ESSAYS ON GOVERNMENT GROWTH, FISCAL POLICY AND DEBT SUSTAINABILITY

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften des Fachbereichs Wirtschaftswissenschaften der Universität Osnabrück

vorgelegt von

Jan Kuckuck

January 2014

Dekan: Prof. Dr. Thomas Gaube

Referenten: Prof. Frank Westermann, Ph.D. und Prof. Dr. Thomas Gaube

Tag der Disputation: 03.12.2014

Abstract

The financial crisis of 2007/8 has triggered a profound debate about public budget finance sustainability, ever-increasing government expenditures and the efficiency of fiscal policy measures. Given this context, the following dissertation provides four contributions that analyze the long-run growth of government spending throughout economic development, discuss potential effects of fiscal policy measures on output, and provide new insights into the assessment of debt sustainability for a variety of industrialized countries.

Since the breakout of the European debt crisis in 2009/2010, there has been a revival of interest in the long-term growth of government expenditures. In this context, the relationship between the size of the public sector and economic growth - often referred to as Wagner's law - has been in the focus of numerous studies, especially with regard to public policy and fiscal sustainability. Using historical data from the mid-19th century, the first chapter analyzes the validity of Wagner's law for five industrialized European countries and links the discussion to different stages of economic development. In line with Wagner's hypothesis, our findings show that the relationship between public spending and economic growth has weakened at an advanced stage of development. Furthermore, all countries under review support the notion that Wagner's law may have lost its economic relevance in recent decades.

As a consequence of the 2007/8 financial crisis, there has been an increasing theoretical and empirical debate about the impact of fiscal policy measures on output. Accordingly, the Structural Vector Autoregression (SVAR) approach to estimating the fiscal multipliers developed by Blanchard and Perotti (2002) has been applied widely in the literature in recent years. In the second chapter, we point out that the fiscal multipliers derived from this approach include the predicted future path of the policy instruments as well as their dynamic interaction. We analyze a data set from the US and document that these interactions are economically and statistically significant. In a counterfactual simulation, we report fiscal multipliers that abstract from these dynamic responses. Furthermore, we use our estimates to analyze the recent fiscal stimulus of the American Recovery and Reinvestment Act (ARRA).

The third chapter contributes to the existing empirical literature on fiscal multipliers by applying a five-variable SVAR approach to a uniform data set for Belgium, France, Germany, and the United Kingdom. Besides studying the effects of expenditure and tax increases on output, we additionally analyze their dynamic effects on inflation and interest rates as well as the dynamic interaction of both policy instruments. By conducting counterfactual simulations, which abstract from the dynamic response of key macroeconomic variables to the initial fiscal shocks, we study the importance of these channels for the transmission of fiscal policy on output. Overall, the results demonstrate that the effects of fiscal shocks are limited and rather different across countries. Further, it is shown that the inflation and interest rate channel are insignificant for the transmission of fiscal policy.

In the field of public finances, governmental budgetary policies are among the most controversial and disputed areas of political and scientific controversy. The sustainability of public debt is often analyzed by testing stationarity conditions of government's budget deficits. The fourth chapter shows that this test can be implemented more effectively by means of an asymmetric unit root test. We argue that this approach increases the power of the test and reduces the likelihood of drawing false inferences. We illustrate this in an application to 14 countries of the European Monetary Union (EMU) as well as in a Monte Carlo simulation. Distinguishing between positive and negative changes in deficits, we find consistency with the intertemporal budget constraint for more countries, i.e. lower persistence of positive changes in some countries, compared to the earlier literature.

Keywords: Asymmetric unit roots, Cointegration, Domar model, Economic development, Euro area, Fiscal deficits, Fiscal multipliers, Fiscal policy, Government expenditures, Intertemporal budget constraint, Net revenues, Public debt, Structural breaks, Structural vector autoregression, Tax revenue elasticities, Threshold autoregressive process, United States, Vector error correction model, Wagner's law

JEL: C22, C24, C51, E62, H10, H20, H50, H53, H62, H63, N43, N44

Acknowledgment

During my dissertation process, I have benefited from the support of many people. First of all, I would like to express my very great appreciation to my mentor and doctoral advisor Frank Westermann for his excellent guidance, valuable suggestions, and enthusiastic encouragement during the planning and implementation of this research work. Furthermore, I would like to offer my special thanks to Thomas Gaube for his willingness to co-supervise this thesis.

My special thanks are extended to Joachim Wilde and the complete staff of the Institute of Empirical Economic Research as well as participants of the Lunch and Paper Seminar at the University of Osnabrück for helpful comments and recommendations. I owe many thanks to my direct colleagues at the Chair of International Economic Policy, especially Andreas Steiner and Sven Steinkamp, for helpful discussions and critical suggestions. I would also like to thank Katharina Diekmann for her continuous advice and assistance in proofreading and editing this dissertation. Also comments from participants at workshops and conferences are gratefully acknowledged.

Finally, I wish to thank my family and friends for their support, patience, and encouragement throughout my study.

Osnabrück, January 2014 *Jan Kuckuck*

Contents

List of Figures

List of Tables

List of Abbreviations

ADF	Augmented Dickey-Fuller
AIC	Akaike Information Criterion
AMECO	Annual Macro-Economic Database
AR	Autoregressive
ARIMA	Autoregressive Integrated Moving Average
ARRA	American Recovery and Reinvestment Act
AU	Austria
BE	Belgium
BEA	Bureau of Economic Analysis
BLS	Bureau of Labor Statistics
CA	Canada
CBO	Congressional Budget Office
CF	Counterfactual Multiplier
CH	Switzerland
CPI	Consumer Price Index
CV	Critical Value
DE	Germany
DH	Doorik-Hansen
DK	Denmark
DSGE	Dynamic Stochastic General Equilibrium
ECT	Error Correction Term
EMU	European Monetary Union

ES	Spain
ESA95	European System of Accounts
EU	European Union
EXP	Expenditure
EXPPC	Expenditure Per Capita
FC	Forecast Multiplier
FI	Finland
FPE	Final Prediction Error
FR	France
G7	Group of Seven
GDP	Gross Domestic Product
GDPPC	Gross Domestic Product Per Capita
GFS	Government Financial Statistics
GFSM	Government Financial Statistics Manual
GR	Greece
IBC	Intertemporal Budget Constraint
IE	Ireland
IFS	International Financial Statistics
IMF	International Monetary Fund
Int$	International Geary-Khamis Dollars
IT	Italy
IVA	Inventory Valuation Adjustment
JP	Japan
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
KR	South Korea
LIBOR	London Interbank Offered Rate
log	Logarithm
LP	Lumsdaine-Papell
LU	Luxembourg

MT	Malta
MTAR	Momentum Threshold Autoregressive
neg.	Negative
NL	Netherlands
NO	Norway
NZ	New Zealand
OECD	Organization for Economic Co-operation and Development
OLS	Ordinary Least Squares
pos.	Positive
PP	Phillips-Perron
Prob.	Probability
PT	Portugal
REV	Revenue
RSS	Residual Sum of Squares
SE	Sweden
SIC	Schwarz Information Criterion
SSC	Social Security Contributions
SupF	Supremum F-statistic
SVAR	Structural Vector Autoregression
TAR	Threshold Autoregressive
TH	Thailand
TW	Taiwan
UK	United Kingdom
US	United States
VAR	Vector Autoregression
VECM	Vector Error Correction Model
WWI	World War I
WWII	World War II
ZA	South Africa

Summary

This dissertation collects four essays on public finances in industrialized countries that apply econometric methods to various data sets and relate them to economic policy questions. The first chapter focuses on the long-run relationship between government growth and economic development from a historical perspective for five western European countries. The findings show that the relationship between government expenditures and economic growth weakens at an advanced stage of development and might have significant implications for the budgetary process of advanced industrialized countries. The second chapter analyzes the impact of fiscal policy measures on output using a data set of the United States. In particular, the interpretation of these estimates with regard to the counter financing of discretionary policy measures are discussed. Overall, the findings illustrate the limits of conventional SVAR estimates of fiscal multipliers for concrete policy advice. Based on this idea, the third chapter provides empirical evidence on fiscal multipliers for some European countries. In addition to the finding that the effects of fiscal shocks are limited and rather different across countries, it is shown that the inflation and interest rate channel are insignificant for the transmission of fiscal policy. Finally, the fourth chapter deals with

the empirical assessment of debt sustainability. In an application to 14 countries of the European Monetary Union (EMU) as well as in a Monte Carlo simulation it is illustrated that asymmetric unit root tests increase the power of the test and reduce the likelihood of drawing false inferences.

In terms of methodology, all chapters apply advanced time-series econometrics. Detailed data definitions and bibliographic references enable the reader to reproduce the studies in detail. Furthermore, every chapter provides in addition to the main analysis, extensive appendices with auxiliary information, regressions, and robustness checks.

The chapters presented in this dissertation are extended versions of research papers that have been published in parts in various sources. Chapter one is based on Kuckuck (2014), "Testing Wagner's Law at Different Stages of Economic Development", *FinanzArchiv/Public Finance Analysis 70(1):128-168* and Kuckuck (2012), "Testing Wagner's Law at Different Stages of Economic Development: A Historical Analysis of Five Western European Countries", *Institute of Empirical Economic Research, Working Paper No. 91, University of Osnabrück*. The second chapter is an extension of Kuckuck and Westermann (2014), "On the Size of Fiscal Multipliers: A Counterfactual Analysis", *Economics Letters 123(1):26-32* as well as Kuckuck and Westermann (2013), "On the Size of Fiscal Multipliers: A Counterfactual Analysis", *Institute of Empirical Economic Research, Working Paper No. 96, University of Osnabrück*.

In the following, I will give a more detailed summary of the four chapters of this dissertation.

The relationship between the size of the public sector and economic growth has been widely discussed in the literature. In this context, Wagner's law of increasing state activity has received much attention, postulating a positive correlation between economic development and government activity. According to the spirit of Wagner's hypothesis, an expanding government accompanies social progress and rising incomes. Still, the driving determinants of Wagner's law have changed throughout economic evolution. While during the 19^{th} century laissez-faire attitudes of governments continued to predominate, public spending was primarily driven by expenditures on military and basic infrastructure. Throughout the 20^{th} century driving forces of spending growth have developed into expenditures on basic welfare, public utilities, and education. Most recently, empirical studies show that Wagner's law primarily operates through demographic change, social welfare policies, and trade openness.

The **first chapter** focuses on the direct relationship between public spending and national income with regard to different epochs of economic development in five diverse European welfare states: United Kingdom, Denmark, Sweden, Finland, and Italy. By using historical data on government expenditure and GDP from the mid-19^{th} century, we classify every country into three individual stages of income and evaluate the validity of Wagner's law from the industrialization period until the present.

To test the hypothesis of a long-run relationship between income and government spending - which is in line with Wagner's interpretation that there is not necessarily a cause-and-effect relationship between the variables - we employ cointegration analysis and allow for possible structural breaks in the data series. To subsequently make a statement about the long-run causal relationship and the speed of adjustment of public spending to changes in economic growth, we estimate vector error correction models (VECM) and compare the results throughout countries and development stages.

In general, our findings exhibit that a long-run equilibrium between public spending and economic growth exists, independent of development stage or functional form. Furthermore, the hypothesis that economic growth is a driving force for government expenditure can be rejected at least for the period of the last 150 years. Nevertheless, the hypothesis that Wagner's law might have a higher validity during early stages of development turns out to be confirmed for the United Kingdom, Denmark, Sweden, and Finland. In addition, all countries under review support the notion that the causal relationship between economic growth and public spending in line with Wagner's hypothesis may have lost its economic relevance in recent decades.

In the spirit of Wagner's hypothesis, the weakened relationship between government expenditure and economic growth can be explained by the expanding role of governments associated with strong changes in the structure of the economy. Well-established welfare states like the United Kingdom, Denmark, Finland, Sweden, and Italy have past through those major structural changes in recent days. With regard to the sustainability of growing public debts, these signs of expenditure decoupling could have implications for the budgetary process of advanced industrialized countries. Nevertheless, it remains to be seen how future main drivers of spending growth, such as ageing and demographic change as well as health care will affect the prospective relationship between economic growth and public spending.

The SVAR approach to estimating the fiscal multipliers developed by Blanchard and Perotti (2002) has been applied widely in the literature in recent years. It was one of the first analysis that solved the identification problem associated with earlier stylized facts on the co-movement of spending, taxes, and income. In the **second chapter**, we point out that while the identification of shocks has been achieved, the approach still includes the predicted future path of the policy instruments as well as their dynamic interaction. The derived multipliers are therefore best characterized as forecasting multipliers because the response of output to fiscal shocks is derived under the assumption that following the initial policy intervention, all variables will behave as they have typically done in the past. In this chapter, we raise the question whether this assumption is reasonable, when using the results for policy advice, as a benchmark for DSGE modeling, or for testing the Keynesian model. We analyze a data set from the US and document that these interactions are economically and statistically significant. In a counterfactual simulation, we report fiscal multipliers that abstract from these dynamic responses. Furthermore, we use our estimates to analyze the recent fiscal stimulus of the American Recovery and Reinvestment Act (ARRA).

The analysis starts by illustrating that there exists a significant and economically sizeable effect of a shock in expenditure on net taxes and vice versa. The effect of a shock in expenditure on net taxes is positive, i.e. expenditures today tend to be financed by tax increases in the immediately following quarters. In the estimation of the spending multiplier, this will have a dampening effect on GDP. With regard to taxes, we have the opposite finding.

After a standard positive shock to net taxes, there is a significant response of expenditure that is negative. Furthermore both series are autocorrelated. A fiscal policy shock will lead to further changes in fiscal policy in the subsequent quarters.

In order to isolate the effects of a pure spending and pure tax shock, we implement the following counterfactual analysis: we first estimate the model using the Blanchard and Perotti (2002) approach. When computing the impulse response functions, however, we shut down the channel that captures the discretionary dynamic interaction among policy instruments as well as each policy instruments' autocorrelation (i.e. restrict their responses to zero). All other responses remain unrestricted. In particular, the indirect effect that government spending has on net taxes - via automatic stabilizers - remains included in the simulation.

The main result of our analysis is that our counterfactual multiplier is substantially larger than the forecasting multiplier from standard SVAR estimates in case of an expenditure shock. The tax multiplier is initially smaller but gets larger at longer horizons. Finally, when both spending and net taxes experience a shock at the same time, the counterfactual multiplier is close to one - as predicted by Haavelmo (1945) - while the forecast multiplier is nearly zero. The robustness of our findings is confirmed by a set of sensitivity tests.

Finally, we use our estimates to analyze the recent fiscal stimulus of the ARRA. In the context of our paper, it provides an interesting case study as the fiscal stimulus was intended to be predominately a reduction in net taxes, which received only little support in the expenditure side, and thus broke from previous paths. We find that in a period from 2009 to 2013, ARRA had on aggregate an overall impact on GDP of 510.68 billion

US-Dollars, which amounts to an average quarterly GDP growth of less than one percent.

As a consequence of the 2007/8 financial crisis and the global economic downturn, there has been an increasing theoretical and empirical debate about the impact of fiscal policy measures on output. Hence, the **third chapter** seeks to contribute to the existing empirical literature by applying a five-variable SVAR approach to a uniform data set for four European countries (Belgium, France, Germany, and the United Kingdom). Besides studying the effects of expenditure and tax increases on output, we additionally analyze their dynamic effects on inflation and interest rates as well as the dynamic interaction of both policy instruments. By conducting counterfactual simulations, which abstract from the dynamic response of key macroeconomic variables to the initial fiscal shock, we study the importance of these channels for the transmission of fiscal policy on output.

Empirical studies analyzing fiscal multipliers of more than one country typically derive the data from various sources that do not follow a uniform classification system. Our analysis benefits from the adoption of a common statistical standard in the EMU, namely the European System of Accounts (ESA95), which collects and classifies accrual fiscal data at quarterly frequencies. This uniform data set, comprising the period from 1991Q1 to 2011Q4, allows us to analyze the efficiency of fiscal policy in recent decades and to compare the effects between the various countries.

The standard Vector Autoregression (VAR) methodology estimates the fiscal multiplier under the assumption that following the initial policy intervention, all variables will behave as they have typically done in the past. In the counterfactual simulations, we assume that price levels, interest rates as well as net taxes (expenditures) are not affected by the initial spending (tax) shock. When computing the impulse response functions, we shut down the price and inflation channels as well as the discretionary interaction among policy instruments.

Our findings show that the effects of fiscal shocks are limited and different across countries. Merely expenditure shocks in France and Germany as well as tax shocks in Germany have a statistically significant effect on output. A one Euro increase in government spending raises GDP by around 1.05 Euros in France and 1.24 Euros in Germany. In this respect, public expenditure increases tend to be more efficient than tax cuts. In the case of Germany, a one Euro increase in net taxes leads to a decline in GDP by around 0.58 Euros. Further, we find that the fiscal multipliers do not always have the sign as predicted by economic theory, e.g. interest rates significantly decrease in response to an expenditure shock and increase in response to a tax shock across countries.

The results of the counterfactual exercise vary from country to country and show that the inflation and interest rate channel have negligible effects on the transmission of fiscal policy. While shutting down the interest channel eliminates the impact of expenditure in Belgium, it slightly strengthens the effect on GDP in the United Kingdom. Regarding the dynamic interaction of policy instruments, the effects of the counterfactual analysis are rather limited and without economic significance.

In the wake of the 2007/8 financial crisis, public budgetary policies are recently among the most controversial and disputed areas of political and scientific controversy. From an empirical perspective, the sustainability of fiscal policy is often analyzed by testing stationarity conditions of government's budget deficits. The empirical framework developed by Hamilton and Flavin (1986) is interested in the question whether the government's creditors could rationally expect that the government budget would be balanced in present-value terms. In an application to 14 countries of the European Monetary Union (EMU), **chapter 4** shows that symmetric unit root tests might lead to inconsistent results regarding debt sustainability and can be implemented more effectively by means of an asymmetric unit root test.

We start our analysis by estimating the unit root properties of government's budget deficits in two samples: a reduced sample that stops in 2007, the year before the crisis (1972-2008), and a sample that includes the post-financial crisis time period (1972-2011). The unit root results suggest that in the latter sample, including above average high deficit years, more countries follow a sustainable fiscal policy path than in a time period excluding the crisis. These contradicting findings can be explained by the fact that until the beginning of the financial crisis 2007/8, declining deficit ratios dominated the data-generating process in most countries. As a consequence of the economic downturn and the return of Keynesian fiscal policies, deficit ratios increased dramatically in 2009, 2010 and 2011. Hence, the unit root tests do not reject unit root behavior in the sample before the crisis - despite decreasing new indebtedness - and reject unit root behavior in the sample including the crisis. Symmetric unit root tests suffer from the fact that they do not distinguish between the asymmetric persistence of positive and negative adjustment paths. In fact, deficit time series with positive or negative stochastic trends are treated equally. However, from an empirical point of view, this specification might lead to erroneous inferences in terms of a balanced intertemporal budget.

In order to capture the asymmetric persistence in a series, we propose - based on Enders' and Granger's (1998) momentum threshold autoregressive model (MTAR) - an asymmetric unit root testing approach, analyzing positive and negative deficit changes separately. This asymmetric approach gives the possibility of testing fiscal sustainability more precisely by separating between the persistence of positive and negative deficit ratio changes. Further, a Monte Carlo simulation demonstrates that the asymmetric unit root approach increases the power of the test, especially in a setting with very small sample sizes. Now, fiscal policy is considered to be sustainable as long as the deficit series is either global stationary or positive deficit changes follow a mean reverting pattern regardless whether the deficit series has a global unit root.

Distinguishing between positive and negative changes in deficits, we find consistency with the intertemporal budget constraint for more countries, i.e. lower persistence of positive changes in some countries compared to the earlier literature. Overall, these results can be explained by the successive cutback of deficit ratios until the turn of the century. This downward sloping curve of deficit series might give an alternative explanation why prior studies on fiscal sustainability, employing unit root or cointegration approaches, commonly do not attest sustainability for the majority of European countries even though this does not appear to be compatible with the introduction of fiscal policy rules, the Maastricht Treaty, and the Stability and Growth Pact.

1 Testing Wagner's law at different stages of economic development[1]

1.1 Introduction

The current European debt crisis has triggered an extensive debate about public budget finance sustainability, ever-increasing government spending, and institution-grounded expenditure rules (see, e.g., the modified European Growth and Stability Pact). In this context, the relationship between the size of the public sector and economic growth - often referred to as Wagner's law - is an important issue, especially with regard to public policy and fiscal sustainability (see Koester and Priesmeier (2013)). The purpose of our study is to analyze the positive relationship between economic development and the scope of government from an empirical perspective for five European countries. Our analysis differs essentially from others by the fact that we investigate long historical time series starting from the mid-19th century and specifically examine different stages of economic develop-

[1]This chapter is based on Kuckuck (2014, 2012).

ment in different countries.[2]

The expansion of the public sector with an ongoing economic development has become a widely accepted stylized fact. In this context, Wagner's law of increasing state activity has received much attention, postulating a positive correlation between economic development and government activity. Wagner explains this nexus with an ongoing "cultural and economic progress" (Wagner (1893), p. 908), which substitutes private economic activity for state activity, increases public cultural and welfare expenditures, and requires the pubic sector to manage and finance natural monopolies. According to the spirit of Wagner's law, an expanding government accompanies social progress and rising incomes.[3]

As a consequence of ongoing economic development, changes in public expenditure components are followed by changes in attitudes towards the role of the state and changes in the institutions that constrain government intervention in the economy (see Tanzi and Schuknecht (2000)). The driving determinants of Wagner's law have changed throughout economic evolution.[4] While during the 19^{th} century laissez-faire attitudes of governments continued to predominate, public spending was primarily driven by expenditures on military and basic infrastructure. Throughout the 20^{th} century the driving forces of spending

[2]Another very interesting link is the discussion on government size and economic growth - the Wagner's law relationship the other way around. An empirical investigation on the optimal size of the state for European countries is provided by Forte and Magazzino (2011). In addition, Bergh and Henrekson (2011) provide an exhaustive survey.

[3]Budget-maximization models on government growth argue that governments will spend as much as it can collect from its citizen (see Brennan and Buchanan (1980)). Wagner's law might also be interpreted as an indirect tax-and-spend hypothesis, as public revenues increase with national income. Holcombe (2005) reasons that governments' growth seems to have been constrained in the past primarily by their ability to raise revenues.

[4]Wagner (1893) explains the direct linkage between public spending and economic growth through permanent changes in the structure of the economy that are associated with new social activities of the state.

growth have expanded into expenditures on basic welfare, public utilities, and education.[5] Most recently, empirical studies have shown that Wagner's law primarily operates through demographic change (see Durevall and Henrekson (2011)), social welfare policies (see Lee and Chang (2006)), and trade openness (see Sobhee and Joysuree (2004)). In addition, various empirical contributions demonstrate that the development of government spending growth is dependent on various country-specific determinants like country size (see Alesina and Wacziarg (1998)), population density (see Dao (1995)), business cycle volatility (see Andres et al. (2008)), electoral systems (see Milesi-Ferretti et al. (2002)), periods of major social disturbances (see Peacock and Wiseman (1961)), and unbalanced sectoral growth (see Baumol (1967)). Shelton (2007) provides a common empirical framework that tests several leading hypotheses on determinants of government expenditure.

This study focuses on the direct relationship between public spending and national income with regard to different epochs of economic development in five diverse European welfare states: the United Kingdom, Denmark, Sweden, Finland, and Italy. By using historical data on government expenditure and GDP from the mid-19^{th} century, we classify every country into three individual stages of income development and evaluate the validity of Wagner's law from the industrialization period until the present. This data classification step allows us to analyze and compare the dynamics of Wagner's law at different stages of economic development from a within-country perspective and additionally enables us to identify commonalities across countries despite differences in size, development pattern,

[5]See Tanzi and Schuknecht (2000) as well as Peltzman (1980) for the growth of governments in the 20^{th} century.

and individual economic and social characteristics.[6]

To test the hypothesis of a long-run relationship between income and government spending that is in line with Wagner's stipulation that there is not necessarily a cause-and-effect relationship between the variables, we employ cointegration analysis and allow for possible structural breaks in the data series. To subsequently make a statement about the long-run causal relationship and the speed of adjustment of public spending to changes in economic growth, we estimate vector error correction models (VECMs) and compare the results throughout countries and development stages.

In general, our findings exhibit that a long-run equilibrium between public spending and economic growth exists, independently of development stage or functional form. Furthermore, the hypothesis that economic growth is a driving force for government expenditure can be rejected at least for the period of the last 150 years. Nevertheless, the hypothesis that Wagner's law might have a higher validity during early stages of development turns out to be confirmed for the United Kingdom, Denmark, Sweden, and Finland. Regarding the classic relationship between public expenditure and GDP, our findings show a decreasing error correction mechanism for expenditure from the first to the last development stage in the United Kingdom(-0.437 to 0.055), Denmark (-0.358 to -0.021), Sweden (-1.554 to 0.052), and Finland (-0.349 to -0.089). In the case of Italy, statistical causality is only

[6]The vast majority of studies focus either on emerging or on industrialized countries in order to make a statement about the relation between development level and Wagner's law (see, e.g., Wu et al. (2010); Chang (2002)). In addition, many of the low and middle income countries under review do not satisfy the requirements of Wagner's definition of a "culture and welfare state", which postulates the development tendency of the public sector explicitly for modern "constitutional and welfare states" (Wagner (1911), p. 743). It remains a matter of doubt if developing countries fulfill these requirements. Studies from Kuznets (1958) and Morris and Adelman (1989) show that there are significant differences between modern states in the 19^{th} century and recent developing countries.

detected in the last development stage with an expenditure adjustment of -0.057 having almost no economic significance. In general, all countries under review support the notion that the causal relationship between economic growth and public spending in line with Wagner's hypothesis may have lost its economic relevance in recent decades.

The remainder of this chapter is organized as follows. The next section reports and discusses some empirical evidence regarding Wagner's law and provides a survey of the empirical literature relying on time-series methods. Section 1.3 describes the historical development of national income and government expenditure and classifies each country into three stages of development. The subsequent section 1.4 presents the analytic framework and data. Cointegration results are displayed in section 1.5, and section 1.6 presents the long-run causal relationship and the adjustment speed of public spending to changes in economic growth. Section 1.7 deals with robustness checks and provides some alternative sample estimations. Concluding remarks are offered in Section 1.8.

1.2 Empirical evidence on Wagner's law

The empirical assessment of Wagner's law has focused on the relationship between government spending and national income in cross-country (see, e.g., Akitoby et al. (2006)), panel data (see, e.g., Lamartina and Zaghini (2011)), and time-series approaches (see, e.g., Magazzino (2012b)). According to a recent review by Durevall and Henrekson (2011), around 35% of these studies fail to find evidence for Wagner's law, while around 30% provide indirect support by controlling for other variables or focusing on specific expenditure components, and around 35% provide direct support.

The origins of time-series studies date back to the early 1970s when seminal contributions by Peacock and Wiseman (1961) as well as Timm (1961) spawned interest in the long-run relationship between economic growth and government expenditure. Since then, a mass of empirical articles has appeared, which provide mixed and partially contradictory results. These conflicting findings have been attributed to differences in econometric methodologies, country-specific characteristics, and investigated time periods. The early strand of literature was dominated by cross-sectional analyses (see, e.g., Gupta (1969); Gandhi (1971)), but the majority of studies have switched to a classic time-series framework since the cointegration revolution at the beginning of the 1990s (see, e.g., Henrekson (1993)). This approach is straightforward when considering Wagner's law. Cross-sectional analyses are not able to appropriately capture the dynamics of various macroeconomic variables and additionally do not allow one to control for country-specific effects in an adequate way.

A summary of several previous studies in mainly industrialized countries relying on time-series methods since 1990 is reported in table 1.1.[7] Generally speaking, Wagner's hypothesis receives considerable support, with few expectations. However, only few analyses investigate Wagner's law in a historical context. Oxley (1994) for the United Kingdom, Thornton (1999) for Denmark, Germany, Italy, Norway, Sweden, and the United Kingdom as well as Durevall and Henrekson (2011) for Sweden and the United Kingdom confirm the validity of Wagner's law in a time period some 40 to 50 years preceding World War I. In contrast,

[7]Note that table 1.1 does not include any panel data studies as in Brückner et al. (2012), Jaen-Garcia (2011), or Lamartina and Zaghini (2011).

studies that analyze a very long time span tend to reject Wagner's hypothesis. Henrekson (1993) and Bohl (1996) find no support for Wagner's law in Sweden from 1861 to 1990 and the United Kingdom from 1870 to 1995, respectively; Ghate and Zak (2002) do not find any empirical evidence in the United States from 1929 to 2000 and Durevall and Henrekson (2011) only find direct evidence in favor of Wagner's law for Sweden and the United Kingdom in a time period from around 1860 to 1970.

