathbf{z}_t \\ \mathbf{x}_t \end{bmatrix}$$

and

$$\mathbf{A}_i = \begin{bmatrix} \mathbf{A}_{11,i} & \mathbf{A}_{12,i} \\ \mathbf{A}_{21,i} & \mathbf{A}_{22,i} \end{bmatrix}, \qquad i = 1, \ldots, p,$$

where the $\mathbf{A}_i$ are partitioned in accordance with the partitioning of $\mathbf{y}_t$, then $\mathbf{x}_t$ does not Granger-cause $\mathbf{z}_t$ if and only if the hypothesis

$$H_0 : \mathbf{A}_{12,i} = 0 \qquad \text{for} \qquad i = 1, \ldots, p,$$

is true. If we then consider an ECM representation of the subvectors $\mathbf{z}_t$ and $\mathbf{x}_t$:

$$\begin{bmatrix} \Delta\mathbf{z}_t \\ \Delta\mathbf{x}_t \end{bmatrix} = \begin{bmatrix} \mathbf{\Pi_{11}} & \mathbf{\Pi_{12}} \\ \mathbf{\Pi_{21}} & \mathbf{\Pi_{22}} \end{bmatrix} \begin{bmatrix} \mathbf{z}_{t-1} \\ \mathbf{x}_{t-1} \end{bmatrix} + \sum_{i=1}^{p-1} \begin{bmatrix} \mathbf{F}_{11,i} & \mathbf{F}_{12,i} \\ \mathbf{F}_{21,i} & \mathbf{F}_{22,i} \end{bmatrix} \begin{bmatrix} \Delta\mathbf{z}_{t-i} \\ \Delta\mathbf{x}_{t-i} \end{bmatrix} + \mathbf{u}_t \tag{3.45}$$

then the Granger noncausality can be characterized as:

$$\mathbf{\Pi}_{12} = 0 \qquad \text{and} \qquad \mathbf{F}_{12} = 0 \qquad \text{for} \qquad i = 1, \ldots, p-1,$$

meaning that to check for Granger causality, one has to just test a set of linear hypotheses. A standard Wald test is suitable for this purpose only when the model restrictions are correctly specified. In case of misspecified restrictions (for example, in the case of a misspecified cointegration rank or misspecified restrictions in the estimation procedure) and/or in the case where (some of) the data are non-stationary, the Wald test might not follow its asymptotic chi-square

distribution under the null hypothesis (Andrews, 1987; Toda and Phillips, 1993). A possible solution is to rewrite the VEC model in such a way that all parameters under test are attached to stationary regressors. The procedure would then ultimately consists in performing the *Wald test based on a Lag Augmented VAR*, equivalently, follow the Toda-Yamamoto (TY) (Toda and Yamamoto, 1995) approach to Granger noncausality. This approach can be summarized as follows. Given two time-series variables, **X** and **Y**, the most simple definition of Granger causality states that **X** Granger-causes **Y** if the prediction of **Y** can be improved more using the histories of both **X** and **Y** than using only the history of **Y**. The absence of Granger causality can then be tested by estimating the following VAR model:

$$\mathbf{Y}_t = \boldsymbol{\alpha}_0 + \boldsymbol{\alpha}_1 \mathbf{Y}_{t-1} + \cdots + \boldsymbol{\alpha}_p \mathbf{Y}_{t-p} + \boldsymbol{\beta}_1 \mathbf{X}_{t-1} + \cdots + \boldsymbol{\beta}_p \mathbf{X}_{t-p} + \mathbf{u}_t \tag{3.46}$$

$$\mathbf{X}_t = \boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1 \mathbf{X}_{t-1} + \cdots + \boldsymbol{\gamma}_p \mathbf{X}_{t-p} + \boldsymbol{\delta}_1 \mathbf{Y}_{t-1} + \cdots + \boldsymbol{\delta}_p \mathbf{Y}_{t-p} + \mathbf{v}_t \tag{3.47}$$

where the null hypothesis $H_0 : \boldsymbol{\beta}_1 = \boldsymbol{\beta}_2 = \cdots = \boldsymbol{\beta}_p = 0$, against H_A : Not H_0 tests that **X** *does not Granger-cause* **Y**. And the null hypothesis $H_0 : \boldsymbol{\delta}_1 = \boldsymbol{\delta}_2 = \cdots = \boldsymbol{\delta}_p = 0$, against H_A : Not H_0 tests that **Y** *does not Granger-cause* **X**.

Within this setup, the Toda-Yamamoto procedure would start with performing ADF and KPSS tests to determine the order of integration of each time series and set the maximum order of integration for the group time series be m. Next, a VAR model in the levels of the data is set up, regardless of the integration of the various time series. The maximum lag length, p, for the variables in the VAR is determined using the usual methods, for example, the method described in section 3.2. The modeler has to make sure that the VAR is well-specified, by checking, for example, the absence of serial correlation in the residuals[18]. If two or more of the time series have the same order of integration (as resulting from the ADF and KPSS tests), a test for cointegration is necessary, preferably using the Johansen method described in section 3.5. The result of the cointegration test serves as a possible cross-check on the validity of the results at the end of the Toda-Yamamoto procedure. Having determined the maximum lag length of the VAR model and ensured that the model is well-specified, the Toda-Yamamoto procedure requires augmenting the VAR model with the m additional lags of each of the variables into each of the equations. The test for Granger noncausality would go equation by equation in the VAR model and test the hypothesis that the coefficients of (only) the first p lagged values of **X** are zero in the **Y** equation (i.e., Granger noncausality from **X** to **Y**) and that the coefficients of (only) the first p lagged values of **Y** are zero in the **X** equation (i.e., Granger noncausality from **Y** to **X**). This tests are standard Wald tests. It is essential that the modeler does not include the coefficients for the "extra" m lags when she performs the Wald tests. The "extra" m lags are only included to fix the asymptotic properties of the Wald test statistics, which will then be asymptotically chi-square distributed with p degrees of freedom, under the null. Rejecting the null implies rejecting Granger noncausality and, thus, concluding that there is Granger causality.

Finally, reconciling with the cointegration results, *"if two or more time-series are cointegrated, then there must be Granger causality between them - either one-way or in both directions. However,*

[18]If serial correlation is present, the lag length, p of the VAR model might need to be increased.

the converse is not true.". Based on these results, the modeler can deduce whether there is a conflict in her results. The presence of cointegration and the absence of causality suggest a conflict in the results[19].

Lütkepohl and Reimers, 1992, show, however, that for bivariate processes with cointegration rank 1, no extra lag is needed if both variables are I(1).

3.7 Structural Analysis with VEC Models: Impulse Response Analysis

Integrated and cointegrated systems must be interpreted cautiously. In cointegrated systems the term $\boldsymbol{\beta}'\mathbf{y}_t$ is usually thought as representing the *long-run equilibrium relations* between the variables. If we suppose there is just one such relation, say

$$\boldsymbol{\beta}_1\mathbf{y}_{1t} + \cdots + \boldsymbol{\beta}_K\mathbf{y}_{Kt} = 0,$$

or, if $\boldsymbol{\beta}_1 \neq 0$,

$$\mathbf{y}_{1t} = -\frac{\boldsymbol{\beta}_2}{\boldsymbol{\beta}_1}\mathbf{y}_{2t} - \cdots - \frac{\boldsymbol{\beta}_K}{\boldsymbol{\beta}_1}\mathbf{y}_{Kt}$$

It is tempting to argue that the long-run effect of a unit increase in $\boldsymbol{y}_2$ will be a change in size $\frac{\boldsymbol{\beta}_2}{\boldsymbol{\beta}_1}$ in $\mathbf{y}_1$. This, however, ignores all the other relations between the variables which are summarized in a VAR(p) model or the corresponding VECM.

A one-time unit innovation in $\mathbf{y}_2$ may affect various other variables which also have an impact on $\mathbf{y}_1$. Therefore, the long-run effect of a $\mathbf{y}_2-$innovation on $\mathbf{y}_1$ may be quite different from $-\frac{\boldsymbol{\beta}_2}{\boldsymbol{\beta}_1}$. The impulse responses may give a better picture of the relations between the variables.

For stationary, stable VAR(p) processes, the impulse responses are the coefficients of specific MA representations and the effect of a one-time impulse dies out asymptotically. This is not the case for unstable, integrated or cointegrated VAR(p) process. However, the tools available for stable processes for *structural analysis*, that is, the *accumulated impulse responses, the responses to orthogonalized residuals* and *the forecast error variance decompositions*, are also available for the *unstable case*. The only quantities that cannot be computed in general are the *total "long-run" effects* or *total multipliers* Ψ_∞ and Ξ_∞ because they may not be finite.

3.8 Principal Component Analysis

Because of the high dimensionality feature of yield data and our interest in studying the commonality in movements of the yield curve drivers, in the sequel, we review the dimension reduction technique of Principal Component Analysis (PCA). The exposition follows Tsay, 2005.

[19]This situation might happen if the sample size is too small to satisfy the asymptotics underlying the test for cointegration and Granger causality.

PCA is one of the most commonly used techniques to study the covariance (or correlation) structure of multivariate time series, in order to understand the source of variations of these time series. Technically, given a k-dimensional random variable $\mathbf{r} = (r_1, \ldots, r_k)'$ with covariance matrix $\boldsymbol{\Sigma_r}$ and correlation matrix $\boldsymbol{\rho}_r$, the PCA finds a few linear combinations of r_i to explain the structure of $\boldsymbol{\Sigma_r}$ or $\boldsymbol{\rho}_r$.

Let $\mathbf{w}_i = (w_{i1}, \ldots, w_{ik})'$ be a k-dimensional vector, where $i = 1, \ldots, k$. Then

$$y_i = \mathbf{w}_i'\mathbf{r} = \sum_{j=1}^{k} w_{ij} r_j \tag{3.48}$$

is a linear combination of the random vector $\mathbf{r}$. If we let $\mathbf{r}$ hold k yield curve drivers, then y_i is the principal component that assigns weight w_{ij} to the jth yield curve driver. The vector $\mathbf{w}_i$ can be standardized so that $\mathbf{w}_i'\mathbf{w}_i = \sum_{j=1}^{k} w_{ij}^2 = 1$. Exploiting the properties of a linear combination of random variables, the variance and covariance of 3.48 are given by:

$$Var(y_i) = \mathbf{w}_i'\boldsymbol{\Sigma}_r\mathbf{w}_i, \qquad i = 1, \ldots, k, \tag{3.49}$$

$$Cov(y_i, y_j) = \mathbf{w}_i'\boldsymbol{\Sigma}_r\mathbf{w}_j, \qquad i, j = 1, \ldots, k, \tag{3.50}$$

The idea of PCA, as Tsay explains, is to find linear combinations $\mathbf{w}_i$ such that y_i and y_j are uncorrelated for $i \neq j$ and the variances of y_i are as large as possible. A theoretical result is that the proportion of total variance in $\mathbf{r}$ explained by the ith principal component is given by the ratio between the ith eigenvalue and the sum of all eigenvalues of $\boldsymbol{\Sigma}_r$:

$$\frac{Var(y_i)}{\sum_{i=1}^{k} Var(r_i)} = \frac{\lambda_i}{\lambda_1 + \cdots + \lambda_k}. \tag{3.51}$$

Whereas, the cumulative proportion of total variance explained by the first i principal components can be computed as:

$$\frac{\sum_{j=1}^{i} \lambda_j}{\sum_{j=1}^{k} \lambda_j}. \tag{3.52}$$

In practice, one selects a small number i of principal components, such that the prior cumulative proportion is large. Also in practice, the covariance matrix $\boldsymbol{\Sigma}_r$ and the correlation matrix $\boldsymbol{\rho}_r$ of the vector $\mathbf{r}$ holding the original raw data are unknown but they can be estimated by the sample covariance and correlation matrices under some regularity conditions. The number i of principal components can then be chosen by examining the so-called *scree plot*, which is the time plot of the eigenvalues $\hat{\lambda}_i$ (of the estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_r$) ordered from the largest to the smallest. The visual examination would consist in looking for an elbow in the scree plot, indicating that the remaining eigenvalues are relatively small and approximately of the same size and, hence, that they contribute little in explaining the variability in the underlying data.

3.9 Conclusion

In the present Chapter, we reviewed the theoretical concepts and methods of multiple time series analysis that we will employ for the derivation of the dynamic properties of the international yield curve drivers.

We introduced the VAR models as suitable tools for exploring lead-lag relationships among variables, we discussed the models'underlying assumptions of stationarity, stability, and invertibility. Furthermore, we showed how VAR(p) models can be simplified to VAR(1) models by adding appropriate variables. We explained how to determine in a systematic manner the lag order of a VAR model, how to conveniently estimate the model with multivariate least squares methods, and how to produce forecasts by minimizing the MSE. In addition, we introduced the concept of Granger causality to draw conclusions about the causal direction with which the variables in a VAR model influence each other. To this regard, the cross-correlation function can provide first suggestions on the existence of Granger causality and lead-lag relationships among given variables. Depending on what kind of realizations of the variables are included in a VAR model, we introduced three different types of Granger causality, i.e., no delayed Granger causality, no instantaneous and no delayed Granger causality, and no delayed but instantaneous Granger causality. Given an estimated VAR model, we explained how to derive the IR function, which can be used to investigate further the dynamic interrelationships between the model's variables.

Since many economic and financial time series tend to exhibit nonstationary behavior, we introduced the VEC models, illustrating their assumptions, properties, and estimation methods. For completeness, we discussed the concept of Granger causality in VEC models and the derivation of IR function.

Finally, we recalled the theoretical concepts behind PCA, as an important tool to deal with the high dimensionality of yield data and understand their source of variation.

Chapter 4

Dynamic Properties of U.S. and German Yield Curve Drivers

4.1 Introduction

"*No single currency dominates global bond markets*"[1]. As of Q4 of 2017, the debt securities outstanding of the central governments of all developed countries amounted to approximately 41,571 billions (amount in US dollars). Of this amount, 17,571 billions were US central government debt securities, over a half were non-US central government debt securities. Germany owned almost 1,292 billions of international debt securities, of which 1,012 billions were from financial corporations, 203 billions from non-financial corporations, and 78 billions from international central governments. The same numbers recorded for the US were 2,430 billions of international debt securities, of which 1,796 billions were from financial corporations, 629 billions from non-financial corporations, and 5 billions were from international central governments. [2] There is no doubt that the global bond markets have reached massive dimensions as a result of the internationalisation of capital markets, which started back in 1980. Today, international fixed income investors are using foreign bonds both tactically, as a substitute for domestic bonds, and strategically, by constructing cross-border bond portfolios benchmarked to the major global indexes. On one hand, benefits arise in terms of diversification, wealth preservation, and attractive returns; on the other hand, challenges arise in terms of which risks are international bond portfolios exposed to, since now spillover effects and macroeconomic shocks to the interest rate markets are transmitted internationally via monetary policy and risk channels. International fixed income investors, risk managers, and central banks, all have a vital interest in understanding how the global fixed income markets interact – the object of interest being the term structure of government bond yields of different world regions.

From a term structure modeling perspective, the global magnitude of bond markets induced an evolution of the term structure models from a domestic setting, in which the term structure of a single country is modeled in isolation, to a global setting, in which term structures of different

[1] Lee, 2006.

[2] Summary of debt securities outstanding, Bank for International Settlements, 2017.

world regions are modeled jointly with the aim of capturing their dependencies. Multiple studies provide strong evidence of cross-border dependencies of yield curves in different world regions (Al Awad and Goodwin, 1998; Ang and Piazzesi, 2003; Frankel, Schmukler, and Serven, 2004; Belke and Gros, 2005; Chinn and Frankel, 2003; Bauer and Rios, 2012; Byrne, Fazio, and Fiess, 2012; Abbritti et al., 2013; Jotikasthira, Le, and Lundblad, 2015; Byrne, Cao, and Korobilis, 2017). These dependencies arise in the form of common factors that drive the yield curves of different countries. The prominent work of Diebold, Li, and Yue, 2008 finds strong evidence about the existence of global yield factors, their high economic importance, and their direct linkage to global macroeconomic fundamentals such as inflation and real activity. The estimation results of Diebold, Li and Yue indicate that global yield factors do indeed exist and are economically important. The global level (relating to global inflation) and the global slope (relating to real economic activity) are found to explain significant fractions of country yield curve dynamics.

In terms of global macroeconomic fundamentals, other works support the results of Diebold, Li and Yue. Global inflation (Borio and Filardo, 2007; Ciccarelli and Mojon, 2010; Byrne, Fazio, and Fiess, 2012) and international business cycles (Lumsdaine and Prasad, 2003; Kose, Otrok, and Whiteman, 2003; Hellerstein, 2011; Dahlquist and Hasseltoft, 2013) explain large portions of the variance of the country-specific inflation and global bond risk premia.

In the tradition of Diebold, Li, and Yue, 2008, Spencer and Liu, 2010; Bauer and Rios, 2012; Abbritti et al., 2013; Jotikasthira, Le, and Lundblad, 2015; Byrne, Cao, and Korobilis, 2017, among others, extend the country-specific term structure models to the multi-country setting to incorporate the international dynamics of the term structure and find that yield curve fluctuations across different currencies are highly correlated, that macroeconomic variables are important drivers of international term and foreign exchange risk premia as well as expected exchange rate changes, and that global factors explain long-term dynamics in yield curves. Yield curve fluctuations are transmitted internationally, across different currencies, through the monetary policy channel and through the risk compensation channel, inducing the creation of lead-lag relationships between world economies. Taking the example of the United States and the European Union, a reciprocal leader-follower relationship does seem to exist, in the long run, between the European Central Bank (ECB) and the US Federal Reserve (the Fed) (Chinn and Frankel, 2003; Belke and Gros, 2005). The ECB follows the Fed in setting its monetary policy; the Fed is also increasingly influenced by the ECB, although the relationship is asymmetric. The international financial integration implies that the monetary policy shocks of the leading countries are transmitted internationally to the follower countries, affecting their financial conditions and giving rise to co-movement of business cycles across countries (Anderton, Di Mauro, and Moneta, 2004, Stock and Watson, 2005, Rey, 2016).

Despite the significant importance of modeling yield curves in a global setting, the yield curve literature is still lacking adequate research on the identification of the international yield curve drivers, on their dynamic properties, and, more importantly, on their co-movement. In both domestic and global yield curve modeling, it is common to disregard the in-sample properties of yield curve variables and make unrealistic assumptions about their dynamic evolution and correlation structure.

Yield curve variables are known to exhibit persistent, unit-root dynamics. This observation might suggest that yield curve variables are integrated of order one, [I(1)]. Nevertheless, yield

curve variables are commonly modeled in levels, thus, disregarding a potential cointegration structure.

With respect to the correlation structure, yield curve factors are often assumed to be uncorrelated, thus, imposing a diagonality assumption on the covariance matrix. Such an assumption excludes the possibility of lead-lag relationships arising from contemporaneous and non-contemporaneous dependencies of yield curves across different world regions. The lead-lag relationships among yield curve variables are also not thoroughly analyzed with the objective of determining whether specific yield curve variables possess explanatory power for other yield curve variables.

In this Chapter, we aim at fulfilling these gaps in the yield curve literature, by providing a rigorous and comprehensive study of what drives yield curves in different world regions, i.e., what are the international yield curve drivers, what are their dynamic properties, and how do they co-move. In our empirical study, we choose to focus on two major economies: US and Germany, assuming Germany as a representative of the euro area. Using US and German government bond yields, we seek to provide answers to the following research questions. Which are the yield curve drivers that jointly move the term structures of US and Germany. What are the dynamic properties of such international yield curve drivers? More specifically, given a system of international yield curve drivers, what are the interrelationships between the variables? How do the variables co-move? Is there commonality in the movements? Are there contemporaneous dynamic interdependencies? Are there non-contemporaneous dependency patterns, like causality linkages or lead-lag relationships? Do the variables have common trends so that they move together to some extent?

The present Chapter makes several important contributions: first, following a structured econometric workflow, we employ a wide range of tools to provide an extensive study on the identification of international yield curve drivers. To the best of our knowledge, the relevant literature was employing specific tools only, no study has reported the joint conclusions of a complete econometric workflow. Second, we screen out the most robust dynamic properties of international yield curve drivers using an extended sample period, which includes very recent observations. Third, we document, for the first time, the dynamic evolution and co-movement of systems of US and German yield curve drivers. Previous yield-curve literature lacks thorough analysis from this perspective. We now provide evidence of interrelationships of US and German yield curve drivers by documenting the nonstationarity/stationarity properties, volatility clustering, correlation and cross-correlation structure, causality linkages and lead-lag relationships, cointegration structure, and impulse-response functions.

These results are instrumental in developing new econometric models for forecasting the co-movement of international yield curve drivers.

The Chapter is organized as follows. Section 4.2 describes, visualizes, and analyzes the yield data, with the aim of screening out the most robust dynamic properties of the US and German yield curve drivers. Section 4.3 introduces our workflow for conducting the empirical study. Following the workflow, Section 4.4 reports the results of the ADF tests for unit root and distinguishes between stationary and nonstationary drivers. Section 4.5 deep-dives into the commonality of movements of the US and German yield curve drivers using cross-correlation analysis. Section 4.6 introduces the data generation process (DGP) of the stationary drivers, i.e.,

the US and German slopes. Upon estimation of the DGP for the slopes, Section 4.7 and Section 4.8 perform, respectively, Granger causality and Impulse Response (IR) analysis to understand the complete story about the interactions between the US and German slopes. The nonstationary drivers, i.e., the levels and curvatures are submitted to Cointegration Analysis in Section 4.9 and their DGPs are introduced in Section 4.10. Upon estimation of the DGPs for the levels and curvatures, Section 4.11 and Section 4.12 perform, respectively, Granger causality and IR analysis for the cointegrated drivers. Section 4.13 conducts Principal Component Analysis on US and German yields, spreads, and estimated drivers, in order to understand the source of variations of US and German yield curve drivers, provide an economic interpretation of the principal components, and discuss interest rate risk management beyond duration and convexity adjustment using the results of the PCA. Finally, Section 4.15 concludes the Chapter.

4.2 Data Description, Visualization, and Preliminary Analysis

In this section, we describe, visualize and analyze the yield data, with the aim of screening out the most robust dynamic properties of the US and German yield curve drivers. Our data consists of actively traded US and German government bond yield curves, retrieved from the Federal Reserve Board[3] and Deutsche Bundesbank[4] databases. The sample period runs from '1999:01' to '2018:01'. The yield data are sampled monthly (229 monthly observations) and the cross sections span over short-, medium-, and long-term maturities, i.e., over 6-month, 1-year, 2-year, 3-year, 5-year, 7-year, and 10-year maturities.

Yield Curves and Spreads Over Space and Time

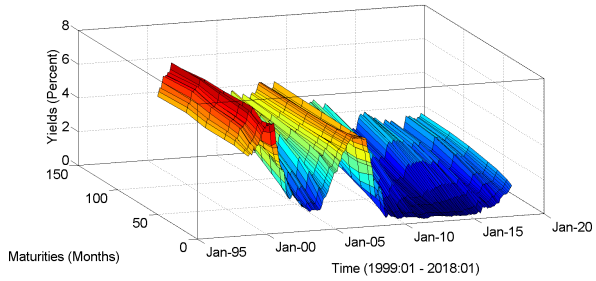
In Figure 4.1, we show the US and German government bond yield curves, in levels and first differences. In this figure it is possible to notice complex movements in all yield curve levels, as well as slopes and curvatures. The yields in levels appear quite persistent and exhibit a decreasing trend, which is more apparent in the German yields. The high persistency of yields in levels suggests nonstationary dynamics. Stationary dynamics can be noticed, instead, in the first differences of the yields. With respect to the volatility of the first differences, the US exhibit higher volatility compared to Germany. Also in Figure 4.1 we show the 3-dimensional surface of the German-US yield spreads, both in levels and first differences. The persistency of the spreads appears weaker compared to that of the underlying yields.

It is easy to spot the period of the 2007-2008 financial crisis, the beginning being signed by a drastic decrease in the magnitude of the yields, followed by the low interest rate environment. In the US, the low interest rate environment started soon after December 2008, when the Federal Reserve System (the Fed) reduced the Fed funds rate to become effectively zero (the lowest Fed funds rate possible). That was the period when the Fed began its first round of quantitative

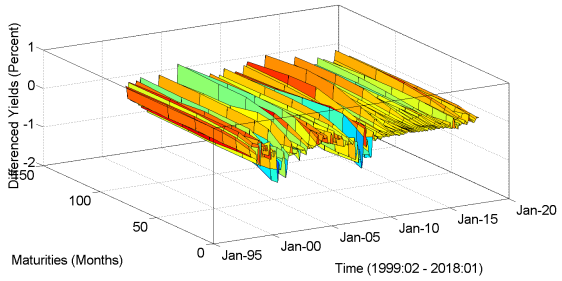
[3]Federal Reserve Board (FRB): Download Program.
[4]Deutsche Bundesbank: Time series databases.

easing. In Germany, the low rates environment started somewhere around October 2008 and May 2009, when the European Central Bank (ECB) decreased its key interest rates and introduced the "Enhanced Credit Support" to the banking sector. During this period, the German yields entered the negative territory. The drastic interventions of the Central Bankers introduced some noise in the dynamics of the term structures. The noise is apparent in the first difference surfaces, where a significant drop is recorded around mid-2008.

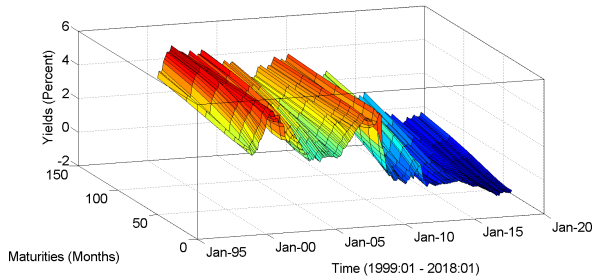
Figure 4.1: Yield curves over space and time. (Notes to figure: All yield data are monthly, [1999:01-2018:01], for 6-month, 1-year, 2-year, 3-year, 5-year, 7-year, 10-year maturities).



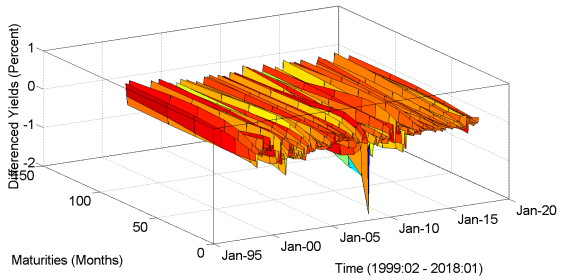
(a) US yields, in levels.



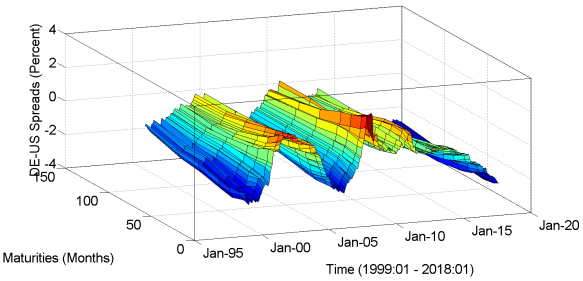
(b) US yields, in first differences.



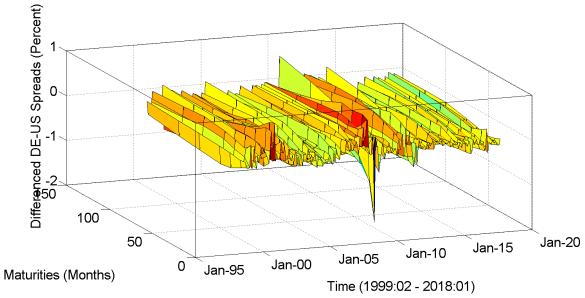
(c) German yields, in levels.



(d) German yields, in first differences.



(e) German-US yield spreads, in levels.



(f) German-US yield spreads, in first differences.

In Table A.1 and Table A.2 in Appendix A, we report descriptive statistics of the yield data. The US yields are 1.9-3.6% on average. The German yields are lowest on average, approximately 1.7-3.1%. All yield curves are upward-sloping. Yield volatility tends to decrease with maturity for both US and Germany. The minimum values show that only the German yields have recorded

negative values during the sample period under analysis. The short-term German yields tend to be less volatile compared to the US counterparts but the feature reverses for the medium- to long-term maturities, where the German yields tend to be more volatile than the US ones. The sample autocorrelations (reported in Table A.3 and Table A.2 in Appendix A) confirm the high persistency of the yields in levels, for both countries. The average first-order autocorrelation is around 0.99, for the US, and around 0.98, for Germany.

With respect to the German-US yield spreads, the statistics show that the spreads are negative, on average, for all maturities. The dynamics is downward-sloping, with the spreads becoming more and more negative. A change in the slope occurs for the 10-year maturity, where the average spread becomes less negative compared to the 7-year maturity. The spread volatility is lower compared to the single-country yields and decreasing with maturity. Similarly to the US and German yields, also the spreads are highly persistent, with an average first-order autocorrelation around 0.96. The descriptive statistics of the first differences of US and German yields and German-US spreads suggest stationarity and mean-reversion. The average yields and spreads and their respective volatility are around zero. The sample autocorrelations suggest weak persistency of all yields and spreads.

(Nelson-Siegel) Estimated Country Factors

As reviewed in Chapter 2, a few unobservable factors account for most of the changes in the shape of the yield curves. These factors are widely known as *level, slope*, and *curvature* (Litterman and Scheinkman, 1991; Dai and Singleton, 2000) and they describe how the yield curve changes shape in response to a shock.

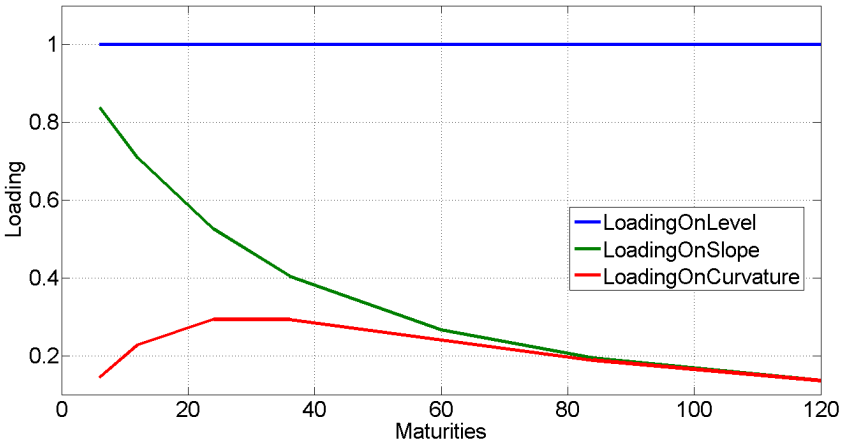
In this section, we estimate these unobservable factors for both the US and German term structures. To do so, we employ the Dynamic Nelson-Siegel model (as developed by Diebold and Li, 2006 and explained in Section 2.3):

$$y_{it}(\tau) = l_{it} + s_{it}\left(\frac{1-e^{-\lambda_{it}\tau}}{\lambda_{it}\tau}\right) + c_{it}\left(\frac{1-e^{-\lambda_{it}\tau}}{\lambda_{it}\tau} - e^{-\lambda_{it}\tau}\right) + \nu_{it}(\tau)$$

where l_{it}, s_{it}, and c_{it} denote the (US or German, $i = US, DE$) country specific level, slope, and curvature factors. The exponential components represent the Nelson-Siegel loading structure, which controls how the three factors affect the yields of different maturities. Figure 4.2 shows the Nelson-Siegel loadings on the estimated country factors. The blue line denotes the loading on the level factor. It equals 1, meaning that it produces an identical impact across all maturities, thus, inducing a parallel shift (up and down) of the whole yield curve.

The green line denotes the loading on the slope factor, which starts at around 1 and decreases to zero, as maturity increases. Such a loading makes so that the slope factor increases short-term interest rates by much larger amounts than long-term interest rates, so that the yield curve changes its steepness (it becomes less steep and its slope decreases). Finally, the red line denotes the loading on the curvature factor, which starts at around zero, increases for the medium-term maturities, and decreases down to zero as the long-term maturity spectrum is reached. Such a loading makes so that the curvature factor focuses its effects on medium-term interest rates, producing "hump-shaped" movements in the yield curve.

Figure 4.2: Nelson-Siegel loadings on estimated country factors.



In Figures 4.3 and 4.4, we show the US and German level, slope, and curvature (in levels and first differences), which we estimated via a series of OLS regressions for each of the two countries[5]. Since we are interested in the co-movement of these factors/drivers, we plot them inter-country and investigate the commonality in movements. Confirming the results of Diebold, Li, and Yue, 2008, the visual analysis suggests commonality in factor dynamics. More specifically, the US and German levels tend to move together and follow similar dynamics with a decreasing trend over time. The decreasing trend might suggest nonstationary behavior of the two series. Divergent dynamics are apparent from the end of 2012 to the end of the sample period, when the US level started an increasing trend but the German level did not follow at the same pace and with the same magnitude. In the early 2017, a significant difference can be observed in the two levels: the German level almost hit zero, whereas the US level stayed at around 2%. The plot in 4.3 suggests some lead-lag structure between the two levels, although it is not clear which country is leading and which one is lagging.

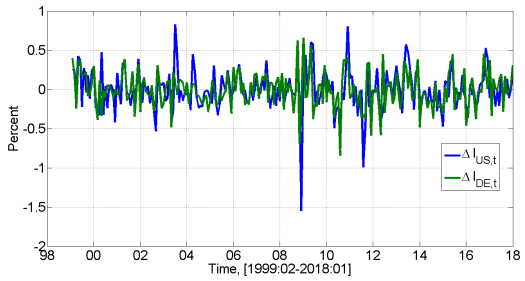
The US and German levels in first differences show a contained volatility, suggesting a stationary behavior. Episodes of volatility clustering seem to have occurred somewhere between 2009 and 2013, when large changes were followed by large changes and small changes were followed by small changes, of either sign. The drastic drop during the financial crisis, which we observed in the yield curves, is also visible in the US and German levels. The descriptive statistics (included in Appendix A, Table A.5 and Table A.6) show that, on average, the US level is higher than the German one and less volatile. Both US and German levels are highly persistent with first-order sample autocorrelations of 0.97 and 0.98, respectively.

Similar observations can be deduced for the US and German slopes and curvatures, which we plot in Figure 4.4.

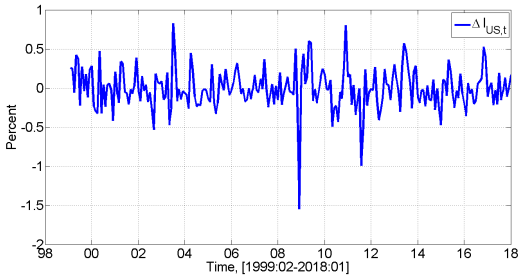
[5]The first step of the process consists in fixing λ at 0.0609, meaning that the value at which the loading on the curvature is maximized occurs at 30 months. The next step equates the level, slope, and curvature factors to the regression coefficients obtained by OLS, and accumulates a 3D time series of estimated country factors by repeating the OLS fit for each observed yield curve.

Figure 4.3: (Dynamic Nelson-Siegel) estimated country factors: US and German levels, [1999:01 -2018:01].

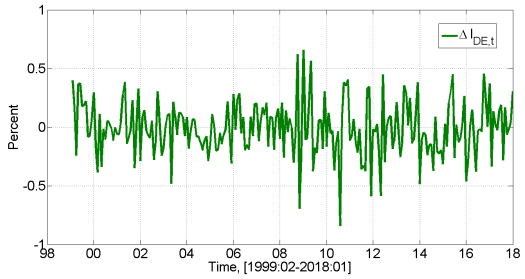
(a) US and German levels, in levels.



(b) US and German levels, in first differences.

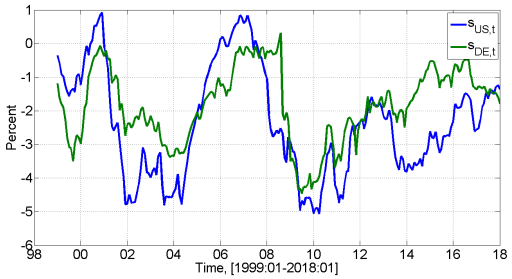


(c) US level, in first differences.

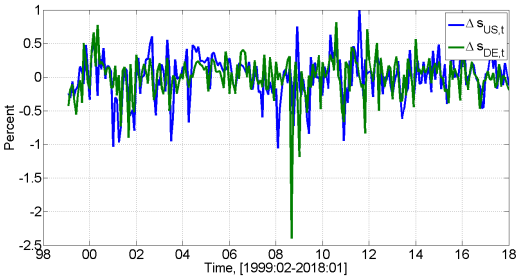


(d) German level, in first differences.

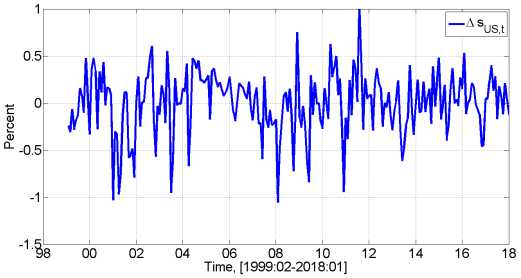
Figure 4.4: (Dynamic Nelson-Siegel) estimated country factors: US and German slopes and curvatures [1999:01-2018:01].



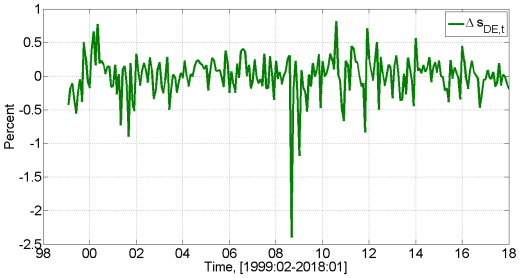
(a) US and German slopes, in levels.



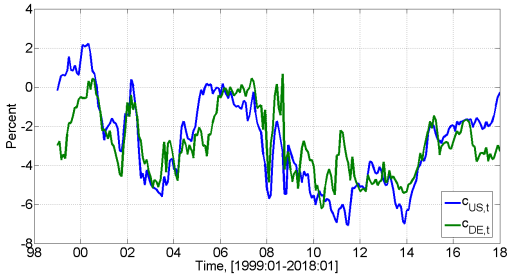
(b) US and German slopes, in first differences.



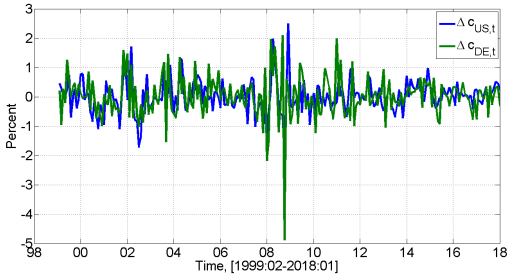
(c) US slope, in first differences.



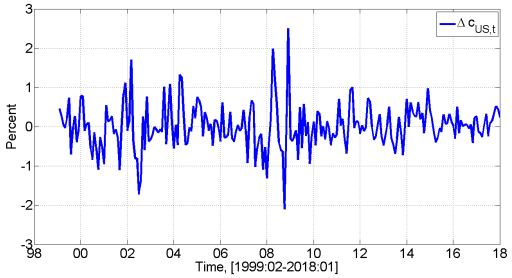
(d) German slope, in first differences.



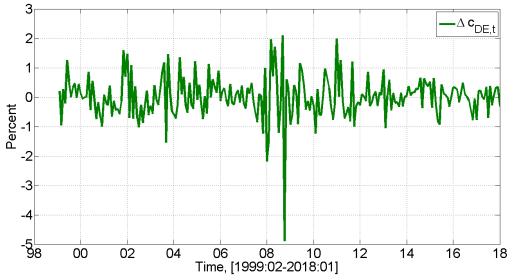
(e) US and German curvatures, in levels.



(f) US and German curvatures, in first differences.

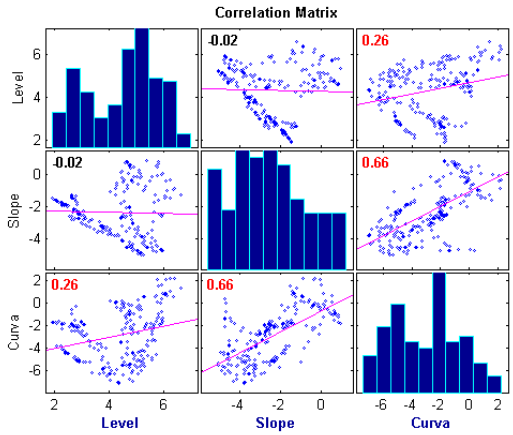


(g) US curvature, in first differences.

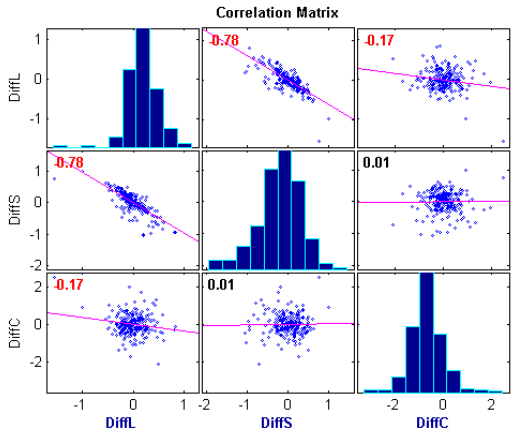


(h) German curvature, in first differences.

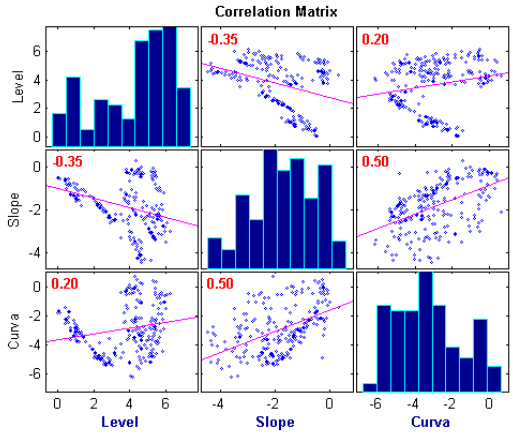
Figure 4.5: Intra-country factor correlation: US and Germany, [1999:01 -2018:01].



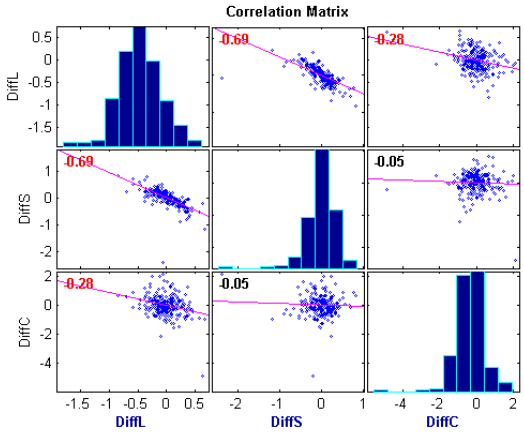
(a) US factors, in levels.



(b) US factors, in first differences.



(c) German factors, in levels.



(d) German factors, in first differences.

The two slopes exhibit common dynamics, with significant change in means around 2004 and 2007. Both slopes are negative for most of the time in the sample period. The average US slope is more negative then the German counterpart and significantly more volatile. Similarly to the US and German levels, the US and German slopes are highly persistent in levels and not so persistent in first differences.

The US and German curvatures also tend to follow each other quite closely. The persistency of the two curvatures in levels is high, although not as high as that of the levels and slopes.

Figure 4.5a is a matrix of plots showing correlations among pairs of US yield curve drivers ($l_{US,t}$, $s_{US,t}$, and $c_{US,t}$).

Histograms of pairs of $l_{US,t}$, $s_{US,t}$, and $c_{US,t}$ appear along the matrix diagonal; scatter plots of driver pairs appear off diagonal. The slope of the least-squares reference line in the scatter plots are equal to the displayed Pearson's correlation coefficients. The correlation coefficients highlighted in red indicate which pairs of variables have correlations significantly different from zero. For the US yield curve drivers, only the $l_{US,t} - c_{US,t}$ and $s_{US,t} - c_{US,t}$ pairs have positive correlations significantly different from zero. The $l_{US,t} - s_{US,t}$ are almost uncorrelated. The same plot with data in first differences (Figure 4.5b) is slightly changed: $\Delta l_{US,t} - \Delta s_{US,t}$ are negatively correlated, $\Delta l_{US,t} - \Delta c_{US,t}$ and $\Delta s_{US,t} - \Delta c_{US,t}$ are almost uncorrelated.

In Figure 4.5c, we can see that all German yield curve drivers, in levels, have correlations significantly different from zero. $l_{DE,t} - s_{DE,t}$ are negatively correlated, whereas, the $s_{DE,t} - c_{DE,t}$ are positively correlated.

The correlations of the German yield curve drivers in first differences (Figure 4.5d) are significantly different from zero only for $\Delta l_{DE,t} - \Delta c_{DE,t}$ and $\Delta l_{DE,t} - \Delta s_{DE,t}$; $\Delta s_{DE,t} - \Delta c_{DE,t}$ are almost uncorrelated.

In Figure 4.6 and Figure 4.7 we reproduce the matrix of correlation plots for pairs of mixed US and German yield curve drivers in levels and first differences, respectively, in order to gain insights into the inter-country factor correlation. Almost all pairs have correlations different from zero, except for $s_{US,t} - l_{DE,t}$ and $c_{US,t} - l_{DE,t}$. The $l_{US,t} - l_{DE,t}$ have the highest positive correlation, 0.92.

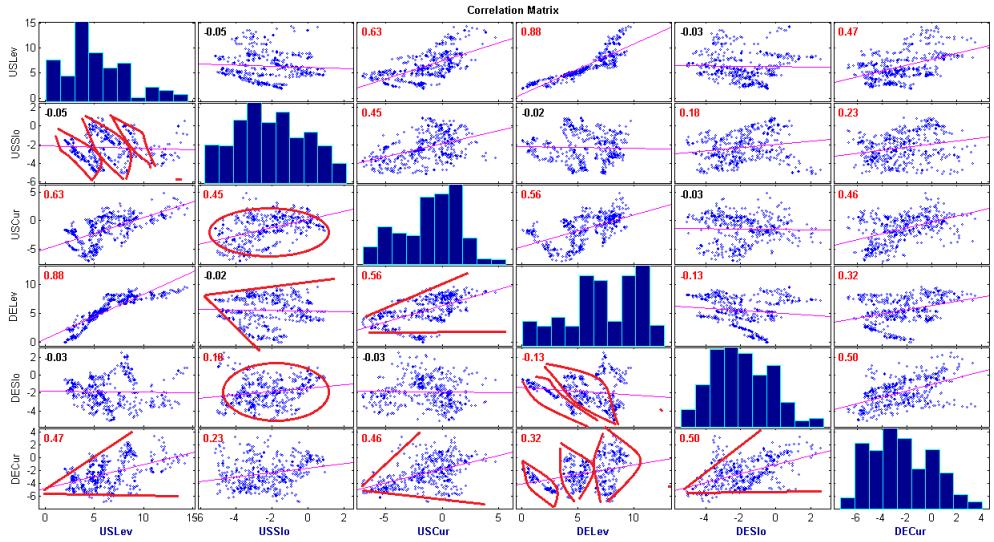


Figure 4.6: Inter-country factor correlation: US and German factors in levels, [1999:01-2018:01]

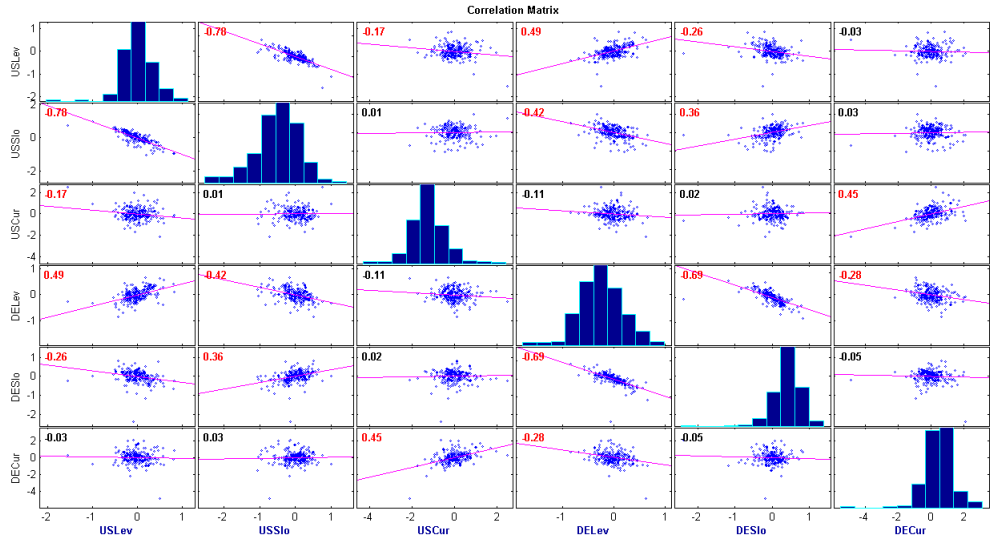


Figure 4.7: Inter-country factor correlation: US and German factors in first differences, [1999:02-2018:01]

The correlation between the two levels decreases if we consider the series in first differences. Highly positive correlations have also the slopes and the curvatures.

It is interesting to note in Figure 4.6 that the correlation for the highly persistent factors, which present a "near-unit-root" behavior, produces distinctive patterns on the plot. There are clear blotches of dots that could suggest the presence of regime switches in the data[6]. For the

[6]We study this aspect in Part II, where we focus on structural breaks and regime switches in the dynamics of international yield curve drivers. In Part II we perform a split-sample correlation analysis (we split the sample

correlation of $l_{US,t} - s_{US,t}$, $l_{DE,t} - s_{DE,t}$, and $l_{DE,t} - c_{DE,t}$ two to three blotches of dots can be recognized.

A heteroskedastic behavior can be noticed for $l_{US,t} - c_{DE,t}$, $s_{US,t} - l_{DE,t}$, $c_{US,t} - c_{DE,t}$, and $c_{US,t} - l_{DE,t}$. Even though the differenced country factors appear to fluctuate around a constant level[7], they might still exhibit autocorrelation in the squared series or volatility clustering. We investigate the presence of volatility clustering by conducting the Engle's autoregressive conditional heteroskedastic (ARCH) test[8] on the residuals of differenced country factors. To conduct the test, we determine a suitable number of lags for the model by fitting the model over a range of plausible lags[9] and comparing the fitted models. We choose the number of lags that yields the best fitting model for the ARCH test (the lowest Bayesian Information Criterion (BIC)).

The results are presented in Table 4.1 and indicate that the null hypothesis of no conditional heteroskedasticity is rejected for $\Delta c_{DE,t}$ (h=1, p-Value = 0) in favor of the ARCH(1) alternative. The F statistic for the test is 6.67, slightly higher than the critical value from the χ^2 distribution with 1 degree of freedom, 6.63.

The null hypothesis is not rejected (h=0,p-Value = 0.39) for $\Delta l_{US,t}$ (similar results hold for $\Delta l_{DE,t}$, $\Delta s_{US,t}$, $\Delta s_{DE,t}$ and $\Delta c_{US,t}$). The F statistic for the test is 0.74, significantly lower than the critical value from the χ^2 distribution with 1 degree of freedom, 6.63. Hence, one can conclude that there is no conditional heteroskedasticity (ARCH effects) in the residuals of $\Delta l_{US,t}$ (and $\Delta l_{DE,t}$, $\Delta s_{US,t}$, $\Delta s_{DE,t}$ and $\Delta c_{US,t}$).

Table 4.1: Engle's ARCH test: Differenced US and German yield curve factors

Time Series/Results	$\Delta l_{US,t}$	$\Delta l_{DE,t}$	$\Delta s_{US,t}$	$\Delta s_{DE,t}$	$\Delta c_{US,t}$	$\Delta c_{DE,t}$
Suitable Nr. of Lags	1	1	1	1	1	1
Fstat	0.74	5.50	2.41	0.00	1.17	6.67
χ^2 Critical	6.63	6.63	6.63	6.63	6.63	6.63
p-Value	0.39	0.02	0.12	0.97	0.28	0.01
α	0.01	0.01	0.01	0.01	0.01	0.01
Decision, h	0	0	0	0	0	1

4.3 Introduction to the Workflow

Having analyzed the yield data and derived the first drivers of the US and German yield curves, i.e., the levels, slopes, and curvatures, we now proceed with a deeper analysis, which seeks to understand what are the dynamic properties of the US and German yield curve drivers, to which we will refer, from now on, as *international yield curve drivers* (IYCD). Our analysis follows the workflow depicted in Figure 4.8.

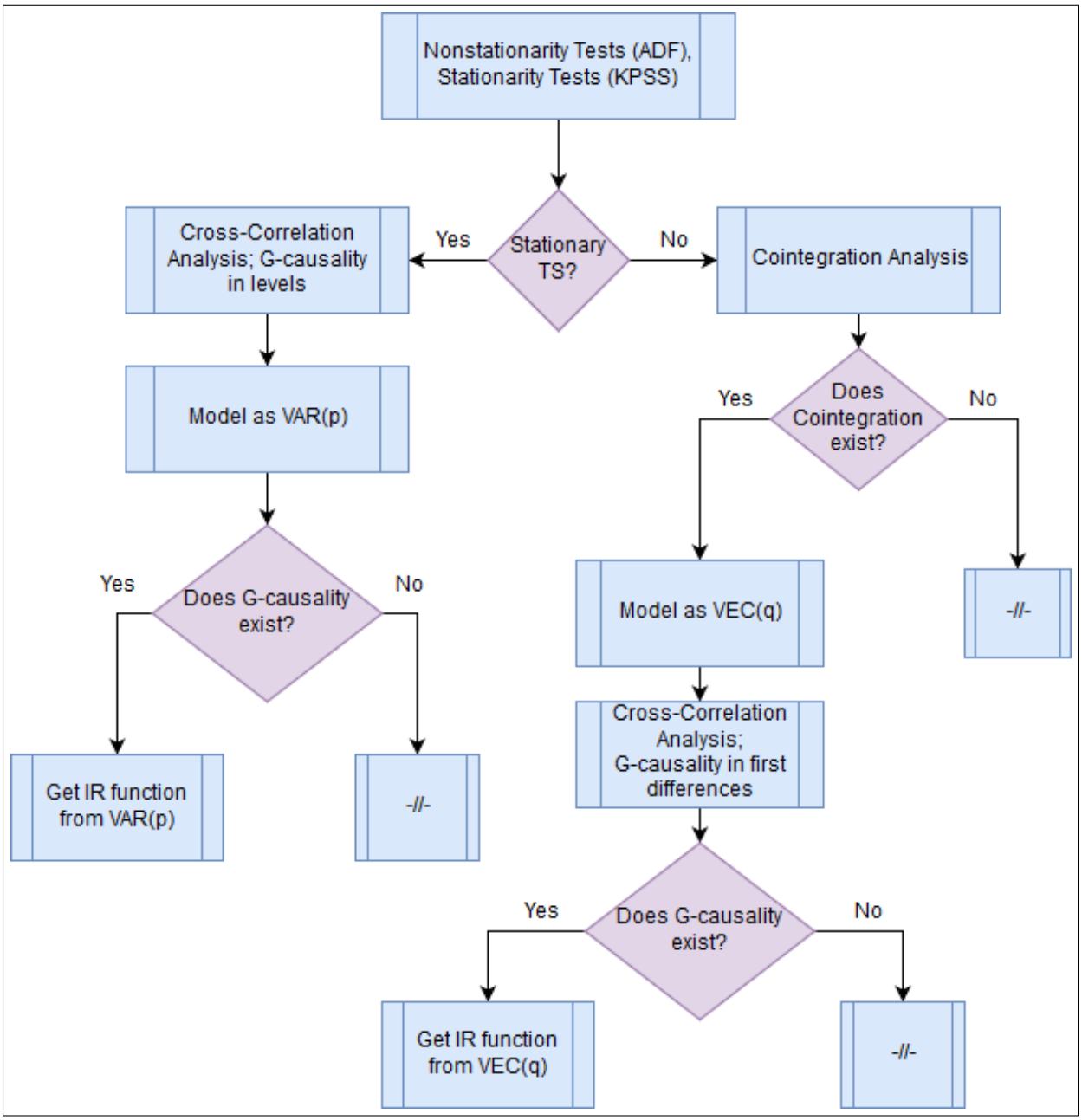
before and after the financial crisis of 2008) to learn whether regime switches occurred in our sample period.

[7] See Figures 4.3c, 4.3d, 4.4c, 4.4d, 4.4g, 4.4h.

[8] Engle, 1982.

[9] We choose a range of 12 months because of the monthly frequency of the data.

Figure 4.8: Dynamic properties: Workflow



We first submit all yield variables to unit root tests, in order to confirm the nonstationary/stationary behavior of the series. We employ the augmented Dickey-Fuller (ADF) unit root test. The ADF test is performed using all three model specifications, i.e., the *autoregressive model variant* (AR), the *autoregressive model with drift variant* (ARD), and the *trend-stationary model variant* (TS), although the decisions of the tests are based on the specification that best describes the data under both the null and alternative hypotheses. Therefore, if a series seems to exhibit a deterministic and stochastic trend, we consider the result of the TS model specification, where the alternative model includes a drift coefficient, c, a deterministic trend coefficient, δ, and an $AR(1)$ coefficient, $\phi < 1$. If the series does not exhibit any trend but seems to have a nonzero mean, we consider the ARD specification, where the alternative model includes only a drift coefficient, c, and an $AR(1)$ coefficient, $\phi < 1$. Finally, if a series does not exhibit any trend and seems to fluctuate around a zero mean, we consider the AR specification, where the alternative model includes only an $AR(1)$ coefficient, $\phi < 1$. When performing the ADF tests, we specify the lag length, p, following the method proposed by Ng and Perron, 1995.

If the unit root tests reveal stationary behavior of the time series under consideration, we follow the left parent branch of the diagram, which leads to the theory of stable vector autoregressive processes (as explained in Section 3.2). If the unit root tests reveal, instead, nonstationary behavior, we follow the right parent branch of the diagram, which leads to the theory of cointegrated processes and vector error correction models (as explained in Section 3.5).

With respect to stationary series, we continue the analysis using data in levels and study the cross-correlation structure, in order to obtain first insights of Granger causality and lead-lag relationships among given variables. The presence of causality is ultimately confirmed/rejected by performing Granger causality analysis (as explained in Section 3.3). The stationary variables are modeled as VAR(p) processes. If causality structure is present, we investigate it further by deriving the IR function from the estimated VAR(p) processes (theory in Section 3.4).

With respect to nonstationary series, we are interested in learning whether there are feedback mechanisms that force the nonstationary processes to stay close together. Hence, we perform cointegration analysis (as explained in Section 3.5) and model the cointegrated variables as VEC(q) models. We gain first insights of Granger causality from cross-correlation analysis in first differences of data. We then confirm/reject the presence of Granger causality (in first differences of data), as explained in Section 3.6. If causality structure is present, we investigate it further by deriving the IR function from the estimated VEC(q) processes (theory in Section 3.7).

4.4 Stationary and Nonstationary IYCDs

Yields and spreads in levels are submitted for ADF tests, which suggest nonstationarity, especially for medium- and long-term maturities (results in Table A.7, A.8, A.9, A.10, A.11). Consequently, one can conclude that yields and yield spreads are integrated of order 1, [I(1)]. To ensure valid statistical inference, the first differences need to be taken to induce stationarity in the yield variables (results in Table A.8, A.10, A.12). These results invalidate the Economic Theory, which postulates that nominal bond yields cannot be I(1), since they have a lower bound support at

zero and an upper bound support lower than infinity. Our unit root test results support Jarrow's argument (Jarrow, 2013) that the belief that there exists a zero-lower bound on interest rates is a myth rather than reality. As Jarrow argues, a negative default-free spot rate of interest is consistent with an arbitrage-free term structure evolution in a competitive and nearly frictionless market. Despite the nonstationarity in-sample property of yields and spreads, many academic papers on yield curve modeling and forecasting choose to model yields in levels[10] and, thus, disregard their real in-sample properties.

The ADF test results for the US and German estimated country factors (reported in Table A.13) suggest that only the $s_{US,t}$ and $s_{DE,t}$ are stationary, if we consider the autoregressive model with drift variant (which seems to be the most plausible description of the data). The levels, $l_{US,t}$, $l_{DE,t}$ and curvatures, $c_{US,t}$, $c_{DE,t}$, are nonstationary, considering the same ADF model variant, i.e., the autoregressive model with drift. Taking the first differences, all estimated country factors become stationary (results reported in Table A.14). These results are somewhat in line with those of Diebold, Li, and Yue, 2008, who also found that the factor roots are not easily distinguished from unity. The Economic theory is again invalidated, since the theory strongly suggests that the roots are less than one and that nominal bond yields cannot go negative but they would eventually go negative almost surely if they contained unit roots (Diebold, Li, and Yue, 2008).

4.5 Cross-Correlation Analysis

In this section, we start exploring deeper the commonality in movements of US and German yield curve drivers, which we observed in their joint plots[11]. In today's well-integrated international capital markets, it is natural to conjecture the existence of global bond yield factors, especially when lead-lag relationships are present among the international yield curve drivers. Numerous academic works (Al Awad and Goodwin, 1998; Solnik, 1974; Thoms, 1993; Dungey, Martin, and Pagan, 2000; Brennan and Xia, 2006; Lumsdaine and Prasad, 2003; Gregory and Head, 1999; Kose, Otrok, and Whiteman, 2003) exploit the presence of cross-border dependencies of yield curves and model the term structure in such a way that certain countries assume the role of global players, as such, leading the economies of other modeled world economies. The presence of lead-lag relationships makes possible that information from foreign yield curves (the "leaders") may have predictive power in forecasting the domestic yield curves (the "laggards").

Wang, Yang, and Li, 2007 study interest rate linkages in the Eurocurrency market to find that, before the European Monetary Union (EMU), the German eurocurrency rate played a strong global role, whereas, after the introduction of the euro, the US rate started to assume an increasing role in affecting eurozone currency interest rates. Diebold, Li, and Yue, 2008 also confirm the leading role of the US and thus, independence of the US market from other modeled economies (Japan, UK, and Germany).

Let x_t denote the German estimated country factors (i.e., $l_{DE,t}$, $s_{DE,t}$, $c_{DE,t}$, $\Delta l_{DE,t}$, $\Delta s_{DE,t}$,

[10] Nelson and Siegel, 1987; Diebold and Li, 2006; Diebold, Rudebusch, and Aruoba, 2006; Diebold, Li, and Yue, 2008; Duffee, 2006; Ang and Piazzesi, 2003; Bansal and Zhou, 2002; Dai and Singleton, 2000.

[11] Figures 4.3 and 4.4.

$\Delta c_{DE,t}$) and y_t the US estimated country factors (i.e., $l_{US,t}$, $s_{US,t}$, $c_{US,t}$, $\Delta l_{US,t}$, $\Delta s_{US,t}$, $\Delta c_{US,t}$), we will now use the *cross correlation function* to determine whether lags of US estimated country factors may be useful predictors of German estimated country factors. This can be deduced if, in the sample cross covariance function

$$c_{xy}(k) = \begin{cases} \frac{1}{T}\sum_{t=1}^{T-k}(x_t - \bar{x})(y_{t+k} - \bar{y}); & \text{if } k = 0, 1, 2, ... \\ \frac{1}{T}\sum_{t=1}^{T+k}(y_t - \bar{y})(x_{t-k} - \bar{x}); & \text{if } k = 0, -1, -2, ... \end{cases}$$

(or, equivalently, in the cross correlation function) we see *peaks at positive lags*, hence at $k = 0, 1, 2, ...$, then US estimated country factors, (y_t), *lag* German estimated country factors, x_t. If we see *peaks at negative lags*, hence at $k = 0, -1, -2, ...$, then US estimated country factors, (y_t), *lead* German estimated country factors, x_t.

In Figure 4.9, we show the sample cross-correlation function (with 12 lags) for US and German estimated country factors of the same class.

The sample cross-correlation function of German country level and US country level, in levels of data, $c_{xy}(k) = c_{l_{DE,t},l_{US,t+k}}(12)$, is slowly decreasing. A similar behavior is visible for the country slopes and curvatures. The sample cross-correlation function of German country level and US country level, in first differences of data, $c_{xy}(k) = c_{\Delta l_{DE,t},\Delta l_{US,t+k}}(12)$, is not symmetrical about zero and has a well-defined peak at $k = +1$, indicating that $\Delta l_{US,t}$ lags one month behind $\Delta l_{DE,t}$.

A similar remark can be made for slopes and curvatures, in first differences, where again a peak at $k = +1$ suggests causal structure from the differenced German factor to the differenced US factor.

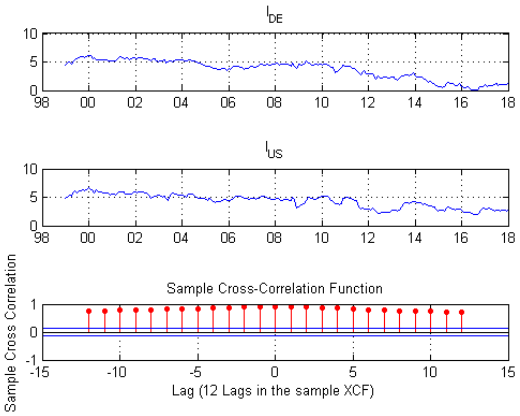
In Figure 4.10, we plot the sample cross-correlation function for pairs of estimated country factors of mixed classes. Well-defined peaks at positive lags (thus, suggesting leading behavior of German country factors with respect to US country factors) are visible for differenced German level and differenced US slope (a negative peak at $k = +1$), for differenced German slope and differenced US level (negative peaks at $k = +1$ and $k = +4$), for differenced German slope and differenced US curvature (a negative peak at $k = +3$), and for differenced German curvature and differenced US level (a positive peak at $k = +2$ and a negative peak at $k = +7$). No special pattern is visible for differenced German curvature and differenced US slope.

4.6 Data Generation Processes for Stationary IYCDs

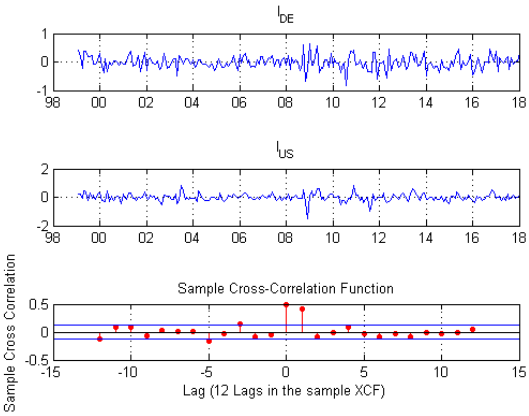
The slopes of the US and German term structure are found to be stationary and positively correlated (the correlation being significantly different from zero). Moreover, the cross-correlation analysis suggests the existence of causality structure from the German slope to the US slope. These findings provide good reasons to fit a VAR model to the two slopes, in order to exploit the linear dependence of the variables on their own lagged values and those of the other variable in the vector and to study the lead-lag relationship between the US and German slopes.

Therefore, for the $s_{US,t}/s_{DE,t}$ system, we fit a 2D-VAR(5) model. The order of the model is chosen following the procedure in Section 3.2 and the results are reported in Table 4.2. When

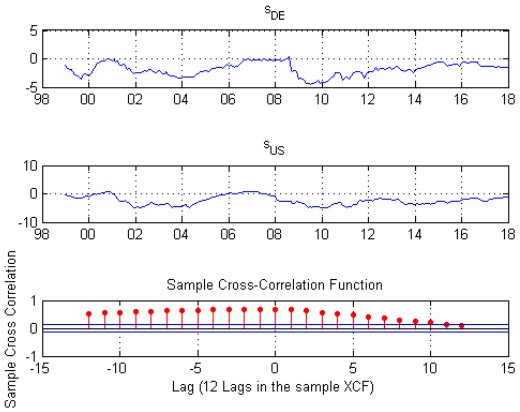
Figure 4.9: Sample cross-correlation function: Estimated country factors, same class.



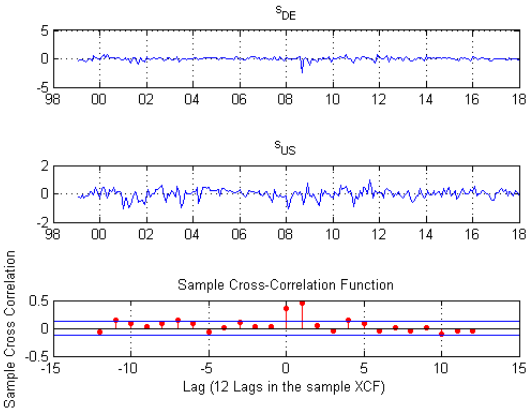
(a) $c_{l_{DE,t},l_{US,t+k}}(12)$



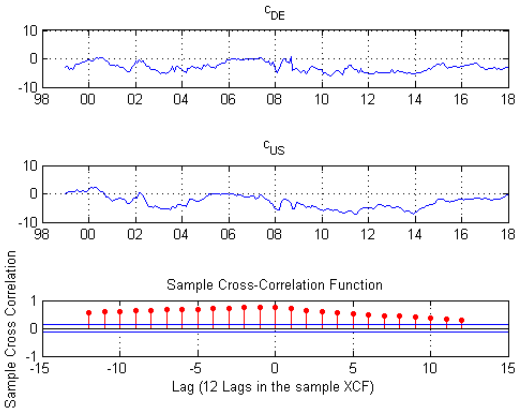
(b) $c_{\Delta l_{DE,t},\Delta l_{US,t+k}}(12)$



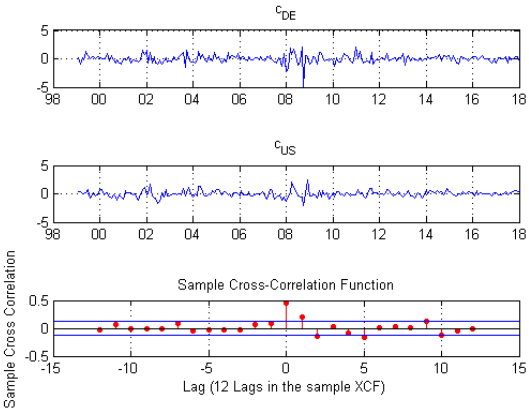
(c) $c_{s_{DE,t},s_{US,t+k}}(12)$



(d) $c_{\Delta s_{DE,t},\Delta s_{US,t+k}}(12)$

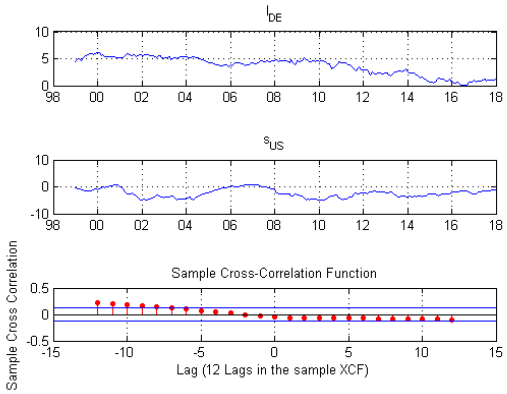


(e) $c_{c_{DE,t},c_{US,t+k}}(12)$

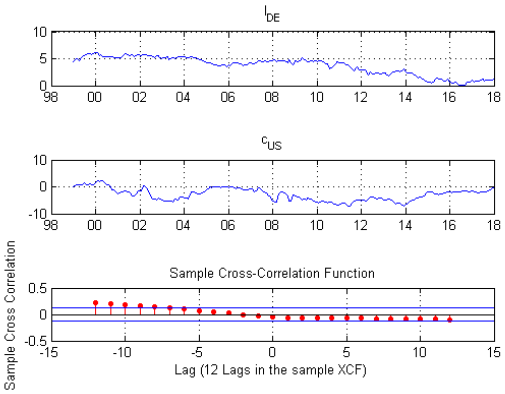


(f) $c_{\Delta c_{DE,t},\Delta c_{US,t+k}}(12)$

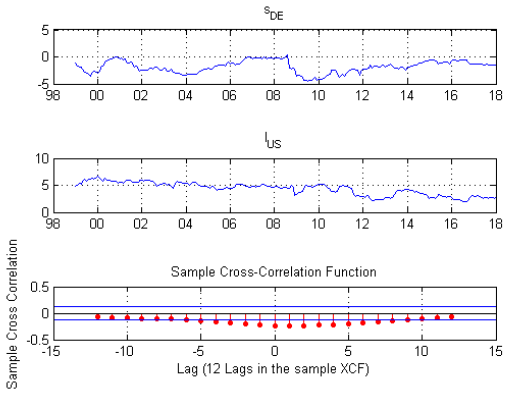
Figure 4.10: Sample cross-correlation function: Estimated country factors (in levels), mixed classes.



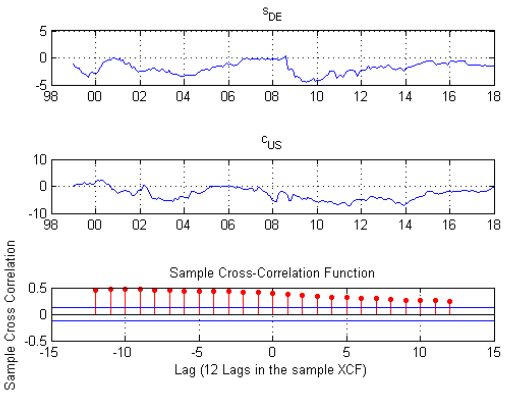
(a) $c_{l_{DE,t},s_{US,t+k}}(12)$



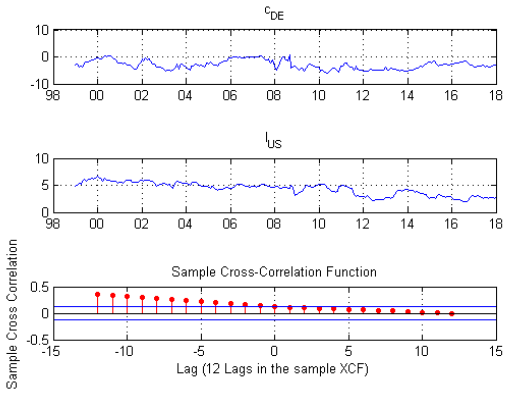
(b) $c_{l_{DE,t},c_{US,t+k}}(12)$



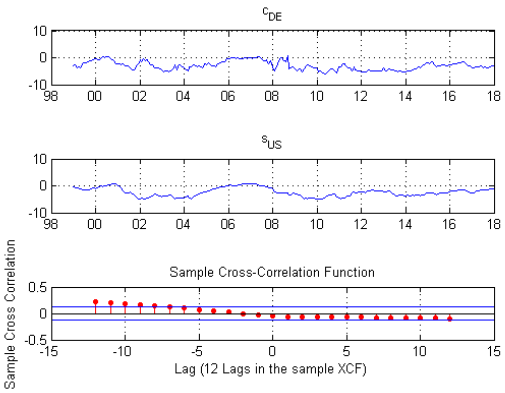
(c) $c_{s_{DE,t},l_{US,t+k}}(12)$



(d) $c_{s_{DE,t},c_{US,t+k}}(12)$

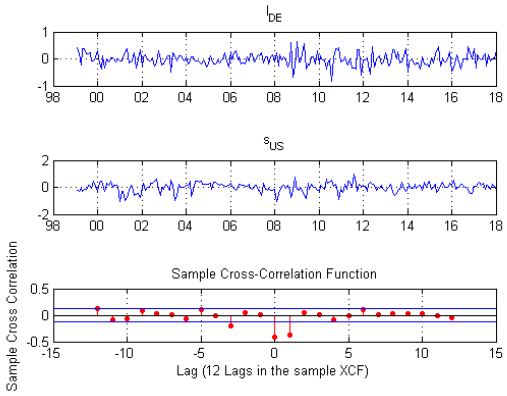


(e) $c_{c_{DE,t},l_{US,t+k}}(12)$

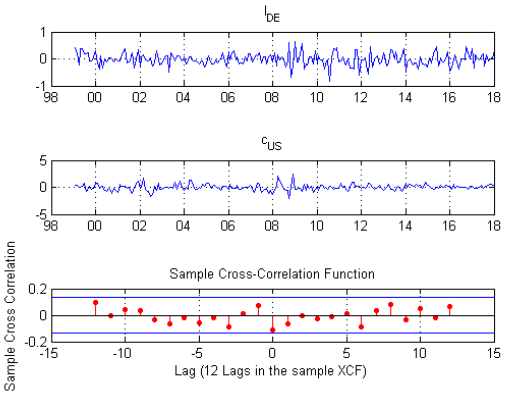


(f) $c_{c_{DE,t},s_{US,t+k}}(12)$

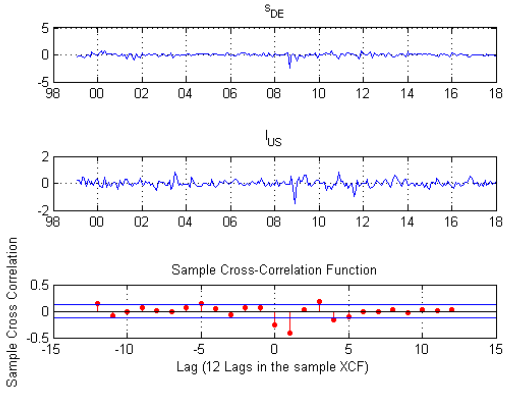
Figure 4.11: Sample cross-correlation function: Estimated country factors (in first differences), mixed classes.