Table 1.1: Some existing empirical evidence, based on time-series methods, regarding Wagner's law in industrialized countries

Author and date	Countries (Period)	Empirical method	Major findings
Gyles (1991)	UK (1946-1985)	Time-domain transfer functions	General support for Wagner's law.
Henrekson (1993)	SE (1861-1990)	Cointegration analysis	No support for Wagner's law.
Courakis et al. (1993)	GR, PT (1958-1985)	OLS regression and hypothesis testing	Ambiguous support for Wagner's law; in GR Wagner's law only holds for transfer expenditure and in PT only for government consumption.
Oxley (1994)	UK (1870-1913)	Cointegration, short-run causality	General support for Wagner's law.
Hondroyiannis and Papapetrou (1995)	GR (1951-1992)	Cointegration, strong causality	No evidence for a long-run relationship between government spending and income.
Ahsan et al. (1996)	CA (1952-1988)	Cointegration analysis	Existence of a cointegrated relationship between income and public spending or the size of the public sector.
Bohl (1996)	G7 countries (1950-1996), DE (1850-1913), UK (1870-1995)	Cointegration, strong causality	Wagner's law only holds for CA and the UK in the post-Word War II period.
Hayo (1996)	DE (1960Q1-1993Q4)	Cointegration analysis	General support for Wagner's law.
Payne and Ewing (1996)	22 countries (AU, CA, FI, DE, GR, IT, JP, SE, CH, UK, US) (1948-1994)	Cointegration, strong causality	Unidirectional causality in line with Wagner's law is found for AU and DE; bidirectional causality exists for SE, CH, UK, and US.
Chletsos and Kollias (1997)	GR (1958-1993)	Cointegration, long-run causality	Wagner's hypothesis is only valid in the case of defense expenditures.
Thornton (1999)	DK, DE, IT, NO, SE, UK (1850-1913)	Cointegration, short-run and long-run causality	Considerable support for Wagner's law.
Biswal et al. (1999)	CA (1950-1995)	Cointegration, short-run and long-run causality	General support for Wagner's law at aggregate level, results for disaggregate expenditure components suggest only short-run causality in the sense of Wagner.
Kolluri et al. (2000)	G7 countries (1960-1993)	Cointegration, short-run and long-run causality	General support for the traditional versions of WagnerŠs Law.
Islam (2001)	US (1929-1996)	Cointegration, short-run and long-run causality	General support for Wagner's hypothesis.
Ghate and Zak (2002)	US (1929-2000)	Cointegration analysis	No statistical support for a long-run relationship between real government expenditures and GDP.
Chang (2002)	JP, KR, TW, TH, UK, US (1951-1996)	Cointegration, strong causality	Wagner's law holds for the selected countries with exception of TH.
Chow et al. (2002)	UK (1948-1997)	Cointegration, short-run and long-run causality	No direct support for Wagner law; indirect support for Wagner's law controlling for money supply.

Table is continued on the next page.

Table 1.1 – continued

Author and date	**Countries (Period)**	**Empirical method**	**Major findings**
Karagianni et al. (2002)	EU-15 (1949-1998)	Cointegration, short-run causality	Support for Wagner's law in the vast majority of EU countries; only GR does not exhibit a causal relationship between public spending and economic growth.
Legrenzi and Milas (2002b)	IT (1959-1996)	Cointegration analysis, persistence profile analysis	No direct support for Wagner's law; indirect support for Wagner's law controlling for a supply-side variable, bureaucratic power, and an institutional factor that captures the division of competencies between local and central government in allocating public expenditure.
Chang et al. (2004)	AU, CA, JP, KR, NZ, ZA, TW, TH, UK, US (1951-1996)	Cointegration, strong causality	Findings for KR, TH, JP, UK, and US support Wagner's law; no support could be found for AU, CA, NZ, ZA, and TW.
Dritsakis and Adamopoulos (2004)	GR (1960-2001)	Cointegration and short-run causality	General support for Wagner's law at aggregate and disaggregate levels.
Loizides and Vamvoukas (2005)	GR, IE, UK (1948-1995)	Cointegration, short-run and long-run causality	Direct support for Wagner's law only in GR; UK provides indirect support for Wagner's law controlling for inflation, results for IE do not indicate any Wagnerian causality effect.
Karagianni and Pempetzoglou (2009)	15 EU countries (1949-1998)	Non-linear Granger causality	Patterns of causality between income and government expenditure display dramatic differences across various countries. However, empirical findings in most countries seem to be favorable to Wagner's hypothesis.
Durevall and Henrekson (2011)	SE (1800-2006), UK (1830-2006)	Cointegration analysis	Wagner's law holds some 40-50 years preceding WWI and a period of 30-35 years after WWII. In more recent decades this relationship only can be maintained controlling for the age structure.
Magazzino (2011)	IT (1990-2010)	Cointegration, short-run causality	Considerable support for Wagner's law at disaggregated spending level; only two out of ten spending series do not share a common trend with real aggregate income.
Magazzino (2012a)	IT (1960-2008)	Cointegration, short-run and long-run causality	Only weak support for Wagner's law; Granger causality tests show evidence only for passive interests spending in the long-run.
Magazzino (2012b)	EU-27 (1970-2009)	Cointegration, short-run causality	Empirical evidence seems to be most favorable to Wagner's hypothesis.
Kumar et al. (2012)	NZ (1960-2007)	Cointegration, short-run and long-run causality	Long-run results exhibit statistically significant evidence in favor of Wagner's law.
Koester and Priesmeier (2013)	DE (1960-2007)	Cointegration analysis between expenditure, revenue, and GDP	Cointegration coefficient estimates between public spending and GDP provide strong evidence in favor of Wagner's law.

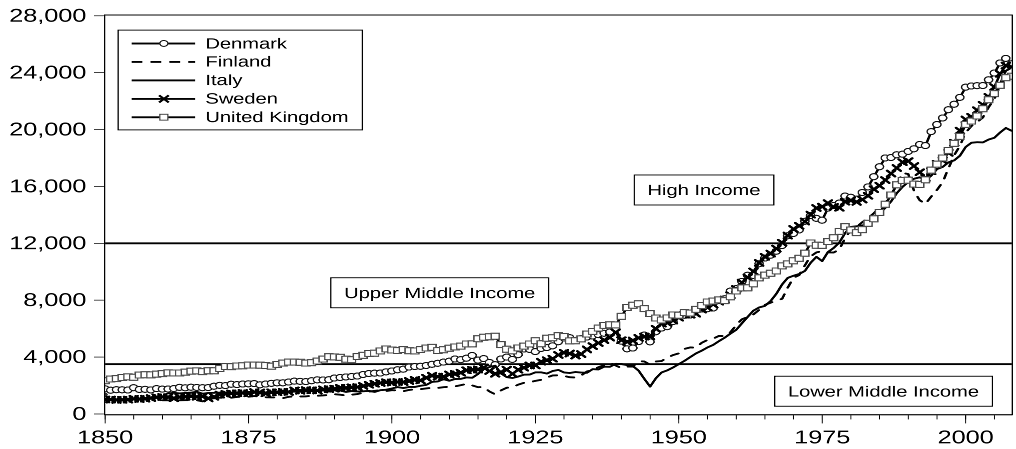
Note: The table reviews all published empirical studies since 1990 that largely cover industrialized countries and apply time-series methods (updated to December 2013).

1.3 Wagner's law and economic development in the 20^{th} century

In general, Wagner's formulations offer three reason for the direct linkage between economic growth and government activity: (i) changes in the structure of the economy associated with new social activities of the state, (ii) increasing administrative and protective functions substituting private for public actions, and (iii) increasing control of externalities and welfare aspects. As mentioned by Timm (1961), Wagner's hypothesis was conceived as applicable to countries throughout the 19^{th} century, beginning with the industrial revolution. Although Wagner suggested that his law would be operative as long there exists cultural and economic progress, his elaborations assume that the changing role of governments is contingent on the development stage of the economy; that the public expenditures of well-established welfare states should not react to changes in income in the same manner as in emerging states that have just started to respond to the challenges induced by increasing prosperity. This implies that according to Wagner's hypothesis the direct linkage between increasing state activity and economic growth might have a higher validity during early stages of development than at a later stage.[8] In order to shed light on the relation between Wagner's law and development stage, we analyze five advanced Western European countries that can be regarded from an income perspective as equally developed at present. Figure 1.1 shows that the GDP per capita in 2008 for the United Kingdom, Denmark, Swe-

[8] A recent study by Lamartina and Zaghini (2011) embracing 23 OECD countries supports this view. The authors find that the correlation between government activity and economic growth is higher in countries with lower per-capita GDP, suggesting that the catching-up period is characterized by a stronger development of government activity than for more advanced economies.

Figure 1.1: Development of GDP per capita in 1990 International Geary-Khamis dollars



Source: Bolt and van Zanden (2013)

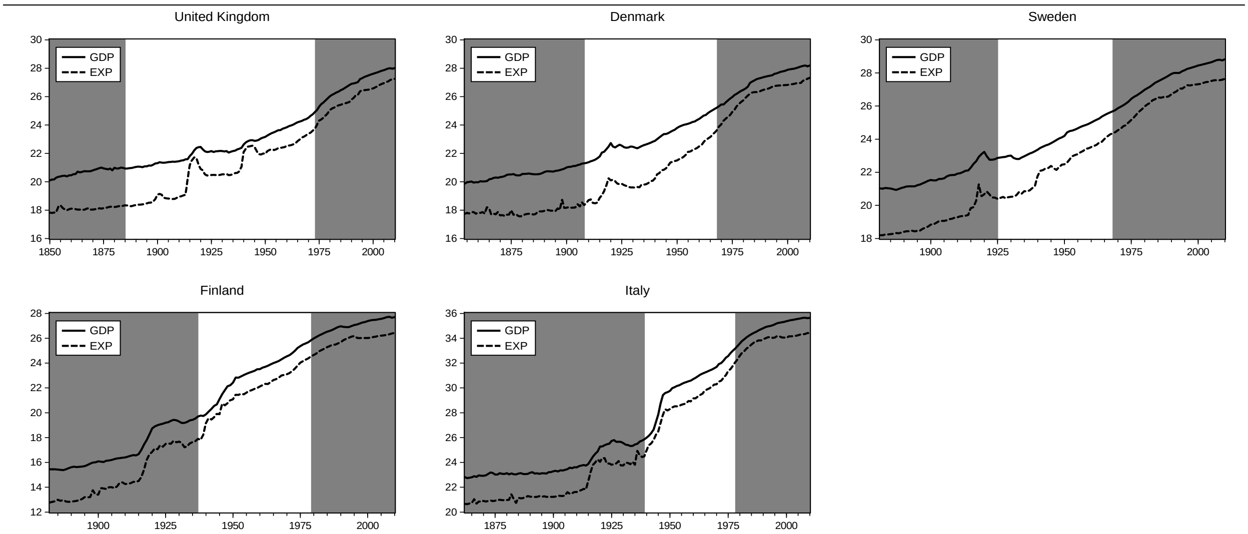
Note: The graph displays the development of gross domestic product per capita from 1850 to 2008 (measured in 1990 International Geary-Khamis dollars). The horizontal lines divide the data set into three stages of economic development: lower middle income (less than 3 500 Int$), upper middle income (3 500 - 12 000 Int$), and high income (more than 12 000 Int$).

den, and Finland ranges from 23 742 to 24 621 International Geary-Khamis dollars (Int$). Only Italy's per capita income exhibits a slightly lower but still comparable value of 19 909 Int$. Nevertheless, regarding the development process over the last 150 years, all countries reveal individual patterns, especially during the late 19^{th} century. In 1850, the United Kingdom, mother country of industrial revolution, had a per capita income of 2 230 Int$, which is more than twice as high as in Finland (911 Int$) and Sweden (1 019 Int$).

In order to provide comparable development stages throughout the countries, we define three development stages based on the World Bank's income group definitions. The first stage is called the *lower-middle-income stage* and is defined by GDP per-capita less than 3 500 Int$. Figure 1.1 shows that the United Kingdom is the first country to hit this threshold, in 1885, followed by Denmark in 1908, Sweden in 1925, Finland in 1937, and

Italy in 1939.[9] The second development stage is called the *upper-middle-income stage* and has per capita GDP between 3 500 and 12 000 Int\$. Compared to the first stage, it can be seen that during this stage the per capita income of all countries converged. Denmark and Sweden reach the upper mark in 1968, followed by the United Kingdom in 1972, Italy in 1977, and Finland in 1978. The third development stage is called the *high-income stage*, comprising a GDP per capita income above 12 000 Int\$.[10]

Figure 1.2: Development of GDP and central government expenditure



Source: 1850 to 1995 from Mitchell (2007); 1996 to 2010 from Eurostat (2012) and IMF (2012b).

Note: The graph displays the logs of gross domestic product (GDP) and central government expenditure (EXP) for the United Kingdom, Denmark, Sweden, Finland, and Italy. The shaded areas highlight the development stages: lower middle income (less than 3 500 Int\$ per capita income), upper middle income (between 3 500 and 12 000 Int\$ per capita income) and high income (above 12 000 Int\$ per capita income).

Figure 1.2 gives a broad historical overview of the development of gross domestic product and central government spending throughout these different income stages. Not surpris-

[9] It can be argued that Italy's per capita income already reaches the 3 500 Int\$ mark in 1918. However, the post-World War I periods caused long-term stagnation of income growth. Hence, the 3 500 Int\$ boundary is deemed to have first been reached in 1939. Nevertheless, this has no effect on the subsequent results.

[10] Our classifications slightly differ from the World Bank income definitions of 2010 in order to provide sufficiently large sample sizes in every development stage.

ingly, all variables have increased considerably over the whole sample period; however, the amount of increase and the stability of the growth pattern differ clearly between the stages and countries. Additionally, it should be noted that the relationship between government expenditure and economic development has changed between the various subsamples. Whereas during the first income stage government spending and GDP rose almost equally, the second stage pictures a catching-up process of expenditure with GDP, especially evident in Denmark and Sweden. In the last stage of development, however, it appears that GDP and expenditure drift slightly apart. Furthermore the spread between nominal GDP and nominal expenditure has narrowed over time. In this regard, periods of major social disturbances (e.g., World Wars I and II) seem to raise expenditures in relation to GDP to a higher level, which is in line with the displacement effect (Peacock and Wiseman (1961)).

1.4 Analytic framework and data

In order to quantify the validity of Wager's law, we concentrate on three functional forms - widespread in the literature - of Wagner's hypothesis, which are summarized in table 1.2.

Table 1.2: Functional forms for testing Wagner's hypothesis

Version	**Functional form**	**Source**
1	$ln(exp) = \alpha + \beta * ln(gdp) + z_t$	Peacock and Wiseman (1961)
2	$ln(exp) = \alpha + \beta * ln(gdppc) + z_t$	Goffman (1968)
3	$ln(exppc) = \alpha + \beta * ln(gdppc) + z_t$	Gupta (1967)

Note: *exp* denotes central government expenditure, *gdp* corresponds to gross domestic product, *gdppc* gross domestic product per capita, and *exppc* central government expenditure per capita.

In an early, classic version, Peacock and Wiseman (1961) model the log of government expenditure in terms of the log of output. Goffman (1968) adopts this version and includes per capita variables in order to control for the development process of the state. Accordingly, Goffman (1968) quantifies government expenditure as a function of per capita output. A related version correcting for the population increase was given by Gupta (1967), who describes the log of per capita government expenditure as a function of the log of per capita output. In general, the literature deals with some additional naive functional forms of Wagner's law (see for example the seminal studies by Mann (1980) as well as Abizadeh and Yousefi (1988)). However, in order to provide a clearly arranged analysis, we confine ourselves to the three well-established versions mentioned above.[11]

[11] We additionally tested the alternative functional forms of Wagner's law given by Musgrave (1969) [$ln(exp/gdp) = \alpha + \beta * ln(gdppc)$] and Mann (1980) [$ln(exp/gdp) = \alpha + \beta * ln(gdp)$]. The cointegration as well as causality results are comparable to those obtained by the versions shown in table 1.2. The estimation results are provided in section 1.A.9 in the appendix.

In order to investigate the relationship and causality between these pairs of variables throughout different periods of economic development, we use annual historical data from Mitchell (2007), who provides data on nominal GNP/GDP and nominal central government expenditure from 1850 to 1995 for the five western European countries Denmark, Finland, Italy, Sweden, and the United Kingdom.[12] Mitchell (2007) uses official publications of the various European governments as the main sources. Therefore, in order to extend the time series and to capture recent behavior of government expenditure and economic development, we interpolate the time series by using data on central government expenditure from Eurostat (2012) and on national income from IMF (2012b) for the periods 1996 to 2010. Overlapping data from all sources for the period 1990 to 1995 display a high degree of data consistency and thus warrant the time series linkage.[13] Data on the total population are taken from Groningen Growth & Development Centre (2012).[14]

Early empirical studies on the evaluation of Wagner's law - which were performed until the early 1990s - used single equation static and dynamic approaches to test the alternative versions of Wagner's hypothesis (see, e.g., Gupta (1967); Mann (1980); Ram (1987)). In general, these studies suffer from various shortcomings. Consequently, at the beginning

[12] Wagner's original definition of government includes local government units as well as public enterprises. As mentioned by Timm (1961), Wagner's law was meant to be valid for every public subsector. Despite the decentralization process of government activities, the central government is still the most important subsector in its expenditure for the services of defense, law and order, welfare, and general structural changes. Therefore, from a historical perspective, the expansion of the central government probably reflects best the traditional government services, which is in line with Wager's hypothesis.

[13] Although all data series generally display a high degree of overlapping, the data are not completely identical. Reasons for this might be data revisions and methodology changes. Deeper insights into the data sources and data interpolation are provided in section 1.A.1 in the appendix.

[14] Historical data are always exposed to criticism concerning data quality. Nevertheless, historical data provided by Mitchell has been used in a number of earlier studies (see, e.g., Dalena and Magazzino (2012); Easterly (2007); Eloranta (2007); Gollin et al. (2004); Thornton (1999); Rousseau and Wachtel (1998)).

of the 1990s the majority of studies adopted a cointegration approach in order to establish a positive long-run relationship between government spending and economic growth (see, e.g., Henrekson (1993); Ahsan et al. (1996)). More recently this approach has been enhanced by the use of endogenous models and advanced causality tests (see, e.g., Iniguez-Montiel (2010); Babatunde (2011); Kumar et al. (2012)). In line with recent empirical studies, our analysis adopts a VECM approach, which allows to study the short- and long-run dynamics between the different variables included in the various functional versions of Wagner's law. In contrast to most other studies, we additionally allow for multiple structural breaks in the unit root and cointegration testing procedures. Thus, the focus of the analysis is thereby on the long-run causal relationship and the adjustment speed of public spending to changes in economic growth throughout countries and development stages. This approach has the advantage of capturing and visualizing the dynamically changing relationship between government activity and economic growth throughout the different stages of economic development.

1.5 Stationarity and cointegration analyses

Because the VECM approach requires the use of difference stationary and cointegrated variables, we start our empirical analysis by testing the unit root properties, applying the Phillips-Perron test (PP) and the Kwiatkowski-Phillips-Schmidt-Shin test (KPSS) in levels and first differences of the logarithmized variables. As historical time series are often associated with changes in the drift parameter or trend function, we additionally utilize the

Lumsdaine and Papell's (1997) unit root approach (LP) allowing for multiple structural breaks at unknown time.[15] To subsequently analyze the existence of a long-run equilibrium relationship among government expenditure and GDP, expenditure and GDP per capita as well as expenditure per capita and GDP per capita, we initially apply the VAR-based cointegration procedure developed by Johansen (1988) and Johansen and Juselius (1990). In order to allow for possible structural breaks and regime shifts in the cointegration analysis, we enhance the basic Johansen testing procedure to allow for multiple structural breaks at unknown time.[16]

The results of the PP and KPSS unit root test are displayed in table 1.7 in the appendix, while the results of the LP unit root test are shown in table 1.10 in the appendix. In most cases, the various test statistics provide coinciding results, concluding that most data series have a unit root in levels and are stationary in first differences. In those cases where the unit root tests do not yield consistent results, at least one test statistic allows to assume unit root behavior in levels and stationarity in first differences. Consequently, all data series in the full sample as well as in the first and second development stage can be treated as integrated of order one. Furthermore, as a consequence of the log transformation, the unit root tests in the third stage of development exhibit only level stationary data. There-

[15]In the presence of structural breaks, PP and KPSS tests have low power and are biased towards non-rejection of a unit root or rejection of stationarity, respectively. The LP approach tests the null hypothesis that the series have a unit root against the alternative of stationarity, with structural changes allowing for multiple structural breaks at unknown time. The breakpoint dates correspond to the minimized value of the sequentially obtained Augmented Dickey-Fuller (ADF) statistic.

[16]Details about the test specifications and lag selection of the unit root and cointegration analysis can be found in section 1.A.2 and 1.A.3 in the appendix. In order to locate possible structural breaks in the cointegration analysis, we apply the multiple structural breakpoint test developed by Bai and Perron (1998).

fore, in order to test for the cointegration relationship in the latest subsample, we use non-transformed level data of all variables. In this case, the data are also integrated of order one.[17]

In the next step, table 1.3 displays a general overview of all pairs of variables in different subsamples where at least one test statistic rejected the null of no cointegration at least at a 10 percent level.[18] The detailed test statistics as well as determined break points are presented in table 1.11 to 1.18 in the appendix. Since the estimated breakpoint dates of expenditure and GDP are in some cases very close to each other, the depicted test statistics only include the public expenditure breakpoints. Nevertheless, the results are robust to considering the GDP breakpoints. In all other cases, the expenditure as well as GDP breakpoints are included. Due to the integrity of the data in periods of major social disturbances (e.g., World Wars I and II, Great Depression, Oil crises) and the impact of country-specific economic crisis (e.g., the Finish and Swedish banking crisis), it is not surprising that during some stages, cointegration is only detected by allowing for structural breaks. As listed in table 1.19 in section 1.A.4 in the appendix, the majority of the structural breaks detected by the Bai-Perron procedure coincide with these major economic crises, as predicted by Peacock and Wiseman's displacement hypothesis (see Henry and Olekalns (2010)). The cointegration results reveal that public spending in the United Kingdom, Sweden, Fin-

[17]The unit root results stay qualitatively unchanged if we include a constant in our level test specification. Furthermore, the unit root approaches of Zivot and Andrews (1992) and of Lanne et al. (2002), allowing for a single structural break, also provide evidence that the series can be treated as integrated of order one. Detailed results are available in the appendix, see section 1.A.2.

[18]A significant test statistic is based on the assumption that the null hypothesis is true. The Johansen testing procedure tests the null hypothesis for no cointegration. Therefore, rejecting no cointegration provides stronger statistical evidence than not rejecting the no-cointegration null hypothesis. A significant test statistic yields a stronger statement than an insignificant statistic.

Table 1.3: Cointegration relationships for different development stages

Country	Variable	Full Sample	Stage I	Stage II	Stage III
United Kingdom	exp and gdp	C	C	C	C
	exp and gdppc	C	C	C	C
	exppc and gdppc	C	C	C	C
Denmark	exp and gdp	C	-	C	C
	exp and gdppc	C	-	C	C
	exppc and gdppc	C	-	C	C
Sweden	exp and gdp	C	C	C	C
	exp and gdppc	C	C	C	C
	exppc and gdppc	C	C	C	C
Finland	exp and gdp	C	C	C	C
	exp and gdppc	C	C	C	C
	exppc and gdppc	C	C	C	C
Italy	exp and gdp	C	C	C	C
	exp and gdppc	C	C	C	C
	exppc and gdppc	C	C	C	C

Note: C denotes that a cointegration vector exists between the set of variables. The cointegration results without structural breaks are based upon the trace and maximum eigenvalue tests derived by Johansen (1988) and Johansen and Juselius (1990). The cointegration results with structural breaks are based upon the trace test derived by Johansen et al. (2000).

land, and Italy is cointegrated with economic growth, independently of development stage or functional form. These findings are in line with Wagner's hypothesis and confirm the statement that the public sector and economic growth display a co-movement as long there is cultural and economic progress. This relationship is maintained throughout every stage of development and is still valid today. However, it does not hold for Denmark. In that country, a cointegration relationship for all three versions of Wagner's law was found in the second and third development stages, but not in the first. This is a contradictory finding to the assumption that the relationship between the public sector and economic growth is characteristic of the early stages of development.[19]

[19]These results are in accordance with other empirical studies that investigate early stages of industrialization (see Thornton (1999); Oxley (1994)). Durevall and Henrekson (2011) detect, for Sweden and the United Kingdom, a cointegration relationship between the public sector and economic growth, esp-

1.6 A VECM approach to test for long-run causality

The model used to test for long-run causality in each subsample is expressed as a restricted VAR in terms of an error correction model:

$$\Delta ln(g)_t = c_{1t} + \sum_{i=1}^{p} \varphi_{1i} \Delta ln(g)_{t-i} + \sum_{i=1}^{p} \vartheta_{1i} \Delta ln(y)_{t-i} + \gamma_1 [ln(g)_{t-1} - \beta_1 \cdot ln(y)_{t-1} + \alpha_1] + \epsilon_{1t} \tag{1.1}$$

$$\Delta ln(y)_t = c_{2t} + \sum_{i=1}^{p} \varphi_{2i} \Delta ln(g)_{t-i} + \sum_{i=1}^{p} \vartheta_{2i} \Delta ln(y)_{t-i} + \gamma_2 [ln(y)_{t-1} - \beta_2 \cdot ln(g)_{t-1} + \alpha_2] + \epsilon_{2t} \tag{1.2}$$

with g representing government expenditure in versions 1 and 2 and government expenditure per capita in version 3, while y denotes GDP in version 1 and GDP per capita in versions 2 and 3. Because of the cointegration relationship, at least one of the variables has to significantly adjust to deviations from the long-run equilibrium, which is captured by γ_i. That parameter describes the speed of adjustment back to the equilibrium and measures the proportion of last period's equilibrium error that is corrected for. Thus, in equation (1.1) and (1.2), the VECM allows for the ascertainment that g granger-causes y or vice versa, as long as the corresponding error correction term γ_i carries a statistically significant coefficient, even if all other coefficients are not jointly significant (see Granger (1988)). Verification of the law is obtained if significant causality is found to run from

ecially between 1860 and the mid-1970s. Comparable country-specific studies of advanced industrialized countries in the post-Bretton Woods era are scarce and provide rather mixed results. While Kolluri et al. (2000) yield support of Wagner's law for Italy and the UK, Durevall and Henrekson (2011) and Chow et al. (2002) detect only long-run relationships controlling for age structure and money supply, respectively. Analyzing disaggregated data on Italy, Magazzino (2012a) finds a cointegration relationship for three out of five items of public spending between 1960 to 2008.

economic growth to government activity. The magnitude of the adjustment parameter γ_i contains information about the capacity of countries to absorb exogenous shocks in different development stages and also about the reaction of expenditure to changes in GDP.[20] In order to test for long-run causality between the different variables and different country sets, we estimate for every detected cointegration pair a VECM and apply a one-sided t-test to the error correction term.[21] A negative statistically significant adjustment parameter in the VECM with expenditure and expenditure per capita on the left-hand side implies validity of Wagner's hypothesis, bespeaking GDP or GDP per capita, respectively, as the driving force of government expenditure.[22]

Table 1.4 presents the estimated error correction terms and the results of the one-sided t-test. According to the estimated VECMs and the corresponding error correction terms, at least one of the coefficients is - in every model - statistically significantly smaller than zero, which is a requirement for the various versions of Wagner's law to be cointegrated. Only Denmark does not exhibit a cointegration relationship in the first stage of development, so that a feasible error correction model for it could not be estimated.

[20] As mentioned by Granger (1969), VAR-based models are only valid to test for causality if instantaneous causality can be excluded theoretically. Wagner's law implies that government expenditure reacts to a change of income in the long-run, driven by a changing demand for public goods as a result of increasing prosperity. Thus, it can be assumed that a response of government spending to changes in national income does not appear in the same period, but is delayed by some periods.

[21] Because the time series cover different historical epochs and are split into different samples, the data exhibits individual clustered episodes of relatively high variance. In order to allow for cross-equation heteroskedasticity, we employ weighted least squares to sustain consistent and asymptotically efficient estimates.

[22] In general, most empirical studies only interpret unidirectional causality running from economic growth to public spending as a pure statistically confirmation of Wagner's law (see, e.g, Magazzino (2012b)). Yet, if there is a bidirectional causal relationship, then an increase in expenditure may influence GDP as well. Despite this feedback effect between the variables, Wagner's law is still valid as long as the expenditure adjustment coefficient is sufficiently large.

Table 1.4: Long-run causality and short-run adjustment

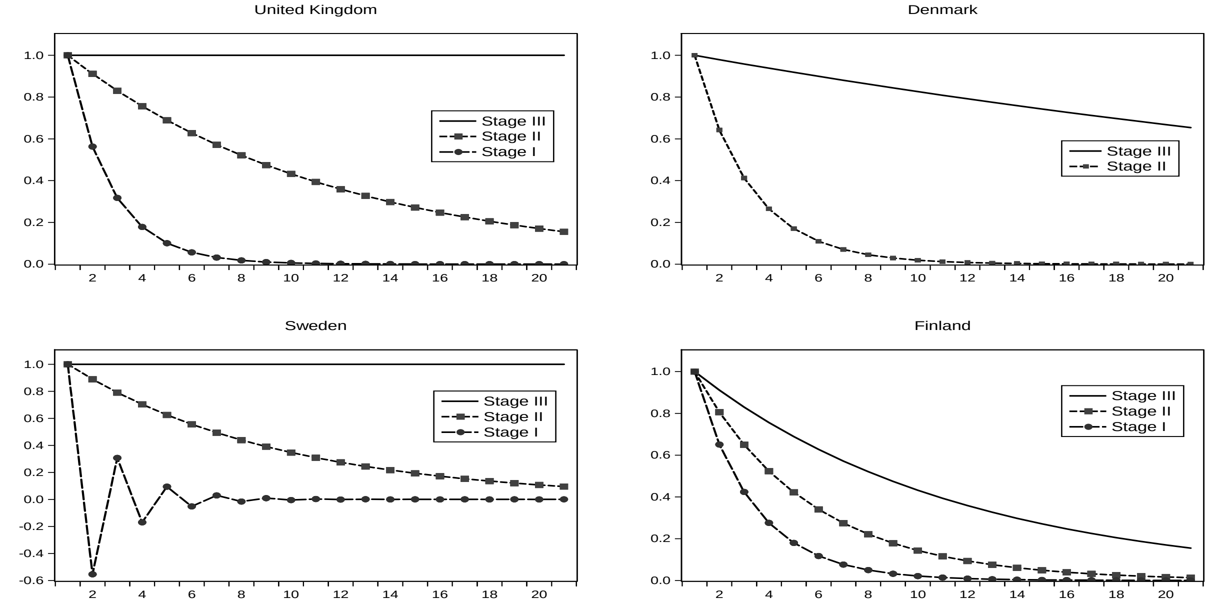
		Full Sample		Stage I		Stage II		Stage III	
Country	G and Y	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$
UK	exp and gdp	-0.042*	-0.016*	-0.437***	0.022	-0.089**	-0.015	0.055	-0.101***
		(-1.769)	(-1.629)	(-3.574)	(1.876)	(-2.350)	(-1.131)	(2.767)	(-4.789)
	exp and gdppc	-0.015	-0.017**	-0.339***	0.002	-0.057**	-0.021**	0.064	-0.115***
		(-0.823)	(-2.267)	(-2.875)	(2.336)	(-1.679)	(-1.926)	(2.774)	(-4.883)
	exppc and gdppc	-0.035*	-0.018*	-0.535***	-0.053**	-0.071**	-0.019**	0.051	-0.098***
		(-1.472)	(-1.948)	(-4.428)	(-2.176)	(-1.965)	(-1.678)	(2.839)	(-4.842)
Denmark	exp and gdp	-0.078***	-0.002	-	-	-0.358***	-0.059	-0.021***	-0.001
		(-2.560)	(-0.157)			(-2.977)	(-0.680)	(-3.655)	(-0.251)
	exp and gdppc	-0.048	-0.066***	-	-	-0.487***	-0.197**	-0.008***	-0.003
		(-1.174)	(-3.021)			(-3.547)	(-1.691)	(-3.666)	(-0.337)
	exppc and gdppc	-0.078**	-0.015	-	-	-0.381***	-0.115*	-0.017***	-0.001
		(-2.286)	(-0.865)			(-3.253)	(-1.307)	(-3.671)	(-0.179)
Sweden	exp and gdp	-0.034	-0.083***	-1.554***	0.338	-0.111*	-0.085***	-0.052	-0.188***
		(0.054)	(-3.681)	(-4.886)	(2.156)	(-1.508)	(-2.741)	(-0.694)	(-4.045)
	exp and gdppc	-0.006	-0.073***	-1.595***	0.164	-0.107*	-0.092***	-0.058	-0.209***
		(-0.121)	(-3.820)	(-3.054)	(0.608)	(-1.468)	(-2.687)	(-0.707)	(-4.219)
	exppc and gdppc	-0.027	-0.081***	-1.592***	0.364	-0.114*	-0.081***	-0.054	-0.191***
		(-0.524)	(-3.774)	(-3.388)	(1.667)	(-1.566)	(-2.577)	(-0.688)	(-4.079)
Finland	exp and gdp	-0.068	-0.090***	-0.349***	-0.153**	-0.194*	-0.228***	-0.089***	0.012
		(-1.101)	(-3.098)	(-2.639)	(-2.062)	(-1.365)	(-3.957)	(-3.110)	(3.316)
	exp and gdppc	-0.024	-0.089***	-0.153*	-0.161***	-0.195*	-0.237***	-0.141***	0.044
		(-0.452)	(-3.254)	(-1.456)	(-2.613)	(-1.367)	(-3.887)	(-3.258)	(3.640)
	exppc and gdppc	-0.066	-0.089***	-0.313***	-0.156**	-0.195*	-0.233***	-0.138***	0.039
		(-1.075)	(-3.058)	(-2.414)	(-2.130)	(-1.359)	(-4.049)	(-3.496)	(4.075)
Italy	exp and gdp	-0.042	-0.096***	-0.095	-0.220***	0.289	-0.442***	-0.057***	-0.019***
		(-0.801)	(-3.553)	(-0.796)	(-4.739)	(3.058)	(-6.079)	(2.398)	(-3.788)
	exp and gdppc	-0.002	-0.094***	-0.011	-0.232***	0.282	-0.412***	-0.073***	-0.009***
		(-0.051)	(-3.839)	(-0.086)	(-4.892)	(3.432)	(-6.106)	(-2.834)	(-3.346)
	exppc and gdppc	-0.039	-0.097***	-0.077	-0.232***	0.292	-0.442***	-0.066***	-0.010***
		(-0.728)	(-3.607)	(-0.635)	(-4.872)	(3.103)	(-6.075)	(-2.621)	(-3.585)

Note: The table displays estimated error correction terms (ect) of corresponding VECMs. The t-statistics are presented in parenthesis. The amount of the regressors included in the VECMs are determined by using Schwarz information criterion. In the presence of autocorrelation the lag length of the VECMs are successively enhanced to remove all serial correlation from the data, up to a maximum of 5 lags. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Starting with the full sample results, it can be seen that only the United Kingdom and Denmark have statistically significant error correction terms (ect) in the first and third versions, which are in line with Wagner's law. However, in both countries the convergence speed of government spending is relatively slow: it takes around 11 periods in Denmark and 19 periods in the United Kingdom until half of the disequilibrium is removed. Thus, Wagner's hypothesis that economic growth is a driving force for government expenditure can be rejected at least in a time period over the last 150 years. Interestingly, at the same time all countries exhibit significant long-run causality running from public spending to economic growth at least in one functional form. These findings support models of economic growth that suggest a possible long-run relationship between the share of government spending in GDP and the growth rate of per capita real GDP (see, e.g., Barro (1990); Devarajan et al. (1996)). Nevertheless, here too the adjustment coefficients are rather low, leaving the economic significance open to question.

These results provide a nuanced picture when dissecting the full sample of the three stages of income development. Particularly striking is that in the United Kingdom, Denmark, Sweden, and Finland, the error correction terms running from public spending to economic growth decrease in statistical significance as well as in adjustment speed with increasing state of development. These findings confirm the hypothesis that in an advanced stage of development, public spending does not react to changes in income as sensitively as in earlier stages. The decreasing speed of adjustment of government expenditure towards long-run equilibrium induced by shocks in GDP is visualized in figure 1.3.

Figure 1.3: Speed of adjustment of government expenditure towards long-run equilibrium induced by shocks in GDP



Note: The graph displays the expenditure convergence to shocks in GDP for the United Kingdom, Denmark, Sweden, and Finland during different stages of economic development. For Denmark, no long-run equilibrium between government expenditure and GDP was detected during the first development stage. The expenditure convergences are calculated by $\sum_{n=0}^{20}(1-ect_{ij})^n$ where $n+1$ denotes the number of periods, i the country, j the development stage and ect the corresponding error correction term from table 1.4.

In early stages of development, the adjustment speed of public expenditure is faster than in latter stages where no adjustment is found in the UK and Sweden and very slow adjustment can be exhibited in Denmark and Finland. The economic relevance of Wagner's law seems to lapse in the high-income stage.[23]

However, the results for Italy provide a different picture and do not follow this pattern. In this case, statistical causality is only detected in the last development stage, which carries a low adjustment coefficient having no economic significance. The invalidity of Wagner's law

[23]For the period before 1913, Thornton (1999) estimates, by the use of single error correction models, long-run adjustment parameters of -3.82 and -3.54 for Denmark and the UK, respectively, which are even higher than our estimates. Focusing on later periods, studies by Chow et al. (2002) and Loizides and Vamvoukas (2005) present error correction terms in line with Wagner's theory: -0.23 and -0.28 for the UK.

in Italy might be explained by the deviating pathway of the Italian economy compared to the other countries. On the one hand, the Italian economy developed rather slowly reaching the upper-middle-income level in 1940 despite a comparable high per capita income of 1 350 Int.$ in 1850. On the other hand, a peculiar Italian welfare system was established in the period following World War II, and a universal welfare model was not introduced until 1978, which might explain the significant results in the last stage of development.[24]
The decreasing validity of Wagner's law may be derived from the public choice literature. Rational-choice models of government size have led to arguments that government has grown because of extensions in voter turnout, which changed the position of the median voter (see, e.g., Meltzer and Richard (1981); Lott and Kenny (1999)). The late 19^{th} and early 20^{th} century is characterized as an era of political enfranchisement where the increased number of voters with relatively low incomes led to a preference (by the median voter) for governmental redistributional programs and thus to an increase in government spending. This might also explain the invalidity of Wagner's law in Italy, where democratic structures developed rather late.

Wagner's hypothesis was meant to explain the growth of expenditure in the long-run (see, e.g., Timm (1961)). Therefore our analysis focuses primarily on the long-run relationship between public spending and economic growth. Yet, it might be interesting to gain some further insights into the short-run relationship between the variables. In addition to a

[24] For further information on the development of the Italian welfare state, see Ferrera (1997). In addition, Legrenzi and Milas (2002a,b) show that a bivariate representation of Wagner's law based on post-World War II data is spurious. In particular, they identify a long-run relationship that involves general government expenditure, the demand-side GDP variable, the supply-side bureaucratic power and the ratio of local to state expenditure, as an institutional factor that captures the division of competencies between local and central government in allocating public expenditure.

short-run validity of Wagner's law, the results provide evidence for Keynes's hypothesis in that significant causality is detected from expenditure to economic growth. Granger causality tests show that even in the short-run, Wagner's law seems to have lost its validity in the last stage of development.[25] Conversely, regarding the full sample, evidence in favor of Wagner's law could be found in Denmark, Sweden, Finland, and Italy. In addition, support for Keynesian policies could be found in the UK, Denmark, Sweden, and Finland. The outcome of the causality results during the first and second development stage are rather mixed. The detailed results of the short run causality test can be found in table 1.20 in section 1.A.5 in the appendix.

We conclude our main analysis by examining some additional VECM diagnostics to provide further insights into the model specification and residual diagnostics. In general, table 1.21 in section 1.A.6 in the appendix displays that the goodness-of-fit measured by the adjusted R^2 is sufficiently large for every VECM. Nevertheless, in the case of Finland (stage II) and Italy (stage I), the adjusted R^2 is negative.[26] The reason for this lies in the fact that the sample beginning in Finland and ending in Italy coincides with extreme values caused by World War II. In both countries the adjusted R^2 gets sufficiently large and positive if the sample is shortened or extended, and in both the estimation results do not change substantially.[27] Additionally, it can be seen that no estimated models evince

[25] The Granger causality tests on the VECMs were performed as Wald style exclusion tests on the lagged difference terms.

[26] After adjustments, the included number of observations in the VECM in Finland at the last stage of development (stage III) includes only 29 observations that could affect the reliability of tests and estimations.

[27] We also estimated the VECMs under alternative specifications of the deterministic components. Adding a trend to the cointegration equation increases slightly the error correction mechanism running from economic growth to government activity. However, qualitatively the speed of adjustment of public

any sign of serial correlation. In those cases where heteroskedasticity could not be rejected, we employ weighted least squares to sustain consistent and asymptotically efficient estimates. However, the statistical significance and point estimators are not changed from the standard ordinary least squares (OLS) estimations. Unfortunately, the test for joint normality of the residuals significantly rejects the normality assumption in the majority of VECMs. As these tests are generally very sensitive to outliers, we additionally estimate the corresponding VECMs including dummy variables to effectively remove extreme observations and improve the chances of error normality. In fact the qualitative outcome for all countries and samples remains unchanged. Detailed results are provided in section 1.A.6 in the appendix. Finally, the stability of the VECM has been ensured through the test of inverse roots of the autoregressive (AR) characteristic polynomial. In addition, the Chow forecast test for parameter stability shows that in most cases the parameter stabilities of the corresponding VECMs are satisfactory (see section 1.A.7 in the appendix for details).

1.7 Robustness analysis

The baseline estimations in the previous section provide evidence of a decreasing response of government expenditure to changes in GDP with an advanced stage of development for the United Kingdom, Denmark, Sweden, and Finland. In this section, we run several robustness estimations to underpin this changing relation between public spending and economic growth throughout economic development.

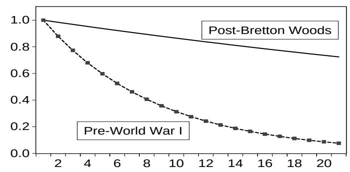
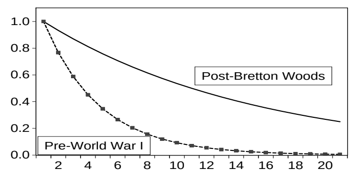
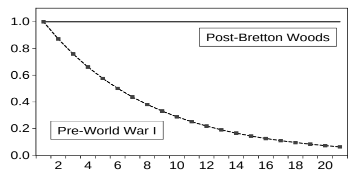
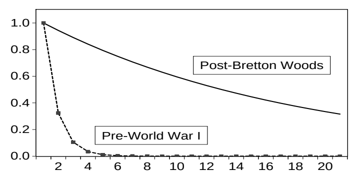
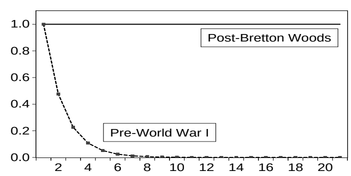
expenditure towards long-run equilibrium induced by shocks in GDP remains unchanged. Detailed estimation results are provided in section 1.A.8 in the appendix.

Time series of historical data are exposed to abnormalities during periods of major social disturbances. With respect to the analysis of Wagner's law, this results in several problems. On the one hand, outliers might have a significant effect on the estimation results, and on the other hand, structural breaks induced by the displacement effect may permanently bias the adjustment coefficients. A particularly crisis-ridden period was from the beginning of World War I until the end of the Bretton Woods era in 1973. During this period, the economies were heavily affected by World Wars I and II, the Great Depression, and the oil crisis. However, the exact time limits of a particular crisis prove very difficult to determine, because the aftermath of the initial crisis may last up to several years (see Reinhart and Rogoff (2009)). Therefore, in order to exclude periods of major social disturbances from the analysis, we split our data set for each country into a pre-World War I and a post-Bretton Woods sample. This approach allows us to compare the relationship between public spending and economic growth in a very low and a high development stage without the influence of several major global economic crises.

Table 1.5 presents the error correction terms for the pre-World War I and post-Bretton Woods sample. It can be seen that the adjustment coefficients with economic growth as the dependent variable are significantly higher during the early pre-World War I sample. In general, this finding applies for all countries. Only Sweden does not provide robust results throughout the different versions of Wagner's law; this might be a result of the small sample size.[28]

[28] For Denmark and Italy, cointegration could only be detected in the pre-World War I period using the Engle-Granger approach. In the case of Denmark not all variables (exppc) seem to fulfill the stationarity requirements. Therefore, the displayed error correction terms have to be interpreted with caution.

Table 1.5: Long-run causality and short-run adjustment without crisis period

		pre-World War I		post-Bretton Woods		Adjustment speed of exp
Country	G and Y	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$	towards long-run equilibrium
UK	exp and gdp	-0.121***	-0.064**	-0.016***	-0.031***	
		(-2.218)	(-1.767)	(-4.413)	(-4.362)	
	exp and gdppc	-0.046	-0.128***	0.028	-0.072***	
		(-1.157)	(-2.478)	(3.307)	(-4.566)	
	exppc and gdppc	-0.122***	-0.068**	-0.023***	-0.027***	
		(-2.213)	(-1.829)	(-3.583)	(-4.423)	
Denmark	exp and gdp	-0.233**	-0.023**	-0.067***	0.004	
		(-1.946)	(-2.115)	(-3.300)	(0.792)	
	exp and gdppc	-0.211**	-0.039***	-0.047***	-0.001	
		(-1.778)	(-2.141)	(-3.179)	(-0.983)	
	exppc and gdppc	-0.274**	-0.004*	-0.059***	0.002	
		(-2.307)	(-1.445)	(-3.303)	(0.821)	
Sweden	exp and gdp	-0.129*	-0.214**	0.040	-0.129***	
		(-1.340)	(-1.913)	(1.734)	(-6.743)	
	exp and gdppc	-0.080	-0.264***	0.053	-0.157***	
		(-0.963)	(-2.399)	(1.497)	(-6.324)	
	exppc and gdppc	-0.113	-0.228**	0.050	-0.145***	
		(-1.188)	(-2.058)	(1.528)	(-6.136)	
Finland	exp and gdp	-0.676***	-0.139**	-0.056**	0.005	
		(-3.580)	(-1.830)	(-2.423)	(3.221)	
	exp and gdppc	-0.693***	-0.266***	-0.055**	0.005	
		(-3.842)	(-2.713)	(-2.431)	(3.226)	
	exppc and gdppc	-0.711***	-0.172**	-0.053**	0.001	
		(-3.809)	(-1.927)	(-2.363)	(3.272)	
Italy	exp and gdp	-0.524***	-0.047	0.007	-0.071***	
		(-3.400)	(-0.503)	(0.642)	(-4.381)	
	exp and gdppc	-0.224**	-0.094	0.009	-0.069***	
		(-2.092)	(-0.985)	(0.758)	(-4.020)	
	exppc and gdppc	-0.486***	-0.056	0.007	-0.072***	
		(-3.189)	(-0.624)	(0.474)	(-4.049)	

Note: The table displays estimated error correction terms (ect) of corresponding VECMs. The t-statistics are presented in parenthesis. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels. The expenditure convergences are calculated for the first functional form of Wagner's law.

For Sweden, Finland, and Italy the pre-World War I period covers an earlier development stage than the lower-middle-income stage used in the baseline estimations in the previous section. This might explain the significant increase of adjustment speed for Finland and Italy. Additionally, it is striking that Finland and Italy - both countries with the lowest economic development in 1913 (measured in terms of GDP per capita) - exhibit the highest speed of adjustment of expenditure towards the long-run equilibrium. The declining adjustment speed of government expenditure towards long-run equilibrium induced by shocks in GDP is visualized in the right column of table 1.5. The response of expenditure to changes in GDP happens much faster during the pre-World War I stage than in the post-Bretton Woods sample, supporting the notion that Wagner's law loses its validity with an advanced stage of development.

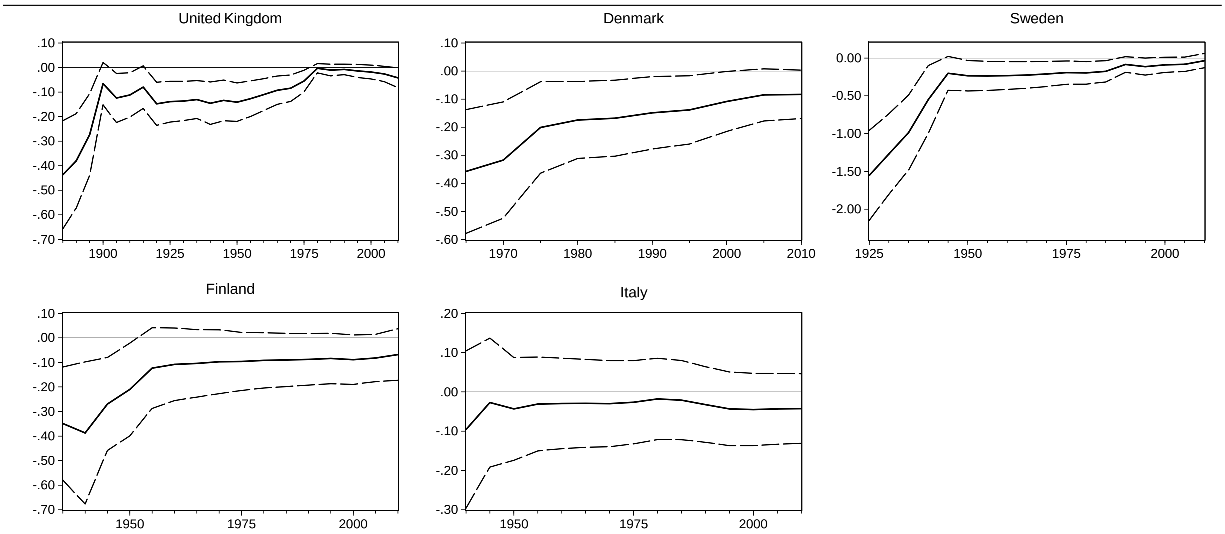
With regard to the development of the relationship between public expenditure and economic growth throughout the last 150 years, figure 1.4 displays the development of the expenditure adjustment by recursive VECM estimation. Starting from the lower-middle-income stage, we added five years in each step and visualized every corresponding error correction term, including the 90% confidence band.[29]

It can be seen that, with an advanced economic evolution, the adjustment coefficient of expenditure with respect to changes in GDP declines, again suggesting a declining causality between economic growth and government activity. This declining path of the error correction mechanism is valid for the United Kingdom, Denmark, Sweden, and Finland.

[29]For Denmark, we started the recursive VECM estimation at the end of the upper-middle-income stage because of the missing cointegration relationship in the first development stage.

In contrast, Italy displays no sign of significant expenditure adjustment throughout the whole sample period. The recursive estimations confirm the result of the previous section that with an advanced degree of development the adjustment speed of expenditure steadily declines. The insignificant error correction terms around the year 1915 in the UK as well as the one around 1945 in Sweden might be the effect of World Wars I and II.[30]

Figure 1.4: Recursive estimation of expenditure adjustment



Note: The graphs display the development of expenditure adjustment by a recursive VECM estimation for the first functional form of Wagner's law. The solid line visualizes the point estimation of the error correction term, while the dashed lines present the 90% confidence band.

[30]In general, the step-by-step reduction of adjustment speed supports the finding by Durevall and Henrekson (2011), who detect a direct linkage between public spending and GDP in a period of 30 to 35 years after World War II for the UK and Sweden. Lamartina and Zaghini (2011), via recursive pooled estimations, also detect a significant decline in long-run elasticity between GDP and public spending for 23 OECD countries from 1990 to 2006.

1.8 Conclusion

In order to test the validity of Wagner's law at different stages of economic development, we apply advanced cointegration and causality approaches on five European advanced welfare states: the United Kingdom, Denmark, Finland, Sweden, and Italy. By using historical data on government expenditure and GDP from the mid-19^{th} century, we classify every country into three individual stages of development in terms of per-capita income. This approach allows us to make statements about the dynamic relationship between public spending and economic growth from a within-country perspective and additionally enables us to identify commonalities across countries despite differences in size and development pattern.

The findings reveal that public spending and economic growth are cointegrated in the United Kingdom, Sweden, Finland, and Italy independently of development stage or functional form. However, in the case of Denmark, a cointegration relationship was only detected in the second and third development stage. The co-movement between the variables is consistent with Wagner's view that there was not necessarily a cause-and-effect relationship between economic development and government activity (see Peacock and Scott (2000)).

To gain further insights into the relation between Wagner's law and development stage, we estimate subsequent VECMs and analyze the adjustment speed of public spending to changes in economic growth. The hypothesis that Wagner's law might have a higher validity during early stages of development turned out to be acceptable for the United Kingdom, Denmark, Sweden, and Finland. The estimations exhibit that with an increasing state of development, the error correction terms running from public spending to economic growth decline in statistical significance as well as in adjustment speed. Regarding the classic relationship between government expenditure and GDP, our VECM estimations show a decreasing error correction mechanism of public spending from the first to the last development stage in the United Kingdom (-0.437 to 0.055), Denmark (-0.358 to -0.021), Sweden (-1.554 to 0.052), and Finland (-0.349 to -0.089). Recursive vector error correction estimations confirm the weakened dynamic relationship between public expenditure and economic growth throughout economic evolution. The United Kingdom, Denmark, Sweden, and Finland display a clear declining trend of the error correction mechanism running from GDP to government spending. In general, the results substantiate that the relationship between public spending and economic growth has weakened over the last century. In the case of Italy, however, statistical causality is only detected in the last development stage, which carries a low adjustment coefficient of -0.057 and thus has no economic significance. Data in recent decades on all countries under review suggest that the Wagnerian relationship between economic growth and public spending may have lost its economic relevance.

As mentioned by Lindert (1996), the relationship between income growth and government spending remains a popular black box to explain the increase of government size throughout time. The detailed reasons why Wagner's law holds in some periods and countries may be various and are beyond the scope of this study. In the spirit of Wagner's hypothesis, the weakened relationship between government expenditure and economic growth can be explained by the expanding role of governments associated with strong changes in the structure of the economy. Well-established welfare states like the United Kingdom, Denmark, Finland, Sweden, and Italy have past through those major structural changes in recent days. With regard to the sustainability of growing public debts, these signs of expenditure decoupling could have implications for the budgetary process of advanced industrialized countries. Nevertheless, it remains to be seen how future main drivers of spending growth, such as ageing and demographic change (see, e.g., European Commission (2012)) as well as health care (see, e.g., Przywara (2010)), will affect the prospective relationship between economic growth and public spending.

1.A Appendix

In the following sections, further information and robustness checks related to the main part of this chapter are provided. The appendix is organized as follows: **section 1.A.1** discusses general data issues and gives deeper insights into the data sources, definitions, and interpolation. The subsequent **section 1.A.2** presents detailed unit root results from various testing procedures. Besides results derived from alternative specifications of the deterministic components (trend-stationarity versus difference-stationarity), this section also discusses unit root results allowing for one or more endogenously chosen structural breaks. Subsequently, **section 1.A.3** displays the detailed test statistics of the cointegration analysis for different development stages and functional forms of Wagner's law, which are summarized in table 1.3 in the main part of this chapter. The detected break points included in the unit root and cointegration analysis are discussed in **section 1.A.4**. Thereafter, **section 1.A.5** provides further insights into the short-run relationship between government expenditure and economic activity in terms of Granger causality results. Additional VECM diagnostics including estimations controlling for outliers are provided in **section 1.A.6**. **Section 1.A.7** controls for the stability conditions of estimates, while **section 1.A.8** obtains alternative VECM estimations with different lag specifications and deterministic components. Finally, the last **section 1.A.9** of this appendix provides cointegration and causality results for some additional functional forms of Wagner's law by Musgrave (1969) and Mann (1980).

1.A.1 Variables, data sources, and data interpolation

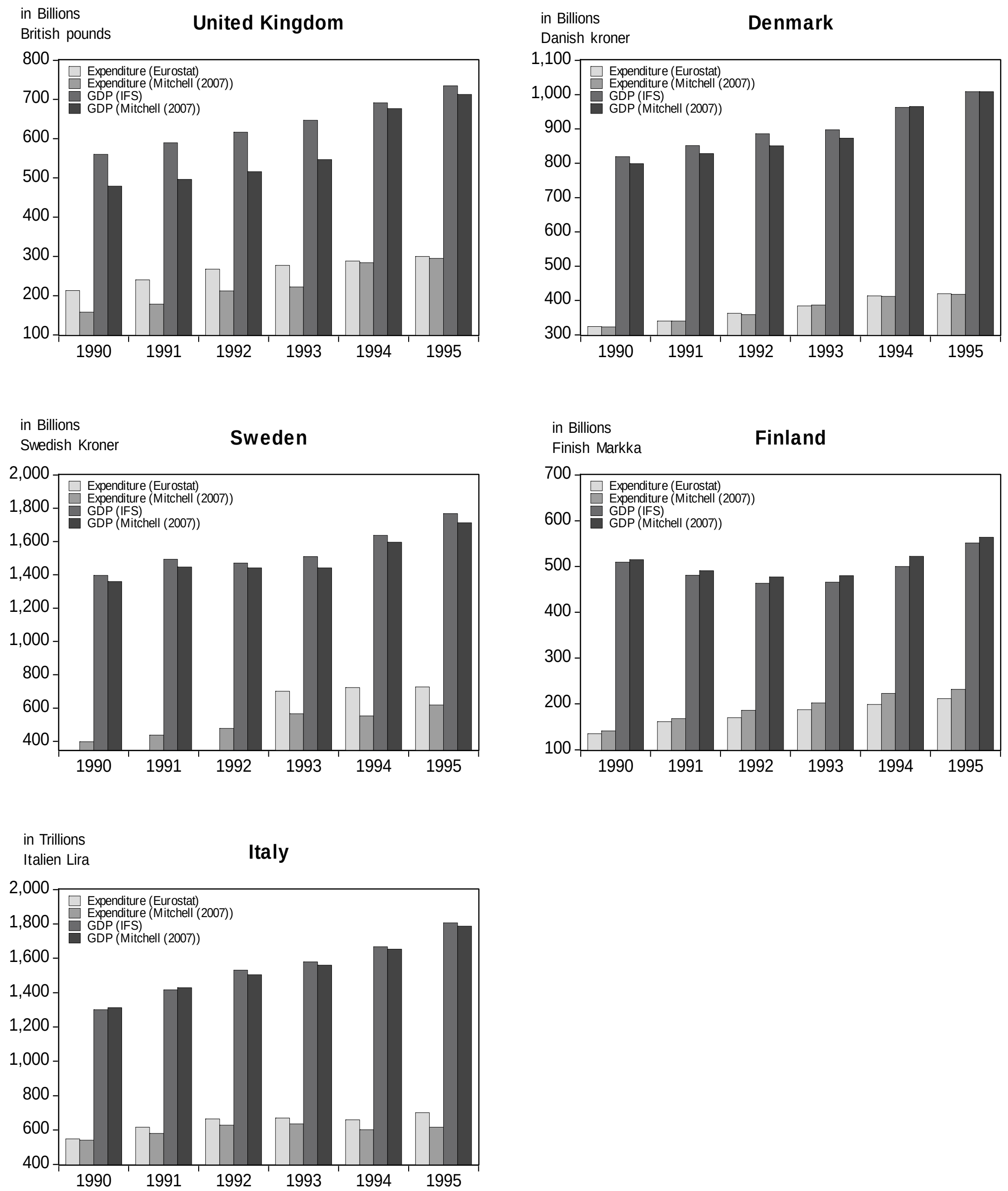
The main data sources used in Mitchell (2007) have been the official publications of the various European governments. Therefore, in order to extend the time series and to capture recent behavior of government expenditure and economic development, we interpolate the time series by using data on central government expenditures from Eurostat (2012) and on national income from IMF (2012b) for the periods 1996 to 2010. An overview of the variables and data sources can be found in table 1.6. The combination of data series from different sources makes it possible to analyze sufficient large sample sizes in the last stage of development. As it can be seen in figure 1.5, overlapping data from all sources for the period 1990 to 1995 display a high degree of data consistency and thus warrant the time series linkage. Small differences between the data series might be the consequence of data revisions and methodology changes. Nevertheless, figure 1.6 displays that the data linkage from different sources has no visual effects on the time series. Hence, the data interpolation should have only very little effect on the estimation results.