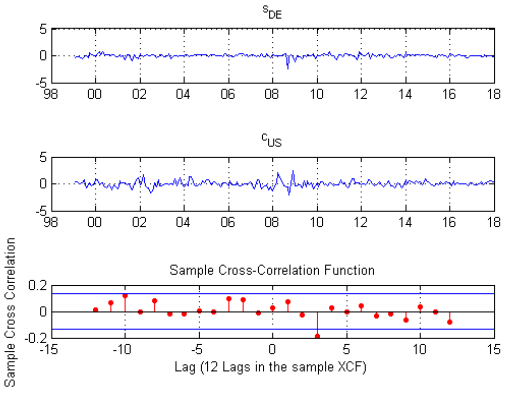
(a) $c_{\Delta l_{DE,t},\Delta s_{US,t+k}}(12)$



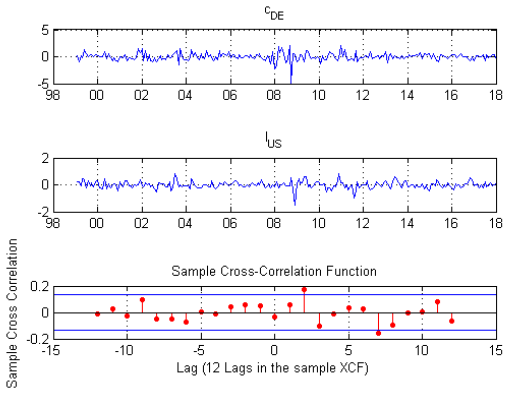
(b) $c_{\Delta l_{DE,t},\Delta c_{US,t+k}}(12)$



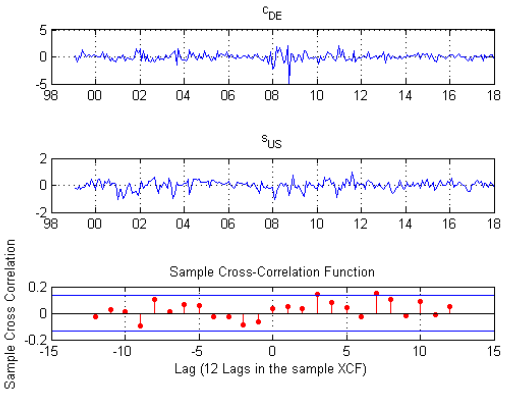
(c) $c_{\Delta s_{DE,t},\Delta l_{US,t+k}}(12)$



(d) $c_{\Delta s_{DE,t},\Delta c_{US,t+k}}(12)$



(e) $c_{\Delta c_{DE,t},\Delta l_{US,t+k}}(12)$



(f) $c_{\Delta c_{DE,t},\Delta s_{US,t+k}}(12)$

estimating the VAR model for the selection of the optimal lag length, a deterministic trend term is included. Hence, the VAR(p) model we estimate is of the following form:

$$\boldsymbol{y_t} = \boldsymbol{\nu} + \mathbf{A_1}\boldsymbol{y_{t-1}} + \mathbf{A_2}\boldsymbol{y_{t-2}} + \cdots + \mathbf{A_p}\boldsymbol{y_{t-p}} + \boldsymbol{\varepsilon_t}. \tag{4.1}$$

Table 4.2: Lag-order selection statistics for VAR model.

System	Optimal Lag	Portmanteau Test (asymptotic)
$s_{US,t}/s_{DE,t}$	5	χ^2 = 58.87, df = 48, p-value = 0.1352

For the $s_{US,t}/s_{DE,t}$ system, an optimal lag length of 5 is selected according to Lütkepohl's version of AIC. The $p-$value of the multivariate Portmanteau- and Breusch-Godfrey test for serially correlated errors is 0.1352, hence we fail to reject the null hypothesis for the absence of up to the order 5 of serially correlated disturbances in a stable VAR(5).

The least squares estimates are reported in Equation 4.2. t-values[12] are reported in parentheses underneath parameter estimates.

In the US slope equation, $s_{US,t}$, only the coefficients of $s_{US,t-1}$, $s_{DE,t-1}$, $s_{DE,t-2}$, and $s_{US,t-4}$ have high t-values and, hence, are significant. For all the other coefficients, the null hypothesis of zero cannot be rejected (i.e., the null hypothesis that these coefficients are not significant cannot be rejected). The equation for $s_{US,t}$ has high explanatory power[13], with almost 97% (R^2=0.9673) of the variance of $s_{US,t}$ explained by the regression equation.

Similar results hold for the German slope equation, in which only the coefficients of $s_{DE,t-1}$, $s_{US,t-3}$, and $s_{US,t-4}$ are significant. For all the other coefficients, the null hypothesis of zero cannot be rejected. The equation for $s_{DE,t}$ has slightly more explanatory power than the equation for $s_{US,t}$. In fact, more than 98% (R^2=0.9838) of the variance is explained.

Despite the presence of many insignificant coefficients in lags 4 and 5, we do not eliminate these lags as they are needed to ensure the absence of serial correlation in the residuals and the stability of the process.

[12]We recall that the t-value of an estimated coefficient measures how many standard deviations that coefficient is far from zero. t-values can be calculated by dividing each coefficient estimate by its respective estimated standard deviation under the assumption that that coefficient is zero. The t-probability of a coefficient estimate tests the significance of that coefficient by estimating the probability of the null hypothesis that that coefficient is zero. The t- probability is the p-value of the t-statistic, i.e., is the probability of the tail beyond the observed value of the t-statistics of the Student-t distribution with T-p (for the $s_{US,t}/s_{DE,t}$ system, T-p = 229 - 5 = 224) degrees of freedom (Rachev et al., 2007).

[13]The R^2 and the adjusted R^2 quantify the explanatory power of a regression equation, in the sense of how much of the variance of the dependent variable is explained by the regression equation.

$$
\begin{aligned}
\begin{bmatrix} s_{US,t} \\ s_{DE,t} \end{bmatrix} &= \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \sum_{i=1}^{5} \mathbf{B}_i \begin{bmatrix} s_{US,t-i} \\ s_{DE,t-i} \end{bmatrix} + \boldsymbol{\varepsilon}_t \qquad (4.2) \\
&= \begin{bmatrix} -0.04884 \\ \text{[-0.855]} \\ 0.06622 \\ \text{[1.547]} \end{bmatrix} + \begin{bmatrix} 1.22048 & -0.38158 \\ \text{[16.123]} & \text{[-3.758]} \\ -0.05904 & 0.91054 \\ \text{[-1.041]} & \text{[11.966]} \end{bmatrix} \begin{bmatrix} s_{US,t-1} \\ s_{DE,t-1} \end{bmatrix} \\
&+ \begin{bmatrix} -0.19972 & 0.61329 \\ \text{[-1.708]} & \text{[4.745]} \\ 0.15486 & -0.04925 \\ \text{[1.767]} & \text{[-0.508]} \end{bmatrix} \begin{bmatrix} s_{US,t-2} \\ s_{DE,t-2} \end{bmatrix} \\
&+ \begin{bmatrix} 0.16175 & -0.09901 \\ \text{[1.371]} & \text{[-0.748]} \\ -0.23489 & 0.18946 \\ \text{[-2.656]} & \text{[1.910]} \end{bmatrix} \begin{bmatrix} s_{US,t-3} \\ s_{DE,t-3} \end{bmatrix} \\
&+ \begin{bmatrix} -0.29238 & -0.11188 \\ \text{[-2.590]} & \text{[-0.842]} \\ 0.17971 & -0.07683 \\ \text{[2.124]} & \text{[-0.772]} \end{bmatrix} \begin{bmatrix} s_{US,t-4} \\ s_{DE,t-4} \end{bmatrix} \\
&+ \begin{bmatrix} 0.08113 & -0.02610 \\ \text{[1.154]} & \text{[-0.253]} \\ -0.01448 & 0.01976 \\ \text{[-0.275]} & \text{[0.256]} \end{bmatrix} \begin{bmatrix} s_{US,t-5} \\ s_{DE,t-5} \end{bmatrix} + \begin{bmatrix} \widehat{\varepsilon}_{1t} \\ \widehat{\varepsilon}_{2t} \end{bmatrix}
\end{aligned}
$$

The estimated covariance matrix of the residuals, $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$, is:

$$
\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} = \begin{bmatrix} 0.08694 & -0.02824 \\ -0.02824 & 0.04883 \end{bmatrix}. \qquad (4.3)
$$

4.7 Granger Causality Analysis of Stationary IYCDs

In this section we focus on interpreting the VAR(5) model for the US and German slopes. We employ the concept of Granger causality to obtain information about the nature of interactions among the US and German slopes and the way they influence each other. Following Corollary 2.2.1 of Lütkepohl, 2005, p. 45, Granger causality can be evaluated by just looking at the VAR representation of the system. In Equation 4.2, the highly significant coefficients of lag 1 and lag 2 of the German slope in the US slope equation and the highly significant coefficient of lag 3 of the US slope in the German slope equation would suggest the presence of causality structure in the system. The estimated covariance matrix of the residuals (Equation 4.3) provides first insights of instantaneous causality.

In the sequel, we test for three types of Granger causality (as described in Section 3.3) in the $s_{US,t}/s_{DE,t}$ system. We test for *no delayed Granger causality* (by including lagged realizations of $s_{US,t}$ in the unrestricted regression of $s_{DE,t}$ and vice versa), *no causality at all*, i.e., no instantaneous and no delayed Granger causality (by including contemporaneous realizations of $s_{US,t}$ in the unrestricted regression of $s_{DE,t}$ and vice versa), and *no delayed but instantaneous Granger causality* (by including contemporaneous realizations of $s_{US,t}$ in both the unrestricted and restricted regressions of $s_{DE,t}$ and vice versa). The results of the tests are reported in Table 4.3.

Table 4.3: Test for Granger causality: $s_{US,t}/s_{DE,t}$ system.

System	Equation	Direction	Type of Causation	p value	Decision
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	$s_{DE,t} \rightarrow s_{US,t}$	Type 1: No delayed G-causality	0.0000	1
$s_{US,t}/s_{DE,t}$	$s_{DE,t}$	$s_{DE,t} \leftarrow s_{US,t}$	Type 1: No delayed G-causality	0.1900	0
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	$s_{DE,t} \rightarrow s_{US,t}$	Type2: No causality at all: no instant. And no delayed G-causality	0.0000	1
$s_{US,t}/s_{DE,t}$	$s_{DE,t}$	$s_{DE,t} \leftarrow s_{US,t}$	Type2: No causality at all: no instant. And no delayed G-causality	0.0000	1
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	$s_{DE,t} \rightarrow s_{US,t}$	Type 3: No delayed but instant. G-causality	0.0000	1
$s_{US,t}/s_{DE,t}$	$s_{DE,t}$	$s_{DE,t} \leftarrow s_{US,t}$	Type 3: No delayed but instant. G-causality	0.0000	1

The p-values smaller than 0.05 suggest rejection of the null hypotheses of no causality and conclusion that there exists causality structure in the $s_{US,t}/s_{DE,t}$ system. The German slope could contain useful information for improving the prediction of the US slope and vice versa.

The existence of causality structure from the German factors to the US ones might be explained by the fact that the *German bund futures market* is the largest futures markets worldwide. Ahn, Cai, and Cheung, 2002.

4.8 Impulse Response Analysis of Stationary IYCDs

In order to understand the complete story about the interactions between the US and German slopes, we trace out the effect of an unexpected shock in one slope on the other. We, therefore, generate impulse responses from the VAR(5) model for the US and German slopes, $s_{US,t}$ vs.

$s_{DE,t}$, in order to assess the dynamic responses to one standard deviation shock in either of the two country factors.

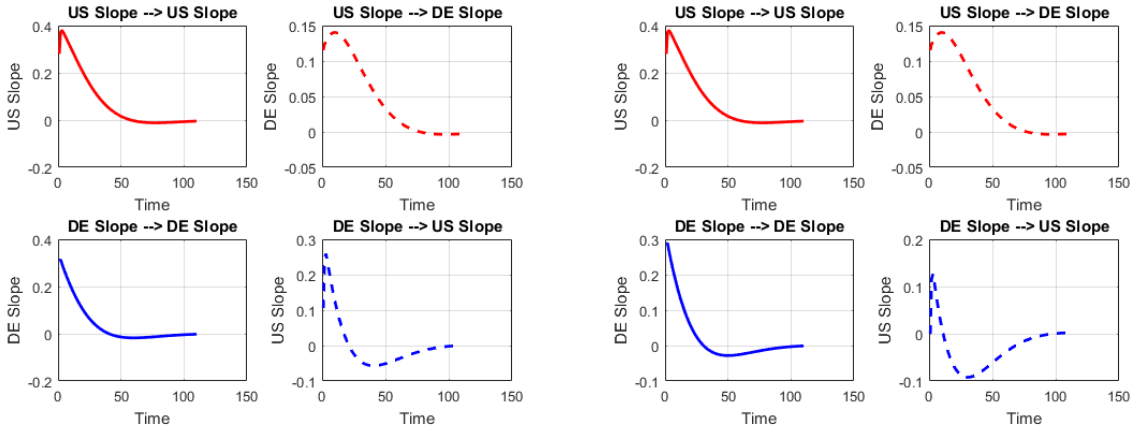
Figure 4.12 shows that a positive one standard deviation shock to the US slope produces an increase in the US slope of almost 0.4%. The increase in the US slope leads to an increase of almost 0.15% in the German slope.

A positive one standard deviation innovation in the German slope corresponds to an increase of almost 0.3% in the German slope. The increase in the German slope leads to an increase of more than 0.25% in the US slope and the responses are relatively short-lived.

Similar results are obtained if the orthogonalized impulse-response method is employed.

Summing up all MA coefficient matrices of the $s_{US,t}/s_{DE,t}$ system, we quantify the accumulated or long-run effects. Figure 4.13 depicts the accumulated responses of the $s_{US,t}/s_{DE,t}$ system, calculated with both the generalized and the orthogonalized methods.

Figure 4.12: Impulse responses of the $s_{US,t}/s_{DE,t}$ system (impulse → response).



(a) Generalized method

(b) Orthogonalized method

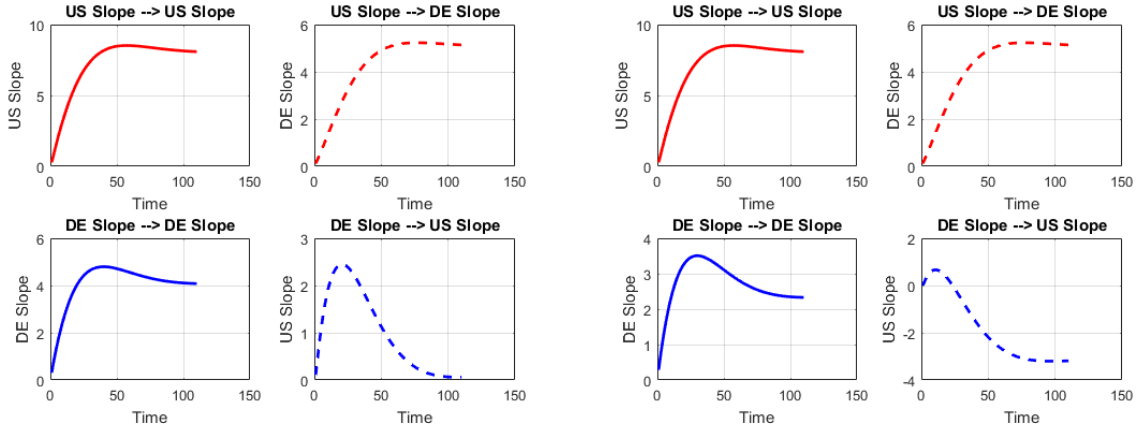
4.9 Cointegration Analysis

We run the cointegration analysis for the estimated country factors that are integrated of order 1, [I(1)], i.e., $l_{US,t}$, $l_{DE,t}$, $c_{US,t}$, and $c_{DE,t}$. More specifically, we investigate two systems for cointegration: the $l_{US,t}/l_{DE,t}$ system and the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system.

In Figure 4.14, we visualize the dynamics of the integrated country factors. In Figure 4.14a, the $l_{US,t}$ and $l_{DE,t}$ stay close to each other even if they "drift about" as individual processes. This is an indication that cointegration might exist. In Figure 4.14b, the two country curvatures tend to follow the dynamics of the country levels, although the curvatures have a significantly higher volatility.

Before applying the Johansen test for cointegration, the optimal lag structure for the VEC(q)

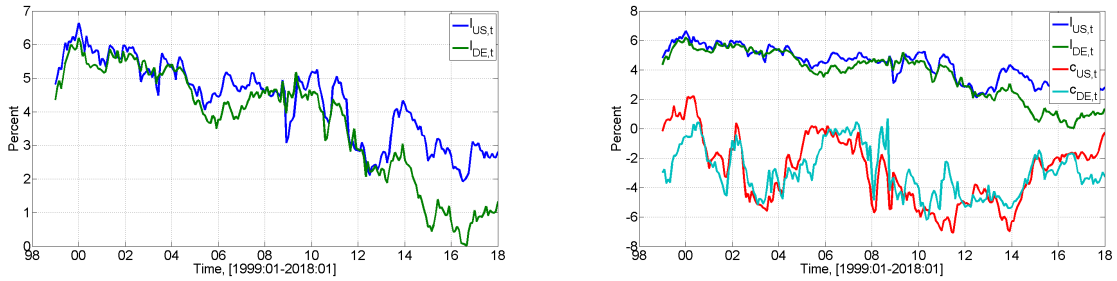
Figure 4.13: Accumulated and long-run responses of the $s_{US,t}/s_{DE,t}$ system (impulse $\rightarrow$ response).



(a) Generalized method

(b) Orthogonalized method

Figure 4.14: US and German integrated country factors, [1999:01-2018:01].



(a) $l_{US,t}/l_{DE,t}$ system

(b) $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system

model is selected based on the equivalent VAR(p) representation of the VEC(q) model[14]. The estimated country factors in Figure 4.14 exhibit a random walk behavior and despite a decreasing trend being visible in the estimated country levels, we do not include a deterministic trend term when estimating the VAR model for the selection of the optimal lag length and include just a constant for both systems of I(1) country factors. The results are reported in Table 4.4.

Table 4.4: Lag-order selection statistics for VECMs.

System	Optimal Lag	Portmanteau Test (asymptotic)
$l_{US,t}/l_{DE,t}$	1	χ^2 = 64.286, df = 56, p-value = 0.2091
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	3	χ^2 = 219.78, df = 192, p-value = 0.08257

For the $l_{US,t}/l_{DE,t}$ system, an optimal lag length of 2 is selected according to Lütkepohl's version of AIC. The p-value of the multivariate Portmanteau and Breusch-Godfrey test for serially correlated errors is 0.2091, hence we fail to reject the null hypothesis for the absence of up to the order 2 of serially correlated disturbances in a stable VAR(2). For cointegration testing, the lag length q is set to 1 ($p = q + 1$), since we are now running the model in first differences. Similar considerations can be made for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, for which an optimal lag length 3 is selected.

Determining Cointegrating Relationships in Estimated Country Factors

Using the information of the optimal lag structure for the VEC(q) models, we now test for the cointegration rank, in order to determine the number of cointegrating relationships in the two systems of I(1) country factors. The Johansen test for cointegration incorporates the testing procedure into the process of model estimation, avoiding in this way conditional estimates. A model specification is required for the cointegrated VAR processes. Among the five cases proposed by Johansen, we choose case $H1^*$, which assumes no intercept in the cointegrating relations and no trends in the data:

$$\Delta x_t = \boldsymbol{\alpha}(\boldsymbol{\beta}' + c_0)x_{t-1} + \sum_{i=1}^{q} \mathbf{B}_i \Delta x_{t-i} + \boldsymbol{\varepsilon}_t. \tag{4.4}$$

This model is appropriate for nontrending data with nonzero mean. The trace test results (for the determination of the cointegration rank, see Section 3.5 for theoretical background) are reported in Table 4.5.

For the $l_{US,t}/l_{DE,t}$ system, the test of 2D-VEC(1) with model $H1^*$ fails to reject a cointegration rank of 1. The trace statistic (15.9133) is below the critical value (20.2619) at the 95% confidence level. The inference is that the US level and the German level have 1 cointegrating relationship.

For the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, the test of 4D-VEC(3) with model $H1^*$ rejects the null hypothesis of no cointegration rank (r $\leq$ 0) at the 95% confidence level. The test for the null hypotheses that r $\leq$ 1, r $\leq$ 2, and r $\leq$ 3, fail to reject the null hypotheses of at most 1, 2, or 3

[14]See Section 3.5.

cointegrating relationships, respectively. For these three tests the trace statistics are below the respective critical values at the 95% confidence level. The inference is that the US level, German level, US curvature, and the German curvature have 3 cointegrating relationships.

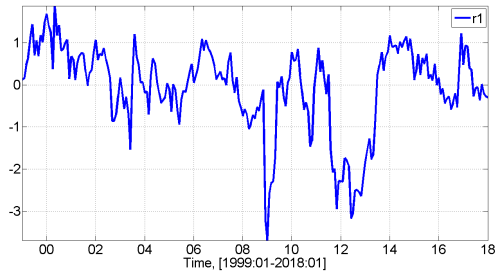
Table 4.5: Johansen cointegration test for I(1) country factors, [1999:01-2018:01].

System	Coint. Relation, r	Decision	Trace Stat	cValue	pValue	eigValue
$l_{US,t}/l_{DE,t}$	r ≤ 0	0	15.9133	20.2619	0.1787	0.0618
$l_{US,t}/l_{DE,t}$	r ≤ 1	0	1.4223	9.1644	0.8869	0.0062
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	r ≤ 0	1	57.5844	54.0779	0.0236	0.1115
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	r ≤ 1	0	30.9969	35.1929	0.1327	0.072
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	r ≤ 2	0	14.1907	20.2619	0.3078	0.0502
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	r ≤ 3	0	2.6027	9.1644	0.6926	0.0115

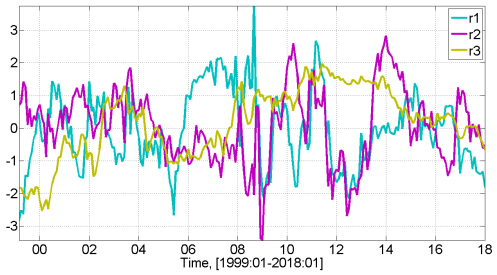
Assessing Stationarity in Cointegrating Relationships of Estimated Country Factors

In Figure 4.15a we plot the Johansen cointegrating relation, $\hat{\beta}' x_t + c_0$, from model $H1^*$, assuming a cointegration rank of 1 (and, thus, 1 cointegrating relation) for the $l_{US,t}/l_{DE,t}$ system. The same plot, in Figure 4.15b, assumes a cointegration rank of 3 (and, thus, 3 cointegrating relations) for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. From the graphical assessment, the cointegrating relations of model $H1^*$ appear to be stationary, for both systems under consideration. Therefore, model $H1^*$ (which assumes that there are intercepts in the cointegrating relations and that there are no trends in the data) is a good representation for the deterministic term in the VEC model for both systems of I(1) country factors.

Figure 4.15: Johansen cointegrating relationships of I(1) country factors, [1999:01-2018:01].

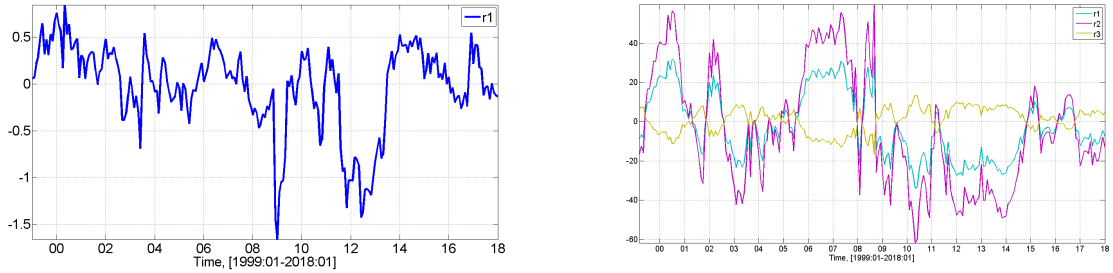


(a) $l_{US,t}/l_{DE,t}$ system



(b) $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system

Figure 4.16: Johansen cointegrating relationships of I(1) country factors, [1999:01-2018:01]. Estimation with $\boldsymbol{\beta}$ coefficients normalized to 1, as per Equations 4.5 and 4.8.



(a) $l_{US,t}/l_{DE,t}$ system

(b) $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system

4.10 Data Generation Processes for Cointegrated IYCDs

The preliminary tests conducted so far provide reasons for the existence of cointegration between the US and German levels and US and German levels and curvatures. We recall that the ADF tests fail to reject the unit root hypotheses for the levels and curvatures, thus suggesting nonstationarity. The nonstationarity is further visible in the slowly-decreasing cross-correlation functions in levels of data. The Johansen tests for cointegration suggest the existence of 1 cointegrating relation for the levels and 3 cointegrating relations for the levels and curvatures.

To analyze the cointegration structure, we fit VEC models to the levels and levels and curvatures. For the $l_{US,t}/l_{DE,t}$ system, we fit a 2D-VEC(1) model with 1 lagged difference[15], an intercept in the cointegrating relation, and no trend in the data. For the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, we fit a 4D-VEC(3) model with 3 lagged differences, an intercept in the cointegrating relation, and no trend in the data.

With respect to the $l_{US,t}/l_{DE,t}$, the ML estimation results are reported in Equation 4.5. t-values are reported in parentheses underneath parameter estimates. The first coefficient of the $\boldsymbol{\beta}$ matrix is normalized to 1.

$$\begin{aligned}\begin{bmatrix}\Delta l_{US,t}\\ \Delta l_{DE,t}\end{bmatrix} &= \boldsymbol{\alpha}\left(\boldsymbol{\beta}'\begin{bmatrix} l_{US,t-1}\\ l_{DE,t-1}\end{bmatrix} + \mathbf{c}_0\right) + \mathbf{B}\begin{bmatrix}\Delta l_{US,t-i}\\ \Delta l_{DE,t-i}\end{bmatrix} + \boldsymbol{\varepsilon}_t \qquad (4.5)\\ &= \begin{pmatrix} -0.120\\ \text{[-3.536]}\\ -0.022\\ \text{[-0.666]}\end{pmatrix}\left(\begin{pmatrix} 1.000 & -0.602\\ \text{[0.000]} & \text{[-8.664]}\end{pmatrix}\begin{bmatrix} l_{US,t-1}\\ l_{DE,t-1}\end{bmatrix} - \begin{pmatrix} 2.141\\ \text{[-7.571]}\end{pmatrix}\right)\\ &+ \begin{pmatrix} 0.0102 & 0.4538\\ \text{[0.147]} & \text{[5.872]}\\ -0.0158 & -0.0366\\ \text{[-0.231]} & \text{[-0.477]}\end{pmatrix}\begin{bmatrix}\Delta l_{US,t-1}\\ \Delta l_{DE,t-1}\end{bmatrix} + \begin{bmatrix}\widehat{\varepsilon}_{1t}\\ \widehat{\varepsilon}_{2t}\end{bmatrix}\end{aligned}$$

[15] As suggested by our statistical procedure for determining the lag length of VEC(q) processes (Section 3.5).

where the estimated covariance matrix of the residuals, $\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}}$, is

$$\widetilde{\boldsymbol{\Sigma}}_{\boldsymbol{\varepsilon}} = \begin{bmatrix} 0.0530 & 0.0293 \\ 0.0293 & 0.0523 \end{bmatrix} \tag{4.6}$$

In the $l_{US,t}$ equation, the term $\boldsymbol{\alpha}_{1,1}\boldsymbol{\beta}'_{1,1}l_{US,t}$ is the lagged error correction term[16]. It is significantly negative ($\boldsymbol{\alpha}_{1,1}\boldsymbol{\beta}'_{1,1}l_{US,t} = -0.120$), representing the negative feedback necessary in the US yield curve level to bring the German yield curve level back to equilibrium. Looking at the short-run matrix $\mathbf{B}$, only the $\mathbf{B}_{1,2} = 0.4538$ is significantly different from zero.

In the $l_{DE,t}$ equation, the lagged error correction term, $\boldsymbol{\alpha}_{2,1}\boldsymbol{\beta}'_{1,2}l_{DE,t} = -0.022 \times -0.602 = 0.0132$, is positive , as it must be for the other variable in the relationship. In other words, if the relationship between the $l_{US,t}$ and $l_{DE,t}$ is above the long-run equilibrium, either the $l_{US,t}$ must fall or the $l_{DE,t}$ must rise. The estimates of the adjustment speeds ($\widehat{\boldsymbol{\alpha}}$) are negative. The significant and negative estimate for the adjustment speed of the $l_{US,t}$ ($\widehat{\boldsymbol{\alpha}}_{1,1} = -0.120$) shows that the $l_{US,t}$ is "caused" by the $l_{DE,t}$.

The estimated Johansen cointegrating relationship, $\widehat{\boldsymbol{\beta}}'x_{t-1} + \mathbf{c}_0$, in the $l_{US,t}/l_{DE,t}$ system is

$$\begin{aligned} ec_t^{ll} &= \widehat{\boldsymbol{\beta}}'x_{t-1} + \mathbf{c}_0 \\ &= l_{US,t} - \underset{[-8.664]}{0.602}\, l_{DE,t} - \underset{[-7.571]}{2.141}. \end{aligned} \tag{4.7}$$

where the superscript ll is short for the $l_{US,t}/l_{DE,t}$ system.

With respect to the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$, the ML estimation results are reported in Equation 4.8. The structure of the impact matrix[17] is such that the US level, US curvature, and German curvature have a negative lagged error correction term, representing the negative feedback necessary to bring the other variables back to equilibrium. The lagged error correction term for the German level (in its own equation) is positive.

An observation that can be made is that there are some insignificant coefficients in the short-run matrices $\mathbf{B}_i$. Given that several parameters in $\widehat{\mathbf{B}}_3$ have rather large t-ratios, reducing

[16]We recall that the component $\boldsymbol{\beta}'x_{t-1}$ reflects common trends present in the system under consideration and it measures the "error" in the data (i.e., the deviation from the stationary mean) at time $t-1$. $\boldsymbol{\beta}$ contains the long-run effects. $\boldsymbol{\alpha}$ contains the loading factors of the common trends. Equivalently, $\boldsymbol{\alpha}$ contains adjustment speeds, i..e, the rate at which the series "correct" from disequilibrium. The component $\boldsymbol{\alpha}\boldsymbol{\beta}'x_{t-1}$ is the error correction term. $\boldsymbol{\alpha}\boldsymbol{\beta}' = \boldsymbol{\Pi}$ is the impact matrix. The impact matrix, $\boldsymbol{\Pi}x_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, for the $l_{US,t}/l_{DE,t}$ is

$$\boldsymbol{\Pi}x_{t-1} = \begin{bmatrix} -0.1200 & 0.0722 \\ -0.0020 & 0.0012 \end{bmatrix} \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \end{bmatrix}.$$

[17]The impact matrix, $\boldsymbol{\Pi}x_{t-1} = \boldsymbol{\alpha}\boldsymbol{\beta}'$, for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ is

$$\boldsymbol{\Pi}x_{t-1} = \begin{bmatrix} -0.1490 & 0.0860 & 0.0190 & -0.0261 \\ -0.0910 & 0.0530 & 0.0070 & 0.0140 \\ 0.1670 & -0.0980 & -0.0030 & -0.0847 \\ 0.1110 & -0.0430 & 0.1210 & -0.2426 \end{bmatrix} \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \\ c_{US,t-1} \\ c_{DE,t-1} \end{bmatrix}.$$

the lag order might not be a good strategy for reducing the number of parameters in the model. It makes sense, however, to impose zero restrictions on some of the parameter values.

$$\begin{bmatrix} \Delta l_{US,t} \\ \Delta l_{DE,t} \\ \Delta c_{US,t} \\ \Delta c_{DE,t} \end{bmatrix} = \boldsymbol{\alpha} \left(\boldsymbol{\beta}' \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \\ c_{US,t-1} \\ c_{DE,t-1} \end{bmatrix} + \mathbf{c}_0 \right) + \sum_{i=1}^{3} \mathbf{B}_i \begin{bmatrix} \Delta l_{US,t-i} \\ \Delta l_{DE,t-i} \\ \Delta c_{US,t-i} \\ \Delta c_{DE,t-i} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{4.8}$$

$$= \begin{pmatrix} -0.149 & 0.086 & 0.019 \\ [-3.963] & [4.022] & [1.766] \\ -0.091 & 0.053 & 0.007 \\ [-2.433] & [2.489] & [0.654] \\ 0.167 & -0.098 & -0.003 \\ [1.851] & [-1.895] & [-0.130] \\ 0.111 & -0.043 & 0.121 \\ [0.970] & [-0.655] & [3.767] \end{pmatrix}$$

$$\left(\begin{pmatrix} 1.000 & 0.000 & 0.000 & 9.739 \\ [0.000] & [0.000] & [0.000] & [3.829] \\ 0.000 & 1.000 & 0.000 & 17.604 \\ [0.00] & [0.000] & [0.000] & [3.895] \\ 0.000 & 0.000 & 1.000 & -4.683 \\ [0.000] & [0.000] & [0.000] & [-5.558] \end{pmatrix} \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \\ c_{US,t-1} \\ c_{DE,t-1} \end{bmatrix} + \begin{bmatrix} 21.723 \\ [2.622] \\ 43.480 \\ [2.954] \\ -10.073 \\ [-3.671] \end{bmatrix} \right)$$

$$+ \begin{pmatrix} -0.025 & 0.511 & -0.106 & 0.121 \\ [-0.304] & [6.136] & [-3.270] & [4.461] \\ 0.091 & -0.048 & -0.026 & 0.072 \\ [1.109] & [-0.570] & [-0.804] & [2.653] \\ 0.048 & 0.079 & 0.267 & 0.097 \\ [0.245] & [0.392] & [3.412] & [1.477] \\ 0.306 & -0.211 & 0.210 & -0.144 \\ [1.234] & [-0.832] & [2.133] & [-1.742] \end{pmatrix} \begin{bmatrix} \Delta l_{US,t-1} \\ \Delta l_{DE,t-1} \\ \Delta c_{US,t-1} \\ \Delta c_{DE,t-1} \end{bmatrix} +$$

$$+ \begin{pmatrix} -0.144 & 0.132 & 0.041 & 0.055 \\ [-1.819] & [1.420] & [1.209] & [1.901] \\ -0.086 & -0.068 & -0.003 & 0.007 \\ [-1.079] & [-0.730] & [-0.095] & [0.253] \\ 0.085 & -0.242 & -0.049 & -0.108 \\ [0.443] & [-1.077] & [-0.604] & [-1.560] \\ 0.027 & -0.320 & -0.048 & 0.052 \\ [0.110] & [-1.131] & [-0.471] & [0.597] \end{pmatrix} \begin{bmatrix} \Delta l_{US,t-2} \\ \Delta l_{DE,t-2} \\ \Delta c_{US,t-2} \\ \Delta c_{DE,t-2} \end{bmatrix} +$$

$$+ \begin{pmatrix} 0.087 & 0.096 & -0.025 & -0.030 \\ [1.297] & [1.071] & [-0.813] & [-1.093] \\ 0.118 & 0.107 & -0.02 & -0.027 \\ [1.764] & [1.196] & [-0.943] & [-0.986] \\ -0.060 & -0.044 & -0.060 & 0.082 \\ [-0.371] & [-0.204] & [-0.808] & [1.249] \\ -0.084 & 0.333 & -0.186 & 0.278 \\ [-0.414] & [1.221] & [-1.991] & [3.369] \end{pmatrix} \begin{bmatrix} \Delta l_{US,t-3} \\ \Delta l_{DE,t-3} \\ \Delta c_{US,t-3} \\ \Delta c_{DE,t-3} \end{bmatrix} + \begin{bmatrix} \widehat{\varepsilon}_{1t} \\ \widehat{\varepsilon}_{2t} \\ \widehat{\varepsilon}_{3t} \\ \widehat{\varepsilon}_{4t} \end{bmatrix}$$

where the estimated covariance matrix of the residuals, $\widetilde{\Sigma}_{\varepsilon}$, is

$$\widetilde{\Sigma}_{\epsilon} = \begin{bmatrix} 0.0446 & 0.0228 & -0.0169 & 0.0019 \\ 0.0228 & 0.0446 & -0.0140 & -0.0354 \\ -0.0169 & -0.0140 & 0.2596 & 0.1546 \\ 0.0019 & -0.0354 & 0.1546 & 0.4125 \end{bmatrix} \tag{4.9}$$

The estimated Johansen cointegrating relationships, $\widehat{\boldsymbol{\beta}}' x_{t-1} + \mathbf{c}_0$, in the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system are

$$\begin{aligned} ec_{1,t}^{llcc} &= \widehat{\boldsymbol{\beta}_1'} x_{t-1} + \mathbf{c}_0 \\ &= l_{US,t} + \underset{[3.829]}{9.739} c_{DE,t} + \underset{[2.622]}{21.723}. \end{aligned} \tag{4.10}$$

$$\begin{aligned} ec_{2,t}^{llcc} &= \widehat{\boldsymbol{\beta}_2'} x_{t-1} + \mathbf{c}_0 \\ &= l_{DE,t} + \underset{[3.895]}{17.604} c_{DE,t} + \underset{[2.954]}{43.480}. \end{aligned} \tag{4.11}$$

$$\begin{aligned} ec_{3,t}^{llcc} &= \widehat{\boldsymbol{\beta}_3'} x_{t-1} + \mathbf{c}_0 \\ &= c_{US,t} - \underset{[-5.558]}{4.683}\, c_{DE,t} - \underset{[-3.671]}{10.073}. \end{aligned} \tag{4.12}$$

where the superscript $llcc$ is short for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system.

4.11 Granger Causality Analysis of Cointegrated IYCDs

To study the presence of Granger causality in the VEC models for the I(1) country factors, we employ the equivalent VAR(p=q+1) of the VEC(q) models and follow the Toda-Yamamoto (TY) approach (as explained in Section 3.6). The results of the TY approach are reported in Table 4.11, for the $l_{US,t}/l_{DE,t}$ system, and in Table 4.11, for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system.

As the TY approach requires augmenting the VAR models with m additional lags (i.e., with the maximum order of integration m for the group time series) in order to fix the asymptotic properties of the Wald test statistics, this information is reported in column "Aug VAR(p+m)", where p is the lag order of the equivalent VAR of the VEC model.

Table 4.6: Test for Granger causality: $l_{US,t}/l_{DE,t}$ system. Toda-Yamamoto approach.

System	Equation	Aug VAR(p+m)	Direction	pValue	stat	cValue	is Rejected?
Type 1: No delayed G-causation							
$l_{US,t}/l_{DE,t}$	$l_{US,t}$	Aug VAR(2+1)	$l_{DE,t} \rightarrow l_{US,t}$	0.0000	39.07	5.992	1
$l_{US,t}/l_{DE,t}$	$l_{DE,t}$	Aug VAR(2+1)	$l_{DE,t} \leftarrow l_{US,t}$	0.752	0.571	5.992	0
Type2: No causation at all: no instant. And no delayed G-Causation							
$l_{US,t}/l_{DE,t}$	$l_{US,t}$	Aug VAR(2+1)	$l_{DE,t} \rightarrow l_{US,t}$	0.0000	153.4	7.815	1
$l_{US,t}/l_{DE,t}$	$l_{DE,t}$	Aug VAR(2+1)	$l_{DE,t} \leftarrow l_{US,t}$	0.0000	98.02	7.815	1
Type 3: No delayed but instant. G-causation							
$l_{US,t}/l_{DE,t}$	$l_{US,t}$	Aug VAR(2+1)	$l_{DE,t} \rightarrow l_{US,t}$	0.0000	32.88	5.992	1
$l_{US,t}/l_{DE,t}$	$l_{DE,t}$	Aug VAR(2+1)	$l_{DE,t} \leftarrow l_{US,t}$	0.0000	51.56	5.992	1

The p-values smaller than 0.05 suggest rejection of the null hypotheses of no causality and conclusion that there exists causality structure in the system. The results show that $l_{DE,t}$ could contain useful information for improving the prediction of $l_{US,t}$ and vice versa. These results are consistent with Johansen cointegration test, which suggested the presence of cointegration structure between US and German levels. Quoting from Toda and Yamamoto, 1995, "*If two or more time series are cointegrated, then there must be Granger causality between them - either one-way or in both directions*". Therefore, the results of causality analysis are in line with the results of cointegration analysis.

Similar results hold for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. However, the null hypothesis of no delayed Granger causality cannot be rejected from $l_{US,t}, l_{DE,t}, c_{DE,t}$, in the $c_{US,t}$ equation, and from $l_{US,t}, l_{DE,t}, c_{US,t}$, in the $c_{DE,t}$ equation.

Table 4.7: Test for Granger causality: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. Toda-Yamamoto approach.

System	Equation	Aug VAR(p+m)	Direction	pValue	stat	cValue	is Rejected?
Type 1: No delayed G-causation							
$l_{US,t}/l_{DE,t}/$ $c_{US,t}/c_{DE,t}$	$l_{US,t}$	Aug VAR(4+1)	$l_{DE,t}, c_{US,t}, c_{DE,t}$ $\rightarrow l_{US,t}$	0.0000	84.4715	21.0261	1

Continued on next page

Table 4.7 – *Continued from previous page*

System	Equation	Aug VAR(p+m)	Direction	pValue	stat	cValue	is Rejected?
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, c_{US,t}, c_{DE,t} \rightarrow l_{DE,t}$	0.0176	24.4617	21.0261	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{DE,t} \rightarrow c_{US,t}$	0.2150	15.5048	21.0261	0
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{US,t} \rightarrow c_{DE,t}$	0.1463	17.0873	21.0261	0
Type2: No causation at all: no instant. And no delayed G-Causation							
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{US,t}$	Aug VAR(4+1)	$l_{DE,t}, c_{US,t}, c_{DE,t} \rightarrow l_{US,t}$	0.0000	133.4024	24.9958	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, c_{US,t}, c_{DE,t} \rightarrow l_{DE,t}$	0.0000	206.4685	24.9958	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{DE,t} \rightarrow c_{US,t}$	0.0000	90.9514	24.9958	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{US,t} \rightarrow c_{DE,t}$	0.0000	116.6907	24.9958	1
Type 3: No delayed but instant. G-causation							
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{US,t}$	Aug VAR(4+1)	$l_{DE,t}, c_{US,t}, c_{DE,t} \rightarrow l_{US,t}$	0.0000	96.7249	21.0261	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, c_{US,t}, c_{DE,t} \rightarrow l_{DE,t}$	0.0000	57.9288	21.0261	1
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{US,t} \rightarrow c_{US,t}$	0.0000	68.5749	21.0261	1
$l_{US,t}/l_{DE,t}/c_{DE,t}/c_{US,t}$	$c_{DE,t}$	Aug VAR(4+1)	$l_{US,t}, l_{DE,t}, c_{US,t} \rightarrow c_{DE,t}$	0.0000	62.9816	21.0261	1

4.12 Impulse Response Analysis of Cointegrated IYCDs

We quantify the effects of unexpected shocks in the two VEC models for the I(1) country factors, by employing the same tools available for stable processes, i.e., the forecast error impulse responses (FEIR) and the orthogonalized impulse responses. The two types of impulse responses are shown in Figure 4.17, for the $l_{US,t}/l_{DE,t}$ system, and in Figures 4.18 and 4.19, for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. Comparing the shapes of the FEIR versus orthogonalized impulse responses, we can notice that the two figures are quite similar (in both systems of I(1) country factors), except for the scaling. This feature could be explained by the almost diagonal residual covariance matrices.

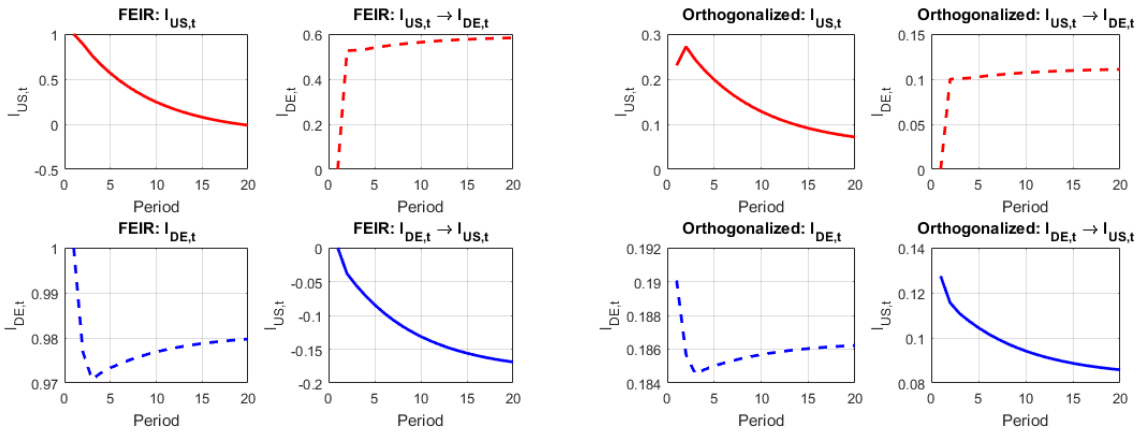
Another feature visible in the figures is that there are cases in which the impulse responses do not die out to zero when increasing the time span after the impulse. The impulse responses,

instead, approach some nonzero value. Such patterns reflect the nonstationarity of the systems, where a one-time impulse can have permanent effects. In other words, the impulses can have permanent effects because $l_{US,t}$, $l_{DE,t}$, $c_{US,t}$ and $c_{DE,t}$ are variables integrated of order one. This conclusion holds even if estimation uncertainty is accounted for.

In the $l_{US,t}/l_{DE,t}$ system, permanent effects are visible for $l_{DE,t}$, after a one standard deviation forecast error impulse in $l_{US,t}$.

In the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, permanent effects are visible for $l_{DE,t}$ and $c_{US,t}$, after a one standard deviation forecast error impulse in $l_{US,t}$; in the $l_{US,t}$, $c_{US,t}$ and $c_{DE,t}$, after a forecast error impulse in $l_{DE,t}$; in the $l_{US,t}$, and $c_{DE,t}$, after a forecast error impulse in $c_{US,t}$; and in $l_{DE,t}$ after a forecast error impulse in $c_{DE,t}$. Similar conclusions hold if the orthogonalized method is employed.

Figure 4.17: Impulse responses of the $l_{US,t}/l_{DE,t}$ system (impulse → response).



(a) Forecast Error IR

(b) Orthogonalized IR

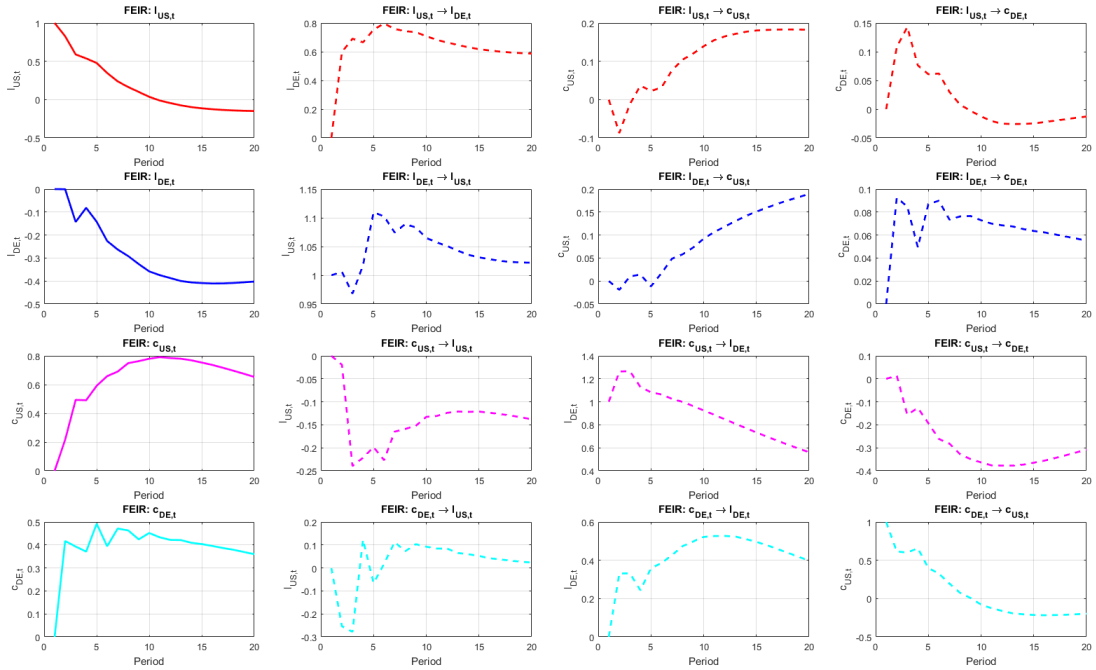


Figure 4.18: Forecast error impulse responses of the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system (impulse $\rightarrow$ response).

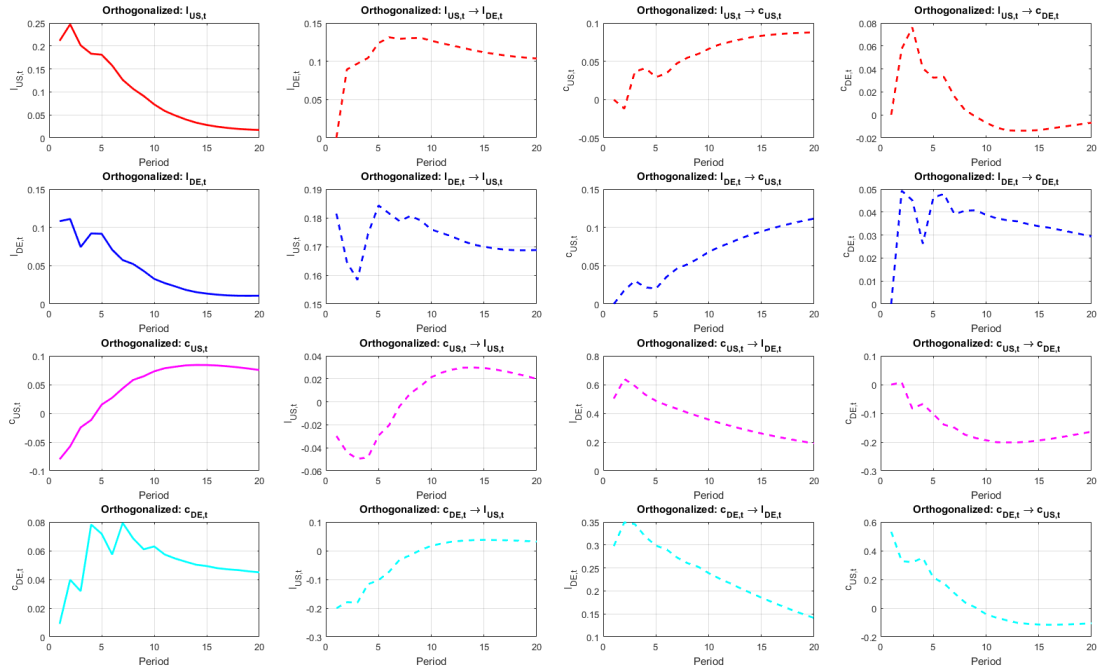


Figure 4.19: Orthogonalized impulse responses of the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system (impulse $\rightarrow$ response).

4.13 Principal Component Analysis

In this section we want to understand the source of variations of IYCD. To do so, we study the covariance structure of IYCD by means of Principal Component Analysis (PCA).

Datasets

We perform PCA on three different datasets. Dataset 1 (D_1) consists of all US and German yields, $y_{it}(\tau)$, where $i = \{\text{US,DE}\}$, $\tau = \{\text{6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y}\}$, and $t = \{\text{1999:01-2018:01}\}$. Hence, for D_1 we have a 14-dimensional vector of multivariate yield observations:

$$\begin{aligned} y_{D_1} &= y_{it}(\tau) \\ &= \left[y^1_{US,t}(6M), y^2_{US,t}(1Y), ..., y^7_{US,t}(10Y), y^8_{DE,t}(6M), y^9_{DE,t}(1Y), ..., y^{14}_{DE,t}(10Y)\right]. \end{aligned} \tag{4.13}$$

We call D_1 the dataset of "All-Yields".

Dataset 2 (D_2) consists of German-US yield curve spreads, $s_{DE-US,t}(\tau)$, defined as $s_{DE-US,t}(\tau) = y_{DE,t}(\tau) - y_{US,t}(\tau)$ where τ and t are defined as for D_1. Hence, for D_2 we have a 7-dimensional vector of multivariate German-US spread observations:

$$
\begin{aligned}
y_{D_2} &= s_{DE-US,t}(\tau) \\
&= \left[s^1_{DE-US,t}(6M), s^2_{DE-US,t}(1Y), s^3_{DE-US,t}(2Y), s^4_{DE-US,t}(3Y), \ldots, s^7_{DE-US,t}(10Y)\right].
\end{aligned} \tag{4.14}
$$

We call D_2 the dataset of "DE-US Spreads".

Dataset 3 (D_3) consists of US and German estimated yield curve level, slope, and curvature, i.e., l_{it}, s_{it}, c_{it} where i and t are defined as for D_1 and D_2. Hence for D_3, we have a 6-dimensional vector of multivariate US and German yield curve factor observations:

$$
y_{D_3} = \left[l^1_{US,t}, l^2_{DE,t}, s^3_{US,t}, s^4_{DE,t}, c^5_{US,t}, c^6_{US,t}\right]. \tag{4.15}
$$

Based on the unit root tests, in D_3, only the slopes are stationary variables. Levels and curvatures are nonstationary, integrated of order 1 [I(1)], and cointegrated variables. Since PCA requires stationary variables, we correct the dataset by dropping the US and German levels and curvatures and replacing them with the four cointegrating relations estimated from the 2D-VEC(1) model (for the US and German levels, Equation 4.5) and from the 4D-VEC(3) model (for the US and German levels and curvatures, Equation 4.8). Hence, for D_3 we have the following 6-dimensional vector of multivariate IYCD observations:

$$
y_{D_3} = \left[s^1_{US,t}, s^2_{DE,t}, ec^{3,ll}_t, ec^{4,llcc}_{1t}, ec^{5,llcc}_{2t}, ec^{6,llcc}_{3t}\right], \tag{4.16}
$$

where ec^{ll}_t is defined as in Equation 4.7 and $ec^{llcc}_{1t}, ec^{llcc}_{2t}, ec^{llcc}_{3t}$ are defined as in Equations 4.10, 4.11, 4.12. We call D_3 the dataset of "Refined All-Factors".

Figure 4.20 shows the graphics of the 3 datasets we use for PCA.

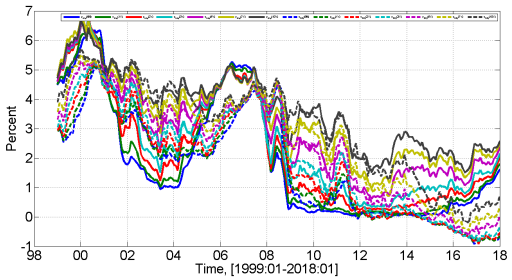
D_1 shall help us identify commonalities in the covariance structure of the high-dimensional dataset of all US and German yields. D_2 shall help us identify the commonalities in the covariance structure of the German-US yield spreads. And finally, D_3 shall help us identify the same information in the covariance structure of the US and German yield curve factors.

Principal Components, Loadings, Explained Variance, and Economic Interpretation

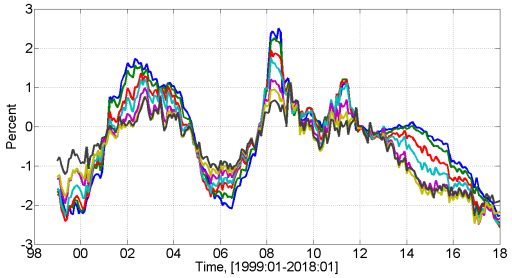
Because of the substantial difference in the variance of different columns, the PCA for the three datasets are run using the inverse variances of the data as weights.

The results of the PCA are contained in Table 4.8, which reports the percentage of total variance of the original data explained by the principal components and the cumulative percentage of total variance. With respect to the "All-Yields" dataset, D_1, the first three principal components ($PC^y_{1,t}, PC^y_{2,t}$, and $PC^y_{3,t}$, where y stands for "yields") explain almost 99.5% of the total variability of the US and German yields. The first principal component alone explains almost 89.7% of the total variability. $PC^y_{1,t}, PC^y_{2,t}$, and $PC^y_{3,t}$ and their corresponding loadings

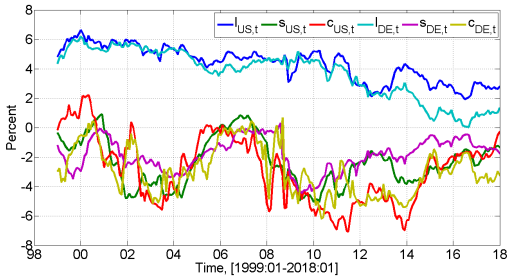
Figure 4.20: Datasets used for PCA, [1999:01-2018:01].



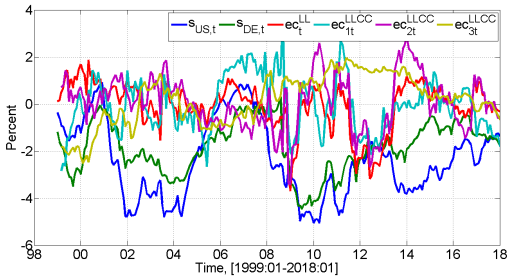
(a) D_1, "All-Yields" dataset



(b) D_2, "DE-US Spread" dataset



(c) All US and German factors



(d) D_3, "Refined All-Factors" dataset

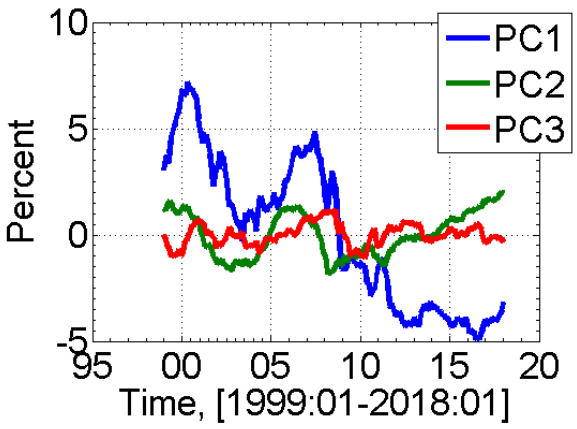
are plotted in Figures 4.21a and 4.21b. In Figure 4.21a, it is clearly visible that $PC^y_{1,t}$ is indeed the principal component with the most structure.

In Table 4.9, we report the pairwise correlation coefficients of the principal components and their underlying datasets. These figures can help us provide an economic interpretation of the principal components. In fact, based on the pairwise correlation coefficients, $PC^y_{1,t}$ is highly correlated with the 10-year US yield, $PC^y_{2,t}$ is highly correlated with the 6-month US yield, and $PC^y_{3,t}$ is highly correlated with the 6-month German yield. Therefore, one could conclude that $PC^y_{1,t}$ and $PC^y_{2,t}$ capture the long- and short-term spectrum, respectively, of the US term structure, whereas, $PC^y_{3,t}$ captures the short-term spectrum of the German term structure.

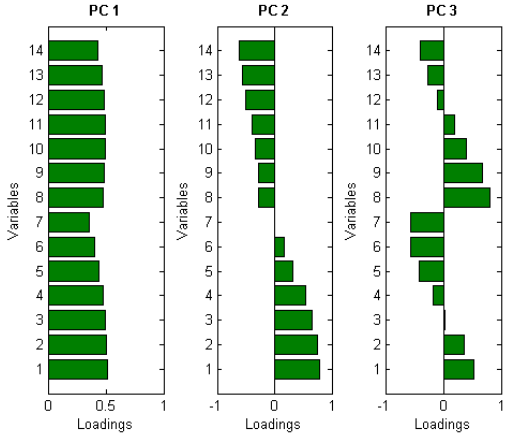
With respect to the "DE-US Spread" dataset, D_2, the first principal component, $PC^s_{1,t}$ (where s stands for "spreads") explains almost 92.5% of the total variability of the German-US spread term structure. The first three principal components explain almost 99.9% of the total variability. The significant structure of $PC^s_{1,t}$ is visible in Figure 4.21c. $PC^s_{1,t}$ seems to capture the 3-year spread, with which the principal component is almost perfectly correlated. $PC^s_{2,t}$ and $PC^s_{3,t}$ capture the short- and long-end of the spread term structure, respectively.

Finally, with respect to the "Refined All-Factors" dataset, D_3, the first five principal components explain almost 99.8% of the total variability of the underlying data. $PC^f_{1,t}$ and $PC^f_{5,t}$ seem to capture the structure of the US slope, $PC^f_{2,t}$ seems to capture the cointegrating relation of the US and German levels, $PC^f_{3,t}$ seems to capture the first cointegrating relation of the US and German levels and curvatures, and finally, $PC^f_{4,t}$ seems to capture the German slope.

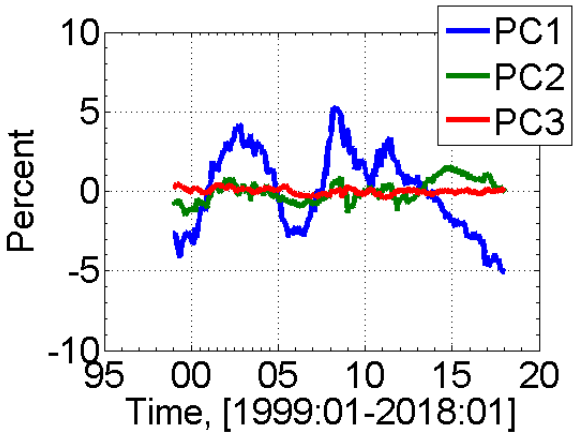
Figure 4.21: Principal Components and their corresponding loadings, [1999:01-2018:01].



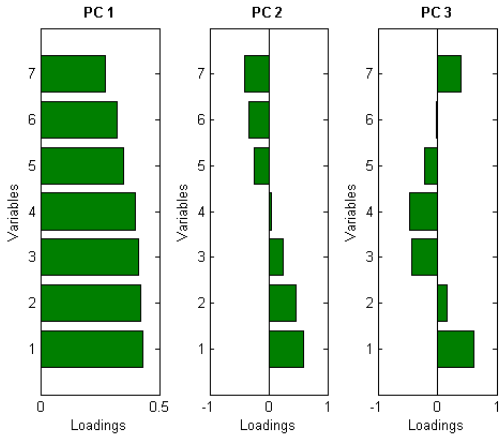
(a) First 3 "All-Yields" PCs



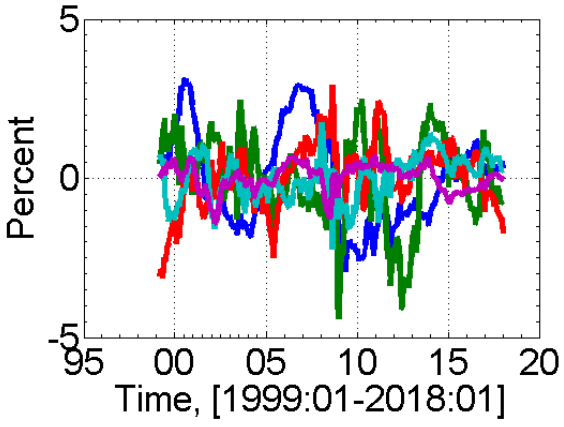
(b) Loadings on first 3 "All-Yields" PCs



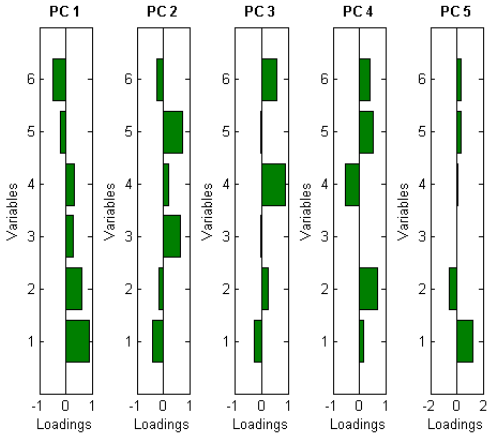
(c) First 3 "DE-US Spread" PCs



(d) Loadings on first 3 "DE-US Spread" PCs



(e) First 5 "Refined All-Factors" PCs



(f) Loadings on first 5 "Refined All-Factors" PCs

Table 4.8: Variance explained by the principal components of all yields, all German-US spreads, and refined factors.

(Percentage of) Total Variance of Original Data Explained by the Principal Components						
Dataset	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6
D_1	89.6949	7.9798	1.8120	0.3683	0.0909	0.0337
D_2	92.4654	6.9224	0.4805	0.0814	0.0267	0.0187
D_3	38.1745	32.4658	17.6347	8.4468	3.1129	0.1654
Cumulative Percentage of Total Variance Explained by the Principal Components						
Dataset	PC_1	PC_2	PC_3	PC_4	PC_5	PC_6
D_1	89.6949	97.6747	99.4867	99.8550	99.9459	99.9795
D_2	92.4654	99.3878	99.8683	99.9497	99.9764	99.9951
D_3	38.1745	70.6403	88.2750	96.7218	99.8346	100.0000

Table 4.9: Pairwise correlation coefficients of the principal components and their underlying datasets.

"All-Yields" PCs								
	$y_{US,t}(6M)$	$y_{US,t}(1Y)$	$y_{US,t}(2Y)$	$y_{US,t}(3Y)$	$y_{US,t}(5Y)$	$y_{US,t}(7Y)$	$y_{US,t}(10Y)$	…
$PC^y_{1,t}$	0.8983	0.9100	0.9310	0.9462	0.9671	0.9705	**0.9739**	**…**
$PC^y_{2,t}$	**0.4081**	0.3973	0.3616	0.3158	0.2073	0.1242	0.0072	…
$PC^y_{3,t}$	0.1296	0.0934	0.0072	-0.0490	-0.1301	-0.1933	-0.2135	…
	$y_{DE,t}(6M)$	$y_{DE,t}(1Y)$	$y_{DE,t}(2Y)$	$y_{DE,t}(3Y)$	$y_{DE,t}(5Y)$	$y_{DE,t}(7Y)$	$y_{DE,t}(10Y)$	
$PC^y_{1,t}$	0.9541	0.9649	0.9724	0.9701	0.9545	0.9348	0.9064	
$PC^y_{2,t}$	-0.1647	-0.1696	-0.1989	-0.2331	-0.2930	-0.3397	-0.3909	
$PC^y_{3,t}$	**0.2314**	0.1910	0.1121	0.0519	-0.0278	-0.0760	-0.1184	
"DE-US Spreads"								
	$s_t(6M)$	$s_t(1Y)$	$s_t(2Y)$	$s_t(3Y)$	$s_t(5Y)$	$s_t(7Y)$	$s_t(10Y)$	
$PC^s_{1,t}$	0.9308	0.9562	0.9833	**0.9957**	0.9813	0.9592	0.9223	
$PC^s_{2,t}$	**0.3512**	0.2892	0.1634	0.0275	-0.1844	-0.2784	-0.3724	
$PC^s_{3,t}$	0.0952	0.0268	-0.0729	-0.0863	-0.0429	-0.0034	**0.0961**	
"Refined All-Factors"								
	$s_{US,t}$	$s_{DE,t}$	ec^{ll}_t	$ec^{llcc}_{1,t}$	$ec^{llcc}_{2,t}$	$ec^{llcc}_{3,t}$		
$PC^f_{1,t}$	**0.8588**	0.8031	0.4278	0.4447	-0.2596	-0.6780		
$PC^f_{2,t}$	-0.3454	-0.2212	**0.9009**	0.2508	0.8992	-0.3108		
$PC^f_{3,t}$	-0.1727	0.2449	-0.0176	**0.7925**	-0.0079	0.5830		
$PC^f_{4,t}$	0.0730	**0.4507**	0.0041	-0.3311	0.3262	0.2869		
$PC^f_{5,t}$	**0.3285**	-0.2071	-0.0072	0.0329	0.1220	0.1411		

Note to table: **bold** values represent the highest positive pairwise correlation.

Interest Rate Risk Management Beyond Duration and Convexity Adjustment

Our results from PCA can find applications in cross-country bond portfolio management in explaining the co-movement of the US and German yield curves and measuring and managing yield curve risk. Parallel and non-parallel changes in the term structure of interest rates give rise to interest rate risk, which may erode the value of a bond position or bond portfolio. When market yields change, the prices of fixed-income instruments are impacted, since these instruments have been priced with the initial yield curve, which then changes. When market interest rates, or yields, increase, the price of fixed-income instruments decreases, and vice versa. Investors typically use the *duration* tool, coupled with a second-order approximation of price changes (*convexity*) to measure interest rate risk. The studies of Ho, 1992; Schumacher, Dektar, and Fabozzi, 1994; Golub and Tilman, 1997a; Golub and Tilman, 1997b; Axel and Vankudre, 2002; Martellini, Priaulet, and Priaulet, 2003; Martellini et al., 2006, among others, show that hedging yield curve risk with principal components is superior to using duration, convexity, or even the widely used measures of the yield curve risk, *key rate duration*.

We recall that *duration hedging* relies on the very restrictive assumption of small and parallel shifts in the yield curve. The technique of *convexity adjustment* is also restrictive since it only relaxes the assumption of small changes but not that of parallel shifts (Fabozzi, Martellini, and Priaulet, 2006). Our empirical analysis of the US and German bond markets suggests that large and non-parallel variations can affect the term structure of interest rates and the spread structure and that three principal components are found to drive the yield and spread curves. These findings strongly support the results of the above-mentioned studies, suggesting that duration hedging and convexity adjustment are inefficient in many cases.

Using the principal components, we can go "beyond duration and convexity adjustment" by relaxing the assumptions of small and parallel changes in the yield curve and describing yield curve risk in terms of principal components. In doing so, we can account for non-parallel deformations of the US and German term structures and systematically cope with the dimensionality problem stemming from multiple sources of uncertainty (i.e., risk factors) affecting the term structures.

In the US and German bond markets, three principal components or risk factors are needed to fully explain the joint dynamics of the two yield curves. The first risk factor brings the highest contribution to the explanatory power (explaining almost 90% of the total variability of the US and German yields) and has a very clear, linear relationship with the long-term US yield (a correlation of +0.97). The second risk factor explains almost 8% of yield variability and it is highly-correlated with the 6-month US yield. The third risk factor has the lowest explanatory power (below 2%) and it is highly-correlated with the 6-month German yield. Since the first three principal components explain almost the entire variability in the US and German yields and they are found to have clear, linear dependence with the short-term and long-term yields, these findings suggest that non-parallel shifts in the US and German yield curves are of critical importance. Clearly the three factors can produce multiple yield curve shapes, i.e., flat, inverted, humped (where both negative and positive butterflies are possible).

To construct hedges that neutralize exposure to changes in the direction of interest rates

and exposure to changes in non-parallel shifts in the yield curve, the factor loadings (displayed in Figure 4.21b) can be used. This is the PCA hedging scheme, which turns out to be a reliable method for all kinds of yield curve scenarios (Fabozzi, Martellini, and Priaulet, 2006; Axel and Vankudre, 2002; Martellini et al., 2006; Golub and Tilman, 1997b).

In the German-US spread market, over all seven maturities, the first principal component explains 92.5% of the spreads, the second principal component 6.9%, and the third principal component 0.5%. The first principal component or risk factor is almost perfectly correlated (with a correlation of +0.99) with the 3-year spread, the second with the 6-month spread (with a correlation of 0.09), and the third one with the 10-year spread (with a correlation of 0.09). The German-US spread market can be fully described in terms of three risk factors, which account for almost the entire variability of the spreads.

4.14 Pool of IYCDs

After completing our econometric workflow described in Section 4.3, we are able to gather the following pool of international yield curve drivers:

$$x \in [l_{US,t}, l_{DE,t}, s_{US,t}, s_{DE,t}, c_{US,t}, c_{DE,t}, ec_t^{ll}, ec_{1t}^{llcc}, ec_{2t}^{llcc}, ec_{3t}^{llcc}, PC_{1,t}^{y}, PC_{2,t}^{y}, PC_{3,t}^{y}, PC_{1,t}^{s}, PC_{2,t}^{s}, PC_{3,t}^{s}, PC_{1,t}^{f}, PC_{2,t}^{f}, PC_{3,t}^{f}, PC_{4,t}^{f}, PC_{5,t}^{f}] \tag{4.17}$$

where $l_{US,t}, l_{DE,t}, s_{US,t}, s_{DE,t}, c_{US,t}, c_{DE,t}$ are the estimated Nelson-Siegel US and German levels, slopes, and curvatures; ec_t^{ll} is the cointegrating relation of $l_{US,t}$ and $l_{DE,t}$; $ec_{1t}^{llcc}, ec_{2t}^{llcc}, ec_{3t}^{llcc}$ are the cointegrating relations of $l_{US,t}$, $l_{DE,t}$, $c_{US,t}$, and $c_{DE,t}$; $PC_{1,t}^{y}, PC_{2,t}^{y}, PC_{3,t}^{y}$ are the first three principal components of all US and German yields; $PC_{1,t}^{s}, PC_{2,t}^{s}, PC_{3,t}^{s}$ are the first three principal components of all German-US yield spreads; and $PC_{1,t}^{f}, PC_{2,t}^{f}, PC_{3,t}^{f}, PC_{4,t}^{f}, PC_{5,t}^{f}$ are the first five principal components of the so-called "refined" US and German yield curve factors.