Table 1.6: List of variables and data sources

Country	Variable	Source and Data Codes
United Kingdom	Total central government expenditure (exp)	[1850-1995] Mitchell (2007); [1996-2010] Eurostat (2012) (UK-TE-S1313)
	Gross domestic product (gdp)	[1850-1995] Mitchell (2007); [1996-2010] IMF (2012b) (Gross national income)
	Population	[1850-2010] Groningen Growth & Development Centre (2012)
Denmark	Total central government expenditure (exp)	[1854-1995] Mitchell (2007); [1996-2010] Eurostat (2012) (DK-TE-S1313)
	Gross domestic product (gdp)	[1854-1995] Mitchell (2007); [1996-2010] IMF (2012b) (Gross national income)
	Population	[1854-2010] Groningen Growth & Development Centre (2012)
Sweden	Total central government expenditure (exp)	[1881-1995] Mitchell (2007); [1996-2010] Eurostat (2012) (SE-TE-S1313)
	Gross domestic product (gdp)	[1881-1995] Mitchell (2007); [1996-2010]IMF (2012b) (Gross national income)
	Population	[1881-2010] Groningen Growth & Development Centre (2012)
Finland	Total central government expenditure (exp)	[1882-1995] Mitchell (2007); [1996-2010] Eurostat (2012) (FI-TE-S1313)
	Gross domestic product (gdp)	[1882-1995] Mitchell (2007); [1996-2010] IMF (2012b) (Gross national income)
	Population	[1882-2010] Groningen Growth & Development Centre (2012)
Italy	Total central government expenditure (exp)	[1862-1995] Mitchell (2007); [1996-2010] Eurostat (2012) (IT-TE-S1313)
	Gross domestic product (gdp)	[1862-1995] Mitchell (2007); [1996-2010] IMF (2012b) (Gross national income)
	Population	[1862-2010] Groningen Growth & Development Centre (2012)

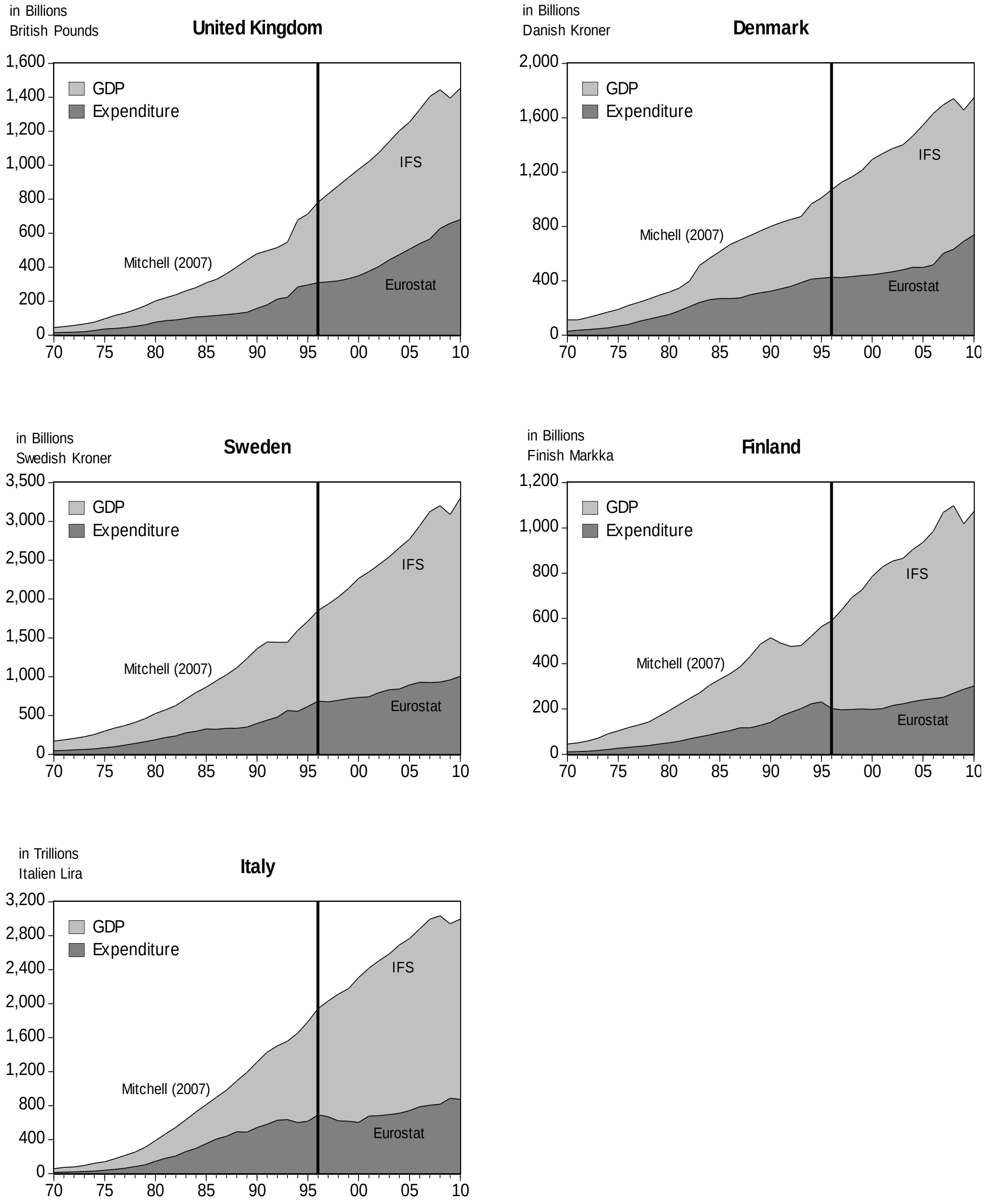
Note: The main sources used in Mitchell (2007) have been the official publications of the various European governments. A few gaps were filled from the League of Nations, Public Finance Statistics.

Figure 1.5: Comparison of different data sources



Note: The graphs display overlapping data on total central government expenditure from Eurostat (2012) and Mitchell (2007) as well as data on GDP from the IMF (2012b) and Mitchell (2007) in a time period from 1990 to 1995.

Figure 1.6: Interpolation of data from different sources



Note: The graphs display the development of total central government expenditure and GDP from 1970 to 2010. The vertical line highlights the interface between data provided by Mitchell (2007) and data provided by Eurostat (2012) and IMF (2012b), respectively.

1.A.2 Unit root results

In this section, we conduct a battery of integration tests in order to analyze the stationarity properties of the data series. Table 1.7 as well as table 1.8 display the results of the Phillips and Perron (1988) test (PP) and the Kwiatkowski et al. (1992) test (KPSS) in levels and first differences with different deterministic components in the test equations. In general, the test statistics provide coinciding results, concluding that most data series have a unit root in levels and are stationary in first differences.

In the presence of structural breaks, PP and KPSS tests have low power and are biased towards a non-rejection of a unit root or a rejection of stationarity, respectively. As historical time series are often associated with changes in the drift parameter or trend function, we additionally perform a couple of unit root tests allowing for single and multiple structural breaks. Table 1.9 displays the results of the Zivot and Andrews (1992) test and the Lanne et al. (2002) test, which allow for one endogenously chosen structural break. Again, it can be seen that the data can be treated as integrated of order one.

Finally, we apply the Lumsdaine and Papell (1997) unit root approach (LP) allowing for multiple structural breaks at unknown time (see table 1.10). The LP approach tests the null hypothesis that the series has a unit root against the alternative of stationarity with structural changes allowing for multiple structural breaks at unknown time. The breakpoint dates correspond to the minimized value of the sequentially obtained ADF statistic. These test statistics also indicate that the variables can be treated as integrated of order one.

Table 1.7: Results of unit root tests for different sample sizes (with constant and trend)

		Philips-Perron Test								Kwiatkowski-Phillips-Schmidt-Shin Test							
		Full Sample		Stage I		Stage II		Stage III		Full Sample		Stage I		Stage II		Stage III	
Country	Variable	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
United Kingdom	exp	-2.319	-6.398***	-2.367	-6.761***	-2.990	-4.889***	0.169	-3.837***	0.305***	0.193	0.509**	0.500**	0.045	0.035	0.193**	0.630**
	exppc	-2.223	-6.569***	-2.439	-6.665***	-3.003	-4.933***	-0.304	-4.087***	0.313***	0.220	0.502**	0.500**	0.044	0.037	0.213**	0.576**
	gdp	-0.723	-8.703***	-1.735	-8.318***	-1.294	-5.315***	-1.993	-4.678***	0.373***	0.745***	0.198**	0.248	0.234***	0.308	0.191**	0.581**
	gdppc	-0.727	-8.615***	-1.799	-8.149***	-1.363	-5.275***	-2.133	-4.747***	0.376***	0.764***	0.199**	0.279	0.240***	0.339	0.187**	0.452*
Denmark	exp	-2.039	-13.351***	-3.746**	-15.739***	-1.377	-5.350***	-1.206	-3.527**	0.345***	0.591**	0.235***	0.500**	0.164**	0.142	0.066	0.295
	exppc	-1.993	-13.318***	-3.792**	-16.007***	-1.345	-5.367***	-1.539	-3.491**	0.352***	0.664**	0.235***	0.500**	0.168**	0.150	0.076	0.235
	gdp	-1.511	-8.873***	-1.666	-7.068***	-1.135	-5.482***	-2.934	-5.299***	0.361***	0.483**	0.114	0.102	0.151**	0.183	0.179**	0.468*
	gdppc	-1.383	-8.791***	-1.786	-7.026***	-1.073	-4.889***	-3.096	-5.522***	0.369***	0.599**	0.109	0.093	0.157**	0.198	0.164**	0.359*
Sweden	exp	-2.667	-11.726***	-2.539	-7.687***	-2.543	-5.132***	-2.695	-6.027***	0.244***	0.168	0.157**	0.093	0.067	0.093	0.174**	0.373*
	exppc	-2.673	-11.717***	-2.545	-7.696***	-2.504	-5.119***	-2.917	-6.409***	0.253***	0.174	0.156**	0.090	0.069	0.084	0.153**	0.261
	gdp	-1.906	-6.290***	-2.143	-3.709**	-2.664	-3.249**	-2.005	-4.551***	0.337***	0.579**	0.166**	0.145	0.152**	0.437*	0.212**	0.850***
	gdppc	-1.918	-6.248***	-2.135	-3.724**	-2.621	-3.349**	-2.444	-4.359***	0.341***	0.648**	0.164**	0.137	0.151**	0.458*	0.205**	0.678**
Finland	exp	-1.778	-9.669***	-2.046	-5.932***	-3.076	-6.502***	-1.746	-3.373**	0.153**	0.148	0.154**	0.142	0.175**	0.315	0.117	0.089
	exppc	-1.992	-9.585***	-2.054	-5.881***	-3.045	-6.487***	-1.731	-3.475**	0.143**	0.140	0.157**	0.149	0.167**	0.303	0.125*	0.099
	gdp	-2.345	-5.234***	-1.894	-2.674*	-1.343	-4.296***	-2.283	-4.813***	0.248***	0.284	0.160**	0.186	0.162**	0.235	0.132*	0.112
	gdppc	-2.351	-5.233***	-1.899	-2.661*	-1.390	-4.343***	-2.255	-4.389***	0.258***	0.305	0.163**	0.196	0.158***	0.223	0.122*	0.095
Italy	exp	-2.229	-11.106***	-1.867	-9.305***	-2.819	-3.888***	-1.541	-5.054***	0.279***	0.281	0.219***	0.154	0.133*	0.310	0.163**	0.298
	exppc	-2.239	-11.069***	-1.871	-9.248***	-2.809	-3.885***	-1.428	-4.953***	0.283***	0.288	0.218***	0.150	0.133*	0.309	0.167**	0.339
	gdp	-2.145	-4.893***	-1.465	-6.286***	-1.937	-2.235	-1.471	-3.522**	0.295***	0.192	0.246***	0.251	0.159**	0.277	0.126*	0.191
	gdppc	-2.149	-4.983***	-1.488	-6.168***	-1.934	-2.230	0.934	-1.845	0.299***	0.339	0.243***	0.241	0.159**	0.276	0.145*	0.434*

Note: The Phillips-Perron (PP) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the levels are calculated including a constant and a trend in the test equation. The test statistics for the first differences are calculated including a constant. The bandwidth for the PP and KPSS test is selected based on Newey-West using Bartlett kernel. All variables in the full sample, stage I as well as stage II, are log-transformed, while stage III displays the results of non-transformed level data. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from MacKinnon (1996) and Kwiatkowski et al. (1992).

Table 1.8: Results of unit root tests for different sample sizes (with constant, no trend)

		Philips-Perron Test								Kwiatkowski-Phillips-Schmidt-Shin Test							
		Full Sample		Stage I		Stage II		Stage III		Full Sample		Stage I		Stage II		Stage III	
Country	Variable	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
United Kingdom	exp	0.654	-6.398***	-1.924	-6.761***	-0.661	-4.889***	4.850	-3.837***	1.500***	0.192	0.657**	0.500**	1.127***	0.035	0.719***	0.630**
	exppc	0.682	-6.569***	-2.573	-6.665***	-0.690	-4.933***	3.998	-4.087***	1.479***	0.220	0.151	0.500*	1.119***	0.037	0.723***	0.576**
	gdp	2.472	-8.703***	-2.515	-8.318***	1.452	-5.315***	2.223	-4.678***	1.448***	0.745***	0.681**	0.248	1.146***	0.308	0.724***	0.581**
	gdppc	2.511	-8.615***	-2.790*	-8.194***	1.519	-5.275***	1.639	-4.748***	1.410***	0.764***	0.630**	0.279	1.138***	0.339	0.727***	0.452*
Denmark	exp	1.543	-13.351***	-2.522	-15.739***	0.371	-5.350***	1.709	-3.527**	1.429***	0.591**	0.586**	0.500**	0.899***	0.142	0.829***	0.295
	exppc	1.616	-13.318***	-3.858***	-16.007***	0.382	-5.367***	1.314	-3.491**	1.391***	0.664**	0.237	0.500**	0.885***	0.150	0.827***	0.235
	gdp	1.573	-8.873***	-0.001	-7.068***	0.475	-5.482***	1.757	-5.299***	1.510***	0.483**	0.956***	0.102	0.925***	0.183	0.813***	0.468*
	gdppc	1.858	-8.791***	-0.327	-7.026***	0.553	-4.889***	1.385	-5.522***	1.474***	0.599**	0.893***	0.093	0.912***	0.198	0.815***	0.359*
Sweden	exp	0.372	-11.726***	-0.458	-7.687***	0.257	-5.132***	1.522	-6.027***	1.386***	0.168	0.773***	0.093	0.796***	0.093	0.812***	0.373*
	exppc	0.395	-11.717***	-0.563	-7.696***	0.162	-5.119***	0.944	-6.409***	1.382***	0.174	0.758***	0.090	0.794***	0.084	0.813***	0.261
	gdp	1.849	-6.290***	0.067	-3.709***	1.912	-3.249**	3.789	-4.551***	1.423***	0.579**	0.788***	0.145	0.794***	0.437*	0.807***	0.850***
	gdppc	1.932	-6.248***	-0.061	-3.724***	1.801	-3.349**	2.691	-4.359***	1.410***	0.648**	0.773***	0.137	0.793***	0.458*	0.809***	0.678**
Finland	exp	-0.496	-9.669***	0.001	-5.932***	-2.544	-6.502***	-0.441	-3.373**	1.379***	0.148	0.845***	0.142	0.793***	0.315	0.720***	0.089
	exppc	-0.394	-9.585***	-0.029	-5.881***	-2.441	-6.487***	-0.719	-3.475**	1.377***	0.140	0.831***	0.149	0.793***	0.303	0.712**	0.099
	gdp	0.544	-5.234***	0.229	-2.674*	-1.449	-4.296***	0.099	-4.813***	1.447***	0.284	0.832***	0.186	0.768***	0.235	0.739***	0.112
	gdppc	0.607	-5.233***	0.176	-2.661*	-1.397	-4.343***	-0.355	-4.389***	1.437***	0.305	0.812***	0.196	0.767***	0.223	0.738***	0.095
Italy	exp	0.541	-11.106***	-0.062	-9.305***	-2.392	-3.888***	-1.752	-5.054***	1.403***	0.281	1.043***	0.154	0.726***	0.310	0.726***	0.298
	exppc	0.548	-11.069***	-0.199	-9.248***	-2.419	-3.885***	-2.096	-4.953***	1.395***	0.288	0.999***	0.150	0.724***	0.309	0.717***	0.339
	gdp	0.700	-4.893***	0.451	-6.286***	-2.216	-2.235	-0.672	-3.522**	1.500***	0.331	1.037***	0.251	0.699**	0.277	0.658**	0.191
	gdppc	0.703	-4.850***	0.251	-6.168***	-2.229	-2.230	-1.369	-1.845	1.388***	0.339	0.978***	0.241	0.694**	0.276	0.657**	0.434*

Note: The Phillips-Perron (PP) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the levels and first differences of the variables are calculated including a constant in the test equation. The bandwidth for the PP and KPSS test is selected based on Newey-West using Bartlett kernel. All variables in the full sample, stage I as well as stage II are log-transformed, while stage III displays the results of non-transformed level data. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% level using critical values from MacKinnon (1996) and Kwiatkowski et al. (1992).

Table 1.9: Results of unit root tests for different sample sizes with one structural break

		Zivot-Andrews Test								Lanne, Lütkepohl and Saikkonen Test							
		Full Sample		Stage I		Stage II		Stage III		Full Sample		Stage I		Stage II		Stage III	
Country	Variable	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
UK	exp	-4.628	-7.306***	-4.869	-7.235***	-4.344	-5.681***	-1.836	-6.229***	-2.284	-6.024***	-2.529	-5.196***	-2.033	-5.131***	-0.679	-5.218***
		(1914)	(1918)	(1869)	(1873)	(1922)	(1918)	(1999)	(1996)	(1914)	(1918)	(1869)	(1873)	(1922)	(1918)	(1999)	(1996)
	exppc	-4.588	-7.291***	-4.848	-7.899***	-5.467**	-5.676***	-1.927	-5.909***	-2.136	-5.877***	-2.455	-5.475***	-2.072	-5.169***	-0.770	-4.841***
		(1914)	(1918)	(1869)	(1867)	(1914)	(1918)	(1999)	(1996)	(1914)	(1918)	(1869)	(1867)	(1914)	(1918)	(1999)	(1996)
	gdp	-3.734	-7.306***	-3.532	-5.318**	-4.096	-6.194***	-3.251	-7.361***	-1.392	-3.406**	-1.418	-3.072**	-2.658	-5.933***	-0.663	-3.204**
		(1921)	(1918)	(1876)	(1875)	(1926)	(1920)	(1994)	(1994)	(1921)	(1918)	(1876)	(1875)	(1926)	(1920)	(1994)	(1994)
	gdppc	-3.887	-6.549***	-3.554	-5.377**	-3.933	-5.915***	-2.958	-7.051***	-1.392	-4.064***	-0.620	-2.863**	-2.672	-5.978***	-0.759	-3.148**
		(1920)	(1971)	(1876)	(1875)	(1926)	(1920)	(1994)	(1994)	(1920)	(1971)	(1876)	(1875)	(1926)	(1920)	(1994)	(1994)
Denmark	exp	-2.796	-8.003***	-4.791	-12.722***	-3.823	-8.369***	-2.623	-5.572***	-0.822	-13.181***	-2.685	-9.080***	-1.417	-2.806*	-1.958	-3.959***
		(1899)	(1982)	(1878)	(1866)	(1922)	(1920)	(1981)	(1997)	(1899)	(1982)	(1878)	(1866)	(1922)	(1920)	(1981)	(1997)
	exppc	-2.899	-8.023***	-4.812	-12.712***	-3.835	-8.408***	-2.751	-5.423**	-0.728	-6.897***	-4.231***	-9.094***	-1.422	-2.790*	-1.994	-3.931***
		(1999)	(1962)	(1878)	(1866)	(1922)	(1920)	(1981)	(1995)	(1999)	(1962)	(1878)	(1866)	(1922)	(1920)	(1981)	(1995)
	gdp	-3.259	-5.279**	-4.094	-6.502***	-4.664	-8.876***	-3.105	-6.325***	-1.783	-2.799*	-1.943	-3.966***	-1.552	-3.534***	-1.152	-5.109***
		(1921)	(1972)	(1884)	(1877)	(1941)	(1921)	(1977)	(1982)	(1921)	(1972)	(1884)	(1877)	(1941)	(1921)	(1977)	(1982)
	gdppc	-3.552	-5.393**	-4.379	-6.465***	-4.607	-8.903***	-3.658	-6.512***	-1.721	-6.492***	-2.440	-3.998***	-1.539	-3.556***	-1.156	-4.818***
		(1921)	(1972)	(1884)	(1877)	(1941)	(1921)	(1983)	(1982)	(1921)	(1972)	(1884)	(1877)	(1941)	(1921)	(1983)	(1982)
Sweden	exp	-3.544	-5.336**	-3.345	-5.977***	-6.083***	-6.439***	-3.761	-7.489***	-1.586	-4.019***	-2.019	-5.865***	-2.008	-4.969***	-1.295	-4.898***
		(1923)	(1919)	(1905)	(1917)	(1940)	(1942)	(1992)	(1997)	(1923)	(1919)	(1905)	(1917)	(1940)	(1942)	(1992)	(1997)
	exppc	-3.662	-5.312**	-3.381	-5.950***	-5.847***	-6.443***	-4.2931	-4.126	-1.615	-3.973***	-2.077	-5.881***	-2.284	-5.011***	-1.5770	-4.941***
		(1923)	(1919)	(1905)	(1917)	(1940)	(1942)	(1995)	(1998)	(1923)	(1919)	(1905)	(1917)	(1940)	(1942)	(1995)	(1998)
	gdp	-4.534	-7.260***	-4.631	-5.681***	-4.542	-6.098***	-2.607	-6.647***	-1.107	-6.702***	-2.463	-3.204**	-1.416	-5.452***	-0.337	-4.228***
		(1921)	(1921)	(1915)	(1918)	(1935)	(1934)	(1996)	(1992)	(1921)	(1921)	(1915)	(1918)	(1935)	(1934)	(1996)	(1992)
	gdppc	-4.588	-7.276***	-4.735	-5.702***	-5.063*	-6.182***	-2.881	-6.704***	-0.891	-6.560***	-2.533	-3.211**	-1.514	-5.467***	-0.406	-3.128**
		(1921)	(1921)	(1915)	(1918)	(1935)	(1934)	(1996)	(1992)	(1921)	(1921)	(1915)	(1918)	(1935)	(1934)	(1996)	(1992)
Finland	exp	-4.027	-5.474**	-5.993***	-7.170***	-4.464	-8.263***	-3.492	-6.883***	-1.124	-10.267***	-1.438	-5.993***	-0.380	-5.813***	-1.351	-3.119**
		(1939)	(1916)	(1917)	(1916)	(1952)	(1952)	(1996)	(1996)	(1939)	(1916)	(1917)	(1916)	(1952)	(1952)	(1996)	(1996)
	exppc	-4.103	-5.444**	-6.047***	-7.122***	-4.414	-8.302***	-4.004	-6.735	-1.127	-7.932***	-1.417	-5.945***	-0.436	-5.799***	-1.329	-2.854*
		(1939)	(1922)	(1917)	(1916)	(1969)	(1952)	(1996)	(1996)	(1939)	(1922)	(1917)	(1916)	(1969)	(1952)	(1996)	(1996)
	gdp	-4.022	-5.811***	-6.625***	-4.744	-3.008	-6.975***	-4.878*	-5.711***	-1.170	-6.140***	-2.258	-3.011*	-1.245	-4.015***	-1.718	-4.435***
		(1945)	(1915)	(1916)	(1915)	(1951)	(1952)	(1993)	(1996)	(1945)	(1915)	(1916)	(1915)	(1951)	(1952)	(1993)	(1996)
	gdppc	-3.952	-5.805***	-6.625***	-4.654	-2.968	-6.978***	-4.849*	-5.608***	-1.070	-6.102***	-2.209	-2.969*	-1.241	-4.071***	-2.566	-4.162***
		(1945)	(1915)	(1916)	(1915)	(1972)	(1952)	(1991)	(1996)	(1945)	(1915)	(1916)	(1915)	(1972)	(1952)	(1991)	(1996)
Italy	exp	-3.335	-5.198**	-6.726***	-10.205***	-3.663	-9.300***	-3.145	-6.135***	-0.830	-3.972***	-2.226	-6.361***	-1.505	-5.337***	-1.512	-5.873***
		(1940)	(1936)	(1915)	(1915)	(1959)	(1949)	(1997)	(1993)	(1940)	(1936)	(1915)	(1915)	(1959)	(1949)	(1997)	(1993)
	exppc	-3.383	-7.764***	-6.707***	-10.169***	-3.385	-9.358***	-3.169	-6.044***	-0.866	-9.759***	-2.198	-6.389***	-1.237	-5.392***	-1.574	-5.116***
		(1940)	(1940)	(1915)	(1915)	(1972)	(1949)	(1997)	(1993)	(1940)	(1940)	(1915)	(1915)	(1972)	(1949)	(1997)	(1993)
	gdp	-3.882	-6.281***	-4.849	-5.508**	-4.218	-9.546***	-2.857	-5.749***	-1.255	-5.267***	-1.009	-6.015***	-2.382	-1.371	-1.314	-4.077***
		(1943)	(1940)	(1916)	(1927)	(1952)	(1948)	(2005)	(2004)	(1943)	(1940)	(1916)	(1927)	(1952)	(1948)	(2005)	(2004)
	gdppc	-3.870	-6.209***	-5.055*	-5.359**	-4.219	-9.514***	-2.521	-5.016*	-1.262	-5.229***	-1.009	-5.949***	-2.392	-1.351	-1.495	-4.216***
		(1943)	(1940)	(1916)	(1927)	(1952)	(1948)	(2006)	(2005)	(1943)	(1940)	(1916)	(1927)	(1952)	(1948)	(2006)	(2005)

Note: The model of the Zivot-Andrews test allows for a structural break in the intercept and the slope of the trend function ($\Delta y_t = \mu + \beta t + \theta DU1_t + \gamma DT1_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i}$). The model of the Lanne, Lütkepohl and Saikkonen test allows for a level shift ($\Delta y_t = \mu + \beta t + \theta DU1_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i}$). The break dates are chosen endogenously as the value that minimizes the t-statistic in the Zivot-Andrews test regression (values in parenthesis). The lag length for both test specifications is selected by the Schwarz information criterion. All variables in the full sample, stage I as well as stage II are log-transformed, while stage III displays the results of non-transformed level data. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from Zivot and Andrews (1992) and Lanne et al. (2002).

Table 1.10: Results of Lumsdaine and Papell unit root test for different sample sizes

		Full Sample				Stage I				Stage II				Stage III			
Country	Variable	Breaks	Level	Breaks	1st Diff.	Breaks	Level	Breaks	1st Diff.	Breaks	Level	Breaks	1st Diff.	Breaks	Level	Breaks	1st Diff.
UK	exp	1914	-0.167	1914	-0.903	1865	-0.668	1856	-2.103	1914	-0.173	1914	-0.863	1986	-0.194	1987	-1.581
		1931	(-6.154)	1923	(-10.532)***	1869	(-5.041)	1879	(-12.648)***	1939	(-6.440)	1923	(-8.618)***	1997	(-2.699)	1997	(-8.076)***
	exppc	1914	-0.199	1914	-0.855	1859	-0.713	1856	-1.401	1914	-0.357	1914	-0.869	1986	-0.199	1987	-1.556
		1931	(-6.274)	1923	(-11.311)***	1869	(-5.079)	1879	(-7.661)***	1939	(-6.912)**	1923	(-8.549)***	1997	(-2.685)	1996	(-7.462)***
	gdp	1894	-0.066	1915	-0.638	1859	-0.485	1875	-1.519	1915	-0.199	1915	-0.904	1978	-0.871	1994	-1.379
		1921	(-4.731)	1924	(-7.023)***	1876	(-3.184)	1880	(-10.898)***	1933	(-5.191)	1927	(-6.938)**	1991	(-4.063)	2003	(-8.364)***
	gdppc	1908	-0.070	1915	-0.602	1861	-0.539	1875	-1.514	1915	-0.286	1915	-0.819	1978	-0.774	1994	-1.336
		1921	(-4.874)	1924	(-6.744)*	1876	(-3.493)	1880	(-10.869)***	1940	(-6.088)	1924	(-8.609)***	1990	(-3.588)	2003	(-8.272)***
Denmark	exp	1899	-0.095	1912	-1.064	1870	-1.416	1866	-2.513	1917	-0.404	1916	-0.992	1977	-0.221	1977	-1.055
		1969	(-2.942)	1920	(-9.252)***	1879	(-5.668)	1900	(-7.382)***	1935	(-5.421)	1920	(-8.763)***	2005	(-2.568)	2008	(-6.165)
	exppc	1912	-0.088	1912	-1.068	1866	-2.321	1866	-2.518	1917	-0.417	1916	-0.997	1977	-0.168	1977	-1.026
		1922	(-3.232)	1920	(-9.280)***	1886	(-5.580)	1900	(-7.372)***	1935	(-5.593)	1920	(-8.808)***	2005	(-2.196)	2004	(-6.072)
	gdp	1895	-0.062	1914	-0.807	1865	-0.399	1865	-1.074	1916	-0.343	1921	-1.316	1974	-0.364	1983	-1.246
		1921	(-4.203)	1933	(-10.488)***	1884	(-4.953)	1880	(-7.903)***	1936	(-4.730)	1945	(-9.705)***	2005	(-3.337)	1994	(-7.353)***
	gdppc	1905	-0.073	1914	-0.802	1865	-0.411	1865	-1.065	1916	-0.362	1916	-0.997	1973	-0.530	1983	-1.298
		1921	(-4.849)	1933	(-10.419)***	1884	(-4.792)	1880	(-7.812)***	1936	(-4.862)	1920	(-8.807)***	1983	(-3.791)	1994	(-7.584)***
Sweden	exp	1914	-0.088	1917	-1.167	1894	-1.050	1915	-2.026	1931	-0.666	1940	-1.844	1973	-0.643	1990	-1.239
		1919	(-3.079)	1919	(-16.114)***	1918	(-5.761)	1919	(-15.503)***	1940	(-5.945)	1948	(-8.881)***	1993	(-4.654)	1997	(-8.175)***
	exppc	1914	-0.102	1917	-1.167	1895	-1.087	1913	-3.925	1931	-0.643	1931	-1.836	1973	-0.609	1990	-1.244
		1919	(-3.419)	1919	(-16.125)***	1918	(-5.982)	1920	(-7.397)***	1940	(-5.720)	1939	(-8.936)***	1993	(-4.484)	1997	(-8.289)***
	gdp	1902	-0.085	1915	-0.739	1890	-0.380	1888	-1.168	1931	-0.483	1931	-1.269	1976	-0.457	1973	-1.846
		1921	(-4.884)	1925	(-8.252)***	1915	(-5.066)	1915	(-5.654)	1951	(-9.112)***	1937	(-8.134)***	1992	(-3.602)	1992	(-7.292)***
	gdppc	1905	-0.088	1915	-0.747	1893	-0.543	1888	-1.174	1931	-0.513	1931	-1.281	1976	-0.496	1982	-1.634
		1921	(-5.109)	1925	(-8.324)***	1918	(-5.160)	1915	(-5.673)	1951	(-9.434)***	1937	(-8.283)***	1993	(-3.217)	1996	(-6.734)*
Finland	exp	1916	-0.101	1917	-1.324	1894	-0.366	1917	-1.579	1946	-0.412	1943	-1.288	1982	-0.301	1990	-1.679
		1939	(-3.859)	1934	(-10.306)***	1917	(-5.804)	1931	(-7.921)***	1973	(-6.097)	1947	(-7.639)***	1995	(-3.990)	1995	(-10.521)***
	exppc	1916	-0.116	1920	-1.311	1898	-0.368	1917	-1.556	1946	-0.416	1943	-1.294	1982	-0.337	1990	-1.711
		1919	(-4.156)	1937	(-10.206)***	1917	(-5.852)	1931	(-7.790)***	1969	(-6.044)	1947	(-7.709)***	1995	(-4.483)	1995	(-10.313)***
	gdp	1915	-0.076	1916	-0.605	1889	-0.232	1916	-0.718	1943	-0.253	1945	-1.101	1983	-1.133	1991	-1.769
		1945	(-4.729)	1932	(-8.319)***	1917	(-5.719)	1929	(-6.991)**	1945	(-4.465)	1952	(-9.481)***	1993	(-4.379)	2001	(-6.792)*
	gdppc	1916	-0.089	1916	-0.605	1889	-0.234	1916	-0.715	1943	-0.254	1945	-1.104	1983	-1.056	1991	-1.660
		1945	(-5.012)	1932	(-8.331)***	1917	(-5.757)	1929	(-6.885)**	1945	(-4.238)	1952	(-9.483)***	1993	(-4.174)	2001	(-6.776)*
Italy	exp	1915	-0.101	1915	-1.057	1885	-0.459	1916	-2.444	1946	-0.387	1946	-1.312	1983	-0.371	1993	-1.925
		1940	(-3.663)	1926	(-13.191)***	1915	(-6.355)	1923	(-8.829)***	1970	(-5.003)	1949	(-11.057)***	1997	(-3.236)	2001	(-7.181)***
	exppc	1915	-0.112	1915	-1.056	1904	-0.591	1916	-2.429	1946	-0.389	1943	-1.312	1983	-0.396	1993	-1.946
		1940	(-3.872)	1926	(-13.175)***	1916	(-6.459)	1923	(-8.801)***	1970	(-5.029)	1947	(-11.017)***	1997	(-3.279)	2001	(-6.753)*
	gdp	1915	-0.099	1939	-0.712	1892	-0.202	1916	-1.232	1945	-0.224	1946	-0.713	1982	-0.519	1996	-1.832
		1944	(-5.579)	1948	(-10.257)***	1916	(-5.015)	1929	(-11.179)***	1973	(-3.968)	1948	(-7.094)**	2007	(-3.389)	2006	(-6.580)*
	gdppc	1915	-0.110	1938	-0.688	1892	-0.214	1916	-1.223	1945	-0.224	1946	-0.711	1982	-0.754	1988	-1.442
		1944	(-5.891)	1948	(-10.091)***	1916	(-5.339)	1929	(-11.170)***	1973	(-3.971)	1948	(-7.075)**	2007	(-4.857)	2006	(-6.808)*

Note: The model allows for two breaks in the intercept and the slope of the trend function ($\Delta y_t = \mu + \beta t + \theta DU1_t + \gamma DT1_t + \varpi DU2_t + \varphi DT2_t + \alpha y_{t-1} + \sum_{i=1}^{k} c_i \Delta y_{t-i}$). T-values of α are shown in parenthesis. The lag length is selected by a general-to-specific method following Lumsdaine and Papell (1997) with k_{max}=3. The trimming parameter is set to 0.1. All variables in the full sample, stage I as well as stage II, are log-transformed, while stage III displays the results of non-transformed level data. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from Lumsdaine and Papell (1997).

1.A.3 Cointegration results

A weak interpretation of Wagner's theory suggests that there is not necessarily a cause-and-effect relationship between government expenditure and economic output (see, e.g., Peacock and Scott (2000)). Table 1.3 in the main part of this chapter summarizes the cointegration results for different countries, functional forms, and development stages. This section discusses these findings in more detail and provides additional methodological information on the testing procedures.

To analyze the existence of a long-run equilibrium relationship among government expenditure and GDP, we apply the VAR-based cointegration procedure developed by Johansen (1988) and Johansen and Juselius (1990). The approach of testing for a cointegration vector relies on a first-difference VAR of order p:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-1} + C x_t + \epsilon_t \tag{1.3}$$

where y_t represents a vector of non-stationary I(1) variables containing GDP and expenditure in version 1, expenditure and GDP per capita in version 2, and expenditure per capita and GDP per capita in version 3. The vector X_t contains deterministic variables and ϵ_t normally distributed random error terms.

The cointegration results are very sensitive to the deterministic trend assumption and the choice of the order p in equation (1.3). According to the data of expenditure and GDP, it can be seen that the time series follow a linear trend in the log level data. Therefore, as suggested by Franses (2001), our test specification allows for a linear trend in the level data and a constant in the cointegration space $[H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0]$. In order to additionally include the case that an individual series might be trend-stationary, we also apply the Johansen test specification allowing for a constant and a trend in the cointegration space $[H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0]$.[31] The optimal lag length in the test specifications were chosen by the Schwarz information criterion. To obviate spurious cointegration, the lag length of the VAR was successively enhanced to remove all serial correlation from the data considering a maximum of 5 lags.

To test for the number of cointegration vectors, Johansen (1988) and Johansen and Juselius (1990) propose two maximum likelihood test statistics (L_{Eigen} and L_{Trace}). In the bivariate case of L_{Eigen}, the null hypothesis of r cointegrating vectors is tested against the alternative of $r+1$ cointegrating vectors, while in the bivariate case of L_{Trace}, the null hypothesis is tested that there are at most r cointegration vectors in the system against its general alternative. The test statistics to test for the reduced rank of the π matrix are computed by

$$L_{Eigen} = -T \cdot ln(1 - \hat{\lambda}_{r+1}) \qquad \text{and} \qquad L_{Trace} = -T \sum_{i=r+1}^{p-2} ln(1 - \hat{\lambda}_i), \tag{1.4}$$

where T is the sample size and $\hat{\lambda}_{r+1}$,..., $\hat{\lambda}_n$ are the smallest characteristic roots.[32]

[31] As outlined by Franses (2001), this specification seems to be the most important case for practical purposes.

[32] Lütkepohl et al. (2001) found that the local power of corresponding maximum eigenvalue and trace tests is very similar. In small samples, however, the trace test tends to have superior power. Yet, the authors recommend to apply both tests simultaneously in empirical works.

Testing for a long-run relationship considering structural breaks

The characteristics of historical time series covering data of major social disturbances (World War I and II, Great Depression etc.) make the conventional cointegration procedure particularly vulnerable to a non-rejection of the no cointegration hypothesis although the true data generation process of the variables share a common stochastic trend. In order to account for possible structural breaks and regime shifts into the cointegration analysis, we enhance the basic Johansen testing procedure allowing for multiple structural breaks at unknown time.

According to Johansen et al. (2000), the first-difference VAR can be rewritten with q equations, assuming that the data contains $q-1$ breaks. By introducing dummy variables, equation (1.3) can be rearranged as

$$\Delta y_t = \alpha \begin{pmatrix} \beta \\ \\ \gamma \end{pmatrix}' \begin{pmatrix} y_{t-1} \\ \\ t \cdot E_t \end{pmatrix} + \mu \cdot E_t + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-1} + \sum_{i=0}^{k-1} \sum_{j=2}^{q} \Theta_{j,i} D_{j,t-i} + \epsilon_t. \tag{1.5}$$

with $j = 1, ..., q$ and the defined matrices $E_t = (E_{1,t}, ...E_{q,t})'$, $\mu = (\mu_1, ...\mu_q)$ and $\gamma = (\gamma_1', ..., \gamma_q')'$ of dimension (q $\times$ 1), (p $\times$ q), (q $\times$ r), respectively. The $q-1$ intervention dummies are defined as $D_{j,t} = 1$ for $t = T_{j-1}$; $= 0$ otherwise for all $j = 2, ..., q$. $D_{j,t-i}$ is an indicator function for the i-th observation in the j-th period, that is, $D_{j,t-i} = 1$ if $t = T_{j-1}+i$. Further, the effective sample of the j-th period is defined as $E_{j,t} = \sum_{i=k+1}^{T_j - T_{j-1}} D_{j,t} = 1$ for $T_{j-1} + k + 1 \leq t \leq T_j$; $= 0$ otherwise with k determining the order of the vector autoregressive model.[33]

The likelihood ratio test statistics remain unchanged, while the computation of the critical values depends on the number of non-stationary relations and the location of the break points (see Johansen et al. (2000)). As with the basic cointegration procedure, we again assume that the time series follow a trend in levels. Under this condition, we consider two different models of structural breaks: 1) breaks in levels only, which are restricted to the error correction terms, and 2) breaks in level and trend jointly (regime shift), while the trend shifts are restricted to the error correction term and the level shifts are unrestricted in the model.

In order to locate possible structural breaks, we apply the multiple structural breakpoint test developed by Bai and Perron (1998). The intuition behind this testing procedure is an algorithm that searches all possible sets of breaks and calculates a goodness-of-fit measure for each number. By implementing a sequential supremum F-statistic (SupF) testing procedure, the null of l breaks is tested against the alternative of $l+1$ breaks. The number of break dates selected is the number associated with the overall minimum error sum of squares.[34] The model specification to test for parameter instability in the various variables of expenditure and GDP follows an AR(p) process with a constant. In order to guarantee sufficiently large subsamples, the trimming parameter was set to 0.3 allowing for a maximum of two possible breaks in each analyzed sample.

[33] A detailed theoretical as well as practical application of the Johansen et al. (2000) procedure is provided by Joyeux (2007).

[34] For a detailed and formal presentation of the Bai-Perron framework see Bai and Perron (1998, 2003).

Full sample results

Table 1.11 and table 1.12 display the cointegration results over the full sample period. The outcome indicates that there is evidence of cointegration for all three examined versions of Wager's law across countries at least at a significance niveau of 10 percent. The significance of these results increases if structural breaks are included in the analysis.

Table 1.11: Results of Johansen cointegration test (Full sample)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	exp and gdp	r=0	11.216	r=0	8.479	r=0	22.275*	r=0	16.747
(1850-2010)		r≤1	2.736	r=1	2.736	r≤1	5.528	r=1	5.528
	exp and gdppc	r=0	9.634	r=0	8.004	r=0	24.417*	r=0	17.968*
		r≤1	1.629	r=1	1.629	r≤1	6.449	r=1	6.449
	exppc and gdppc	r=0	11.257	r=0	9.008	r=0	22.968*	r=0	16.760
		r≤1	2.249	r=1	2.249	r≤1	6.207	r=1	6.207
Denmark	exp and gdp	r=0	8.317	r=0	7.913	r=0	23.445*	r=0	15.992
(1854-2010)		r≤1	0.404	r=1	0.404	r≤1	7.453	r=1	7.453
	exp and gdppc	r=0	11.765	r=0	11.475	r=0	22.365	r=0	14.781
		r≤1	0.289	r=1	0.289	r≤1	7.584	r=1	7.584
	exppc and gdppc	r=0	9.031	r=0	8.355	r=0	23.326*	r=0	15.877
		r≤1	0.676	r=1	0.676	r≤1	6.875	r=1	6.875
Sweden	exp and gdp	r=0	16.738**	r=0	16.719**	r=0	24.817*	r=0	18.371*
(1881-2010)		r≤1	0.019	r=1	0.019	r≤1	6.447	r=1	6.447
	exp and gdppc	r=0	16.179**	r=0	15.915**	r=0	25.567**	r=0	19.105**
		r≤1	0.264	r=1	0.264	r≤1	6.462	r=1	6.462
	exppc and gdppc	r=0	17.072**	r=0	17.037**	r=0	25.579**	r=0	19.138**
		r≤1	0.034	r=1	0.034	r≤1	6.440	r=1	6.440
Finland	exp and gdp	r=0	14.804*	r=0	13.876*	r=0	22.833*	r=0	13.933
(1882-2010)		r≤1	0.928	r=1	0.928	r≤1	8.901	r=1	8.901
	exp and gdppc	r=0	13.763*	r=0	12.437*	r=0	21.579	r=0	12.442
		r≤1	1.326	r=1	1.326	r≤1	9.138	r=1	9.138
	exppc and gdppc	r=0	14.500*	r=0	13.621*	r=0	23.385*	r=0	13.799
		r≤1	0.879	r=1	0.879	r≤1	9.587	r=1	9.587
Italy	exp and gdp	r=0	18.678**	r=0	18.651***	r=0	25.804**	r=0	19.269**
(1862-2010)		r≤1	0.027	r=1	0.027	r≤1	6.535	r=1	6.535
	exp and gdppc	r=0	18.295**	r=0	18.179**	r=0	27.221**	r=0	20.657**
		r≤1	0.115	r=1	0.115	r≤1	6.564	r=1	6.564
	exppc and gdppc	r=0	18.692**	r=0	18.652***	r=0	25.995**	r=0	19.382**
		r≤1	0.040	r=1	0.040	r≤1	6.613	r=1	6.613

Note: Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

Table 1.12: Johansen cointegration test with structural breaks (Full sample)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	exp and gdp	1916	1936	r=0	34.11*	r=0	25.81
(1850-2010)				r≤1	12.50	r=1	9.04
	exp and gdppc	1916	1936	r=0	34.76*	r=0	26.12
				r≤1	13.33	r=1	8.41
	exppc and gdppc	1916	1936	r=0	41.77**	r=0	36.15
				r≤1	12.07	r=1	9.37
Denmark	exp and gdp	1900	1920	r=0	26.94	r=0	43.54*
(1854-2010)				r≤1	11.28	r=1	12.66
	exp and gdppc	1900	1920	r=0	27.09	r=0	45.35**
				r≤1	11.84	r=1	13.14
	exppc and gdppc	1900	1920	r=0	26.88	r=0	44.57**
				r≤1	11.38	r=1	12.93
Sweden	exp and gdp	-	1920	r=0	40.69**	r=0	58.67***
(1881-2010)			1966	r≤1	10.90	r=1	18.38
	exp and gdppc	-	1920	r=0	41.67**	r=0	58.00***
			1966	r≤1	11.12	r=1	18.76
	exppc and gdppc	-	1920	r=0	41.40**	r=0	57.69***
			1966	r≤1	11.34	r=1	18.31
Finland	exp and gdp	-	1920	r=0	36.00***	r=0	36.28**
(1882-2010)				r≤1	11.06	r=1	11.12
	exp and gdppc	-	1920	r=0	34.84***	r=0	35.27*
				r≤1	11.89	r=1	11.29
	exppc and gdppc	-	1920	r=0	36.47***	r=0	36.27**
				r≤1	11.58	r=1	11.29
Italy	exp and gdp	1935	1942	r=0	53.32***	r=0	73.50***
(1862-2010)				r≤1	8.31	r=1	13.77
	exp and gdppc	1935	1942	r=0	48.70***	r=0	72.83***
				r≤1	8.38	r=1	14.13
	exppc and gdppc	1935	1942	r=0	52.29***	r=0	73.59***
				r≤1	8.23	r=1	14.14

Note: The cointegration approach allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. Break specification: Johansen (1) allows only for breaks in levels that are restricted to the error correction term. Johansen (2) allows for breaks in trend and constant jointly with trend shifts restricted to the error correction term and level shifts unrestricted in model. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. Break points were estimated by Bai and Perron's (1998) procedure considering the following linear regression for each variable: $y_t = \mu + \psi_i \sum_{i=1}^{p} y_{t-i}$. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Johansen et al. (2000).

Stage I: Lower middle income

Table 1.13 and table 1.14 display the cointegration results of the first development stage. The formal test results detect a cointegration relationship for all countries with the exception of Denmark.

Table 1.13: Results of Johansen cointegration test (Stage I)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	exp and gdp	r=0	15.968**	r=0	12.779*	r=0	30.936***	r=0	27.242***
(1850-1885)		r≤1	3.189	r=1	3.189	r≤1	3.694	r=1	3.694
	exp and gdppc	r=0	15.671**	r=0	13.184**	r=0	30.652***	r=0	26.977***
		r≤1	2.487	r=1	2.487	r≤1	3.675	r=1	3.675
	exppc and gdppc	r=0	26.613***	r=0	23.252***	r=0	30.555***	r=0	26.932***
		r≤1	3.361	r=1	3.361	r≤1	3.622	r=1	3.622
Denmark	exp and gdp	r=0	6.788	r=0	6.055	r=0	12.941	r=0	6.967
(1854-1908)		r≤1	0.733	r=1	0.733	r≤1	5.975	r=1	5.975
	exp and gdppc	r=0	6.012	r=0	5.763	r=0	13.326	r=0	7.777
		r≤1	0.249	r=1	0.249	r≤1	5.549	r=1	5.549
	exppc and gdppc	r=0	5.734	r=0	5.655	r=0	13.436	r=0	7.819
		r≤1	0.079	r=1	0.079	r≤1	5.618	r=1	5.618
Sweden	exp and gdp	r=0	16.146**	r=0	13.153*	r=0	29.237**	r=0	16.998*
(1881-1925)		r≤1	2.992	r=1	2.992	r≤1	12.239	r=1	12.239
	exp and gdppc	r=0	17.398**	r=0	14.833**	r=0	27.639**	r=0	16.432
		r≤1	2.564	r=1	2.564	r≤1	11.207	r=1	11.207
	exppc and gdppc	r=0	15.938**	r=0	13.198*	r=0	28.296**	r=0	16.695
		r≤1	2.739	r=1	2.739	r≤1	11.600	r=1	11.600
Finland	exp and gdp	r=0	5.585	r=0	5.545	r=0	16.057	r=0	11.936
(1882-1937)		r≤1	0.039	r=1	0.039	r≤1	4.120	r=1	4.120
	exp and gdppc	r=0	4.449	r=0	4.107	r=0	15.428	r=0	11.423
		r≤1	0.342	r=1	0.342	r≤1	4.005	r=1	4.005
	exppc and gdppc	r=0	5.232	r=0	5.149	r=0	16.056	r=0	12.008
		r≤1	0.083	r=1	0.083	r≤1	4.048	r=1	4.048
Italy	exp and gdp	r=0	25.375***	r=0	25.345***	r=0	30.151**	r=0	25.345***
(1862-1939)		r≤1	0.029	r=1	0.029	r≤1	1.632	r=1	1.632
	exp and gdppc	r=0	21.908***	r=0	21.724***	r=0	30.540***	r=0	25.606***
		r≤1	0.183	r=1	0.183	r≤1	4.935	r=1	4.935
	exppc and gdppc	r=0	25.269***	r=0	25.113***	r=0	30.210***	r=0	25.296***
		r≤1	0.156	r=1	0.156	r≤1	4.914	r=1	4.914

Note: Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

Table 1.14: Johansen cointegration test with structural breaks (Stage I)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	exp and gdp	-	1875	r=0	35.66***	r=0	35.16**
(1850-1885)				r≤1	9.19	r=1	8.98
	exp and gdppc	-	1875	r=0	35.78***	r=0	35.39*
				r≤1	9.48	r=1	9.40
	exppc and gdppc	-	1875	r=0	35.62***	r=0	35.31*
				r≤1	9.32	r=1	9.32
Finland	exp and gdp	1916	1915	r=0	41.58***	r=0	46.06***
(1882-1937)				r≤1	7.47	r=1	11.24
	exp and gdppc	1916	1915	r=0	42.24***	r=0	47.91***
				r≤1	7.25	r=1	14.63
	exppc and gdppc	1916	1915	r=0	42.39***	r=0	46.29***
				r≤1	7.44	r=1	11.35
Italy	exp and gdp	1914	1915	r=0	30.61**	r=0	38.11**
(1862-1939)				r≤1	8.82	r=1	10.36
	exp and gdppc	1914	1915	r=0	30.08**	r=0	37.84**
				r≤1	9.13	r=1	10.23
	exppc and gdppc	1914	1915	r=0	31.34**	r=0	38.85**
				r≤1	8.97	r=1	10.04

Note: The cointegration approach allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. Break specification: Johansen (1) allows only for breaks in levels that are restricted to the error correction term. Johansen (2) allows for breaks in trend and constant jointly with trend shifts restricted to the error correction term and level shifts unrestricted in model. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. Break points were estimated by Bai and Perron's (1998) procedure considering the following linear regression for each variable: $y_t = \mu + \psi_i \sum_{i=1}^{p} y_{t-i}$. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Johansen et al. (2000). Only cointegration results with at least a 10% significance level are reported.

Stage II: Upper middle income

Table 1.15 and table 1.16 display the cointegration results of the second development stage. As this sample includes some periods of major social disturbances, it is not surprising that for some countries cointegration could only be detected including structural breaks. Still, all countries provide evidence for an existing long-run relationship between the variables.

Table 1.15: Results of Johansen cointegration test (Stage II)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	exp and gdp	r=0	15.798**	r=0	12.058	r=0	26.865**	r=0	16.531
(1886-1972)		r≤1	3.740	r=1	3.740	r≤1	10.335	r=1	10.335
	exp and gdppc	r=0	15.477**	r=0	13.241*	r=0	28.583**	r=0	18.706*
		r≤1	2.236	r=1	2.236	r≤1	9.877	r=1	9.877
	exppc and gdppc	r=0	15.858**	r=0	13.002*	r=0	28.154**	r=0	18.156*
		r≤1	2.855	r=1	2.855	r≤1	9.999	r=1	9.999
Denmark	exp and gdp	r=0	7.619	r=0	7.619	r=0	17.072	r=0	10.459
(1909-1967)		r≤1	0.000	r=1	0.000	r≤1	6.613	r=1	6.613
	exp and gdppc	r=0	9.389	r=0	9.381	r=0	16.114	r=0	10.013
		r≤1	0.009	r=1	0.009	r≤1	6.101	r=1	6.101
	exppc and gdppc	r=0	7.795	r=0	7.792	r=0	16.268	r=0	10.136
		r≤1	0.003	r=1	0.003	r≤1	6.133	r=1	6.133
Sweden	exp and gdp	r=0	11.360	r=0	10.778	r=0	21.502	r=0	12.577
(1926-1967)		r≤1	0.582	r=1	0.582	r≤1	8.924	r=1	8.924
	exp and gdppc	r=0	10.729	r=0	10.195	r=0	20.874	r=0	12.137
		r≤1	0.534	r=1	0.534	r≤1	8.736	r=1	8.736
	exppc and gdppc	r=0	10.769	r=0	10.187	r=0	20.906	r=0	12.169
		r≤1	0.582	r=1	0.582	r≤1	8.737	r=1	8.737
Finland	exp and gdp	r=0	25.108***	r=0	20.323***	r=0	30.385**	r=0	21.684**
(1938-1978)		r≤1	4.785**	r=1	4.785**	r≤1	8.701	r=1	8.701
	exp and gdppc	r=0	26.879***	r=0	24.545***	r=0	29.969**	r=0	21.473**
		r≤1	4.709**	r=1	4.709**	r≤1	8.496	r=1	8.496
	exppc and gdppc	r=0	25.613***	r=0	21.052***	r=0	30.902***	r=0	22.119**
		r≤1	4.561**	r=1	4.561**	r≤1	8.783	r=1	8.783
Italy	exp and gdp	r=0	25.522***	r=0	25.475***	r=0	44.654***	r=0	36.055***
(1940-1977)		r≤1	0.047	r=1	0.047	r≤1	8.599	r=1	8.599
	exp and gdppc	r=0	27.176***	r=0	27.170***	r=0	44.611***	r=0	36.009***
		r≤1	0.006	r=1	0.006	r≤1	8.602	r=1	8.602
	exppc and gdppc	r=0	25.682***	r=0	25.626***	r=0	44.596***	r=0	36.015***
		r≤1	0.056	r=1	0.056	r≤1	8.582	r=1	8.582

Note: Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

Table 1.16: Johansen cointegration test with structural breaks (Stage II)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	exp and gdp	1915	-	r=0	28.19*	r=0	30.25
(1886-1972)				r≤1	7.44	r=1	9.47
	exp and gdppc	1915	1918	r=0	28.74**	r=0	31.27
				r≤1	7.64	r=1	10.04
	exppc and gdppc	1915	1918	r=0	28.48*	r=0	31.23
				r≤1	7.72	r=1	10.01
Denmark	exp and gdp	1919	1920	r=0	42.28**	r=0	54.22***
(1909-1967)		1933	1932	r≤1	13.88	r=1	16.76
	exp and gdppc	1919	1920	r=0	43.69***	r=0	53.75***
		1933	1933	r≤1	13.18	r=1	17.70
	exppc and gdppc	1919	1920	r=0	43.01**	r=0	54.15***
		1933	1933	r≤1	13.37	r=1	17.36
Sweden	exp and gdp	1939	-	r=0	36.96**	r=0	44.37
(1926-1967)		1949		r≤1	14.24	r=1	13.73
	exp and gdppc	1939	-	r=0	36.57*	r=0	44.35
		1949		r≤1	14.40	r=1	13.88
	exppc and gdppc	1939	-	r=0	35.98*	r=0	45.43*
		1949		r≤1	14.30	r=1	14.08
Finland	exp and gdp	-	1951	r=0	31.21**	r=0	31.62
(1938-1978)				r≤1	10.53	r=1	14.89
	exp and gdppc	-	1951	r=0	31.48**	r=0	31.13
				r≤1	10.67	r=1	15.15
	exppc and gdppc	-	1951	r=0	31.93**	r=0	30.74
				r≤1	11.00	r=1	14.83
Italy	exp and gdp	1950	-	r=0	64.07***	r=0	66.89***
(1940-1977)				r≤1	22.82***	r=1	21.72**
	exp and gdppc	1950	-	r=0	64.33***	r=0	67.10***
				r≤1	22.97***	r=1	21.80**
	exppc and gdppc	1950	-	r=0	64.00***	r=0	66.64***
				r≤1	22.91***	r=1	21.83**

Note: The cointegration approach allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. Break specification: Johansen (1) allows only for breaks in levels that are restricted to the error correction term. Johansen (2) allows for breaks in trend and constant jointly with trend shifts restricted to the error correction term and level shifts unrestricted in model. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. Break points were estimated by Bai and Perron's (1998) procedure considering the following linear regression for each variable: $y_t = \mu + \psi_i \sum_{i=1}^{p} y_{t-i}$. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Johansen et al. (2000).

Stage III: High income

Table 1.17 and table 1.18 display the cointegration results of the last development stage. The results show that even without considering structural breaks most pairs of variables display a cointegration relationship at least at a 10 percent level. Controlling for structural breaks the outcome becomes more significant.