4.15 Conclusion

In this Chapter, we addressed empirically the research questions of what drives the US and German term structure of interest rates and how do the drivers co-move.

We run our empirical analysis on actively traded US and German government bond yields, for the sample period from '1999:01' to '2018:01' and for seven of the most liquid maturities. A preliminary analysis of the US and German yields, German-US spreads, and Nelson-Siegel estimated country factors provided first insights into the dynamic evolution of yield curve variables. The yields and spreads in levels appear highly persistent. The estimated country factors appear to follow a common dynamics. The US and German level move together to some extent and exhibit a lead-lag structure. Similar observations could be deduced for the US and German slopes and curvatures. The correlation analysis showed positive correlations, significantly different from zero, for the US level and curvature and US slope and curvature. The same analysis for Germany showed that all German yield curve drivers have correlations

significantly different from zero. Inter-country factor correlation is significantly different from zero for all pairs of mixed US and German yield curve drivers, except for $s_{US,t} - l_{DE,t}$ and $c_{US,t} - l_{DE,t}$. Because some heteroskedasticity could be visible in the matrix plot of inter-country factor correlation, we investigated the presence of volatility clustering in the residuals of differenced country factors with Engle's ARCH test. The null hypothesis of no conditional heteroskedasticity was rejected only for $\Delta c_{DE,t}$.

After the preliminary analysis, we run the workflow depicted in the diagram in Section 4.3, with the aim of understanding deeper the dynamic properties of US and German yield curve drivers. The theoretical concepts and methods behind our workflow have been thoroughly described in Chapter 3.

The starting point of the workflow consisted in assessing the nonstationary/stationary behavior of yield curve variables via unit root tests. At that step, we concluded that yields, yield spreads, US and German country levels, and US and German country curvatures are nonstationary variables, integrated of order 1. This finding invalidates the Economic Theory which postulates that nominal bond yields cannot be I(1), since they have a lower bound support at zero and an upper bound support lower than infinity.

Further on in our workflow, we run a cross-correlation analysis, with the objective of exploring deeper the commonality in movements and the presence of lead-lag structure between the US and German yield curve drivers. In the cross-correlation plots of German and US country levels, in first differences, we noticed a well-defined peak at the first positive lag, indicating that $\Delta l_{US,t}$ lags one month behind $\Delta l_{DE,t}$. We derived similar conclusions for the slopes, curvatures, and pairs of estimated country factors of mixed classes, where again well-defined peaks at positive lags suggested causal structure from the differenced German factors to the differenced US factors, thus, a leading behavior of the German country factors with respect to the US country factors.

Since the slopes of the US and German term structure were found to be stationary, positively correlated, and potentially causing each other, we had good reasons to fit a VAR(5) model to the two slopes, in order to exploit the linear dependence of the variables on their own lagged values and those of the other variable in the vector and to study the lead-lag relationship between the US and German slopes. The tests for Granger causality failed to find evidence against the existence of causality structure in the $s_{US,t}/s_{DE,t}$ system. We concluded that the German slope could contain useful information for improving the prediction of the US slope and vice versa. We derived the complete story about the interactions between the US and German slopes, by tracing out the effect of an unexpected shock in one slope on the other. A positive standard deviation shock to the US slope produced an increase in the German slope of almost 0.15%. The same shock in the German slope increased the US slope by more than 0.25%.

The ADF tests for the US and German levels and curvatures failed to reject the unit root hypotheses, suggesting nonstationarity. The nonstationarity of the levels and curvatures was also visible in the slowly-decreasing cross-correlation functions in levels of data. Based on the Johansen tests for cointegration, we concluded that the US and German levels and curvatures are cointegrated variables. We found one cointegrating relation in the $l_{US,t}/l_{DE,t}$ system and three cointegrating relations in the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. We analyzed the cointegration structure in a VEC(1) model for the levels and in a VEC(3) model for the levels and curvatures.

In the VEC(1) model for the US and German levels, the lagged error correction component suggested that a negative feedback is necessary in $l_{US,t}$ to bring $l_{DE,t}$ back to equilibrium. In other words, we found that if the relationship between the $l_{US,t}$ and $l_{DE,t}$ is above the long-run equilibrium, either $l_{US,t}$ must fall or $l_{DE,t}$ must rise. The significant and negative estimate for the adjustment speed of $l_{US,t}$ showed that $l_{US,t}$ is "caused" by $l_{DE,t}$. In the VEC(3) model for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, we found that a negative feedback is necessary in $l_{US,t}$, $c_{US,t}$ and $c_{DE,t}$ to bring the other variables in the system back to equilibrium.

In the two VEC models, we studied the presence of Granger causality following the Toda-Yamamoto approach, to find that $l_{DE,t}$ could contain useful information for improving the prediction of $l_{US,t}$ and vice versa. We arrived to similar results in the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. From the two estimated VEC models, we derived the impulse response functions. We noticed that the impulse responses do not die out to zero when increasing the time span after the impulse. Such pattern led us to conclude that the systems composed of $l_{US,t}/l_{DE,t}$ and $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ are indeed nonstationary, and that one-time impulses can have permanent effects.

Finally, we performed PCA on all US and German yields, on all German-US spreads, and on stationary country factors and cointegrating relations of nonstationary country factors, in an attempt to improve our understanding of the source of variations in the international yield curve drivers. We found that three principal components can explain the variability in all US and German yields. Similar conclusions were reached for the spreads. For the stationary country factors and cointegrating relations, five principal components are needed. Based on the pairwise correlation coefficients of the principal components and their underlying datasets, we provided an economic interpretation of the principal components and discussed how the principal components can be used to manage interest rate risk beyond duration and convexity, in the so-called principal component hedging scheme.

Part II

Outliers and Structural Breaks

Outliers and Structural Breaks

Part II of this thesis is dedicated to a comprehensive econometric study of outliers and structural breaks in the dynamics of US and German yield curve drivers. Part II consists of Chapters 5, 6, and 7.

In Chapter 5, we start our study with a univariate analysis of structural breaks in the 2D-VAR(5) model for the slopes and the 4D-VEC(3) model for the levels and curvatures. Using the methods of Bai and Perron, 1998 and Perron and Zhou, 2008, we test for the presence of multiple structural breaks of unknown timing to find evidence supporting the existence of breaks in all drivers.

In Chapter 6, we question the nature of structural breaks by conducting a study of the US Fed and ECB monetary policy predictability. The assumption we want to investigate is whether the 2008 Financial Crisis signed a monetary policy regime change and, therefore, a change in the ability of market participants in predicting monetary policy decisions. Using 18 years of money market and policy rates, we document the transition of the Fed and the ECB from a traditional to a more accommodative and nontraditional monetary policy. This transition is signed by the 2008 Financial Crisis and corresponds to an increased predictability of Fed's and ECB's actions by market participants. These findings provide good reasons to believe that the root causes of structural breaks are linked to a change in monetary policy regimes and increased predictability of Central Banks.

In Chapter 7, we investigate whether the presence of structural breaks is due to variables with predictive power missing in the univariate dynamics of the US and German yield curve drivers. To verify this assumption, we adopt a multivariate state-space setting and develop a new data-driven state-space model, the FSSM, for the co-movement of US and German yield curve drivers. The novelty of the FSSM is that it is designed to preserve the dynamic properties of the yield curve drivers embodied in their underlying data generation processes. We test for the presence of outliers and structural breaks in the FSSM to find that the structural alterations resemble of patches of outliers rather of structural breaks. We explain how to adjust the FSSM for the most blatant outliers by including intervention variables in the measurement equation. We call this new version of the FSSM the MShock-FSSM.

Chapter 5

Univariate Analysis

5.1 Introduction

Constancy of parameters in dynamic econometric models is a necessary condition for accurate forecasting and reliable econometric inference. Very often, however, economic time series exhibit "regime shifts" or "structural breaks" that undermine the assumed property of stationary parameters. For instance, economic time series may exhibit changes in the serial correlation, mean, and volatility, and these changes might be due to sudden and unexpected external events in particular time periods, thus, not easy to handle with simple transformations, Lütkepohl, 2005.

Because model stability and, hence, stationarity is an important assumption in time series modeling, in the past decades, a large literature has emerged developing tests for structural breaks. The bibliography of Hackl and Westlund, 1989 lists more than 500 studies, most of which revolve around the classical Chow, 1960 Breakpoint Test. The test, somehow restrictive, is designed to test the null hypothesis of stationary parameters against an alternative of a one-time shift in the parameters at some known time. As in most empirical applications, the breakdate is not known a priori, the solution is to employ the Quandt, 1960 test, by taking the largest Chow statistic over all possible breakdates. In this setting, significance should be assessed with Andrews, 1993 and Andrews and Ploberger, 1994 asymptotic critical values. Theoretical and computational extensions of the Quandt-Andrews framework have been put forward by Bai, 1997a; Bai and Perron, 1998; Bai and Perron, 2003; and Perron, 2006, who allow for testing for multiple unknown breakpoints.

In econometric practice, the Quandt-Andrews and Andrews-Ploberger family of statistics are very popular. Comprehensive applications include Stock and Watson, 1996; Ben-David and Papell, 1998; and McConnell and Perez-Quiros, 2000. In many applications, the breakdate is an unknown parameter and can be estimated applying the theory of least squares estimation developed by Bai, 1994 and Bai, 1997b. Chong, 1995 discusses the consequences of underspecifying the number of change points in a simple structural change model and shows how to estimate multiple breakdates sequentially. Bai, 1997a investigates sequential estimation of multiple breaks and finds that reestimation of breakdates based on refined samples can lead to important

computational improvements.

Despite the wide range of tools and procedures available for detecting structural breaks, past literature contains very few empirical studies of structural breaks in yield curves and yield curve variables. Estrella, Rodrigues, and Schich, 2003, drawing on evidence from Germany and the United States, perform break testing to examine whether the empirical relationships between the slope of the yield curve, real activity, and inflation is stable over time. Rapach and Wohar, 2005 employ the Bai-Perron (1998) sequential method to test for multiple structural breaks in the mean real interest rate and mean inflation rate of 13 industrialized countries. Their empirical results confirm the existence of structural breaks and a coincidence of breaks among the inflation and real interest rates. Barassi, Caporale, and Hall, 2005 detect structural breaks in the causal linkages that generate the cointegrating relations between the G-7 short-term rates. Their evidence supports the hypothesis of a break in the causal linkages between the UK and other EU countries after the third-fourth quarter of 1992, a world-wide leadership position of the US, and a weak leadership position of Germany within the Eurozone. Schrimpf and Wang, 2010 study whether the yield spread may still be considered a predictor of real activity in the presence of structural change. Using test for multiple structural breaks on yield data for Canada, Germany, the UK, and the US, Schrimpf and Wang find evidence that the power of the yield curve in predicting the output growth has been decreasing in recent years. Aguiar-Conraria, Martins, and Soares, 2012 employ cross-wavelet tools, such as coherency, with bootstrap intervals, and phase difference, to study the relation between the level, slope, and curvature of the US yield curve and macroeconomic activity, unemployment, inflation, and the policy rate for the time period between early 1960s and 2009. The authors find a clear structural break in the second half of the 1980s in the relation between the US yield curve slope and real economic activity, a structural break in the late 1980s/early 1990s in the relation between the slope and inflation, and no systematic pattern in the relation between the curvature and economic activity.

To the best of our knowledge, the presence of unusual behavior or structural breaks in the relations between the yield curve drivers of different world regions has not yet been studied. Popular global yield curve models, such as the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" Model, extended to the global context by Diebold, Li, and Yue, 2008, assume parameter stability and fit the global yield curve factors to a VAR(1) process. Disregarding the existence of potential structural breaks might undermine the model's forecasting accuracy and lead to unreliable inference. The goal of this Chapter is to contribute to the existing literature with a comprehensive study of structural breaks in the data generation processes of the US and German yield curve drivers. Adopting a univariate setting, that is, going equation-wise in the 2D-VAR(5) model for the US and German slopes and the 4D-VEC(3) model for the US and German levels and curvatures, we test for the presence of structural breaks in the sample period [1999:01-2018:01] using the methods of Bai and Perron, 1998 and Perron and Zhou, 2008. The decision of which of the two methods to apply is based on whether the Chow test model assumptions of normal, serially uncorrelated, and homoskedastic errors are satisfied. As such, our study employs the most suitable tools for the detection of structural breaks and estimation of their timing by taking into account the dynamic properties of the data.

The Chapter proceeds as follows. Section 5.2 explains the theoretical concepts behind the sequential method of Bai and Perron, 1998 to test for multiple structural breaks of unknown

timing. This method is suitable for detecting structural breaks and estimating their timing only for the cases where the Chow test model assumptions are satisfied. For the cases where the Chow test model assumptions are not satisfied, Section 5.3 explains the Perron and Zhou, 2008 battery of tests to detect jointly structural changes in the regression coefficients and error variance. Section 5.4 reports our empirical results from applying the Bai-Perron (1998) and Perron-Zhou (2008) procedures to the US and German yield curve drivers. Section 5.5 concludes the Chapter.

5.2 Testing for Multiple Structural Breaks of Unknown Timing

The Chow, 1960 Breakpoint Test is the classical test used to check whether the parameters of the model are stable over time, or equivalently, across various subsamples of the data. Assuming Gaussian errors with mean zero and constant variance, σ^2, the test consists in checking for differences between two or more regressions. This is done by performing an F test defined by:

$$F_{m,n_1+n_2-2m} = \frac{[S_1 - (S_2 + S_3)]/m}{[(S_2 + S_3)/(n_1 + n_2 - 2m)]}. \tag{5.1}$$

where n_1 and n_2 are the sizes of the two sub-samples, S_1 is the sum of squared residuals from running a regression on the combined sample, S_2 and S_3 are the sums of squared residuals of the two sub-samples, and m is the number of parameters being estimated, including the intercept.

If the F statistic exceeds the critical F, we reject the null hypothesis that the two regressions are equal.

In the sequel, we choose to work in a setting that allows us to test for multiple structural breaks of unknown timing and estimate the timing of the structural breaks. The workflow we use is built on the sequential method proposed by Bai and Perron, 1998 and, more specifically, consists in the following. We start by testing for a single structural break of unknown timing. As the Chow test requires that the breakdate be known a priori, we employ the solution proposed by Quandt, 1960, by taking the largest Chow statistic over all possible breakdates[1].

Formally, the Quandt test for structural change reads as follows. An $m \times 1$ parameter β, describing some aspect of a time series ${x_t}$[2], takes the value β_1, for $t < k$ and the value β_2 for $t \geq k$, where $m \leq k \leq n - m$. Let $F_n(k)$ denote a Wald, Lagrange multiplier (LM), or likelihood ratio statistic of the hypothesis of no structural change $(\beta_1 = \beta_2)$ for given k, where k denotes the date of structural change. When k is known to lie in the range $[k_1, k_2]$, the Quandt or "Sup" test statistic[3] is defined as:

$$\text{Sup}F_n = \sup_{k_1 \leq k \leq k_2} F_n(k). \tag{5.2}$$

[1]Breakdates too close to the beginning or end of the sample cannot be considered, as there are not enough observations to identify the subsample parameters. The common convention suggests to consider all breakdates in the interior τ percent to $(1 - \tau)$ percent of the sample. In our analysis, we use a τ = 5% trimming.

[2]The notation follows Hansen, 1997.

[3]Variations of the "Sup" test statistic include the "Exp" and "Ave" test statistics presented in Andrews and Ploberger, 1994.

As in Hansen, 2001, we report the Quandt statistic visually by plotting the sequence of Chow statistics as a function of candidate breakdates[4]. We calculate the Quandt statistic by performing single Chow Breakpoint Tests at every observation between k_1 and k_2. If the true parameters are constant, the subsample estimates should be constant across candidate breakdates. If, instead, there is a structural break, the subsample estimates will exhibit a systematic variation across candidate breakdates, and this behavior will be visible in the Chow test sequence.

Since we choose to treat the breakdate as unknown a priori, the χ^2-square critical values are inappropriate for deciding the outcome of the "Sup" test statistic. Andrews, 1993 and Andrews and Ploberger, 1994 show that, under a wide set of regularity conditions, the "Sup" statistic has an asymptotic null distribution:

$$\text{Sup}F_{n\to d}\text{Sup}F(\pi_0) = \sup_{\pi_1\leq\tau\leq\pi_2} F(\tau), \tag{5.3}$$

[5] based on which, Andrews, 1993 tabulated a selected set of asymptotic critical values, which can be used to assess significance of the Quandt statistic when the breakdate is unknown a priori[6].

On the lines of Hansen, 2001, we assess significance of the Quandt statistic in a visual way, by checking whether the Chow test sequence breaks above the critical value. If this happens, the hypothesis of no structural break is easily rejected[7].

After testing for a single structural break with the use of the Quandt statistic and Andrews asymptotic critical values, if the conclusion is to reject the null hypothesis of no structural break, the sequential method of Bai-Perron suggests to split the sample further in two and reapply the test to each subsample. The sequential exercise continues until each subsample test fails to find evidence of a break. The sample is usually split at the breakdate estimate. Following Bai, 1994 and Bai, 1997b, the breakdate estimate can be obtained by the method of least squares. More specifically, the least squares breakdate estimate is obtained by splitting the sample at each possible breakdate, estimating the parameters by ordinary least squares and calculating and storing the sum of squared errors. The least squares breakdate estimate is the date that minimizes the full-sample sum of squared errors, or equivalently, the date that minimizes the

[4]The candidate breakdates are along the $x-$axis and the values of the Chow statistic along the $y-$axis.

[5] In 5.3,

$$F(\tau) = \frac{(W(\tau) - \tau W(1))' (W(\tau) - \tau W(1))}{\tau(1-\tau)}, \tag{5.4}$$

$W(\tau)$ is an $m \times 1-$vector Brownian motion, $\pi_1 = \frac{k_1}{n}$, and $\pi_2 = \frac{k_2}{n}$. These asymptotic distributions of the tests are nonstandard and depend on two parameters: the number of parameter tested, m, and the range of the sample, π_1 and π_2, which is examined for the break date. More specifically, the distributions depend on π_1 and π_2 through the single index

$$\pi_0 = \frac{1}{1+\sqrt{\lambda_0}}, \qquad \text{where} \qquad \lambda_0 = \frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}. \tag{5.5}$$

[6]From Andrews, 1993, one can retrieve the asymptotic critical values for the "Sup" test and from Andrews and Ploberger, 1994, one can retrieve the asymptotic critical values for the "Exp" test ($c = \infty$) and "Ave" test ($c = 0$).

[7]More informed conclusions about the tests of structural breaks can be reached by calculating the asymptotic p-values of the tests as proposed by Hansen, 1997.

residual variance. The assessment can be done visually by plotting the residual variance[8] as a function of the breakdates. The expected patterns are the following. If the true parameters are constant, the subsample estimates (and hence, the sum of squared errors) exhibit a random and erratic behavior across candidate breakdates. If the true parameters are not constant, and therefore, a structural break occurs, the subsample estimates register a well-defined minimum near the true breakdate.

5.3 Testing Jointly for Structural Change in the Regression Coefficients and Error Variance

The standard Chow-type tests for structural change rely on the assumptions of normal, serially uncorrelated, and homoskedastic errors. If a linear model violates these assumptions, the Chow test result might not be correct, or the Chow test might lack power. Perron and Zhou, 2008 provide a comprehensive workflow for testing jointly for structural change in both the regression coefficients and the variance of the errors. Within their workflow, the errors can be non-normal, serially correlated and heteroskedastic.

Since in many empirical applications, changes in the regression coefficients and residual variance may occur at *different dates*, Bai and Perron, 2003 and Perron and Zhou, 2008 propose the following *general to specific* type of sequential procedure to determine the appropriate number and types of breaks. The procedure is called "*general to specific*", in the sense that, one should first start with a "double maximum" test, *UDmax*, to test the null hypothesis of no structural breaks against an unknown number of breaks, given some upper bound for each of the regression coefficients and variance. If the *UDmax* test indicates the presence of at least one break, the number of breaks can be decided based on a sequential examination of the sup$Seq_T(l+1|l)$ statistics, which test the null hypothesis of l breaks versus the alternative of $l+1$ breaks.

Analytically, let m and n denote the number of breaks in coefficients and variance, respectively; M and N the upper bound for number of breaks in coefficients and variance, respectively. The following testing problems (TP)[9] need to be carried out:

1. (TP-8) $H_0 : \{m = n = 0\}$ versus $H_1 : \{1 \leq m \leq M, 1 \leq n \leq N\}$, to test the null hypothesis of no structural breaks against an unknown number of breaks, given the upper bound M and N for each of the regression coefficients and variance, respectively. The test statistic is the *equal-weight double maximum test*[10] given by

$$\begin{aligned} \text{UDmax}LR^*_{4,T} &= \max_{1\leq n_a\leq N} \max_{1\leq m_a\leq M} \text{sup}LR^*_{4,T}(m_a, n_a, \epsilon | n = m = 0) \\ &\Rightarrow \max_{1\leq n_a\leq N} \max_{1\leq m_a\leq M} H_{c,v}(m_a, n_a). \end{aligned} \tag{5.6}$$

[8] The reidual variance is calculated as the sum of squared errors divided by the sample size.

[9] The complete list of testing problems, for which Perron and Zhou, 2008 derive asymptotic statistics, is provided in Section B.1.

[10] See Theorem 2 in Perron and Zhou, 2008 for the relevant assumptions.

where sup$LR^*_{4,T}$ is the modified sup-Likelihood ratio test[11][12] with asymptotic distribution free of nuisance parameters given by

$$\begin{aligned}\text{sup}LR^*_{4,T} \Rightarrow & \sup_{(\lambda^c_1,\ldots,\lambda^c_{m_a};\lambda^v_1,\ldots,\lambda^v_{n_a})\in\Lambda_\epsilon} \left[\begin{array}{l} \sum_{j=1}^{m_a} \frac{||\lambda^c_j W_q(\lambda^c_{j+1})-\lambda^c_{j+1}W_q(\lambda^c_j)||^2}{\lambda^c_{j+1}\lambda^c_j(\lambda^c_{j+1}-\lambda^c_j)} \\ +\sum_{i=1}^{n_a} \frac{(\lambda^v_i W(\lambda^v_{i+1})-\lambda^v_{i+1}W(\lambda^v_i))^2}{\lambda^v_{i+1}\lambda^v_i(\lambda^v_{i+1}-\lambda^v_i)} \end{array}\right] \\ & \equiv H_{c,v}(m_a,n_a).\end{aligned} \tag{5.7}$$

If the null hypothesis of (TP-8) cannot be rejected, (TP-9) needs to be carried out to assess whether too few coefficient breaks are included:

2. (TP-9) $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a + 1, n = n_a\}$. The test statistic reads as follows:

$$\begin{aligned}\text{sup}Seq_T(m+1,n|m,n) = 2[\max_{1\le j\le m+1}\sup_{\tau\in\Lambda^c_{j,\epsilon}} & \log\hat{L}_T(\widetilde{T}^c_1,\ldots,\widetilde{T}^c_{j-1},\tau,\widetilde{T}^c_j,\ldots,\widetilde{T}^c_m,\widetilde{T}^v_1,\ldots,\widetilde{T}^v_n) \\ & -\log\hat{L}_T(\widetilde{T}^c_1,\ldots,\widetilde{T}^c_m;\widetilde{T}^v_1,\ldots,\widetilde{T}^v_n)],\end{aligned} \tag{5.8}$$

where $\Lambda^c_{j,\epsilon} = \{\tau : \widetilde{T}^c_{j-1} + (\widetilde{T}^c_j - \widetilde{T}^c_{j-1})\epsilon \le \tau \le \widetilde{T}^c_j - (\widetilde{T}^c_j - \widetilde{T}^c_{j-1})\epsilon\}$[13].

If the null hypothesis of (TP-8) cannot be rejected, (TP-10) needs to be carried out to assess whether too few variance breaks are included:

3. (TP-10) $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a + 1\}$. The test statistic reads as follows:

$$\begin{aligned}\text{sup}Seq_T(m,n+1|m,n) = \frac{2}{\hat{\psi}}[\max_{1\le j\le n+1}\sup_{\tau\in\Lambda^v_{j,\epsilon}} & 2\log\hat{L}_T(\widetilde{T}^c_1,\ldots,\widetilde{T}^c_m,\tau,\widetilde{T}^v_1,\ldots,\widetilde{T}^v_{j-1},\widetilde{T}^v_1,\ldots,\widetilde{T}^v_m) \\ & -\log\hat{L}_T(\widetilde{T}^c_1,\ldots,\widetilde{T}^c_m;\widetilde{T}^v_1,\ldots,\widetilde{T}^v_n)],\end{aligned} \tag{5.9}$$

where $\Lambda^v_{j,\epsilon} = \{\tau : \widetilde{T}^v_{j-1} + (\widetilde{T}^v_j - \widetilde{T}^v_{j-1})\epsilon \le \tau \le \widetilde{T}^v_j - (\widetilde{T}^v_j - \widetilde{T}^v_{j-1})\epsilon\}$[14].

In addition to (TP-9) and (TP-10), (TP-3) and (TP-2) can be carried out to test the following.

4. (TP-3) $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a\}$. This is the testing problem where there are n_a breaks in the variance under both the null and the alternative

[11] In 5.7, Λ_ϵ denotes the search set for possible values of the break fractions in coefficients (i.e., the $(\lambda^c_1,\ldots,\lambda^c_{m_a})$) and variance (i.e., the $(\lambda^v_1,\ldots,\lambda^v_{n_a})$), ϵ is the trimming parameter, and $W(\cdot)$ are Wiener processes.

[12] Asymptotic critical values of the sup$LR^*_{4,T}$ test are reported in Perron and Zhou, 2008.

[13] $\{\widetilde{T}^c_1,\ldots,\widetilde{T}^c_m\}$ are the estimates of the break dates in the regression coefficients.

[14] $\{\widetilde{T}^v_1,\ldots,\widetilde{T}^v_n\}$ are the estimates of the break dates in the variance of the errors.

hypotheses so that the test boils down to assessing whether there are 0 or m_a breaks in the regression coefficients. The test statistic[15][16] reads as follows:

$$\begin{aligned} \sup LR_{3,T}(m_a, n_a, \epsilon | m = 0, n_a) &\Rightarrow \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c)\in\Lambda_{c,\epsilon}^v} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &\equiv H_c^*(m_a) \\ &\leq \sup_{(\lambda_1^c,\ldots,\lambda_{m_a}^c)\in\Lambda_{c,\epsilon}} \sum_{j=1}^{m_a} \frac{||\lambda_j^c W_q(\lambda_{j+1}^c) - \lambda_{j+1}^c W_q(\lambda_j^c)||^2}{\lambda_{j+1}^c \lambda_j^c (\lambda_{j+1}^c - \lambda_j^c)} \\ &\equiv H_c(m_a), \end{aligned} \tag{5.10}$$

where

$$\begin{aligned} \Lambda_{c,\epsilon}^v = \{(\lambda_1^c, \ldots, \lambda_{m_a}^c) : \quad for \quad (\lambda_1, \ldots, \lambda_K) = (\lambda_1^c, \ldots, \lambda_{m_a}^c) \cup (\lambda_1^{0v}, \ldots, \lambda_{n_a}^{0v}) \\ |\lambda_{j+1} - \lambda_j| \geq \epsilon (j = 1, \ldots, K-1), \lambda_1 \geq \epsilon, \lambda_K \leq 1 - \epsilon\} \end{aligned} \tag{5.11}$$

$$\Lambda_{c,\epsilon} = \{(\lambda_1^c, \ldots, \lambda_{m_a}^c) : |\lambda_{j+1}^c - \lambda_j^c| \geq \epsilon (j = 1, \ldots, m_a - 1), \lambda_1^c \geq \epsilon, \lambda_{m_a}^c \leq 1 - \epsilon\}.$$

5. (TP-2) $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$. This is a testing problem where there are m_a breaks in the regression coefficients under both the null and the alternative hypotheses so that the test boils down to assessing whether there are 0 or n_a breaks in the variance. The test statistic reads as follows:

$$\begin{aligned} \sup LR_{2,T}^* &= (2/\hat{\phi}) \sup LR_{2,T} \\ &\Rightarrow \sup_{(\lambda_1^v,\ldots,\lambda_{n_a}^v)\in\Lambda_{v,\epsilon}^c} \sum_{i=1}^{n_a} \frac{(\lambda_i^v W(\lambda_{i+1}^v) - \lambda_{i+1}^v W(\lambda_i^v))^2}{\lambda_{i+1}^v \lambda_i^v (\lambda_{i+1}^v - \lambda_i^v)} \equiv H_v^*(n_a) \leq H_v(n_a). \end{aligned} \tag{5.12}$$

5.4 Empirical Results

5.4.1 Workflow

Our workflow for testing and dating structural breaks is depicted in Figure 5.1 and revolves around the classical Chow, 1960 Breakpoint Test, provided the Chow test model assumptions hold. Before starting the univariate analysis, we get first insights about the presence of structural breaks from a visual investigation of the univariate evolution over time of the IYCDs. In the graphs of US and German levels, slopes, and curvatures, episodes of different mean, persistence, and variability indicate potential structural breaks.

[15] In 5.10 and 5.12, Λ denotes the search set for possible values of the break fractions in coefficients $(\lambda_1^c, \ldots, \lambda_m^c)$ and variance $(\lambda_1^v, \ldots, \lambda_n^v)$. Λ also specifies the trimming parameter ϵ, which affects the limiting distribution of the tests. $W(\cdot)$ denotes Wiener processes and $(\phi/2)$ a scaling factor, estimated to $\hat{\phi}$. For details, see Perron and Zhou, 2008, p. 14.

[16] See Theorem 1 in Perron and Zhou, 2008 for the relevant assumptions.

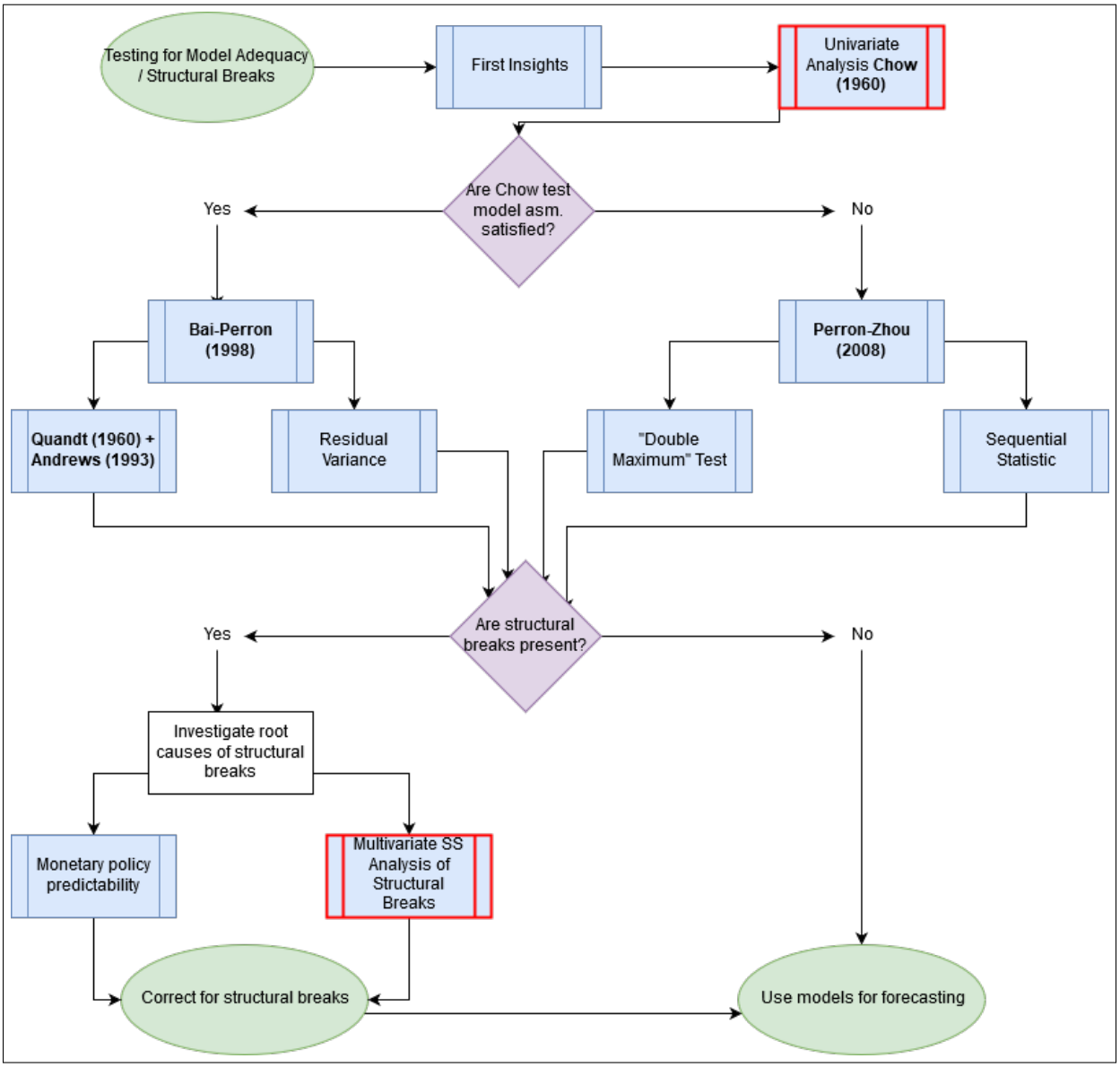
Since the Chow test relies on independent, Gaussian innovations, constancy of innovations variance within subsamples, and constancy of the innovations across any structural breaks, the first step of our workflow consists in checking whether the Chow test model assumptions hold.

We do so by extracting and analyzing the residuals of the estimated linear models. We draw two histogram plots using the residuals: one plot of residuals with respect to fitted values in case (i.e., row) order, and another plot with respect to the previous (i.e., lagged) residual. A visual inspection of these histograms can provide first suggestions about the presence of autocorrelation and/or heteroskedasticity in the residuals. We conduct Engle's ARCH tests to ultimately assess whether the innovations are heteroskedastic and Kolmogorov-Smirnov (KS) tests to assess whether the innovations are Gaussian.

If the Chow test assumptions appear valid, i.e., we fail to reject the null hypothesis of no ARCH effects (Engle's ARCH test) and we fail to reject the null hypothesis that the data come from a standard normal distribution (KS test), we test for multiple structural breaks of unknown timing by following the sequential method of Bai and Perron, 1998. We start the method by testing for a single break of unknown timing. We employ the Quandt statistic, i.e., we take the largest Chow test over all possible breakdates, and assess the significance using the Andrews, 1993 asymptotic critical values. As in Hansen, 2001, we report the Quandt statistic visually by plotting the sequence of Chow statistic as a function of candidate breakdates and by looking for systematic variations across candidate breakdates. The Quandt statistic is significant if the Chow test sequence breaks above the Andrews asymptotic critical values. If this happens, the null hypothesis of no structural break is rejected. If the conclusion is to reject the null hypothesis of no structural break, we continue with the sequential method of Bai and Perron, 1998 by splitting the sample further in two and reapplying the test to each subsample. We continue this exercise until each subsample test fails to find evidence of a break. We split the sample at the least squares breakdate estimate. The least squares breakdate estimate is the date that minimizes the full-sample sum of squared errors, equivalently, the date that minimizes the residual variance. We make this assessment visually by plotting the residual variance as a function of breakdates.

If the Chow test model assumptions are violated, i.e., we reject the null hypothesis of no ARCH effects (Engle's ARCH test) and we reject the null hypothesis that the data come from a standard normal distribution (KS test), we first try to correct for the presence of autocorrelation and/or conditional heteroskedasticity in the residuals. If we fail to do so, we employ the battery of tests proposed by Perron and Zhou, 2008 to test jointly for structural change in the regression coefficients and error variance, allowing the errors to be non-normal, and/or serially correlated, and/or conditionally heteroskedastic. Within this context of relaxed assumptions about the errors's distribution, we follow the practical recommendation of Bai and Perron, 2003 and Perron and Zhou, 2008 of using a general to specific type of procedure to determine the appropriate number and type of breaks. Therefore, we start with a "double maximum" test, *UDmax*, to test the null hypothesis of no structural breaks against an unknown number of breaks, given some upper bound for each of the regression coefficients and variance. If the *UDmax* test indicates the presence of at least one break, we decide the number of breaks based on the sequential statistic, $\sup Seq_T(l+1|l)$.

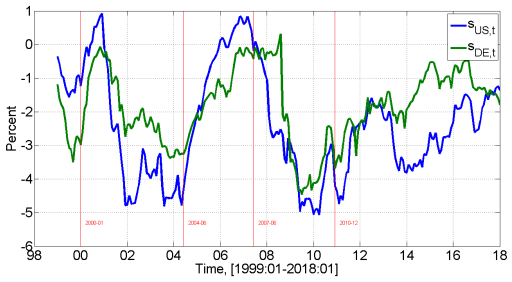
Figure 5.1: Univariate Analysis: Workflow



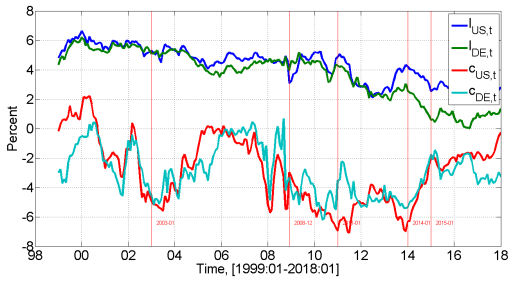
5.4.2 First Suggestions of Structural Breaks

Figures 5.2a and 5.2b show the univariate evolution over time of IYCDs. In these figures, we can spot episodes of different mean, persistence, and variability[17] for all yield curve factors. The slopes appear quite volatile, with changes in mean and persistence around 2000, mid-2004, mid-2007, and end of 2010. The levels exhibit a decreasing trend over time and slight persistence till the end of 2008. From 2003 onwards, we can observe an increased variability and persistence. Similarly to the slopes, the curvatures are also quite volatile throughout the sample period. Episodes of particular persistence seem to happen from 2003 till 2011 and from 2014 till 2015. From 2015 onwards, the curvatures appear weakly persistent.

Figure 5.2: Univariate evolution of IYCDs.



(a) $s_{US,t}/s_{DE,t}$, [1999:01-2018:01]



(b) $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$, [1999:01-2018:01]

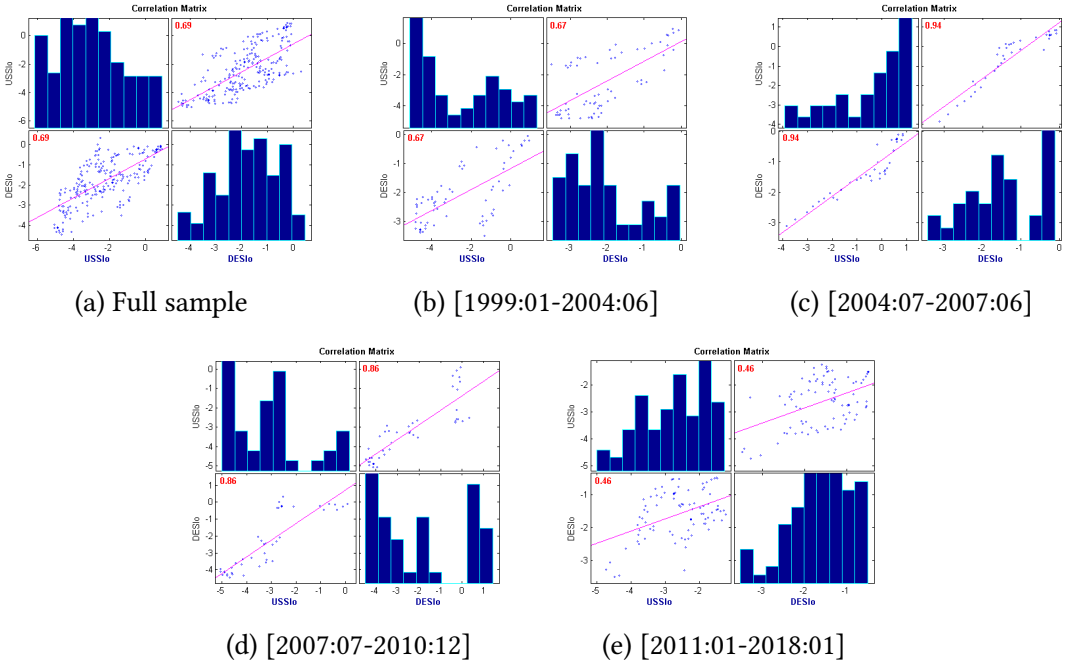
The split-sample correlation analysis for the $s_{US,t}/s_{DE,t}$ system, showed in Figure 5.3, and for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, showed in Figure B.1, provides further evidence about the existence of structural breaks and regime switches in the IYCDs.

Figure 5.3 displays the cross-correlation between the US slope and German slope. A slightly heteroskedastic behavior can be noticed for the full sample, [1999:01-2018:01]. The dots on the plot seem to form an increasing trend and flare out. For the sub-sample [1999:01-2004:06], two blotches of dots can be noticed. For the sub-sample [2004:07-2007:06], the two slopes are almost perfectly correlated. The correlation, however, starts to decrease quite significantly for the sub-sample [2007:07-2010:12], leading again to a heteroskedastic pattern for the sub-sample [2011:01-2018:01].

Figure B.1 displays the cross-correlation between the US level, German level, US curvature and German curvature. In the full sample, we can observe a heteroskedastic behavior between the levels and curvatures, for both intra- and inter-country pairs. Indeed, blotches of dots can be clearly spotted in the correlation plot of US level and German level, US level and US curvature, German level and German curvature, and German level and US curvature. Compared to the full sample, the correlations are higher in the sub-sample [1999:01-2002:12] and lower

[17] Episodes of changing intercept suggest changes in the mean of the series, equivalently, changes in the trend. Episodes of different persistence suggest changes in the autoregressive parameters, thus, they reflect changes in the serial correlation of the series. Finally, episodes of different variability in different periods suggest changes in the volatility of the series.

Figure 5.3: Split-sample correlation analysis: $s_{US,t}/s_{DE,t}$ system.



(a) Full sample (b) [1999:01-2004:06] (c) [2004:07-2007:06]

(d) [2007:07-2010:12] (e) [2011:01-2018:01]

in the sub-sample [2003:01-2011:12]. In the sub-sample [2004:01-2015:12], the levels are highly positively correlated and the curvatures are highly negatively correlated. A similar, though not so strong, correlation structure characterizes the sub-sample [2016:01-2018:01].

The observations derived from the time series plot analysis and split-sample correlation analysis suggest that when performing tests for structural breaks, we might want to allow for the possibility of changes in the mean, changes in the autoregressive parameters, and changes in the structure of the correlation in the errors.

5.4.3 Model Assumptions for Chow Test

The model assumptions for Chow test are satisfied for both the US and German slopes[18], in the $s_{US,t}/s_{DE,t}$ system, and for the US and German levels[19], in the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. For these variables we test for multiple structural breaks of unknown timing following the method of Bai and Perron, 1998. The model assumption of homoskedastic residuals does not hold for the regressions of US and German curvatures, for which we therefore employ the battery of tests presented in Perron and Zhou, 2008 to test jointly for structural change in the error variance and regression coefficients in the presence of heteroskedastic errors.

[18]See Figure B.2 and Table B.1.

[19]See Figure B.3 and Table B.1. To test for structural change in the VEC model for the US and German levels and curvatures, we employ the equivalent VAR(p=q+1) representation. Hence, for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$, for which we found a VEC(3) model, we test for structural change in the equivalent VAR(4) model, and we test equation-wise.

5.4.4 Bai and Perron, 1998: Testing for Multiple Structural Breaks of Unknown Timing

The test results for multiple structural change of unknown timing in the US and German levels and slopes are reported in Table 5.1 and in Figures B.4, B.5, B.6, B.7, B.8, B.9, B.10, and B.11.

Starting with the US slope in the $s_{US,t}/s_{DE,t}$ system, the test results provide evidence about the existence of 4 structural breaks in the US slope. The earliest break is in 2007:05, followed by breaks in 2008:09, 2011:08, and 2013:08. For the break in 2007:05, the Bai 90% confidence interval is [2006:05,2008:05]. The 90% confidence intervals for the other breaks are: [2008:08,2008:11], [2011:04,2012:05], and [2013:06,2013:10], respectively. For the German slope, there is evidence about at least 1 structural break in 2011:01, with a Bai 90% confidence interval at [2010:11-2011:03][20].

In the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, for the US and German levels, the test results provide evidence about one break. In the US level, the break is in 2001:10 and its Bai 90% confidence interval is [2001:08-2001:11]. In the German level, the break is in 2004:07 and its Bai 90% confidence interval is [2004:05-2004:10].

[20]Even though the Quandt statistic failed to find evidence of structural breaks in the samples [1999:01-2009:05] and [1999:01-2011:01], for these samples the residual variance as a function of breakdates registered well-defined V shapes.

Table 5.1: Results of tests for structural change of unknown timing. Andrews, 1993 asymptotic critical values.

Nr. of Param	Nr. of Split	Sample	SupLM Stat	5% asymp. cValue	5% χ^2 cValue	Estimated Breakdate	Confidence Interval
System: $s_{US,t}/s_{DE,t}$; Equation: $s_{US,t}$							
11	0	[1999:01 - 2018:01]	13.7475	9.84	3.84	2007-05	[2006:05 - 2008:05]
11	1a	[1999:01-2007:05]	9.3139	9.84	3.84	2001-06	[2001:04 - 2001:10]
11	1b	[2007:06-2018:01]	18.1777	9.84	3.84	2008-09	[2008:08 - 2008:11]
11	2a	[1999:01- 2008-09]	8.1217	9.84	3.84	2001-06	[2001:04 - 2001:10]
11	2b	[2008:10-2018:01]	17.4168	9.84	3.84	2011-08	[2011:04 - 2012:05]
11	3	[2011:09-2018:01]	24.8018	9.84	3.84	2013-08	[2013:06 - 2013:10]
11	4a	[1999:01-2013:08]	6.3972	9.84	3.84	2007-07	[2006:08 - 2008:05]
System: $s_{US,t}/s_{DE,t}$; Equation: $s_{DE,t}$							
11	0	[1999:06 - 2018:01]	6.6354	9.84	3.84	2009-05	[2008:04 - 2010:07]
11	1a	[1999:01-2009:05]	3.1935	9.84	3.84	2008-06	[2008:05 - 2008:08]
11	1b	[2009:06-2018:01]	10.3317	9.84	3.84	2011-01	[2010:11 - 2011:03]
11	2a	[1999:01-2011:01]	3.6291	9.84	3.84	2008-08	[2008:07 - 2008:10]
11	2b	[2011:02-2018:01]	7.0717	9.84	3.84	2012-06	[2012:04 - 2012:08]
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: $l_{US,t}$							
17	0	[1999:01-2018:01]	9.3163	9.84	3.84	2003-07	[2003:05 - 2003:12]
17	1a	[1999:01-2003:07]	19.278	9.84	3.84	2001-10	[2001:08 - 2001:11]
17	1b	[2003:08-2018:01]	7.2157	9.84	3.84	2009-01	[2008:11 - 2009:05]
17	2	[2001:10-2018:01]	8.1391	9.84	3.84	2010-04	[2010:07 - 2010:06]
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: $l_{DE,t}$							
17	0	[1999:06 - 2018:01]	6.6575	9.84	3.84	2009-06	[2008:04 - 2009:11]
17	1a	[1999:01-2009:06]	27.2673	9.84	3.84	2004-07	[2004:05 - 2004:10]
17	1b	[2009:07-2018:01]	8.2866	9.84	3.84	2011-11	[2011:09 - 2012:01]
17	2a	[2004:07-2018:01]	4.9077	9.84	3.84	2011-06	[2011:04 - 2011:08]

5.4.5 Perron and Zhou, 2008: Testing Jointly for Structural Changes in the Error Variance and Coefficients in the Presence of Heteroskedastic Errors

We recall that the Chow test model assumption of homoskedastic residuals does not hold for the linear regression models of US and German curvatures. The general to specific procedure of Perron and Zhou, 2008 applied to these regressions produces the results reported in Tables 5.2 and B.2.

For both curvatures, we start with an $UDmax$ test[21] with an upper bound of M=N=2 for each of the regression coefficients and variance[22].

For the US curvature, the $UDmax$ test is highly significant (value of 178.6823). The sequential sup$Seq_T(2,1|1,1)$ test is also significant at all levels (value of 17.4294), indicating that, given a model with m=n=1 breaks, including a second break in the coefficients is warranted. The equivalent test for the inclusion of a second break in the variance is significant at the 10% level (value of 8.3184), thus, given a model with m=n=1 breaks, including a second break in variance is warranted. The sup$LR_{3,T}(1,1|0,1)$ test assesses whether there are zero or 1 break in the regression coefficients. The test is insignificant at all levels (value of 7.0606). The equivalent test for the variance, sup$LR^*_{2,T}(1,1|1,0)$, assesses whether there are zero or 1 break in variance. The test is significant at 10% level (value of 7.9536). The estimated break dates in coefficients are 2007:10 and 2014:05 and the estimated break dates in variance are 2007:11 and 2011:10[23].

Slightly different results are obtained for the German curvature. The $UDmax$ test with M=N=2 is highly significant (value of 114.4483). However, the sequential tests performed afterwards provide evidence on the existence of only one break both in the regression coefficients and variance. More specifically, the sequential sup$Seq_T(2,1|1,1)$ test is insignificant at the 1% level (value of 13.122), indicating that, given a model with m=n=1 breaks, including a second break in the coefficients is unwarranted. The sequential sup$Seq_T(1,2|1,1)$ test is also insignificant at the 5% level (value of 7.8397), indicating that a second break in variance is also unwarranted. Lastly, the sup$LR_{3,T}(1,1|0,1)$ test is significant at the 5% level (value of 14.0779) and the sup$LR^*_{2,T}(1,1|1,0)$ is significant at all levels (value of 25.3325), indicating that one break exists in each tested coefficients and variance. The estimated break date in coefficients is 2014:02 and in variance is 2009:11.

[21] The $UDmax$ test checks the null hypothesis of no structural breaks against an unknown number of breaks, given an upper bound for each of the regression coefficients and variance.

[22] For the $UDmax$ test with the upper bound M=N=2, asymptotic critical values are available in Perron and Zhou, 2008 for a number of regressors up to q=5. For this reason, we test for partial structural change in the regressions of US and German curvatures, in the sense that, we allow only for breaks in the intercept and in the lags of the dependent variable (for a total of q=5 regressors).

[23] A joint conclusion on the results of the tests for the US curvatures would be that there are either no breaks or two breaks both in the regression coefficients and variance. Given that the $UDmax$ test is thought to be the most useful test for trying to determine if structural breaks are present and our $UDmax$ test with an upper bound of M=N=2 breaks resulted highly significant, we could conclude that two breaks are present, both in the regression coefficients and variance.

Table 5.2: Results of testing jointly for structural change in the regression coefficients and error variance.

TP	H0	Ha	test	Stat	cValue at 10%	cValue at 5%	cValue at 1%	d
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: c_{US}								
TP-8	$m = n = 0$	$1 \leq m \leq 2, 1 \leq n \leq 2$	UDmax$LR^{*}_{4,T}$	178.6823	37.12	39.82	44.88	1
TP-9	$m = 1, n = 1$	$m = 1 + 1, n = 1$	sup$Seq(2, 1\|1, 1)$	17.4294	8.51	10.13	13.89	1
TP-10	$m = 1, n = 1$	$m = 1, n = 1 + 1$	sup$Seq(1, 2\|1, 1)$	8.3184	7.04	8.58	12.29	0
TP-3	$m = 0, n = 1$	$m = 1, n = 1$	sup$LR_{3,T}(1, 1\|0, 1)$	7.0606	12.359	14.21	16.94	0
TP-2	$m = 1, n = 0$	$m = 1, n = 1$	sup$LR^{*}_{2,T}(1, 1\|1, 0)$	7.9536	7.04	8.58	12.29	1
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: c_{DE}								
TP-8	$m = n = 0$	$1 \leq m \leq 2, 1 \leq n \leq 2$	UDmax$LR^{*}_{4,T}$	114.4483	37.12	39.82	44.88	1
TP-9	$m = 1, n = 1$	$m = 1 + 1, n = 1$	sup$Seq(2, 1\|1, 1)$	13.122	8.51	10.13	13.89	0
TP-10	$m = 1, n = 1$	$m = 1, n = 1 + 1$	sup$Seq(1, 2\|1, 1)$	7.8397	7.04	8.58	12.29	0
TP-3	$m = 0, n = 1$	$m = 1, n = 1$	sup$LR_{3,T}(1, 1\|0, 1)$	14.0779	12.359	14.21	16.94	1
TP-2	$m = 1, n = 0$	$m = 1, n = 1$	sup$LR^{*}_{2,T}(1, 1\|1, 0)$	25.3325	7.04	8.58	12.29	1

5.5 Conclusion

In the present Chapter, we performed a univariate analysis to verify whether structural breaks are present in the data generation processes of US and German yield curve drivers. We employed tools and procedures that are most suitable given the dynamic properties of our data and we found evidence supporting the existence of multiple structural breaks in all yield curve drivers.

In the 2D-VAR(5) model for the US and German slopes, the Chow test model assumptions of normal, serially uncorrelated, and homoskedastic residuals hold. In the 4D-VEC(3) model for the US and German levels and curvatures, these assumptions hold only for the US and German levels. For the slopes and levels, we employed the Bai-Perron (1998) sequential method to test for multiple structural breaks of unknown timing. For the US slope, we found evidence about 4 breaks, with the least squares breakpoint estimates in 2007:05, 2008:09, 2011:08, and 2013:08. For the German slope and US and German levels, we found evidence about at least 1 break. The breakdate estimates for the US slope, US and German levels are in 2011:01, 2001:10, and 2004:07, respectively.

For the US and German curvatures, for which the Chow test model assumption of homoskedastic residuals does not hold, we employed the general to specific procedure of Perron-Zhou (2008) to test jointly for structural changes in the error variance and regression coefficients in the presence of heteroskedastic errors. For the US curvature we estimated two breaks in the regression coefficients, in 2007:10 and 2014:05, and 2 breaks in the error variance, in 2007:11 and 2011:10. For the German curvature we estimated one break in the regression coefficients, in 2014:02, and one break in the error variance, in 2009:11.

Given the results of this univariate analysis, two research questions arise. What is the nature of the structural breaks in the US and German yield curve drivers and how to account for their presence in our models. We provide answers to these questions in Chapter 6 and Chapter 7.

Table 5.3: Univariate Analysis of Structural Breaks: Summary of Estimated Breakdates

System	Eq.	Chow Test Asm.	Method	Coefficient Breaks	Variance Breaks
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	satisfied	Bai-Perron (1998)	2007:05	
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	satisfied	Bai-Perron (1998)	2008:09	
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	satisfied	Bai-Perron (1998)	2011:08	
$s_{US,t}/s_{DE,t}$	$s_{US,t}$	satisfied	Bai-Perron (1998)	2013:08	
$s_{US,t}/s_{DE,t}$	$s_{DE,t}$	satisfied	Bai-Perron (1998)	2011:01	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{US,t}$	satisfied	Bai-Perron (1998)	2001:10	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$l_{DE,t}$	satisfied	Bai-Perron (1998)	2004:07	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	not satisfied	Perron-Zhou (2008)	2007:10	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	not satisfied	Perron-Zhou (2008)	2014:05	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	not satisfied	Perron-Zhou (2008)		2007:11
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{US,t}$	not satisfied	Perron-Zhou (2008)		2011:10
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{DE,t}$	not satisfied	Perron-Zhou (2008)	2014:02	
$l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$	$c_{DE,t}$	not satisfied	Perron-Zhou (2008)		2009:11

Chapter 6

Monetary Policy, Interest Rates, and Structural Breaks

6.1 Introduction

The empirical results of Chapter 5 suggest that structural breaks are present in the US and German yield curve drivers in the sample period [1999:01-2018:01].

In this chapter, our objective is to understand the root causes or nature of the structural breaks. Past academic research provides strong evidence that changes in business cycle conditions and monetary policy impact interest rates, causing them to behave differently in different time periods (Ang and Bekaert, 2002). Our hypothesis is that the breaks in the time series of IYCDs were caused by significant changes in the monetary policy of the European Central Bank and the US Federal Reserve System, changes that were not anticipated by market participants well in advance. Our hypothesis is founded on the numerous market events that are covered by our sample period and the way the ECB and the Fed responded in terms of monetary policy. In the US, throughout the chairmanships of Alan Greenspan, Ben Bernanke, and Janet Yellen, the 2008 Financial Crisis induced a transition from traditional monetary policy to more accommodative and nontraditional monetary policy. Such a transition provides good reasons to believe that the 2008 Financial Crisis signed a change in monetary policy regimes and an increased predictability of Fed's actions by market participants, mainly due to the forward guidance tool, which was meant to guide market participants in understanding the Federal Open Market Committee's thinking and the future course of monetary policy. In the euro area, the ECB switched from the standard measure of setting its monetary policy via increasing/decreasing the rate on the main refinancing operations to the non-standard measures of Enhanced Credit Support, Securities Markets Programme, Covered Bond Purchase Programmes, Expanded Asset Purchase Programme, and Forward Guidance.

If there are good reasons to believe that a change in monetary policy regime did occur in our sample period and that the regime change affected the ability of market participants in predicting monetary policy decisions, we can formulate the hypothesis that the structural breaks in the US and German yield curve drivers stem from such a regime change. Periods

in which the monetary policy is more predictable than in others are characterized by reduced market volatility to monetary policy announcements. This happens because market participants correctly foresee policy decisions and start to price-in monetary policy changes before they are publicly announced. Periods in which the monetary policy is less predictable are characterized by increased market volatility because market participants perceive monetary policy decisions as "surprises".

In this chapter, our focus is on measuring the predictability of monetary policy decisions in an attempt to explain the nature of structural breaks in the time series of US and German yield curve drivers.

Our study is structured as follows. In Section 6.2, we start with a discussion of how monetary policy affects interest rates and may cause structural breaks. We review the transmission mechanism of monetary policy of the Fed and the ECB and, more specifically, the monetary policy objectives, principles, and practice, the policy implementation process, and the instruments and policy tools employed by the Fed and the ECB. We continue our discussion with a brief description of the Fed chairmanships and the ECB periods of monetary policy covered by our sample period. The aim is to analyze the main market events, how the Fed and the ECB responded and whether the way they responded caused changes in monetary policy regimes. In Section 6.3, we discuss different methods available in the literature for measuring the predictability of Central Banks. These methods include an analysis of the volatility in the money market rates on days of policy meetings, a regression of (absolute) changes in the money market rates on a time dummy accounting for monetary policy meetings, calculation and analysis of the "Hit-Rate", measurement of money market adjustment to monetary policy moves, and measurement of market anticipation and pass-through of monetary policy. Using money market rates and the Fed and the ECB policy rates, in Section 6.4, we measure the US Fed and the ECB predictability and list our empirical findings. Section 6.5 concludes.

The novelty of the study is to assess monetary policy predictability in the context of the term structure of interest rates, in order to investigate and understand the root causes of structural breaks in the US and German yield curve drivers. Moreover, we assess monetary policy predictability on a significantly larger sample period compared to previous literature. From the beginning of the European Monetary Union (EMU), 1999:01, and up to recent days, 2018:01, we analyze 18 years of daily data.

6.2 Monetary Policy and Interest Rates: US Fed vs ECB

In this Section we address the question of how monetary policy affects interest rates and may cause structural breaks.

6.2.1 US Federal Reserve System

The US Federal Open Market Committee (FOMC[1]) sets the US monetary policy to achieve three goals: maximum employment, stable prices, and moderate long-term interest rates. These goals are achieved with the *target for the federal funds rate*[2] (the so-called "Fed funds rate") by managing the level of short-term interest rates and influencing the availability and cost of credit in the economy. As economic conditions change, the FOMC adjusts monetary policy accordingly with an "easing" (decrease in the target for the Fed funds rate) or "tightening" (increase in the target for the Fed funds rate) of monetary policy. The monetary policy has a direct impact on interest rates and an indirect impact on stock prices, wealth, and currency exchange rates. Through these channels, monetary policy influences the real economy, i.e., spending, investment, production, employment, and inflation in the United States[3].

In practice, monetary policy is made at the FOMC meetings[4]. At these meetings, three key questions are addressed: "*How is the US economy likely to evolve in the near and medium term, what is the appropriate monetary policy setting to help move the economy over the medium term to the FOMC's goals of 2% inflation and maximum employment, and how can the FOMC effectively communicate its expectations for the economy and its policy decisions to the public? (Fed, 2018)*". Once the monetary policy is determined, it is implemented in practice via the *monetary policy tools*, i.e., *reserve requirements, open market operations,* and *discount window lending*. These tools characterize the so-called "traditional monetary policy"[5] of the Fed and are employed to achieve the targeted federal funds rate.

Figure 6.1 plots the Federal funds rate for the sample period [1999:01-2018:01] and the main market events, marked with horizontal lines. In this Figure, we can observe that in the first part of the sample period, the Fed funds rate exhibited increasing and decreasing trends, whereas in the second part of the sample period, the rate was almost flat and close to zero, characterizing the so-called "low interest rate environment". In fact, over the years and until the 2007-2009 Financial Crisis, the Fed has relied upon traditional monetary policy tools that involve the setting of the Fed funds rate.

The increasing trend in the Fed funds rate that can be observed in the 2000s was supported by Alan Greenspan's[6] tightening monetary policy. In that period, the Fed raised interest rates

[1]The FOMC is the monetary policymaking arm of the US Federal Reserve System (the Fed). The Fed was established by the US Congress on Dec. 23, 1913, with the Federal Reserve Act, "*as the central bank for the United States to provide the nation with a safer, more flexible, and more stable monetary and financial system*" (Congress, 1913)". The purposes and functions of the Fed are detailed in the 10th edition of "*The Federal Reserve System Purposes & Functions*" (Fed, 2018).

[2]The *effective federal funds rate* is the interest rate for overnight borrowing between banks. More specifically, it is the "*interest rate at which depository institutions – banks, savings institutions (thrifts), and credit unions – and government-sponsored enterprises borrow from and lend to each other overnight to meet short-term business needs*" (Fed, 2018).

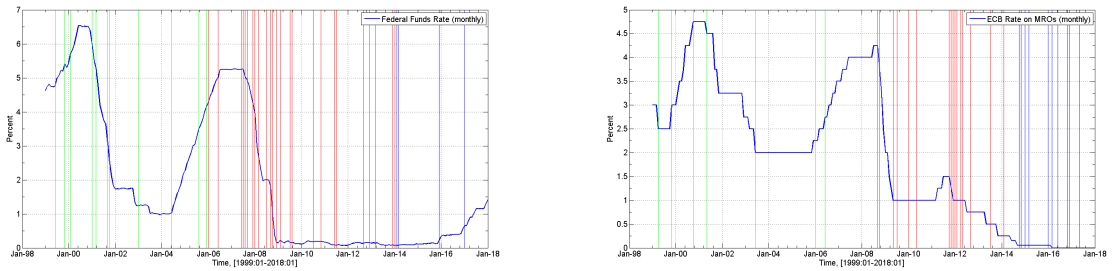
[3]See Section B.3 for a detailed explanation of the transmission mechanism of Fed monetary policy.

[4]Each year there are 8 regularly scheduled meetings of the FOMC. At these meetings, the members of the Board of Governors and the presidents of the 12 Federal Reserve Banks gather at the Board's Office in Washington, D.C. to discuss economic and financial conditions and deliberate on monetary policy.

[5]See Section B.3 for a detailed explanation of the US Fed tools for traditional monetary policy.

[6]Alan Greenspan, Chairmanship Aug. 1987 – Jan. 2006.

Figure 6.1: US Federal funds rate, the ECB rate on the MRO, and the main market events in the sample period [1999:01-2018:01]



(a) Federal funds rate

(b) ECB Rate on MRO

Note: The vertical lines are set at the main market events which occurred in the sample period [1999:01-2018:01]. For the US, the three colors denote the three Fed chairmanships: green – Greenspan, red – Bernanke, and blue – Yellen. For the EU, the three colors denote the three periods of ECB monetary policy: green – first period, red – second period, and blue – third period.

several times. In February 2000, the Fed raised rates despite the stock market decline in March. In January 2001, George Bush took office and, in March 2001, the recession began. The Fed lowered rates to fight the recession. The September 11 attacks and the various corporate scandals of 2001 undermined the US economy. The Fed responded with a series of interest rate cuts that put downward pressure on the federal funds rate.

In December 2005, right after IMF Economist, Dr. Raghuram Rajan warned the World's Central Banks that "*The inter-bank market could freeze up and one could well have a full-blown financial crisis*", similar to the Long-Term Capital Management (LTCM) crisis, the US yield curve started to invert, as investors started to buy more long-term Treasuries (3- to 20-year) than short-term bills (1-month to 2-year). Consequently, the yield on long-term Treasury notes was falling faster than the yield on short-term bills. By December 22, 2005, the yield curve for the US Treasuries inverted. In 2006, the falling housing prices and the difficult conditions in financial markets gave the first signs of the Great Recession of 2008. On February 1, 2006, Ben Bernanke[7] took office as chairman of the Fed. With unemployment at 6% and inflation at 2½%, from January to June 2006, the Bernanke-led Fed raised the federal funds rate to cool the housing market bubble. From mid-2007 till the end of the year, the banking sector was stung by the mortgage securities. One month later, Standard&Poor's placed 612 securities backed by subprime residential mortgages on a credit watch with negative implications. In August 2007, American Home Mortgage Investment Corp. filed for bankruptcy and Fitch Ratings downgraded Countrywide Financial Corp.. In September 2007, as home sales were continuing to fall, the Fed began lowering the federal funds rate to help the economy. By the end of the year, the federal funds rate was 4¼%.

As 2008 began, economic indicators were pointing towards an increased risk of recession. Bernanke testified before Congress, prompting for quick action to stimulate the economy through

[7]Ben Bernanke, Chairmanship Feb. 2006 – Jan 2014.

targeted government spending and tax incentives. In February 2008, the Bush administration received congressional approval for its $168 billion economic stimulus package and the Economic Stimulus Act of 2008 was enacted by Congress.

In March 2008, the Fed bailed out Bear Stearns. In July 2008, Bank of America (BofA) bought Countryside Financial. In September 2008, Lehman Brothers filed for bankruptcy. The stock market crashed[8]. AIG was bailout by the US Treasury. BofA bought Merrill Lynch and regulators closed Washington Mutual.

In October 2008, President Bush signed into law the Troubled Asset Relief Program (TARP), to allow the US Department of the Treasury to infuse cash into the nation's banks to keep them operating. In December 2008, the federal funds rate became effectively zero – the lowest federal funds rate possible. With the federal funds rate at zero, the FOMC could no longer rely on reducing that rate to provide support to economy. To support the US economy during and after the financial crisis, the FOMC turned to two less conventional policy measures – *large-scale asset purchases* (also know as *Quantitative Easing*) and forward guidance[9].

In December 2008, the Fed began its first round of QE by purchasing $600 million in MBS. Between December 2008 and August 2010, the Fed purchased $175 billion in direct obligations of the government-sponsored entities Fannie Mae, Freddie Mac, and the Federal Home Loan Banks as well as $1.25 trillion in MBS guaranteed by Fannie Mae, Freddie Mac, and Gennie Mae. In addition, between March 2009 and October 2009, the Fed purchased $300 billion of longer-term Treasury securities.

In February 2009, the DJIA hit a 12-year low. A few months later, in July 2009, the Fed introduced bank stress tests – the first versions of the Dodd-Frank Wall Street Reform and Consumer Protection Act (Dodd-Frank). The same month, 140 banks failed. In July 2010, the Dodd-Frank became law.

In face of a sluggish economic recovery, in November 2010, the Fed announced the second large-scale asset purchase program, QE2. Between November 2010 and June 2011, the Fed expanded its asset holdings by buying $600 billion of longer-term Treasury securities.

Between September 2011 and December 2012, the Fed introduced the *Maturity Extension Program* (MEP), under which the Fed bought $667 billion of Treasury securities with remaining maturities of 6 to 30 years and sold an equivalent value of Treasury securities with remaining maturities of 3 years or less. The effect of MEP was to add to the downward pressure on longer-term interest rates without affecting the size of the Fed's balance sheet. In July 2011, the Consumer Financial Protection Bureau (CFPB) opened.

With considerable slack remaining in the economy, in September 2012, the Fed announced the third round of asset purchases, QE3, and began purchasing MBS at a pace of $40 billion per month. In January 2013, these MBS purchases were supplemented by $45 billion per month in purchases of longer-term Treasury securities. Unlike QE1, QE2, and MEP, in which the total size of the program was announced at the time of their undertaking, QE3 was announced as an *open-ended* asset purchase program, which would continue until an improved outlook from

[8]The Dow Jones Industrial Average (DJIA) fell 777.68 points in intra-day trading, causing approximately $1.2 trillion loss in market value after the House rejected the $700 billion bailout plan. Source: CNN Money.

[9]See Section B.3 for a detailed description of the US Fed tools for non-traditional monetary policy.

the labor market, stable inflation and expected inflation, and until the benefits of the purchases continued to outweigh their costs and risks.

In December 2013, the FOMC began to taper the economic stimulus by slowing the pace of its asset purchase, from $85 billion per month to $75 billion, but kept the federal funds rate at 0% to 0.25%. In February 2014, Janet Yellen[10] became Fed chair. In March, Yellen held her first FOMC meeting and announced to continue Bernanke's tapering of asset purchases. The FOMC continued to slow the pace of purchases at its subsequent meetings. QE3 was concluded in October 2014.

In late 2015, the US economic growth stabilized. With the unemployment rate nearly consistent with maximum employment, the Fed began the so-called *normalization*[11] process of monetary policy. The Policy Normalization Principles and Plans were announced on September 17, 2014. The monetary policy normalization process began in December 2015. Based on considerable improvement in labor market conditions during 2015 and reasonable confidence that inflation would rise to 2% over the medium term, the FOMC decided to raise the federal funds rate by ¼ percentage points, bringing the target range to 25 to 50 basis points. This was the first change since December 2008.

To keep the federal funds rate in its target range, the Fed indicated that it intended to use two administered rates: the interest rate the Fed pays on excess reserve balance (the IOER rate) and the interest rate it pays on overnight reverse repurchase agreements (the ON RRP rate).

When supporting the economy during the financial crisis, the FOMC employed, in addition to the QE tool, the nontraditional policy tool of forward guidance. The *forward guidance* tool is the communication to the public of how the FOMC intends to adjust policy in the future, in order to help the public understand the Committee's thinking about the future course of policy. In December 2008, when the federal funds rate was effectively nearly close to its lower bound, the FOMC indicated in its postmeeting statement that it expected that "*weak economic conditions are likely to warrant exceptionally low levels of the federal funds rate for some time*"[12]. In March 2009, as the economic downturn worsened, the FOMC amended the forward guidance announcing to the public that the federal funds rate could remain at exceptionally low levels "for an extended period". In August 2011, the FOMC introduced a *date-based forward guidance* by indicating the period of time over which it expected economic conditions to warrant maintaining the federal funds rate near zero. In December 2012, the FOMC replaced the date-based forward guidance with an *economic conditionality*, i.e., with language indicating the economic conditions that the Committee expected to see before it would begin to consider raising its target for the federal funds rate. In December 2015, with the normalization of the monetary policy, the FOMC indicated that "*monetary policy is not on a predetermined path*" and that its policy decisions will "*depend on what incoming information tells policymakers about whether a change in policy is necessary to move the economy toward, or keep it at, maximum employment and 2% inflation*", (Fed, 2018).

[10] Janet Yellen, Chairmanship Feb. 2014 – Jan. 2018.

[11] The term *normalization* refers to the steps the FOMC began taking to return short-term interest rates to more-normal levels and reduce the size of the Fed's balance sheet.

[12] FOMC statement (December 16, 2008).

6.2.2 European Central Bank

The primary objective of the ECB's monetary policy consists in maintaining price stability[13]. In pursuing this objective, the ECB has adopted a specific strategy based on a quantitative definition of price stability and a comprehensive assessment of the risks to price stability. The quantitative definition of price stability is a "*[. . .] a year-on-year increase in the Harmonised Index of Consumer Prices (HICP) for the euro area of below 2% [. . .]*" (ECB, 2011c). Price stability must be maintained over the medium term, over which the ECB aims to maintain inflation rates below but close to 2%.

The ECB channels its monetary policy to the real economy via the transmission mechanism illustrated in Figure B.12[14]. This is "*[. . .] the process through which monetary policy decisions affect the economy in general, and the price level in particular [. . .]*". The two main channels of monetary policy transmission are the *money-market interest rates channel* and the *expectations channel*. The transmission mechanism starts with the ECB changing its official interest rates, the so-called "key ECB interest rates"[15], charged on the funds provided to the banking system. By virtue of its monopoly of supplying the monetary base, the ECB can fully determine the key interest rates. In doing so, the central bank manages the liquidity situation in the money market, influences money market interest rates (short-term), and the expectations of future rate changes (in the medium to long term), Guidolin and Thornton, 2008. Other two channels operating in the transmission mechanism of monetary policy are the *exchange rates channel* and the *asset price channel*. Expectations of future official interest rates changes affect medium and long-term interest rates. Via the so-called *expectations channel*, the central bank can influence directly the price developments by guiding economic agents'expectations of future inflation.

To steer interest rates and signal the monetary policy stance, the Eurosystem[16] employs *open market operations, standing facilities*, and (minimum) *reserve requirements*[17]. Figure 6.1b plots the ECB rate on the main refinancing operations (MRO), together with the most important market events happening in the sample period [1999:01 - 2018:01].

Over the sample period under consideration, [1999:01-2018:01], the conduct of ECB monetary policy can be classified in three different periods: a *first period*, starting with the creation of the European Monetary Union (EMU) until the 2007 financial crisis, a *second period*, covering the 2007 financial crisis and the introduction of non-standard measures, and a *third period*, covering the years post financial crisis until present days. In the sequel, we review the ECB monetary policy and market events, by distinguishing the three periods. The exposition follows mainly Hanspeter, 2004, ECB, 2011a, ECB, 2011b, ECB, 2011c, Wyplosz, 2013, De La Dehesa, 2013, Rodriguez and Carrasco, 2014, Verhelst, 2014 and Delivorias, 2015.

[13] Article 127(1) of the Treaty on the Functioning of the European Union (2008/C115/01, 2010).

[14] See Section B.3 for a detailed explanation of the transmission mechanism of ECB monetary policy.

[15] The Governing Council (GC) of the ECB sets the three key interest rates for the euro area, namely, the *interest rate on the main refinancing operations (MRO)*, which provide the bulk of liquidity to the banking system; the *rate on the deposit facility (DF)*, which banks may use to make overnight deposits with the Eurosystem; the *rate on the marginal lending facility (MLF)*, which offers overnight credit to banks from the Eurosystem (Official Interes Rates, ECB.)