Table 1.17: Results of Johansen cointegration test (Stage III)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	exp and gdp	r=0	26.527***	r=0	26.489***	r=0	34.334***	r=0	29.073***
(1973-2010)		r≤1	0.037	r=1	0.037	r≤1	5.261	r=1	5.261
	exp and gdppc	r=0	29.444***	r=0	28.355***	r=0	34.432***	r=0	29.386***
		r≤1	1.089	r=1	1.089	r≤1	5.046	r=1	5.046
	exppc and gdppc	r=0	26.885***	r=0	26.788***	r=0	34.026***	r=0	28.818***
		r≤1	0.097	r=1	0.097	r≤1	5.208	r=1	5.208
Denmark	exp and gdp	r=0	18.658**	r=0	15.327**	r=0	24.437*	r=0	17.221*
(1968-2010)		r≤1	3.331	r=1	3.331	r≤1	7.216	r=1	7.216
	exp and gdppc	r=0	20.448***	r=0	17.545**	r=0	27.137**	r=0	20.088**
		r≤1	2.903	r=1	2.903	r≤1	7.049	r=1	7.049
	exppc and gdppc	r=0	17.516**	r=0	15.188**	r=0	25.247*	r=0	18.069*
		r≤1	2.328	r=1	2.328	r≤1	7.178	r=1	7.178
Sweden	exp and gdp	r=0	21.034***	r=0	20.828***	r=0	35.532***	r=0	29.034***
(1968-2010)		r≤1	0.206	r=1	0.206	r≤1	6.498	r=1	6.498
	exp and gdppc	r=0	16.709**	r=0	15.844**	r=0	32.987***	r=0	26.658***
		r≤1	0.864	r=1	0.864	r≤1	6.329	r=1	6.329
	exppc and gdppc	r=0	18.720***	r=0	18.716***	r=0	33.111***	r=0	26.487***
		r≤1	0.005	r=1	0.005	r≤1	6.625	r=1	6.625
Finland	exp and gdp	r=0	12.549	r=0	10.586	r=0	21.973	r=0	11.640
(1979-2010)		r≤1	1.963	r=1	1.963	r≤1	10.332	r=1	10.332
	exp and gdppc	r=0	12.263	r=0	11.483	r=0	24.484*	r=0	14.875
		r≤1	0.779	r=1	0.779	r≤1	9.609	r=1	9.609
	exppc and gdppc	r=0	11.361	r=0	10.726	r=0	24.023*	r=0	15.509
		r≤1	0.635	r=1	0.635	r≤1	8.514	r=1	8.514
Italy	exp and gdp	r=0	6.539	r=0	6.448	r=0	24.859*	r=0	19.215**
(1978-2010)		r≤1	0.091	r=1	0.091	r≤1	5.644	r=1	5.644
	exp and gdppc	r=0	6.441	r=0	5.280	r=0	20.303	r=0	15.084
		r≤1	1.161	r=1	1.161	r≤1	5.219	r=1	5.219
	exppc and gdppc	r=0	7.602	r=0	5.557	r=0	20.234	r=0	14.681
		r≤1	2.045	r=1	2.045	r≤1	5.552	r=1	5.552

Note: Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

Table 1.18: Johansen cointegration test with structural breaks (Stage III)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	exp and gdp	-	1993	r=0	63.46***	r=0	56.65***
(1973-2010)				r≤1	18.28***	r=1	13.56
	exp and gdppc	-	1993	r=0	58.90***	r=0	56.57***
				r≤1	16.05**	r=1	15.08
	exppc and gdppc	-	1993	r=0	58.90***	r=0	56.57***
				r≤1	16.05**	r=1	15.08
Denmark	exp and gdp	1994	-	r=0	32.28	r=0	59.26***
(1968-2010)				r≤1	8.05	r=1	12.54
	exp and gdppc	1994	-	r=0	25.75	r=0	59.79***
				r≤1	7.86	r=1	12.64
	exppc and gdppc	1994	-	r=0	25.44	r=0	59.05***
				r≤1	9.31	r=1	12.53
Finland	exp and gdp	1990	-	r=0	28.65	r=0	48.35*
(1979-2010)		2001	-	r≤1	7.46	r=1	20.80
	exp and gdppc	1990	-	r=0	30.59	r=0	49.06**
		2001		r≤1	7.38	r=1	20.39
	exppc and gdppc	1990	-	r=0	30.59	r=0	49.06**
		2001		r≤1	7.38	r=1	20.39
Italy	exp and gdp	-	2001	r=0	42.10***	r=0	63.49***
(1978-2010)				r≤1	11.87	r=1	22.50**
	exp and gdppc	-	2001	r=0	39.85***	r=0	62.04***
				r≤1	8.21	r=1	22.00**
	exppc and gdppc	-	2001	r=0	40.55***	r=0	62.12***
				r≤1	8.36	r=1	21.88**

Note: The cointegration approach allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^{*}(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_{\perp}\gamma_0$. Break specification: Johansen (1) allows only for breaks in levels that are restricted to the error correction term. Johansen (2) allows for breaks in trend and constant jointly with trend shifts restricted to the error correction term and level shifts unrestricted in model. The basic lag length in the test specifications is chosen by the Schwarz information criterion. In the presence of autocorrelation the lag length of the VARs are successively enhanced to remove all serial correlation from the data, considering a maximum of 5 lags. Break points were estimated by Bai and Perron's (1998) procedure considering the following linear regression for each variable: $y_t = \mu + \psi_i \sum_{i=1}^{p} y_{t-i}$. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Johansen et al. (2000). Only cointegration results with at least a 10% significance level are reported.

1.A.4 Bai-Perron break points and economic crisis

The Peacock-Wiseman hypothesis states that during times of major social disturbances (e.g. World War I and II) tax rates are increased by the government to generate more funds to meet the increase in defense expenditure (see Peacock and Wiseman (1961)). The phenomenon that after a crisis earlier lower tax and expenditure levels are displaced by new and higher budgetary levels is called the displacement effect. For our set of countries, this correlation is shown in figure 1.2, where periods of major social disturbances (e.g., World War I and II) seem to raise expenditures in relation to GDP to a higher level.[35]

As seen in table 1.19, the majority of the detected structural breaks by the Bai-Perron procedure coincide with these major economic crises as predicted by Peacock and Wiseman's displacement hypothesis. Most of the break dates across countries represent significant economic and social crises. Especially World Wars I and II had permanent effects on public spending in all countries. Hence, our estimation provides additional evidence on the fact that major crisis permanently influence public finances and economic growth.

[35] Henry and Olekalns (2010) analyze for the United Kingdom whether the temporal increase in the government expenditure to GDP ratio has been associated with significant structural changes to the underlying data generating process and at what time these changes occur.

Table 1.19: Bai-Perron break points and economic crisis

Country	**Estimated Breaks**	**Crisis**
United Kingdom	1875 (GDP)	Long Depression of 1873
	1915 (EXP), 1916 (EXP), 1918 (GDP)	World War I
	1936 (GDP)	World War II
	1993 (GDP)	Exchange Rate Mechanism crisis
Denmark	1892 (GDP)	-
	1900 (EXP)	-
	1919 (EXP), 1920 (GDP)	World War I/ Danish banking crisis
	1933 (EXP/GDP)	Great Depression
	1994 (EXP)	Scandinavian banking crisis
Sweden	1920 (GDP)	World War I
	1939 (EXP), 1949 (GDP)	World War II
	1966 (GDP)	-
Finland	1915 (GDP), 1916 (EXP), 1920 (GDP)	Word War I
	1949 (EXP), 1951 (GDP)	World War II
	1990 (EXP)	Finish banking crisis
	2001 (EXP)	Early 2000s recession
Italy	1914 (EXP), 1915 (GDP)	World War I
	1935 (EXP)	Great Depression
	1942 (GDP), 1950 (EXP)	World War II
	2001 (GDP)	Early 2000s recession

Note: The table displays estimated structural breaks detected by the Bai-Perron procedure. EXP and GDP denote if the structural break belongs to a public spending or an economic growth variable. Details on the estimation process can be found in section 1.A.3.

1.A.5 Granger causality results

Wagner's hypothesis was meant to explain the growth of expenditure in the long run (see, e.g., Timm (1961); Koester and Priesmeier (2013)). Therefore the main part of the analysis focuses primarily on the long-run relationship between public spending and economic growth as the methodological core of the chapter. Classic Granger tests investigate the short-run causality between the variables and do not appropriately capture the long-run relationship in line with Wagner's theory from a historical perspective. However, complementary to the main analysis it is very interesting to gain some further insights into the short-run relationship between the variables. In addition to a short run validity of Wagner's law, the results provide evidence for Keynes' hypothesis (significant causality is detected from expenditure to economic growth).

Table 1.20 shows the results of classic Granger causality tests, which have been already briefly described in the main part of this chapter. It is shown that even in the short-run, Wagner's law seems to have lost its validity in the last stage of development. Conversely, regarding the full sample, evidence in favor of Wagner's law could be found in Denmark, Sweden, Finland, and Italy. In addition, support for Keynesian policies could be found in the UK, Denmark, Sweden, and Finland over the full sample. The outcome of the causality results during the first and second development stage are rather mixed.

Table 1.20: Results of short-run causality test in VECM

		Full Sample		Stage I		Stage II		Stage III	
Country	G and Y	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$
UK	exp and gdp	0.868	21.331***	0.188	0.713	0.027	17.638***	0.691	0.155
	exp and gdppc	0.409	21.839***	0.213	0.848	0.187	18.702***	0.646	0.287
	exppc and gdppc	0.964	20.535***	0.720	0.892	0.059	17.457***	0.637	0.175
Denmark	exp and gdp	10.739***	23.846***	-	-	4.533	25.588***	0.027	6.311**
	exp and gdppc	13.775***	13.311***	-	-	0.125	22.412***	0.003	5.293**
	exppc and gdppc	12.257***	21.815***	-	-	0.524	25.785***	0.029	6.505**
Sweden	exp and gdp	14.547***	34.462***	12.549**	36.218***	2.434	0.271	5.889	3.859
	exp and gdppc	18.962***	20.814***	11.231**	26.511***	2.332	0.218	4.555	4.307
	exppc and gdppc	17.855***	38.109***	12.476**	33.837***	2.319	0.188	5.691	3.670
Finland	exp and gdp	22.471***	14.685***	28.774***	0.678	0.499	1.492	2.282	4.448**
	exp and gdppc	23.107***	14.841***	30.393***	0.648	0.432	1.338	3.623	6.327**
	exppc and gdppc	22.058***	14.872***	28.729***	0.605	0.461	1.493	2.907	6.208**
Italy	exp and gdp	28.753***	0.808	1.723	0.013	34.481***	20.635***	0.754	0.340
	exp and gdppc	26.888***	0.871	1.278	1.445	36.681***	20.323***	0.912	1.663
	exppc and gdppc	28.692***	0.871	1.685	0.005	34.756***	20.329***	1.754	1.715

Note: The table displays the block exogeneity Wald tests from the Granger procedure. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels by employing critical values from the χ^2-distribution.

1.A.6 Diagnostics and outlier correction

This section presents some general VECM diagnostics, which have been briefly discussed in the main part of this chapter (see section 1.4). Table 1.21 displays insights into the model specification and residual diagnostics in terms of autocorrelation, heteroskedasticity, and normality. Firstly, it can be noticed that the included number of observations in all estimations for the different development stages is sufficient.[36] Secondly, the goodness-of-fit measured by the adjusted R^2 is acceptable for almost every VECM.[37] Finally, it can be seen that no estimated models evince any sign of serial correlation. In those cases where heteroskedasticity could not be rejected, we employ weighted least squares to sustain consistent and asymptotically efficient estimates. Unfortunately, the test for joint normality of the residuals significantly rejects the normality assumption in the majority of VECMs. As these tests are generally very sensitive to outliers, we additionally estimate the corresponding VECMs including dummy variables to effectively remove extreme observations and improve the chances of error normality. In fact, the qualitative outcome for all countries and samples remains unchanged, which is shown in table 1.22.

[36] Only in case of Finland the number of observations implemented in the third stage of development is relatively small (29 observations), which could affect the reliability of tests.

[37] In case of Finland (stage II) and Italy (stage I) the adjusted R^2 is negative. The reason for this lies in the fact that the sample beginning in Finland and ending in Italy, respectively, coincides with extreme values caused by World War II. In both countries the estimation results remain qualitatively unchanged and the adjusted R^2 gets sufficiently large if the sample is shortened or extended.

Table 1.21: VECM diagnostics

		Full Sample						Stage I						Stage II						Stage III					
Country	Version	Obs.	Lag	Adj. R^2	Ljung-Box test	White test	DH test	Obs.	Lag	Adj. R^2	Ljung-Box test	White test	DH test	Obs.	Lag	Adj. R^2	Ljung-Box test	White test	DH test	Obs.	Lag	Adj. R^2	Ljung-Box test	White test	DH test
UK	1	161	2	0.341	1.876	62.221	178.447	36	1	0.250	3.648	41.848	15.079	87	1	0.358	4.026	30.650	136.123	38	1	0.277	1.976	27.396	16.708
	2		2	0.309	2.034	51.613	144.352		1	0.193	4.259	37.628	16.546		1	0.342	4.474	26.034	110.208		1	0.279	1.949	27.910	16.887
	3		2	0.309	1.922	55.764	157.211		1	0.348	6.077	46.248	20.012		1	0.339	4.408	28.788	121.375		1	0.295	1.983	27.733	16.564
5% CV					12.59	79.08	9.49				12.59	40.11	9.49				12.59	40.11	9.49				12.59	40.11	9.49
1% CV					16.81	83.30	13.28				16.81	46.96	13.28				16.81	46.96	13.28				16.81	46.96	13.28
DK	1	157	2	0.154	7.765	201.889	164.059	-	-	-	-	-	-	59	4	0.327	10.294	173.755[A]	11.174	43	1	0.618	7.590	28.751	36.856
	2		2	0.127	7.342	208.193	184.514		-	-	-	-	-		2	0.320	6.423	136.163	20.216		1	0.616	7.555	27.179	38.397
	3		2	0.151	7.879	205.491	165.893		-	-	-	-	-		2	0.298	8.558	134.846	21.034		1	0.610	7.476	30.217	35.646
5% CV					12.59	79.08	9.49										12.59	40.11	9.49				12.59	40.11	9.49
1% CV					16.81	83.30	13.28										16.81	46.96	13.28				16.81	46.96	13.28
SE	1	130	4	0.169	5.499	276.604	221.456	45	4	0.490	6.921	69.767	16.467	42	1	0.069	2.475	22.440	62.069	43	2	0.354	4.519	67.583	2.489
	2		4	0.166	5.147	268.950	210.201		4	0.456	7.555	70.189	13.660		1	0.069	2.484	22.642	60.057		2	0.343	4.466	65.211	2.249
	3		4	0.168	5.381	273.988	218.237		4	0.484	7.125	69.925	14.040		1	0.074	2.514	22.539	61.347		2	0.360	4.670	67.083	1.507
5% CV					12.59	234.00	9.49				12.59	67.50	9.49				12.59	40.11	9.49				12.59	79.08	9.49
1% CV					16.81	249.40	13.28				16.81	76.15	13.28				16.81	46.96	13.28				16.81	88.38	13.28
FI	1	129	3	0.182	5.967	216.312	83.316	56	1	0.450	8.017	50.926	10.446	41	1	-0.029	3.413	67.335	21.028	31	1	0.514	7.173	28.256	20.149
	2		3	0.176	7.175	215.489	70.139		1	0.401	9.187	52.173	7.793		1	-0.029	3.643	65.388	21.443		2	0.523	7.544	80.801	10.788
	3		3	0.181	7.105	213.981	72.028		1	0.438	7.579	50.868	9.119		1	-0.030	3.349	67.465	20.509		2	0.507	7.495	81.056	11.745
5% CV					12.59	124.34	9.49				12.59	40.11	9.49				12.59	40.11	9.49				12.59	79.08	9.49
1% CV					16.81	135.81	13.28				16.81	46.96	13.28				16.81	46.96	13.28				16.81	88.38	13.28
IT	1	149	1	0.166	4.752	96.639	180.990	78	1	-0.009	2.774	33.131	23.519	38	2	0.554	9.833	95.609	7.099	33	1	0.724	7.094	33.485	5.871
	2		1	0.161	4.389	99.472	106.876		2	-0.043	2.033	85.594[B]	24.083		2	0.573	10.093	94.049	7.032		1	0.734	6.356	35.294	5.207
	3		1	0.167	4.538	97.439	107.702		1	-0.012	2.556	32.551	23.411		2	0.557	9.878	95.502	7.011		1	0.739	6.173	35.212	5.049
5% CV					12.59	40.11	9.49				12.59	40.11	9.49				12.59	79.08	9.49				12.59	40.11	9.49
1% CV					16.81	46.96	13.28				16.81	46.96	13.28				16.81	88.38	13.28				16.81	46.96	13.28

Note: The adjusted R^2 belongs to the VECM with the set of expenditure variables on the left-hand side. The Ljung-Box Q-statistics are computed for every lag $p+1$ with the corresponding VECM of order p. The White LM-statistics are calculated including a cross-term. The Doorik-Hansen (DH) test reports the multivariate test statistic for joint normality.

[A] 5% critical value [234.0]; 1% critical value [249.4].

[B] 5% critical value [79.08]; 1% critical value [88.38].

Table 1.22: Outlier correction, long-run causality and short-run adjustment

		Full Sample				Stage I				Stage II				Stage III			
Country	Version	Dummy	$Y \to G$	$G \to Y$	DH test	Dummy	$Y \to G$	$G \to Y$	DH test	Dummy	$Y \to G$	$G \to Y$	DH test	Dummy	$Y \to G$	$G \to Y$	DH test
UK	1	1854,1914,15, 1919,39,40	-0.049***	-0.012	42.679	1854,1863	-0.355***	0.011	6.062	1914,21,33 1940,41	-0.083***	-0.013	23.429	1980,1990	0.071	-0.140***	9.149
	2	1854,1914,15, 1919,39,40	-0.039***	-0.004	40.291	1854,1863	-0.313***	0.029	6.923	1914,21,33 1940,41	-0.072***	-0.020	23.781	1980,1990	0.079	-0.158***	9.649
	3	1854,1914,15, 1919,39,40	-0.051***	-0.003	38.742	1854,1863	-0.428***	-0.003	5.117	1914,21,33 1940,41	-0.078***	-0.017	23.133	1980,1990	0.067	-0.137***	9.523
DK	1	1864,98,99 1919,21,83	-0.083***	0.005	36.365	-	-	-	-	1913,1919	-0.349***	-0.023	6.068	1969,77,83 2007	0.049	-0.055***	13.059
	2	1864,98,99 1919,21,83	-0.076**	-0.039**	45.392	-	-	-	-	1913,1919	-0.385***	-0.142	8.204	1969,77,83 2007	0.062	-0.068***	14.001
	3	1864,98,99 1919,21,83	-0.089***	0.003	35.747	-	-	-	-	1913,1919	-0.266***	-0.078	10.683	1969,77,83 2007	0.059	-0.063***	13.516
SE	1	1915,18,19 1931,40,46	-0.026	-0.103***	19.049	1918	-1.249***	0.224	9.819	1931,40,51	-0.076*	-0.055***	4.663	1975	-0.058	-0.190***	2.515
	2	1915,18,19 1931,40,46	-0.011	-0.092***	18.727	1918	-1.255***	0.032	8.790	1931,40,51	-0.085*	-0.049***	4.855	1975	-0.064	-0.212***	1.443
	3	1915,18,19 1931,40,46	-0.0211	-0.101***	19.206	1918	-1.254***	0.194	8.865	1931,40,51	-0.088*	-0.050***	4.431	1975	-0.059	-0.195***	2.084
FI	1	1916,17,40 1945,51,52	-0.085	-0.042**	22.199	1916,1921	-0.355***	-0.139**	2.239	1940,46,51	-0.069	-0.243***	9.798	1988,91,96 2009	-0.032***	-0.019***	3.182
	2	1916,17,40 1945,51,52	-0.053	-0.037**	22.757	1916,1921	-0.486***	-0.123**	2.774	1940,46,51	-0.066	-0.257***	10.175	1988,91,96 2009	-0.029***	-0.022***	2.813
	3	1916,17,40 1945,51,52	-0.081	-0.042***	22.437	1916,1921	-0.328***	-0.133**	2.289	1940,46,51	-0.067	-0.249***	9.763	1988,91,96 2009	-0.028***	-0.025***	2.832
IT	1	1915,16,36 1944,46,48	-0.065	-0.098***	39.014	1936	-0.092	-0.214***	11.571	1945,1946	0.311	-0.364***	4.264	2009	-0.042***	-0.026***	5.721
	2	1915,16,36 1944,46,48	-0.036	-0.095***	41.304	1936	-0.015	-0.227***	15.902	1945,1946	0.299	-0.335***	4.425	2009	-0.062***	-0.014***	4.619
	3	1915,16,36 1944,46,48	-0.063	-0.099***	39.289	1936	-0.078	-0.226***	12.055	1945,1946	0.314	-0.365***	4.243	2009	-0.052***	-0.017***	4.926

Note: The table displays estimated error correction terms of corresponding VECMS including exogenous dummy variables to remove outliers. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels. The Doorik-Hansen (DH) test reports the multivariate test statistic for joint normality (10% critical value [7.78]; 5% critical value [9.49]; 1% critical value [13.28]).

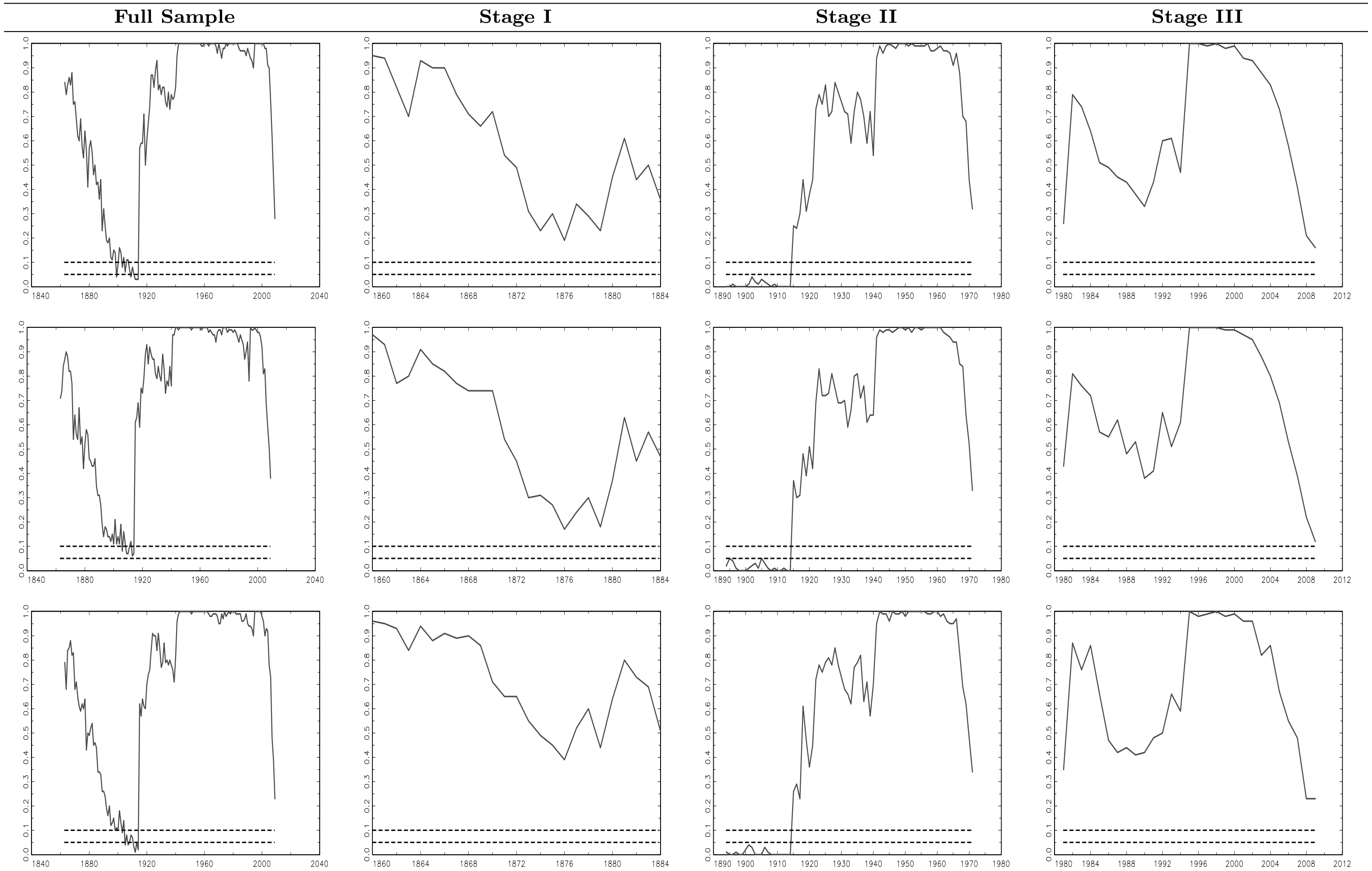
1.A.7 Stability conditions of estimates

Parameter constancy throughout the various sample periods is a key assumption in econometric models. Hence, this section controls for the stability conditions of estimates. To test the parameter stability of the various estimated VECMs shown in section 1.6, we apply the Chow forecast test. The intuition behind this testing procedure is that it splits the data into sub-periods and then estimates VECMs for each of the sub-parts. Finally, it compares the residual sum of squares (RSS) of each of the models and tests for parameter stability against the alternative that all coefficients including the residual covariance matrix may vary.[38]

Figure 1.7, figure 1.8, figure 1.9, figure 1.10 as well as figure 1.11 visualize the p-values of the Chow testing procedure for parameter stability in VECMs calculated for the different versions of Wagner's law and the corresponding development stages. The estimations show that in most cases the parameter stabilities of the corresponding VECMs are satisfactory. Nevertheless, especially in the early development stage (stage I), the test statistic allows to reject parameter stability for some break points. In those cases, the estimated error correction terms in table 1.4 should be interpreted with caution. Yet, estimations from alternative sample definitions (e.g. dropping World War data) provide qualitatively comparable results to the one shown in the main section of this article (see, for instance, table 1.22 and 1.5).

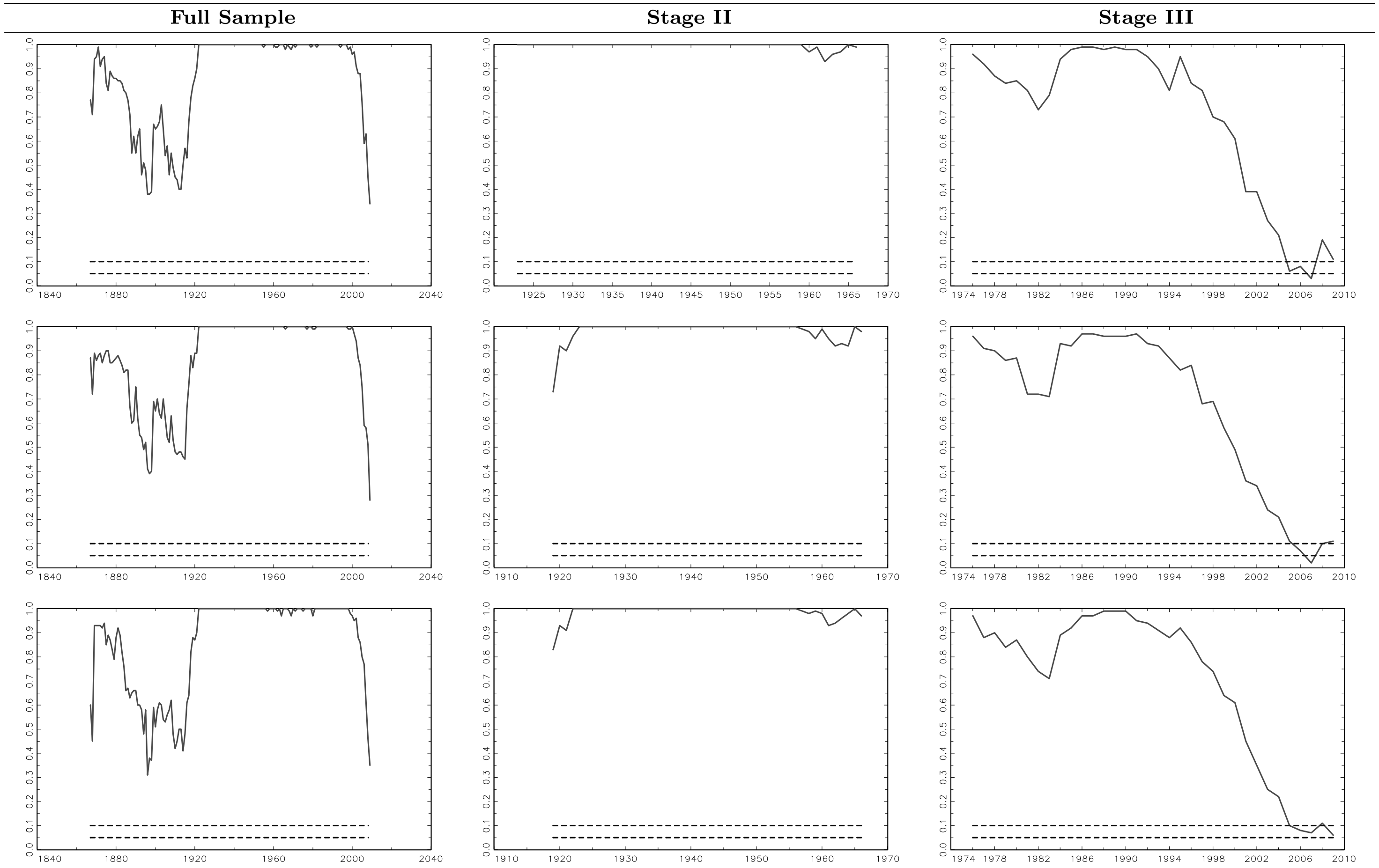
[38]Because especially small sample distributions of the test statistics under the null hypothesis may be quite different from the asymptotic Chi- or F-distributions (see Candelon and Lütkepohl (2001)), we use bootstrapped critical values calculated from 1 000 replications.