[16] I.e., the ECB and the national central banks of the EU Member States whose currency is the euro.

[17] See Section B.3 for a detailed explanation of the ECB tools for traditional monetary policy.

The first period is designated by the transition to EMU and until the 2007 financial crisis. During this period, the ECB pursued its primary objective of price stability via its standard tool, i.e., the rate on the main refinancing operations (MRO). In April 1999, as a response to decreasing inflation, the GC reduced the rate on MRO from 3% to 2½%. From November 1999 to October 2000, the opposite action was taken: the GC increased the MRO rate, from 3% to 4¾%, in order to contain inflationary pressures created by strong economic growth. Post dot-com crash in the US and 9/11 terrorist attacks, the GC cut the MRO rate to support the economic growth. The key ECB interest rates were left unchanged until December 2005. From January 2006 and until mid-2007, the GC raised the key interest rates from 2¼% to 4¼%, in order to counter the fast growing economy and expanding supply of money and credit in the euro area.

The second period was signed by the 2007 Financial Crisis and the introduction of non-standard measures. With the collapse of Lehman Brothers in September 2008 the US financial turmoil became a financial crisis of global scale. The ECB first measures were "*unprecedented in terms of their nature, scope, and magnitude*"[18], as they were a combination of standard measures, in terms of drastic reduction[19] of key interest rates, and non-standard measures, representing the *Enhanced Credit Support* to the banking sector. The Enhanced Credit Support was defined as the "*special and primarily bank-based measures (...) taken to enhance the flow of credit above and beyond what could be achieved through policy interest rate reductions alone*"[20]. These measures included *liquidity management measures* and *covered bond purchases*[21]. More specifically, the *maximum maturity of LTROs was extended* from 3 to 12 months, in order to keep the money-market interest rates at low levels and provide a longer liquidity planning horizon to banks. In addition, the ECB engaged in *currency swap agreements* with the Fed, in order to maintain the US-dollar funding supply.

These liquidity management measures were complemented by the First Covered Bond Purchase Programme (CBPP1), launched in July 2009 and aiming at reviving the long-term funding market for Eurosystem banks. Between June 2009 and June 2010, the Eurosystem purchased covered bonds denominated in euro and issued in the euro area for a total amount of €60 billion.

Between 2010 and 2011, markets witnessed the first stage of the European sovereign debt crisis (or Eurozone crisis), in response to which the ECB introduced, in May 2010, the Securities Markets Programme. Under the Securities Markets Programme, the ECB purchased €210 billion (mainly) sovereign bonds on the secondary markets. Even though the program led to "*stabilization in markets as well as to an immediate and substantial decline in government bond yields*" (Cour-Thimann and Winkler, 2012), between 2011-2012, the European sovereign bond crisis intensified and a new banking crisis required for additional ECB measures.

In October 2011, to bring banks in a stronger position, the GC agreed on a capital package proposed by the European Banking Authority (EBA), requiring banks to reach a ratio of 9%

[18] Keynote address by Jean-Claude Trichet, President of the ECB at the University of Munich, 13 July 2009.

[19] The rate on MROs was reduced from 3¾% to 1%.

[20] See the aforementioned keynote address by Jean-Claude Trichet.

[21] *Covered bonds* are debt securities issued by banks. This type of bonds allow banks to access funding of a longer-term maturity compared to the ECB's refinancing operations. As such, covered bonds allow banks to manage the maturity mismatch between their assets and liabilities.

Core Tier 1 (CT1) capital (EBA, 2011). The capital package was accompanied by two LTROs (one in December 2011 and one in February 2012) with a three-year maturity each and for a total amount of approximately €1 trillion, a reduction in the minimum reserve ratio requirement from 2% to 1%, an increased in collateral availability (ECB, 2012), and a Second Covered Bond Purchase Programme (CBPP2), launched in November 2011 and ended in October 2012, for a total nominal amount of €16.4 billion.

Between 2012 and 2014, the financial crisis reached a third stage. The main market events in this period were the Greek referendum on the EU financing package, government crises in Greece and Italy, the Standard&Poor's downgrades of 9 euro-area sovereigns and 16 Spanish banks (including Santander and BBVA).

With the euro zone in the throes of crisis, on 26 July 2012, at the Global Investment Conference in London[22], the ECB President Mario Draghi declared that "*within our mandate, the ECB is ready to do whatever it takes to preserve the euro*". In September 2012, the ECB announced the Outright Monetary Transactions (OMTs) Programme, under which it would intervene along with NCBs to "*undertake outright transactions in secondary, sovereign bond markets, aimed at safeguarding an appropriate monetary policy transmission and the singleness of the monetary policy*"[23].

Mario Draghi's remarks and the OMT Programme have reduced market volatility in the euro area.

In July 2013, in response to disinflation[24] and slow growth of the euro-area economy, the ECB adopted the non-standard measure of *forward guidance*. The introduction of forward guidance was made on 4 July 2013, in a press conference[25], in which ECB President Draghi declared that "[...] *Looking ahead, our monetary policy stance will remain accommodative for as long as necessary. The Governing Council expects the key ECB interest rates to remain at present or lower levels for an extended period of time*[...]". The aim of the forward guidance is to "[...] *influence private expectations about short-term rates, which in turn will influence expectations about long-term rates, in order to strengthen the transmission of monetary policy, and thus support the economy*"[26].

The third period of ECB monetary policy designates the present trends, which include the ECB as a direct supervisor of significant banks in the context of the Single Supervisory Mechanism (SSM), the Third Covered Bond Purchase Programme (CBPP3), the Asset-Backed Securities Purchase Programme (ABS PP), and the Expanded Asset Purchase Programme (EAPP). CBPP3 and ABS PP were meant to support financial conditions in the euro area, by facilitating credit provision to the real economy, and generating positive spillovers to other markets. The EAPP (commonly referred to as "quantitative easing") was meant to support the euro-area economy and counter receding inflationary pressures. Today, the EAPP includes "*all purchase*

[22]Speech by Mario Draghi, President of the European Central Bank at the Global Investment Conference in London 26 July 2012

[23]"Technical features of Outright Monetary Transactions", ECB Press Release, 6 September 2012.

[24]ECB: Measuring inflation – the Harmonised Index of Consumer Prices (HICP).

[25]Introductory statement to the press conference, Mario Draghi, President of the ECB, Frankfurt am Main, 4 July 2013

[26]Hubert, P. and Labondance, F., "The chiaroscuro of the ECB's "forward guidance"", 13 November 2013 in monetary policy, OFCE blog

programs under which private sector securities and public sector securities are purchased to address the risks of too prolonged a period of low inflation". It consists of the Corporate Sector Purchase Programme (CSPP), the Public Sector Purchase Programme (PSPP), the Asset-Backed Securities Purchase Programme (ABS PP), and the Third Covered Bond Purchase Programme (CBPP3). From March 2015 until March 2016, the monthly purchases were conducted at average pace of €60 billion, from April 2016 until March 2017, at average pace of €80 billion, and from April 2017 to December 2017, at average pace of €60 billion.

6.2.3 Root Causes of Structural Breaks in IYCDs

Both in the US and in the euro area, the analysis of monetary policy and market events over the sample period [1999:01-2018:01] shows how the central banks transitioned from a "standard" way of setting the monetary policy to a "non-standard" way. The transition being signed mainly by the 2007-2008 Financial Crisis.

In the US, the Fed led by the chairmanships of Alan Greenspan, Ben Bernanke, and Janet Yellen had to cope with different market events and economic conditions. Alan Greenspan, despite having to deal with the aftermath of the LTCM crisis and the dot-com bubble, remained fundamentally monetarist in orientation on the economy and followed the standard Taylor rule prescriptions when deciding the monetary policy. His monetary policy decisions were implemented mainly via open market operations. Ben Bernanke's chairmanship was shaped by the 2008 Financial Crisis. Under his guidance, the Fed supplemented the traditional tools of open market operations, reserve requirements, and discount window lending, with the nontraditional tools of forward guidance and large-scale asset purchases. Janet Yellen continued mainly on the steps of Bernanke.

In the euro area, the ECB switched from the standard measure of setting its monetary policy, via increasing/decreasing the rate on the MROs, to the non-standard measures of Enhanced Credit Support, Securities Markets Programme, Covered Bond Purchase Programmes, Expanded Asset Purchase Programme, and Forward Guidance.

The transition from traditional monetary policy to more accommodative and nontraditional monetary policy provides good reasons to believe that the 2008 Financial Crisis signed a change in monetary policy regimes and an increased predictability of Fed's and ECB's actions by market participants, mainly due to the forward guidance tool, which was meant to guide market participants in understanding the central banks' thinking and the future course of monetary policy.

If there are good reasons to believe that a change in monetary policy regime did occur in our sample period and that the regime change affected the ability of market participants in predicting monetary policy decisions, we can formulate the hypothesis that the structural breaks in the international yield curve drivers stem from such a regime change. In the sequel, we strengthen this hypothesis by measuring empirically the predictability of the Fed and the ECB.

6.3 Measures for Monetary Policy Predictability

Following Perez-Quiros and Sicilia, 2002; Coppel and Connolly, 2003; and Wilhelmsen and Zaghini, 2011, and the references therein, the ability of markets to anticipate monetary policy decisions can be measured empirically with an analysis of *money market* and *policy rates*. A first method consists in calculating the absolute value of the changes in money market interest rates on the days of the policy meetings:

$$\delta_t = |i_k - i_{k-1}|, \tag{6.1}$$

where i_k is the market interest rate on the day of the meeting. The rationale behind using δ_t as a measure of monetary policy predictability is that when a policy decision is correctly foreseen by market participants, also the market volatility should not be influenced by the announcement.

The predictability of Central Banks can also be measured by comparing the changes in the money market on the days of policy meetings to a benchmark. This is the so-called "Hit-Rate" and the intuition behind it is that changes in excess of the benchmark would signal a "surprise" and thus the failure of the market in anticipating the Central Bank behavior. The "Hit-Rate" is computed as the number of times (in per cent) the market was able to correctly anticipate the monetary policy announcement. As in Wilhelmsen and Zaghini, 2011, two criteria can be used to set the benchmark and identify a surprise. A first criterion compares the absolute value of the changes in the money market rates on policy meeting days to 2 times the standard deviation of all daily changes:

$$\delta_k = |i_k - i_{k-1}| > 2\sigma_\delta. \tag{6.2}$$

A second criterion compares the same changes to 12.5 basis points:

$$\delta_k = |i_k - i_{k-1}| > 0.0125, \tag{6.3}$$

where k refers to the day of the selected meeting and σ_δ is the standard deviation of the change in interest rates on all days of the sample[27].

The reaction of financial markets to monetary policy moves can be estimated via a regression of the daily changes in the 1-month money market rate, Δi_t, on a constant α and the changes in the key policy rate, Δp_t:

$$\Delta i_t = \alpha + \gamma \Delta p_t + \epsilon_t. \tag{6.4}$$

A low value of γ can be interpreted as a small market response to the policy announcement, suggesting that the market was already pricing-in and, therefore, anticipating the monetary

[27] Wilhelmsen and Zaghini, 2011 explain that the measure defined in (6.2) compares market rate changes around monetary policy meetings with the general behavior of the market. A change outside the "confidence" bands of two times the standard deviation is considered a significant deviation from the "normal" market rate volatility, thus it can be said that the market has been surprised by the Central Bank. The measure defined in (6.3) is based on the idea that a standard monetary policy action is an increase or decrease of *minimum* 25 basis points in the policy rate. Thus, a change of more than 12.5 basis points – 50% of the overall change – in the market rates on the day of monetary policy meetings suggests that market participants were surprised by the policy announcement.

policy decision. How much in advance the market is able to price-in the expected monetary policy decision can be assessed by estimating the daily differences between the 1-month market interest rate i_t and the key policy rate p_t as a function of a constant β_0, and the change in the key policy rate Δp, led by 1, 5 and 10 business days, and lagged by 5 business days:

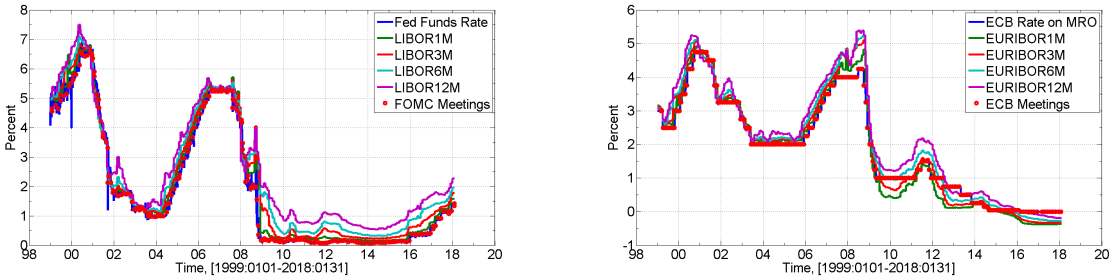
$$i_t - p_t = \beta_0 + \beta_1 \Delta p_{t+1} + \beta_2 \Delta p_{t+5} + \beta_3 \Delta p_{t+10} + \beta_4 \Delta p_{t-5}. \tag{6.5}$$

This method follows Coppel and Connolly, 2003.

6.4 Empirical Results: Predictability of US Fed vs ECB

In this Section, we employ the methods described in Section 6.3 and money market and policy rates to assess Fed and ECB monetary policy predictability in the context of the term structure of interest rates. The aim is to understand the root causes of structural breaks in the US and German yield curve drivers. More specifically, for the US market, we use 1-month, 3-month, 6-month, and 12-month LIBOR based on U.S.-dollar and the Effective Federal Funds Rate[28]. For the euro area, we use 1-month, 3-month, 6-month, and 12-month EURIBOR and the ECB Rate on Main Refinancing Operations (MRO)[29]. All data are sampled daily, for the sample period [1999:01-2018:01].

Figure 6.2: Federal Funds Rate, LIBOR and all monetary policy meetings, [1999:0101-2018:0131]. ECB Rate on Main Refinancing Operations, EURIBOR and all monetary policy meetings, [1999:0101-2018:0131]



(a) US Fed

(b) ECB

Figures 6.2a and 6.2b illustrate the US and euro money market rates, together with the Fed and ECB policy rates. In addition, the red dots designate all policy meetings that took place in the period [1999:01-2018:01]. On these days, the volatility of the LIBOR rates is larger compared to that of EURIBOR rates (Table 6.1). Table 6.1 also reports the "Hit-Rate" for both the EURIBOR and LIBOR rates, calculated according to the two times the standard deviation criterion and the 12.5 basis points criterion. The "Hit-Rate", that is, the number of times (in per cent) the

[28]Source: FRED Economic Data
[29]Source: ECB, Statistical Data Warehouse

market was able to correctly anticipate the monetary policy announcement, is higher for the euro area money market, compared to the U.S. money market. This result is in line with the study of Wilhelmsen and Zaghini, 2011, who find that the euro area has the best score, in the case of the 12.5 basis points criterion, and that the U.S. follow closely.

The *two times standard deviation criterion* for the definition of the benchmark appears to be stricter than the *12.5 basis points criterion*, as the overall hit rates for all maturities of the money market are lower compared to the hit rates obtained according to the 12.5 basis point criterion.

A hit rate of 25.3% for the 1-month LIBOR indicates that markets were able to correctly predict the outcome of the FOMC policy meetings only 25 times out of 100. For 75 times out of 100, the Fed monetary policy announcement was a surprise to the financial markets. With the same interpretation, a hit-rate of 32.13% for the 1-month EURIBOR indicates that markets were able to correctly predict the outcome of the ECB policy meeting only 32 times out of 100. 68 times out of 100, the ECB monetary policy announcement was a surprise to the financial markets.

It can also be observed that the hit-rate increases with the maturity of the money market. This observation indicates that the monetary policy surprises are less frequent at the longer maturities.

Better scores are obtained using the 12.5 basis points criterion, according to which 52 times out of 100 the 1-month LIBOR market was able to correctly predict the monetary policy decision. The 12-month LIBOR market was able to do so almost 82 times out of 100. 61 times out of 100 the 1-month EURIBOR market was able to correctly predict the monetary policy decision. The 12-month EURIBOR market was able to do so 73 times out of 100.

Table 6.1: Comparison between EURIBOR and LIBOR based on U.S.-dollar: Volatility in the money market rates and the "Hit-Rate".

Rate	σ (bps) all days	σ (bps) all meetings	Hit-Rate*	Hit-Rate**
EURIBOR1M	2.9518	1.2313	32.13%	61.04%
EURIBOR3M	2.8046	0.9185	36.95%	68.67%
EURIBOR6M	2.7296	1.7591	40.93%	72.29%
EURIBOR12M	2.8798	2.1533	46.99%	73.49%
LIBOR1M	4.8305	4.747	25.3%	52.21%
LIBOR3M	4.1472	3.3271	33.33%	62.25%
LIBOR6M	3.8907	2.9742	38.15%	73.09%
LIBOR12M	3.9406	3.3532	44.98%	81.53%

Note: (*) 2 Times Standard Deviation Criterion. (**) 12.5 basis points criterion.

Figures B.13, B.14, B.15, and B.16 plot the monetary policy surprises (red circles) for each of the four maturities of LIBOR and EURIBOR and for the two definitions of the benchmark for the calculation of the hit rate. In these Figures, we can observe that the red dots are "concentrated" more on the first part of the sample period. The accommodative, nontraditional monetary policy and the forward guidance tool, which was meant to guide market participants in understanding

the future course of monetary policy, increased the predictability of both Central Banks, thus the less frequent "surprises" in the second part of the sample period.

The estimate of γ in the following two regressions of the daily changes in the 1-month money market rate on a constant and on the changes in the key policy rate

$$\Delta i_{EURIBOR1M,t} = -0.0002 + 0.7401\Delta p_t + \epsilon_t, \tag{6.6}$$

$$\Delta i_{LIBOR1M,t} = -0.0005 + 0.0274\Delta p_t + \epsilon_t. \tag{6.7}$$

suggest that financial markets reaction to ECB monetary policy moves is quite high, indicating low predictability of the ECB, whereas in the FOMC case, the low estimate of γ indicates high predictability of the FOMC.

Concerning the ECB pass-through monetary policy, in regression 6.8 below, the 1-day, 5-day, and 10-day anticipation coefficients for the EU money market are not statistically significant and the lag indicator suggests that the ECB policy moves are generally not fully passed-through within a week. The case for the Fed is different. In regression 6.9, the 1-day, 5-day, and 10-day anticipation coefficients for the U.S. money market are statistically significant, suggesting that the U.S. money market was already pricing-in the policy change one day, one week and even two weeks ahead of the meeting. The low lag indicator suggests, however, that the FOMC policy moves are generally not fully passed-through within a week.

$$i_t - p_t = -0.0561 + 0.0454\Delta p_{t+1} + 0.0454\Delta p_{t+5} + +0.0454\Delta p_{t+10} - 0.2144\Delta p_{t-5}. \tag{6.8}$$

$$i_t - p_t = 0.1549 + 0.4905\Delta p_{t+1} + 0.1742\Delta p_{t+5} + +0.1049\Delta p_{t+10} + 0.0086\Delta p_{t-5}. \tag{6.9}$$

6.5 Conclusion

In the present Chapter, we provided a study of the Fed and the ECB monetary policy and monetary policy predictability, in an attempt to understand the root causes of structural breaks in the US and German yield curve drivers.

Using 18 years of money market and policy rates, we documented the transition of the Fed and the ECB from a traditional to a more accommodative and nontraditional monetary policy. This transition is signed by the 2008 Financial Crisis and corresponds to an increased predictability of Fed's and ECB's actions by market participants. The "Hit-Rate", calculated using LIBOR based on US-dollar and EURIBOR rates, provides evidence of more frequent monetary policy surprises in the period prior to the 2008 Financial Crisis and less frequent surprises in the period after.

We also found that, compared to the ECB, the FOMC is more predictable, with the US money market being able to price-in the policy change already one day, one week, and even two weeks ahead of the meeting.

These findings support the results of the univariate analysis of Chapter 5, namely, that structural breaks did occur in the US and German yield curve drivers in the sample period [1999:01-2018:01]. There are good reasons to believe that the root causes of structural breaks are linked to a change in monetary policy regimes and increased predictability of Central Banks.

Chapter 7

Multivariate State-Space Analysis

7.1 Introduction

In Part I, we found that the DGPs best describing the US and German yield curve drivers are a 2D-VAR(5) model for the US and German slopes, a 2D-VEC(1) model for for the US and German levels, and a 4D-VEC(3) model for the US and German levels and curvatures.

In an univariate analysis (i.e., going equation-wise in the multivariate models), in Chapter 5 we find structural breaks in all drivers. In Chapter 6 we find evidence that the root causes of these structural breaks can be linked to a regime change in the Fed and the ECB monetary policy and an increased monetary policy predictability after the 2007-2008 Financial Crisis. However, a correlation analysis[1] of the residuals from the three fitted models reveals correlation coefficients significantly different from zero. This finding might suggest that the breaks found in the univariate analysis might actually be due to missing variables with explanatory power. That is, the structural breaks in the univariate dynamics of the US and German levels might be due to the missing US and German slopes and curvatures. Similarly, the structural breaks in the univariate dynamics of US and German slopes might have occurred because the US and German levels and curvatures are missing.

In this Chapter, we test for structural breaks in a multivariate setting, in an attempt to derive more accurate results with respect to the existence of structural breaks.

Substantial payoffs of techniques for inference about breaks in multivariate systems have been already acknowledged back in 1998, by Bai, Lumsdaine, and Stock, 1998, when evidence was found against conventional associations of the slowdown in growth in the US with the oil shock. Bai et al. derive theoretical and empirical econometric results concerning multivariate change-point tests and confidence intervals for I(0) dynamic models, address the change-point problem in a cointegrating system, and present a Monte Carlo study of the tests and internal estimates. Within their theoretical framework, the first step consists in determining whether a break occurred at an unknown date. To this regard, a variety of tests can be conducted, for example, the maximum Wald statistic (Quandt, 1960; Hansen, 1992) and the logarithm of the Andrews-Ploberger exponential Wald statistic (Andrews and Ploberger, 1994).

[1] See Figure B.17.

If a break can be found, then the next step consists in constructing confidence intervals for the true break date. As Bai et al. recall, the problem of constructing confidence intervals for the true break date has been considered by various authors, following different approaches (Hinkley, 1970; Picard, 1985; Yao, 1987; Siegmund, 1988; Kim and Siegmund, 1989).

Picard, 1985, specifically, developed an asymptotic distribution for the Gaussian MLE of the breakpoint. Even though Picard's results are limited to the univariate setting, for finite order autoregressive processes, the results permitted the construction of asymptotic confidence intervals for the break point in the univariate setting. Bai et al. extend Picard's results to the multivariate setting, by allowing for the following. The covariance matrix is explicitly treated as unknown and estimated. The normality assumption is relaxed and no assumption is made with respect to the underlying density function. It is assumed, however, that the disturbances form a sequence of martingale differences with some moment conditions, and pseudo-Gaussian maximum likelihood estimation is employed. The results are limited to partial structural change models estimated with the full sample but are extended to models with I(1) and trending regressors and with serially correlated errors (i.e., broken-trend stationary variables and broken cointegrating relations).

With respect to the change-point problem in a cointegrating system, Bai et al. use the results of Gregory and Hansen, 1996 and Campos, Ericsson, and Hendry, 1996 (who develop theoretical results for testing for cointegration allowing for a possible break) to study how to estimate the break date (if a break indeed occurred) and investigate the statistical property of the estimated break point.

A Monte Carlo study of a break date statistics for an I(0) model (with a break in the intercept only) and a cointegrated model (with a break in the intercept and the cointegrating coefficient) provides good evidence in favor of the asymptotic theory developed by Bai et al., when the break is of moderate size.

Other works on structural break issues in the context of a system of multivariate equations include Bai, 2000, who obtained consistency, rate of convergence, and limiting distributions for the estimated break points in a segmented VAR model. The model is estimated by the quasi-maximum likelihood (QML) method, allowing for the break to occur in the parameters of the conditional mean, the variance of the error term, or both. Hansen, 2003 generalized the cointegrated vector autoregressive model of Johansen (Johansen, 1988) to allow for multiple structural changes. In his analysis, the break dates must be known a priori.

The work of Qu and Perron, 2007 provides a comprehensive treatment of issues related to estimation, inference, and computation with multiple structural breaks occurring at unknown dates in multivariate regression models. In the techniques proposed by Qu et al., changes can occur in the parameters of the conditional mean, the covariance matrix of the errors, or both. The distribution of the regressors is allowed to change across regimes and the assumptions about the distribution of the errors are relaxed such that the presence of conditional heteroskedasticity and autocorrelation is allowed. Using a QML method based on normal errors, Qu et al. derive the consistency, rate of convergence, and limiting distribution of all parameters. The result of Bai, Lumsdaine, and Stock, 1998 of common breaks across equations can be derived as a special case from the results of Qu et al.. In this case, the precision of the estimates would increase with the number of equations in the system. Furthermore, Qu et al. argue that standard distributions

apply for the parameters of the conditional mean and the covariance matrix of the errors because the limiting distributions are the same with either estimated break dates or known break dates.

To determine the number of breaks in the system, Qu et al. consider the use of testing procedures (as in Bai and Perron, 1998). The authors consider tests that allow for changes in the coefficients of the conditional mean or in the variance of the error term or in both. In addition, the tests allow for only a subset of coefficients to change across segments. Moreover, the authors consider a sequential procedure that tests for l changes versus $l + 1$ changes and no change versus some unknown number up to some upper bound, m. For important classes of restrictions (which include partial structural change models and models where breaks occur only in a subset of the equation), the limiting distributions of the tests are dependent only on a parameter related to the number of coefficients allowed to change. This finding justifies the usage of the already available critical values.

In practice, following the testing procedure of Qu and Perron, 2007, one could determine the number of breaks in the system with a $UD_{max}LRT(M)$ statistics to test whether at least one break is present (Gómez-Loscos, Montañés, and Gadea, 2011; Gómez-Loscos, Gadea, and Montañés, 2012; and Gadea, Gómez-Loscos, and Montañés, 2016). If the test statistics rejects the presence of at least one break, the test $Seq_t(l+1|l)$ can be sequentially applied for $l = 1, 2 \ldots, m$, until it fails to reject the null hypothesis of no additional structural break. Additionally, one can calculate the $SupLR$ test statistics to test $l = 1, 2 \ldots, m$ versus $l = 0$, i.e., the null hypothesis of no change versus some unknown number of changes up to some upper bound, m.

Dynamic factor models are popular for analyzing the variability in large datasets. The presence of structural breaks in the model parameters has been a topic of debate. On one hand, there are studies arguing in favor of no accountability of structural breaks, as factor models are able to account for both breaks in the factor models of a subset of the underlying variables and moderate parameter drift in all the underlying variables (Stock and Watson, 2002). On the other hand, there are important studies arguing that the presence of structural breaks in the factor loadings is an issue that needs to be accounted for in the estimation methods if of main interest are the estimation of the common components or the transmission of common shocks to specific variables (Breitung and Eickmeier, 2011). Breitung et al. argue that structural breaks in the factor loadings increase the dimension of the factor space. The explanation is that if a structural break occurs in a sample, to represent the common components in the two subsamples (before and after the break) two sets of common factors are needed. Structural breaks in the factor loadings lead to inconsistent estimates of the loadings and to a larger dimension of the factor space. If the object of analysis is to find a more parsimonious factor representation of a large dataset that would allow to recover the original factors, the estimation method of the factor loadings has to account for the presence of structural breaks in the factor loadings themselves. Breitung et al. propose a battery of tests to inform about the existence of structural breaks in the factor loadings of dynamic factor models.

Hanzon, 1993; Commandeur, Koopman, and Ooms, 2011; Durbin and Koopman, 2012 and Ribarits and Hanzon, 2014a argue of numerous advantages of state-space modeling compared to VARMA representations. First of all, the state-space framework represents a larger and more flexible model class, in which VAR models are special cases. Second of all, the techniques available for the parametrization and estimation of the state-space models are much simpler compared

to those available for VARMA models. The methods to compute likelihood functions and their gradients are readily available. In practical applications, the processes describing the data will typically be neither VAR nor state-space models. However, the state-space approximation of these processes will in general result more parsimonious compared to the VAR approximation. The model selection procedures of state-space models are computationally cheap and can be efficiently used to derive good guesses for the initial parameter estimates for the likelihood optimization problem. Other convenient features of state-space methods are that they can handle two important aspects in time series analysis, i.e., forecasting and missing observations.

Commandeur and Koopman, 2007 and Commandeur, Koopman, and Ooms, 2011 explain that all significance tests in linear Gaussian models are based on the three assumptions of *independence*, *homoskedasticity*, and *normality*. In a state-space model, these assumptions should be checked on the so-called *standardized prediction errors*. *Structural breaks* and *outlier observations* in a state-space model can be detected by investigating the so-called *standardized smoothed state disturbances* (SSSD) and *standardized smoothed observation disturbances* (SSOD), respectively. Given an estimated state-space model, the standardized prediction errors can be calculated from a *forward pass* through the data with the Kalman filter, whereas the standardized smoothed observation disturbances and the standardized smoothed state disturbances can be calculated from a *backward pass* through the data using the output of the Kalman filter and smoothing algorithms.

Adopting a multivariate state-space framework, this Chapter advances further empirical results about the existence of outliers and structural breaks in the US and German yield curve drivers. To reach the results, we develop new data-driven state-space models for the co-movement of US and German yield curve drivers. The novelty of the models is that they are designed to preserve the dynamic properties of the US and German yield curve drivers embodied in the VAR model for the slopes and the VEC model for the levels and curvatures. We call the main version of the models the Full State-Space Model (FSSM). We estimate the FSSM via the Kalman filter and maximum likelihood and test for outliers and structural breaks in the FSSM using the SSOD and SSSD, respectively. It turns out that both outlying values and structural breaks are present in the FSSM; however, the alterations in structure resemble more of patches of outliers rather than structural breaks. We explain how to adjust the FSSM for the most blatant outliers by including intervention variables in the measurement equation. We call this new version of the FSSM the MShock-FSSM.

The Chapter is organized as follows. Section 7.2 summarizes the theory of state-space models and the Kalman filter needed for the development of the FSSM. The key points covered here are the state-space representation of a dynamic system, the derivation of the Kalman filter, maximum likelihood estimation of a state-space model's parameters and related issues, and backward recursion of the Kalman filter for the estimation of smoothed states and related quantities. Section 7.3 recalls the theoretical methods, which we will employ in our empirical analysis, for detecting outliers and structural breaks in state-space models. Section 7.4, dedicated to our empirical results, introduces the workflow, develops the FSSM, explains the initialization and estimation of the FSSM, tests and reports the results of outliers and structural breaks in the FSSM, and adjusts the FSSM for outliers and structural breaks by developing the MShock-FSSM. Section 7.5. concludes the Chapter.

7.2 State-Space Models and the Kalman Filter

This Section summarizes the theoretical concepts of state-space analysis and the Kalman filter, as an algorithm for sequentially updating a linear projection for a state-space system. Following mainly Mittnik, 1989; Mittnik, 1990; Harvey, 1990; Harvey, 1993; Hamilton, 1994; Commandeur and Koopman, 2007; Commandeur, Koopman, and Ooms, 2011; and Durbin and Koopman, 2012; we recall how a dynamic system can be written in state-space form and analyzed using the Kalman filter. Further on, we recall the derivation of the Kalman filter, the use of the filter in forecasting, the estimation of a state-space system via maximum likelihood, and the smoothing algorithm used to form the best inference about the unobserved state.

7.2.1 State-Space Representation of a Dynamic System

State-space models were originally developed in the field of control engineering, starting with the famous work of Kalman, 1960. In practice, these models are used to accurately track the position and velocity of moving objects such as rockets, missiles, airplanes, and ships. The usage of state-space models goes beyond control engineering, to find application in a wide range of time series analysis subjects in economics, finance, political science, environmental science, road safety, and medicine. Any autoregressive integrated moving average (ARIMA) model of Box et al., 2015 can be cast in state-space form, fit by state-space method, and analyzed by the Kalman filter.

The idea behind the models is that the object of analysis is a set of r *state variables* which change over time. Recalling the example from Harvey, 1993, these variables may be a signal, for example, the position of a rocket. As the signal, in most cases, is not observable, it is subject to systematic distortion and contamination by "noise".

Let $\boldsymbol{y}_t$ denote an $(n \times 1)$ vector of variables which are actually observed at date t and $\boldsymbol{\xi}_t$ the $(r \times 1)$ vector of *state variables*. The observed variables are related to the *state variables* by a *measurement equation*. The movements of the state variables (equivalently, the *state vector*) are governed by a well-defined process, called the *transition equation*. The measurement equation (7.2) and the transition equation (7.1) compose the *state-space representation*[2] of the dynamics of $\boldsymbol{y}$:

$$\underbrace{\boldsymbol{\xi_{t+1}}}_{r\times 1} = \underbrace{\mathbf{F}}_{r\times r}\underbrace{\boldsymbol{\xi_t}}_{r\times 1} + \underbrace{\mathbf{B}}_{r\times r}\underbrace{\mathbf{v_t}}_{r\times 1}, \tag{7.1}$$

$$\underbrace{\mathbf{y_t}}_{n\times 1} = \underbrace{\mathbf{A}'}_{n\times k}\underbrace{\mathbf{x_t}}_{k\times 1} + \underbrace{\mathbf{H}'}_{n\times r}\underbrace{\boldsymbol{\xi_t}}_{r\times 1} + \underbrace{\mathbf{D}}_{n\times n}\underbrace{\mathbf{w_t}}_{n\times 1} \tag{7.2}$$

where $\boldsymbol{F}$, $\boldsymbol{B}$, $\boldsymbol{A}'$, and $\boldsymbol{H}'$, and $\boldsymbol{D}$ are matrices of parameters of dimension $(r \times r)$, $(r \times r)$, $(n \times k)$, $(n \times r)$, and $(n \times r)$, respectively. Following Durbin and Koopman, 2012, matrix $\mathbf{F}$ is called the *state transition coefficient matrix*, specifying how the r states are expected to transition from period $t-1$ to t, for all $t = 1, \ldots, T$. Matrix $\mathbf{B}$ is the *state disturbance loading coefficient*

[2] Hamilton, 1994, p. 372.

matrix, specifying the additive error model for the state transition from period $t-1$ to t, for all $t = 1, \ldots, T$. Matrix $\mathbf{H}$ is the *measurement sensitivity coefficient matrix*, specifying how the r states are expected to combine at period t to form the n observations. Matrix $\mathbf{D}$ is the *observation innovation coefficient matrix*, specifying the error model for the observations for period t, for all $t = 1, \ldots, T$.

Following Hamilton, $\boldsymbol{x}_t$ is a $(k \times 1)$ vector of *exogenous* or *predetermined variables*, in the sense that, $\boldsymbol{x}_t$ provides no information about $\boldsymbol{\xi}_{t+s}$ or $\boldsymbol{w}_{t+s}$ for $s = 0, 1, 2, \ldots$ beyond that contained in $\boldsymbol{y}_{t-1}, \boldsymbol{y}_{t-2}, \ldots, \boldsymbol{y}_1$. $\boldsymbol{x}_t$ could include, for example, lagged values of the measurements $\boldsymbol{y}$ or variables that are uncorrelated with $\boldsymbol{\xi}_t$ and $\boldsymbol{w}_t$ for all τ.

The $(r \times 1)$ vector $\boldsymbol{v}_t$ and the $(n \times 1)$ vector $\boldsymbol{w}_t$ represent the disturbances in the transition and measurement equations, respectively, and are assumed to be vector white noise:

$$E(\boldsymbol{v}_t \boldsymbol{v}_\tau') = \begin{cases} \underbrace{\boldsymbol{Q}}_{r \times r} & \text{for} \quad t = \tau \\ \mathbf{0} & \text{otherwise} \end{cases}$$

$$E(\boldsymbol{w}_t \boldsymbol{w}_\tau') = \begin{cases} \underbrace{\boldsymbol{R}}_{n \times n} & \text{for} \quad t = \tau \\ \mathbf{0} & \text{otherwise} \end{cases}$$

where $\boldsymbol{Q}$ and $\boldsymbol{R}$ are $(r \times r)$ and $(n \times n)$ matrices, respectively. If we also assume that $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are *unit-variance* white noise processes, their covariance matrices $\boldsymbol{Q}$ and $\boldsymbol{R}$ are identity matrices. The disturbances $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ are assumed to be uncorrelated at all lags:

$$E(\boldsymbol{v}_t \boldsymbol{w}_\tau') = 0 \quad \text{for all} \quad t \quad \text{and} \quad \tau.$$

For given values of all system matrices and initial conditions for the state means and covariance matrix, the Kalman filter (Kalman, 1960) enables the estimation of the state vector in three different ways, to produce the *filtered*, the *predicted*, and the *smoothed* estimates of the state vector (Durbin and Koopman, 2012; Commandeur, Koopman, and Ooms, 2011). These estimates can be obtained by running one or two passes through the data. More specifically, a *forward pass* through the data, from $t = 1, \ldots, n$, with the Kalman filter calculates the *predicted states* (based on $y_1, \ldots, y_{t+1}$), *filtered states* (based on $y_1, \ldots, y_t$), and *observation prediction errors*. A *backward pass* through the data, from $t = n, \ldots, 1$, using the output of the Kalman filter and *state* and *disturbance smoothers*, calculates *smoothed estimates of the states* and *disturbances* at time point t, considering all available observations $\{y_1, y_2, \ldots, y_n\}$.

7.2.2 Derivation of the Kalman Filter

Forecasting the States $\widehat{\boldsymbol{\xi}}_{t+1|t}$

In the sequel, we recall the derivation of the Kalman filter as explained in Hamilton, 1994. Among many other uses, the Kalman filter[3] can be used as an algorithm for calculating linear least squares forecasts of the state vector, given the observed data through date t. Technically, given

[3] Hamilton, 1994, p. 377.

the general state-space system in 7.2 and 7.1, and known numerical values of all system matrices, the *linear least squares forecasts of the state vector* are defined as

$$\widehat{\boldsymbol{\xi}}_{t+1|t} \equiv \widehat{E}(\boldsymbol{\xi}_{t+1}|\boldsymbol{\mathcal{Y}}), \tag{7.3}$$

where $\boldsymbol{\mathcal{Y}} \equiv (\boldsymbol{y}_t', \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_1', \boldsymbol{x}_t', \boldsymbol{x}_{t-1}', \ldots, \boldsymbol{x}_1')'$ and $\widehat{\boldsymbol{\xi}}_{t+1|t}$ is the best forecast of $\boldsymbol{\xi}_{t+1}$ based on a constant and a linear function of $(\boldsymbol{y}_t', \boldsymbol{y}_{t-1}', \ldots, \boldsymbol{y}_1', \boldsymbol{x}_t', \boldsymbol{x}_{t-1}', \ldots, \boldsymbol{x}_1')$. With a forward pass through the data, the Kalman filter calculates the state forecasts recursively as $\widehat{\boldsymbol{\xi}}_{1|0}, \widehat{\boldsymbol{\xi}}_{2|1}, \ldots, \widehat{\boldsymbol{\xi}}_{T|T-1}$, and their associated $(r \times r)$ mean squared error (MSE) matrix:

$$\boldsymbol{P}_{t+1|t} \equiv E[(\boldsymbol{\xi}_{t+1} - \boldsymbol{\xi}_{t+1|t})(\boldsymbol{\xi}_{t+1} - \boldsymbol{\xi}_{t+1|t})']. \tag{7.4}$$

Starting the Kalman Filter

Hamilton recommends to start the Kalman filter recursion by considering the following.

If the eigenvalues of the state transition coefficient matrix $\boldsymbol{F}$ are inside the unit circle, the Kalman filter can be started with the unconditional mean of $\boldsymbol{\xi}_1$ (i.e., the forecast of $\boldsymbol{\xi}_1$ based on no observations of $\boldsymbol{y}$ or $\boldsymbol{x}$):

$$\widehat{\boldsymbol{\xi}}_{1|0} = E(\boldsymbol{\xi}_1) = 0 \tag{7.5}$$

and unconditional variance of $\boldsymbol{\xi}_1$,

$$\boldsymbol{P}_{1|0} = E\{[\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)][\boldsymbol{\xi}_1 - E(\boldsymbol{\xi}_1)]'\}. \tag{7.6}$$

The elements of the $(r \times r)$ matrix $\boldsymbol{P}_{1|0}$ can be expressed as a column vector given by

$$\text{vec}(\boldsymbol{P}_{1|0}) = [\boldsymbol{I}_{r^2} - (\boldsymbol{F} \otimes \boldsymbol{F})]^{-1} \cdot \text{vec}(\boldsymbol{Q}). \tag{7.7}$$

If some of the eigenvalues of the state transition coefficient matrix $\boldsymbol{F}$ are on or outside the unit circle, or if the initial state $\boldsymbol{\xi}_1$ is not considered as an arbitrary draw from the process implied by the transition matrix 7.1, the Kalman filter can be started with the analyst's best guess for the initial value of $\boldsymbol{\xi}_1$. In these cases, matrix $\boldsymbol{P}_{1|0}$ is a positive definite matrix reflecting the confidence of analyst's initial guess. Greater uncertainty about the true value of $\boldsymbol{\xi}_1$ is reflected in larger values for the diagonal elements of $\boldsymbol{P}_{1|0}$.

Given starting values $\widehat{\boldsymbol{\xi}}_{1|0}$ and $\boldsymbol{P}_{1|0}$, the Kalman filter is iterated to obtain state forecasts for $t = 2, 3, \ldots, T$. In general terms:

$$\begin{aligned} \widehat{\boldsymbol{\xi}}_{t+1|t} =& \boldsymbol{F}\widehat{\boldsymbol{\xi}}_{t|t-1} \\ &+ \boldsymbol{F}\boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}(\boldsymbol{y}_t - \boldsymbol{A}'\boldsymbol{x}_t - \boldsymbol{H}'\widehat{\boldsymbol{\xi}}_{t|t-1}) \end{aligned} \tag{7.8}$$

for $t = 1, 2, \ldots, T$. The MSE of these forecasts are given by matrix $\boldsymbol{P}_{t+1|t}$.

Forecasting the Measurements $\boldsymbol{y}_t$

The next step consists in producing *measurements forecasts*. The forecast of $\boldsymbol{y}_{t+1}$ is given by

$$\widehat{\boldsymbol{y}}_{t+1|t} \equiv \widehat{E}(\boldsymbol{y}_{t+1}|\boldsymbol{x}_{t+1}, \mathcal{Y}) = \boldsymbol{A}'\boldsymbol{x}_{t+1} + \boldsymbol{H}'\widehat{\boldsymbol{\xi}}_{t+1|t} \tag{7.9}$$

The corresponding *MSE of the observation forecasts* is given by

$$E[(\boldsymbol{y}_{t+1} - \widehat{\boldsymbol{y}}_{t+1|t})(\boldsymbol{y}_{t+1} - \widehat{\boldsymbol{y}}_{t+1|t})'] = \boldsymbol{H}'\boldsymbol{P}_{t+1|t}\boldsymbol{H} + \boldsymbol{R}. \tag{7.10}$$

Updating the Current Value of the State Vector $\boldsymbol{\xi}_t$

The current value of the state vector $\boldsymbol{\xi}_t$ is updated on the basis of the observation of the measurements $\boldsymbol{y}_t$. Technically, this is achieved by using the formula for updating a linear projection, producing the following *updated projection* of $\boldsymbol{\xi}_t$

$$\widehat{\boldsymbol{\xi}}_{t|t} = \widehat{\boldsymbol{\xi}}_{t|t-1} + \boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}(\boldsymbol{y}_t - \boldsymbol{A}'\boldsymbol{x}_t - \boldsymbol{H}'\widehat{\boldsymbol{\xi}}_{t|t-1}). \tag{7.11}$$

The corresponding MSE is given by

$$\boldsymbol{P}_{t|t} = \boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}\boldsymbol{H}'\boldsymbol{P}_{t|t-1}. \tag{7.12}$$

Producing a Forecast of $\boldsymbol{\xi}_{t+1}$ and the Raw Kalman Gain

The transition equation in 7.1 is used to forecast $\boldsymbol{\xi}_{t+1}$:

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{t+1|t} &= \widehat{E}(\boldsymbol{\xi}_{t+1|t}|\boldsymbol{\mathcal{Y}_t}) && (7.13)\\
&= \boldsymbol{F}\widehat{E}(\boldsymbol{\xi}_t|\boldsymbol{\mathcal{Y}}_t) + \widehat{E}(\boldsymbol{v}_{t+1}|\boldsymbol{\mathcal{Y}}_t) && (7.14)\\
&= \boldsymbol{F}\widehat{\boldsymbol{\xi}}_{t|t} + \boldsymbol{0}.
\end{aligned}$$

Substituting $\widehat{\boldsymbol{\xi}}_{t|t}$ with the equation in 7.11, the forecast of $\boldsymbol{\xi}_{t+1}$ can be written as

$$\begin{aligned}
\widehat{\boldsymbol{\xi}}_{t+1|t} =& \boldsymbol{F}\hat{\boldsymbol{\xi}}_{t|t-1} \\
&+ \boldsymbol{F}\boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}(\boldsymbol{y}_t - \boldsymbol{A}'\boldsymbol{x}_t - \boldsymbol{H}'\widehat{\boldsymbol{\xi}}_{t|t-1})
\end{aligned} \tag{7.15}$$

where the coefficient matrix denotes the *gain matrix*, $\boldsymbol{K}_t$:

$$\boldsymbol{K}_t \equiv \boldsymbol{F}\boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}. \tag{7.16}$$

where $\boldsymbol{P}_{t|t-1}$ is the estimated variance-covariance matrix of the state forecasts, given all information up to period $t-1$. The raw Kalman gain is a matrix that designates how much to weigh the observations during recursions of the Kalman filter, so that the filtered states at period t are close to the corresponding state forecasts.

With the definition of the Kalman gain, the *forecast* of $\boldsymbol{\xi}_{t+1}$ can be written in the following more compact way:

$$\widehat{\boldsymbol{\xi}}_{t+1|t} = \boldsymbol{F}\widehat{\boldsymbol{\xi}}_{t|t-1} + \boldsymbol{K}_t(\boldsymbol{y}_t - \boldsymbol{A}'\boldsymbol{x}_t - \boldsymbol{H}'\widehat{\boldsymbol{\xi}}_{t|t-1}). \tag{7.17}$$

The corresponding *MSE* of this forecast is given by

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{F}[\boldsymbol{P}_{t|t-1} - \boldsymbol{P}_{t|t-1}\boldsymbol{H}(\boldsymbol{H}'\boldsymbol{P}_{t|t-1}\boldsymbol{H} + \boldsymbol{R})^{-1}\boldsymbol{H}'\boldsymbol{P}_{t|t-1}]\boldsymbol{F}' + \boldsymbol{Q}. \tag{7.18}$$

The *s-period-ahead state forecasts*[4] are estimates of the states at period t using all information (e.g., the measurements) up to period $t-s$ and are given by the following equation

$$\boldsymbol{\xi}_{t+s} = \mathbf{F}^s\boldsymbol{\xi}_t + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^{s-2}\mathbf{v}_{t+2} + \cdots + \mathbf{F}^1\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s} \tag{7.19}$$

for $s = 1, 2, \ldots$. From which the *s-period-ahead forecast error* for the state vector is

$$\begin{aligned}\boldsymbol{\xi}_{t+s} - \widehat{\boldsymbol{\xi}}_{t+s|t} =& \mathbf{F}^s(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_{t|t}) + \mathbf{F}^{s-1}\mathbf{v}_{t+1} + \mathbf{F}^{s-2}\mathbf{v}_{t+2} \\ &+ \cdots + \mathbf{F}^s\mathbf{v}_{t+s-1} + \mathbf{v}_{t+s}\end{aligned} \tag{7.20}$$

with *MSE*

$$\begin{aligned}\mathbf{P}_{t+s|t} =& \mathbf{F}^s\mathbf{P}_{t|t}(\mathbf{F}')^s + \mathbf{F}^{s-1}\mathbf{Q}(\mathbf{F}')^{s-1} + \mathbf{F}^{s-2}\mathbf{Q}(\mathbf{F}')^{s-2} \\ &+ \cdots + \mathbf{F}\mathbf{Q}\mathbf{F}' + \mathbf{Q}.\end{aligned} \tag{7.21}$$

If the state vector is defined in such a way that $\mathbf{x}_t$ is deterministic, the *s-period-ahead-forecast* of $\mathbf{y}$ is

$$\widehat{\mathbf{y}}_{t+s|t} \equiv \widehat{E}(\mathbf{y}_{t+s}|\mathcal{Y}_t) = \mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\widehat{\boldsymbol{\xi}}_{t+s|t}, \tag{7.22}$$

from which the *s-period-ahead forecast error* is

$$\begin{aligned}\mathbf{y}_{t+s} - \widehat{\mathbf{y}}_{t+s|t} &= (\mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\boldsymbol{\xi}_{t+s} + \mathbf{w}_{t+s}) - (\mathbf{A}'\mathbf{x}_{t+s} + \mathbf{H}'\boldsymbol{\xi}_{t+s|t}) \\ &= \mathbf{H}'(\boldsymbol{\xi}_{t+s} - \widehat{\boldsymbol{\xi}}_{t+s|t}) + \mathbf{w}_{t+s}\end{aligned} \tag{7.23}$$

with *MSE*

$$E[(\mathbf{y}_{t+s} - \widehat{\mathbf{y}}_{t+s|t})(\mathbf{y}_{t+s} - \widehat{\mathbf{y}}_{t+s|t})'] = \mathbf{H}'\mathbf{P}_{t+s|t}\mathbf{H} + \mathbf{R}. \tag{7.24}$$

[4] Hamilton, 1994, p. 384.

7.2.3 Maximum Likelihood Estimation of Parameters

Evaluation of the Likelihood Function via the Kalman Filter

The unknown parameters in **F, B, Q, A, H, D, R** can be estimated using the Kalman filter and maximum likelihood[5]. The maximization of the sample log likelihood[6] is started by assigning an initial guess to the numerical values of the unknown parameters. For the given initial values of the unknown parameters, matrices **F, B, Q, A, H, D, R** are constructed from the expressions in 7.15, 7.18, 7.22, and 7.24 and iterated for $t = 1, 2, \ldots, T-1$ to produce the sequences $\{\widehat{\boldsymbol{\xi}}_{t|t-1}\}_{t=1}^{T}$ and $\{\mathbf{P}_{t|t-1}\}_{t=1}^{T}$. These sequences can then be used in the distribution function of $\mathbf{y}_t$ conditional on $(\mathbf{x}_t, \mathcal{Y}_{t-1})$[7] and the sample log likelihood to compute the value for the log likelihood function that results from the assigned initial parameter values.

To aid the maximum likelihood estimation of the state-space model, the crude set of initial parameters can be refined using refinement algorithms such as *loose bound interior point*, *Nelder-Mead algorithm*, *Quasi-Newton*, *starting value perturbation*, or *starting value shrinkage*. These algorithms fit the state-space model to the crude set of initial parameter values and produce sets of initial parameter values. Hamilton suggests that the numerical search is better behaved if $\boldsymbol{\Omega}$ is initialized in terms of its Cholesky factorization. The analyst should choose the refined set that yields the highest loglikelihood.

State-Space Model Identification

The state-space representation of a system provides a very convenient way to calculate the exact likelihood function. Nevertheless, in the absence of restrictions on **F, B, Q, A, H, D, R**, identification issues[8] are very common. More than one set of parameter values can give rise to the same value of the likelihood function. Rothenberg, 1971 explains two types of absence of identification: *global identification* and *local identification*. Global identification implies local identification and it can be shown that a model is locally identified at a particular parameter value $\boldsymbol{\theta}_0$ if and only if the information matrix is nonsingular in a neighborhood around $\boldsymbol{\theta}_0$. Hamilton, 1994 explains that, when estimating a state-space model, difficulty with inverting the matrix of second derivatives of the log likelihood function is an indication that an unidentified model is tried to be estimated. If such an indication exists, the analyst can check for local identification by converting the state-space representation back to a vector ARMA model and then check that the conditions in Hannan, 1971 are satisfied. Local identification can also be checked directly with the state-space representation, by following the approach described in Gevers and Wertz, 1984 and Wall, 1987.

From Hamilton, 1994, we recall that the maximum likelihood estimate $\widehat{\boldsymbol{\theta}}_T$ based on a sample of size T is consistent and asymptotically normal (as shown in Caines, 2018) if the following regularity conditions are satisfied: (1) the model is identified; (2) all eigenvalues of $\mathbf{F}$ lie inside the unit circle; (3) the regressors in $\mathbf{x}_t$ behave asymptotically like a full-rank linearly indeterministic

[5] Hamilton, 1994, p. 385.

[6] The sample log likelihood is given in Equation [13.4.2] in Hamilton, 1994, p. 386.

[7] The distribution function of $\mathbf{y}_t$ conditional on $(\mathbf{x}_t, \mathcal{Y}_{t-1})$ is given in Equation [13.4.1] in Hamilton, 1994, p. 385.

[8] Hamilton, 1994, p. 387.

covariance-stationary process; and (4) the true value of $\boldsymbol{\theta}$ are inside the boundaries of the allowable parameter space.

Moreover, the Kalman filter can be used to compute the linear projection of $\mathbf{y}_{t+s}$ on past observations, even in the presence of non-Gaussian state and observation disturbances (i.e., $\mathbf{v}_t$ and $\mathbf{w}_t$, respectively, in 7.1 and 7.2). The sample log likelihood function can still be maximized with respect to θ, even for non-Gaussian systems. The estimation of the elements in **F, B, Q, A, H, D** and **R** would still be consistent and asymptotically normal.

7.2.4 State Smoothing

In some cases, it might be of interest to give a structural interpretation to the state vector $\boldsymbol{\xi}_t$[9]. A backward pass through the data with the Kalman filter allows to form an inference about the value of $\boldsymbol{\xi}_t$, considering the full set of data collected, including observations on $\mathbf{y}_t, \mathbf{y}_{t+1}, \ldots, \mathbf{y}_T, \mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_T$. This kind of inference is called the *smoothed state*, denoted

$$\widehat{\boldsymbol{\xi}}_{t|T} \equiv \widehat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T), \tag{7.25}$$

with associated *MSE*:

$$\mathbf{P}_{t|t-1} \equiv E[(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_{t|t-1})(\boldsymbol{\xi}_t - \widehat{\boldsymbol{\xi}}_{t|t-1})']. \tag{7.26}$$

To recall the recursive formulae for *state smoothing*, as in Hamilton, 1994, p. 394, for convenience, we first list the key formulae for the Kalman filter:

$$\widehat{\boldsymbol{\xi}}_{t|t} = \widehat{\boldsymbol{\xi}}_{t|t-1} + \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}(\mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}'\widehat{\boldsymbol{\xi}}_{t|t-1}) \tag{7.27}$$

$$\widehat{\boldsymbol{\xi}}_{t+1|t} = \mathbf{F}\widehat{\boldsymbol{\xi}}_{t|t} \tag{7.28}$$

$$\mathbf{P}_{t|t} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1}\mathbf{H}(\mathbf{H}'\mathbf{P}_{t|t-1}\mathbf{H} + \mathbf{R})^{-1}\mathbf{H}'\mathbf{P}_{t|t-1} \tag{7.29}$$

$$\mathbf{P}_{t+1|t} = \mathbf{F}\mathbf{P}_{t|t}\mathbf{F}' + \mathbf{Q}. \tag{7.30}$$

If we consider the estimate of $\widehat{\boldsymbol{\xi}}_t$ based on observations through date t, $\widehat{\boldsymbol{\xi}}_{t|t}$ and we suppose that subsequently we know the true value of $\boldsymbol{\xi}_{t+1}$, the *new estimate of* $\widehat{\boldsymbol{\xi}}_t$ could be expressed as

$$\widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_t) = \widehat{\boldsymbol{\xi}}_{t|t} + \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}^{-1}_{t+1|t}(\boldsymbol{\xi}_{t+1} - \widehat{\boldsymbol{\xi}}_{t+1|t}). \tag{7.31}$$

Defining

$$\mathbf{J}_t \equiv \mathbf{P}_{t|t}\mathbf{F}'\mathbf{P}^{-1}_{t+1|t}\boldsymbol{\xi}_{t+1}, \tag{7.32}$$

[9]For example, in the model of Stock and Watson for the business cycle, it is of interest to known the state of the business cycle at any historical date t.

7.31 can be written more compactly as

$$\widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_t) = \widehat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1} - \widehat{\boldsymbol{\xi}}_{t+1|t}). \tag{7.33}$$

One point to notice is that, knowing $\mathbf{y}_{t+j}$ or $\mathbf{x}_{t+j}$ for $j > 0$ does not add any value if we already know the value of $\boldsymbol{\xi}_{t+1}$. Therefore, the linear projection in 7.33 is identical to $\widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_T)$. Thus,

$$\widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_T) = \widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_t) = \widehat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1} - \widehat{\boldsymbol{\xi}}_{t+1|t}). \tag{7.34}$$

The *smoothed estimate* of $\boldsymbol{\xi}_t$ can be obtained by projecting $\widehat{E}(\widehat{\boldsymbol{\xi}}_t|\widehat{\boldsymbol{\xi}}_{t+1}, \mathcal{Y}_T)$ on $\mathcal{Y}_T$. This is a trivial result that reads as follows

$$\widehat{E}(\boldsymbol{\xi}_t|\mathcal{Y}_T) = \widehat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t[\widehat{E}(\boldsymbol{\xi}_{t+1}|\mathcal{Y}_T) - \boldsymbol{\xi}_{t+1|t}], \tag{7.35}$$

or

$$\boldsymbol{\xi}_{t|T} = \widehat{\boldsymbol{\xi}}_{t|t} + \mathbf{J}_t(\boldsymbol{\xi}_{t+1|T} - \boldsymbol{\xi}_{t+1|t}). \tag{7.36}$$

The steps to perform, in order to generate the sequence of smoothed estimates, $\{\widehat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$, can be summarized in the following way. Step 1: Calculate the Kalman filter, i.e., equations 7.27 to 7.30 and store the sequences $\{\widehat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$, $\{\widehat{\boldsymbol{\xi}}_{t+1|t}\}_{t=0}^{T-1}$, $\{\mathbf{P}_{t|t}\}_{t=1}^{T}$, and $\{\mathbf{P}_{t+1|t}\}_{t=0}^{T-1}$. The smoothed estimate for the last date in the sample, $\widehat{\boldsymbol{\xi}}_{T|T}$, is the last entry in $\{\widehat{\boldsymbol{\xi}}_{t|t}\}_{t=1}^{T}$. Step 2: Use the definition in 7.32 to generate $\{\mathbf{J}_t\}_{t=1}^{T-1}$. Step 3: Use $\{\mathbf{J}_t\}_{t=1}^{T-1}$ with 7.36 for $t = T-1$ to calculate

$$\widehat{\boldsymbol{\xi}}_{T-1|T} = \widehat{\boldsymbol{\xi}}_{T-1|T-1} + \mathbf{J}_{T-1}(\widehat{\boldsymbol{\xi}}_{T|T} - \widehat{\boldsymbol{\xi}}_{T|T-1}). \tag{7.37}$$

Step 4: Having $\widehat{\boldsymbol{\xi}}_{T-1|T}$ from Step 3, use 7.36 for $t = T-2$ to calculate

$$\widehat{\boldsymbol{\xi}}_{T-2|T} = \widehat{\boldsymbol{\xi}}_{T-2|T-2} + \mathbf{J}_{T-2}(\widehat{\boldsymbol{\xi}}_{T-1|T} - \widehat{\boldsymbol{\xi}}_{T-1|T-2}). \tag{7.38}$$

Step 5: Iterate backward through the data to calculate the *full set of smoothed states*, $\{\widehat{\boldsymbol{\xi}}_{t|T}\}_{t=1}^{T}$. Step 6: Calculate the *MSE* associated with the smoothed estimate as

$$\mathbf{P}_{t|T} = \mathbf{P}_{t|t} + \mathbf{J}_t(\mathbf{P}_{t+1|T} - \mathbf{P}_{t+1|t})\mathbf{J}_t'. \tag{7.39}$$

Step 7: Iterate backward through the data, starting with $t = T-1$, to calculate the *full set of MSE*, $\{\mathbf{P}_{t|T}\}_{t=1}^{T}$.

7.2.5 State-Space Modeling vs. Bayesian Econometrics

In analogy to state-space models (in which the object of analysis is a set of r state variables, changing over time, unobservable, and subject to systematic distortion and contamination by "noise"), Bayesian econometrics assumes that the quantity of interest is not measured directly and the measured data are corrupted by noise (Grewal, 2011; Greenberg, 2012). Following Bayes

rule, this implies that uncertainty in input is transformed into uncertainty in inference. In Bayesian inference, the Bayes theorem is used to compute the probability of an output B given measurements A. More specifically, the Bayes theorem states that

$$P(B|A) = \frac{P(A|B)P(B)}{P(A)}, \tag{7.40}$$

where $P(B|A)$ is the *posterior probability*, $P(A|B)$ is the *likelihood*, and $P(B)$ is the *prior probability* without any evidence from measurements. The likelihood $P(A|B)$ evaluates the measurements given an output B. The posterior probability $P(B|A)$ is the probability of B after taking the measurement A into account.

As explained above, the unknown parameters of a state-space model can be estimated with the Kalman filter and maximum likelihood. The Kalman filter estimates the true values of states recursively over time using incoming measurements and a mathematical process model (in our case, the state-space model). Analogously, recursive Bayesian estimation computes estimates of an unknown probability density function (PDF) recursively over time using incoming measurements and a mathematical process model. In the Bayesian framework, the true state is assumed to be an unobserved Markov process, whereas, the measurements are the observed states of a hidden Markov model (HMM). The Kalman filter consists of four steps, namely, *initialization*, *prediction*, *correction* (or update), and *forecasting* (Lütkepohl, 2005).

In the initialization step, in order to start the maximization of the sample log likelihood, initial guesses are assigned to the numerical values of the unknown parameters. The equivalent concept in Bayesian econometrics is the prior density, which is not dependent on the data and contains any non-data information available about the parameters of the model to be estimated. The next steps of the Kalman filter are the prediction and correction steps. In the prediction step, the state estimate from the previous timestep is used to produce an estimate of the state at the current timestep. Since the predicted state estimate does not include observation information from the current timestep, the predicted state estimate is also known as the *a priori* state estimate. In the correction step, the state prediction is refined by combining the current a priori prediction with current observation information. In the forecasting step of the Kalman filter, s-step-ahead state and observation forecasts are calculated at period t using all information available up to period $t - s$.

Likewise, in Bayesian econometrics, the distribution of a new, unobserved data point is predicted using the posterior predictive distribution. As new information becomes available, the posterior distribution is updated in the so-called *Bayesian updating step*, i.e., the posterior distribution becomes the prior for the next prediction step.

7.3 Structural Breaks and Outliers in State-Space Models

7.3.1 Diagnostic Checking using the Auxiliary Residuals

The adequacy of a fitted time series model is normally assessed using the innovations, equivalently, the *one-step-ahead prediction errors*. The adequacy of a state-space model can be assessed

by investigating the so-called *auxiliary residuals*, i.e., the *standardized smoothed observation disturbances* (SSODs) and the *standardized smoothed state disturbances* (SSSDs). The auxiliary residuals of a state-space model are the estimators of the disturbances associated with the unobserved components and are functions of the innovations. An investigation of these residuals can unveil features of a fitted model that are not explicitly available from the innovations themselves. Aberrant behavior commonly observed in time series, such as outlier observations, level shifts, and switches, can easily be detected using plots of test statistics based on the auxiliary residuals. Harvey and Koopman, 1992; De Jong and Penzer, 1998; Commandeur and Koopman, 2007; and Commandeur, Koopman, and Ooms, 2011, explain the use of SSODs and SSSDs to detect outliers and structural breaks in state-space models.

Given an estimated state-space model, the SSODs and SSSDs can be obtained from a backward recursion through the data with the output of the Kalman filter and the state and disturbance smoothing algorithms. Before recalling the relevant formulae from Commandeur, Koopman, and Ooms, 2011, for convenience, we first report the key equations of the Kalman filter:

$$\begin{aligned} \mathbf{m}_t &= \mathbf{y}_t - \mathbf{A}'\mathbf{x}_t - \mathbf{H}_t\boldsymbol{\xi}_t & \mathbf{V}_t &= \mathbf{H}_t\mathbf{P}_t\mathbf{H}_t^T + \mathbf{R}_t \\ \mathbf{K}_t &= \mathbf{F}_t\mathbf{P}_t\mathbf{H}_t^T\mathbf{V}_t^{-1} & \mathbf{L}_t &= \mathbf{F}_t - \mathbf{K}_t\mathbf{H}_t \\ \boldsymbol{\xi}_{t+1} &= \mathbf{F}_t\boldsymbol{\xi}_t + \mathbf{K}_t\mathbf{m}_t & \mathbf{P}_{t+1} &= \mathbf{F}_t\mathbf{P}_t\mathbf{L}_t^T + \mathbf{B}_t\mathbf{Q}_t\mathbf{R}_t^T \end{aligned} \tag{7.41}$$

for $t = 1, \ldots, n$. In 7.41, $\mathbf{m}_t$ are the *one-step-ahead prediction errors*[10]; $\mathbf{V}_t$ the *variances* of $\mathbf{m}_t$; $\boldsymbol{\xi}_t$ denote the predicted states and $\mathbf{P}_t$ the *estimated error variance matrix* of the predicted state $\boldsymbol{\xi}_t$. Assuming normality, $\mathbf{P}_t$ is employed in the construction of confidence intervals for the predicted state. For example, the 90% confidence limits for the predicted state can be calculated as

$$\boldsymbol{\xi}_t \pm 1.64\sqrt{\mathbf{P}_t} \tag{7.42}$$

for $t = 1, \ldots, n$.

The main purpose of state and disturbance smoothing algorithms is to produce estimated values of the state and disturbance vectors at time point t, considering all available observations $\{y_1, y_2, \ldots, y_n\}$. The recursive formulas for state smoothing are the following:

$$\begin{aligned} \mathbf{r}_{t-1} &= \mathbf{H}_t^T\mathbf{V}_t^{-1}\mathbf{m}_t + \mathbf{H}_t^T\mathbf{r}_t & \mathbf{N}_{t-1} &= \mathbf{H}_t^T\mathbf{V}_t^{-1}\mathbf{H}_t + \mathbf{L}_t^T\mathbf{N}_t\mathbf{L}_t \\ \widehat{\boldsymbol{\xi}}_t &= \boldsymbol{\xi}_t + \mathbf{P}_t\mathbf{r}_{t-1} & \mathbf{W}_t &= \mathbf{P}_t - \mathbf{P}_t\mathbf{N}_{t-1}\mathbf{P}_t \end{aligned} \tag{7.43}$$

for $t = 1, \ldots, n$. In 7.43, $\widehat{\boldsymbol{\xi}}_t$ is the *smoothed state estimate*, i.e., optimal estimate of $\boldsymbol{\xi}_t$ using the full set of observations $\{y_1, y_2, \ldots, y_n\}$; $\mathbf{W}_t$ is the smoothed state estimation error variance matrix. Assuming normality, $\mathbf{W}_t$ is employed in the construction of confidence intervals for the smoothed state components. For example, the 90% confidence limits for the smoothed state components can be calculated as

$$\widehat{\boldsymbol{\xi}}_t \pm 1.64\sqrt{\mathbf{W}_t} \tag{7.44}$$

[10] In case of exogenous variables (i.e., the term $\mathbf{x}_t$) in the measurement equation of the state-space model, the term $\mathbf{A}'\mathbf{x}_t$ needs to be subtracted as well from $\boldsymbol{y}_t$ in the equation of $\mathbf{m}_t$. The term $\mathbf{A}'\mathbf{x}_t$, however, does not change the variance $\mathbf{V}_t$, as it is assumed to be deterministic at time t.

for $t = 1, \ldots, n$. The recursions for $\mathbf{r}_{t-1}$ and $\mathbf{N}_{t-1}$ in 7.43 allow the computation of the *smoothed observation disturbances* ($\widehat{\mathbf{w}}_t$) and their corresponding *smoothed estimation error variance matrix* (Var($\widehat{\mathbf{w}}_t$)):

$$\begin{aligned} \widehat{\mathbf{w}}_t &= \mathbf{R}_t(\mathbf{V}_t^{-1}\mathbf{m}_t - \mathbf{K}_t^T\mathbf{r}_t) \\ \text{Var}(\widehat{\mathbf{w}}_t) &= \mathbf{R}_t(\mathbf{V}_t^{-1} + \mathbf{K}_t^T\mathbf{N}_t\mathbf{K}_t)\mathbf{R}_t, \end{aligned} \tag{7.45}$$

and the *smoothed state disturbances* ($\widehat{\mathbf{v}}_t$) and their corresponding *smoothed estimation error variance matrix* (Var($\widehat{\mathbf{v}}_t$)):

$$\begin{aligned} \widehat{\mathbf{v}}_t &= \mathbf{Q}_t\mathbf{B}_t^T\mathbf{r}_t \\ \text{Var}(\widehat{\mathbf{v}}_t) &= \mathbf{Q}_t\mathbf{B}_t^T\mathbf{N}_t\mathbf{B}_t\mathbf{Q}_t, \end{aligned} \tag{7.46}$$

for $t = n, \ldots, 1$. *Outlier observations* and *structural breaks* in a state-space model can be detected using the *standardized smoothed observation disturbances* ($\mathbf{e}_t^*$) and the *standardized smoothed state disturbances* ($\mathbf{r}_t^*$), respectively:

$$\mathbf{e}_t^* = \frac{\hat{\mathbf{w}}_t}{\sqrt{Var(\hat{\mathbf{w}}_t)}} \quad \text{and} \quad \mathbf{r}_t^* = \frac{\hat{\mathbf{v}}_t}{\sqrt{Var(\hat{\mathbf{v}}_t)}}, \tag{7.47}$$

for $t = 1, \ldots, n$. Harvey and Koopman, 1992 show that these residuals should follow (asymptotically) a t-distribution. By plotting them together with the 95% confidence intervals for a t-distribution, one can easily detect observation outliers and/or structural breaks. More specifically, $\mathbf{e}_t^*$ can be inspected for the detection of observation outliers, by considering them as a t test for the null hypothesis that there was no outlier observation. In the same vein, $\mathbf{r}_t^*$ can be inspected for the detection of structural breaks, by considering them as a t test for the null hypothesis that there was no structural break in the corresponding unobserved component of the observed time series. Harvey and Koopman, 1992 explain that the basic detection procedure of outliers and/or structural breaks is to plot the auxiliary residuals after they have been standardized. In a Gaussian model, values greater than 2 in absolute value would provide indications of outliers and/or structural breaks. The residuals at the end and at the beginning of the sample period will tend to have a higher variance. Harvey et al. explain further that the standardized innovations may also indicate outliers and structural change but will not normally indicate the source of the problem. A formal detection procedure for unusually large residuals would require conducting a test for kurtosis and skewness, which boils down to the Bowman-Shenton test for normality. An allowance for serial correlation is necessary, in order to make the tests asymptotically valid.

7.3.2 Introducing Shocks in State-Space Models

Outliers or changes in the structure of a time series can create bias in the sample autocorrelation function (Lefrancois, 1991) and lead to problems with model identification (Tsay, 1986), thus undermining the reliability of parameter estimates and forecasts (Harvey, 1990). The seminal work in the field of outliers in time series is attributed to Fox, 1972, who proposes two models to characterize outliers as additive or innovative. Box and Tiao, 1975 discuss the possibility of permanent changes in structure and employ difference equation models to represent possible

intervention effects. De Jong and Penzer, 1998 develop methods based on state-space forms to detect and model unusual behaviors commonly observed in practice, such as, outlier observations, level shifts, and switches. Their methods require comparing a fitted null state-space model to an alternative model, which incorporates a vector of shocks, $\boldsymbol{\delta}$, to represent the suspected inadequacy in the null.

To illustrate the theoretical concepts of the methods of de Jong et al., let us assume that an appropriate representation of the process that generates the data $\boldsymbol{y} = (\boldsymbol{y}_1', \ldots, \boldsymbol{y}_n')'$ is the state-space model in 7.1 and 7.2, which we recall here for convenience:

$$\underbrace{\boldsymbol{\xi}_{t+1}}_{r\times 1} = \underbrace{\mathbf{F}}_{r\times r}\underbrace{\boldsymbol{\xi}_t}_{r\times 1} + \underbrace{\mathbf{B}}_{r\times r}\underbrace{\mathbf{v_t}}_{r\times 1}, \tag{7.48}$$

$$\underbrace{\mathbf{y_t}}_{n\times 1} = \underbrace{\mathbf{A}'}_{n\times k}\underbrace{\mathbf{x_t}}_{k\times 1} + \underbrace{\mathbf{H}'}_{n\times r}\underbrace{\boldsymbol{\xi}_t}_{r\times 1} + \underbrace{\mathbf{D}}_{n\times n}\underbrace{\mathbf{w_t}}_{n\times 1}. \tag{7.49}$$

The alternative model is the null model extended to include the vector of shocks, $\boldsymbol{\delta}$. More specifically,

$$\underbrace{\boldsymbol{\xi}_{t+1}}_{r\times 1} = \boldsymbol{\Gamma}_t\boldsymbol{\delta} + \underbrace{\mathbf{F}}_{r\times r}\underbrace{\boldsymbol{\xi}_t}_{r\times 1} + \underbrace{\mathbf{B}}_{r\times r}\underbrace{\mathbf{v_t}}_{r\times 1}, \tag{7.50}$$

$$\underbrace{\mathbf{y_t}}_{n\times 1} = \boldsymbol{\Lambda}_t\boldsymbol{\delta} + \underbrace{\mathbf{A}'}_{n\times k}\underbrace{\mathbf{x_t}}_{k\times 1} + \underbrace{\mathbf{H}'}_{n\times r}\underbrace{\boldsymbol{\xi}_t}_{r\times 1} + \underbrace{\mathbf{D}}_{n\times n}\underbrace{\mathbf{w_t}}_{n\times 1}, \tag{7.51}$$

where $\boldsymbol{\Gamma}_t$ and $\boldsymbol{\Lambda}_t$ are called the *shock design matrices* and $\boldsymbol{\delta}$ is the shock magnitude. Outliers, level shifts, and switches can be accounted for by including in the state-space model these intervention variables (also called *intervention signature*). For example, an outlying value in the measurements[11] can be modeled by a measurement intervention. The intervention signature in this case is such that $\boldsymbol{\Lambda}_t$ is taken to be zero everywhere except at the single point $t = i$ (denoting the outlying value), where $\boldsymbol{\Lambda}_t$ equals one. $\boldsymbol{\Gamma}_t$ is taken to be zero everywhere. A permanent shift[12] in the mean of a series can be modeled by an intervention signature that takes the value zero up to the point of the shift and one thereafter. Consecutive extreme values[13] on either side of the current level of a time series can be modeled by a switch intervention.

More generally, a *pure measurement shock* at time $t = i$ can be implemented by setting $\boldsymbol{\Gamma}_t = \mathbf{0}$ and $\boldsymbol{\Lambda}_t = \boldsymbol{I}$; whereas a *pure state shock*, in which each component of the state vector is shocked separately, can be implemented by setting $\boldsymbol{\Gamma}_t = \boldsymbol{I}$ and $\boldsymbol{\Lambda}_t = \mathbf{0}$.

de Jong et al. explain that once a state-space model is extended with the additional structure to account for outliers and structural breaks, in the spirit of Box and Tiao, 1975, the new state-space model should be estimated again with the Kalman filter and the diagnostic procedure based on the auxiliary residuals should be repeated.