Figure 1.7: Chow test for parameter stability in VECM for the United Kingdom



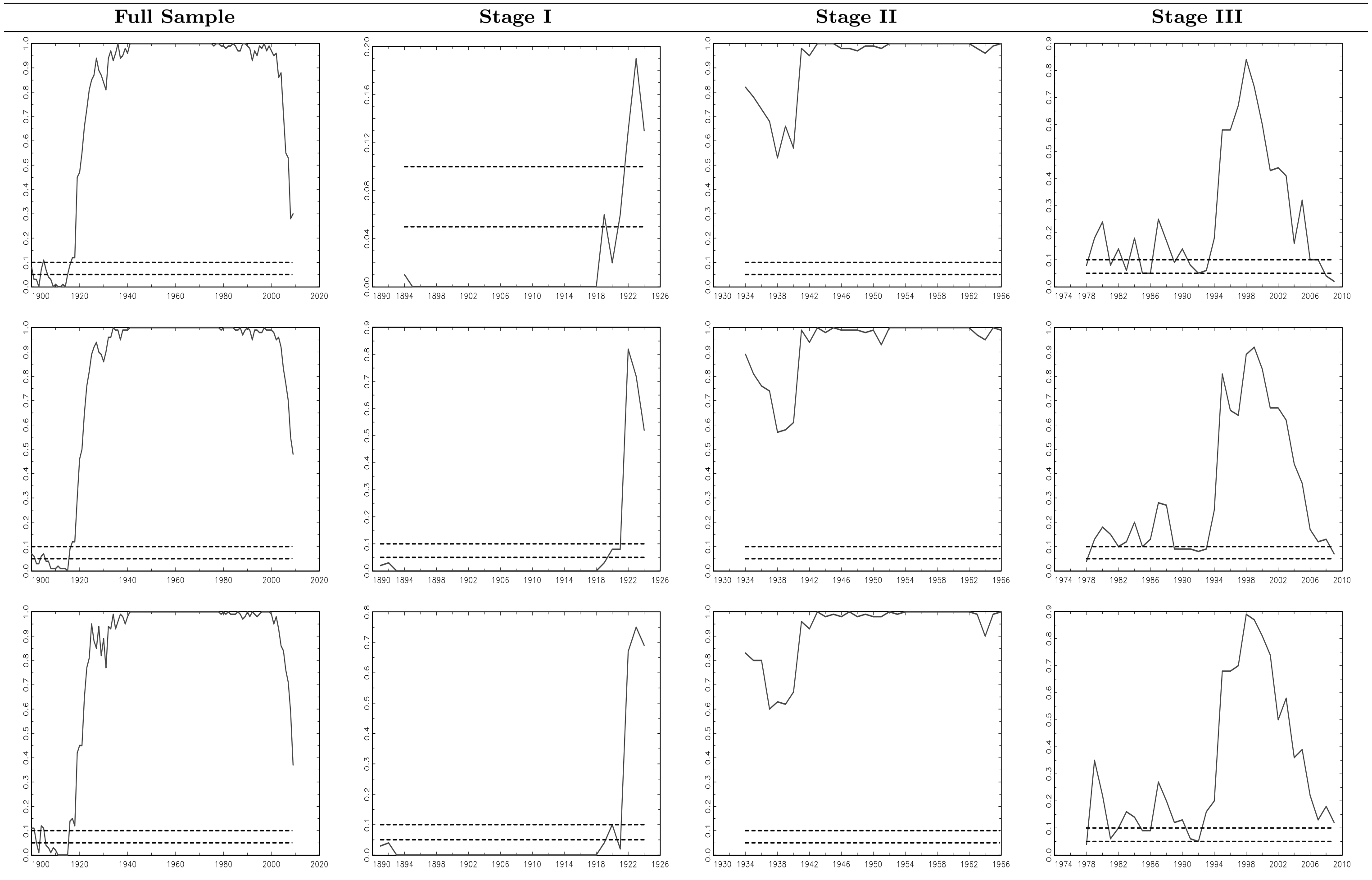
Note: The graph displays the bootstrapped p-values (1 000 replications) of the Chow test for parameter stability in VECMs calculated over a range of time points. Results for the first functional version of Wagner's law (see table 1.2) are displayed in the top row, results for the second functional version are shown in the middle row, and results for the third functional version are displayed in the bottom row.

Figure 1.8: Chow test for parameter stability in VECM for Denmark



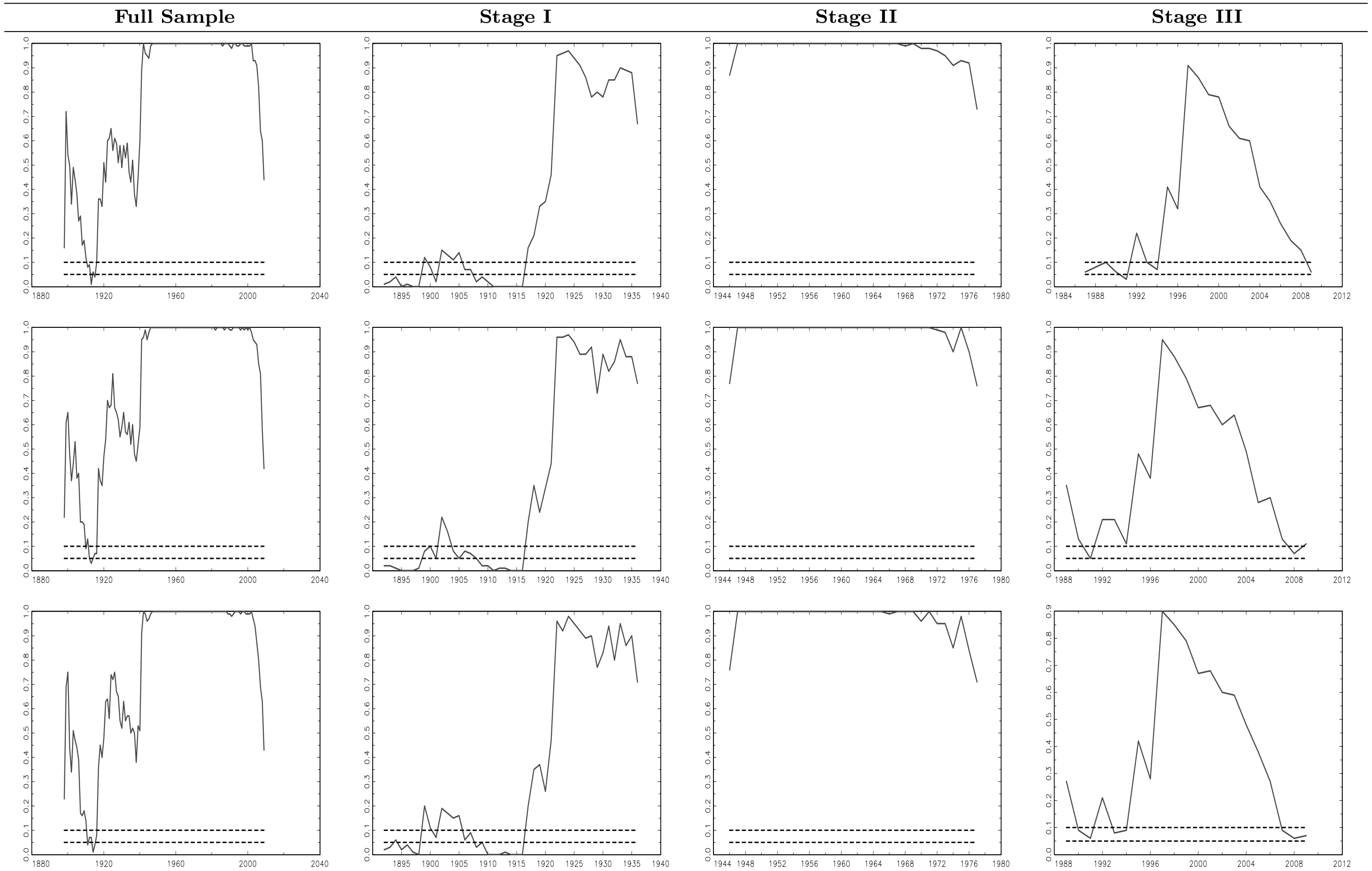
Note: The graph displays the bootstrapped p-values (1 000 replications) of the Chow test for parameter stability in VECMs calculated over a range of time points. Results for the first functional version of Wagner's law (see table 1.2) are displayed in the top row, results for the second functional version are shown in the middle row, and results for the third functional version are displayed in the bottom row.

Figure 1.9: Chow test for parameter stability in VECM for Sweden



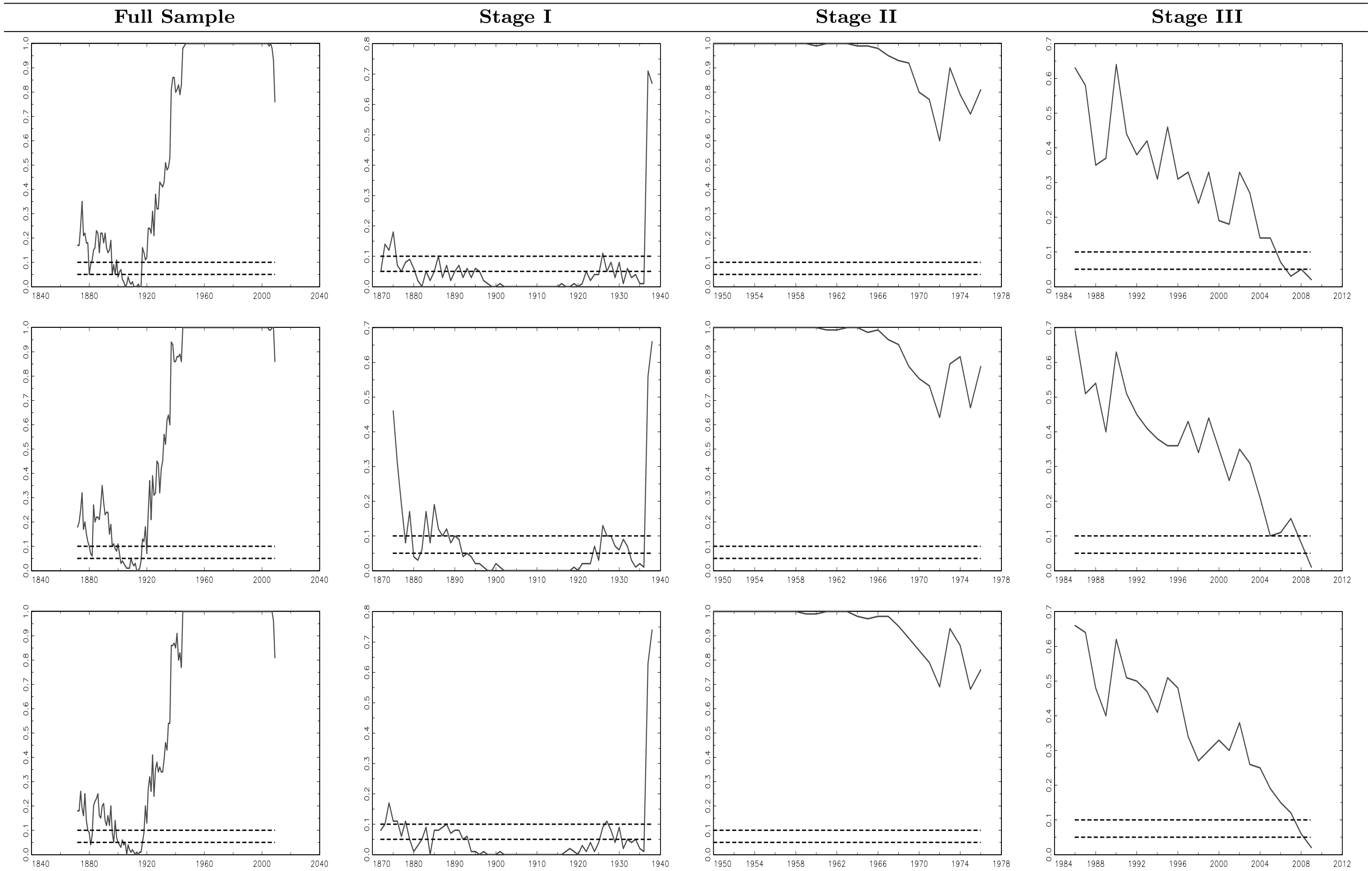
Note: The graph displays the bootstrapped p-values (1 000 replications) of the Chow test for parameter stability in VECMs calculated over a range of time points. Results for the first functional version of Wagner's law (see table 1.2) are displayed in the top row, results for the second functional version are shown in the middle row, and results for the third functional version are displayed in the bottom row.

Figure 1.10: Chow test for parameter stability in VECM for Finland



Note: The graph displays the bootstrapped p-values (1 000 replications) of the Chow test for parameter stability in VECMs calculated over a range of time points. Results for the first functional version of Wagner's law (see table 1.2) are displayed in the top row, results for the second functional version are shown in the middle row, and results for the third functional version are displayed in the bottom row.

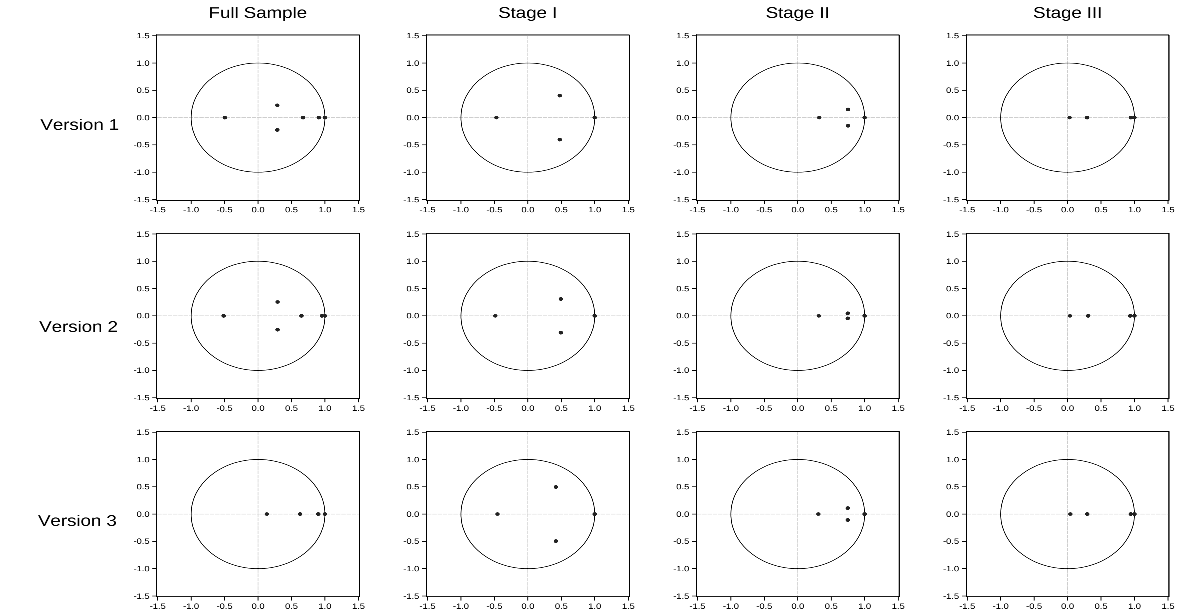
Figure 1.11: Chow test for parameter stability in VECM for Italy



Note: The graph displays the bootstrapped p-values (1 000 replications) of the Chow test for parameter stability in VECMs calculated over a range of time points. Results for the first functional version of Wagner's law (see table 1.2) are displayed in the top row, results for the second functional version are shown in the middle row, and results for the third functional version are displayed in the bottom row.

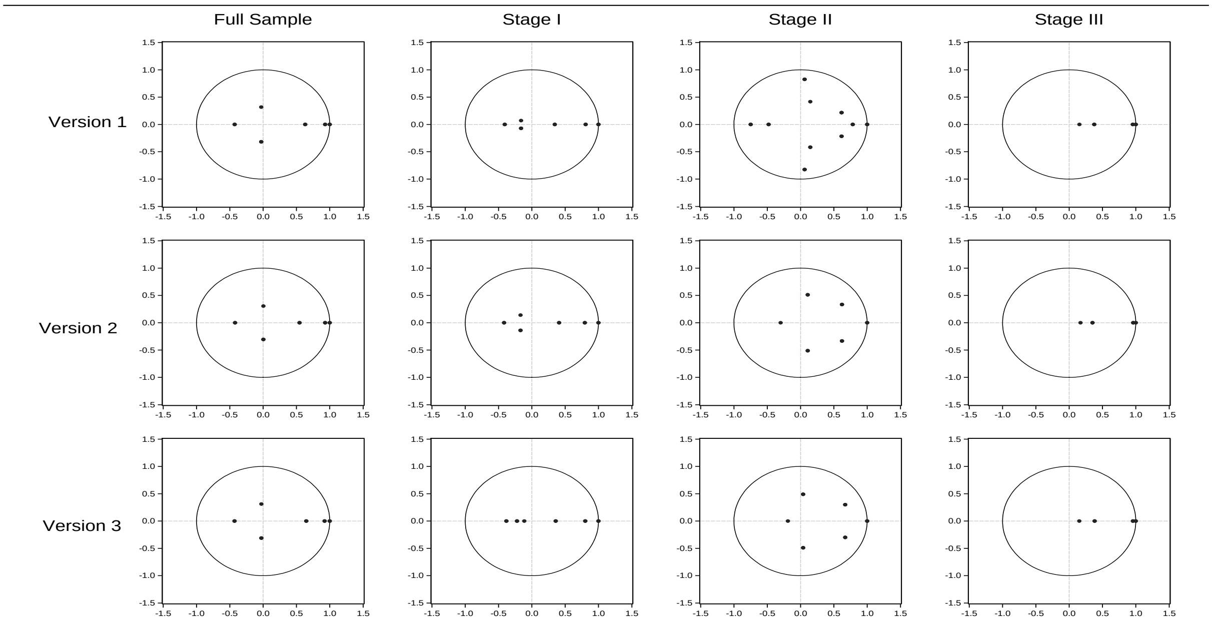
Besides parameter stability, it is also important to verify if the estimated VECMs are stable in terms of stationarity. As already seen in section 1.A.2, the battery of unit root tests generally provides evidence that the various data series can be treated as integrated of order one. Still, in some cases the unit root tests do not yield consistent results. Hence, we additionally ensure the stability of the VECMs through the test of inverse roots of the AR characteristic polynomial. The estimated VECMs are stable if exactly one root lies on the unit circle (cointegrated relation), while the remaining roots have modulus less than one and lie inside the unit circle (differenced data). Figure 1.12, figure 1.13, figure 1.14, figure 1.15 as well as figure 1.16 display the AR root graphs of the estimated VECMs for the different versions of Wagner's law as well as the different development stages. It can be seen that in every estimation exactly one root lies on the unit circle while all other roots lie inside the circle, indicating that the various estimated VECMs meet the stability conditions. Consequently, the displayed standard errors in the baseline estimations are valid.

Figure 1.12: AR roots graph for the United Kingdom



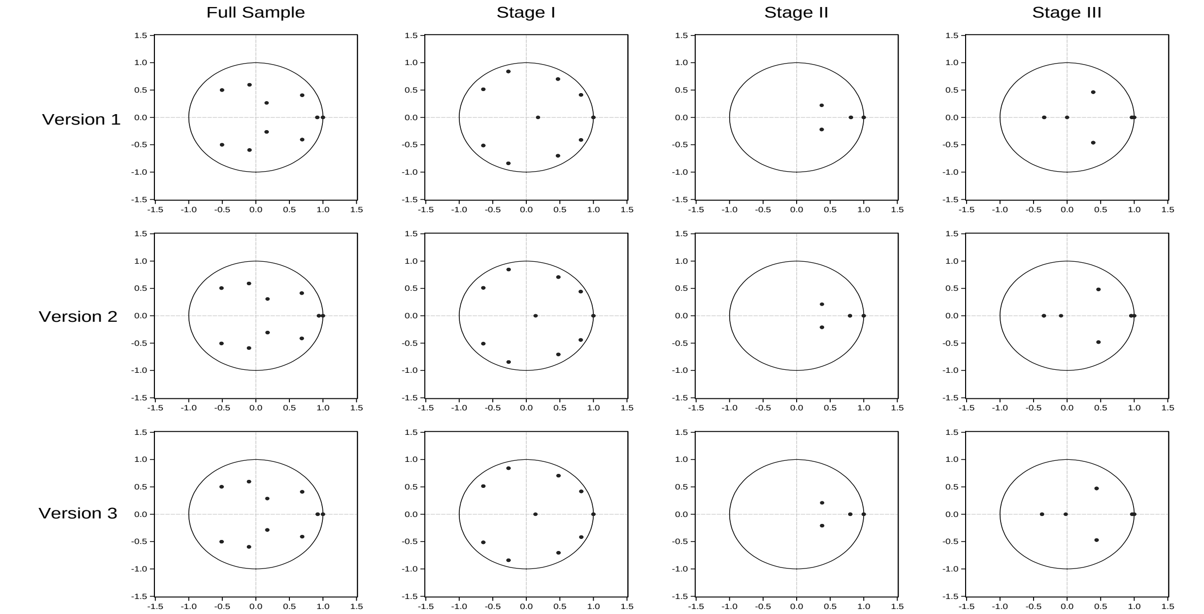
Note: The figure displays the inverse roots of AR characteristic polynomial of VECMs for different versions of Wagner's law (row) and different development stages (column). The estimated VECMs are stable if exactly one root lies on the unit circle, while the remaining roots have modulus less than one and lie inside the unit circle.

Figure 1.13: AR roots graph for Denmark



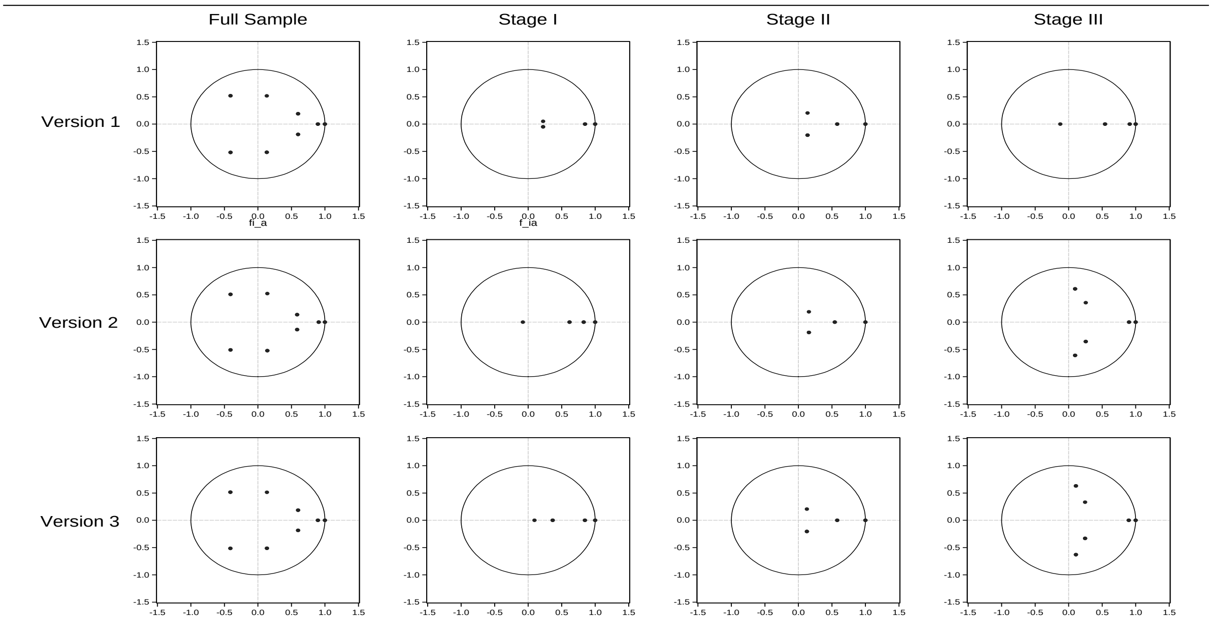
Note: The figure displays the inverse roots of AR characteristic polynomial of VECMs for different versions of Wagner's law (row) and different development stages (column). The estimated VECMs are stable if exactly one root lies on the unit circle, while the remaining roots have modulus less than one and lie inside the unit circle.

Figure 1.14: AR roots graph for Sweden



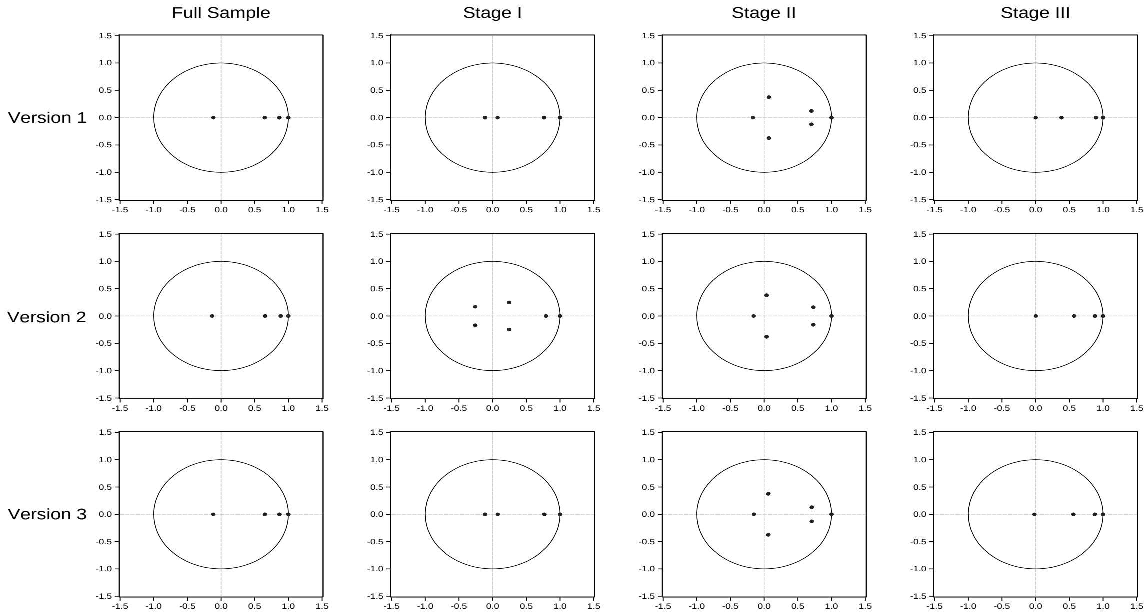
Note: The figure displays the inverse roots of AR characteristic polynomial of VECMs for different versions of Wagner's law (row) and different development stages (column). The estimated VECMs are stable if exactly one root lies on the unit circle, while the remaining roots have modulus less than one and lie inside the unit circle.

Figure 1.15: AR roots graph for Finland



Note: The figure displays the inverse roots of AR characteristic polynomial of VECMs for different versions of Wagner's law (row) and different development stages (column). The estimated VECMs are stable if exactly one root lies on the unit circle, while the remaining roots have modulus less than one and lie inside the unit circle.

Figure 1.16: AR roots graph for Italy



Note: The figure displays the inverse roots of AR characteristic polynomial of VECMs for different versions of Wagner's law (row) and different development stages (column). The estimated VECMs are stable if exactly one root lies on the unit circle, while the remaining roots have modulus less than one and lie inside the unit circle.

1.A.8 Alternative model specifications

As causality tests are known to be very sensitive to methodological issues, this section provides some additional VECM estimations with different lag specifications and deterministic components. In the benchmark estimations in section 1.6, the lag structure is chosen by a specific-to-general method. Table 1.23 presents error correction terms of additional VECMs where the lag length is chosen by standard Schwarz information criterion (SIC), Akaike information criterion (AIC), and the final prediction error (FPE). It can be noticed that the results are robust to alternative lag specifications.

Further, the benchmark VECMs in the main analysis (see table 1.4) are estimated allowing for a constant in the cointegration equation, a constant in the VAR and a linear trend in the level data. This standard model was chosen for every stage and country in order to compare the results throughout countries and development stages. Table 1.24 displays VECMs with alternative specifications of the deterministic components. Adding a trend to the cointegration equation increases slightly the error correction mechanism running from economic growth to government activity. A model including a constant and a trend in the cointegrating equation as well as in the VAR provides similar results. However, qualitatively the adjustment speed of public expenditure towards long-run equilibrium induced by shocks in GDP remains unchanged. This correlation is visualized in figure 1.17, which looks very similar to figure 1.3 in section 1.6 of this chapter.