[11] For example, an incorrect recording of the data.

[12] Examples include the Nile data of Cobb, 1978 and the British seatbelt data of Harvey and Durbin, 1986.

[13] Examples include an increased production after a strike or a sudden rise in a stock market followed by a collapse.

In regression analysis, a highly accepted idea is that unusual observations may occur in patches. Such structures are normally handled with leave-k-out diagnostics that involve deletion of observations (Cook and Weisberg, 1982 and Atkinson, 1985). In time series data, Justel, Peña, and Tsay, 2001 identify the beginning and end of potential outlier patches using Gibbs sampling and then handle the patches of outliers by means of an adaptive procedure with block interpolation. Proietti, 2003 develops an efficient and easy to implement algorithm based on a reverse run of the Kalman filter on the smoothing errors to calculate leave-k-out diagnostics for the detection of patches of outliers in state-space models.

Generally, the leave-k-out approaches assume that the dynamics of a process are more or less identical on either side of any unusual points. However, Penzer, 2007 explains that, in time series data, a patch of outliers may rise from a level shift, a seasonal break, or any other permanent alteration in structure. To this regard, Penzer argues that patches of unusual behavior should be represented by allowing shocks in the measurement equation of a state-space model. A more general approach, which accounts for more persistent behavior, is to allow shocks to the transition equation.

For the detection of persistent departures from a null state-space model, Penzer proposes the so-called put-k-shocks-in framework, which includes the leave-k-out as a special case. Building on the results of De Jong and Penzer, 1998 and Proietti, 2003, Penzer shows that diagnostics associated with deleting k observations are identical to those derived by introducing k measurement shocks. Therefore, put-k-shocks-in is a good generalized procedure, which can be simplified to leave-k-out where needed.

7.4 Empirical Results

7.4.1 Workflow

Our multivariate state-space analysis is motivated by the results of the univariate analysis, which revealed the presence of multiple structural breaks. Furthermore, a correlation analysis of the residuals from the fitted 2D-VAR(5) model for the slopes, the 2D-VEC(1) model for the levels, and the 4D-VEC(3) model for the levels and curvatures revealed correlation coefficients significantly different from zero. These findings might suggest that the multiple alterations in the structure of US and German yield curve drivers are due to missing variables with potentially predictive power. Therefore, in order to derive more accurate results with respect to the existence of structural breaks in the US and German yield curve drivers, we proceed with the workflow depicted in Figure 7.1.

Our modeling idea is to check for structural breaks in one *full state-space model* (the FSSM) for the US and German yield curve drivers. The FSSM is composed of two sub-models: the *state-space VAR* (SSVAR) model for the slopes and the *state-space VEC* (SSVEC) model for the levels and curvatures. Given this structure, the FSSM has the nice feature of preserving the original dynamics found for the US and German yield curve drivers. We estimate the FSSM with the Kalman filter and maximum likelihood.

A *forward pass* through the data with the Kalman filter enables the calculation of the *stan-*

dardized one-step prediction errors (SPEs) of the FSSM. In the spirit of De Jong and Penzer, 1998; Commandeur and Koopman, 2007; and Commandeur, Koopman, and Ooms, 2011, these residuals can be used to check that the most important assumption for significance tests, i.e., the *assumption of independence* is satisfied. For the sake of efficiency, we choose to test for independence directly on the sub-models of the FSSM. To this regard, we employ the Box-Ljung test statistic to check the null hypothesis of independent residuals versus the alternative of serially correlated residuals. If the null hypothesis of independence is eventually rejected, the correction for serial correlation in the SPEs is done by increasing the lag order of the problematic variables only. The corrected sub-model is then re-estimated and the test repeated.

A *backward pass* through the data with the output of the Kalman filter and state and disturbance smoothing algorithms enables the calculation of the *standardized smoothed observation disturbances* (SSODs) and of the *standardized smoothed state disturbances* (SSSDs) of the FSSM. We use these two quantities to conduct t-tests for *outlier observations* and *structural breaks* in the FSSM. The presence of outlier observations and structural breaks is handled in the FSSM with the addition of shocks in the measurement equation. We call the FSSM with shocks in the measurement equation the *MShock-FSSM*.

7.4.2 State-Space Models with I(0) and I(1) Variables

A generic vector autoregressive model of the form

$$\boldsymbol{x}_t = \mathbf{A}_1 \boldsymbol{x}_{t-1} + \cdots + \mathbf{A}_p \boldsymbol{x}_{t-p} + \boldsymbol{\varepsilon}_t \tag{7.52}$$

can be cast into state-space form in the following way (Hamilton, 1994; Lütkepohl, 2005). The first step consists in defining the state vector $\boldsymbol{\xi}_t$, which is not unique. This implies that there are several representations of the same ARMA or VARMA models. The most favorable representation should be chosen depending on the application in mind. In (7.52) we can notice that $\boldsymbol{x}_t$ only depends on $\boldsymbol{x}_{t-1}, \ldots, \boldsymbol{x}_{t-p}$; thus, the state vector $\boldsymbol{\xi}_t$ can be taken as the stacked vectors $\boldsymbol{x}_t, \ldots, \boldsymbol{x}_{t-p+1}$. If there exists a constant in the time series model, a one needs to be included in the state vector. Since there is no constant in the process in (7.52), we define $\boldsymbol{\xi}_t$ as

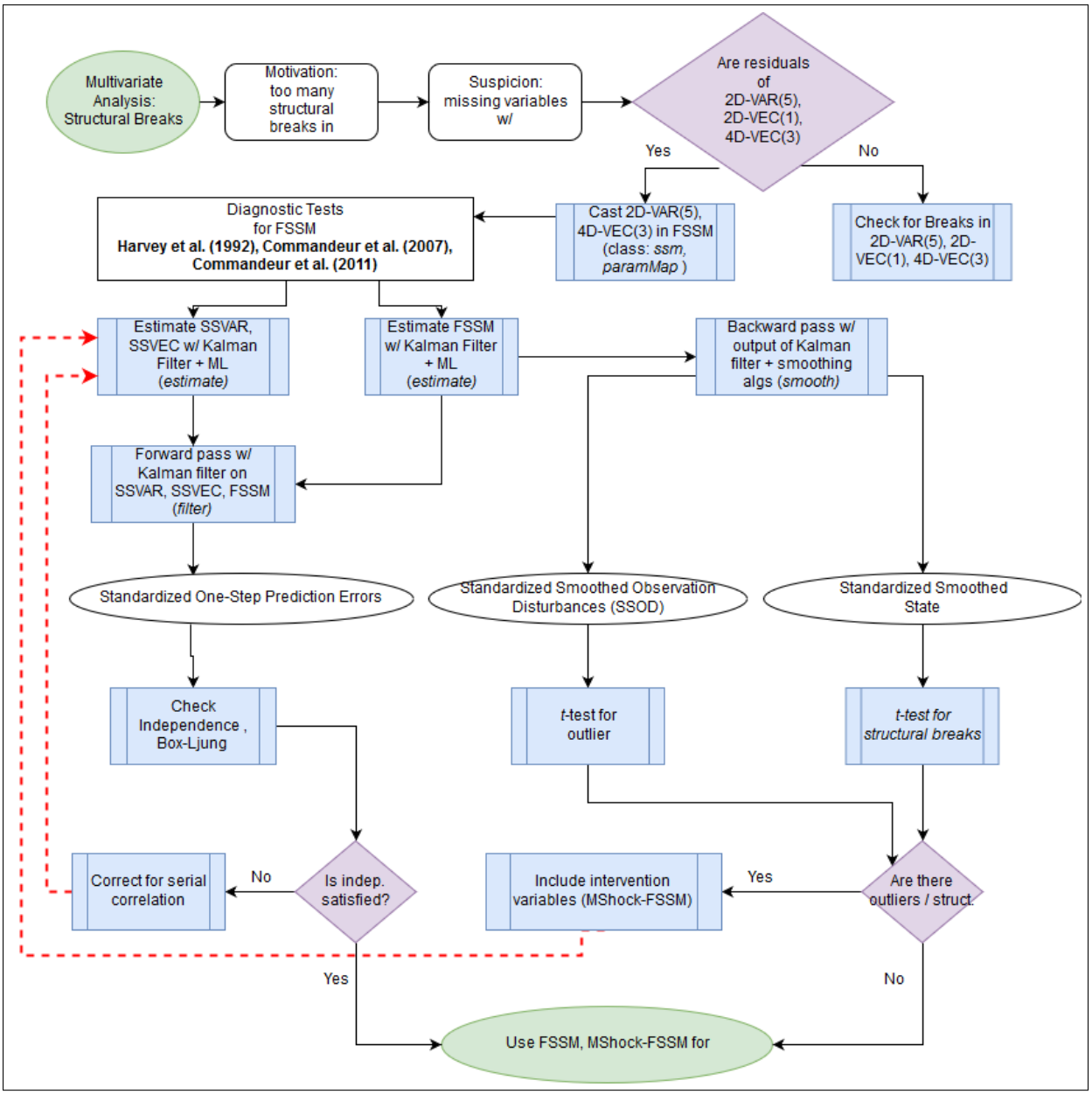
$$\boldsymbol{\xi}_t = [\boldsymbol{x}_t' \ldots \boldsymbol{x}_{t-p+1}']'. \tag{7.53}$$

To complete the transition equation in (7.1), the matrices $\mathbf{F}$ and $\mathbf{B}$ need to be defined. The matrix $\mathbf{F}$ is a square matrix, whose structure is driven by the dynamics of the time series model:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \boldsymbol{x}_{t-1} \\ \vdots \\ \boldsymbol{x}_{t-p+1} \end{bmatrix}}_{\boldsymbol{\xi}_t} = \underbrace{\begin{bmatrix} \mathbf{A}1_{n\times n} & \mathbf{A}2_{n\times n} & \cdots & \mathbf{A}p_{n\times n} \\ \mathbf{I}_{n\times n} & \mathbf{0}_{n\times n} & \cdots & \mathbf{0}_{n\times n} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{n\times n} & \mathbf{0}_{n\times n} & \mathbf{I}_{n\times n} & \mathbf{0}_{n\times n} \end{bmatrix}}_{\boldsymbol{F}\text{matrix}} \begin{bmatrix} \boldsymbol{x}_{t-1} \\ \boldsymbol{x}_{t-2} \\ \vdots \\ \boldsymbol{x}_{t-p} \end{bmatrix}. \tag{7.54}$$

The first row of matrix $\mathbf{F}$ is given by the time series model. The vector $\boldsymbol{x}_t$ is on the left-hand side and the lagged state vector, holding the lagged values of $\boldsymbol{x}_t$, is on the right-hand side. Thus,

Figure 7.1: Multivariate Analysis: Workflow



the first row of matrix **F** is filled in with the autoregressive coefficient matrices from $\mathbf{A}_1$ to $\mathbf{A}_p$. The following rows make the **F** matrix a square matrix of dimensions equal to the number of states. Matrix **B** has rows equal to the number of states, and columns equal to the number of state disturbances. This matrix consists of zeros everywhere except in the first column (corresponding to the noise term ε_t of the time series model), where it is an identity matrix:

$$[\boldsymbol{I}_{n\times n}\mathbf{0}_{n\times n}\cdots\mathbf{0}_{n\times n}]'. \tag{7.55}$$

In the next step, we need to complete the measurement equation in (7.2), by defining the matrices **H** and **D**. Matrix **H** has rows equal to the number of observations, and columns equal to the number of states. This matrix consists of zeros everywhere except in the first column (corresponding to the time series process), where it is an identity matrix:

$$[\boldsymbol{I}_{n\times n}\mathbf{0}_{n\times n}\cdots\mathbf{0}_{n\times n}]. \tag{7.56}$$

Matrix **D** has rows equal to the number of observations, and columns equal to the number of observations innovations. An empty **D** matrix indicates that there are no observation innovations in the state-space model.

The literature on state-space models in the context of cointegration is rather limited. Aoki and Havenner, 1991; Bauer and Wagner, 2002; and Aoki, 2013 contribute with their studies to the problems related to estimating state-space models using subspace algorithms. Wagner, 2010 presents a survey on cointegration analysis with state-space models by exemplifying the results obtained by Bauer and Wagner, 2002. The survey discusses cointegration analysis with state-space models considering both structure and statistical theory and concludes with a list of multiple open questions that require further research, in order to make cointegration analysis with state-space models an equivalent alternative to VAR cointegration analysis. The survey reveals that, with respect to structure theory, the inclusion of exogenous variables lacks adequate treatment. Many economic variables are characterized by both stochastic and deterministic trend components. Therefore, cointegration analysis with state-space models needs to allow for the inclusion of certain deterministic trend components. With respect to statistical aspects, the theory lacks the treatment of order estimation, parameter estimation, testing for the number of common trends and cycles. Furthermore, subspace algorithms are not yet available beyond the I(1) case, despite their great potential in reducing computational complexity.

The state-space equivalent to the celebrated VAR error correction model has been developed only recently by Ribarits and Hanzon, 2006 and Ribarits and Hanzon, 2014a. The two authors introduce the *state-space error correction model* (SSECM) and cover in great detail the open topics of model selection, parametrization, and maximum likelihood estimation of the SSECM. In a separate "companion paper", Ribarits and Hanzon, 2014b present simulation studies and applications of the SSECM. Ribarits and Hanzon, 2014a generalize VAR cointegration to cointegration analysis in state-space form by proceeding along the lines of Johansen's celebrated VAR error correction model. In the sequel, we recall the main steps in the derivation of SSECMs.

The starting point is to consider the following state-space model for $t = t_0, t_0+1, t_0+2, \ldots$:

$$x_{t+1} = Ax_t + B\varepsilon_t, \qquad x_{t_0} = x_0 \tag{7.57}$$
$$y_t = Cx_t + \varepsilon_t, \tag{7.58}$$

where x_t is the unobserved n-dimensional state vector; $A \in R^{n\times n}$, $B \in R^{n\times p}$ and $C \in R^{p\times n}$ are parameter matrices; y_t is the observed p-dimensional output. $(\varepsilon_t)_{t\in\mathbb{N}}$ is a p-dimensional Gaussian discrete-time white noise process with $\mathbb{E}\varepsilon_t = 0$ and $\mathbb{E}\varepsilon_s\varepsilon_t' = \delta_{s,t}\Sigma$ where $\delta_{s,t} = 1$ for $s = t$ and $\delta_{s,t} = 0$ otherwise and $'$ denotes the transpose.

In the next step, a set of assumptions are made concerning the system in (7.57, 7.58). The system is assumed to be minimal, in the sense that $rk[B, AB, \ldots, A^{n-1}B] = rk[C', A'C', \ldots, (A')^{n-1}C']' = n$. Second, matrix A is assumed to be stable, with eigenvalues either equal to 1 or lying within the open unit disc. Afterwards, the exposition is simplified by making the following formal definitions: $\varepsilon_t = 0$ and $y_t = 0$, for all t with $t < t_0$. Furthermore, z denotes the lag operator, so $zy_t = y_{t-1}$ and the following notation is used:

$$y_t = k(z)\varepsilon_t,$$

where

$$k(z) = \sum_{j=1}^{\infty} CA^{j-1}Bz^j + I = Cz(I - zA)^{-1}B + I$$
$$z \in C, z^{-1} \notin \sigma(A),$$

where $\sigma(A)$ denotes the spectrum of A; $k(z)$ is a $p \times p$ matrix of rational functions and is called the *transfer function* corresponding to the state-space model (7.57, 7.58).

At this point, let $\bar{k}(z) := k(z)^{-1} = z\bar{C}(I - z\bar{A})^{-1}\bar{B} + I$. $(\bar{A}, \bar{B}, \bar{C})$ can be computed with a simple inversion of the state-space model in (7.57, 7.58) as follows:

$$x_{t+1} = \overbrace{A - BC}^{\bar{A}} x_t + \overbrace{B}^{\bar{B}} y_t \tag{7.59}$$
$$\varepsilon_t(\bar{A}, \bar{B}, \bar{C}) = \underbrace{-C}_{\bar{C}} x_t + y_t. \tag{7.60}$$

Consequently,

$$\varepsilon_t = \bar{k}(z)y_t = \sum_{j=1}^{\infty} \bar{C}\bar{A}^{j-1}\bar{B}y_{t-j} + y_t. \tag{7.61}$$

The process in (7.61) can be rewritten in error correction form by defining $\tilde{k}(z) := I + (\bar{k}(z) - \bar{k}(1)z)/(z-1)$ and knowing that $\tilde{k}(0) = 0$. More specifically, the error correction form of (7.61) is given by:

$$\Delta y_t = \alpha\beta' y_{t-1} + \tilde{k}(z)\Delta y_t + \varepsilon_t \tag{7.62}$$

where $\Delta y_t = (1-z)y_t = y_t - y_{t-1}$. Finally, working out the state-space matrices for $\tilde{k}(z)$, the *state-space error correction* of (7.62) or its *state-space error correction model* (SSECM) is given by:

$$\begin{aligned} \Delta y_t &= \alpha\beta' y_{t-1} + \bar{C}\bar{A}(I-\bar{A})^{-1}\tilde{x}_t + \varepsilon_t \\ \tilde{x}_{t+1} &= \bar{A}\tilde{x}_t + \bar{B}\Delta y_t \\ \text{s.t.} -\alpha\beta' &= \bar{C}(I-\bar{A})^{-1}\bar{B} + I_p \qquad \text{(SSECM)} \end{aligned} \tag{7.63}$$

The idea behind the FSSM for the US and German yield curve drivers is a model that assembles together the state-space representations of the dynamics of the US and German levels, slopes, and curvatures. As such, the main peculiarity of the FSSM is that it mixes both I(0) and I(1) variables. The I(0) variables are the slopes, which follow a vector autoregressive process and the I(1) variables are the levels and curvatures, which follow a vector error correction process. To the best of our knowledge, the past literature on state-space methods lacks the treatment of state-space models mixing both I(0) and I(1) variables. We attempt to close this gap with the development of the *full state-space model* for IYCDs.

7.4.3 The Full State-Space Model (FSSM) for IYCDs

Our modeling idea is to check for structural breaks in one *full state-space model* (the FSSM) for the US and German yield curve drivers. The FSSM is composed of two sub-models: the *state-space VAR* (SSVAR) model for the slopes and the *state-space VEC* (SSVEC) model for the levels and curvatures. Given this structure, the FSSM has the nice feature of preserving the original dynamics found for the US and German yield curve drivers.

For convenience, before illustrating the development of the FSSM, we recall here the three processes of interest, namely:

1. the 2D-VAR(5) model for the $s_{US,t}/s_{DE,t}$ system:

$$\begin{bmatrix} s_{US,t} \\ s_{DE,t} \end{bmatrix} = \begin{bmatrix} \nu_1 \\ \nu_2 \end{bmatrix} + \sum_{i=1}^{5} \mathbf{B}_i \begin{bmatrix} s_{US,t-i} \\ s_{DE,t-i} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{7.64}$$

2. the 2D-VEC(1) model for the $l_{US,t}/l_{DE,t}$ system:

$$\begin{bmatrix} \Delta l_{US,t} \\ \Delta l_{DE,t} \end{bmatrix} = \boldsymbol{\alpha}\left(\boldsymbol{\beta}' \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \end{bmatrix} + \mathbf{c}_0\right) + \mathbf{B} \begin{bmatrix} \Delta l_{US,t-1} \\ \Delta l_{DE,t-1} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{7.65}$$

3. the 4D-VEC(3) model for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system:

$$\begin{bmatrix} \Delta l_{US,t} \\ \Delta l_{DE,t} \\ \Delta c_{US,t} \\ \Delta c_{DE,t} \end{bmatrix} = \boldsymbol{\alpha} \left(\boldsymbol{\beta}' \begin{bmatrix} l_{US,t-1} \\ l_{DE,t-1} \\ c_{US,t-1} \\ c_{DE,t-1} \end{bmatrix} + \mathbf{c}_0 \right) + \sum_{i=1}^{3} \mathbf{B}_i \begin{bmatrix} \Delta l_{US,t-i} \\ \Delta l_{DE,t-i} \\ \Delta c_{US,t-i} \\ \Delta c_{DE,t-i} \end{bmatrix} + \boldsymbol{\varepsilon}_t \tag{7.66}$$

The first step in the development of the FSSM is to notice that in (7.66) the 2D-VEC(1) model for the levels is nested in the 4D-VEC(3) model for the levels and curvatures. In other words, the 2D-VEC(1) model can be easily obtained from the 4D-VEC(3) model, by imposing the appropriate restrictions. To avoid issues related to overlapping variables, in the development of the FSSM, we choose to leave the 2D-VEC(1) model for the US and German levels aside. Such a choice is reasonable also because the estimated matrices $\boldsymbol{\alpha}\boldsymbol{\beta}'$ in the 2D-VEC(1) and the 4D-VEC(3) are quite similar, meaning that the structure of the impact matrix of the 2D-VEC(1) model is more or less preserved in the 4D-VEC(3), thus providing an additional justification to consider only the 4D-VEC(3) model in the FSSM.

In the next step, we derive the sub-models of the FSSM, i.e., we cast into state-space form the 2D-VAR(5) model for the US and German slopes, to obtain the SSVAR sub-model, and the 4D-VEC(3) model for the US and German levels and curvatures, to obtain the SSVEC sub-model.

The 2D-VAR(5) model in (7.64) can be cast into state-space form by choosing the following *state equation*:

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \end{bmatrix}}_{10\times 1} = \underbrace{\begin{bmatrix} \mathbf{A}1_{2\times 2} & \mathbf{A}2_{2\times 2} & \mathbf{A}3_{2\times 2} & \mathbf{A}4_{2\times 2} & \mathbf{A}5_{2\times 2} \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} \end{bmatrix}}_{10\times 10} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\xi}_{t-5} \end{bmatrix}}_{10\times 1} + \underbrace{\mathbf{B}}_{10\times 2} \boldsymbol{\epsilon}_t, \tag{7.67}$$

where $\boldsymbol{\xi}_t$ denotes a linear combination of current and lagged values of $s_{US,t}$ and $s_{DE,t}$.

Instead of modeling the constant in the VAR model as a separate state, we choose to centralize the slopes and, therefore, work with demeaned[14] data. The *measurement equation* of the SSVAR model is:

$$\underbrace{\boldsymbol{x}_t}_{2\times 1} = \underbrace{\begin{bmatrix} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \end{bmatrix}}_{2\times 10} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \end{bmatrix}}_{10\times 1} + \underbrace{\mathbf{D}}_{2\times 2} \boldsymbol{u}_t. \tag{7.68}$$

[14]Working with demeaned data turned out to be a requirement in order to fix issues with numerical optimization failing to converge when estimating the FSSM with the Kalman filter and maximum likelihood.

where $\boldsymbol{x}_t$ denotes the vector of measurements, i.e., the $s_{US,t}$ and the $s_{DE,t}$.
The process in (7.66) can be written more compactly as:

$$\Delta \boldsymbol{y}_t = \boldsymbol{\Pi} \boldsymbol{y}_{t-1} + \sum_{i=1}^{3} \boldsymbol{B}_i \Delta \boldsymbol{y}_{t-i} + \boldsymbol{\eta}_t, \tag{7.69}$$

where $\boldsymbol{\Pi} = \boldsymbol{\alpha}\boldsymbol{\beta}'$ denotes the impact matrix of the VEC model and the term $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ denotes the error-correction term. Along the lines of Hamilton, 1994; Ribarits and Hanzon, 2014a; and Ribarits and Hanzon, 2014b, we choose to treat the error-correction term $\boldsymbol{\Pi}\boldsymbol{y}_{t-1}$ as a *vector of exogenous variables* or a *regression component* in the measurement equation of the SSVEC model for the levels and curvatures. More specifically, the *state equation* of the SSVEC model is chosen to be:

$$\underbrace{\begin{bmatrix} \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{12\times 1} = \underbrace{\begin{bmatrix} \mathbf{B}1_{4\times 4} & \mathbf{B}2_{4\times 4} & \mathbf{B}3_{4\times 4} \\ \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \\ \mathbf{0}_{4\times 4} & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} \end{bmatrix}}_{12\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \\ \boldsymbol{\varphi}_{t-3} \end{bmatrix}}_{12\times 1} + \underbrace{\mathbf{B}}_{12\times 4} \boldsymbol{\varepsilon}_t \tag{7.70}$$

where $\boldsymbol{\varphi}_t$ denotes a linear combination of current and lagged values of $l_{US,t}, l_{DE,t}, c_{US,t}$, and $c_{DE,t}$, in first differences. The *measurement equation* is:

$$\underbrace{\Delta \boldsymbol{y}_t}_{4\times 1} = \underbrace{\begin{bmatrix} \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \end{bmatrix}}_{4\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{12\times 1} + \boldsymbol{\Pi}\boldsymbol{y}_{t-1} + \underbrace{\mathbf{D}}_{4\times 4} \boldsymbol{\eta}_t, \tag{7.71}$$

where $\Delta \boldsymbol{y}_t$ denotes the vector of measurements, i.e., the $l_{US,t}, l_{DE,t}, c_{US,t}$, and $c_{DE,t}$, in first differences.

Having defined the sub-models, the final step in the development of the FSSM consists in assembling together the SSVAR model for the slopes and the SSVEC model for the levels and curvatures into a unique state-space model, which we call the *full state-space model* (FSSM) for the international yield curve drivers. The *state equation* of the FSSM is:

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1} = \underbrace{\left[\begin{array}{cccc|ccc} \mathbf{A}1_{2\times 2} & \mathbf{A}2_{2\times 2} & \dots & \mathbf{A}5_{2\times 2} & & & \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & \mathbf{0} & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \hline & & & & \mathbf{B}1_{4\times 4} & \mathbf{B}2_{4\times 4} & \mathbf{B}3_{4\times 4} \\ & \mathbf{0} & & & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \\ & & & & \mathbf{0}_{4\times 4} & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\xi}_{t-5} \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \\ \boldsymbol{\varphi}_{t-3} \end{bmatrix}}_{22\times 1}$$

$$+ \underbrace{\mathbf{B}}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -- \\ \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{22\times 1}, \tag{7.72}$$

and the *measurement equation* is:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{6\times 1} = \underbrace{\left[\begin{array}{cccc|ccc} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \mathbf{0}_{4\times 2} & \dots & \mathbf{0}_{4\times 2} & \boldsymbol{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{6\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1}$$

$$+ \underbrace{\begin{bmatrix} \mathbf{0}_{2\times 4} \\ \boldsymbol{\Pi}_{4\times 4} \end{bmatrix}}_{6\times 4} \underbrace{\begin{bmatrix} \boldsymbol{y}_{t-1} \end{bmatrix}}_{4\times 1} + \underbrace{\mathbf{D}}_{6\times 6} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{6\times 1}. \tag{7.73}$$

Therefore, the state vector of the FSSM holds 22 states, which are the US and German slopes in levels plus 4 lagged states for each of the two slopes and the US and German levels and curvatures in first differences, plus 2 lagged states for each of the two levels and curvatures. The 22-dimensional state equation is collapsed to a 6-dimensional measurement equation.

7.4.4 MATLAB Implementation, Initialization, and Estimation of the FSSM

MATLAB SSM Functionality

We implement in MATLAB (Release R2014a) the FSSM defined in (8.2,8.3) using the SSM[15] functionality available in the Econometrics Toolbox™.

The MATLAB `ssm` class creates a linear, state-space model object with independent Gaussian state disturbances and observations innovations. The software supports specifications of time-invariant or time-varying models. The states can be specified to be either stationary, static, or nonstationary. A state-space model is created with the `ssm` function by providing the system matrices either *explicitly* or *implicitly*.

A model is created explicitly by directly specifying the parameters in the state transition, state disturbance loading, measurement sensitivity, and observation innovation matrices.

A model is created implicitly by providing a function that maps the input parameter vector to the matrices. This is the so-called *parameter-to-matrix mapping function*.

The explicit approach has the advantage of simplicity and the feature that each estimated parameter affects and is uniquely associated with a single element of a coefficient matrix. The implicit approach is more suitable for estimating complex models since the mapping function allows not only for the complete definition of the model but also for the imposition of various parameter constraints and the execution of additional steps, such as the deflation of observations before estimation, in order to account for regression components in the measurement equation.

Once a state-space model is specified using the `ssm` function, the unknown parameters can be estimated with the Kalman filter and maximum likelihood by passing the model and data to the `estimate`[16] function.

The estimated model (or the model without unknown parameters) can be passed to the `filter`[17] function, to implement forward recursion and obtain filtered states, to the `forecast`[18] function, to obtain forecasted states and observations, to the `smooth`[19] function, to implement backward recursion and obtain smoothed states, and to the `simulate`[20] function, to simulate states and observations from the state-space model.

The SSM framework of MATLAB supports regression of exogenous predictors, however, it does not store any non-zero offsets of state variables or any parameters associated with regression components in the measurement equation. This implies that all other related SSM functions assume that the measurements have already been deflated to account for any offsets or regression components. Therefore, after running the `filter`, `smooth`, `forecast`, and `simulate` function, the deflation must be unwinded by adding back the offsets or the regression components.

[15] MathWorks® Documentation: `ssm` class.
[16] MathWorks® Documentation: Maximum likelihood parameter estimation of state-space models.
[17] MathWorks® Documentation: Forward recursion of state-space models.
[18] MathWorks® Documentation: Forecast states and observations of state-space models.
[19] MathWorks® Documentation: Backward recursion of state-space models.
[20] MathWorks® Documentation: Monte Carlo simulation of state-space models.

MATLAB Implementation of the FSSM

Given its complexity and the presence of a regression component (coming from the error-correction term of the SSVEC sub-model), we choose to adopt an implicit approach for the MATLAB implementation of the FSSM in (8.2,8.3). More specifically, we implement a parameter-to-matrix mapping function that does the following. It maps the input parameter vector to the system matrices, specifies the initial state mean, the initial state covariance matrix, and the initial state distribution type, and finally, it deflates the measurements during the estimation, so that the estimation is performed on the deflated measurements.

Initialization of the Kalman Filter

We estimate the FSSM with the Kalman filter and maximum likelihood. In general, the maximum likelihood estimation (MLE) of state-space models via the Kalman filter is sensitive to the initial parameter values. To have the algorithm behaving well, a good initial guess of parameter values must be made. We choose to use the results of the estimated 2D-VAR(5) model for the slopes and the 4D-VEC(3) model for the levels and curvatures to initialize the estimation. More specifically, we initialize the *state transition coefficient matrix*, $\mathbf{F}$, with the 5 estimated 2-by-2 AR[21] coefficient matrices of 2D-VAR(5) for the $s_{US,t}/s_{DE,t}$ system and 3 short-run[22] matrices of 4D-VEC(3) for the $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system. The *measurement sensitivity coefficient matrix*, $\mathbf{H}$, has no parameters to estimate, as it consists of zeros except for columns corresponding to the measurement variables. For these columns, the $\mathbf{H}$ matrix is an identity matrix. The matrix $\mathbf{A}$ contains zero elements everywhere except for columns corresponding to the error-correction term of the SSVEC model, at these places, we initialize the $\mathbf{A}$ matrix with the estimated impact matrix $\mathbf{\Pi}$. For the parameterization of the $\mathbf{B}$ and $\mathbf{D}$ matrices, we assume that the disturbances $\boldsymbol{v}_t$ and $\boldsymbol{w}_t$ of the transition and measurement equations, respectively, are unit-variance white noise processes. Therefore, we assume that their covariance matrices, $\mathbf{Q}$ and $\mathbf{R}$, are identity matrices[23]. With respect to the *state disturbance loading coefficient matrix*, $\mathbf{B}$, we impose a diagonality constraint such that $\mathbf{Q} = \mathbf{B}\mathbf{B}'$ and initialize the $\mathbf{B}$ matrix with the identity matrix. Therefore, the $\mathbf{B}$ matrix has no parameters to estimate, as it consists of zeros except for columns corresponding to the noise term in the time series models. For these columns, the $\mathbf{B}$ matrix is an identity matrix. With respect to the *observation innovation coefficient matrix*, $\mathbf{D}$, we impose a non-diagonality constraint such that $\mathbf{R} = \mathbf{D}\mathbf{D}'$ and initialize the $\mathbf{D}$ matrix in terms of the Cholesky factorization of the sample covariance matrices of the residuals of VAR and VEC models. Therefore, the $\mathbf{D}$ matrix has below-diagonal elements, implying that shocks to the US and German slopes and to the US and German levels and curvatures are correlated. The *vector of initial state means* is initialized with the sample means of the states and the *matrix of initial state covariances* is initialized with the sample covariances of the states.

[21] Five autoregressive coefficient matrices in the 2D-VAR(5) model for the US and German slopes because of the optimal lag order of 5.

[22] Three short-run coefficient matrices in the 4D-VEC(3) model for the US and German levels and curvatures because of the optimal lag order of 3.

[23] In other words, we choose not to estimate the covariance matrix $\mathbf{R}$, instead, we estimate the observation innovation coefficient matrix, $\mathbf{D}$.

Parameter Restrictions

In order to avoid issues with model identification and to increase estimation precision, zero restrictions can be imposed on insignificant coefficients (i.e., coefficients with small t-ratios). The most adequate restricted model can be chosen based on likelihood ratio tests and Akaike or Bayesian information criteria. We choose not to impose any restrictions on the AR coefficients of the SSVAR sub-model for the following reasons. Firstly, we cannot assume diagonal autoregressive coefficient matrices since we found evidence of Granger-causality in both directions between the US and German slopes. Secondly, estimating the SSVAR model with parameter restrictions yields standardized one-step prediction errors that fail to satisfy the assumption of independence. Zero restrictions are imposed, instead, in the SSVEC sub-model. The FSSM has a total of 49 parameters[24] to be estimated.

Given the mix of I(0) and I(1) variables in the FSSM, in the spirit of Harvey and Koopman, 1992, we handle nonstationary components by means of a diffuse prior on the states.

Optimization Options

For the estimation of the FSSM with the Kalman filter and maximum likelihood, we choose to solve an unconstrained optimization problem[25] with quasi-Newton methods, for which we allow a maximum of `25000` function evaluations, maximum `1000` iterations, and a termination tolerance on the function value and on x of `1e-8`.

A first run of the FSSM estimation warns about numerical optimization failing to converge and imprecise computation of the covariance matrix of the estimates due to inversion difficulty.

With respect to convergence failure, in general, the presence of multiple local maxima induces complicated likelihood surfaces of state-space models. Very often the maximum likelihood estimation via the Kalman filter fails to converge or converges to an unsatisfactory solution. In this case, one solution might consists in refining the set of initial parameters with a refinement algorithm. The refined, initial parameter values returned by the refinement algorithm might appear similar to each other and/or to the crude set of initial parameter values. Conventional approaches suggest choosing the set yielding estimates that make economic sense and correspond to relatively large likelihood values. To aid the estimation of the state-space model, we decide to refine the initial parameters by several refinement algorithms[26] when fitting the state-space model to the response data using the crude set of initial parameter values. We then choose the output of the algorithm that yields the highest likelihood value.

With respect to inversion difficulty when computing the covariance matrix of the estimates, solutions include checking for parameter identifiability, trying different starting values, and trying different methods to compute the covariance matrix, i.e., *negative, inverted Hessian matrix*, *Outer Product of Gradients* (OPG), and *both Hessian and OPG*.

[24] This implies that there are on average $49/6 \approx 8$ parameters to be estimated for each of the 6 measurement equations.

[25] In MATLAB, this is the `SolverName` `fminunc` given to the `optimoptions` function. See MathWorks® Documentation: Find minimum of unconstrained multivariable function.

[26] I.e., *Loose bound interior point*, *Nelder-Mead algorithm*, *Quasi-Newton*, *Starting value perturbation*, and *Starting value shrinkage*.

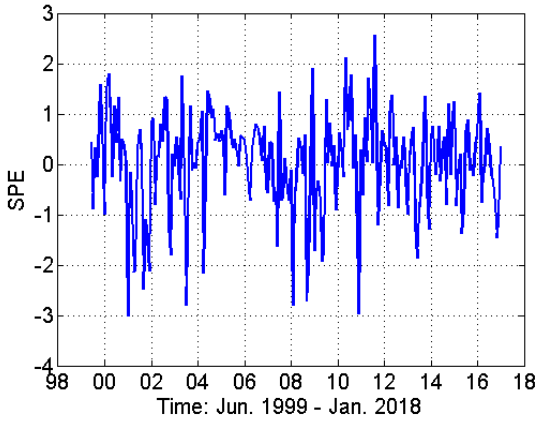
Working with Deflated Measurements

The presence of the *regression component* in the measurement equation implies working with deflated observations. The FSSM is therefore estimated using deflated observations and all subsequent algorithms (i.e., state and disturbance smoothing and forecasting algorithms) need to account for this peculiarity by unwinding the deflation effect (i.e., adding back the regression component to derive the inflated counterparts of the relevant quantities).

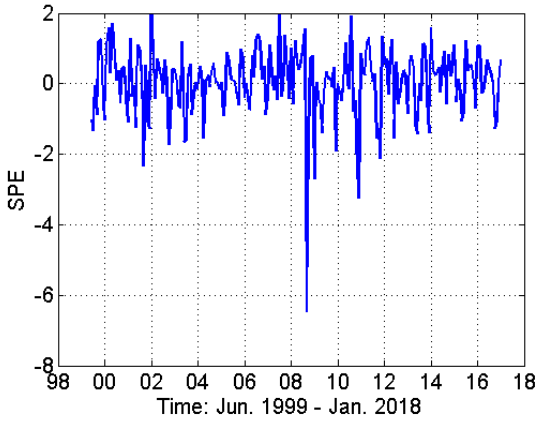
Forward Recursion of the FSSM

A forward pass through the data with the Kalman filter on the estimated SSVAR and SSVEC models enables the calculation of the Kalman filtered states, $\widehat{\boldsymbol{\xi}}_{t|t}$. The Kalman filtered states are useful in calculating the standardized one-step prediction errors (SPEs) of the sub-models of the FSSM. We test these residuals for independence using the Ljung-Box Q-test for serial autocorrelation. The test requires the definition of the number of lagged terms, which, as suggested by Box et al., 2015, can be set to $\min[20, T-1]$, where T is the sample size. All SPEs of the SSVAR and SSVEC are independent for the first 20 lags. As a crosscheck, we calculate also the Kalman filtered states of the FSSM and, from these, the SPEs of the FSSM (plotted in Figure 7.2). The Ljung-Box test performed on the SPEs of the FSSM confirms the results obtained for the sub-models.

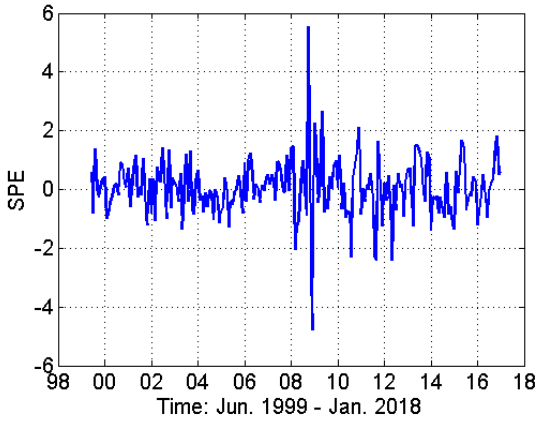
Figure 7.2: Standardized one-step prediction errors of the FSSM



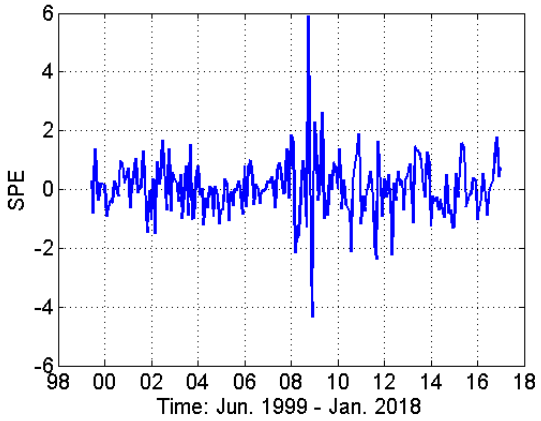
(a) $e_{s_{US},t}$



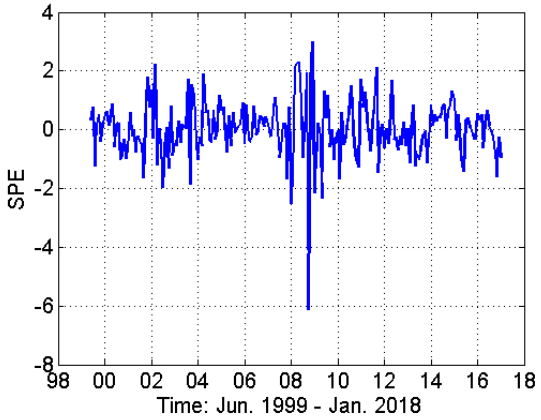
(b) $e_{s_{DE},t}$



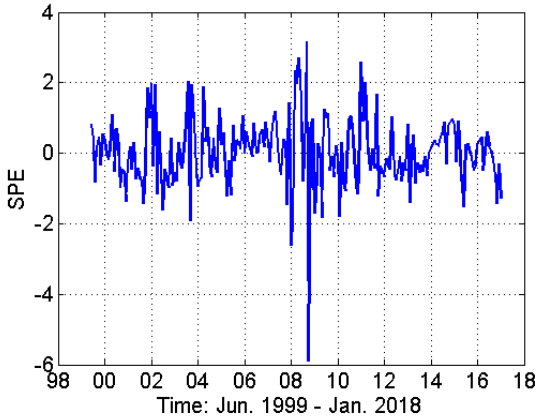
(c) $e_{\Delta l_{US},t}$



(d) $e_{\Delta l_{DE},t}$



(e) $e_{\Delta c_{US},t}$



(f) $e_{\Delta c_{DE},t}$

7.4.5 Structural Breaks in FSSM

Backward Recursion of the FSSM

A backward pass through the data with the output of the Kalman filter and state and disturbance smoothing algorithms enables the calculation of the smoothed state values, of the standardized smoothed observation disturbances, e_t^* (plotted in Figure 7.3), and of the standardized smoothed state disturbances, r_t^* (plotted in Figure 7.4).

Detecting Outliers and Structural Breaks Using the Auxiliary Residuals

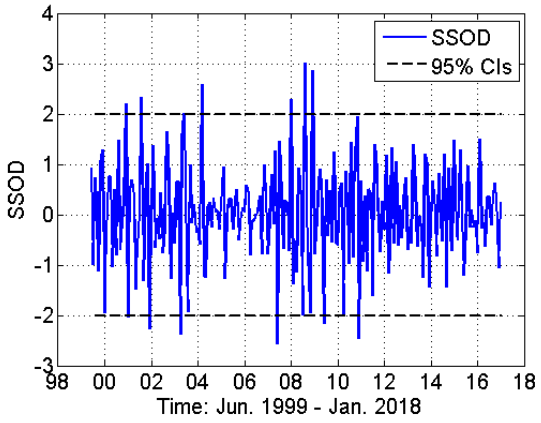
Along the lines of De Jong and Penzer, 1998; Commandeur and Koopman, 2007; and Commandeur, Koopman, and Ooms, 2011, we employ e_t^* and r_t^*, i.e., the auxiliary residuals of the FSSM to test for outlier observations and structural breaks in the FSSM. The assessment is done visually by plotting e_t^* and r_t^* with their 95% confidence intervals for a t-distribution. Values of e_t^* and r_t^* greater than 2 in absolute value provide indications of outlier observations and structural breaks, respectively.

Figure 7.3a plots $e^*_{s_{US},t}$, i.e., the standardized smoothed observations disturbances of the US slope. In this Figure, no values seem to exceed extremely the confidence limits. The case is different in Figures 7.3b, 7.3c, 7.3d, 7.3e, and 7.3f, plotting the standardized smoothed observations disturbances of the German slope in levels, of the first differences of US and German levels, and of the first differences of US and German curvatures. In these Figures, one can notice extreme values of opposite sign on either side of the series. The extreme values are registered predominantly in 2008:08-2008:09 in all 5 residuals, thus suggesting a sort of synchronicity of outliers across the 5 yield curve drivers.

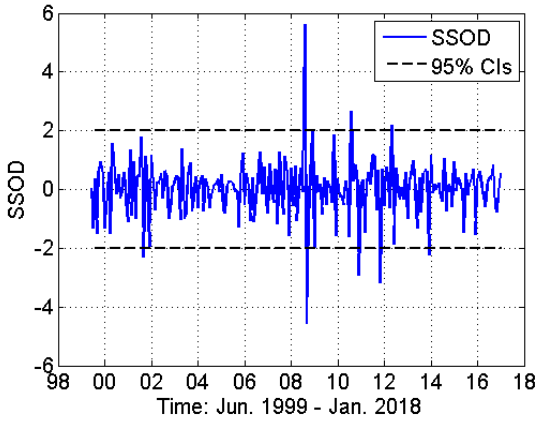
In Figure 7.4, we assess the presence of structural breaks in the FSSM. Extreme values occur mostly in the r_t^* of the German slope in levels (Figure 7.4b), in the r_t^* of the first differences of the US level (Figure 7.4c), and in the r_t^* of the first differences of the German curvature (Figure 7.4f). Similarly to the plots of the standardized smoothed observation disturbances, also in the standardized smoothed state disturbances the extreme values occur predominantly in 2008:08-2008:09, thus suggesting a sort of synchronicity of structural breaks across the US and German yield curve drivers.

Commandeur and Koopman, 2007 advise accounting for outlier observations with the insertion of a *pulse intervention variable*, consisting of ones at the time points corresponding to the outlier observations, and zeroes elsewhere. Structural breaks, on the other hand, can be handled with the insertion of *shift intervention variables*. However, Commandeur et al. explain that care should be taken not to indiscriminately add pulse and/or shift intervention variables for each and every outlier and structural break detected in the auxiliary residuals. The risk is that the improved fit of the model (resulted from the addition of pulse intervention variables) might provide a false sense of confidence in the forecasts. The insertion of an intervention variable to account for an observed structural break in the auxiliary residuals should be justified by a theory concerning the possible cause of the structural break.

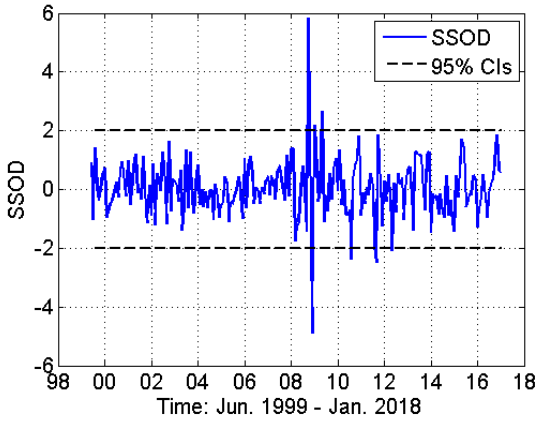
Figure 7.3: Detecting outlier observations in the FSSM: Plot of standardized smoothed observation disturbances and their 95% confidence intervals for a t-distribution.



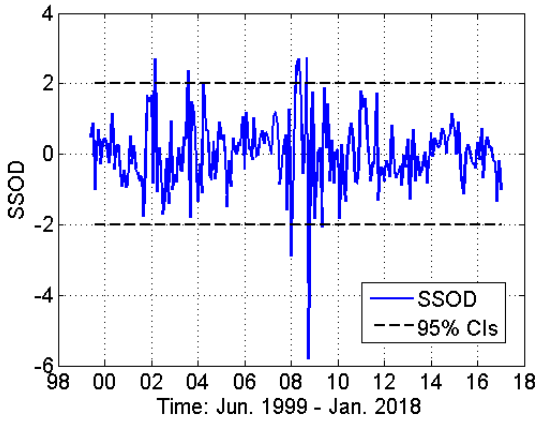
(a) $e^*_{s_{US},t}$ and 95% CIs



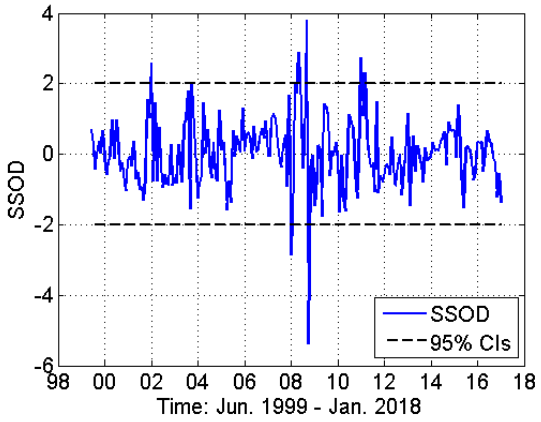
(b) $e^*_{s_{DE},t}$ and 95% CIs



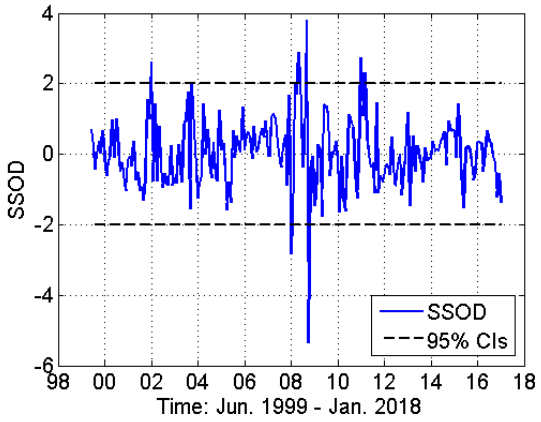
(c) $e^*_{\Delta l_{US},t}$ and 95% CIs



(d) $e^*_{\Delta l_{DE},t}$ and 95% CIs

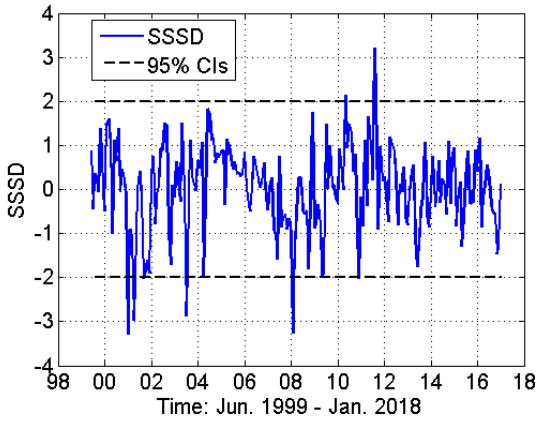


(e) $e^*_{\Delta c_{US},t}$ and 95% CIs

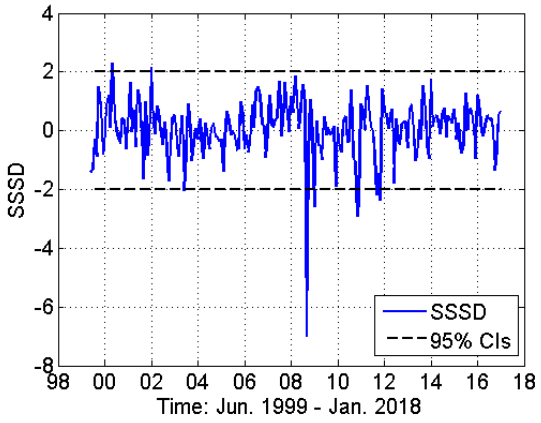


(f) $e^*_{\Delta c_{DE},t}$ and 95% CIs

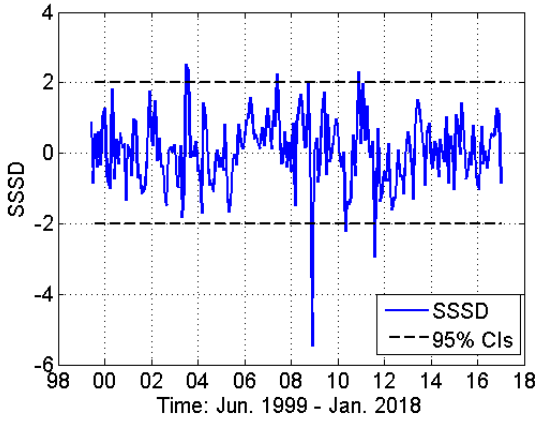
Figure 7.4: Detecting structural breaks in the FSSM: Plot of standardized smoothed state disturbances and their 95% confidence intervals for a t-distribution.



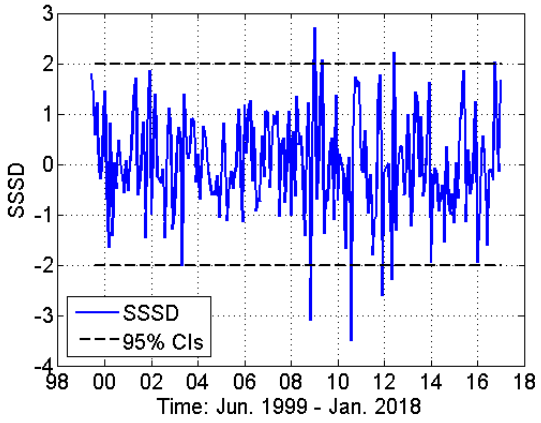
(a) $r^*_{s_{US},t}$ and 95% CIs



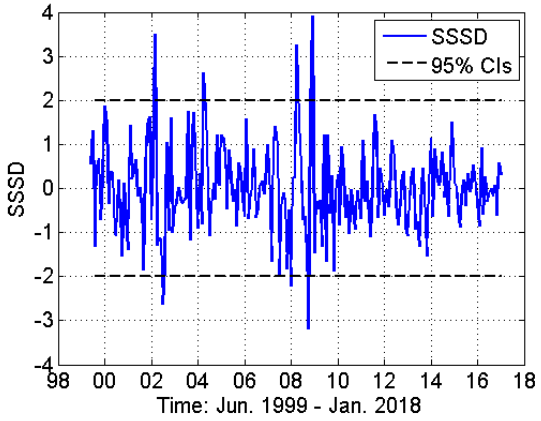
(b) $r^*_{s_{DE},t}$ and 95% CIs



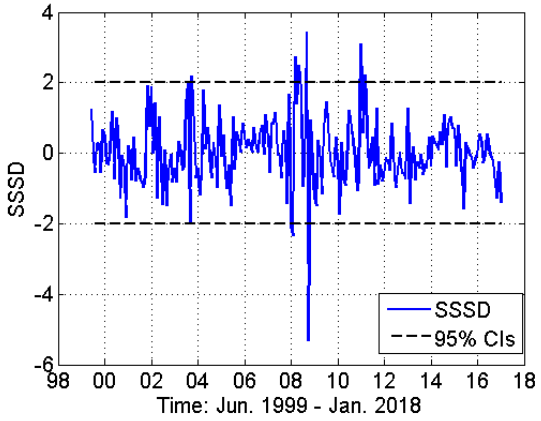
(c) $r^*_{\Delta l_{US},t}$ and 95% CIs



(d) $r^*_{\Delta l_{DE},t}$ and 95% CIs



(e) $r^*_{\Delta c_{US},t}$ and 95% CIs



(f) $r^*_{\Delta c_{DE},t}$ and 95% CIs

7.4.6 Adjusting for Outliers and Structural Breaks in the FSSM: the MShock-FSSM

In Figures 7.3 and 7.4, we noticed extreme values on either side of the current level of the series. According to De Jong and Penzer, 1998, such alterations in structure should be handled with switch intervention variables. Since the unusual points in Figures 7.3 and 7.4 appear more as patches of outliers rather than as persistent alterations or structural breaks, in the spirit of Penzer, 2007, we decide to account for their presence in the FSSM with the inclusion of shocks in the measurement equation.

Therefore, in this section we develop the *MShock-FSSM* model (where M stands for "measurement"), whose *state equation* is equal to that of the FSSM:

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1} = \underbrace{\left[\begin{array}{cccc|ccc} \mathbf{A}1_{2\times 2} & \mathbf{A}2_{2\times 2} & \dots & \mathbf{A}5_{2\times 2} & & & \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & \mathbf{0} & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \hline & & & & \mathbf{B}1_{4\times 4} & \mathbf{B}2_{4\times 4} & \mathbf{B}3_{4\times 4} \\ & \mathbf{0} & & & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \\ & & & & \mathbf{0}_{4\times 4} & \mathbf{I}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\xi}_{t-5} \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \\ \boldsymbol{\varphi}_{t-3} \end{bmatrix}}_{22\times 1}$$
$$+ \underbrace{\mathbf{B}}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \hline \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{22\times 1}. \tag{7.74}$$

The intervention variables are incorporated in the regression component of the measurement equation in such a way as to shock separately each component of the measurement vector. More

specifically, the *measurement equation* of the MShock-FSSM is defined as:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{6\times 1} = \underbrace{\left[\begin{array}{ccccc|ccc} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \mathbf{0}_{4\times 2} & \dots & \mathbf{0}_{4\times 2} & & \boldsymbol{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{6\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1}$$
$$+ \underbrace{\begin{bmatrix} \boldsymbol{\delta}^{\text{VAR}}_{2\times 2} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \boldsymbol{\delta}^{\text{VEC}}_{4\times 4} & \boldsymbol{\Pi}_{4\times 4} \end{bmatrix}}_{6\times 10} \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}^{\text{VAR}}_t \\ \boldsymbol{\Lambda}^{\text{VEC}}_t \\ \boldsymbol{y}_{t-1} \end{bmatrix}}_{10\times 1} + \underbrace{\mathbf{D}}_{6\times 6} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{6\times 1}, \tag{7.75}$$

where $\boldsymbol{\Lambda}^{\text{VAR}}_t$ and $\boldsymbol{\Lambda}^{\text{VEC}}_t$ are the shock variables, defined as:

$$\boldsymbol{\Lambda}^{\text{VAR}}_{t,s_{US}} = 0 \tag{7.76}$$

$$\boldsymbol{\Lambda}^{\text{VAR}}_{t,s_{DE}} = \begin{cases} 1, & t = \text{'2008:08'} \\ -1, & t = \text{'2008:09'} \\ 0, & \text{otherwise} \end{cases} \tag{7.77}$$

$$\boldsymbol{\Lambda}^{\text{VEC}}_{t,l_{US}} = \begin{cases} 1, & t = \text{'2008:10'} \\ -1, & t = \text{'2008:12'} \\ 0, & \text{otherwise} \end{cases} \tag{7.78}$$

$$\boldsymbol{\Lambda}^{\text{VEC}}_{t,l_{DE}} = \begin{cases} 1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{7.79}$$

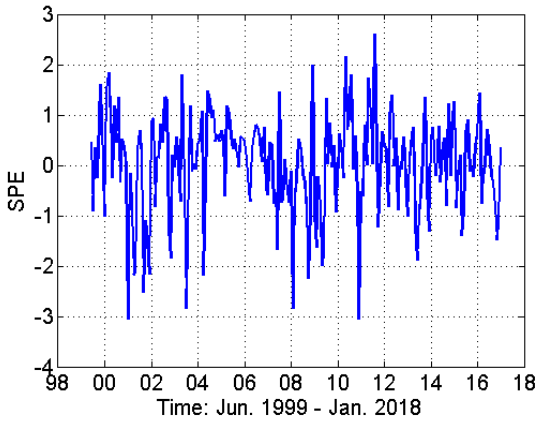
$$\boldsymbol{\Lambda}^{\text{VEC}}_{t,c_{US}} = \begin{cases} -1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{7.80}$$

$$\boldsymbol{\Lambda}^{\text{VEC}}_{t,c_{DE}} = \begin{cases} -1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{7.81}$$

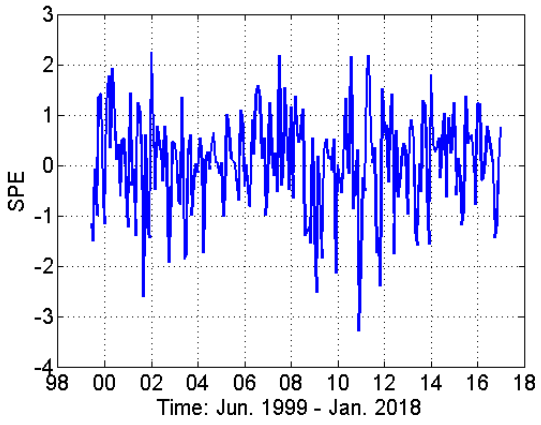
The shock variables are equal to 1 or -1 at date points corresponding to outlying measurements, and 0 at all other date points. In order not to undermine the fit of the model, we account only for the most blatant outliers. The quantities $\boldsymbol{\delta}^{\text{VAR}}_{2\times 2}$ and $\boldsymbol{\delta}^{\text{VEC}}_{4\times 4}$ are the magnitudes of the shocks and are determined by estimation. For the initialization of the Kalman filter for the estimation of the MShock-FSSM, both $\boldsymbol{\delta}^{\text{VAR}}_{2\times 2}$ and $\boldsymbol{\delta}^{\text{VEC}}_{4\times 4}$ are initialized with the identity matrix.

Using the full sample of observations, i.e., [1999:01-2018:01], a forward pass through the data with the Kalman filter on the estimated MShock-FSSM yields the Kalman filtered states, which we use to calculate the standardized one-step prediction errors, plotted in Figure 7.5. Ljung-Box tests on these latter residuals support the assumption of independence.

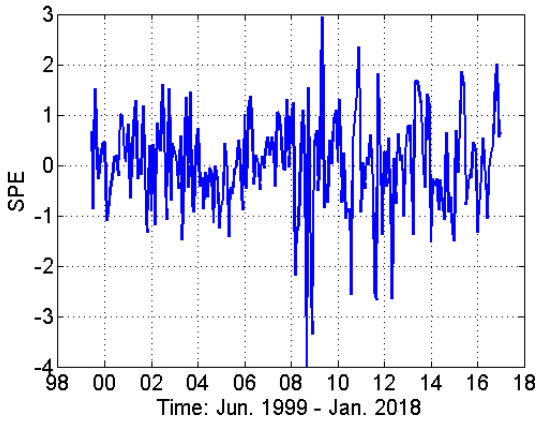
Figure 7.5: Standardized one-step prediction errors of the MShock-FSSM.



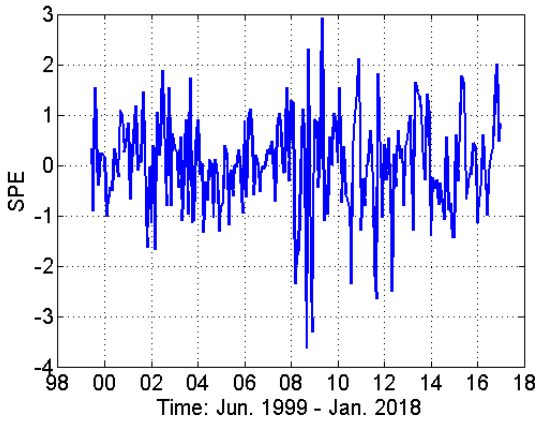
(a) $e_{s_{US},t}$



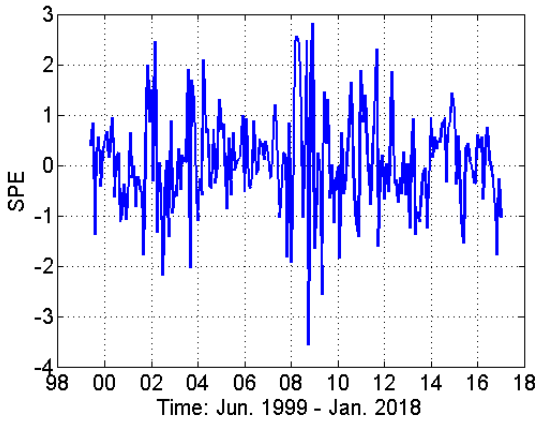
(b) $e_{s_{DE},t}$



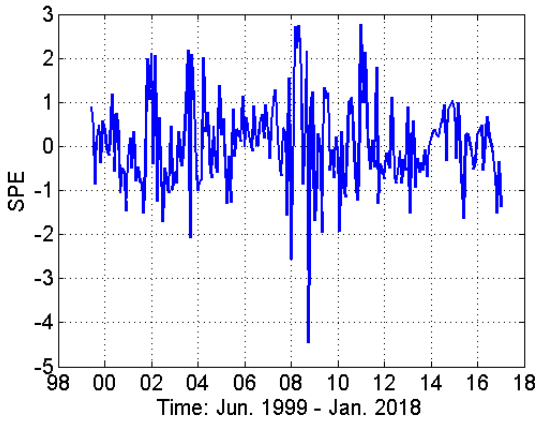
(c) $e_{s_{US},t}$



(d) $e_{s_{DE},t}$



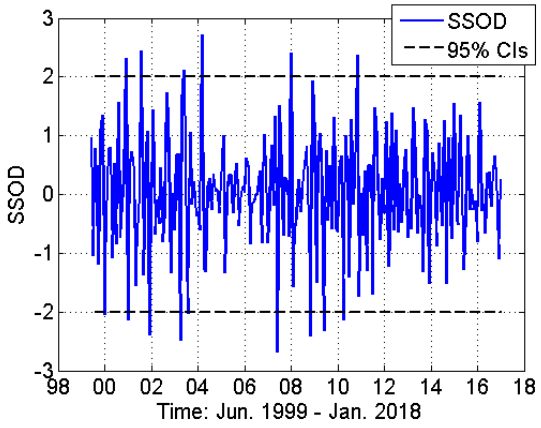
(e) $e_{s_{US},t}$



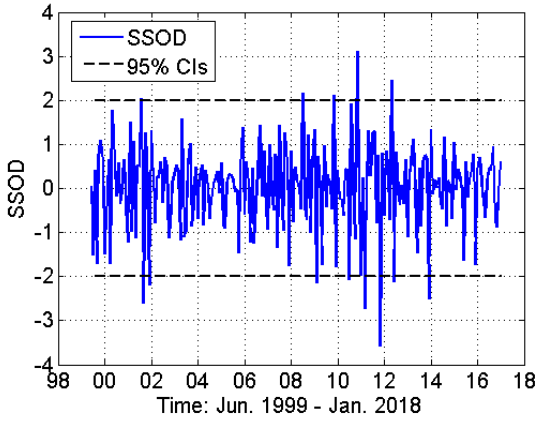
(f) $e_{s_{DE},t}$

With a backward pass through the data with the output of the Kalman filter and state and disturbance smoothing algorithms we calculate and plot the standardized smoothed observations disturbances (Figure 7.6) and the standardized smoothed state disturbances (Figure 7.7), respectively, of the MShock-FSSM. The extreme values observed in the e_t^* and in the r_t^* of the FSSM are now significantly reduced, confirming the theory behind the outlying values. The theory is that the outliers were caused by the Financial Crisis of 2007-2008 and by a regime change in the Fed and ECB monetary policy. In Figures 7.6 and 7.7 some residual values still exceed the confidence limits. However, since the auxiliary residuals are plotted with the 95% confidence intervals, it is expected that 5% of the residuals still break the confidence limits.

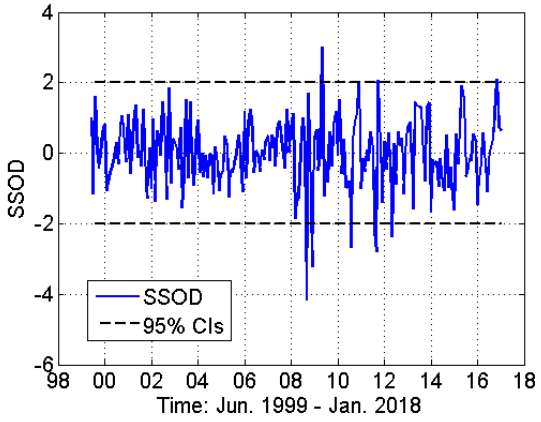
Figure 7.6: Detecting outlier observations in the MShock-FSSM: Plot of standardized smoothed observation disturbances and their 95% confidence intervals for a t-distribution.



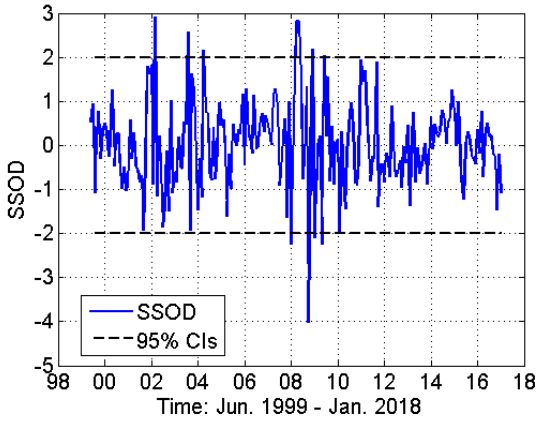
(a) $e^*_{s_{US},t}$ and 95% CIs



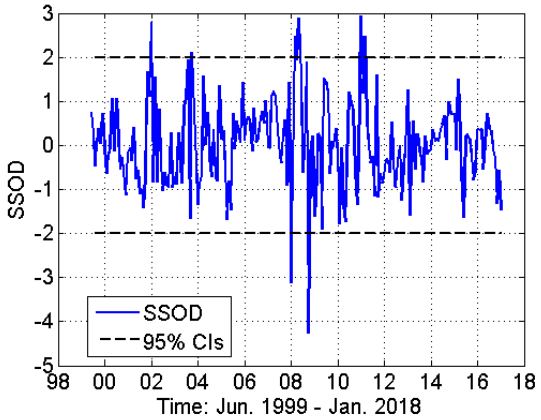
(b) $e^*_{s_{DE},t}$ and 95% CIs



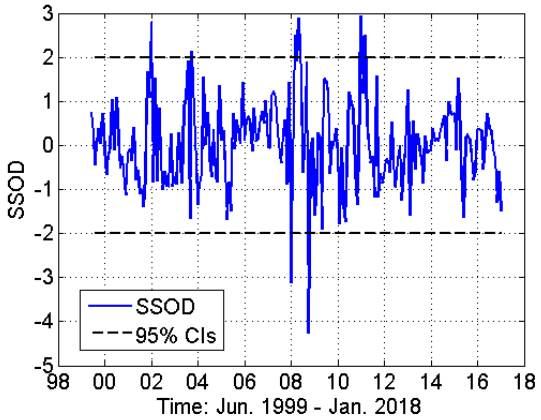
(c) $e^*_{\Delta l_{US},t}$ and 95% CIs



(d) $e^*_{\Delta l_{DE},t}$ and 95% CIs

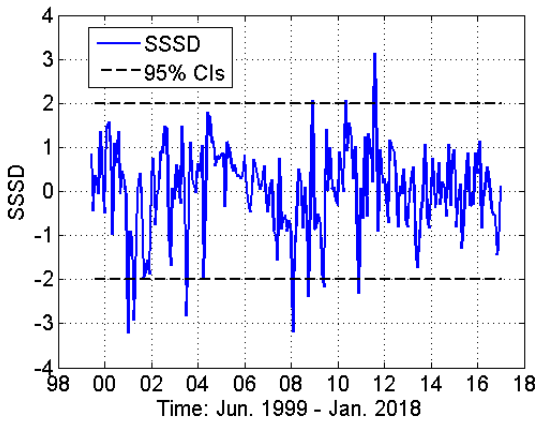


(e) $e^*_{\Delta c_{US},t}$ and 95% CIs

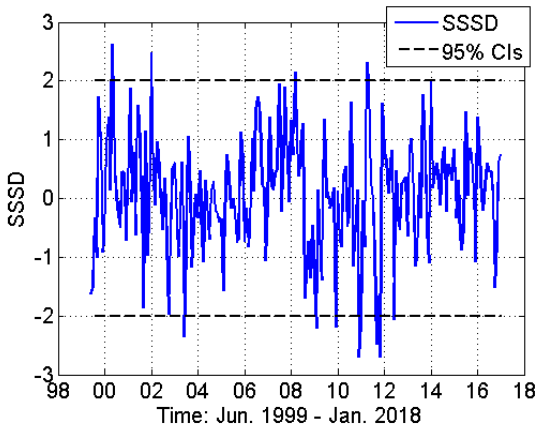


(f) $e^*_{\Delta c_{DE},t}$ and 95% CIs

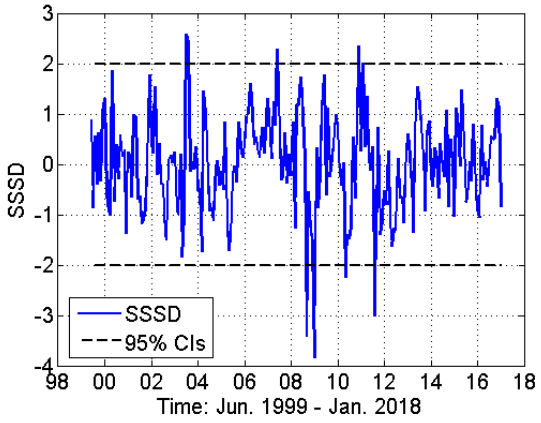
Figure 7.7: Detecting structural breaks in the MShock-FSSM: Plot of standardized smoothed state disturbances and their 95% confidence intervals for a t-distribution.



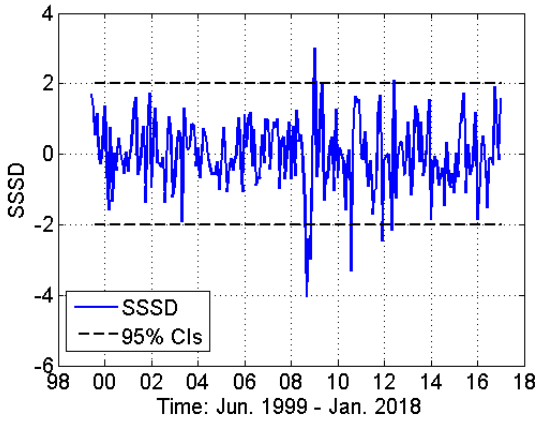
(a) $r^*_{s_{US},t}$ and 95% CIs



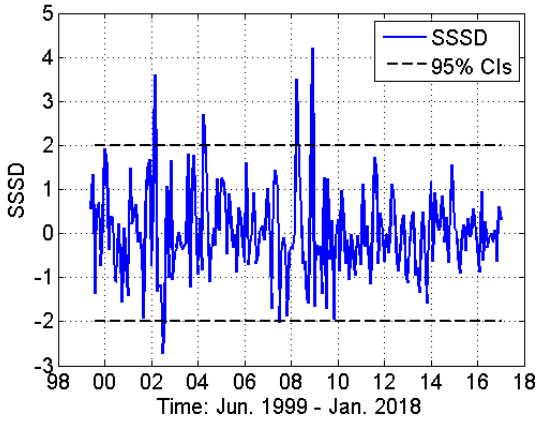
(b) $r^*_{s_{DE},t}$ and 95% CIs



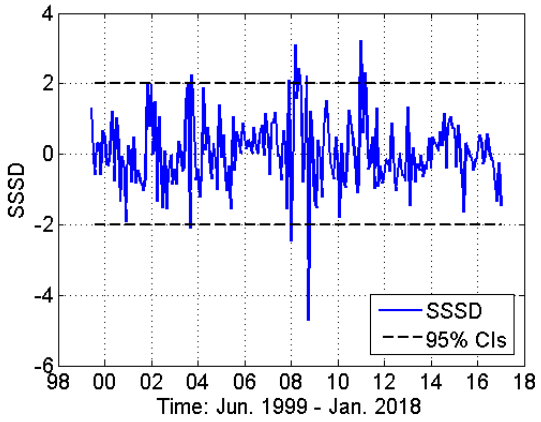
(c) $r^*_{\Delta l_{US},t}$ and 95% CIs



(d) $r^*_{\Delta l_{DE},t}$ and 95% CIs



(e) $r^*_{\Delta c_{US},t}$ and 95% CIs



(f) $r^*_{\Delta c_{DE},t}$ and 95% CIs

7.5 Conclusion

In the present Chapter, we adopted a multivariate state-space framework to test for the presence of outliers and structural breaks in the US and German yield curve drivers. The motivation stemmed from the results of the univariate analysis, which revealed the presence of multiple structural breaks in all drivers.

In an attempt to derive more accurate results, we developed a new state-space model, the FSSM, for the US and German yield curve drivers. The model assembles together the VAR dynamics of the slopes and the VEC dynamics of the levels and curvatures. As such, differently from the Diebold-Li model, the FSSM has the novelty of being designed to preserve the dynamic properties of the yield curve drivers. Another novelty of the FSSM is that it models in one system both I(0) and I(1) variables. The I(0) variables being the slopes and the I(1) variables being the levels and curvatures. From a modeling perspective, the FSSM handles the error-correction term, $\boldsymbol{\alpha}\boldsymbol{\beta}'\boldsymbol{y}_{t-1}$, of the VEC model as a regression component in the measurement equation. The advantage of this modeling choice is not to increase the dimensionality of the state vector, by modeling the regression component as an additional state. The disadvantage is that the regression component in the measurement equation must deflate the observations before estimating the model. Therefore, all state-space modeling and forecasting algorithms, i.e., the Kalman filter and maximum likelihood estimation, forward and backward recursions, and forecasting algorithms work with deflated observations.

We estimated the FSSM model with the Kalman filter and maximum likelihood. To this regard, we discussed the initialization of the Kalman filter, the restrictions imposed on the parameters of the FSSM, and the optimization options chosen for the estimation of the model. For the sake of efficiency, we explained how to work with the SSVAR and SSVEC sub-models, on which we performed a forward recursion through the data with the Kalman filter to calculate the standardized one-step prediction errors. Using these latter quantities, we confirmed that the assumed property of independent errors is satisfied for the FSSM.

A backward pass through the data with the output of the Kalman filter and state and disturbance smoothing algorithms allowed us to calculate the auxiliary residuals of the FSSM, i.e., the standardized smoothed observation disturbances (SSODs) and the standardized smoothed state disturbances (SSSDs). We used these two quantities to detect outliers and structural breaks in the FSSM. We concluded that the unusual points in the plots of SSODs and SSSDs resemble more of patches of outliers rather than of structural breaks. Therefore, we adjusted the FSSM for the presence of most blatant outliers by including intervention variables in the measurement equation. The adjusted model is the MShock-FSSM model.

In conclusion, the results of the univariate and multivariate state-space analysis on our set of yield data provide evidence supporting the existence of alterations in the structure of the yield curve drivers. This evidence undermines the AR(1) dynamics assumed for the yield curve drivers in Diebold and Li, 2006; Diebold, Rudebusch, and Aruoba, 2006; and Diebold, Li, and Yue, 2008.

Part III

Forecasting

Forecasting

The FSSM and the MShock-FSSM for the US and German yield curve drivers are designed to preserve the dynamic properties of the drivers embodied in their individual VAR and VEC dynamics. In addition, in the spirit of Diebold and Li, 2006, the two models allow forecasting the US and German yield curves by forecasting their drivers.

In Part III of this thesis, we explore the performance of the FSSM and MShock-FSSM in out-of-sample yield curve forecasting. To this regard, we perform a recursive out-of-sample forecasting exercise with re-estimation of the parameters every 12 months with the Kalman filter and maximum likelihood and produce term-structure forecasts at both short and long horizons, for US and Germany. We compare the forecasting performance of our models to the benchmark Diebold-Li "Yields-Only" Model. The aim is to understand how do models that account for all dynamic properties of yield data perform compared to the state-of-the-art Diebold-Li model.

In addition, we verify the predictive power of the curvatures by developing and forecasting with two additional models: the FSSM^{LS} and the MShock-FSSM^{LS}, which are the original FSSM and the MShock-FSSM for the US and German levels and slopes only.

The forecasting results of the models for all drivers are promising, with multiple cases where the FSSM and MShock-FSSM outperform the benchmark, whereas the forecasting results of the models without the curvatures are rather poor, supporting the idea that the curvatures do contribute to the forecasting accuracy of the US and German yield curves.

Chapter 8

Forecasting Jointly the US and German Yield Curves

8.1 Introduction

The yield curve – i.e., the price of money today, tomorrow, and many years from now[1] – provides answers to questions such as, should we invest in a money market fund or a bond fund? Which bonds are "rich" and which ones are "cheap"? What is the trade-off between risk and return for fixed income investments? What macroeconomic risks are accounted for in bond returns? How can expectations of short-term interest rates and term premium be estimated? How can the real-interest rate, inflation expectations, and the inflation risk premium be estimated? How are changes in the policy rate channeled to other interest rates? How do we get out of a financial crisis? Why do interest rates influence economic activity? (Campbell, 1995).

The yield curve gains special importance in today's historical market circumstances, where the yields have been following a decreasing trend for more than three decades. Since such a trend is causing a potential inversion of the business cycle, thus a higher market risk, many investors in search for yield are looking outside traditional asset classes, which are attractive for their good performance and diversification benefits (Becker and Ivashina, 2015 and Kräussl, Lehnert, and Rinne, 2017). Furthermore, the central banks' non traditional monetary policy undermines the prediction ability of conventional methods for policy making and asset management. Clearly, modeling and forecasting the future movement of yield curves is paramount for many tasks, including financial assets and derivative security pricing, portfolio allocation, financial risk management, fiscal debt structuring, and monetary policy decisions (Diebold and Rudebusch, 2013). Although there are many studies on yield curve modeling (see Chapter 2), the literature on yield curve forecasting remains limited. The arbitrage-free (Hull and White, 1990; Heath, Jarrow, and Morton, 1992) and affine models (Vasicek, 1977; Cox, Ingersoll Jr, and Ross, 1977; Duffie and Kan, 1996), focusing primarily on the in-sample fit, are known to perform poorly out-of-sample (Duffee, 2002). When the goal is to forecast the yield curve out-of-sample, the domestic term structure factor models (Nelson and Siegel, 1987; Litzenberger, Squassi, and Weir, 1995; Balduzzi

[1] A 3-D View of a Chart That Predicts The Economic Future: The yield curve, *The New York Times*.

et al., 1996; Chen, 1996; Bliss, 1997a; Bliss, 1997b; Andersen and Lund, 1997; Dai and Singleton, 2000; De Jong and Santa-Clara, 1999; Jong, 2000; Brandt and Yaron, 2003; Duffee, 2002) and global (Diebold, Li, and Yue, 2008; Jotikasthira, Le, and Lundblad, 2015) are very often preferred. Of this group of models, the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model is state-of-the-art, given its nice features such as parsimony, accurate parameter estimation, and strong forecastability at long horizons.

Promising results come from machine learning applications. Machine learning algorithms are extensively applied in equity (Agrawal, Chourasia, and Mittra, 2013; Ballings et al., 2015; Booth, Gerding, and Mcgroarty, 2014; Dunis et al., 2016; Eilers et al., 2014; and Vui et al., 2013) and foreign exchange markets (Choudhry et al., 2012; Fletcher, 2012; Gradojevic and Yang, 2006; and Wang and Huang, 2005), although their application remains scarcely explored in the fixed income markets. The first comprehensive study using artificial neural networks in the context of yield curve forecasting has been provided only recently by Nunes et al., 2018. The study employs a multivariate linear regression and multilayer perceptron (MLP) to forecast the European yield curve. Considering forecasting horizons from next day to 20 days ahead, the best results are obtained with the MLP and the addition of synthetic data improves the accuracy.

With respect to the forecasting power of the yield curve itself, the yield curve is well-known to be a great predictor of future economic activity (Estrella and Hardouvelis, 1991; Haubrich and Dombrosky, 1996; Estrella and Mishkin, 1996; Dueker, 1997; Chauvet and Potter, 2005; Wright, 2006; Estrella and Trubin, 2006; Ang, Piazzesi, and Wei, 2006; Rudebusch and Williams, 2009; Zaloom, 2009) and future levels of inflation (Frankel and Lown, 1994; Kozicki, 1997; Evans and Marshall, 1998; Chopin and Pelgrin, 2004; Estrella, 2005; Gürkaynak, Sack, and Wright, 2010; and Joyce, Lildholdt, and Sorensen, 2010). The future economic activity and the future levels of inflation directly influence the price of everything from equities and real estate to household items[2].

In today's global capital markets, bond yields of different countries interact in a dynamic fashion, giving rise to contemporaneous and non-contemporaneous dependency patterns. Central banks, international fixed income investors, and risk managers, all have a vital interest in modeling and forecasting the yield curve in an international setting. To this regard, the literature is again very limited. The main contribution is the Dynamic Nelson-Siegel model of Diebold and Li, 2006, extended to the global context by Diebold, Li, and Yue, 2008. The idea behind the model is a hierarchy for global yields, in the sense that, country yield curves depend on country factors, which in turn depend on global factors. An empirical application of the model to the term structure of government bond yields for Germany, Japan, the UK, and the US, finds evidence supporting the existence, economic importance, and explanatory power of global yield factors.

In addition to these gaps in the literature on yield curve forecasting, it is worth noting that the models mentioned above rely on very restrictive assumptions concerning the dynamic properties of the yield data (see Section 4.1). In Chapter 4, using US and German yield data, we found evidence about the stationarity of the US and German slopes, nonstationarity of the levels and curvatures, cointegration structure between the levels and curvatures, and existence

[2]See the transmission mechanism of monetary policy (ECB).

of Granger causality among all US and German yield curve drivers. Conventional modeling approaches in the literature assume stationarity of yields and refrain from exploring the dynamic properties from a forecasting perspective.

In this Chapter, we employ our newly developed state-space models (i.e., the FSSM and the MShock-FSSM) to forecast the co-movement of the US and German yield curve drivers and, from the drivers, the US and German yield curves. We recall that our state-space models are designed to preserve the dynamic properties of the US and German yield curve drivers embodied in their underlying processes. Therefore, this Chapter contributes to the existing literature by providing forecasting results of our new data-driven state-space models designed to model the co-movement of the yield curves of different world regions, while preserving their dynamic properties. In addition, we verify the predictive power of the yield curve curvatures, by developing and forecasting with two other models for the US and German yield curve levels and slopes only. These are the FSSM^{LS} and the MShock-FSSM^{LS} models. The ultimate aim is to understand how do models that account for all dynamic properties of the yield data perform compared to the state-of-the-art Diebold-Li model.

The remainder of the Chapter is organized as follows. Section 8.2 explains how we employ the Diebold-Li "Yields-Only" Model in its state-space form, estimate it with the Kalman filter and maximum likelihood, and produce forecasts of the US and German yield curves for each maturity in our sample. Section 8.3 recalls for convenience the FSSM and the MShock-FSSM, from which we develop in Section 8.4 their respective versions without the Curvatures. Section 8.5 explains the procedure of forecasting recursively out-of-sample with the Kalman filter, re-estimation of the parameters every 12 months, and the additional step we perform to derive forecasts of the US and German yield curves from the forecasts of their drivers. Section 8.6 reports and discusses the results of our out-of-sample forecasting exercise. Finally, Section 8.7 concludes the Chapter.

8.2 Forecasting with the Diebold-Li "Yields-Only" Model

The Diebold-Li "Yields-Only" Model (Diebold, Rudebusch, and Aruoba, 2006) has gained significant popularity in the aftermath of the 2008 Financial Crisis, when regulators placed greater emphasis on the market valuation and accounting of liabilities. The Diebold-Li model is known to have strong forecastability (Diebold and Li, 2006) at longer horizons, where the model performs noticeably better than standard benchmarks, such as the random walk, slope regression, Fama-Bliss forward rate regression, and other autoregressive models. Its parsimony, accurate parameter estimation, and superior forecasting power, make the Diebold-Li model very attractive from the vantage point of market valuation, long-term liability pricing, active bond trading, and credit portfolio risk management, just to name a few. For these reasons, the model is widely acknowledged as state-of-the-art for yield curve modeling and forecasting.

In our forecasting exercise, we choose to benchmark the FSSM, FSSM^{LS}, MShock-FSSM, and MShock-FSSM^{LS} to the Diebold-Li "Yields-Only" Model, which we introduced in its natural state-space form in (2.6,2.7) in Part I of this thesis.

We estimate the Diebold-Li model in (2.6,2.7) with the Kalman filter and maximum likelihood, for both US and German yield data separately. Given that the Diebold-Li model in Diebold,

Rudebusch, and Aruoba, 2006 requires working with mean-adjusted factors and imposes special constraints on the covariance matrices of the state and observation disturbances, for the MATLAB implementation we choose to define a parameter-to-matrix mapping function that does the following. The function maps the input parameter vector to the model matrices, imposes a symmetry constraint on the covariance matrix $\mathbf{Q} = \mathbf{BB}'$ and a diagonality constraint on the covariance matrix $\mathbf{R} = \mathbf{DD}'$. In the MATLAB SSM formulation, the disturbances $\mathbf{v}_t$ and $\mathbf{w}_t$ are defined as uncorrelated, unit-variance white noise vector processes. Therefore, their covariance matrices are identity matrices.

We account for the presence of factor offsets in the state equation of the Diebold-Li model by including a regression component in the measurement equation. In doing so, we avoid increasing the dimensionality of the state vector by modeling the factor offsets as additional states. The disadvantage is that the regression component must deflate the yields during estimation. This step is performed by the mapping function. All other SSM functions must account for this additional step by first deflating and then inflating the yields.

To initialize the Kalman filter, we use the results of the two-step estimation approach presented in Diebold and Li, 2006. The first step of this approach consists in fixing the λ parameter to 0.0609 and running OLS regressions of all observed yields on the Nelson and Siegel, 1987 loadings. The output of this step is a 3-D time series of estimates of the unobserved level, slope, and curvature factors. At the next second step, a VAR(1) model is fit to the three factors. We use the AR coefficient matrix of the VAR(1) model to initialize the transition matrix, **F**, of the state-space model. The matrix **B** of the state-space model is a 3-by-3 matrix constrained such that $\mathbf{Q} = \mathbf{BB}'$ and the estimate of **B** is the lower Cholesky factor of **Q**. We initialize the matrix **B** with the square root of the estimated innovation variances of the VAR(1) model. Similarly to Diebold, Rudebusch, and Aruoba, 2006, we constrain the matrix **D** to be diagonal such that $\mathbf{R} = \mathbf{DD}'$ and initialize it with the square root of the diagonal elements of the sample covariance matrix of the residuals of the VAR(1) model. The measurement sensitivity coefficient matrix **H** is not estimated directly, since it is a fully-parameterized function of the λ parameter.

For the estimation of the model we use the same optimization parameters described in 7.4.4. Once the Diebold-Li model is estimated with the Kalman filter and maximum likelihood, we invoke the `forecast` function to produce level, slope, and curvature factor forecasts, $\{\widehat{l}_{it}, \widehat{s}_{it}, \widehat{c}_{it}\}$, where $i \in \{US, DE\}$. Because in the Diebold-Li model the yield curve depends only on $\{\widehat{l}_{it}, \widehat{s}_{it}, \widehat{c}_{it}\}$, the yield curve is forecasted by forecasting the factors $\{\widehat{l}_{it}, \widehat{s}_{it}, \widehat{c}_{it}\}$. In other words, once the factor forecasts are available, it is sufficient to plug them in the dynamic Nelson-Siegel functional form

$$\widehat{y}_{t+h/t}(\tau) = \widehat{l}_{t+h/t} + \widehat{s}_{t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau}\right) + \widehat{c}_{t+h/t}\left(\frac{1-e^{-\lambda\tau}}{\lambda\tau} - e^{-\lambda\tau}\right) + \nu_t(\tau). \tag{8.1}$$

to derive h-step-ahead yield curve forecasts for each maturity $\tau \in \{6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y\}$.

8.3 Forecasting with the FSSM and MShock-FSSM

In this Chapter, we employ the FSSM and the MShock-FSSM developed in Part II, to forecast the US and German yield curve drivers and the US and German yield curves from the forecasts of the drivers. For convenience, we recall that the state equation of the FSSM is:

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1} = \underbrace{\left[\begin{array}{cccc|ccc} \mathbf{A}1_{2\times 2} & \mathbf{A}2_{2\times 2} & \dots & \mathbf{A}5_{2\times 2} & & & \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & \mathbf{0} & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & & \\ \hline & & & & & & \\ & & \mathbf{0} & & & \mathbf{B}1_{4\times 4} \; \mathbf{B}2_{4\times 4} \; \mathbf{B}3_{4\times 4} & \\ & & & & & \mathbf{I}_{4\times 4} \; \mathbf{0}_{4\times 4} \; \mathbf{0}_{4\times 4} & \\ & & & & & \mathbf{0}_{4\times 4} \; \mathbf{I}_{4\times 4} \; \mathbf{0}_{4\times 4} & \end{array}\right]}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\xi}_{t-5} \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \\ \boldsymbol{\varphi}_{t-3} \end{bmatrix}}_{22\times 1}$$

$$+ \underbrace{\mathbf{B}}_{22\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -- \\ \boldsymbol{\varepsilon}_t \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}}_{22\times 1}, \tag{8.2}$$

and the measurement equation of the FSSM is:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{6\times 1} = \underbrace{\left[\begin{array}{cccc|ccc} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \mathbf{0}_{4\times 2} & \dots & \mathbf{0}_{4\times 2} & \boldsymbol{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{6\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1}$$

$$+ \underbrace{\begin{bmatrix} \mathbf{0}_{2\times 4} \\ \boldsymbol{\Pi}_{4\times 4} \end{bmatrix}}_{6\times 4} \underbrace{\begin{bmatrix} \boldsymbol{y}_{t-1} \end{bmatrix}}_{4\times 1} + \underbrace{\mathbf{D}}_{6\times 6} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{6\times 1}. \tag{8.3}$$

The state equation of the MShock-FSSM is equal to that of the FSSM, hence to 8.2, whereas the measurement equation contains intervention variables incorporated in the regression component and accounting for the most blatant outliers:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{6\times 1} = \underbrace{\left[\begin{array}{ccccc|ccc} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \mathbf{0}_{4\times 2} & \dots & \mathbf{0}_{4\times 2} & & \boldsymbol{I}_{4\times 4} & \mathbf{0}_{4\times 4} & \mathbf{0}_{4\times 4} \end{array}\right]}_{6\times 22} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \\ \boldsymbol{\varphi}_{t-1} \\ \boldsymbol{\varphi}_{t-2} \end{bmatrix}}_{22\times 1}$$
$$+ \underbrace{\begin{bmatrix} \boldsymbol{\delta}_{2\times 2}^{\text{VAR}} & \mathbf{0}_{2\times 4} & \mathbf{0}_{2\times 4} \\ \mathbf{0}_{4\times 2} & \boldsymbol{\delta}_{4\times 4}^{\text{VEC}} & \boldsymbol{\Pi}_{4\times 4} \end{bmatrix}}_{6\times 10} \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}_t^{\text{VAR}} \\ \boldsymbol{\Lambda}_t^{\text{VEC}} \\ \boldsymbol{y}_{t-1} \end{bmatrix}}_{10\times 1} + \underbrace{\mathbf{D}}_{6\times 6} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{6\times 1}, \tag{8.4}$$

where $\boldsymbol{\Lambda}_t^{\text{VAR}}$ and $\boldsymbol{\Lambda}_t^{\text{VEC}}$ are the intervention variables, defined as:

$$\boldsymbol{\Lambda}_{t,s_{US}}^{\text{VAR}} = 0 \tag{8.5}$$

$$\boldsymbol{\Lambda}_{t,s_{DE}}^{\text{VAR}} = \begin{cases} 1, & t = \text{'2008:08'} \\ -1, & t = \text{'2008:09'} \\ 0, & \text{otherwise} \end{cases} \tag{8.6}$$

$$\boldsymbol{\Lambda}_{t,l_{US}}^{\text{VEC}} = \begin{cases} 1, & t = \text{'2008:10'} \\ -1, & t = \text{'2008:12'} \\ 0, & \text{otherwise} \end{cases} \tag{8.7}$$

$$\boldsymbol{\Lambda}_{t,l_{DE}}^{\text{VEC}} = \begin{cases} 1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{8.8}$$

$$\boldsymbol{\Lambda}_{t,c_{US}}^{\text{VEC}} = \begin{cases} -1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{8.9}$$

$$\boldsymbol{\Lambda}_{t,c_{DE}}^{\text{VEC}} = \begin{cases} -1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{8.10}$$

Upon estimation of the FSSM and MShock-FSSM, we invoke the `forecast` function to produce level (in first differences), slope (in levels), and curvature (in first differences) factor forecasts, i.e., $\{\Delta\widehat{l}_{it}, \widehat{s}_{it}, \Delta\widehat{c}_{it}\}$, where $i \in \{US, DE\}$. In the spirit of Diebold-Li, from the drivers'forecasts we derive yield curve forecasts h-step-ahead and for each maturity $\tau \in \{6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y\}$ using the dynamic Nelson-Siegel functional form in (8.1)[3].

[3]Before deriving forecasts of the yield curves, we bring the forecasts of the first differences of the levels and curvatures back in levels with an additional step explained in Section 8.5.

8.4 Forecasting with the FSSM^{LS} and MShock-FSSM^{LS}

In the Dynamic Nelson-Siegel framework (Diebold and Li, 2006), the *long-term factor* β_{1t} governs the yield curve level and is defined as $y_t(\infty) = \beta_{1t}$. The *short-term factor* β_{2t} governs the yield curve slope and is defined as the ten-year yield minus the three-month yield, i.e., $y_t(120) - y_t(3) = \beta_{2t}$. The *medium-term factor* β_{3t} is closely related to the yield curve curvature, which Diebold-Li define as twice the two-year yield minus the sum of the ten-year and three-month yields, i.e., $2y_t(24) - y_t(3) - y_t(120) = \beta_{3t}$. Many studies (Ang and Piazzesi, 2003; Diebold and Li, 2006; and Evans and Marshall, 2007) attempt to provide an economic interpretation to the yield curve factors by observing that these factors are linked and interact dynamically with macroeconomic variables.

Generally, inflation is found to have a strong influence on the yield curve level. More specifically, Evans and Marshall, 2007 found that changes in households'consumption preferences produce large and persistent shifts in the level of the yield curve. Since long-term nominal interest rates are the sum of expected long-run inflation and long-term real interest rates, any structural macroeconomic movement contributing to the determinations of long-run expected inflation or long-term real interest rates (e.g., a change in an inflation-targeting monetary regime or long-term changes in the structural economy, such as, technological innovations) will induce a substantial influence on the yield curve level.

The yield curve slope is closely related to the real economy (Hu, 1993; Peel and Taylor, 1998; Berk, 1998; Rudebusch, 2010; Kurmann and Otrok, 2013). For example, a positive slope is associated with a future increase in consumption, consumer durables, and investment (Estrella and Hardouvelis, 1991). Monetary-policy shocks can explain a large part of variability of the slope (Wu, 2001; Rudebusch and Wu, 2008; Borio, Gambacorta, and Hofmann, 2017).

The predictive power and economic interpretation of the yield curve curvature remains a topic of debate. On one side, several studies show that that the curvature carries predictive information both about the future evolution of the yield curve and the macroeconomy (Giese, 2008; Almeida et al., 2009; Mönch, 2012). The curvature is often linked to the volatility in the interest rates. Christiansen and Lund, 2005 examine the relationship between interest-rate volatility and the shape of the yield curve in a trivariate GARCH-M model for the yield curve level, slope, and curvature, where the conditional short-term volatility is included in the mean specification. The two authors find that the slope and curvature depend positively and significantly on the short-rate volatility, the effect being more pronounced for the curvature than for the slope. Relying on the results of impulse response analysis, Mönch, 2012 explains that unexpected increases of the curvature factor precede a flattening of the yield curve and announce a decrease in output more than 1 year ahead. On the other side, Diebold and Li, 2006 argue that the curvature lacks clear links to macroeconomic variables. Furthermore, missing data at very short and/or very long maturities reduce the estimation precision of the curvature. For the sake of parsimony, these facts motivate the choice of not considering the curvature when estimating term structure factor models. In their empirical application to the term structure of Germany, Japan, the UK, and the US, without loss of information, Diebold, Li, and Yue, 2008 leave out the curvatures when estimating the global Diebold-Li model.

In this Chapter, we verify the predictive power of the US and German curvatures by develop-

ing and forecasting with two additional models, i.e., the FSSM^{LS} and the MShock-FSSM^{LS}. These two models are developed from the original FSSM and MShock-FSSM, respectively, from which we exclude the curvatures. More specifically, let $\boldsymbol{x}_t$ hold the US and German slopes and $\boldsymbol{y}_t$ the US and German levels. From Part I, we recall that the DGP for the levels is a 2D-VEC(1). Therefore, this latter process is cast into state-space form, to compose, together with the SSVAR(5) model for the slopes, the FSSM^{LS}. The state equation of the FSSM^{LS} is:

$$\underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \end{bmatrix}}_{12\times 1} = \underbrace{\left[\begin{array}{ccccc|c} \mathbf{A}1_{2\times 2} & \mathbf{A}2_{2\times 2} & \dots & \mathbf{A}5_{2\times 2} & | & \mathbf{0}_{2\times 2} \\ \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & | & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{I}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & | & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & | & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & | & \mathbf{0}_{2\times 2} \\ -- & -- & -- & -- & -- & -- \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & | & \mathbf{B}1_{2\times 2} \end{array}\right]}_{12\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\xi}_{t-5} \\ \boldsymbol{\varphi}_{t-1} \end{bmatrix}}_{12\times 1}$$
$$+ \underbrace{\mathbf{B}}_{12\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\epsilon}_t \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ \mathbf{0} \\ -- \\ \boldsymbol{\varepsilon}_t \end{bmatrix}}_{12\times 1}, \tag{8.11}$$

and the measurement equation is:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{4\times 1} = \underbrace{\left[\begin{array}{cccc|c} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \boldsymbol{I}_{2\times 2} \end{array}\right]}_{4\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \end{bmatrix}}_{12\times 1}$$
$$+ \underbrace{\begin{bmatrix} \mathbf{0}_{2\times 2} \\ \boldsymbol{\Pi}_{2\times 2} \end{bmatrix}}_{4\times 4} \underbrace{\begin{bmatrix} \boldsymbol{y}_{t-1} \end{bmatrix}}_{2\times 1} + \underbrace{\mathbf{D}}_{4\times 4} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{4\times 1}. \tag{8.12}$$

The state equation of the MShock-FSSM^{LS} is identical to that of the FSSM^{LS} in (8.11), whereas the measurement equation is:

$$\underbrace{\begin{bmatrix} \boldsymbol{x}_t \\ \Delta \boldsymbol{y}_t \end{bmatrix}}_{4\times 1} = \underbrace{\left[\begin{array}{cccc|c} \mathbf{I}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} & \dots & \mathbf{0}_{2\times 2} & \boldsymbol{I}_{2\times 2} \end{array}\right]}_{4\times 12} \underbrace{\begin{bmatrix} \boldsymbol{\xi}_t \\ \boldsymbol{\xi}_{t-1} \\ \boldsymbol{\xi}_{t-2} \\ \boldsymbol{\xi}_{t-3} \\ \boldsymbol{\xi}_{t-4} \\ \boldsymbol{\varphi}_t \end{bmatrix}}_{12\times 1} + \underbrace{\begin{bmatrix} \boldsymbol{\delta}_{2\times 2}^{\text{VAR}} & \mathbf{0}_{2\times 2} & \mathbf{0}_{2\times 2} \\ \mathbf{0}_{2\times 2} & \boldsymbol{\delta}_{2\times 2}^{\text{VEC}} & \boldsymbol{\Pi}_{2\times 2} \end{bmatrix}}_{4\times 6} \underbrace{\begin{bmatrix} \boldsymbol{\Lambda}_t^{\text{VAR}} \\ \boldsymbol{\Lambda}_t^{\text{VEC}} \\ \boldsymbol{y}_{t-1} \end{bmatrix}}_{6\times 1} + \underbrace{\mathbf{D}}_{4\times 4} \underbrace{\begin{bmatrix} \boldsymbol{u}_t \\ \boldsymbol{\eta}_t \end{bmatrix}}_{4\times 1}, \tag{8.13}$$

where $\boldsymbol{\Lambda}_t^{\text{VAR}}$ and $\boldsymbol{\Lambda}_t^{\text{VEC}}$ are defined as:

$$\boldsymbol{\Lambda}_{t,s_{US}}^{\text{VAR}} = 0 \tag{8.14}$$

$$\boldsymbol{\Lambda}_{t,s_{DE}}^{\text{VAR}} = \begin{cases} 1, & t = \text{'2008:08'} \\ -1, & t = \text{'2008:09'} \\ 0, & \text{otherwise} \end{cases} \tag{8.15}$$

$$\boldsymbol{\Lambda}_{t,l_{US}}^{\text{VEC}} = \begin{cases} 1, & t = \text{'2008:10'} \\ -1, & t = \text{'2008:12'} \\ 0, & \text{otherwise} \end{cases} \tag{8.16}$$

$$\boldsymbol{\Lambda}_{t,l_{DE}}^{\text{VEC}} = \begin{cases} 1, & t = \text{'2008:10'} \\ 0, & \text{otherwise} \end{cases} \tag{8.17}$$

We estimate the FSSM^{LS} and the MShock-FSSM^{LS} with the Kalman filter and maximum likelihood. The Kalman filter is initialized in the same way as the FSSM and MShock-FSSM, excluding the curvatures. The assumption of independent residuals is tested on the standardized one-step prediction errors of the equivalent SSVEC^{L} (i.e., the SSVEC model for the US and German levels only), from which we take out all the zero restrictions imposed for the estimation of the FSSM. The assumption of independence holds for the standardized one-step prediction errors of the unrestricted SSVEC^{L} model.

Upon estimation of the FSSM^{LS} and the MShock-FSSM^{LS}, we invoke the `forecast` function to produce level (in first differences) and slope (in levels) factor forecasts, i.e., $\{\Delta\widehat{l}_{it}, \widehat{s}_{it}\}$, where $i \in \{US, DE\}$. In the spirit of Diebold-Li, from the drivers'forecasts we derive yield curve forecasts h-step-ahead and for each maturity $\tau \in \{6M, 1Y, 2Y, 3Y, 5Y, 7Y, 10Y\}$ using the dynamic Nelson-Siegel functional form in (8.1)[4].

[4]Before deriving forecasts of the yield curves, we bring the forecasts of the first differences of the levels back in levels with an additional step explained in Section 8.5.

8.5 Recursive Out-Of-Sample Forecasting with the Kalman Filter

Using the FSSM, MShock-FSSM, FSSM^{LS}, and MShock-FSSM^{LS}, we perform a *recursive out-of-sample forecasting with re-estimation* for the period from 1999:01 through 2018:01, in order to produce forecasts of the US and German yield curve drivers. From the forecasts of the drivers, we calculate forecasts for the US and German yield curves. The estimation period includes major market events of the last decades, such as the dot-com bubble in 2000, the 2008 Financial Crisis, the European Sovereign Debt crisis, the European recession in 2012-2013, and the non-traditional monetary policy interventions of the US Fed and the ECB, i.e., the phases of quantitative easing.

The forecasting results are compared to the state-of-the-art Diebold-Li "Yields-Only" Model using the *Root Mean Square Errors* (RMSEs). We calculate the RMSE statistic using the following procedure. We start with a sub-sample of the data, more specifically, the first 1/3 of data[5]: [1999:06-2005:09]. We then employ the Kalman filter and maximum likelihood to estimate the parameters of the model and produce 1-month-, 2-month-, 3-month-, 2-quarter-, 3-quarter-, and 4-quarter-ahead forecasts of the US and German yield curve drivers, i.e., forecasts of the US and German level, slope, and curvature. In the spirit of Diebold-Li, given forecasts of the drivers we can calculate forecasts of the US and German yield curves using the dynamic Nelson and Siegel, 1987 functional form in 8.1.

We recall that, in our state-space models, the levels and curvatures are in first differences (i.e., $\Delta\mathbf{x}_t = \mathbf{x}_t - \mathbf{x}_{t-1}$), therefore, their forecasts are the forecasts of the first differences (i.e., $\Delta\widehat{\mathbf{x}}_{t+h/t}$). Before calculating forecasts of the yield curves, we bring the forecasts of the first differences of the levels and curvatures back in levels with the following step[6]:

$$\widehat{\mathbf{x}}_t = \mathbf{x}_{t-1} + \Delta\widehat{\mathbf{x}}_t. \tag{8.18}$$

Given the yield forecasts, $\widehat{\mathbf{y}}_{t+h/t}(\tau)$, we compare the forecasted value with the actual value of the yield and calculate the *Squared Forecast Error* (SFE) as follows:

$$\text{SFE} = \left(y_{t+h}(\tau) - \hat{y}_{t+h/t}(\tau)\right)^2 \tag{8.19}$$

Keeping the beginning of the in-sample period fixed[7] for all forecasts, we extend the in-sample period with one more (actual) observation and start again the forecasting exercise, given the new in-sample and out-of-sample period.

Because the parameters change slowly, we choose to re-estimate the parameters every 12 months. The repetitions of the forecasting procedure last until we reach the full length of the sample, by that time we accumulate a *series of SFEs*. It takes 137 repetitions to reach the full length of the sample. Over these repetitions, we estimate the model 10 times, for the following estimation periods:

[5]The first 1/3 of the data equals 76 monthly observations

[6]We get the forecast in levels by adding the forecast of the first differences to the last in-sample observation of the US and German levels and curvatures.

[7]This is in contrast to the rolling-window approach, in which the in-sample period is shifted.

Table 8.1: Estimation Periods

Nr	Estimation Period	Nr	Estimation Period
1	[1999:06 - 2005:09]	6	[1999:06 - 2010:09]
2	[1999:06 - 2006:09]	7	[1999:06 - 2011:09]
3	[1999:06 - 2007:09]	8	[1999:06 - 2012:09]
4	[1999:06 - 2008:09]	9	[1999:06 - 2013:09]
5	[1999:06 - 2009:09]	10	[1999:06 - 2017:01]

In an attempt to understand the stability of the parameters over time, we calculate the eigenvalues[8] of the estimated state transition coefficient matrices of the FSSM. Since we performed 10 estimations, we have 10 such coefficient matrices per each sub-model (i.e., SSVAR and SSVEC[9]) of the FSSM to analyze. Because the eigenvalues of the matrices are complex numbers, we report, in Figure 8.1, their absolute values. In the SSVAR model, the magnitudes across the 10 estimations vary mostly for the 4th, 5th, 6th, and 9th eigenvalues, whereas in the SSVEC, for the 4th, 5th, 6th, 7th, and 8th eigenvalues.

For the comparison of the models'forecasting performance, we report the RMSE as a measure of each model 's forecasting accuracy. The RMSEs are calculated as follows:

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{137} \left(y_{t+h}(\tau) - \hat{y}_{t+h/t}(\tau)\right)^2}{137}} \tag{8.20}$$

In addition to the RMSE, we also report the *mean* and *standard deviation* as descriptive statistics for the forecast errors of the yields.

8.6 Out-Of-Sample Forecasting Performance Comparison

This Section reports the results of the out-of-sample forecasting exercise. Similarly to Diebold and Li, 2006, in Tables 8.6.1, 8.6.2, 8.6.3, 8.6.4, 8.6.5, and 8.6.6 we compare the h-step-ahead out-of-sample forecasting results of the FSSM (and its related versions) to those of Diebold-Li "Yields-Only" model, for maturities of 6 months, 1, 2, 3, 5, 7, and 10 years, and forecast horizons of h = 1,2, 3 months, 2, 3, 4 quarters. In addition to the RMSE, we also report the mean and standard deviations of the forecast errors.

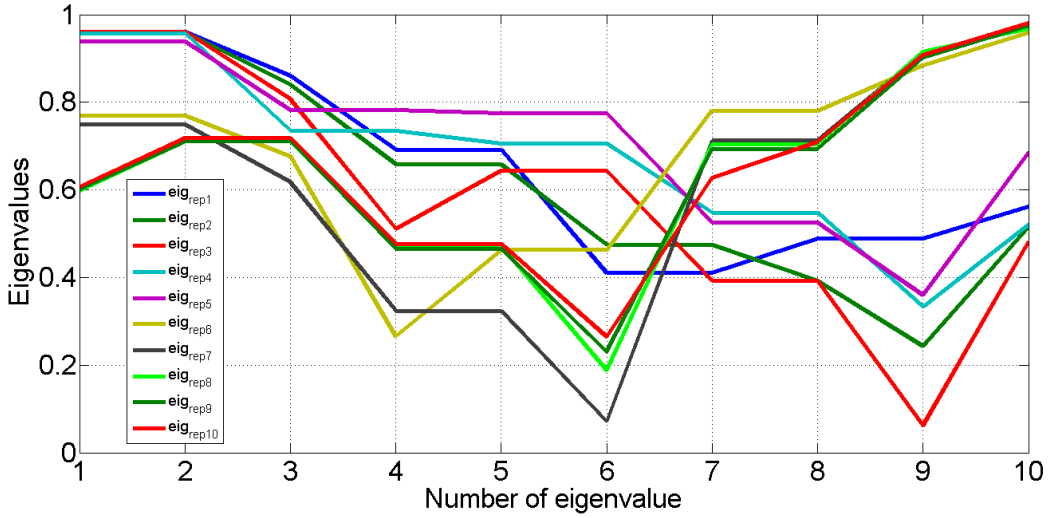
For an easier assessment of the changes in the forecasting accuracy among the models, in Tables 8.6.1, 8.6.2, 8.6.3, 8.6.4, 8.6.5, and 8.6.6, we report ΔRMSE%, i.e., the percentage changes in the RMSEs among the models.

Starting with the 1-month-ahead forecasting results (reported in Tables 8.6.1 and 8.6.1), the US forecasts with the FSSM are better than the benchmark only for the 10-year maturity and

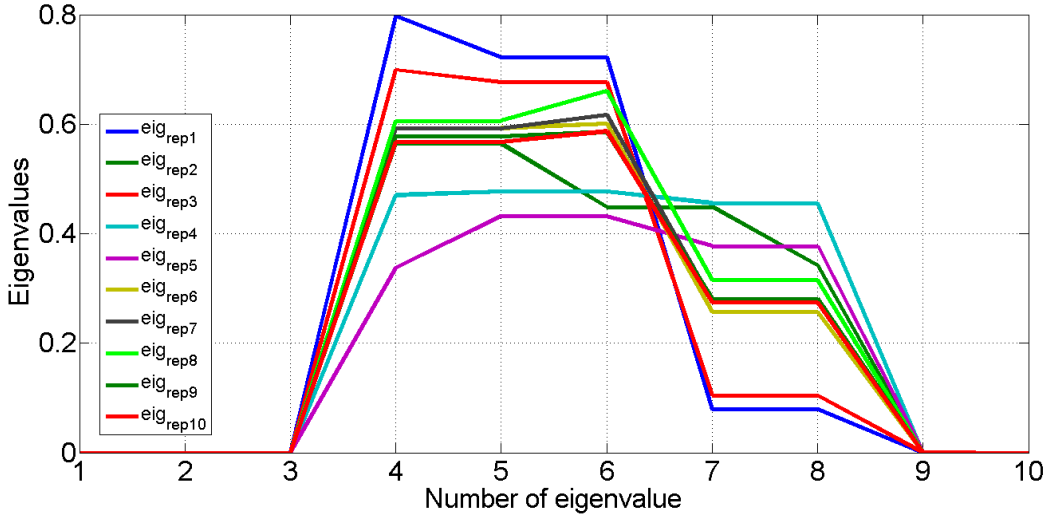
[8]The eigenvalues are a special set of scalars associated with a linear system of equations (equivalently, a matrix equation) and are also known as *characteristic roots* (Hoffman and Kunze, 1971; Marcus and Minc, 1988, p. 144). The eigenvalues and eigenvectors are commonly applied in practice for stability analysis.

[9]The first three and the last three eigenvalues are always zero because of zero restrictions imposed on the short-run coefficient matrices of the SSVEC model.

Figure 8.1: Absolute values of the eigenvalues of the estimated state transition coefficient matrices of the SSVAR and SSVEC in the FSSM.



(a) abs($\text{eig}_{\text{SSVAR}}$)



(b) abs($\text{eig}_{\text{SSVEC}}$)

the German forecasts are better for the 6-month and 1-year maturities. This observation is also visible in the negative sign of the ΔRMSE% reported in Table 8.6.1. Similar remarks hold for the forecasting results of the MShock-FSSM model, for which, however, the German forecasts are better than the benchmark only for the 6-month maturity. Comparing the FSSM with the MShock-FSSM, we can notice that the MShock-FSSM performs better than the FSSM for the US at long-term maturities (5-year, 7-year, and 10-year). To a certain extent, this improved forecasting accuracy of the MShock-FSSM is expected, since the MShock-FSSM is the FSSM, extended with additional exogenous variables, i.e., the measurement shocks accounting for the patches of outlying values. The versions of the FSSM without the curvatures, i.e., the FSSM^{LS} and MShock-FSSM^{LS}, show poor results compared to both the FSSM and Diebold-Li.

The 2-month-ahead forecasting results (reported in Tables 8.6.2 and 8.6.2), of the FSSM are better than the Diebold-Li only for the short- and long-term German maturities, whereas, the results of the MShock-FSSM are better for all German maturities and only for the US 10-year maturity. Comparing the FSSM with the MShock-FSSM, we can notice a slight improvement for the US forecasts and a significant improvement for the German forecasts, where the MShock-FSSM always outperforms the FSSM. Once again, these results are expected since the MShock-FSSM includes additional exogenous variables, which might cause a false sense of confidence in the improved forecasts.

The results improve significantly from the 3-month-ahead horizon (Tables 8.6.3 and 8.6.3), where the FSSM outperforms the benchmark for all German maturities and the MShock-FSSM always outperforms the FSSM for all US and German maturities.

For the US, the FSSM starts to perform better than the benchmark from 2-quarter-ahead (for 1-year, 2-year, 3-year maturity, Tables 8.6.4 and 8.6.4). At this forecast horizon, the MShock-FSSM always performs better than the FSSM. Similar results hold for the 3-quarter-ahead horizon (Tables 8.6.5 and 8.6.5).

At the 4-quarter-ahead horizon (Tables 8.6.6 and 8.6.6), the FSSM performs better than the benchmark for all US maturities and German mid- to long-term maturities. And once again, the MShock-FSSM performs better than the FSSM for all US and German maturities.

Overall, the following conclusions can be drawn about the forecasting performance of our FSSM and its related versions.

- At short forecast horizons, the FSSM outperforms the Diebold-Li model at short-term maturities, for Germany, and at the long-term maturity, for the US;
- At the 1-month-ahead horizon, the MShock-FSSM performs better than the FSSM only for the US and only for long-term maturities;
- At the 3-month-ahead horizon, the FSSM outperforms the benchmark for all German maturities;
- At the 4-quarter-ahead horizon, the FSSM performs better than the Diebold-Li for all US maturities and German mid- and long-term maturities. The model shows, therefore, a strong yield curve forecastability at the longest horizon.

- For all horizons from 2-month-ahead to 4-quarter-ahead, the MShock-FSSM always performs better than the FSSM for Germany. For the US, only from 3-month-ahead horizon onwards.
- The models without the curvatures perform poorly at all forecast horizons, for both US and Germany, and for all maturities, suggesting that for our dataset, the curvatures hold predictive power.

The FSSM and its related versions can be considered as extensions of the Diebold-Li "Yields-Only" Model in that all models represent the yield curve in terms of its drivers, i.e., the level, slope, and curvature, and forecast the yield curve by forecasting its drivers. In line with Diebold and Li, 2006, the strong yield curve forecastability of the FSSM at the 4-quarter-ahead is an important and attractive feature for many tasks, including active bond trading, fixed income portfolio management, credit portfolio risk management, insurance and pension analysis. An important reason to prefer the FSSM is that, on our sample of yields, the FSSM is able to forecast more accurately than the Diebold-Li model while at the same time preserving the dynamic properties of the yield drivers.

8.6.1 Out-Of-Sample 1-Month-Ahead Forecasting Results

Table 8.2: Out-of-sample 1-month-ahead forecasting results

Diebold-Li "Yields-Only" Model							
			US				
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0423	-0.0075	-0.0094	-0.0424	-0.0595	-0.0482	-0.0566
Std. Dev.	0.1660	0.1615	0.1921	0.1935	0.2073	0.2208	0.2397
RMSE	0.1708	0.1611	0.1916	0.1974	0.2149	0.2252	0.2454
			Germany				
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.1721	-0.1060	-0.0630	-0.0597	-0.0700	-0.0649	-0.0330
Std. Dev.	0.2658	0.2230	0.2176	0.2157	0.2102	0.2078	0.2128
RMSE	0.3159	0.2462	0.2258	0.2231	0.2208	0.2170	0.2145
FSSM							
			US				
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0461	0.0317	0.0329	-0.0238	-0.0545	-0.0114	0.0481
Std. Dev.	0.2214	0.1916	0.2094	0.2308	0.2631	0.2591	0.2407
RMSE	0.2254	0.1935	0.2112	0.2312	0.2678	0.2584	0.2446

Continued on next page

Table 8.2 – *Continued from previous page*

Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0477	0.0543	0.0579	-0.0102	-0.0914	-0.0581	0.0856
Std. Dev.	0.2748	0.2321	0.2294	0.2312	0.2315	0.2202	0.2086
RMSE	0.2779	0.2376	0.2358	0.2306	0.2481	0.2270	0.2248

FSSM^{LS}

US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5489	-0.7793	-1.0188	-1.0750	-0.9124	-0.6776	-0.4270
Std. Dev.	0.3733	0.4798	0.6073	0.6504	0.5782	0.4799	0.3699
RMSE	0.6634	0.9145	1.1852	1.2555	1.0793	0.8296	0.5643

Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5378	-0.7042	-0.9099	-0.9726	-0.8797	-0.6782	-0.3674
Std. Dev.	0.3841	0.4539	0.5315	0.5485	0.5012	0.4198	0.3206
RMSE	0.6535	0.8315	1.0450	1.1088	1.0065	0.7928	0.485

MShock-FSSM

US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0439	0.0341	0.0353	-0.0212	-0.0522	-0.0097	0.0492
Std. Dev.	0.2334	0.2044	0.2227	0.2397	0.2562	0.2456	0.2204
RMSE	0.2367	0.2065	0.2247	0.2398	0.2605	0.2449	0.2250

Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0357	0.0664	0.0697	0.0012	-0.0810	-0.0485	0.0946
Std. Dev.	0.2868	0.2389	0.2290	0.2325	0.2415	0.2376	0.2314
RMSE	0.2880	0.2472	0.2385	0.2317	0.2539	0.2416	0.2492

MShock-FSSM^{LS}

US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5491	-0.7809	-1.0219	-1.0780	-0.9152	-0.6806	-0.4299
Std. Dev.	0.3881	0.4903	0.6168	0.6583	0.5793	0.4780	0.3623
RMSE	0.6716	0.9211	1.1924	1.2619	1.0820	0.8307	0.5614

Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y

Continued on next page

Table 8.2 – *Continued from previous page*

Mean	-0.4932	-0.6551	-0.8581	-0.9263	-0.8429	-0.6463	-0.3395
Std. Dev.	0.4199	0.5004	0.5846	0.5992	0.5436	0.4571	0.3509
RMSE	0.6468	0.8233	1.0371	1.1020	1.0020	0.7906	0.4873

Note: The table reports the results of out-of-sample 1-month-ahead forecasting using the state-space models previously described. The models are estimated recursively, with yearly re-estimation, from 1999:06 till 2017:01. The forecast errors are defined at $t+h$ as $y_{t+h}(\tau) - \hat{y}_{t+\frac{h}{t}}(\tau)$, where $h = 1, 2, 3, 6, 9, 12$. The table reports the mean, standard deviation and RMSE of the forecast errors.

Table 8.3: Accuracy changes: Out-of-sample 1-month-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	31.97%	20.11%	10.23%	17.12%	24.62%	14.74%	-0.33%
Germany	-12.03%	-3.49%	4.43%	3.36%	12.36%	4.61%	4.80%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	288.41%	467.66%	518.58%	536.02%	402.23%	268.38%	129.95%
Germany	106.87%	237.73%	362.80%	397.00%	355.84%	265.35%	126.11%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	38.58%	28.18%	17.28%	21.48%	21.22%	8.75%	-8.31%
Germany	-8.83%	0.41%	5.62%	3.85%	14.99%	11.34%	16.18%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	293.21%	471.76%	522.34%	539.26%	403.49%	268.87%	128.77%
Germany	104.75%	234.40%	359.30%	393.95%	353.80%	264.33%	127.18%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	194.32%	372.61%	461.17%	443.04%	303.02%	221.05%	130.70%
Germany	135.16%	249.96%	343.17%	380.83%	305.68%	249.25%	115.75%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	5.01%	6.72%	6.39%	3.72%	-2.73%	-5.22%	-8.01%
Germany	3.63%	4.04%	1.15%	0.48%	2.34%	6.43%	10.85%
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	197.96%	376.02%	464.58%	445.80%	304.03%	221.48%	129.52%
Germany	132.75%	246.51%	339.82%	377.88%	303.87%	248.28%	116.77%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							

Continued on next page

Table 8.3 – *Continued from previous page*

	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	183.73%	346.05%	430.66%	426.23%	315.36%	239.20%	149.51%
Germany	124.58%	233.05%	334.84%	375.62%	294.64%	227.24%	95.55%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

8.6.2 Out-Of-Sample 2-Month-Ahead Forecasting Results

Table 8.4: Out-of-sample 2-month-ahead forecasting results

Diebold-Li "Yields-Only" Model							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0735	-0.0425	-0.0489	-0.0835	-0.1024	-0.0921	-0.1024
Std. Dev.	0.2997	0.2904	0.3199	0.3207	0.3180	0.3306	0.3452
RMSE	0.3075	0.2924	0.3225	0.3303	0.3330	0.3420	0.3588
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.2315	-0.1675	-0.1274	-0.1254	-0.1354	-0.1284	-0.0938
Std. Dev.	0.4045	0.3749	0.3612	0.3475	0.3223	0.3083	0.3074
RMSE	0.4648	0.4094	0.3818	0.3683	0.3485	0.3330	0.3203
FSSM							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0606	0.0157	0.0172	-0.0368	-0.0623	-0.0154	0.0469
Std. Dev.	0.3644	0.3350	0.3474	0.3731	0.4032	0.4022	0.3808
RMSE	0.3681	0.3342	0.3466	0.3736	0.4065	0.4010	0.3823
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0475	0.0554	0.0583	-0.0113	-0.0951	-0.0630	0.0802
Std. Dev.	0.4179	0.3917	0.3814	0.3703	0.3462	0.3202	0.2942
RMSE	0.4190	0.3942	0.3845	0.3691	0.3578	0.3252	0.3039
FSSM^{LS}							

Continued on next page

Table 8.4 – *Continued from previous page*

				US			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5431	-0.7722	-1.0091	-1.0627	-0.8966	-0.6597	-0.4081
Std. Dev.	0.4445	0.5237	0.6430	0.6913	0.6388	0.5630	0.4684
RMSE	0.7015	0.9324	1.1958	1.2668	1.1000	0.8664	0.6204
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5519	-0.7170	-0.9216	-0.9832	-0.8879	-0.6840	-0.3705
Std. Dev.	0.4861	0.5374	0.5906	0.5966	0.5464	0.4732	0.3877
RMSE	0.7267	0.8884	1.0839	1.1405	1.0356	0.8265	0.5344
				MShock-FSSM			
				US			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0483	0.0282	0.0300	-0.0237	-0.0490	-0.0019	0.0605
Std. Dev.	0.3714	0.3396	0.3530	0.3732	0.3866	0.3742	0.3460
RMSE	0.3732	0.3395	0.3530	0.3726	0.3883	0.3728	0.3500
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0330	0.0697	0.0719	0.0013	-0.0839	-0.0527	0.0896
Std. Dev.	0.4080	0.3780	0.3669	0.3572	0.3379	0.3167	0.2955
RMSE	0.4079	0.3830	0.3725	0.3559	0.3470	0.3199	0.3078
				MShock-FSSM^{LS}			
				US			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5319	-0.7624	-1.0009	-1.0548	-0.8900	-0.6544	-0.4039
Std. Dev.	0.4577	0.5302	0.6481	0.6939	0.6338	0.5527	0.4525
RMSE	0.7006	0.9275	1.1911	1.2611	1.0912	0.8553	0.6053
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4815	-0.6455	-0.8516	-0.9218	-0.8399	-0.6429	-0.3353
Std. Dev.	0.5204	0.5865	0.6501	0.6537	0.5918	0.5087	0.4104
RMSE	0.7076	0.8707	1.0699	1.1287	1.0262	0.8187	0.5288

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the benchmark model, Model B, and the alternative model, Model A, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

Table 8.5: Accuracy changes: Out-of-sample 2-month-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	19.71%	14.30%	7.47%	13.11%	22.07%	17.25%	6.55%
Germany	-9.85%	-3.71%	0.71%	0.22%	2.67%	-2.34%	-5.12%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	128.13%	218.88%	270.79%	283.53%	230.33%	153.33%	72.91%
Germany	56.35%	117.00%	183.89%	209.67%	197.16%	148.20%	66.84%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	21.37%	16.11%	9.46%	12.81%	16.61%	9.01%	-2.45%
Germany	-12.24%	-6.45%	-2.44%	-3.37%	-0.43%	-3.93%	-3.90%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	127.84%	217.20%	269.33%	281.80%	227.69%	150.09%	68.70%
Germany	52.24%	112.68%	180.23%	206.46%	194.46%	145.86%	65.10%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	90.57%	178.99%	245.01%	239.08%	170.60%	116.06%	62.28%
Germany	73.44%	125.37%	181.90%	208.99%	189.44%	154.15%	75.85%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	1.39%	1.59%	1.85%	-0.27%	-4.48%	-7.03%	-8.45%
Germany	-2.65%	-2.84%	-3.12%	-3.58%	-3.02%	-1.63%	1.28%
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	90.33%	177.53%	243.65%	237.55%	168.44%	113.29%	58.33%
Germany	68.88%	120.88%	178.26%	205.80%	186.81%	151.75%	74.00%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	87.73%	173.20%	237.42%	238.46%	181.02%	129.43%	72.94%
Germany	73.47%	127.34%	187.22%	217.14%	195.73%	155.92%	71.80%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes among the models. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

8.6.3 Out-Of-Sample 3-Month-Ahead Forecasting Results

Table 8.6: Out-of-sample 3-month-ahead forecasting results

Diebold-Li "Yields-Only" Model							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.1048	-0.0779	-0.0895	-0.1265	-0.1468	-0.1371	-0.1487
Std. Dev.	0.4287	0.4196	0.4441	0.4417	0.4174	0.4180	0.4229
RMSE	0.4398	0.4253	0.4514	0.4578	0.4411	0.4384	0.4468
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.2910	-0.2296	-0.1923	-0.1916	-0.2010	-0.1920	-0.1544
Std. Dev.	0.5286	0.5039	0.4824	0.4584	0.4149	0.3878	0.3758
RMSE	0.6017	0.5521	0.5177	0.4953	0.4597	0.4314	0.4050
FSSM							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0759	-0.0010	0.0008	-0.0507	-0.0703	-0.0189	0.0468
Std. Dev.	0.4863	0.4522	0.4574	0.4835	0.5090	0.5052	0.4777
RMSE	0.4904	0.4505	0.4557	0.4844	0.5120	0.5037	0.4782
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0463	0.0555	0.0565	-0.0153	-0.1017	-0.0706	0.0724
Std. Dev.	0.5424	0.5230	0.5039	0.4811	0.4356	0.3955	0.3555
RMSE	0.5424	0.5240	0.5052	0.4796	0.4457	0.4003	0.3615
FSSM^{LS}							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5428	-0.7703	-1.0046	-1.0556	-0.8856	-0.6461	-0.3929
Std. Dev.	0.5269	0.5749	0.6781	0.7279	0.6908	0.6301	0.5437
RMSE	0.7557	0.9603	1.2110	1.2811	1.1219	0.9012	0.6695
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5686	-0.7326	-0.9360	-0.9970	-0.8988	-0.6928	-0.3764
Std. Dev.	0.5782	0.6153	0.6455	0.6387	0.5796	0.5078	0.4266

Continued on next page

Table 8.6 – *Continued from previous page*

RMSE	0.7978	0.9443	1.1207	1.1693	1.0597	0.8520	0.5671

MShock-FSSM							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0523	0.0218	0.0227	-0.0292	-0.0491	0.0022	0.0677
Std. Dev.	0.4800	0.4465	0.4509	0.4748	0.4889	0.4769	0.4483
RMSE	0.4810	0.4453	0.4498	0.4739	0.4896	0.4752	0.4517
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0318	0.0698	0.0699	-0.0028	-0.0906	-0.0604	0.0817
Std. Dev.	0.5295	0.5086	0.4908	0.4696	0.4276	0.3905	0.3538
RMSE	0.5285	0.5115	0.4940	0.4679	0.4356	0.3938	0.3618

MShock-FSSM^{LS}							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5200	-0.7493	-0.9860	-1.0382	-0.8706	-0.6332	-0.3815
Std. Dev.	0.5181	0.5651	0.6675	0.7173	0.6754	0.6114	0.5232
RMSE	0.7327	0.9372	1.1894	1.2604	1.1004	0.8786	0.646
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4697	-0.6365	-0.8459	-0.9183	-0.8378	-0.6407	-0.3322
Std. Dev.	0.6195	0.6722	0.7134	0.7050	0.6343	0.5503	0.4532
RMSE	0.7756	0.9240	1.1049	1.1561	1.0494	0.8433	0.5606

Note: The table reports the results of out-of-sample 3-month-ahead forecasting using the state-space models previously described. The models are estimated recursively, with yearly re-estimation, from 1999:06 till 2017:01. The forecast errors are defined at $t+h$ as $y_{t+h}(\tau) - \hat{y}_{t+\frac{h}{t}}(\tau)$, where $h = 1, 2, 3, 6, 9, 12$. The table reports the mean, standard deviation and RMSE of the forecast errors.

Table 8.7: Accuracy changes: Out-of-sample 3-month-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	11.51%	5.93%	0.95%	5.81%	16.07%	14.90%	7.03%
Germany	-9.86%	-5.09%	-2.41%	-3.17%	-3.05%	-7.21%	-10.74%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y

Continued on next page

Table 8.7 – *Continued from previous page*

US	71.83%	125.79%	168.28%	179.84%	154.34%	105.57%	49.84%
Germany	32.59%	71.04%	116.48%	136.08%	130.52%	97.50%	40.02%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	9.37%	4.70%	-0.35%	3.52%	11.00%	8.39%	1.10%
Germany	-12.17%	-7.35%	-4.58%	-5.53%	-5.24%	-8.72%	-10.67%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	66.60%	120.36%	163.49%	175.32%	149.47%	100.41%	44.58%
Germany	28.90%	67.36%	113.42%	133.41%	128.28%	95.48%	38.42%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	54.10%	113.16%	165.75%	164.47%	119.12%	78.92%	40.00%
Germany	47.09%	80.21%	121.83%	143.81%	137.76%	112.84%	56.87%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	1.92%	1.15%	1.29%	2.17%	4.38%	5.66%	5.54%
Germany	2.56%	2.39%	2.22%	2.44%	2.27%	1.62%	-0.08%
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	49.41%	108.04%	161.01%	160.20%	114.92%	74.43%	35.09%
Germany	42.99%	76.34%	118.71%	141.06%	135.45%	110.67%	55.08%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	52.33%	110.46%	164.43%	165.96%	124.75%	84.89%	43.02%
Germany	46.75%	80.65%	123.66%	147.08%	140.91%	114.14%	54.95%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

8.6.4 Out-Of-Sample 2-Quarter-Ahead Forecasting Results

Table 8.8: Out-of-sample 2-quarter-ahead forecasting results

Diebold-Li "Yields-Only" Model							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y

Continued on next page

Table 8.8 – *Continued from previous page*

Mean	-0.1950	-0.1792	-0.2043	-0.2468	-0.2700	-0.2605	-0.2738
Std. Dev.	0.7459	0.7434	0.7409	0.7149	0.6482	0.6132	0.5949
RMSE	0.7683	0.7621	0.7659	0.7538	0.7000	0.6641	0.6529

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4610	-0.4078	-0.3796	-0.3819	-0.3884	-0.3722	-0.3242
Std. Dev.	0.7897	0.7764	0.7379	0.6966	0.6252	0.5784	0.5512
RMSE	0.9119	0.8745	0.8275	0.7921	0.7341	0.6860	0.6378

FSSM

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.1245	-0.0547	-0.0529	-0.0968	-0.0988	-0.0338	0.0431
Std. Dev.	0.7948	0.7440	0.7087	0.7244	0.7567	0.7510	0.7202
RMSE	0.8016	0.7432	0.7081	0.7282	0.7604	0.7490	0.7188

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0407	0.0615	0.0588	-0.0182	-0.1127	-0.0853	0.0560
Std. Dev.	0.8974	0.8810	0.8358	0.7869	0.6957	0.6231	0.5522
RMSE	0.8950	0.8799	0.8348	0.7842	0.7022	0.6267	0.5530

FSSM^{LS}

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5384	-0.7607	-0.9870	-1.0305	-0.8492	-0.6024	-0.3443
Std. Dev.	0.7906	0.7625	0.7990	0.8396	0.8360	0.7995	0.7288
RMSE	0.9548	1.0758	1.2687	1.3279	1.1901	0.9992	0.8041

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.6384	-0.7973	-0.9956	-1.0513	-0.9394	-0.7237	-0.3953
Std. Dev.	0.9243	0.9238	0.8964	0.8542	0.7629	0.6873	0.6082
RMSE	1.0677	1.1624	1.2764	1.2987	1.1746	0.9727	0.7156

MShock-FSSM

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0944	-0.0263	-0.0266	-0.0714	-0.0743	-0.0099	0.0665
Std. Dev.	0.7804	0.7309	0.6943	0.7095	0.7371	0.7316	0.6991

Continued on next page

Table 8.8 – *Continued from previous page*

RMSE	0.7833	0.7287	0.6922	0.7105	0.7381	0.7290	0.6997
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0238	0.0778	0.0737	-0.0046	-0.1008	-0.0746	0.0658
Std. Dev.	0.8620	0.8494	0.8113	0.7675	0.6839	0.6170	0.5521
RMSE	0.8591	0.8498	0.8117	0.7647	0.6888	0.6192	0.5540
MShock-FSSM^{LS}							
				US			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5103	-0.7357	-0.9664	-1.0124	-0.8346	-0.5898	-0.3333
Std. Dev.	0.7762	0.7495	0.7871	0.8297	0.8267	0.7927	0.7206
RMSE	0.9265	1.0483	1.2446	1.3070	1.1726	0.9857	0.7915
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4338	-0.6094	-0.8299	-0.9079	-0.8304	-0.6319	-0.3198
Std. Dev.	0.9315	0.9505	0.9385	0.9033	0.8133	0.7279	0.6357
RMSE	1.0244	1.1261	1.2503	1.2784	1.1602	0.9619	0.7095

Note: The table reports the results of out-of-sample 2-quarter-ahead forecasting using the state-space models previously described. The models are estimated recursively, with yearly re-estimation, from 1999:06 till 2017:01. The forecast errors are defined at $t+h$ as $y_{t+h}(\tau) - \hat{y}_{t+\frac{h}{t}}(\tau)$, where $h = 1, 2, 3, 6, 9, 12$. The table reports the mean, standard deviation and RMSE of the forecast errors.

Table 8.9: Accuracy changes: Out-of-sample 2-quarter-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	4.33%	-2.48%	-7.55%	-3.40%	8.63%	12.78%	10.09%
Germany	-1.85%	0.62%	0.88%	-1.00%	-4.35%	-8.64%	-13.30%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	24.27%	41.16%	65.65%	76.16%	70.01%	50.46%	23.16%
Germany	17.09%	32.92%	54.25%	63.96%	60.01%	41.79%	12.20%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	1.95%	-4.38%	-9.62%	-5.74%	5.44%	9.77%	7.17%
Germany	-5.79%	-2.82%	-1.91%	-3.46%	-6.17%	-9.74%	-13.14%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							

Continued on next page

Table 8.9 – *Continued from previous page*

	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	20.59%	37.55%	62.50%	73.39%	67.51%	48.43%	21.23%
Germany	12.34%	28.77%	51.09%	61.39%	58.04%	40.22%	11.24%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	19.11%	44.75%	79.17%	82.35%	56.51%	33.40%	11.87%
Germany	19.30%	32.11%	52.90%	65.61%	67.27%	55.21%	29.40%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-2.28%	-1.95%	-2.25%	-2.43%	-2.93%	-2.67%	-2.66%
Germany	-4.01%	-3.42%	-2.77%	-2.49%	-1.91%	-1.20%	0.18%
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	15.58%	41.05%	75.77%	79.48%	54.21%	31.60%	10.11%
Germany	14.46%	27.98%	49.77%	63.02%	65.22%	53.49%	28.30%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	18.28%	43.86%	79.80%	83.95%	58.87%	35.21%	13.12%
Germany	19.24%	32.51%	54.03%	67.18%	68.44%	55.35%	28.07%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

8.6.5 Out-Of-Sample 3-Quarter-Ahead Forecasting Results

Table 8.10: Out-of-sample 3-quarter-ahead forecasting results

Diebold-Li "Yields-Only" Model							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.2905	-0.2820	-0.3167	-0.3641	-0.3890	-0.3789	-0.3932
Std. Dev.	0.9806	0.9876	0.9741	0.9170	0.7919	0.7124	0.6735
RMSE	1.0192	1.0236	1.0208	0.9835	0.8797	0.8046	0.7777
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.6290	-0.5851	-0.5664	-0.5714	-0.5746	-0.5515	-0.4934

Continued on next page

Table 8.10 – *Continued from previous page*

Std. Dev.	0.9114	0.9027	0.8555	0.8050	0.7175	0.6589	0.6245
RMSE	1.1046	1.0730	1.0234	0.9848	0.9171	0.8574	0.7941

FSSM

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.1874	-0.1233	-0.1228	-0.1613	-0.1466	-0.0681	0.0198
Std. Dev.	1.0292	0.9465	0.8677	0.8583	0.8710	0.8614	0.8320
RMSE	1.0424	0.9510	0.8732	0.8702	0.8801	0.8610	0.8292

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0470	0.0527	0.0446	-0.0375	-0.1387	-0.1143	0.0260
Std. Dev.	1.1494	1.1259	1.0606	0.9905	0.8602	0.7601	0.6649
RMSE	1.1462	1.1229	1.0576	0.9875	0.8682	0.7659	0.6629

FSSM^{LS}

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5342	-0.7504	-0.9687	-1.0067	-0.8161	-0.5632	-0.3014
Std. Dev.	1.0574	0.9742	0.9362	0.9428	0.9268	0.8935	0.8242
RMSE	1.1820	1.2274	1.3451	1.3772	1.2325	1.0536	0.8748

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.7325	-0.8862	-1.0725	-1.1212	-0.9929	-0.7666	-0.4251
Std. Dev.	1.2390	1.2009	1.1113	1.0310	0.8960	0.8057	0.7163
RMSE	1.3095	1.3651	1.4242	1.4187	1.2726	1.0671	0.8168

MShock-FSSM

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.1529	-0.0905	-0.0923	-0.1319	-0.1191	-0.0421	0.0446
Std. Dev.	0.9989	0.9198	0.8468	0.8401	0.8566	0.8486	0.8176
RMSE	1.0069	0.9209	0.8488	0.8474	0.8617	0.8465	0.8158

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0320	0.0673	0.0583	-0.0248	-0.1275	-0.1040	0.0354
Std. Dev.	1.0944	1.0777	1.0252	0.9642	0.8460	0.7536	0.6655
RMSE	1.0908	1.0759	1.0231	0.9610	0.8525	0.7580	0.6640

Continued on next page

Table 8.10 – *Continued from previous page*

MShock-FSSM^{LS}							
				US			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5034	-0.7225	-0.9461	-0.9887	-0.8059	-0.5565	-0.2956
Std. Dev.	1.0272	0.9481	0.9176	0.9287	0.9171	0.8845	0.8116
RMSE	1.1405	1.1893	1.3156	1.3541	1.2183	1.0422	0.861
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4108	-0.5961	-0.8274	-0.9101	-0.8343	-0.6339	-0.3180
Std. Dev.	1.1857	1.1777	1.1218	1.0596	0.9436	0.8485	0.7501
RMSE	1.2507	1.3161	1.3906	1.3938	1.2570	1.0567	0.8122

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

Table 8.11: Accuracy changes: Out-of-sample 3-quarter-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	2.28%	-7.09%	-14.46%	-11.52%	0.05%	7.01%	6.62%
Germany	3.77%	4.65%	3.34%	0.27%	-5.33%	-10.67%	-16.52%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	15.97%	19.91%	31.77%	40.03%	40.10%	30.95%	12.49%
Germany	18.55%	27.22%	39.16%	44.06%	38.76%	24.46%	2.86%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-1.21%	-10.03%	-16.85%	-13.84%	-2.05%	5.21%	4.90%
Germany	-1.25%	0.27%	-0.03%	-2.42%	-7.04%	-11.59%	-16.38%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	11.90%	16.19%	28.88%	37.68%	38.49%	29.53%	10.71%
Germany	13.23%	22.66%	35.88%	41.53%	37.06%	23.24%	2.28%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y

Continued on next page

Table 8.11 – *Continued from previous page*

US	13.39%	29.06%	54.04%	58.26%	40.04%	22.37%	5.50%
Germany	14.25%	21.57%	34.66%	43.67%	46.58%	39.33%	23.22%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-3.41%	-3.17%	-2.79%	-2.62%	-2.09%	-1.68%	-1.62%
Germany	-4.83%	-4.19%	-3.26%	-2.68%	-1.81%	-1.03%	0.17%
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	9.41%	25.06%	50.66%	55.61%	38.43%	21.05%	3.84%
Germany	9.12%	17.21%	31.49%	41.14%	44.78%	37.97%	22.52%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	13.27%	29.15%	55.00%	59.79%	41.38%	23.12%	5.54%
Germany	14.66%	22.33%	35.92%	45.04%	47.45%	39.41%	22.32%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes among the models. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

8.6.6 Out-Of-Sample 4-Quarter-Ahead Forecasting Results

Table 8.12: Out-of-sample 4-quarter-ahead forecasting results

Diebold-Li "Yields-Only" Model							
US							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.3866	-0.3811	-0.4204	-0.4705	-0.4959	-0.4857	-0.5014
Std. Dev.	1.2259	1.2501	1.2416	1.1776	1.0193	0.9072	0.8265
RMSE	1.2811	1.3025	1.3065	1.2641	1.1301	1.0261	0.9641
Germany							
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.7942	-0.7582	-0.7448	-0.7497	-0.7481	-0.7185	-0.6517
Std. Dev.	1.0193	1.0077	0.9538	0.8984	0.8016	0.7348	0.6927
RMSE	1.2892	1.2581	1.2074	1.1676	1.0942	1.0258	0.9492
FSSM							
US							

Continued on next page

Table 8.12 – *Continued from previous page*

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.2615	-0.2034	-0.2043	-0.2373	-0.2062	-0.1150	-0.0166
Std. Dev.	1.2153	1.1014	0.9767	0.9439	0.9378	0.9304	0.9027
RMSE	1.2388	1.1161	0.9943	0.9699	0.9569	0.9341	0.8996

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0641	0.0313	0.0191	-0.0656	-0.1702	-0.1472	-0.0071
Std. Dev.	1.3750	1.3330	1.2333	1.1374	0.9730	0.8546	0.7456
RMSE	1.3715	1.3285	1.2289	1.1351	0.9843	0.8641	0.7429

FSSM^{LS}

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.5254	-0.7343	-0.9436	-0.9762	-0.7775	-0.5202	-0.2562
Std. Dev.	1.3280	1.2125	1.1179	1.0914	1.0538	1.0151	0.9362
RMSE	1.4249	1.4148	1.4607	1.4620	1.3071	1.1379	0.9678

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.8155	-0.9654	-1.1421	-1.1810	-1.0385	-0.8028	-0.4506
Std. Dev.	1.4753	1.3969	1.2535	1.1412	0.9792	0.8796	0.7859
RMSE	1.5278	1.5447	1.5512	1.5185	1.3502	1.1393	0.8918

MShock-FSSM

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.2255	-0.1697	-0.1738	-0.2086	-0.1796	-0.0896	0.0077
Std. Dev.	1.1807	1.0741	0.9573	0.9279	0.9243	0.9172	0.8884
RMSE	1.1977	1.0835	0.9694	0.9477	0.9382	0.9182	0.8852

Germany

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.0540	0.0418	0.0297	-0.0552	-0.1605	-0.1381	0.0015
Std. Dev.	1.2999	1.2659	1.1810	1.0972	0.9490	0.8400	0.7388
RMSE	1.2962	1.2619	1.1771	1.0946	0.9590	0.8482	0.7361

MShock-FSSM^{LS}

US

Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.4944	-0.7065	-0.9219	-0.9592	-0.7686	-0.5155	-0.2524

Continued on next page

Table 8.12 – *Continued from previous page*

Std. Dev.	1.2876	1.1793	1.0939	1.0731	1.0406	1.0022	0.9202
RMSE	1.3749	1.3710	1.4275	1.4363	1.2906	1.1237	0.951
				Germany			
Maturity (τ)	6M	1Y	2Y	3Y	5Y	7Y	10Y
Mean	-0.3957	-0.5893	-0.8273	-0.9117	-0.8355	-0.6332	-0.3142
Std. Dev.	1.4064	1.3673	1.2654	1.1779	1.0377	0.9321	0.8268
RMSE	1.4561	1.4842	1.5080	1.4861	1.3293	1.1240	0.8816

Note: The table reports the results of out-of-sample 4-quarter-ahead forecasting using the state-space models previously described. The models are estimated recursively, with yearly re-estimation, from 1999:06 till 2017:01. The forecast errors are defined at $t+h$ as $y_{t+h}(\tau) - \hat{y}_{t+\frac{h}{t}}(\tau)$, where $h = 1, 2, 3, 6, 9, 12$. The table reports the mean, standard deviation and RMSE of the forecast errors.

Table 8.13: Accuracy changes: Out-of-sample 4-quarter-ahead forecasts

ΔRMSE%: FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-3.30%	-14.31%	-23.90%	-23.27%	-15.33%	-8.97%	-6.69%
Germany	6.38%	5.60%	1.78%	-2.78%	-10.04%	-15.76%	-21.73%
ΔRMSE%: FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	11.22%	8.62%	11.80%	15.66%	15.66%	10.90%	0.38%
Germany	18.51%	22.78%	28.47%	30.05%	23.40%	11.06%	-6.05%
ΔRMSE%: MShock-FSSM vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-6.51%	-16.81%	-25.80%	-25.03%	-16.98%	-10.52%	-8.18%
Germany	0.54%	0.30%	-2.51%	-6.25%	-12.36%	-17.31%	-22.45%
ΔRMSE%: MShock-FSSM^{LS} vs Diebold-Li							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	7.32%	5.26%	9.26%	13.62%	14.20%	9.51%	-1.36%
Germany	12.95%	17.97%	24.90%	27.28%	21.49%	9.57%	-7.12%
ΔRMSE%: FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	15.02%	26.76%	46.91%	50.74%	36.60%	21.82%	7.58%
Germany	11.40%	16.27%	26.23%	33.78%	37.17%	31.85%	20.04%
ΔRMSE%: MShock-FSSM vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	-3.32%	-2.92%	-2.50%	-2.29%	-1.95%	-1.70%	-1.60%
Germany	-5.49%	-5.01%	-4.22%	-3.57%	-2.57%	-1.84%	-0.92%

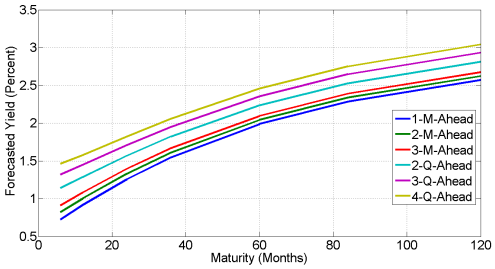
Continued on next page

Table 8.13 – *Continued from previous page*

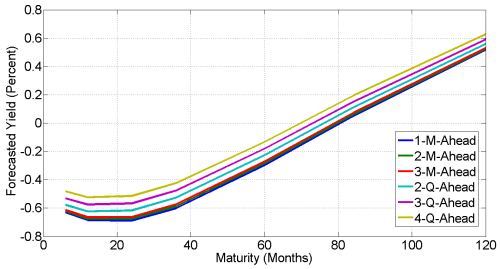
ΔRMSE%: MShock-FSSM^{LS} vs FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	10.99%	22.84%	43.57%	48.09%	34.87%	20.30%	5.71%
Germany	6.17%	11.72%	22.71%	30.92%	35.05%	30.08%	18.67%
ΔRMSE%: MShock-FSSM^{LS} vs MShock-FSSM							
	6M	1Y	2Y	3Y	5Y	7Y	10Y
US	14.80%	26.53%	47.26%	51.56%	37.56%	22.38%	7.43%
Germany	12.34%	17.62%	28.11%	35.77%	38.61%	32.52%	19.77%

Note: The table reports ΔRMSE%, i.e., the percentage changes in the RMSEs among the models as a measure of forecasting accuracy changes. Given two models, i.e., the alternative model, Model A, and the benchmark model, Model B, we calculate ΔRMSE% as $\Delta\text{RMSE}\% = ((\text{RMSE}_{\text{Model A}} - \text{RMSE}_{\text{Model B}})/\text{RMSE}_{\text{Model B}}) \times 100$. A positive (negative) sign denotes a decrease (increase) in forecasting accuracy of the alternative model compared to the benchmark model. In this table, the alternative and the benchmark models are reported in the following format: Model A vs Model B.

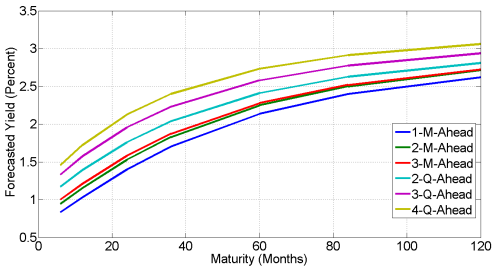
Figure 8.2: Comparison plot of forecasted US and German yield curves



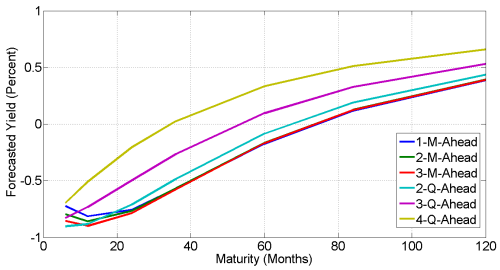
(a) Diebold-Li: US



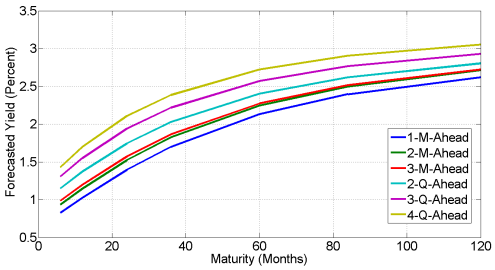
(b) Diebold-Li: DE



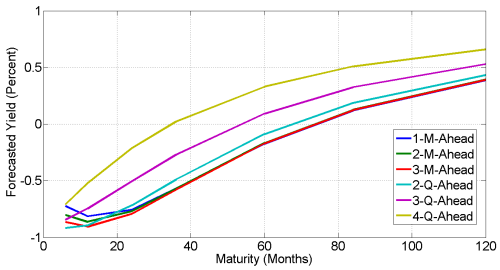
(c) FSSM: US



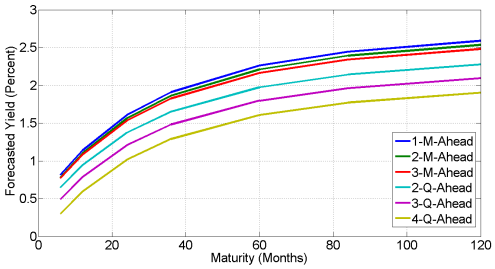
(d) FSSM: DE



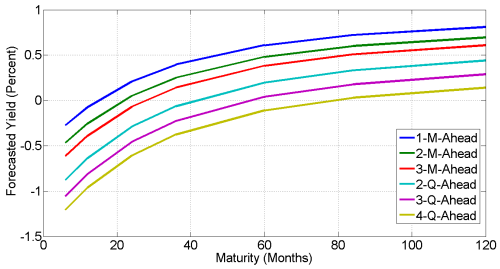
(e) MShock-FSSM: US



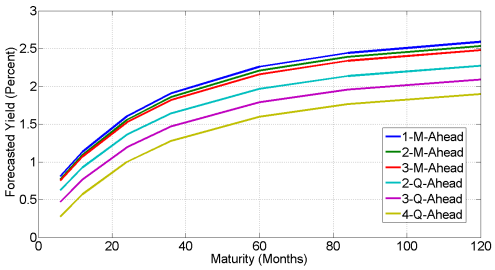
(f) MShock-FSSM: DE



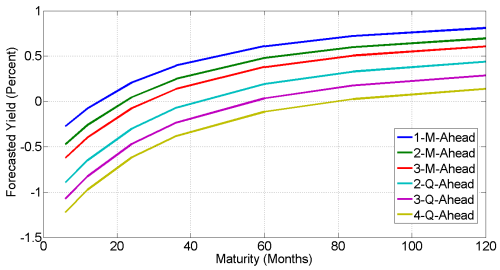
(g) FSSM^{LS}: US



(h) FSSM^{LS}: DE



(i) MShock-FSSM^{LS}: US



(j) MShock-FSSM^{LS}: DE

8.7 Conclusion

In this Chapter, we explored the performance of the FSSM and MShock-FSSM in out-of-sample yield curve forecasting. To this regard, we performed a recursive out-of-sample forecasting exercise with re-estimation of the parameters every 12 months with the Kalman filter and maximum likelihood. At short forecast horizons, the FSSM is able to outperform the state-of-the-art Diebold-Li model only at short-term maturities, for Germany, and at long-term maturities, for the US. Our results provide evidence about an increased forecast accuracy of the FSSM as the forecast horizon increases. At the 3-month-ahead horizon, the FSSM outperforms the Diebold-Li model for all German maturities. At the 4-quarter-ahead horizon, the FSSM outperforms the Diebold-Li for all US maturities and German mid- and long-term maturities. Upon these results, we can conclude that the FSSM has a strong yield curve forecastability at the longer horizons, an appealing feature to fixed income traders, portfolio risk managers, insurance companies, and pension funds.

The MShock-FSSM (i.e., the FSSM, extended with intervention variables accounting for patches of outliers in the measurement equation) surpasses the FSSM in terms of forecasting accuracy. However, the improved accuracy might be due to the inclusion of additional exogenous variables in the MShock-FSSM that could induce a false sense of confidence in the forecasts.

Furthermore, we verified the forecasting power of the curvatures, by developing and forecasting with two additional models, i.e., the FSSM^{LS} and the MShock-FSSM^{LS}. These two models are the versions of the original FSSM and MShock-FSSM without the US and German curvatures, thus, modeling only the levels and slopes. The forecasting results are rather poor at all forecast horizons, thus supporting the idea that, for our sample of yields, the curvatures do have predictive power for the US and German yield curves.

Chapter 9

Conclusions and Perspectives

In today's global capital markets, term structure models for the joint evolution of yield curves of different world regions are of critical importance. The academic literature provides a wide range of term-structure models, however, multiple gaps and shortcomings can be identified. Popular models, such as the Diebold, Rudebusch, and Aruoba, 2006 "Yields-Only" model, rely on simplifying assumptions concerning the dynamic properties of yield data. Persistent, unit-root dynamics, correlation and cointegration structure, lead-lag relationships, presence of outliers and structural breaks are very often disregarded and not accounted for by the data generation processes underlying yield data. Very few term structure models tackle explicitly the problem of out-of-sample yield curve forecasting.

With the development of new data-driven state-space models for forecasting the co-movement of yield curve drivers of different world regions and, from the drivers, the yield curves, this thesis aims at fulfilling the above-mentioned gaps in the literature.

Using actively traded government bond yields for US and Germany, the modeling approach consists in first conducting, in Part I of this thesis, a comprehensive study of the dynamic properties of the US and German yield curve drivers. This study provides evidence about the stationarity of the US and German slopes, nonstationarity of the levels and curvatures, cointegration structure between the levels and curvatures, and existence of Granger causality among all US and German yield curve drivers. The most suitable data generation processes to capture these dynamic properties are found to be a 2D-VAR(5) model, for the US and German slopes, and a 4D-VEC(3) model, for the levels and curvatures.

Further on in our modeling approach, a study of outliers and structural breaks in the dynamics of US and German yield curve drivers is provided in Part II of this thesis. In a univariate setting and adopting the methods of Bai and Perron, 1998 and Perron and Zhou, 2008, tests for the presence of multiple structural breaks of unknown timing are performed to find evidence supporting the existence of breaks in all drivers. The nature of these breaks is investigated in a study of the US Fed and ECB monetary policy predictability. Using money market and policy rates, it is investigated and confirmed that the 2008 Financial Crisis signed a monetary policy regime change and a change in the ability of market participants in predicting monetary policy decisions. These findings provide good reasons to believe that the root causes of structural breaks are linked to a change in monetary policy regimes and increased predictability of Central

Banks. In addition, it is investigated whether the presence of structural breaks is due to variables with predictive power missing in the univariate dynamics of the US and German yield curve drivers. To verify this assumption, we switch to a multivariate state-space setting and develop the FSSM. The novelty of the FSSM is that it is designed to preserve the VAR dynamics for the US and German slopes and the VEC dynamics for the US and German levels and curvatures. Running a backward recursion with the Kalman filter on the estimated FSSM, the model's auxiliary residuals are calculated and tested for the presence of outliers and structural breaks. The results obtained suggest that the structural alterations in the auxiliary residuals of the FSSM resemble of patches of outliers rather than of structural breaks. As a consequence of these results, the FSSM is adjusted to account for the most blatant outliers. The adjustment consists in including intervention (or shock) variables in the measurement equation of the FSSM. This adjusted version of the FSSM is the MShock-FSSM.

In Part III of this thesis, the performance of the FSSM and MShock-FSSM is explored in out-of-sample yield curve forecasting. To this regard, a recursive out-of-sample forecasting exercise is performed with re-estimation of the parameters every 12 months with the Kalman filter and maximum likelihood. US and German term-structure forecasts are produced at both short and long horizons. The forecasting performance of our models is benchmarked to the Diebold-Li "Yields-Only" Model, with the aim of understanding how do models that account for all dynamic properties of yield data perform compared to the state-of-the-art Diebold-Li model. The forecasting results are promising, providing evidence that the FSSM, accounting for all the dynamic properties of the yield data, outperforms the Diebold-Li model. In addition, the predictive power of the curvatures is confirmed by developing and forecasting with the FSSM^{LS} and the MShock-FSSM^{LS}, i.e., the original FSSM and MShock-FSSM for the US and German levels and slopes only.

Following the modeling and forecasting approaches presented in this thesis, future research could be devoted to the development of further data-driven international term-structure models, using the other yield curve drivers in 4.17. More specifically, further state-space models can be developed using as yield curve drivers either the first three principal components of all US and German yields (i.e., $PC^{y}_{1,t}, PC^{y}_{2,t}, PC^{y}_{3,t}$) or the first three principal components of all German-US yield spreads (i.e., $PC^{s}_{1,t}, PC^{s}_{2,t}, PC^{s}_{3,t}$) or the first five principal components of the so-called "refined" US and German yield curve factors (i.e., $PC^{f}_{1,t}, PC^{f}_{2,t}, PC^{f}_{3,t}, PC^{f}_{4,t}, PC^{f}_{5,t}$).

Future studies could fruitfully explore the performance of the FSSM (and its related versions) by extending the number of countries in the yield dataset. To this regard, an important point should be considered. In estimating the FSSM (and its related versions), the normality condition was assumed to hold for the measurement and transition shocks. Gaussian maximum likelihood estimates were then obtained via application of the Kalman filter to our state-space models. For future models for the co-movement of the yield curves of a large set of countries, maximum likelihood might be particularly difficult to implement because of the large number of parameters to be estimated. In these settings, a Bayesian approach might be preferable (Kim and Nelson, 1999; West and Harrison, 2006; Greenberg, 2012). In the tradition of recent advances in Bayesian estimation of large-scale dynamic factor models, Markov Chain Monte Carlo (MCMC) methods – effectively just Gibbs sampling – might be efficiently employed to perform a posterior analysis

of the models that condition on a large information set (Kose, Otrok, and Whiteman, 2003; Bernanke, Boivin, and Eliasz, 2005; Diebold, Li, and Yue, 2008). The Bayesian approach is robustly grounded as suitable for large specifications (Sims and Zha, 1998; Bańbura, Giannone, and Reichlin, 2010), over-parametrization (Koop and Korobilis, 2010; De Mol, Giannone, and Reichlin, 2008), forecasting accuracy (Litterman, 1986; Koop and Potter, 2004; Koop, 2013) and structural analysis.

Since in this thesis we provide empirical evidence of structural alterations in the behavior of the US and German yield curve drivers, an interesting topic for future research is to examine alternative ways of accounting for outliers and/or structural breaks in the yield data. These alternatives might consider the inclusion of key policy rates in the model, as variables explaining the stance of the monetary policy (Roley and Sellon, 1995; Krueger and Kuttner, 1996; Muller and Zelmer, 1999; Haldane and Read, 2000; Ellingsen and Soderstrom, 2001; Kuttner, 2001; Piazzesi, 2001; Brand, Buncic, and Turunen, 2010). Macroeconomic variables should also be investigated for their potential predictive power for the yield curves (Ang and Piazzesi, 2003; Diebold, Piazzesi, and Rudebusch, 2005; Dewachter and Lyrio, 2006; Diebold, Rudebusch, and Aruoba, 2006; Hördahl, Tristani, and Vestin, 2006; Evans and Marshall, 2007; Rudebusch and Wu, 2008). Non-linear models, such as Markov switching latent variable models might improve the forecasting accuracy of the yield curves. Estimation and forecasting studies of Markov-switching versions of the dynamic Nelson-Siegel model show that these versions outperform, from a forecasting perspective, the single-regime Nelson-Siegel model, while remaining parsimonious, relatively easy to estimate and sufficiently flexible to match the changing shapes of the yield curves over time (Xiang and Zhu, 2013; Hevia et al., 2015; Zhu and Rahman, 2015; Levant and Ma, 2017).

A further desirable goal for future research is to explore the capabilities of the FSSM (and its related versions) in economic, financial, and risk management applications. The economic value of the forecasted yields of the FSSM can be assessed in fixed-income portfolio optimization using the mean-variance approach of Markovitz, 1952 (Leibowitz and Henriksson, 1988; Zenios et al., 1998; Bertocchi, Moriggia, and Dupačová, 2000; Caldeira, Moura, and Santos, 2016; Schnorrenberger, 2017). Since the FSSM captures the simultaneous evolution of yield curves of multiple currency areas, the model ensures that expected return distributions are calculated in a manner consistent with historical correlations across multiple currencies (Mulvey and Zenios, 1994). Furthermore, the shape, location, and tail risk measures of the return distributions can be estimated and employed as input to portfolio optimization problems (Black and Litterman, 1992; McNeil and Frey, 2000; Bolder, 2015). From a risk management perspective, the FSSM can be efficiently employed in stress testing of international bond portfolios, in order to assess how effects in the yield curve of one currency area propagate to the yield curves of the other modeled currency areas (Golub and Tilman, 2000; Sorge, 2004; Drehmann, Sorensen, and Stringa, 2010).

Appendix A

Supplementary Material: Part I

A.1 Descriptive Statistics for Yield Data

Table A.1: Yield data, in levels: Sample central moments.

Maturities	$y_t(6M)$	$y_t(1Y)$	$y_t(2Y)$	$y_t(3Y)$	$y_t(5Y)$	$y_t(7Y)$	$y_t(10Y)$
US							
Mean	1.9078	2.0111	2.2767	2.5115	2.9692	3.3383	3.6385
Variance	4.0696	3.8610	3.6081	3.2280	2.5655	2.1816	1.7195
Skewness	0.8330	0.7726	0.7159	0.6291	0.4475	0.3629	0.1713
Kurtosis	2.2216	2.1624	2.1935	2.1572	2.0996	2.1361	1.9907
Min	0.0400	0.1000	0.2100	0.3300	0.6200	0.9800	1.5000
Max	6.3900	6.3300	6.8100	6.7700	6.6900	6.7200	6.6600
Germany							
Mean	1.6859	1.7543	1.9052	2.0748	2.4207	2.7341	3.1099
Variance	3.1246	3.1942	3.2772	3.3171	3.2526	3.0856	2.8271
Skewness	0.2152	0.1529	0.0300	-0.0770	-0.2381	-0.3535	-0.4762
Kurtosis	1.6824	1.6451	1.6196	1.6227	1.6764	1.7614	1.9117
Min	-0.9200	-0.9200	-0.9200	-0.8600	-0.6300	-0.4700	-0.2100
Max	5.1100	5.1700	5.2600	5.3100	5.3400	5.5200	5.7200
DE-US Spreads							
Mean	-0.2219	-0.2568	-0.3715	-0.4367	-0.5485	-0.6041	-0.5286
Variance	1.3866	1.2654	1.1492	1.0294	0.8272	0.7266	0.5745
Skewness	0.0043	-0.0291	-0.0791	-0.0704	-0.1750	-0.2313	-0.3358
Kurtosis	2.2073	2.1685	2.1058	2.0375	2.0468	2.0797	2.2206
Min	-2.3400	-2.4600	-2.5700	-2.5000	-2.4900	-2.4700	-2.2700
Max	2.5100	2.2600	1.9400	1.7600	1.2100	1.1100	1.1300

Table A.2: Yield data, in first differences: Sample central moments.

Maturities	$\Delta y_t(6M)$	$\Delta y_t(1Y)$	$\Delta y_t(2Y)$	$\Delta y_t(3Y)$	$\Delta y_t(5Y)$	$\Delta y_t(7Y)$	$\Delta y_t(10Y)$
US							
Mean	-0.0126	-0.0119	-0.0114	-0.0108	-0.0097	-0.0100	-0.0094
Variance	0.0318	0.0312	0.0391	0.0459	0.0506	0.0503	0.0472
Skewness	-1.5644	-1.3989	-0.4344	-0.0825	0.0638	0.0232	-0.3059
Kurtosis	7.1486	6.9455	4.3986	3.7009	3.5490	4.1483	5.6649
Min	-0.7700	-0.7900	-0.6400	-0.6600	-0.7700	-0.9300	-1.1100
Max	0.3400	0.3600	0.5400	0.5900	0.6000	0.6400	0.6500
Germany							
Mean	-0.0156	-0.0154	-0.0150	-0.0147	-0.0142	-0.0139	-0.0139
Variance	0.0370	0.0335	0.0413	0.0429	0.0408	0.0381	0.0368
Skewness	-4.5979	-1.2077	-0.5124	-0.2093	-0.0781	-0.1172	-0.2205
Kurtosis	46.0203	8.5485	5.2305	3.8361	2.9511	2.8850	3.0171
Min	-1.9400	-1.0400	-0.9700	-0.8000	-0.5700	-0.5600	-0.6500
Max	0.4300	0.4800	0.5800	0.5700	0.5600	0.5300	0.4300
DE-US Spreads							
Mean	-0.0030	-0.0036	-0.0037	-0.0039	-0.0044	-0.0039	-0.0045
Variance	0.0421	0.0322	0.0349	0.0392	0.0429	0.0411	0.0388
Skewness	-1.3871	0.6485	0.5912	0.4883	0.0745	-0.0120	0.1736
Kurtosis	20.1834	6.6052	4.0700	3.6563	3.3056	3.6077	4.4707
Min	-1.6100	-0.7700	-0.5000	-0.4900	-0.7000	-0.7200	-0.6200
Max	0.7800	0.7400	0.7000	0.7400	0.5300	0.6800	0.8600

Table A.3: Yield data, in levels: Sample autocorrelations.

Maturities	$y_t(6M)$	$y_t(1Y)$	$y_t(2Y)$	$y_t(3Y)$	$y_t(5Y)$	$y_t(7Y)$	$y_t(10Y)$
US							
1	0.9924	0.9924	0.9912	0.9898	0.9876	0.9857	0.9834
2	0.9801	0.9798	0.9773	0.9741	0.9694	0.9653	0.9603
3	0.9645	0.9644	0.9608	0.9561	0.9501	0.9445	0.9384
4	0.9464	0.9469	0.9430	0.9372	0.9299	0.9231	0.9159
5	0.9260	0.9276	0.9237	0.9174	0.9087	0.9004	0.8924
6	0.9030	0.9066	0.9037	0.8982	0.8891	0.8794	0.8715
Germany							
1	0.9892	0.9899	0.9890	0.9891	0.9894	0.9893	0.9885
2	0.9752	0.9754	0.9741	0.9746	0.9759	0.9761	0.9749

Continued on next page

Table A.3 – *Continued from previous page*

Maturities	$y_t(6M)$	$y_t(1Y)$	$y_t(2Y)$	$y_t(3Y)$	$y_t(5Y)$	$y_t(7Y)$	$y_t(10Y)$
3	0.9589	0.9581	0.9572	0.9588	0.9619	0.9631	0.9621
4	0.9406	0.9387	0.9381	0.9409	0.9459	0.9483	0.9478
5	0.9210	0.9194	0.9200	0.9236	0.9298	0.9328	0.9324
6	0.9012	0.9002	0.9023	0.9069	0.9144	0.9179	0.9173
DE-US Spreads							
1	0.9753	0.9755	0.9721	0.9690	0.9639	0.9631	0.9588
2	0.9437	0.9434	0.9426	0.9400	0.9361	0.9372	0.9320
3	0.9095	0.9065	0.9082	0.9077	0.9089	0.9144	0.9115
4	0.8673	0.8617	0.8652	0.8674	0.8750	0.8858	0.8871
5	0.8186	0.8139	0.8207	0.8262	0.8414	0.8569	0.8628
6	0.7651	0.7639	0.7755	0.7854	0.8083	0.8270	0.8378

Table A.4: Yield data, in first differences: Sample autocorrelations.

Maturities	$\Delta y_t(6M)$	$\Delta y_t(1Y)$	$\Delta y_t(2Y)$	$\Delta y_t(3Y)$	$\Delta y_t(5Y)$	$\Delta y_t(7Y)$	$\Delta y_t(10Y)$
US							
1	0.5662	0.5474	0.3984	0.3254	0.2355	0.2089	0.1886
2	0.3812	0.3022	0.1838	0.1080	0.0175	-0.0262	-0.0735
3	0.3568	0.2851	0.1270	0.0694	0.0496	0.0372	0.0249
4	0.2237	0.1620	0.0610	-0.0031	-0.0158	-0.0129	-0.0200
5	0.2024	0.1161	-0.0388	-0.1241	-0.1385	-0.1253	-0.1294
6	0.2622	0.1770	0.0363	-0.0640	-0.1366	-0.1531	-0.1290
Germany							
1	0.2169	0.3746	0.2584	0.2022	0.1440	0.1053	0.0651
2	0.2116	0.2739	0.1580	0.0979	0.0103	-0.0413	-0.0892
3	0.2004	0.2099	0.1654	0.1573	0.1593	0.1460	0.1294
4	0.1030	-0.0098	-0.0781	-0.0633	-0.0020	0.0433	0.0570
5	-0.0047	-0.0507	-0.0894	-0.1062	-0.1263	-0.1263	-0.0939
6	0.0484	-0.0302	-0.0488	-0.0574	-0.0707	-0.0803	-0.0886
DE-US Spreads							
1	0.2186	0.3015	0.0754	-0.0425	-0.1602	-0.1987	-0.2146
2	0.0908	0.1741	0.1179	0.0454	-0.0368	-0.0768	-0.1038
3	0.2600	0.3165	0.2854	0.2112	0.1190	0.0906	0.0473
4	0.2143	0.1531	0.0796	0.0506	0.0234	0.0347	0.0285
5	0.1432	0.0948	0.0438	0.0082	-0.0016	0.0104	-0.0059
6	0.0825	0.1117	0.0834	0.0052	-0.0949	-0.1045	-0.1061

A.2 Descriptive Statistics of Estimated Country Factors

Table A.5: (Nelson-Siegel) estimated country factors: Sample central moments.

Factor	Mean	Variance	Skewness	Kurtosis	Min	Max
$l_{US,t}$	4.3313	1.4016	-0.3154	1.9558	1.9205	6.6338
$s_{US,t}$	-2.3803	2.5421	0.2800	2.1434	-5.0621	0.9209
$c_{US,t}$	-2.8512	5.1378	0.1076	2.0668	-7.0798	2.2216
$l_{DE,t}$	3.6872	2.8483	-0.6346	2.1463	0.0151	6.1977
$s_{DE,t}$	-1.8389	1.2432	-0.2877	2.3538	-4.4612	0.3155
$c_{DE,t}$	-2.9330	2.7264	0.3206	2.1497	-6.2160	0.6701
$\Delta l_{US,t}$	-0.0088	0.0683	-0.6305	8.5688	-1.5468	0.8264
$\Delta s_{US,t}$	-0.0044	0.1070	-0.6196	4.0530	-1.0545	0.9991
$\Delta c_{US,t}$	-0.0004	0.3113	0.3720	5.9531	-2.1065	2.5018
$\Delta l_{DE,t}$	-0.0132	0.0531	-0.0549	3.6773	-0.8422	0.6596
$\Delta s_{DE,t}$	-0.0026	0.1005	-2.1943	17.1472	-2.3958	0.8178
$\Delta c_{DE,t}$	-0.0013	0.5120	-1.0826	12.2400	-4.8823	2.1008

Table A.6: (Nelson-Siegel) estimated country factors: Sample autocorrelations.

Lags	l_{it}	s_{it}	c_{it}	Δl_{it}	Δs_{it}	Δc_{it}
$i = US$						
1	0.9717	0.9746	0.9639	0.1639	0.3138	0.2561
2	0.9342	0.9367	0.9119	-0.1408	0.0690	-0.0443
$i = Germany$						
1	0.9861	0.9590	0.9064	-0.0461	0.0501	-0.1449
2	0.9714	0.9149	0.8401	-0.1312	-0.0404	0.0320

A.3 Nonstationarity Test Results

Table A.7: ADF test results: US yields, in levels ($y_{US,t}(\tau)$).

TS / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	10	12	9	9	9	6	5
ADF Statistic	-2.4520	-2.6908	-2.1488	-1.9942	-1.8643	-1.6229	-1.5190

Continued on next page

Table A.7 – *Continued from previous page*

TS	6M	1Y	2Y	3Y	5Y	7Y	10Y
p-value	0.0142	0.0076	0.0308	0.0444	0.0595	0.0987	0.1209
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	0	0	0
BIC	-214.42	-182.11	-92.49	-42.35	-8.49	-19.87	-36.37
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	10	12	9	9	10	1	1
ADF Statistic	-2.9966	-3.2311	-2.6140	-2.3493	-2.1785	-1.6438	-1.6117
p-value	0.0369	0.0196	0.0917	0.1578	0.2180	0.4535	0.4677
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	0	0	0	0	0
BIC	-212.26	-180.49	-89.74	-39.14	-1.05	-33.24	-45.39
Trend-stationary model variant (TS)							
# Lagged Diff.	10	9	11	9	10	10	10
ADF Statistic	-3.2369	-3.3611	-2.9718	-2.0711	-2.1889	-2.4312	-2.8187
p-value	0.0803	0.0597	0.1432	0.5565	0.4987	0.3799	0.1926
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	0	0	0	0
BIC	-209.68	-189.57	-77.96	-34.49	2.68	1.48	-9.24

Table A.8: ADF test results: US yields, in first differences ($\Delta y_{US,t}(\tau)$).

TS / Res	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	9	11	8	8	8	5	4
ADF Statistic	-2.8814	-3.0779	-3.2795	-3.8383	-4.6681	-7.2939	-7.6513
p-value	0.0045	0.0029	0.0013	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-213.61	-179.98	-93.13	-43.66	-10.31	-22.60	-39.44
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	9	11	8	8	8	5	4
ADF Statistic	-2.9131	-3.1465	-3.3322	-3.9103	-4.7718	-7.3720	-7.7217
p-value	0.0455	0.0248	0.0147	0.0031	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-208.40	-175.04	-88.10	-38.84	-5.91	-18.32	-35.11
Trend-stationary model variant (TS)							

Continued on next page

Table A.8 – *Continued from previous page*

TS / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
# Lagged Diff.	9	11	8	8	8	5	4
ADF Statistic	-3.0526	-3.3949	-3.5363	-4.1006	-4.9137	-7.4124	-7.7172
p-value	0.1211	0.0549	0.0383	0.0077	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	1	1	1	1	1
BIC	-204.26	-172.17	-84.68	-35.44	-2.28	-13.71	-30.06

Table A.9: ADF test results: German yields, in levels ($y_{DE,t}(\tau)$).

Time Series / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	3	8	4	4	5	5	3
ADF Statistic	-1.2109	-1.3412	-1.1426	-1.1415	-1.3821	-1.4160	-1.1526
p-value	0.2076	0.1666	0.2326	0.2330	0.1551	0.1460	0.2289
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	0	0	0	0
BIC	-105.81	-127.04	-80.64	-63.15	-73.90	-85.62	-93.86
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	3	8	4	4	5	5	3
ADF Statistic	-1.1380	-1.1168	-0.8354	-0.8232	-0.7131	-0.6379	-0.5396
p-value	0.6763	0.6856	0.8062	0.8097	0.8395	0.8577	0.8792
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	0	0	0	0
BIC	-100.53	-121.67	-75.21	-57.72	-68.47	-80.20	-88.43
Trend-stationary model variant (TS)							
# Lagged Diff.	9	9	4	3	3	3	3
ADF Statistic	-3.3486	-3.1215	-3.2642	-3.9243	-4.0741	-3.9883	-3.7323
p-value	0.0616	0.1044	0.0753	0.0129	0.0083	0.0105	0.0224
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	1	1	1	1
BIC	-75.93	-124.58	-80.16	-69.35	-78.95	-92.06	-97.12

Table A.10: ADF test results: German yields, in first differences ($\Delta y_{DE,t}(\tau)$).

Time Series / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	2	7	3	3	4	4	2
ADF Statistic	-6.0398	-4.4005	-6.9169	-6.8200	-6.8799	-6.7868	-7.5901
p-value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-109.76	-130.61	-84.74	-67.25	-77.38	-89.00	-97.95
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	2	3	3	3	4	4	2
ADF Statistic	-6.0543	-6.4875	-6.9544	-6.8600	-6.9873	-6.9130	-7.6580
p-value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-104.65	-146.48	-79.93	-62.47	-73.39	-85.22	-93.56
Trend-stationary model variant (TS)							
# Lagged Diff.	2	3	3	3	4	4	2
ADF Statistic	-6.0691	-6.5021	-6.9690	-6.8717	-6.9555	-6.8828	-7.6832
p-value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-99.52	-141.37	-74.86	-57.37	-68.06	-79.93	-88.68

Table A.11: ADF test results: German-US yield spreads, in levels ($s_{DE-US,t}(\tau)$).

Time Series / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	9	9	9	9	10	10	10
ADF Statistic	-3.0888	-2.9944	-2.3481	-1.8360	-1.3322	-1.1772	-0.8896
p-value	0.0028	0.0036	0.0187	0.0633	0.1692	0.2199	0.3252
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	0	0	0	0
BIC	-68.70	-137.39	-97.21	-60.96	-35.43	-45.69	-57.18
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	9	9	9	9	10	10	10
ADF Statistic	-3.1310	-3.0681	-2.5253	-2.0828	-1.7533	-1.6824	-1.3818

Continued on next page

Table A.11 – *Continued from previous page*

Time Series / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
p-value	0.0259	0.0306	0.1114	0.2601	0.4052	0.4364	0.5688
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	0	0	0	0	0
BIC	-63.58	-132.47	-92.73	-56.61	-31.49	-41.90	-53.13
Trend-stationary model variant (TS)							
# Lagged Diff.	9	9	9	9	10	10	10
ADF Statistic	-3.2689	-3.2748	-2.8176	-2.5077	-2.3743	-2.4042	-2.2488
p-value	0.0745	0.0735	0.1930	0.3423	0.4078	0.3931	0.4694
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	0	0	0	0
BIC	-60.23	-129.78	-90.65	-54.95	-30.10	-40.71	-52.01

Table A.12: ADF test results: German-US yield spreads, in first differences ($\Delta s_{DE-US,t}(\tau)$).

Time Series / Results	6M	1Y	2Y	3Y	5Y	7Y	10Y
Autoregressive model variant (AR)							
# Lagged Diff.	12	12	12	8	8	9	9
ADF Statistic	-3.9261	-3.8772	-3.6881	-3.5230	-3.9125	-3.3795	-3.6306
p-value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-47.48	-114.34	-75.03	-62.89	-43.47	-49.67	-61.78
Autoregressive model with drift variant (ARD)							
# Lagged Diff.	12	12	12	8	8	9	8
ADF Statistic	-3.9160	-3.8675	-3.6802	-3.5199	-3.9183	-3.3874	-4.2613
p-value	0.0030	0.0034	0.0051	0.0087	0.0030	0.0127	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-42.04	-108.91	-69.61	-57.53	-38.18	-44.36	-60.80
Trend-stationary model variant (TS)							
# Lagged Diff.	12	12	12	8	8	9	8
ADF Statistic	-4.1921	-4.2135	-4.1092	-3.7545	-4.1331	-3.5400	-4.4196
p-value	0.0056	0.0051	0.0075	0.0211	0.0070	0.0379	0.0033
α	0.05	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1	1
BIC	-38.94	-106.37	-67.66	-53.84	-34.53	-40.10	-56.80

Table A.13: ADF test results: (Nelson-Siegel) estimated country factors, in levels.

Time Series / Results	$l_{US,t}$	$s_{US,t}$	$c_{US,t}$	$l_{DE,t}$	$s_{DE,t}$	$c_{DE,t}$
Autoregressive model variant (AR)						
# Lagged Diff.	2	8	5	11	4	10
ADF Statistic	-1.0002	-1.5363	-1.0401	-1.7685	-1.6363	-0.6863
p-value	0.2847	0.1172	0.2701	0.0732	0.0961	0.3996
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	0	0	0	0	0
BIC	36.70	150.15	382.91	22.62	131.69	511.52
Autoregressive model with drift variant (ARD)						
# Lagged Diff.	3	8	8	11	4	10
ADF Statistic	-1.6979	-2.9590	-2.2667	-0.7620	-3.0985	-2.1890
p-value	0.4297	0.0405	0.1840	0.8269	0.0282	0.2134
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	0	1	0	0	1	0
BIC	44.14	149.05	390.62	28.06	130.18	512.36
Trend-stationary model variant (TS)						
# Lagged Diff.	3	8	5	11	4	10
ADF Statistic	-4.6932	-2.9292	-1.5547	-2.2313	-3.0904	-2.2931
p-value	0.0010	0.1558	0.8074	0.4779	0.1119	0.4476
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	0	0	0	0
BIC	30.90	154.46	390.47	28.86	135.57	517.22

Table A.14: ADF test results: (Nelson-Siegel) estimated country factors, in first differences.

Time Series / Results	$\Delta l_{US,t}$	$\Delta s_{US,t}$	$\Delta c_{US,t}$	$\Delta l_{DE,t}$	$\Delta s_{DE,t}$	$\Delta c_{DE,t}$
Autoregressive model variant (AR)						
# Lagged Diff.	1	7	4	10	12	9
ADF Statistic	-11.5489	-3.9499	-8.0532	-4.4349	-4.3254	-6.5067
p-value	0.0010	0.0010	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1
BIC	32.28	147.16	378.59	20.47	167.17	506.58
Autoregressive model with drift variant (ARD)						

Continued on next page

Table A.14 – *Continued from previous page*

Time Series / Results	$\Delta l_{US,t}$	$\Delta s_{US,t}$	$\Delta c_{US,t}$	$\Delta l_{DE,t}$	$\Delta s_{DE,t}$	$\Delta c_{DE,t}$
# Lagged Diff.	1	7	4	2	12	9
ADF Statistic	-11.5506	-3.9399	-8.0368	-8.2982	-4.3148	-6.4968
p-value	0.0010	0.0028	0.0010	0.0010	0.0010	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1
BIC	37.31	152.59	383.99	-11.21	172.59	511.93
Trend-stationary model variant (TS)						
# Lagged Diff.	1	7	4	2	12	9
ADF Statistic	-11.5265	-3.9516	-8.1635	-8.3125	-4.3067	-6.4813
p-value	0.0010	0.0119	0.0010	0.0010	0.0042	0.0010
α	0.05	0.05	0.05	0.05	0.05	0.05
decision	1	1	1	1	1	1
BIC	42.68	157.87	387.52	-6.25	178.00	517.29

Appendix B

Supplementary Material: Part II

B.1 Testing Jointly for Structural Change in the Regression Coefficients and Error Variance: Perron-Zhou (2008)

Perron and Zhou, 2008 propose a set of 10 different testing problems to test jointly for structural change in the regression coefficients and error variance. Assuming normally distributed and serially uncorrelated errors, *quasi-likelihood ratio tests* are considered for the following 4 testing problems:

- (TP-1) $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, n = n_a\}$, this is the testing problem where one specifies no change in the regression coefficients ($m = q = 0$) but tests for a given number n_a of changes in the variance of the errors.
- (TP-2) $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$, this is a testing problem where there are m_a breaks in the regression coefficients under both the null and the alternative hypotheses so that the test boils down to assessing whether there are 0 or n_a breaks in the variance.
- (TP-3) $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a\}$, this is the testing problem where there are n_a breaks in the variance under both the null and the alternative hypotheses so that the test boils down to assessing whether there are 0 or m_a breaks in the regression coefficients.
- (TP-4) $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = m_a, n = n_a\}$, this is the testing problem where the null hypothesis specifies no break in either coefficients or variance and the alternative hypothesis specifies m_a breaks in coefficients and n_a breaks in the variance of the errors.

In these testing problems, m_a and n_a are some positive numbers, selected a priori.

Using the limit distributions of the quasi-likelihood ratio tests of testing problems TP-1 to TP-4, Perron and Zhou propose modified tests with asymptotic distributions free of nuisance parameters. These modified tests can be used to test problems in which the alternatives specify

some unknown number of breaks, up to some maximum. These additional testing problems are the following:

- (TP-5) $H_0 : \{m = n = 0\}$ versus $H_1 : \{m = 0, 1 \leq n \leq N\}$.
- (TP-6) $H_0 : \{m = m_a, n = 0\}$ versus $H_1 : \{m = m_a, 1 \leq n \leq N\}$.
- (TP-7) $H_0 : \{m = 0, n = n_a\}$ versus $H_1 : \{1 \leq m \leq M, n = n_a\}$.
- (TP-8) $H_0 : \{m = n = 0\}$ versus $H_1 : \{1 \leq m \leq M, 1 \leq n \leq N\}$.

Finally, having a model with a particular number of breaks, the following two testing problems can be used to assess the adequacy of the model by looking at whether too few coefficient breaks are included:

- (TP-9) $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a + 1, n = n_a\}$.
- (TP-10) $H_0 : \{m = m_a, n = n_a\}$ versus $H_1 : \{m = m_a, n = n_a + 1\}$.

B.2 Empirical Results

B.2.1 First Suggestions of Structural Breaks

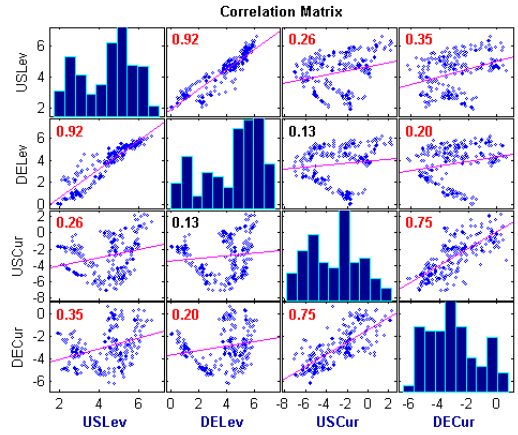
B.2.2 Model Assumptions for Chow Test

For both the US and German slopes, the Chow test model assumptions are satisfied. The scatter plot of residual vs. lagged residual (Figure B.2a and Figure B.2c, for the US and German slopes, respectively) and residual vs. case order (Figure B.2b and Figure B.2d, for the US and German slopes, respectively) do not show any special pattern that would indicate the presence of heteroskedasticity.

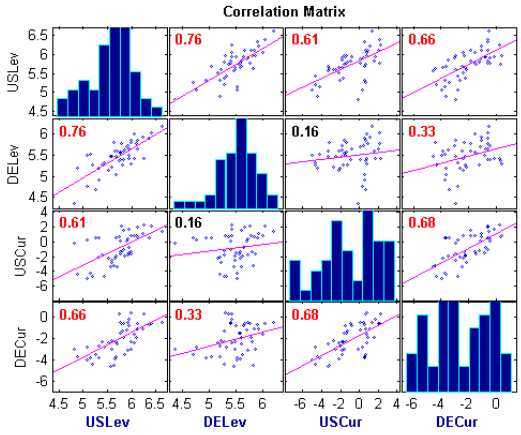
Indeed, the Engle's ARCH test at the 5% level of significance (Table B.1) favors the null hypothesis of no ARCH effects in the residual series of both US slope (p-value = 0.9271) and German slope (p-value = 0.5082). The KS test (Table B.1) suggests that the innovations are Gaussian for both the US slope (p-value = 0.1799) and German slope (p-value = 0.0505).

For the US and German level equations, the Chow test model assumptions are fully satisfied. The scatter plots of residual vs. lagged residual (Figure B.3e and Figure B.3c, for the US and German levels, respectively) and the case order plot (Figure B.3b and Figure B.3d, for the US and German levels, respectively) are quite erratic and do not form any special pattern that would suggest the presence of autocorrelation and/or heteroskedasticity. Engle's ARCH test at 5% level of significance indicates failure to reject the no ARCH effects null hypothesis in the residual series of both US and German levels (p-values equal to 0.9681 and 0.0822, respectively). The KS test also suggests to not reject the null hypothesis that the innovations of the original models for the US and German levels (p-values equal to 0.7787 and 0.8953, respectively) are Gaussian. These results are reported in Table B.1.

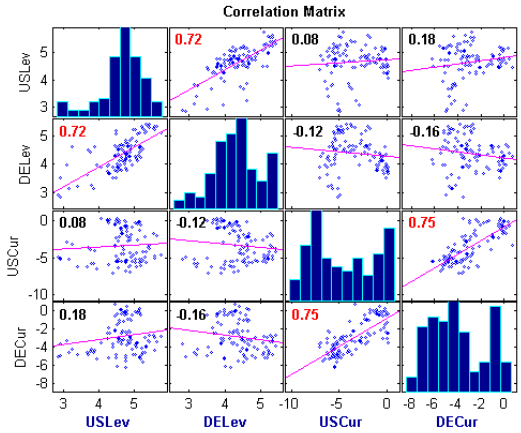
Figure B.1: Split-sample correlation analysis: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system.



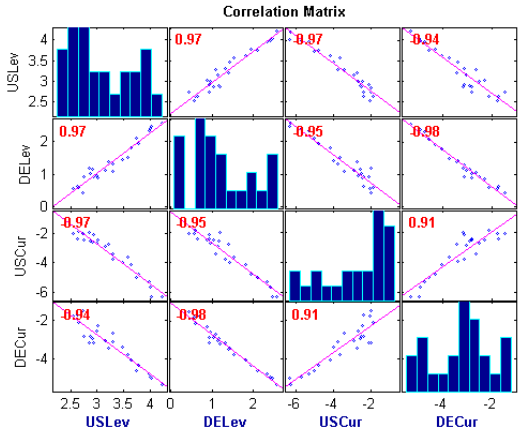
(a) Full sample, [1999:01-2018:01]



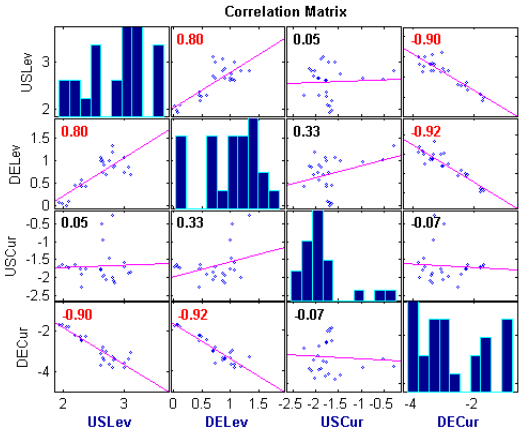
(b) [1999:01-2002:12]



(c) [2003:01-2011:12]

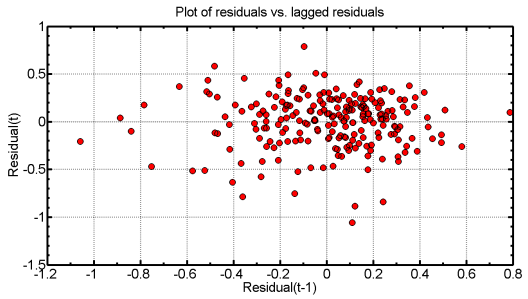


(d) [2004:01-2015:12]

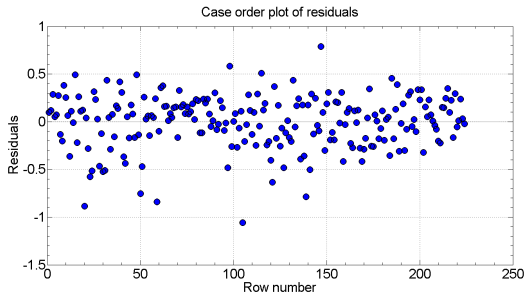


(e) [2016:01-2018:01]

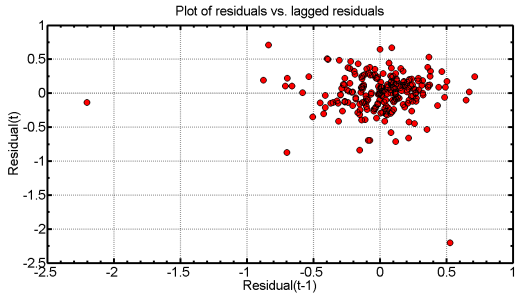
Figure B.2: Histogram plots of residuals: $s_{US,t}/s_{DE,t}$ system, [1999:01 -2018:01].



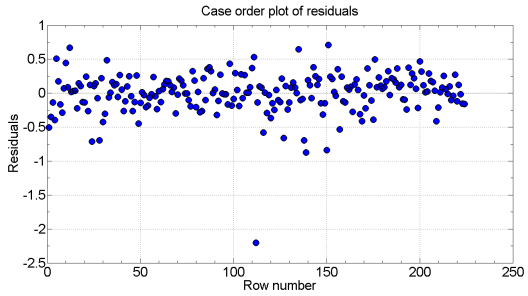
(a) $s_{US,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)



(b) $s_{US,t}$: Res. vs case order

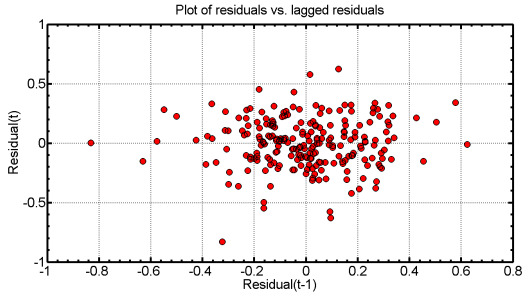


(c) $s_{DE,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)

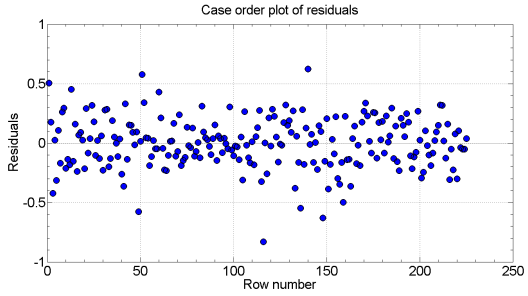


(d) $s_{DE,t}$: Res. vs case order

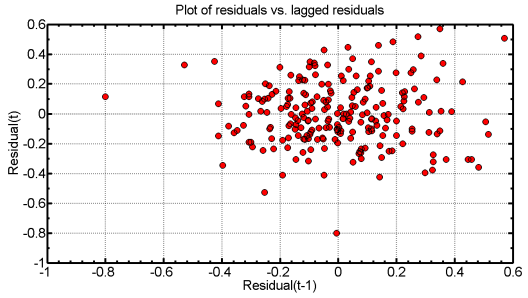
Figure B.3: Histogram plots of residuals: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ system, [1999:01 -2018:01].



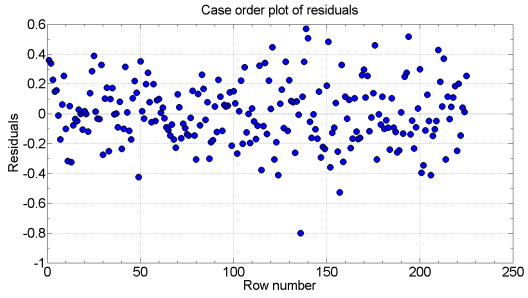
(a) $l_{US,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)



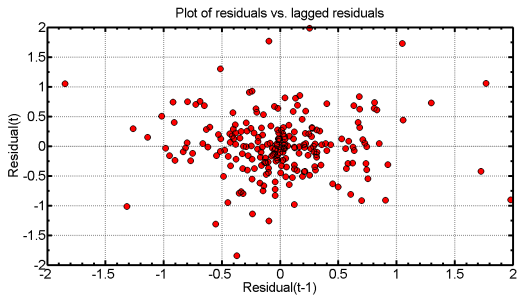
(b) $l_{US,t}$: Res. vs case order



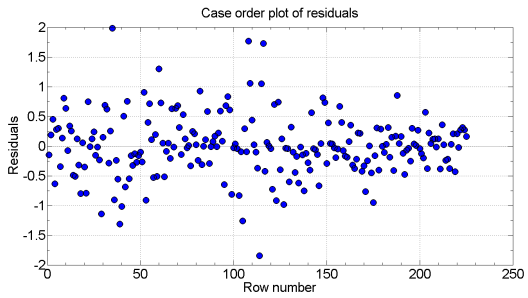
(c) $l_{DE,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)



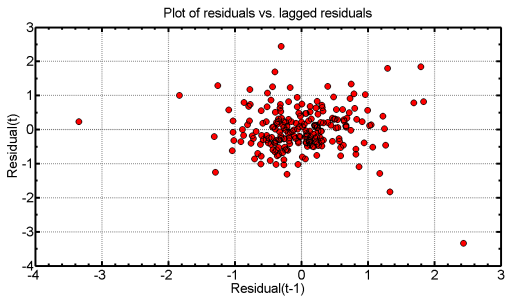
(d) $l_{DE,t}$: Res. vs case order



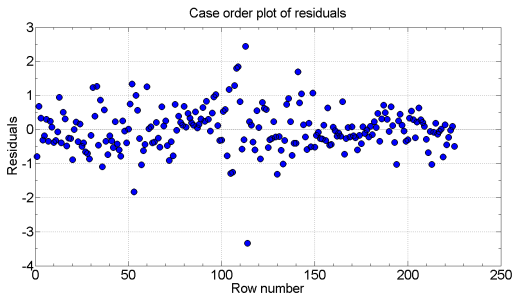
(e) $c_{US,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)



(f) $c_{US,t}$: Res. vs case order



(g) $c_{DE,t}$: Res. vs lagged res. ($r(t)$ versus $r(t-1)$)



(h) $c_{DE,t}$: Res. vs case order

Table B.1: Model assumptions for Chow Test

System	$s_{US,t}/s_{DE,t}$	$s_{US,t}/s_{DE,t}$	$l_{US,t}/l_{DE,t}/$ $c_{US,t}/c_{DE,t}$	$l_{US,t}/l_{DE,t}/$ $c_{US,t}/c_{DE,t}$	$l_{US,t}/l_{DE,t}/$ $c_{US,t}/c_{DE,t}$	$l_{US,t}/l_{DE,t}/$ $c_{US,t}/c_{DE,t}$
Model	2D-VAR(5)	2D-VAR(5)	4D-VAR(4)	4D-VAR(4)	4D-VAR(4)	4D-VAR(4)
Equation	$s_{US,t}$	$s_{DE,t}$	$l_{US,t}$	$l_{DE,t}$	$c_{US,t}$	$c_{DE,t}$
Residuals Histogram, 'caseorder'	Random, no trend.	Random, no trend.	Random, no trend.	Random, no trend.	Random, no trend.	Random, no trend.
Residuals Histogram, 'lagged'	Random, no special pattern	Random, no special pattern	Random, no special pattern	Random, no special pattern	Random, no special pattern	Random, no special pattern
p-value ARCH	0.9271	0.5082	0.9681	0.0822	0.0061	8.27E-09
Decision ARCH	0	0	0	0	1	1
p-value KS	0.1799	0.0505	0.7787	0.8953	0.1152	0.1511
Decision KS	0	0	0	0	0	0

B.2.3 Bai-Perron (1998)

For the US slope, the Quandt test on the full sample finds evidence for a structural break. In Figure B.4a, we can notice significant variation of the Chow test sequence across candidate breakdates. The sequence reaches a high of 13.75 towards the end of the sample. This value denotes the Quandt statistic, which, compared with the Andrews asymptotic critical value, is significant at the 5% level. The estimated breakdate for the full sample is 2007:05, this is the point that minimizes the residual variance and represents the least squares breakdate estimate. In Figure B.5a, we can notice that the residual variance does not vary randomly across candidate breakdates. Instead, it varies systematically by first increasing and then decreasing, thus, forming a V-shape in the correspondence of 2007:05. This systematic variation of the residual variance casts doubts on the constancy of the parameters. Therefore, we break the sample at 2007:05 and test for the structural breaks on the two subsamples: [1999:01-2007:05] and [2007:06-2018:01]. We find no evidence for a break in the period [1999:01-2007:05], but we find evidence for a break in the period [2007:06-2018:01]. For this second period, the least square breakdate estimate is obtained in 2008-09. Now we split the sample in 2008:09 and reestimate on the sample period [1999:01-2008:09]. The Quandt test fails to find evidence for a structural break. The point estimate of the breakdate is 2001:06. Now taking the sample [2008:10-2018:01], the Quandt test rejects the hypothesis of parameter constancy at the 5% level, indicating a structural break. In Figure B.4e, it is visible how the Chow sequence breaks above the Andrews asymptotic critical value three times, with a maximum value of 17.42. For this period, the residual variance has a strong V shape (Figure B.5e) as a function of the breakdate, indicating good identification, and the minimum is obtained in 2011:08. Now we split the sample in 2011:08 and reestimate on the sample period [1999:01-2011:08]. The Quandt test fails to find evidence for a structural break. The point estimate of the breakdate is 2008:01. Testing on the sample period [2011:09-2018:01], the Quandt test rejects the hypothesis of parameter constancy and the point estimate of the breakdate is 2013:08. Now we split the sample at 2013:08. For the sample period [1999:01-2013:08], the Quandt test provides no evidence of structural break, the breakdate estimate is 2007:07. We cannot test the sample period [2013:09-2018:01] because the ending sample is smaller than the number of parameters.

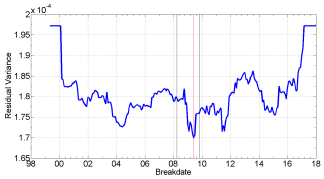
For the German slope, the Quandt test on the full sample finds no evidence for a structural break. In Figure B.6a, the Chow sequence exhibits very little variation in the first part of the sample period and then increased variation for the second part of the sample period. The sequence never breaks above the Andrews asymptotic critical value. Nevertheless, the residual variance as a function of breakdate registers a well-defined V-shape (Figure B.7a), where we could estimate the breakdate to be 2009:05. Because of this pattern in the residual variance, we break the sample at the estimated breakdate (2009:05) and test for the structural breaks on the two subsamples, [1999:01-2009:05] and [2009:06-2018:01]. For the first subsample, [1999:01-2009:05], the Quandt test fails to find evidence of a structural break. The point estimate of the breakdate is 2008:06, for this date, the residual variance as a function of breakdates has a well-defined V-shape. Taking the subsample [2009:06-2018:01], the Quandt test rejects the hypothesis of parameter constancy at the 5% level, indicating a structural break. In Figure B.6c, we can observe how the Chow sequence breaks above the Andrews asymptotic critical value at the 5% level of

significance, reaching a high of 10.3317, which denotes the Quandt statistic. The least squares estimate of the breakdate is 2011:01, i.e., the minimum reached by the residual variance as a function of breakdate (Figure B.7c). Now we split the sample in 2011:01 and reestimate on the sample period [1999:01-2011:01]. The Quandt test fails to find evidence for a structural break. The point estimate of the breakdate is 2008:08. Now taking the sample [2011:02-2018:01], the Quandt test fails again to find evidence of a structural break. The point estimate of the breakdate is 2012:06.

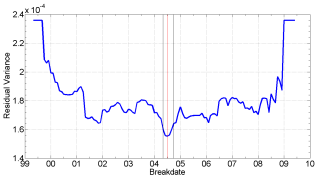
For the US level, the Quandt test on the full sample finds no evidence for a structural break. Compared to Andrews asymptotic critical value, in Figure B.8a, we can see that the Chow sequence never breaks above the critical value. However, in Figure B.9a, we can observe a systematic decrease and then an increase in the residual variance. A global minimum is reached in the correspondence of 2003:07, which represents the least squares breakdate estimate for the full sample. Because of this observation, we decide to split the full sample at 2003:07 and perform breakpoint tests on the two sub-samples, [1999:01-2003:07] and [2003:08-2018:01]. We find evidence for a break in the first sub-sample – the Quandt statistic is well higher than the Andrews asymptotic critical value (Figure B.8b) – and we find no evidence for a break in the second sub-sample. For the first sub-sample, the breakdate estimate is obtained in 2001:10. Finally, taking the sample [2001:10-2018:01], we again find no evidence of structural breaks.

On the full sample of the German level, the Quandt test fails to find evidence for a structural break (Figure B.10a). Nevertheless, because of the systematic variation in the residual variance (Figure B.11a), we decide to split the full sample at the breakdate estimate, 2009:06, and rerun the Quandt test on the sub-samples [1999:01-2009:06] and [2009:07-2018:01]. We find evidence for structural break in the first sub-sample. The Chow test sequence breaks above the Andrews asymptotic critical value (Figure B.10b). The breakdate estimate is obtained in 2004:07. We find no evidence for structural break in the second sub-sample. Finally, we take the sample [2004:07-2018:01] and rerun the Quandt test to conclude that no structural breaks are present.

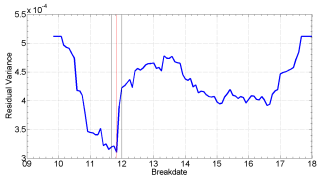
Figure B.11: Residual variance as a function of breakdates: German level



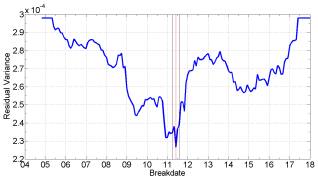
(a) Full sample



(b) Split 1a, [1999:01-2009:06]

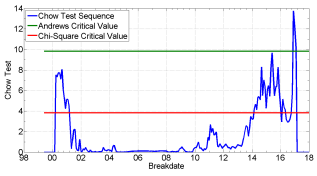


(c) Split 1b, [2009:07-2018:01]

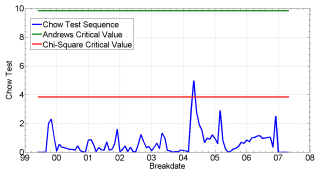


(d) Split 2, [2004:07-2018:01]

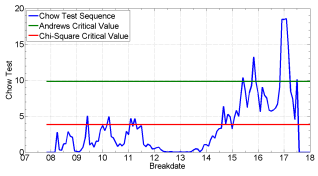
Figure B.4: Quandt statistic with Andrews asymptotic critical values: US Slope



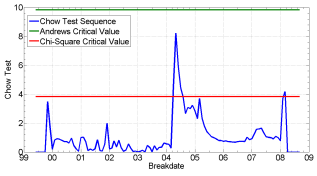
(a) Full sample



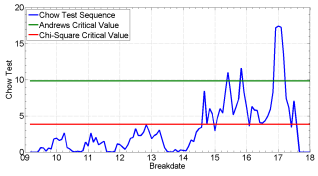
(b) Split 1a, [1999:01-2007:05]



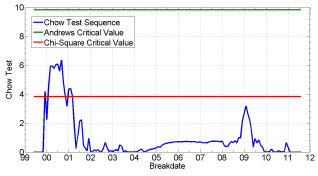
(c) Split 1b, [2007:05 - 2018:01]



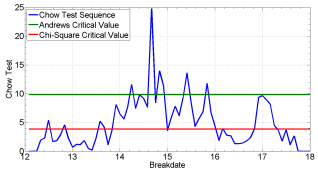
(d) Split 2a, [1999:01 - 2008:09]



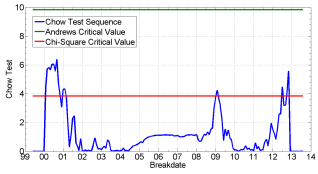
(e) Split 2b, [2008:10 - 2018:01]



(f) Split 3a, [1999:01 - 2011:08]



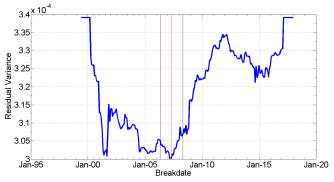
(g) Split 3b, [2011:09 - 2018:01]



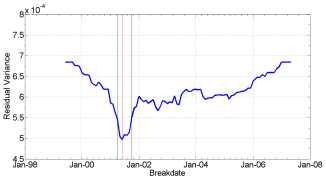
(h) Split 4a, [1999:01 - 2013:08]

Note: The horizontal green line denotes the Andrews (1993) asymptotic critical value. The horizontal red line denotes the χ^2 critical value.

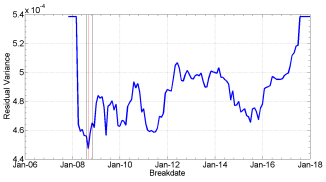
Figure B.5: Residual variance as a function of breakdates: US slope



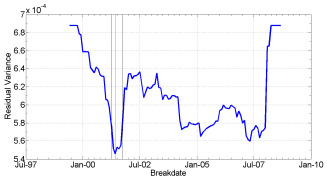
(a) Full sample



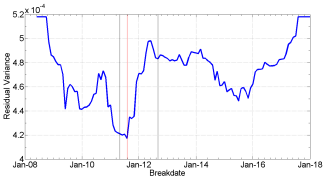
(b) Split 1a, [1999:01-2007:05]



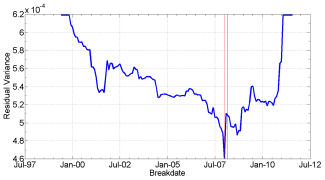
(c) Split 1b, [2007:05 - 2018:01]



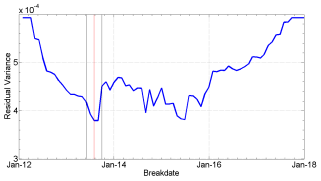
(d) Split 2a, [1999:01 - 2008:09]



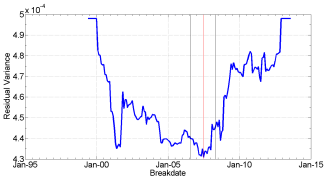
(e) Split 2b, [2008:10 - 2018:01]



(f) Split 3a, [1999:01 - 2011:08]

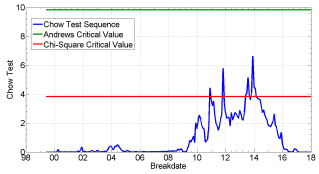


(g) Split 3b, [2011:09 - 2018:01]

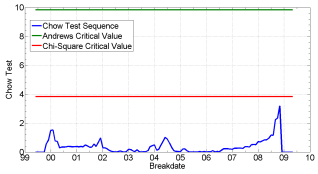


(h) Split 4a, [1999:01 - 2013:08]

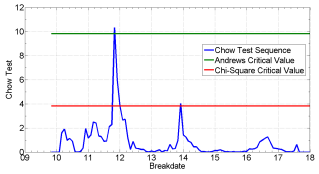
Figure B.6: Quandt statistic with Andrews asymptotic critical values: German slope



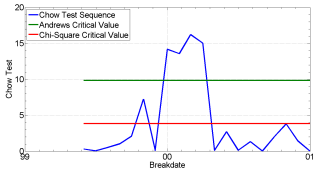
(a) Full sample



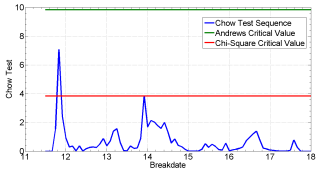
(b) Split 1a, [1999:01-2009:05]



(c) Split 1b, [2009:06 - 2018:01]



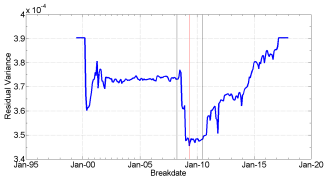
(d) Split 2a, [1999:01 - 2011:01]



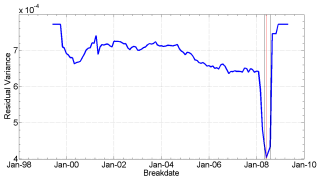
(e) Split 2b, [2011:02 - 2018:01]

Note: The horizontal green line denotes the Andrews (1993) asymptotic critical value. The horizontal red line denotes the χ^2 critical value.

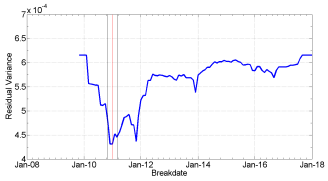
Figure B.7: Residual variance as a function of breakdates: German slope



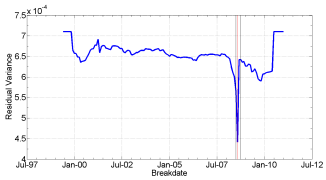
(a) Full sample



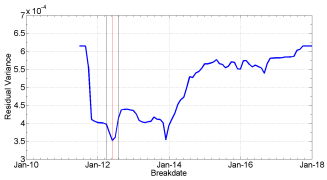
(b) Split 1a, [1999:01-2009:05]



(c) Split 1b, [2009:06 - 2018:01]

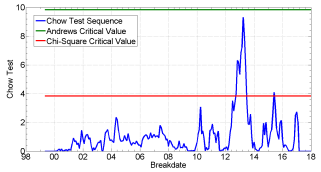


(d) Split 2a, [1999:01 - 2011:01]

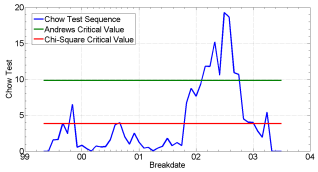


(e) Split 2b, [2011:02 - 2018:01]

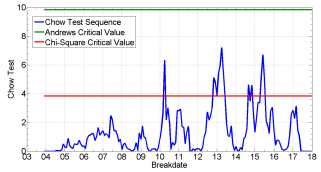
Figure B.8: Quandt statistic with Andrews asymptotic critical values: US level



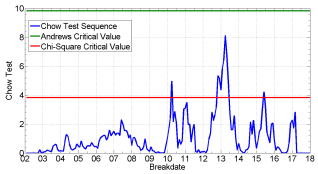
(a) Full sample



(b) Split 1a, [1999:01-2003:07]



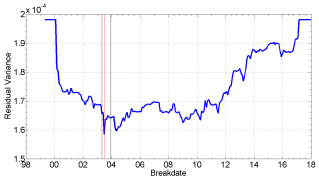
(c) Split 1b, [2003:08-2018:01]



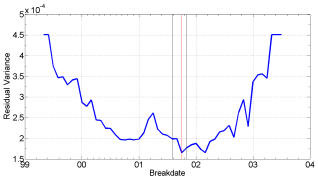
(d) Split 2, [2001:10-2018:01]

Note: The horizontal green line denotes the Andrews (1993) asymptotic critical value. The horizontal red line denotes the χ^2 critical value.

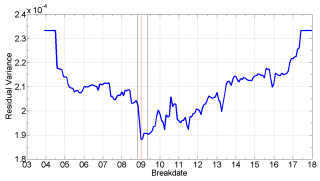
Figure B.9: Residual variance as a function of breakdates: US level



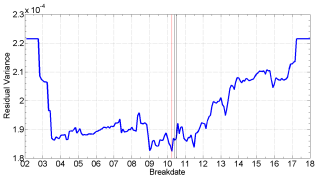
(a) Full sample



(b) Split 1a, [1999:01-2003:07]

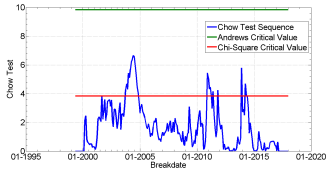


(c) Split 1b, [2003:08-2018:01]

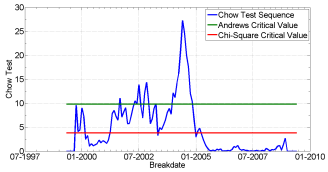


(d) Split 2, [2001:10-2018:01]

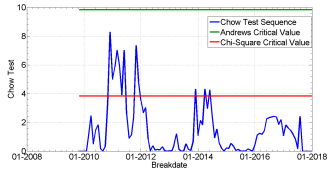
Figure B.10: Quandt statistic with Andrews asymptotic critical values: German level



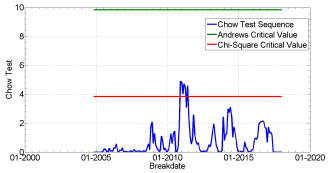
(a) Full sample



(b) Split 1a, [1999:01-2009:06]



(c) Split 1b, [2009:07-2018:01]



(d) Split 2, [2004:07-2018:01]

Note: The horizontal green line denotes the Andrews (1993) asymptotic critical value. The horizontal red line denotes the χ^2 critical value.

B.2.4 Perron-Zhou (2008)

For the US curvature, the estimated break dates in coefficients are 2007:10 and 2014:05 and the estimated break dates in variance are 2007:11 and 2011:10. The two breaks in coefficients designate three regimes: [1999:01-2007:10], [2007:11-2014:05], and [2014:06-2018:01], for which the changes in the model parameters are reported in Table B.2. The intercept of the US curvature series in the three regimes is -0.7323, -1.0338, and 0.7353, registering a reduction by 41.17%, from the first to the second regime, and then an increase by 171.13%, from the second to the third regime. The sum of the autoregressive coefficients in each regime is 0.8967, 0.988, and 1.0186, indicating a first increase in persistence by 9.24%, from the first to the second regime, followed by another increase of 3%, from the second to the third regime. A change in variance also occurred: the standard deviation of the errors is 0.4982 before 2007:11, 0.5409 after 2007:11 and till 2011:10, and 0.2857 after 2011:10 till the end of the sample period. For these three regimes, the variance first increased by 7.89% and then decreased by 89.32%.

For the German curvature, the estimated break date in coefficients is 2014:02 and in variance is 2009:11. The break date in coefficients denotes the presence of two regimes, [1999:01-2014:02] and [2014:03-2018:01], for which the changes in the model parameters are reported in Table B.2. For the two regimes, the value of the intercept is -0.5198 and -1.4568, registering a change of 180.26%. The sum of the autoregressive coefficients in each regime is 0.7763 and 0.533, indicating a reduction in persistence in 2014:02 by 31.34%. The standard deviation of the errors before 2009:11 is 0.7228 and 0.4546 after, indicating a reduction in variance by 37.11%.

Table B.2: Results of testing jointly for structural change in the regression coefficients and error variance. Parameter changes across regimes.

Regime	Type	Intercept	% Change	Σ AR Coeff.s.	% Change	Std. Dev of Errors	% Change
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: c_{US}							
$[1999:01-2007:10]$	Coefficients	-0.7323		0.8967			
$[2007:11-2014:05]$	Coefficients	-1.0338	41.17%	0.988	9.24%		
$[2014:06-2018:01]$	Coefficients	0.7353	-171.13%	1.0186	3.00%		
$[1999:01-2007:11]$	Variance					0.4982	
$[2007:12-2011:10]$	Variance					0.5409	7.89%
$[2011:11-2018:01]$	Variance					0.2857	-89.32%
System: $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$; Equation: c_{DE}							
$[1999:01-2014:02]$	Coefficients	-0.5198		0.7763			
$[2014:03-2018:01]$	Coefficients	-1.4568	180.26%	0.533	-31.34%		
$[1999:01-2009:11]$	Variance					0.7228	
$[2009:12-2018:01]$	Variance					0.4546	-37.11%

B.3 Monetary Policy and Interest Rates: US Fed vs ECB

Transmission Mechanism of Fed Monetary Policy

Changes in the Fed funds rate target impact directly short-term interest rates. An increase in the Fed funds rate target would likely increase short-term interest rates, while a decrease in the Fed funds rate target would decrease short-term interest rates. The changes in the short-term market rates are transmitted to medium- and longer-term interest rates (e.g., Treasury notes and bonds, corporate bonds, fixed-income mortgages, auto and consumer loans). Medium- and longer-term interest rates are also affected by economic agents'expectations about how the Fed funds rate will change in the future. Changes in the longer-term interest rates usually affect stock prices and changes in stock prices ultimately affect individuals'wealth. Changes in US monetary policy affect dollar exchange rates and international trade. For example, declining US interest rates would cause the yield on US dollar assets to look less appealing to international investors and, thus, reduce investments in dollar-denominated assets. Less investments in dollar-denominated assets would cause a dollar depreciation in foreign exchange markets. A dollar depreciation would make US goods and services cheaper on foreign markets, thus, it would boost US exports, on one hand, and it would reduce purchases of imported products and increase purchases of domestic products, on the other hand. The ultimate effect of changes in longer-term interest rates, stock prices, and the foreign exchange value of the dollar is on the spending decisions made by households and business.

US Fed Tools for Traditional Monetary Policy

The *Reserve Requirements* represent the policy tool, which requires all depository institutions to hold cash in their vaults or reserve balances at the Fed for an amount equal to a certain fraction of their deposits. The reserve balances are used by and between banks for overnight borrowing and lending. The interest rate on federal funds transactions is called the *federal funds rate. Open Market Operations* (OMOs) are a key tool used by the Fed in the implementation of monetary policy and consist in the purchase or sale of securities in the open market by a central bank. OMOs directly affect the volume of reserves in the banking system and thus the level of the federal funds rate. *Discount Window Lending* is the tool used by the Fed to relieve pressures in reserve markets and supply liquidity to depository institutions and the banking system as a whole, in periods of systemic stress.

US Fed Tools for Non-Traditional Monetary Policy

Large-scale asset purchases, also know as *Quantitative Easing* (QE), represent a tool of nontraditional monetary policy aiming at providing additional stimulus to interest-sensitive spending. Between November 2008 and October 2014, the Fed conducted three rounds of QE: US QE1, QE2, and QE3. During normalization, the FOMC is using an overnight reverse repurchase (ON RRP) facility as a supplementary tool to control the federal funds rate. The policy implementation during normalization consists in paying interest on reserves and offering ON RRPs. Other sup-

plementary tool used with the same aim of putting upward pressure on money market interest rates and help to control the federal funds rate are term deposits offered through the Fed's Term Deposit Facility and term reverse repurchase agreements. *Forward guidance* is the nontraditional monetary policy tool employed by the Fed to provide information about its intentions for the federal funds rate, in order to influence expectations about the future course of monetary policy.

Transmission Mechanism of ECB Monetary Policy

Figure B.12[1] illustrates the main transmission channels of ECB monetary policy decisions. The money market is the first financial market to be influenced by the monetary policy. When the ECB changes the official interest rates (i.e., the funding cost of liquidity for banks), the ECB steers the money market interest rates, as it affects the money market conditions, and it impacts the nominal market interest rates (e.g., the interest rates set by banks on short-term loans and deposits), as banks pass the positive or negative imbalance (created by the changes in the funding cost of liquidity) on to their customers. This latter process is the so-called "bank channel" as it allows banks to adjust their own rates for loans and savings, in order to preserve their spread[2]. Through the *bank channel*, the saving and investment decisions of households and firms are affected. The next effect is that changes in consumption and investment lead to a change in the aggregate demand, prices, and conditions in the labour markets. All other things being equal, when demand exceeds supply, prices tend to increase. Changes in aggregate demand may produce changes in the labour and intermediate product markets, thus affecting the wage and price-setting mechanisms in the respective markets. According to Economic Theory, these factors may lead to *inflation*, *deflation*, or *disinflation*[3].

Other two channels operating in the transmission mechanism of monetary policy are the *exchange rates channel* and the *asset price channel*.

[1] Source: ECB, `http://www.ecb.europa.eu/mopo/intro/transmission/html/index.en.html`

[2] This is known as the *Net Interest Income*, which is the excess revenue of banks generated from the interest earned on assets over the interest paid out on deposits. A typical bank's *assets* are represented by all forms of personal and commercial loans, mortgages, and securities. The *liabilities* are the customer deposits. (Investopedia, Net Interest Income)

[3] *Inflation* is defined as a sustained increase in the general level of prices for goods and services. *Deflation* is the general decline in prices for goods and services occurring when the inflation rate falls below 0%. *Disinflation* describes periods of slowing inflation. Inflation and deflation refer to the direction of prices. Disinflation refers to the rate of change in the rate of inflation (Investopedia).

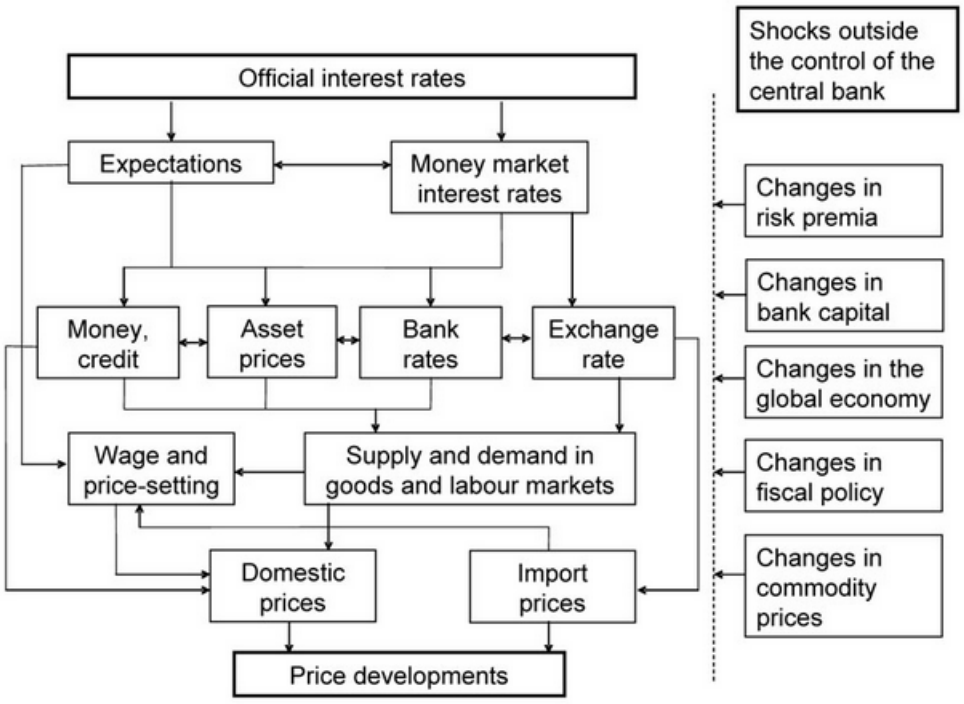


Figure B.12: The main transmission channels of ECB monetary policy decisions. Source: ECB

ECB Tools for Traditional Monetary Policy

The main open market operations are the following. The *Main Refinancing Operations* (MROs) are regular, open market, *reverse transactions*, executed by the Eurosystem with the aim of lending funds to its counterparties (i.e., banks). In order to protect the Eurosystem against financial risks, lending is always based on adequate collateral. MROs are liquidity-providing transactions (via *reverse transactions*), of one-week maturity, and weekly frequency. *Longer-Term Refinancing Operations* (LTROs) are regular, open market, *reverse transactions*, executed by the Eurosystem with the aim of lending long-term liquidity to its counterparties. LTROs are liquidity-providing transactions, of three-month maturity, and monthly frequency. *Fine-Tuning Operations* (FTOs) are ad hoc transactions, aimed at providing (liquidity-providing, via *reverse transactions* and *foreign exchange swaps*) or absorbing (liquidity-absorbing, via *reverse transactions*, *collection of fixed-term deposits*, and *foreign exchange swaps*) liquidity in the money market and at steering interest rates, in order to smooth the effects of unexpected liquidity fluctuations in the banking sector. FTOs are transactions of non-standardized maturity and non-regular frequency. *Structural operations* aim at adjusting the structural positions of the Eurosystem with respect to the financial sector. Structural operations can take place as liquidity-providing transactions (via *reverse transactions* and *outright purchases*) and as liquidity-absorbing transactions (via *issuance of ECB debt certificates* and *outright sales*). Structural operations can be of standardized/non-standardized maturity and regular/non-regular maturity.

In addition to the open market operations, the Eurosystem also implements monetary policy

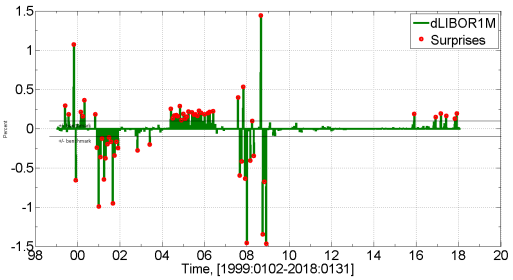
by setting the interest rates on its standing facilities. *Standing facilities* are monetary policy operations initiated by the counterparties[4] (i.e., credit institutions) and employed to provide or absorb liquidity with an overnight maturity. Eligible counterparties can access at their discretion the *marginal lending facility* (which are liquidity-providing reverse transactions, with which banks can borrow overnight funds from their national central banks, against eligible collateral) and the *deposit facility* (which are liquidity-absorbing transactions, with which banks can make overnight deposits with their national central banks).

B.4 Empirical Results: Predictability of US Fed vs ECB

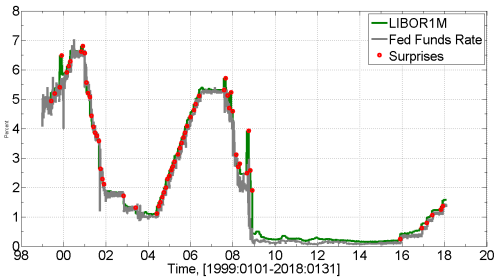
Fed Hit-Rate

[4]Contrary to the open market operations, which are initiated by the ECB.

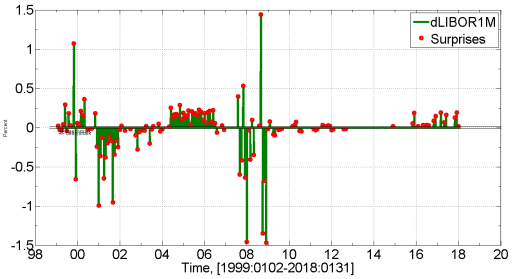
Figure B.13: Fed Hit-Rate: LIBOR1M and LIBOR3M



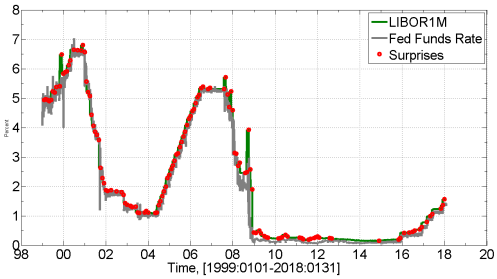
(a) $\Delta LIBOR1M^{*}$



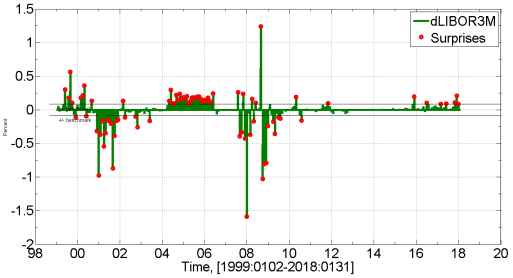
(b) $LIBOR1M^{*}$



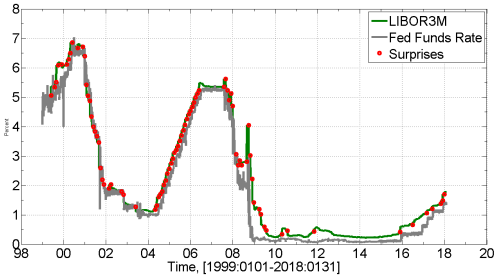
(c) $\Delta LIBOR1M^{**}$



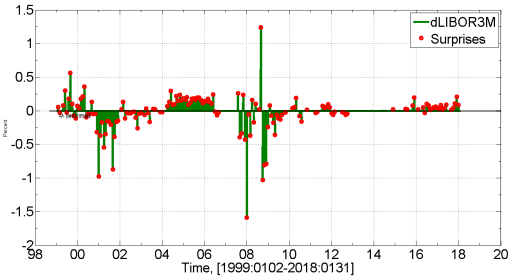
(d) $LIBOR1M^{**}$



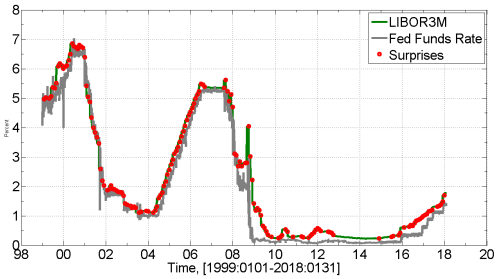
(e) $\Delta LIBOR3M^{*}$



(f) $LIBOR3M^{*}$



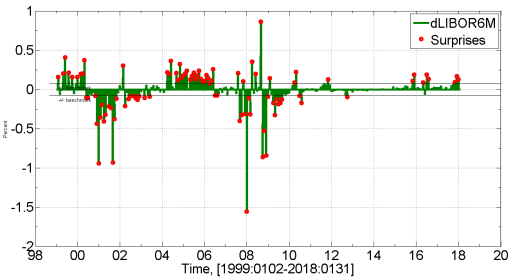
(g) $\Delta LIBOR3M^{**}$



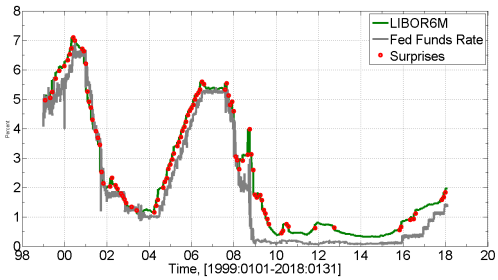
(h) $LIBOR3M^{**}$

Note: (*) Surprises defined according to the 2 times standard deviation criterion. (**) Surprises defined according to the 12.5 basis points criterion.

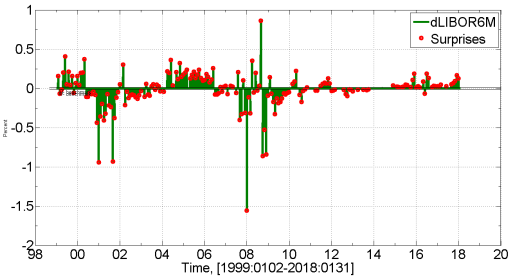
Figure B.14: Fed Hit-Rate: LIBOR6M and LIBOR12M



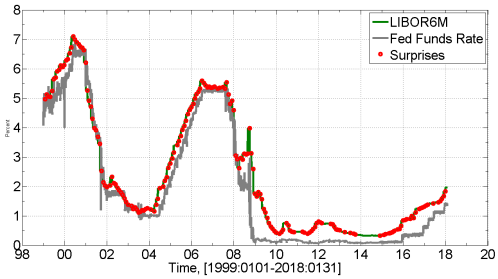
(a) $\Delta LIBOR6M^{*}$



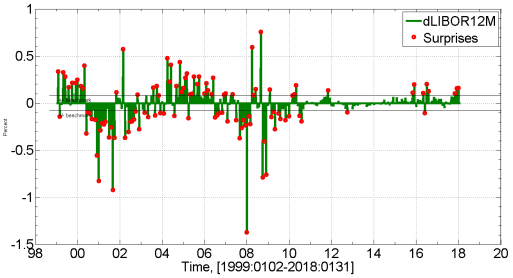
(b) $LIBOR6M^{*}$



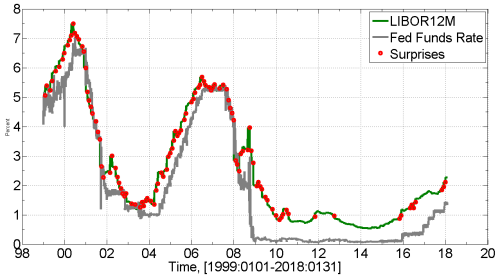
(c) $\Delta LIBOR6M^{**}$



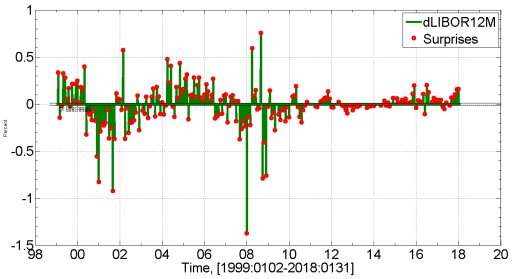
(d) $LIBOR6M^{**}$



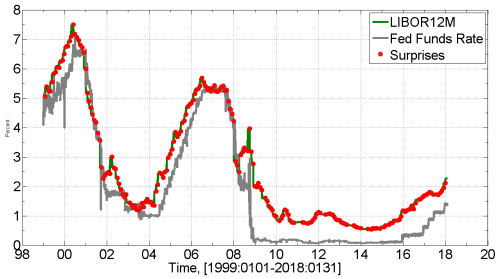
(e) $\Delta LIBOR12M^{*}$



(f) $LIBOR12M^{*}$



(g) $\Delta LIBOR12M^{**}$

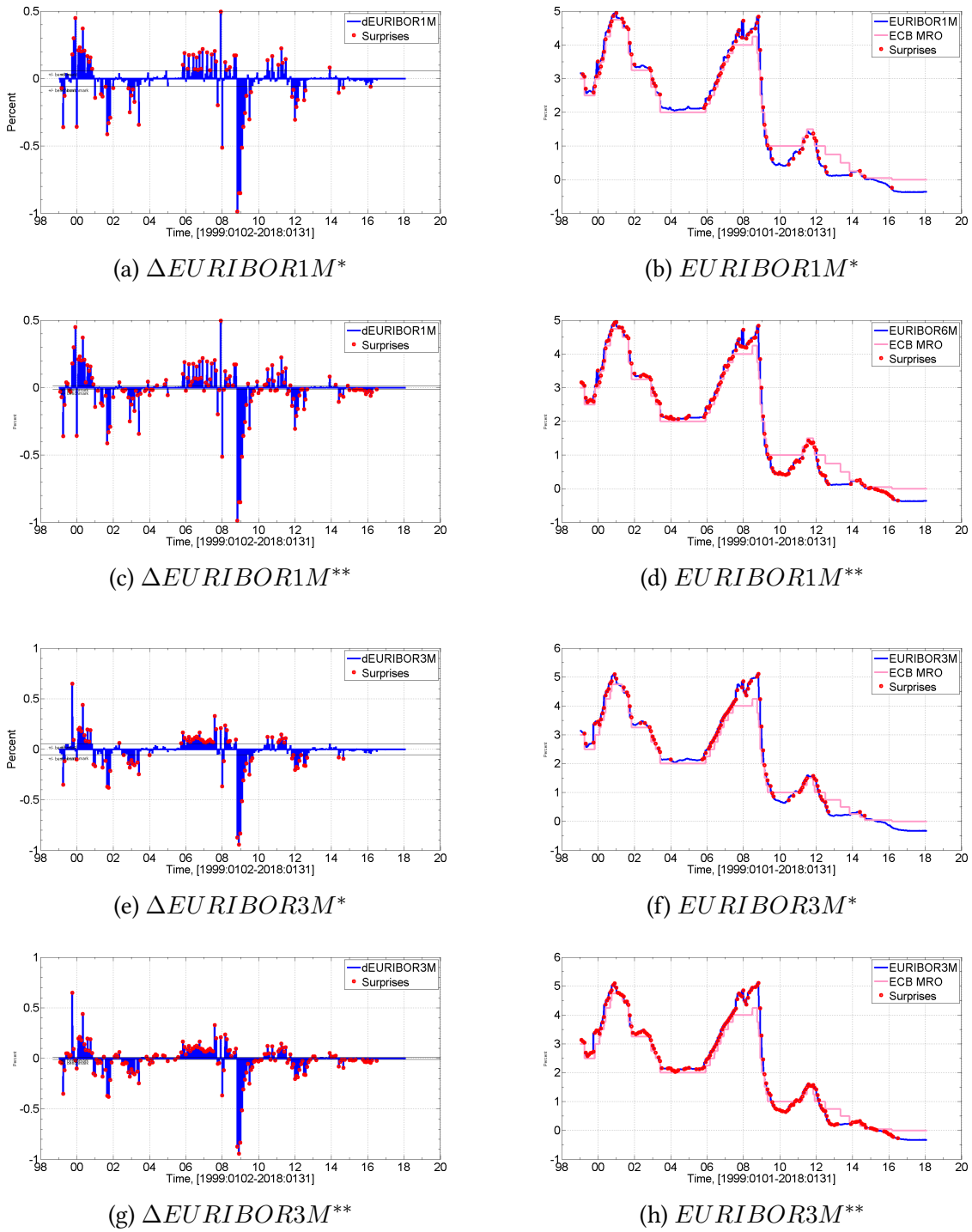


(h) $LIBOR12M^{**}$

Note: (*) Surprises defined according to the 2 times standard deviation criterion. (**) Surprises defined according to the 12.5 basis points criterion.

ECB Hit-Rate

Figure B.15: ECB Hit-Rate: EURIBOR1M and EURIBOR3M



(a) $\Delta EURIBOR1M^{*}$ (b) $EURIBOR1M^{*}$

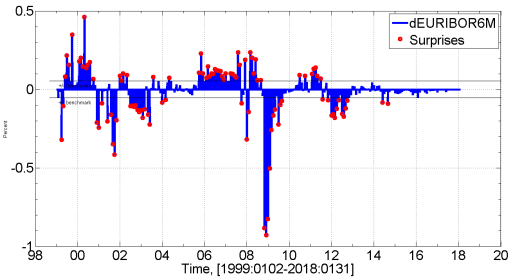
(c) $\Delta EURIBOR1M^{**}$ (d) $EURIBOR1M^{**}$

(e) $\Delta EURIBOR3M^{*}$ (f) $EURIBOR3M^{*}$

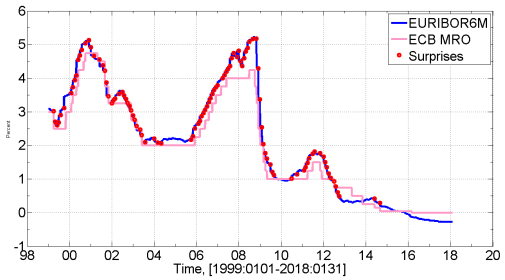
(g) $\Delta EURIBOR3M^{**}$ (h) $EURIBOR3M^{**}$

Note: (*) Surprises defined according to the 2 times standard deviation criterion. (**) Surprises defined according to the 12.5 basis points criterion.

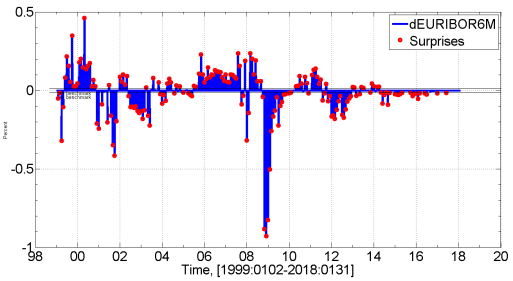
Figure B.16: ECB Hit-Rate: EURIBOR6M and EURIBOR12M



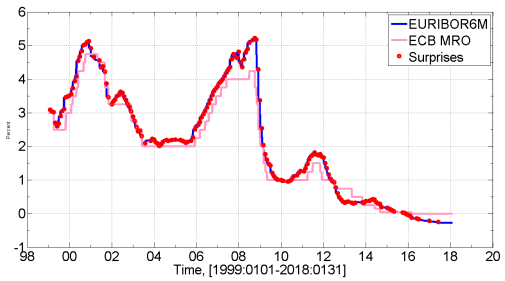
(a) $\Delta EURIBOR6M^{*}$



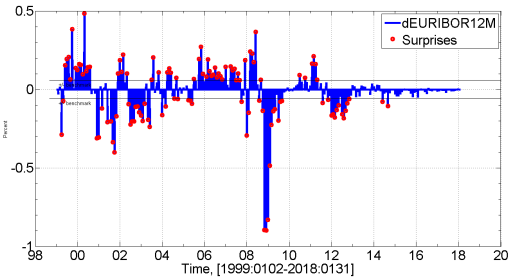
(b) $EURIBOR6M^{*}$



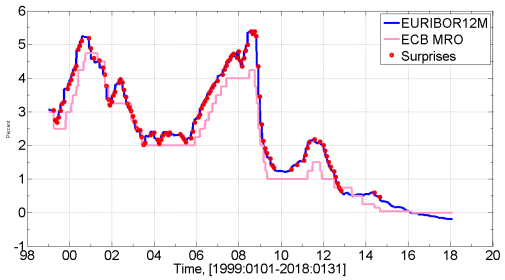
(c) $\Delta EURIBOR6M^{**}$



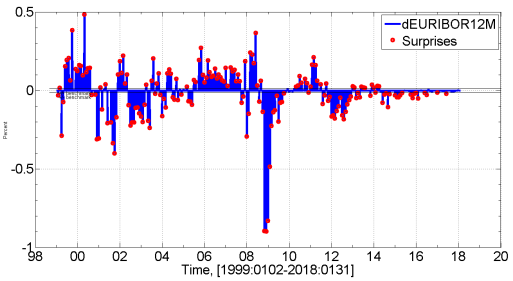
(d) $EURIBOR6M^{**}$



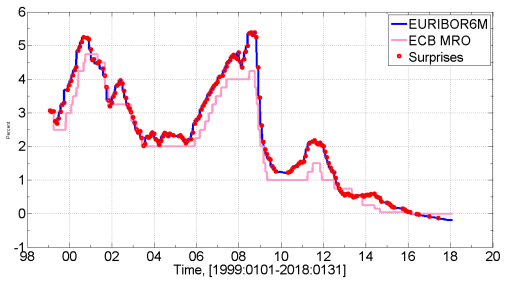
(e) $\Delta EURIBOR12M^{*}$



(f) $EURIBOR12M^{*}$



(g) $\Delta EURIBOR12M^{**}$



(h) $EURIBOR12M^{**}$

Note: (*) Surprises defined according to the 2 times standard deviation criterion. (**) Surprises defined according to the 12.5 basis points criterion.

B.5 Structural Breaks: Multivariate State-Space Analysis

B.5.1 Correlation of Residuals

Figure B.17 is a matrix of plots showing correlations among pairs of residuals from the three fitted multivariate models. Histograms of the variables are located along the matrix diagonal, the scatter plots of residuals pairs appear off diagonal. The slopes of the least-squares reference lines in the scatter plots are equal to the displayed correlation coefficients.

There are 6 residuals series denoted as follows. M1sUS and M1sDE are the residuals from the $s_{US,t}$ and $s_{DE,t}$ equations, respectively, in the 2D-VAR(5) model; M2lUS and M2lDE are the residuals from the $l_{US,t}$ and $l_{DE,t}$ equations, respectively, in the 2D-VEC(1) model; and M3lUS, M3lDE, M3cUS, and M3cDE are the residuals from the $l_{US,t}$, $l_{DE,t}$, $c_{US,t}$, and $c_{DE,t}$ equations, respectively, in the 4D-VEC(3) model.

The correlation coefficients highlighted in red indicate which pairs of residuals have correlations significantly different from zero. We can notice that most of the residuals have correlations significantly different from zero. This observation suggest that a joint analysis of the three multivariate processes might provide more accurate results concerning the presence of structural breaks.

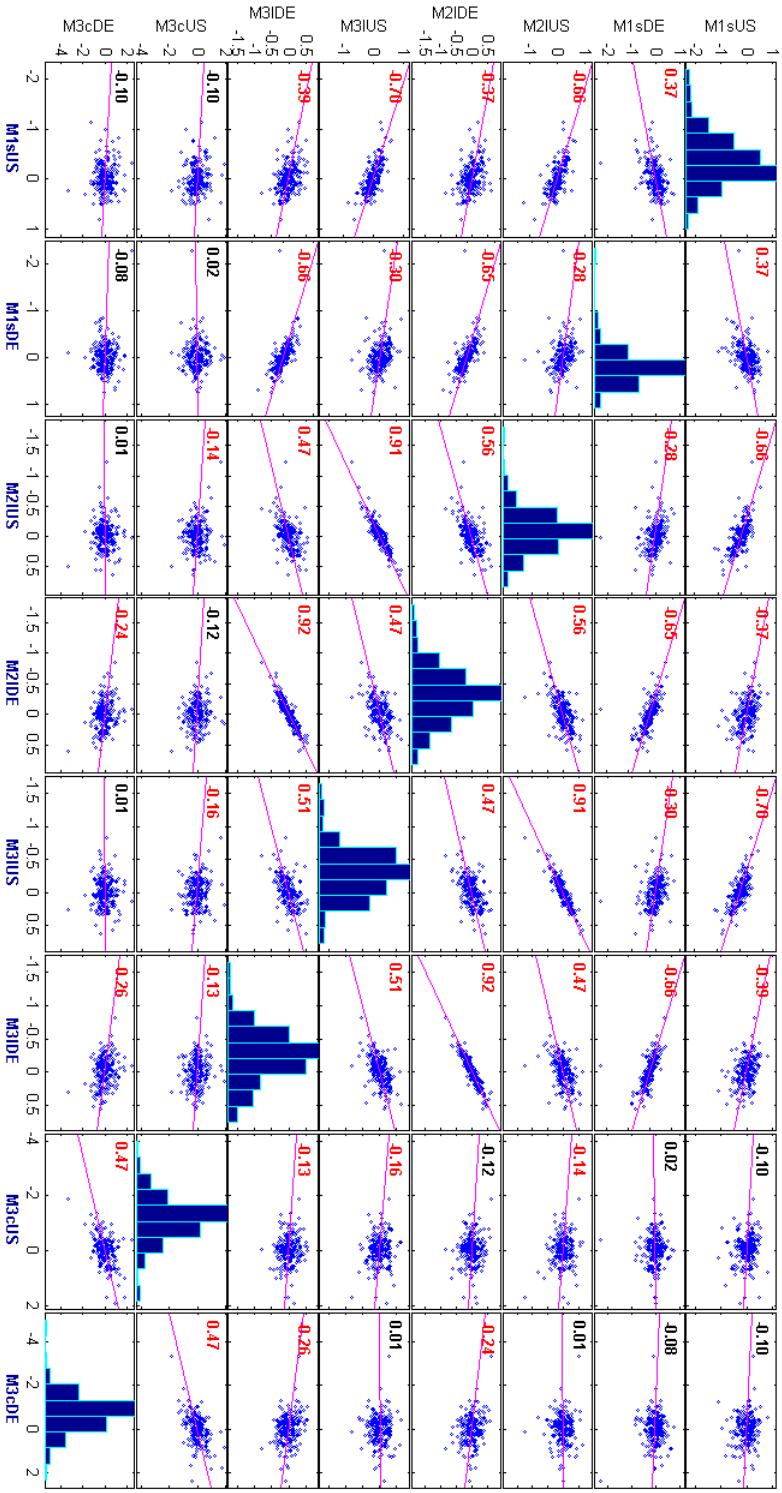


Figure B.17: Correlation matrix of residuals from the $s_{US,t}/s_{DE,t}$, $l_{US,t}/l_{DE,t}$, and $l_{US,t}/l_{DE,t}/c_{US,t}/c_{DE,t}$ systems.

Bibliography

2008/C115/01 (2010). *Consolidated Versions of the Treaty on European Union and of the Treaty on the Functioning of the European Union: Charter of Fundamental Rights of the European Union*. Office for Official publications of the European Communities.

Abbritti, Mirko et al. (2013). "Global factors in the term structure of interest rates". In:

Agrawal, JG, VS Chourasia, and AK Mittra (2013). "State-of-the-art in stock prediction techniques". In: *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering* 2.4, pp. 1360–1366.

Aguiar-Conraria, Luís, Manuel MF Martins, and Maria Joana Soares (2012). "The yield curve and the macro-economy across time and frequencies". In: *Journal of Economic Dynamics and Control* 36.12, pp. 1950–1970.

Ahn, Hee-Joon, Jun Cai, and Yan-Leung Cheung (2002). "What moves German Bund futures contracts on the Eurex?" In: *Journal of Futures Markets* 22.7, pp. 679–696.

Al Awad, Mouawiya and Barry K Goodwin (1998). "Dynamic linkages among real interest rates in international capital markets". In: *Journal of International Money and Finance* 17.6, pp. 881–907.

Almeida, Caio et al. (2009). "Does curvature enhance forecasting?" In: *International Journal of Theoretical and Applied Finance* 12.08, pp. 1171–1196.

Andersen, Torben G and Jesper Lund (1997). *Stochastic volatility and mean drift in the short rate diffusion: sources of steepness, level and curvature in the yield curve*. Tech. rep. working paper, Northwestern University.

Anderton, Robert, Filippo Di Mauro, and Fabio Moneta (2004). "Understanding the impact of the external dimension on the euro area: trade, capital flows and other international macroeconomic linkages". In:

Andrews, Donald WK (1987). "Asymptotic results for generalized Wald tests". In: *Econometric Theory* 3.3, pp. 348–358.

— (1993). "Tests for parameter instability and structural change with unknown change point". In: *Econometrica: Journal of the Econometric Society*, pp. 821–856.

Andrews, Donald WK and Werner Ploberger (1994). "Optimal tests when a nuisance parameter is present only under the alternative". In: *Econometrica: Journal of the Econometric Society*, pp. 1383–1414.

Ang, Andrew and Geert Bekaert (2002). "Regime switches in interest rates". In: *Journal of Business & Economic Statistics* 20.2, pp. 163–182.

Ang, Andrew and Monika Piazzesi (2003). "A no-arbitrage vector autoregression of term structure dynamics with macroeconomic and latent variables". In: *Journal of Monetary economics* 50.4, pp. 745–787.

Ang, Andrew, Monika Piazzesi, and Min Wei (2006). "What does the yield curve tell us about GDP growth?" In: *Journal of econometrics* 131.1-2, pp. 359–403.

Aoki, Masanao (2013). *State space modeling of time series*. Springer Science & Business Media.

Aoki, Masanao and Arthur Havenner (1991). "State space modeling of multiple time series". In: *Econometric Reviews* 10.1, pp. 1–59.

Atkinson, Anthony Curtes (1985). *Plots, transformations and regression; an introduction to graphical methods of diagnostic regression analysis*. Tech. rep.

Axel, Ralph and Prashant Vankudre (2002). "Managing the Yield Curve with Principal Component Analysis". In: *in Frank J. Fabozzi (ed.), Professional Perspectives on Fixed Income Portfolio Management, Vol. 3 (Hoboken, NJ: John Wiley & Sons)*.

Bai, Jushan (1994). "Least squares estimation of a shift in linear processes". In: *Journal of Time Series Analysis* 15.5, pp. 453–472.

— (1997a). "Estimating multiple breaks one at a time". In: *Econometric theory* 13.3, pp. 315–352.

— (1997b). "Estimation of a change point in multiple regression models". In: *The review of economics and statistics* 79.4, pp. 551–563.

Bai, Jushan, Robin L Lumsdaine, and James H Stock (1998). "Testing for and dating common breaks in multivariate time series". In: *The Review of Economic Studies* 65.3, pp. 395–432.

Bai, Jushan and Pierre Perron (1998). "Estimating and testing linear models with multiple structural changes". In: *Econometrica*, pp. 47–78.

— (2003). "Computation and analysis of multiple structural change models". In: *Journal of applied econometrics* 18.1, pp. 1–22.

Bai, Jushan et al. (2000). *Vector autoregressive models with structural changes in regression coefficients and in variance-covariance matrices*. Tech. rep. China Economics, Management Academy, Central University of Finance, and Economics.

Balduzzi, Pierluigi et al. (1996). "A simple approach to three-factor affine term structure models". In: *The Journal of Fixed Income* 6.3, pp. 43–53.

Ballings, Michel et al. (2015). "Evaluating multiple classifiers for stock price direction prediction". In: *Expert Systems with Applications* 42.20, pp. 7046–7056.

Bańbura, Marta, Domenico Giannone, and Lucrezia Reichlin (2010). "Large Bayesian vector auto regressions". In: *Journal of Applied Econometrics* 25.1, pp. 71–92.

Banerjee, Anindya and David F Hendry (1992). "Testing integration and cointegration: An overview". In: *Oxford Bulletin of Economics and Statistics* 54.3, pp. 225–255.

Bansal, Ravi and Hao Zhou (2002). "Term structure of interest rates with regime shifts". In: *The Journal of Finance* 57.5, pp. 1997–2043.

Barassi, Marco R, Guglielmo Maria Caporale, and Stephen G Hall (2005). "Interest rate linkages: a Kalman filter approach to detecting structural change". In: *Economic Modelling* 22.2, pp. 253–284.

Bauer, Dietmar and Martin Wagner (2002). "Estimating cointegrated systems using subspace algorithms". In: *Journal of Econometrics* 111.1, pp. 47–84.

Bauer, Gregory H and Antonio Diez de los Rios (2012). *An international dynamic term structure model with economic restrictions and unspanned risks*. Tech. rep. Bank of Canada working paper.

Becker, Bo and Victoria Ivashina (2015). "Reaching for yield in the bond market". In: *The Journal of Finance* 70.5, pp. 1863–1902.

Belke, Ansgar and Daniel Gros (2005). "Asymmetries in Transatlantic Monetary Policy-making: Does the ECB Follow the Fed?" In: *JCMS: Journal of Common Market Studies* 43.5, pp. 921–946.

Ben-David, Dan and David H Papell (1998). "Slowdowns and meltdowns: postwar growth evidence from 74 countries". In: *Review of Economics and Statistics* 80.4, pp. 561–571.

Berk, Jan Marc (1998). "The information content of the yield curve for monetary policy: A survey". In: *De Economist* 146.2, pp. 303–320.

Bernanke, Ben S, Jean Boivin, and Piotr Eliasz (2005). "Measuring the effects of monetary policy: a factor-augmented vector autoregressive (FAVAR) approach". In: *The Quarterly journal of economics* 120.1, pp. 387–422.

Bertocchi, Marida, Vittorio Moriggia, and Jitka Dupačová (2000). "Sensitivity of bond portfolio's behavior with respect to random movements in yield curve: A simulation study". In: *Annals of Operations Research* 99.1-4, pp. 267–286.

Black, Fischer and Robert Litterman (1992). "Global portfolio optimization". In: *Financial analysts journal* 48.5, pp. 28–43.

Bliss, Robert R (1997a). "Movements in the term structure of interest rates". In: *Economic Review-Federal Reserve Bank of Atlanta* 82.4, p. 16.

— (1997b). "Testing term structure estimation methods". In: *Advances in Futures and Options Research 9*, pp. 97–231.

Bolder, David J (2015). "Fixed-income portfolio analytics". In: *Suiza: Springer*.

Booth, Ash, Enrico Gerding, and Frank Mcgroarty (2014). "Automated trading with performance weighted random forests and seasonality". In: *Expert Systems with Applications* 41.8, pp. 3651–3661.

Borio, Claudio, Leonardo Gambacorta, and Boris Hofmann (2017). "The influence of monetary policy on bank profitability". In: *International Finance* 20.1, pp. 48–63.

Borio, Claudio EV and Andrew J Filardo (2007). "Globalisation and inflation: New cross-country evidence on the global determinants of domestic inflation". In:

Box, George EP and David A Pierce (1970). "Distribution of residual autocorrelations in autoregressive-integrated moving average time series models". In: *Journal of the American statistical Association* 65.332, pp. 1509–1526.

Box, George EP and George C Tiao (1975). "Intervention analysis with applications to economic and environmental problems". In: *Journal of the American Statistical association* 70.349, pp. 70–79.

Box, George EP et al. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.

Brand, Claus, Daniel Buncic, and Jarkko Turunen (2010). "The impact of ECB monetary policy decisions and communication on the yield curve". In: *Journal of the European Economic Association* 8.6, pp. 1266–1298.

Brandt, Michael W and Amir Yaron (2003). *Time-consistent no-arbitrage models of the term structure*. Tech. rep. National Bureau of Economic Research.

Breitung, Jörg and Sandra Eickmeier (2011). "Testing for structural breaks in dynamic factor models". In: *Journal of Econometrics* 163.1, pp. 71–84.

Brennan, Michael J and Yihong Xia (2006). "International capital markets and foreign exchange risk". In: *The Review of Financial Studies* 19.3, pp. 753–795.

Brockwell, Peter J and Richard A Davis (2013). *Time series: theory and methods*. Springer Science & Business Media.

Byrne, Joseph P, Shuo Cao, and Dimitris Korobilis (2017). "Decomposing Global Yield Curve Co-Movement". In:

Byrne, Joseph P, Giorgio Fazio, and Norbert Fiess (2012). "Interest rate co-movements, global factors and the long end of the term spread". In: *Journal of Banking & Finance* 36.1, pp. 183–192.

Caines, Peter E (2018). *Linear stochastic systems*. Vol. 77. SIAM.

Caldeira, João F, Guilherme V Moura, and André AP Santos (2016). "Bond portfolio optimization using dynamic factor models". In: *Journal of Empirical Finance* 37, pp. 128–158.

Campbell, John Y (1995). "Some lessons from the yield curve". In: *Journal of economic perspectives* 9.3, pp. 129–152.

Campbell, John Y and Robert J Shiller (1987). "Cointegration and tests of present value models". In: *Journal of political economy* 95.5, pp. 1062–1088.

Campos, Julia, Neil R Ericsson, and David F Hendry (1996). "Cointegration tests in the presence of structural breaks". In: *Journal of Econometrics* 70.1, pp. 187–220.

Carstensen, Kai (2003). "Nonstationary term premia and cointegration of the term structure". In: *Economics Letters* 80.3, pp. 409–413.

Castle, Jennifer L and David F Hendry (2010). "A low-dimension portmanteau test for non-linearity". In: *Journal of Econometrics* 158.2, pp. 231–245.

Chauvet, Marcelle and Simon Potter (2005). "Forecasting recessions using the yield curve". In: *Journal of Forecasting* 24.2, pp. 77–103.

Chen, Lin (1996). *Stochastic mean and stochastic volatility: a three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management*. Blackwell publishers.

Chinn, Menzie David and Jeffrey A Frankel (2003). "The euro area and world interest rates". In: *Department of Economics, UCSC*.

Chong, Terence Tai-leung (1995). "Partial parameter consistency in a misspecified structural change model". In: *Economics Letters* 49.4, pp. 351–357.

Chopin, Nicolas and Florian Pelgrin (2004). "Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation". In: *Journal of Econometrics* 123.2, pp. 327–344.

Choudhry, Taufiq et al. (2012). "High-frequency exchange-rate prediction with an artificial neural network". In: *Intelligent Systems in Accounting, Finance and Management* 19.3, pp. 170–178.

Chow, Gregory C (1960). "Tests of equality between sets of coefficients in two linear regressions". In: *Econometrica: Journal of the Econometric Society*, pp. 591–605.

Christiansen, Charlotte and Jesper Lund (2005). "Revisiting the shape of the yield curve: The effect of interest rate volatility". In:

Ciccarelli, Matteo and Benoit Mojon (2010). "Global inflation". In: *The Review of Economics and Statistics* 92.3, pp. 524–535.

Cobb, George W (1978). "The problem of the Nile: conditional solution to a changepoint problem". In: *Biometrika* 65.2, pp. 243–251.

Commandeur, Jacques JF and Siem Jan Koopman (2007). *An introduction to state space time series analysis*. Oxford University Press.

Commandeur, Jacques JF, Siem Jan Koopman, Marius Ooms, et al. (2011). "Statistical software for state space methods". In:

Congress, United States. (1913). "Federal Reserve Act : Public Law 63-43, 63d Congress, H.R. 7837: An Act to Provide for the Establishment of Federal Reserve Banks, to Furnish an Elastic Currency, to Afford Means of Rediscounting Commercial Paper, to Establish a More Effective Supervision of Banking in the United States, and for Other Purposes". In: *Federal Reserve Act*.

Cook, R Dennis and Sanford Weisberg (1982). *Residuals and influence in regression*. New York: Chapman and Hall.

Coppel, Jonathan, Ellis Connolly, et al. (2003). *What do financial market data tell us about monetary policy transparency?* Reserve Bank of Australia.

Cour-Thimann, Philippine and Bernhard Winkler (2012). "The ECB's non-standard monetary policy measures: the role of institutional factors and financial structure". In: *Oxford Review of Economic Policy* 28.4, pp. 765–803.

Cox, John C, Jonathan E Ingersoll Jr, and Stephen A Ross (1977). "A theory of the term structure of interest rates". In: *Econometrica* 53, pp. 385–407.

Dahlquist, Magnus and Henrik Hasseltoft (2013). "International bond risk premia". In: *Journal of International Economics* 90.1, pp. 17–32.

Dai, Qiang and Kenneth J Singleton (2000). "Specification analysis of affine term structure models". In: *The Journal of Finance* 55.5, pp. 1943–1978.

De Jong, Frank and Pedro Santa-Clara (1999). "The dynamics of the forward interest rate curve: A formulation with state variables". In: *Journal of Financial and Quantitative Analysis* 34.1, pp. 131–157.

De Jong, Piet and Jeremy Penzer (1998). "Diagnosing shocks in time series". In: *Journal of the American Statistical Association* 93.442, pp. 796–806.

De La Dehesa, Guillermo (2013). "Non-Standard and Unconventional Monetary Policy Measures". In: *Non-Standard Monetary Policy Measures-An Update, European Parliament Directorate General for International Policies Policy Department*, pp. 43–54.

De Mol, Christine, Domenico Giannone, and Lucrezia Reichlin (2008). "Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components?" In: *Journal of Econometrics* 146.2, pp. 318–328.

Delivorias, Angelos (2015). "Monetary policy of the European Central Bank". In: *EPRS|European Parliament Research Center*.

Dewachter, Hans and Marco Lyrio (2006). "Macro factors and the term structure of interest rates". In: *Journal of Money, Credit, and Banking* 38.1, pp. 119–140.

Diebold, Francis X and Canlin Li (2006). "Forecasting the term structure of government bond yields". In: *Journal of econometrics* 130.2, pp. 337–364.

Diebold, Francis X, Canlin Li, and Vivian Z Yue (2008). "Global yield curve dynamics and interactions: a dynamic Nelson–Siegel approach". In: *Journal of Econometrics* 146.2, pp. 351–363.

Diebold, Francis X, Monika Piazzesi, and Glenn Rudebusch (2005). *Modeling bond yields in finance and macroeconomics*. Tech. rep. National Bureau of Economic Research.

Diebold, Francis X and Glenn D Rudebusch (2013). *Yield curve modeling and forecasting: the dynamic Nelson-Siegel approach*. Princeton University Press.

Diebold, Francis X, Glenn D Rudebusch, and S Boragan Aruoba (2006). "The macroeconomy and the yield curve: a dynamic latent factor approach". In: *Journal of econometrics* 131.1, pp. 309–338.

Drehmann, Mathias, Steffen Sorensen, and Marco Stringa (2010). "The integrated impact of credit and interest rate risk on banks: A dynamic framework and stress testing application". In: *Journal of Banking & Finance* 34.4, pp. 713–729.

Dueker, Michael J (1997). "Strengthening the Case for the Yield Curve as a Predictor of US Recessions". In: *Federal Reserve Bank of St. Louis Review* 79.2, p. 41.

Duffee, Gregory R (2002). "Term premia and interest rate forecasts in affine models". In: *The Journal of Finance* 57.1, pp. 405–443.

— (2006). "Term structure estimation without using latent factors". In: *Journal of Financial Economics* 79.3, pp. 507–536.

Duffie, Darrell and Rui Kan (1996). "A yield-factor model of interest rates". In: *Mathematical finance* 6.4, pp. 379–406.

Dungey, Mardi, Vance L Martin, and Adrian R Pagan (2000). "A multivariate latent factor decomposition of international bond yield spreads". In: *Journal of Applied Econometrics* 15.6, pp. 697–715.

Dunis, Christian L et al. (2016). *Artificial Intelligence in Financial Markets: Cutting Edge Applications for Risk Management, Portfolio Optimization and Economics*. Springer.

Durbin, James and Siem Jan Koopman (2012). *Time series analysis by state space methods*. Vol. 38. OUP Oxford.

EBA (2011). "Capital buffers for addressing market concerns over sovereign exposures: Methodological Note". In:

ECB (2011a). "The European Central Bank, the Eurosystem, the European System of Central Banks". In: *European Central Bank, Eurosystem*.

— (2011b). "The implementation of monetary policy in the euro area: general documentation on Eurosystem monetary policy instruments and procedures". In: *European Central Bank*.

— (2011c). "The monetary policy of the ECB". In: *European Central Bank, Eurosystem*.

— (2012). "Implementation of new collateral rules and reserve requirements". In: *ECB Monthly Bulleting*.

Eilers, Dennis et al. (2014). "Intelligent trading of seasonal effects: A decision support algorithm based on reinforcement learning". In: *Decision support systems* 64, pp. 100–108.

Ellingsen, Tore and Ulf Soderstrom (2001). "Monetary policy and market interest rates". In: *American Economic Review* 91.5, pp. 1594–1607.

Engle, Robert F (1982). "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation". In: *Econometrica: Journal of the Econometric Society*, pp. 987–1007.

Engle, Robert F and Clive WJ Granger (1987). "Co-integration and error correction: representation, estimation, and testing". In: *Econometrica: journal of the Econometric Society*, pp. 251–276.

Estrella, Arturo (2005). "Why does the yield curve predict output and inflation?" In: *The Economic Journal* 115.505, pp. 722–744.

Estrella, Arturo and Gikas A Hardouvelis (1991). "The term structure as a predictor of real economic activity". In: *The journal of Finance* 46.2, pp. 555–576.

Estrella, Arturo and Frederic S Mishkin (1996). "The yield curve as a predictor of US recessions". In: *Current issues in economics and finance* 2.7.

Estrella, Arturo, Anthony P Rodrigues, and Sebastian Schich (2003). "How stable is the predictive power of the yield curve? Evidence from Germany and the United States". In: *The review of Economics and Statistics* 85.3, pp. 629–644.

Estrella, Arturo and Mary Trubin (2006). "The yield curve as a leading indicator: Some practical issues". In:

Evans, Charles L and David A Marshall (1998). "Monetary policy and the term structure of nominal interest rates: evidence and theory". In: *Carnegie-Rochester Conference Series on Public Policy*. Vol. 49. Elsevier, pp. 53–111.

— (2007). "Economic determinants of the nominal treasury yield curve". In: *Journal of Monetary Economics* 54.7, pp. 1986–2003.

Fabozzi, Frank J, Lionel Martellini, and Philippe Priaulet (2006). *Advanced bond portfolio management: best practices in modeling and strategies*. Vol. 143. John Wiley & Sons.

Fed (2018). "The Federal Reserve System: Purposes & Functions". In: *Federal Reserve System Publication*.

Fletcher, TSB (2012). "Machine learning for financial market prediction". PhD thesis. UCL (University College London).

Fox, Anthony J (1972). "Outliers in time series". In: *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 350–363.

Frankel, Jeffrey, Sergio L Schmukler, and Luis Serven (2004). "Global transmission of interest rates: monetary independence and currency regime". In: *Journal of International Money and Finance* 23.5, pp. 701–733.

Frankel, Jeffrey A and Cara S Lown (1994). "An indicator of future inflation extracted from the steepness of the interest rate yield curve along its entire length". In: *The Quarterly Journal of Economics* 109.2, pp. 517–530.

Gadea, María Dolores, Ana Gómez-Loscos, and Antonio Montañés (2016). "Oil price and economic growth: A long story?" In: *Econometrics* 4.4, p. 41.

Gevers, Michel and Vincent Wertz (1984). "Uniquely identifiable state-space and ARMA parametrizations for multivariable linear systems". In: *Automatica* 20.3, pp. 333–347.

Giese, Julia (2008). "Level, slope, curvature: Characterising the yield curve in a cointegrated VAR model". In:

Golub, Bennett W and Leo M Tilman (1997a). "Measuring plausibility of hypothetical interest rate shocks". In: *Advanced Bond Portfolio Management: Best Practices in Modeling and Strategies*, pp. 247–266.

— (1997b). "Measuring Yield Curve Risk Using Principal Components, Analysis, Value, At Risk, And Key Rate Durations". In: *The Journal of Portfolio Management* 23.4, pp. 72–84.

— (2000). *Risk management: approaches for fixed income markets*. Vol. 73. John Wiley & Sons.

Gómez-Loscos, Ana, María Dolores Gadea, and Antonio Montañés (2012). "Economic growth, inflation and oil shocks: are the 1970s coming back?" In: *Applied Economics* 44.35, pp. 4575–4589.

Gómez-Loscos, Ana, Antonio Montañés, and M Dolores Gadea (2011). "The impact of oil shocks on the Spanish economy". In: *Energy Economics* 33.6, pp. 1070–1081.

Gonzalo, Jesus (1994). "Five alternative methods of estimating long-run equilibrium relationships". In: *Journal of econometrics* 60.1-2, pp. 203–233.

Gradojevic, Nikola and Jing Yang (2006). "Non-linear, non-parametric, non-fundamental exchange rate forecasting". In: *Journal of Forecasting* 25.4, pp. 227–245.

Granger, Clive WJ (1969). "Investigating causal relations by econometric models and cross-spectral methods". In: *Econometrica: Journal of the Econometric Society*, pp. 424–438.

— (1981). "Some properties of time series data and their use in econometric model specification". In: *Journal of econometrics* 16.1, pp. 121–130.

Greenberg, Edward (2012). *Introduction to Bayesian econometrics*. Cambridge University Press.

Gregory, Allan W and Bruce E Hansen (1996). "Residual-based tests for cointegration in models with regime shifts". In: *Journal of econometrics* 70.1, pp. 99–126.

Gregory, Allan W and Allen C Head (1999). "Common and country-specific fluctuations in productivity, investment, and the current account". In: *Journal of Monetary Economics* 44.3, pp. 423–451.

Grewal, Mohinder S (2011). *Kalman filtering*. Springer.

Guidolin, Massimo and Daniel L Thornton (2008). "Predictions of short-term rates and the expectations hypothesis of the term structure of interest rates". In: *ECB Working Paper Series No 977*.

Gürkaynak, Refet S, Brian Sack, and Jonathan H Wright (2010). "The TIPS yield curve and inflation compensation". In: *American Economic Journal: Macroeconomics* 2.1, pp. 70–92.

Hackl, Peter and Anders H Westlund (1989). "Statistical analysis of "structural change": An annotated bibliography". In: *Empirical Economics* 14.2, pp. 167–192.

Haldane, Andrew and Vicky Read (2000). "Monetary policy surprises and the yield curve". In:

Hall, Anthony D, Heather M Anderson, and Clive WJ Granger (1992). "A cointegration analysis of treasury bill yields". In: *The review of Economics and Statistics*, pp. 116–126.

Hamilton, James Douglas (1994). *Time series analysis*. Vol. 2. Princeton university press Princeton.

Hannan, EJ (1971). "The identification problem for multiple equation systems with moving average errors". In: *Econometrica: Journal of the Econometric Society*, pp. 751–765.

Hansen, Bruce E (1992). "Testing for parameter instability in linear models". In: *Journal of policy Modeling* 14.4, pp. 517–533.

— (1997). "Approximate asymptotic p values for structuras-change tests". In: *Journal of Business & Economic Statistics* 15.1, pp. 60–67.

— (2001). "The new econometrics of structural change: Dating breaks in US labor productivity". In: *The Journal of Economic Perspectives* 15.4, pp. 117–128.
Hansen, Peter Reinhard (2003). "Structural changes in the cointegrated vector autoregressive model". In: *Journal of Econometrics* 114.2, pp. 261–295.
Hanspeter, Sheller K (2004). "The European Central Bank–history, role and functions". In: *European Central Bank*.
Hanzon, Bernard (1993). "On the closure of several sets of ARMA and linear state space models with a given structure". In: *New directions in time series analysis*. Springer, pp. 239–253.
Harvey, Andrew C (1990). *Forecasting, structural time series models and the Kalman filter*. Cambridge university press.
— (1993). "Time series models". In:
Harvey, Andrew C and James Durbin (1986). "The effects of seat belt legislation on British road casualties: A case study in structural time series modelling". In: *Journal of the Royal Statistical Society. Series A (General)*, pp. 187–227.
Harvey, Andrew C and Siem Jan Koopman (1992). "Diagnostic checking of unobserved-components time series models". In: *Journal of Business & Economic Statistics* 10.4, pp. 377–389.
Haubrich, Joseph G, Ann M Dombrosky, et al. (1996). "Predicting real growth using the yield curve". In: *Economic Review* 32.1, pp. 26–35.
Heath, David, Robert Jarrow, and Andrew Morton (1992). "Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation". In: *Econometrica* 60, pp. 77–105.
Hellerstein, Rebecca (2011). "Global bond risk premiums". In:
Hevia, Constantino et al. (2015). "Estimating and forecasting the yield curve using a Markov switching dynamic Nelson and Siegel model". In: *Journal of Applied Econometrics* 30.6, pp. 987–1009.
Hinkley, David V (1970). "Inference about the change-point in a sequence of random variables". In:
Ho, Thomas SY (1992). "Key rate durations: Measures of interest rate risks". In: *The Journal of Fixed Income* 2.2, pp. 29–44.
Hoffman, K and R Kunze (1971). "Characteristic values in linear algebra". In: *Prentice-Hall, New Jersey*.
Hördahl, Peter, Oreste Tristani, and David Vestin (2006). "A joint econometric model of macroeconomic and term-structure dynamics". In: *Journal of Econometrics* 131.1-2, pp. 405–444.
Hu, Zuliu (1993). "The yield curve and real activity". In: *Staff Papers* 40.4, pp. 781–806.
Hull, John and Alan White (1990). "Pricing interest-rate-derivative securities". In: *The Review of Financial Studies* 3.4, pp. 573–592.
Jarrow, Robert A (2013). "The zero-lower bound on interest rates: Myth or reality?" In: *Finance Research Letters* 10.4, pp. 151–156.
Johansen, Søren (1988). "Statistical analysis of cointegration vectors". In: *Journal of economic dynamics and control* 12.2-3, pp. 231–254.
— (1991). "Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models". In: *Econometrica: Journal of the Econometric Society*, pp. 1551–1580.

Johansen, Søren (1992). "Cointegration in partial systems and the efficiency of single-equation analysis". In: *Journal of econometrics* 52.3, pp. 389–402.

— (1995). *Likelihood-based inference in cointegrated vector autoregressive models*. Oxford University Press on Demand.

Jong, Frank de (2000). "Time series and cross-section information in affine term-structure models". In: *Journal of Business & Economic Statistics* 18.3, pp. 300–314.

Jotikasthira, Chotibhak, Anh Le, and Christian Lundblad (2015). "Why do term structures in different currencies co-move?" In: *Journal of Financial Economics* 115.1, pp. 58–83.

Joyce, Michael AS, Peter Lildholdt, and Steffen Sorensen (2010). "Extracting inflation expectations and inflation risk premia from the term structure: a joint model of the UK nominal and real yield curves". In: *Journal of Banking & Finance* 34.2, pp. 281–294.

Justel, Ana, Daniel Peña, and Ruey S Tsay (2001). "Detection of outlier patches in autoregressive time series". In: *Statistica Sinica*, pp. 651–673.

Kalman, Rudolph Emil (1960). "A new approach to linear filtering and prediction problems". In: *Journal of basic Engineering* 82.1, pp. 35–45.

Kim, Chang-Jin, Charles R Nelson, et al. (1999). "State-space models with regime switching: classical and Gibbs-sampling approaches with applications". In: *MIT Press Books* 1.

Kim, Hyune-Ju and David Siegmund (1989). "The likelihood ratio test for a change-point in simple linear regression". In: *Biometrika* 76.3, pp. 409–423.

Koop, Gary, Dimitris Korobilis, et al. (2010). "Bayesian multivariate time series methods for empirical macroeconomics". In: *Foundations and Trends® in Econometrics* 3.4, pp. 267–358.

Koop, Gary and Simon Potter (2004). "Forecasting in dynamic factor models using Bayesian model averaging". In: *The Econometrics Journal* 7.2, pp. 550–565.

Koop, Gary M (2013). "Forecasting with medium and large Bayesian VARs". In: *Journal of Applied Econometrics* 28.2, pp. 177–203.

Kose, M Ayhan, Christopher Otrok, and Charles H Whiteman (2003). "International business cycles: World, region, and country-specific factors". In: *The American Economic Review* 93.4, pp. 1216–1239.

Kozicki, Sharon et al. (1997). "Predicting real growth and inflation with the yield spread". In: *Economic Review-Federal Reserve Bank of Kansas City* 82, pp. 39–58.

Kräussl, Roman, Thorsten Lehnert, and Kalle Rinne (2017). "The search for yield: Implications to alternative investments". In: *Journal of Empirical Finance* 44.

Krueger, Joel T and Kenneth N Kuttner (1996). "The fed funds futures rate as a predictor of Federal Reserve policy". In: *Journal of Futures Markets* 16.8, pp. 865–879.

Kurmann, André and Christopher Otrok (2013). "News shocks and the slope of the term structure of interest rates". In: *American Economic Review* 103.6, pp. 2612–32.

Kuttner, Kenneth N (2001). "Monetary policy surprises and interest rates: Evidence from the Fed funds futures market". In: *Journal of monetary economics* 47.3, pp. 523–544.

Lee, Thomas R. (2006). "Global bond investing for the 21st century". In: *Chapter 18 in Frank J. Fabozzi, Lionel Martellini, and Philippe Priaulet (eds), Advanced Bond Portfolio Management: Best Practices in Modeling and Strategies (Hoboken, NJ: John Wiley & Sons).*

Lefrancois, Bernard (1991). "Detecting over-influential observations in time series". In: *Biometrika* 78.1, pp. 91–99.

Leibowitz, Martin L and Roy D Henriksson (1988). "Portfolio optimization within a surplus framework". In: *Financial Analysts Journal* 44.2, pp. 43–51.

Levant, Jared and Jun Ma (2017). "A dynamic Nelson-Siegel yield curve model with Markov switching". In: *Economic Modelling* 67, pp. 73–87.

Litterman, Robert B (1986). "Forecasting with Bayesian vector autoregressions—five years of experience". In: *Journal of Business & Economic Statistics* 4.1, pp. 25–38.

Litterman, Robert B and Jose Scheinkman (1991). "Common factors affecting bond returns". In: *The Journal of Fixed Income* 1.1, pp. 54–61.

Litzenberger, R, G Squassi, and N Weir (1995). *Spline models of the term structure of interest rates and their applications*. Tech. rep. Working Paper, Goldman, Sachs and Company.

Ljung, Greta M and George EP Box (1978). "On a measure of lack of fit in time series models". In: *Biometrika* 65.2, pp. 297–303.

Lucas, André (1997). "Cointegration testing using pseudolikelihood ratio tests". In: *Econometric Theory* 13.2, pp. 149–169.

Lumsdaine, Robin L and Eswar S Prasad (2003). "Identifying the common component of international economic fluctuations: a new approach". In: *The Economic Journal* 113.484, pp. 101–127.

Lütkepohl, Helmut (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

Lütkepohl, Helmut and Markus Krätzig (2004). *Applied time series econometrics*. Cambridge university press.

Lütkepohl, Helmut and Hans-Eggert Reimers (1992). "Impulse response analysis of cointegrated systems". In: *Journal of economic dynamics and control* 16.1, pp. 53–78.

Malava, A (1999). "Principal Component Analysis on Term Structure of Interest Rates". In: *Helsinki University of Technology Department of Engineering Physics and Mathematics Working Paper*.

Marcus, Marvin and Henryk Minc (1988). *Introduction to linear algebra*. Courier Corporation.

Markovitz, H. (1952). "Portfolio selection". In: *The Journal of finance* 7.1, pp. 77–91.

Martellini, Lionel, Philippe Priaulet, and Stephane Priaulet (2003). "The Euro Benchmark Yield Curve: Principal Component Analysis of Yield Curve Dynamics". In: *in Frank J. Fabozzi (ed.), Professional Perspectives on Fixed Income Portfolio Management: Volume 4 (Hoboken, NJ: John Wiley & Sons)*.

Martellini, Lionel et al. (2006). "Hedging Interest Rate Risk with Term Structure Factor Models". In: *Chapter 11 in Frank J. Fabozzi, Lionel Martellini, and Philippe Priaulet (eds), Advanced Bond Portfolio Management: Best Practices in Modeling and Strategies (Hoboken, NJ: John Wiley & Sons)*.

McConnell, Margaret M and Gabriel Perez-Quiros (2000). "Output fluctuations in the United States: What has changed since the early 1980's?" In: *American Economic Review* 90.5, pp. 1464–1476.

McNeil, Alexander J and Rüdiger Frey (2000). "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach". In: *Journal of empirical finance* 7.3-4, pp. 271–300.

Mittnik, Stefan (1989). "Multivariate time series analysis with state space models". In: *Computers & Mathematics with Applications* 17.8, pp. 1189–1201.

— (1990). "Forecasting with balanced state space representations of multivariate distributed lag models". In: *Journal of Forecasting* 9.3, pp. 207–218.

Mönch, Emanuel (2012). "Term structure surprises: the predictive content of curvature, level, and slope". In: *Journal of Applied Econometrics* 27.4, pp. 574–602.

Muller, Philippe, Mark Zelmer, et al. (1999). *Greater transparency in monetary policy: impact on financial markets*. 86. Citeseer.

Mulvey, John M and Stavros A Zenios (1994). "Capturing the correlations of fixed-income instruments". In: *Management Science* 40.10, pp. 1329–1342.

Nelson, C and A Siegel (1987). "Common factors affecting bond returns". In: *Journal of Business*.

Ng, Serena and Pierre Perron (1995). "Unit root tests in ARMA models with data-dependent methods for the selection of the truncation lag". In: *Journal of the American Statistical Association* 90.429, pp. 268–281.

Nunes, Manuel et al. (2018). "Artificial Neural Networks in Fixed Income Markets for Yield Curve Forecasting". In:

Peel, David A and Mark P Taylor (1998). "The slope of the yield curve and real economic activity: tracing the transmission mechanism". In: *Economics Letters* 59.3, pp. 353–360.

Penzer, Jeremy (2007). "State space models for time series with patches of unusual observations". In: *Journal of Time Series Analysis* 28.5, pp. 629–645.

Perez-Quiros, Gabriel and Jorge Sicilia (2002). "Is the European Central Bank (and the United States Federal Reserve) predictable?" In:

Perron, Pierre, Jing Zhou, et al. (2008). "Testing jointly for structural changes in the error variance and coefficients of a linear regression model". In: *Unpublished Manuscript, Department of Economics, Boston University*.

Perron, Pierre et al. (2006). "Dealing with structural breaks". In: *Palgrave handbook of econometrics* 1.2, pp. 278–352.

Phoa, Wesley (2000). "Yield curve risk factors: domestic and global contexts". In: *The Professional's Handbook of Financial Risk Management. Oxford: Butterworth-Heinemann*, pp. 155–184.

Piazzesi, Monika (2001). *An econometric model of the yield curve with macroeconomic jump effects*. Tech. rep. National Bureau of Economic Research.

Picard, Dominique (1985). "Testing and estimating change-points in time series". In: *Advances in applied probability* 17.4, pp. 841–867.

Proietti, Tommaso (2003). "LEAVE-K-OUT DIAGNOSTICS IN STATE-SPACE MODELS". In: *Journal of Time Series Analysis* 24.2, pp. 221–236.

Qu, Zhongjun and Pierre Perron (2007). "Estimating and testing structural changes in multivariate regressions". In: *Econometrica* 75.2, pp. 459–502.

Quandt, Richard E (1960). "Tests of the hypothesis that a linear regression system obeys two separate regimes". In: *Journal of the American statistical Association* 55.290, pp. 324–330.

Rachev, Svetlozar T et al. (2007). *Financial econometrics: from basics to advanced modeling techniques*. Vol. 150. John Wiley & Sons.

Rapach, David E and Mark E Wohar (2005). "Regime changes in international real interest rates: Are they a monetary phenomenon?" In: *Journal of Money, Credit and Banking*, pp. 887–906.

Rey, Hélène (2016). "International channels of transmission of monetary policy and the mundellian trilemma". In: *IMF Economic Review* 64.1, pp. 6–35.

Ribarits, Thomas and Bernard Hanzon (2006). "On a New Approach to Cointegration–The State-Space Error Correction Model". In:

— (2014a). "The state-space error correction model: Definition, estimation and model selection". In:

— (2014b). "The State-Space Error Correction Model: Simulations and Applications". In:

Rodriguez, Carlos, Carlos A Carrasco, et al. (2014). *ECB Policy Responses between 2007 and 2014: a chronological analysis and a money quantity assessment of their effects*. Tech. rep.

Roley, V Vance, Gordon Sellon, et al. (1995). "Monetary policy actions and long-term interest rates". In: *Federal Reserve Bank of Kansas City Economic Quarterly* 80.4, pp. 77–89.

Rothenberg, Thomas J (1971). "Identification in parametric models". In: *Econometrica: Journal of the Econometric Society*, pp. 577–591.

Rudebusch, Glenn D (2010). "Macro-finance models of interest rates and the economy". In: *The Manchester School* 78, pp. 25–52.

Rudebusch, Glenn D and John C Williams (2009). "Forecasting recessions: the puzzle of the enduring power of the yield curve". In: *Journal of Business & Economic Statistics* 27.4, pp. 492–503.

Rudebusch, Glenn D and Tao Wu (2008). "A macro-finance model of the term structure, monetary policy and the economy". In: *The Economic Journal* 118.530, pp. 906–926.

Schnorrenberger, Richard et al. (2017). "Fixed-income portfolio optimization based on dynamic Nelson-Siegel models with macroeconomic factors for the Brazilian yield curve". In:

Schrimpf, Andreas and Qingwei Wang (2010). "A reappraisal of the leading indicator properties of the yield curve under structural instability". In: *International Journal of Forecasting* 26.4, pp. 836–857.

Schumacher, Michael P., Daniel C. Dektar, and Frank J. Fabozzi (1994). "Yield Curve Risk of CMO Bonds". In: *in Frank J. Fabozzi (ed.), CMO Portfolio Management (Hoboken, NJ: John Wiley & Sons)*.

Shea, Gary S (1992). "Benchmarking the expectations hypothesis of the interest-rate term structure: An analysis of cointegration vectors". In: *Journal of Business & Economic Statistics* 10.3, pp. 347–366.

Shumway, Robert H and David S Stoffer (2000). "Time series analysis and its applications". In: *Studies In Informatics And Control* 9.4, pp. 375–376.

Siegmund, David (1988). "Confidence sets in change-point problems". In: *International Statistical Review/Revue Internationale de Statistique*, pp. 31–48.

Sims, Christopher A and Tao Zha (1998). "Bayesian methods for dynamic multivariate models". In: *International Economic Review*, pp. 949–968.

Solnik, Bruno H (1974). "An equilibrium model of the international capital market". In: *Journal of economic theory* 8.4, pp. 500–524.

Sorge, Marco (2004). "Stress-testing financial systems: an overview of current methodologies". In:

Spencer, Peter and Zhuoshi Liu (2010). "An open-economy macro-finance model of international interdependence: The OECD, US and the UK". In: *Journal of banking & finance* 34.3, pp. 667–680.

Stock, James H and Mark W Watson (1988). "Testing for common trends". In: *Journal of the American statistical Association* 83.404, pp. 1097–1107.

— (1989). "New indexes of coincident and leading economic indicators". In: *NBER macroeconomics annual* 4, pp. 351–394.

— (1996). "Evidence on structural instability in macroeconomic time series relations". In: *Journal of Business & Economic Statistics* 14.1, pp. 11–30.

— (1998). *Diffusion indexes*. Tech. rep. National bureau of economic research.

— (2002). "Forecasting using principal components from a large number of predictors". In: *Journal of the American statistical association* 97.460, pp. 1167–1179.

— (2005). "Understanding changes in international business cycle dynamics". In: *Journal of the European Economic Association* 3.5, pp. 968–1006.

Thoms, SH (1993). "An international CAPM for bonds and equities". In: *Journal of International Money and Finance* 12.4, pp. 390–412.

Toda, Hiro Y and Peter CB Phillips (1993). "Vector autoregressions and causality". In: *Econometrica: Journal of the Econometric Society*, pp. 1367–1393.

Toda, Hiro Y and Taku Yamamoto (1995). "Statistical inference in vector autoregressions with possibly integrated processes". In: *Journal of econometrics* 66.1, pp. 225–250.

Tsay, Ruey S (1986). "Time series model specification in the presence of outliers". In: *Journal of the American Statistical Association* 81.393, pp. 132–141.

— (2005). *Analysis of financial time series*. Vol. 543. John Wiley & Sons.

Vasicek, Oldrich (1977). "An equilibrium characterization of the term structure". In: *Journal of financial economics* 5.2, pp. 177–188.

Verhelst, Stijn (2014). "All monetary policy has become 'unconventional'. Egmont Commentary, 4 June 2014". In:

Vui, Chang Sim et al. (2013). "A review of stock market prediction with Artificial neural network (ANN)". In: *2013 IEEE International Conference on Control System, Computing and Engineering*. IEEE, pp. 477–482.

Wagner, Martin (2010). "Cointegration analysis with state space models". In: *AStA Advances in Statistical Analysis* 94.3, pp. 273–305.

Wall, Kent D (1987). "IDENTIFICATION THEORY FOR VARYING COEFFICIENT REGRESSION MODELS 1". In: *Journal of Time Series Analysis* 8.3, pp. 359–371.

Wang, Dan and Jie Huang (2005). "Neural network-based adaptive dynamic surface control for a class of uncertain nonlinear systems in strict-feedback form". In: *IEEE Transactions on Neural Networks* 16.1, pp. 195–202.

Wang, Zijun, Jian Yang, and Qi Li (2007). "Interest rate linkages in the Eurocurrency market: Contemporaneous and out-of-sample Granger causality tests". In: *Journal of International Money and Finance* 26.1, pp. 86–103.

Wellmann, Dennis and Stefan Trück (2018). "Factors of the term structure of sovereign yield spreads". In: *Journal of International Money and Finance* 81, pp. 56–75.

West, Mike and Jeff Harrison (2006). *Bayesian forecasting and dynamic models*. Springer Science & Business Media.

Wilhelmsen, Bjørn-Roger and Andrea Zaghini (2011). "Monetary policy predictability in the euro area: an international comparison". In: *Applied Economics* 43.20, pp. 2533–2544.

Wold, Herman (1938). "A study in the analysis of stationary time series". PhD thesis. Almqvist & Wiksell.

Wright, Jonathan H (2006). "The yield curve and predicting recessions". In:

Wu, Tao (2001). "Monetary policy and the slope factor in empirical term structure estimations". In:

— (2003). "What makes the yield curve move?" In: *FRBSF Economic Letter* jun6.

Wyplosz, Charles (2013). "Non-Standard Monetary Policy Measures-An Update". In: *Non-Standard Monetary Policy Measures-An Update*, p. 7.

Xiang, Ju and Xiaoneng Zhu (2013). "A regime-switching Nelson–Siegel term structure model and interest rate forecasts". In: *Journal of Financial Econometrics* 11.3, pp. 522–555.

Yao, Yi-Ching et al. (1987). "Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables". In: *The Annals of Statistics* 15.3, pp. 1321–1328.

Zaloom, Caitlin (2009). "How to read the future: the yield curve, affect, and financial prediction". In: *Public Culture* 21.2, pp. 245–268.

Zenios, Stavros A et al. (1998). "Dynamic models for fixed-income portfolio management under uncertainty". In: *Journal of Economic Dynamics and Control* 22.10, pp. 1517–1541.

Zhang, Hua (1993). "Treasury yield curves and cointegration". In: *Applied Economics* 25.3, pp. 361–367.

Zhu, Xiaoneng and Shahidur Rahman (2015). "A regime-switching Nelson–Siegel term structure model of the macroeconomy". In: *Journal of Macroeconomics* 44, pp. 1–17.

Eidesstattliche Versicherung

(Siehe Promotionsordnung vom 12.07.11, § 8 Abs. 2 Pkt. 5)

Hiermit erkläre ich an Eidesstatt, dass die Dissertation von mir selbstständig, ohne unerlaubte Beihilfe angefertigt ist.

München, den 28. November 2019

Maria Sprincenatu