Table 1.23: Long-run causality and short-run adjustment with different lag specification

			Full Sample			Stage I			Stage II			Stage III		
Country	G and Y	Lag	$Y \rightarrow G$	$G \rightarrow Y$		$Y \rightarrow G$	$G \rightarrow Y$		$Y \rightarrow G$	$G \rightarrow Y$		$Y \rightarrow G$	$G \rightarrow Y$	
UK	exp and gdp	SIC	-0.038**	-0.028***	(1)	-0.437***	0.022	(1)	-0.089**	-0.015	(1)	0.055	-0.101***	(1)
		AIC	-0.042*	-0.016*	(2)	-0.509***	0.005	(3)	-0.089**	-0.015	(1)	0.055	-0.101***	(1)
		FPE	-0.042*	-0.016*	(2)	-0.509***	0.005	(3)	-0.089**	0.015	(1)	0.055	-0.101***	(1)
		COR	-0.042*	-0.016*	(2)	-0.437***	0.022	(1)	-0.089**	-0.015	(1)	0.055	-0.101***	(1)
	exp and gdppc	SIC	-0.015	-0.017**	(2)	-0.437***	0.022	(1)	-0.057**	-0.021**	(1)	0.064	-0.115***	(1)
		AIC	-0.015	-0.017**	(2)	-0.249**	-0.049***	(2)	-0.057**	-0.021**	(1)	-0.052***	0.002	(2)
		FPE	-0.015	-0.017**	(2)	-0.249***	-0.049**	(2)	-0.057**	-0.021**	(1)	-0.052***	0.002	(2)
		COR	-0.015	-0.017**	(2)	-0.437***	0.022	(1)	-0.057**	-0.021**	(1)	0.064	-0.115***	(1)
	exppc and gdppc	SIC	-0.035*	-0.017**	(2)	-0.535***	-0.053**	(1)	-0.071**	-0.019**	(1)	0.051	-0.098***	(1)
		AIC	-0.035*	-0.017**	(2)	-1.074***	-0.157**	(3)	-0.071**	-0.019**	(1)	0.051	-0.098***	(1)
		FPE	-0.035*	-0.017**	(2)	-1.074***	-0.157**	(3)	-0.071**	-0.019**	(1)	0.051	-0.098***	(1)
		COR	-0.035*	-0.017**	(2)	-0.535***	-0.053**	(1)	-0.071**	-0.019**	(1)	0.051	-0.098***	(1)
DK	exp and gdp	SIC	-0.077***	0.005	(1)	-	-		-0.340***	-0.098	(2)	-0.021***	-0.001	(1)
		AIC	-0.081***	-0.007	(4)	-	-		-0.340***	-0.098	(2)	-0.021***	-0.001	(1)
		FPE	-0.081***	-0.007	(4)	-	-		-0.340***	-0.098	(2)	-0.021***	-0.001	(1)
		COR	-0.078***	-0.002	(2)	-	-		-0.358***	-0.059	(4)	-0.021***	-0.001	(1)
	exp and gdppc	SIC	-0.025	-0.065***	(1)	-	-		-0.487***	-0.197**	(2)	-0.008***	-0.003	(1)
		AIC	-0.048	-0.066***	(2)	-	-		-0.487***	-0.197**	(2)	-0.008***	-0.003	(1)
		FPE	-0.048	-0.066***	(2)	-	-		-0.487***	-0.197**	(2)	-0.008***	-0.003	(1)
		COR	-0.048	-0.066***	(2)	-	-		-0.487***	-0.197**	(2)	-0.008***	-0.003	(1)
	exppc and gdppc	SIC	-0.083***	-0.001	(1)	-	-		-0.381***	-0.115*	(2)	-0.017***	-0.001	(1)
		AIC	-0.085**	-0.016	(4)	-	-		-0.381***	-0.115*	(2)	-0.017***	-0.001	(1)
		FPE	-0.085**	-0.016	(4)	-	-		-0.381***	-0.115*	(2)	-0.017***	-0.001	(1)
		COR	-0.078**	-0.015	(2)	-	-		-0.381***	-0.115*	(2)	-0.017***	-0.001	(1)
SE	exp and gdp	SIC	-0.041	-0.044**	(2)	-1.551***	0.398	(4)	-0.111*	-0.085***	(1)	0.056	-0.097***	(1)
		AIC	-0.034	-0.083***	(4)	-1.035***	0.457	(5)	-0.111*	-0.085***	(1)	-0.028	-0.072***	(3)
		FPE	-0.034	-0.083***	(4)	-1.551***	0.398	(4)	-0.111*	-0.085***	(1)	-0.028	-0.072***	(3)
		COR	-0.034	-0.083***	(4)	-1.551***	0.398	(4)	-0.111*	-0.085***	(1)	-0.052	-0.188***	(2)
	exp and gdppc	SIC	-0.027	-0.082***	(1)	-1.595***	0.164	(4)	-0.107*	-0.092***	(1)	-0.058	-0.209***	(2)
		AIC	-0.018	-0.042***	(2)	-1.475***	0.076	(5)	-0.107*	-0.092***	(1)	-0.084	-0.181***	(3)
		FPE	-0.018	-0.042***	(2)	-1.475***	0.076	(5)	-0.107*	-0.092***	(1)	-0.084	-0.181***	(3)
		COR	-0.006	-0.074***	(4)	-1.595***	0.164	(4)	-0.107*	-0.092***	(1)	-0.058	-0.209***	(2)
	exppc and gdppc	SIC	-0.036	-0.043**	(2)	-1.592***	0.364	(4)	-0.114*	-0.081***	(1)	-0.054	-0.191***	(2)
		AIC	-0.027	-0.081***	(4)	-1.291***	0.465	(5)	-0.114*	-0.081***	(1)	0.061	-0.107***	(3)
		FPE	-0.027	-0.081***	(4)	-1.291***	0.465	(5)	-0.114*	-0.081***	(1)	0.061	-0.107***	(3)
		COR	-0.027	-0.081***	(4)	-1.592***	0.364	(4)	-0.114*	-0.081***	(1)	-0.054	-0.191***	(2)
FI	exp and gdp	SIC	-0.077	-0.102***	(1)	-0.349***	-0.153**	(1)	-0.194*	-0.228***	(1)	-0.089***	0.012	(1)
		AIC	-0.068	-0.090***	(3)	-0.248**	-0.071	(2)	-0.191*	-0.267***	(3)	-0.080	0.055	(5)
		FPE	-0.068	-0.090***	(3)	-0.248**	-0.071	(2)	-0.191*	-0.267***	(3)	-0.080	0.055	(5)
		COR	-0.068	-0.090***	(3)	-0.349***	-0.153**	(1)	-0.194*	-0.228***	(1)	-0.089***	0.012	(1)
	exp and gdppc	SIC	-0.041	-0.099***	(1)	-0.153*	-0.161***	(1)	-0.195*	-0.237***	(1)	-0.141***	0.044	(2)
		AIC	-0.024	-0.088***	(3)	-0.112	-0.091*	(2)	-0.185	-0.277***	(3)	-0.082	0.062	(5)
		FPE	-0.024	-0.088***	(3)	-0.112	-0.091*	(2)	-0.185	-0.277***	(3)	-0.082	0.062	(5)
		COR	-0.024	-0.088***	(3)	-0.153*	-0.161***	(1)	-0.195*	-0.237***	(1)	-0.141***	0.044	(2)
	exppc and gdppc	SIC	-0.074	-0.086***	(1)	-0.313***	-0.156**	(1)	-0.195*	-0.233***	(1)	-0.138***	0.039	(2)
		AIC	-0.066	-0.054***	(3)	-0.248**	-0.071	(2)	-0.207*	-0.265***	(3)	-0.065	0.055	(5)
		FPE	-0.066	-0.054***	(3)	-0.248**	-0.071	(2)	-0.207*	-0.265***	(3)	-0.065	0.055	(5)
		COR	-0.066	-0.054***	(3)	-0.313***	-0.156**	(1)	-0.195*	-0.233***	(1)	-0.138***	0.039	(2)
IT	exp and gdp	SIC	-0.065	-0.096***	(1)	-0.096	-0.220***	(1)	0.289	-0.442***	(2)	-0.057***	-0.019***	(1)
		AIC	-0.065	-0.096***	(1)	-0.096	-0.220***	(1)	0.388	-0.473***	(4)	-0.126***	0.002	(5)
		FPE	-0.065	-0.096***	(1)	-0.096	-0.220***	(1)	0.388	-0.473***	(4)	-0.126***	0.002	(5)
		COR	-0.065	-0.096***	(1)	-0.096	-0.220***	(1)	0.289	-0.442***	(2)	-0.057***	-0.019***	(1)
	exp and gdppc	SIC	-0.002	-0.094***	(1)	0.000	-0.232***	(1)	0.282	-0.412***	(2)	-0.073***	-0.009***	(1)
		AIC	-0.002	-0.094***	(1)	0.000	-0.232***	(1)	0.378	-0.446***	(4)	-0.128***	0.005	(5)
		FPE	-0.002	-0.094***	(1)	0.000	-0.232***	(1)	0.378	-0.446***	(4)	-0.061***	-0.024***	(3)
		COR	-0.002	-0.094***	(1)	-0.011	-0.256***	(2)	0.282	-0.412***	(2)	-0.073***	-0.009***	(1)
	exppc and gdppc	SIC	-0.039	-0.097***	(1)	-0.077	-0.232***	(1)	0.292	-0.442***	(2)	-0.066***	-0.010***	(1)
		AIC	-0.039	-0.097***	(1)	-0.077	-0.232***	(1)	0.389	-0.469***	(4)	-0.125***	0.005	(5)
		FPE	-0.039	-0.097***	(1)	-0.077	-0.232***	(1)	0.389	-0.469***	(4)	-0.056**	-0.026***	(3)
		COR	-0.039	-0.097***	(1)	-0.077	-0.232***	(1)	0.292	-0.442***	(2)	-0.066***	-0.010***	(1)

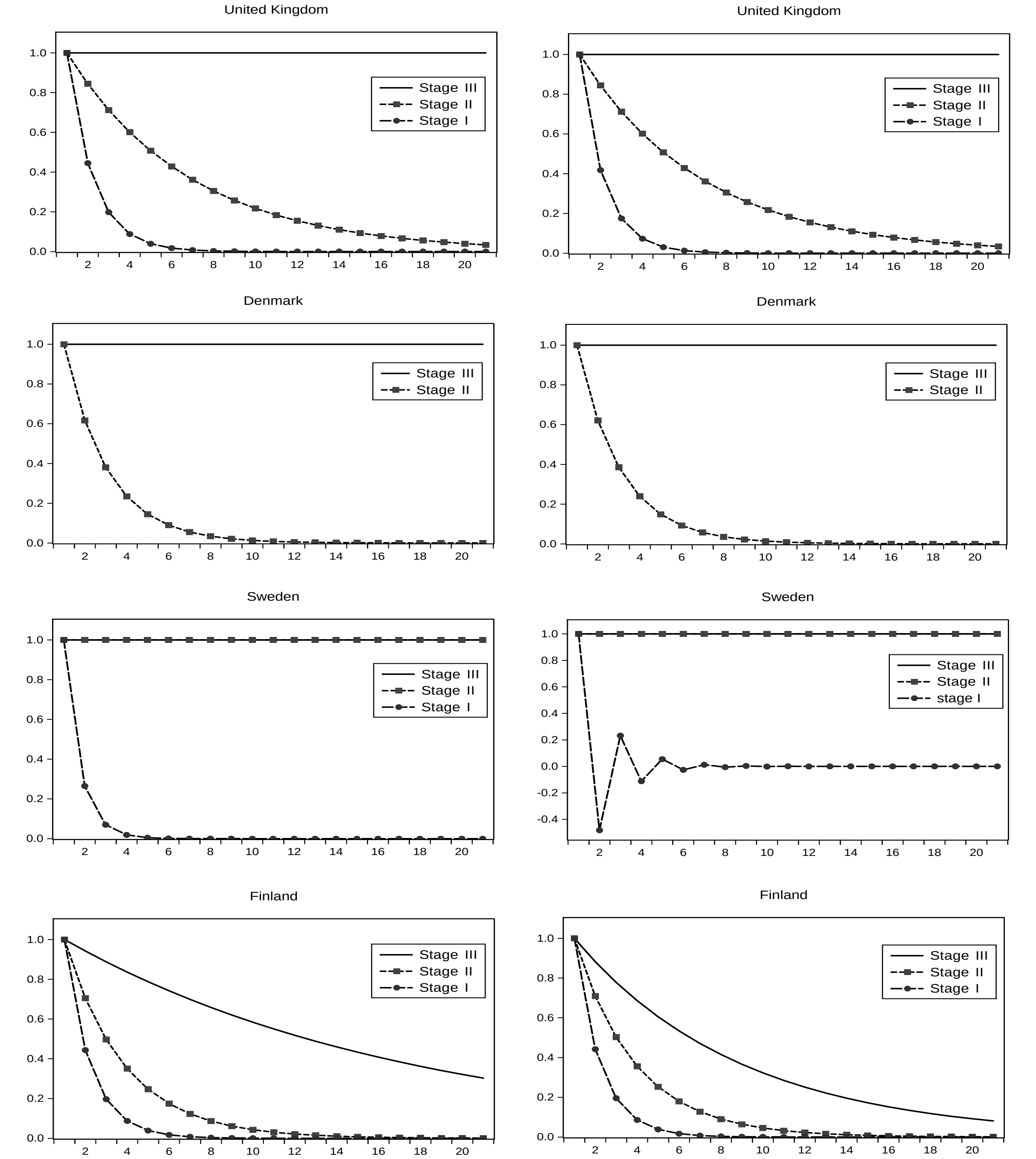
Note: The table displays estimated error correction terms (ect) of corresponding VECMs. The information criteria are computed for reduced form VECMs and searched up a maximum of five lags of first differences in order to determine the optimal number of lags (presented in parenthesis) in each model. The lag length is chosen by Schwarz information criterion (SIC), Akaike information criterion (AIC) as well as the Final prediction error (FPE). COR describes the specific-to-general method used in the benchmark estimations, where the lag length of the VECM was determined by SIC and then successively enhanced until the Ljung-Box test statistics are insignificant at all lags. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Table 1.24: Long-run causality and short-run adjustment with different deterministic components

			Full Sample		Stage I	Stage II	Stage III			
Country	G and Y	Model	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$	$Y \to G$	$G \to Y$
UK	exp and gdp	C	-0.042*	-0.016*	-0.437***	0.022	-0.089**	-0.015	0.055	-0.101***
		C/T	-0.120***	0.002	-0.555***	-0.102**	-0.156***	-0.000	0.007	-0.084***
		C/T/T	-0.118***	-0.001	-0.582***	-0.078*	-0.156***	-0.000	0.031	-0.070***
	exp and gdppc	C	-0.015	-0.017**	-0.339***	0.002	-0.057**	-0.021**	0.064	-0.115***
		C/T	-0.128***	-0.000	-0.545***	-0.106**	-0.152***	0.000	0.013	-0.086***
		C/T/T	-0.126***	-0.002	-0.578***	-0.080*	-0.155***	0.000	0.043	-0.084***
	exppc and gdppc	C	-0.035*	-0.018*	-0.535***	-0.053**	-0.071**	-0.019**	0.051	-0.098***
		C/T	-0.120***	-0.000	-0.544***	-0.094**	-0.155***	0.000	0.008	-0.081***
		C/T/T	-0.118***	-0.002	-0.576***	-0.071*	-0.156***	0.000	0.029	-0.069***
DK	exp and gdp	C	-0.078***	-0.002	-	-	-0.358***	-0.059	-0.021***	-0.001
		C/T	0.021	-0.143***	-	-	-0.383***	-0.246**	0.114	-0.084
		C/T/T	-0.021	-0.139***	-	-	-0.379***	-0.250**	0.107	-0.053
	exp and gdp	C	-0.048	-0.066***	-	-	-0.487***	-0.197**	-0.008***	-0.003
		C/T	0.024	-0.112***	-	-	-0.484***	-0.228**	0.105	-0.071
		C/T/T	-0.017	-0.109***	-	-	-0.482***	-0.231**	0.099	-0.043
	exp and gdp	C	-0.078**	-0.015	-	-	-0.381***	-0.115*	-0.017***	-0.001
		C/T	0.028	-0.130***	-	-	-0.421***	-0.242*	0.114	-0.080
		C/T/T	-0.021	-0.126***	-	-	-0.417***	-0.246**	0.111	-0.068
SE	exp and gdp	C	-0.034	-0.083***	-1.554***	0.338	-0.111*	-0.085***	-0.052	-0.188***
		C/T	-0.102*	-0.073***	-0.735***	-0.263	0.003	-0.116***	-0.076	-0.239***
		C/T/T	-0.135*	-0.063***	-1.482***	0.106	-0.016	-0.099***	-0.119	-0.170***
	exp and gdppc	C	-0.006	-0.073***	-1.595***	0.164	-0.107*	-0.092***	-0.058	-0.209***
		C/T	-0.088*	-0.069***	-0.888***	-0.232	0.001	-0.119***	-0.072	-0.243***
		C/T/T	-0.122**	-0.059***	-1.506***	0.097	-0.029*	-0.095***	-0.101	-0.175***
	exppc and gdppc	C	-0.027	-0.081***	-1.592***	0.364	-0.114*	-0.081***	-0.054	-0.191***
		C/T	-0.100*	-0.073***	-0.856***	-0.227	0.000	-0.119***	-0.081	-0.246***
		C/T/T	-0.137**	-0.062***	-1.499***	0.110	-0.033*	-0.094***	-0.107	-0.178***
FI	exp and gdp	C	-0.068	-0.090***	-0.349***	-0.153**	-0.194*	-0.228***	-0.089***	0.012
		C/T	-0.088*	-0.083***	-0.556***	-0.027	-0.295**	-0.176***	-0.058***	0.047
		C/T/T	-0.083	-0.085***	-0.558***	-0.018	-0.291**	-0.178***	-0.118***	0.007
	exp and gdppc	C	-0.024	-0.089***	-0.153*	-0.161***	-0.195*	-0.237***	-0.141***	0.044
		C/T	-0.018	-0.069***	-0.554***	-0.014	-0.307**	-0.180***	-0.146***	0.051
		C/T/T	-0.009	-0.089***	-0.556***	-0.003	-0.296**	-0.185***	-0.185***	0.053
	exppc and gdppc	C	-0.066	-0.091***	-0.313***	-0.156**	-0.195*	-0.233***	-0.138***	0.039
		C/T	-0.106**	-0.074***	-0.560***	-0.023	-0.284**	-0.185***	-0.149***	0.053
		C/T/T	-0.098	-0.077***	-0.562***	-0.014	-0.281**	-0.186***	-0.189***	0.055
IT	exp and gdp	C	-0.042	-0.096***	-0.095	-0.220***	0.289	-0.442***	-0.057***	-0.019***
		C/T	-0.084*	-0.092***	-0.094	-0.221***	-0.008***	-0.109***	-0.018***	-0.056***
		C/T/T	-0.091*	-0.089***	-0.115	-0.216***	-0.028***	-0.092***	-0.014***	-0.052***
	exp and gdppc	C	-0.002	-0.094***	-0.011	-0.232***	0.282	-0.412***	-0.073***	-0.009***
		C/T	-0.068	-0.099***	-0.089	-0.254***	-0.003***	-0.112***	-0.040***	-0.035***
		C/T/T	-0.075	-0.098***	-0.107	-0.249***	-0.023***	-0.096***	-0.035***	-0.035***
	exppc and gdppc	C	-0.039	-0.097***	-0.077	-0.232***	0.292	-0.442***	-0.066***	-0.010***
		C/T	-0.082*	-0.093***	-0.094	-0.226***	-0.002***	-0.113***	-0.051***	-0.025***
		C/T/T	-0.089*	-0.091***	-0.112	-0.222***	-0.022***	-0.096***	-0.049***	-0.027***

Note: The table displays estimated error correction terms (ect) of corresponding VECMs. Model C includes a constant in the cointegrating equation and in the VAR. Model C/T has a constant and trend in the cointegrating equation, while the VAR only includes a constant. Model C/T/T includes a constant and trend in the cointegrating equation as well as in the VAR. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

Figure 1.17: Adjustment speed of government expenditure towards long-run equilibrium for different model specifications



Note: The graphs display the expenditure convergence to shocks in GDP for the United Kingdom, Denmark, Sweden, and Finland during different stages of economic development. The left column displays the expenditure convergence estimated from the model C/T. The right column displays the expenditure convergence estimated from the model C/T/T. For Denmark no long-run equilibrium between government expenditure and GDP was detected during the first development stage. The expenditure convergences are calculated by $\sum_{n=0}^{20}(1 - ect_{ij})^n$, where $n + 1$ denotes the number of periods, i the country, j the development stage, and ect the corresponding error correction term from table 1.24.

1.A.9 Results for additional functional forms of Wagner's law

As already pointed out in the main part of this chapter, the literature generally deals with some different functional forms of testing Wagner's hypothesis (see, e.g., Magazzino (2012b); Abizadeh and Yousefi (1988)). In this context, the choices of dependent and independent variables supposed to formally represent Wagner's law vary across specifications. Besides the widespread functional forms used in the main analysis (see table 1.2), table 1.25 presents some additional specification on how to quantify the relationship between the public sector and economic growth.[39]

Table 1.25: Additional functional forms of testing Wagner's hypothesis

Version	**Functional form**	**Source**
4	$ln(exp/gdp) = \alpha + \beta * ln(gdppc) + z_t$	Musgrave (1969)
5	$ln(exp/gdp) = \alpha + \beta * ln(gdp) + z_t$	Mann (1980)

Note: *exp* denotes central government expenditure, *gdp* corresponds to gross domestic product, *gdppc* signifies gross domestic product per capita and *exppc* defines central government expenditure per capita.

On the one hand, Musgrave (1969) quantifies in his version of the law that the public sector share to GDP is increasing as the GDP per capita raises. On the other hand, Mann (1980) modifies the original formulation by Peacock and Wiseman (1961) and models the log of public expenditure share to GDP in terms of the log of output. Table 1.26 to table 1.33 display the cointegration results of these additional versions of Wagner's law at different stages of development. Overall, the test results confirm the findings of the

[39] The literature deals with some additional functional forms of Wagner's law (e.g. Pryor (1968); Murthy (1994)), which are ignored in our analysis. Unfortunately, we were not able to test for these alternative specifications because, as to our knowledge, there is no comparable historical data on government final consumption expenditure as well as government deficits available.

main analysis that public spending in the United Kingdom, Sweden, Finland, and Italy is cointegrated with economic growth independently of development stage or functional form. However, in Denmark, a cointegration relationship for all three versions of Wagner's law was only found in the second and third development stage but not in the first.

Full sample

Table 1.26: Johansen cointegration test for additional functional forms (Full Sample)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	expgdp and gdppc	r=0	11.150	r=0	8.931	r=0	23.648*	r=0	15.237
(1850-2010)		r≤1	2.219	r=1	0.014	r≤1	8.410	r=1	8.410
	expgdp and gdp	r=0	10.496	r=0	8.029	r=0	22.265*	r=0	16.516*
		r≤1	2.467	r=1	2.467	r≤1	5.749	r=1	5.749
Denmark	expgdp and gdppc	r=0	12.626	r=0	9.663	r=0	25.974**	r=0	18.684*
(1854-2010)		r≤1	2.964	r=1	2.964	r≤1	7.290	r=1	7.290
	expgdp and gdp	r=0	10.919	r=0	8.307	r=0	25.085*	r=0	17.733*
		r≤1	2.613	r=1	2.613	r≤1	7.352	r=1	7.352
Sweden	expgdp and gdppc	r=0	14.357*	r=0	13.326*	r=0	26.322**	r=0	19.952**
(1881-2010)		r≤1	1.031	r=1	1.031	r≤1	6.369	r=1	6.369
	expgdp and gdp	r=0	14.097*	r=0	13.190*	r=0	25.517*	r=0	19.012**
		r≤1	0.907	r=1	0.907	r≤1	6.505	r=1	6.505
Finland	expgdp and gdppc	r=0	10.751	r=0	9.200	r=0	16.657	r=0	9.228
(1882-2010)		r≤1	1.550	r=1	1.550	r≤1	7.429	r=1	7.429
	expgdp and gdp	r=0	10.937	r=0	9.326	r=0	16.334	r=0	9.363
		r≤1	1.610	r=1	1.610	r≤1	6.971	r=1	6.971
Italy	expgdp and gdppc	r=0	19.214**	r=0	19.121***	r=0	28.356**	r=0	21.524**
(1862-2010)		r≤1	0.093	r=1	0.093	r≤1	6.832	r=1	6.832
	expgdp and gdp	r=0	19.415**	r=0	19.343***	r=0	28.164**	r=0	21.412**
		r≤1	0.071	r=1	0.071	r≤1	6.752	r=1	6.752

Note: See table 1.11.

Table 1.27: Johansen cointegration test with structural breaks for additional functional forms (Full Sample)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	expgdp and gdppc	1913	1993	r=0	36.18**	r=0	44.41**
(1850-2010)				r≤1	14.05	r=1	13.50
	expgdp and gdp	1913	1993	r=0	33.31*	r=0	45.48**
				r≤1	11.62	r=1	13.78
Denmark	expgdp and gdppc	1936	1920	r=0	36.18**	r=0	40.25
(1854-2010)				r≤1	10.95	r=1	12.71
	expgdp and gdp	1936	1920	r=0	33.77*	r=0	41.72*
				r≤1	11.11	r=1	13.66
Sweden	expgdp and gdppc	1920	1920	r=0	53.97***	r=0	60.60***
(1881-2010)			1966	r≤1	12.92	r=1	14.50
	expgdp and gdp	1920	1920	r=0	54.20***	r=0	61.63***
			1966	r≤1	12.63	r=1	14.38
Finland	expgdp and gdppc	1938	1920	r=0	33.24	r=0	49.69**
(1882-2010)				r≤1	9.30	r=1	12.87
	expgdp and gdp	1938	1920	r=0	34.17*	r=0	50.09**
				r≤1	8.91	r=1	13.05
Italy	expgdp and gdppc	-	1942	r=0	45.70***	r=0	66.95***
(1862-2010)				r≤1	7.97	r=1	8.35
	expgdp and gdp	-	1942	r=0	46.78***	r=0	66.85***
				r≤1	7.95	r=1	8.17

Note: See table 1.12.

Stage I: Lower middle income

Table 1.28: Johansen cointegration test for additional functional forms (Stage I)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK (1850-1885)	expgdp and gdppc	r=0	25.162***	r=0	21.853***	r=0	30.367**	r=0	26.723***
		r≤1	0.095	r=1	0.095	r≤1	3.644	r=1	3.644
	expgdp and gdp	r=0	15.992**	r=0	12.787*	r=0	30.734***	r=0	27.022***
		r≤1	3.205	r=1	3.205	r≤1	3.712	r=1	3.712
Denmark (1854-1908)	expgdp and gdppc	r=0	6.875	r=0	6.606	r=0	12.713	r=0	6.834
		r≤1	0.269	r=1	0.269	r≤1	5.879	r=1	5.879
	expgdp and gdp	r=0	7.930	r=0	7.011	r=0	12.487	r=0	7.076
		r≤1	0.919	r=1	0.919	r≤1	5.411	r=1	5.411
Sweden (1881-1925)	expgdp and gdppc	r=0	16.589**	r=0	13.709*	r=0	27.793**	r=0	16.477
		r≤1	2.880	r=1	2.880	r≤1	11.316	r=1	11.316
	expgdp and gdp	r=0	16.599**	r=0	13.363*	r=0	28.807**	r=0	16.814
		r≤1	3.236	r=1	3.236	r≤1	11.992	r=1	11.992
Finland (1882-1937)	expgdp and gdppc	r=0	4.455	r=0	4.268	r=0	15.536	r=0	11.738
		r≤1	0.187	r=1	0.187	r≤1	3.798	r=1	3.798
	expgdp and gdp	r=0	4.713	r=0	4.623	r=0	15.557	r=0	11.705
		r≤1	0.090	r=1	0.090	r≤1	3.852	r=1	3.852
Italy (1862-1939)	expgdp and gdppc	r=0	26.833***	r=0	26.740***	r=0	31.215***	r=0	27.075***
		r≤1	0.093	r=1	0.093	r≤1	4.139	r=1	4.139
	expgdp and gdp	r=0	26.452***	r=0	26.445***	r=0	30.508***	r=0	26.451***
		r≤1	0.007	r=1	0.007	r≤1	4.057	r=1	4.057

Note: Note: See table 1.11.

Table 1.29: Johansen cointegration test with structural breaks for additional functional forms (Stage I)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK (1850-2010)	expgdp and gdppc	-	1875	r=0	35.38***	r=0	35.05*
				r≤1	9.34	r=1	9.32
	expgdp and gdp	-	1875	r=0	35.40***	r=0	34.89*
				r≤1	9.19	r=1	8.96
Denmark (1854-2010)	expgdp and gdppc	1870	-	r=0	20.71	r=0	27.86
				r≤1	6.35	r=1	6.55
	expgdp and gdp	1870	-	r=0	20.63	r=0	27.60
				r≤1	5.76	r=1	5.95
Sweden (1881-1925)	expgdp and gdppc	1912	-	r=0	37.28***	r=0	40.00**
				r≤1	9.93	r=1	11.53
	expgdp and gdp	1912	-	r=0	37.30***	r=0	40.26**
				r≤1	9.95	r=1	11.81
Finland (1882-1937)	expgdp and gdppc	1900	1915	r=0	47.30***	r=0	56.24***
		1917		r≤1	14.76	r=1	22.24
	expgdp and gdp	1900	1915	r=0	46.27***	r=0	48.69**
		1917		r≤1	14.77	r=1	16.58
Italy (1862-1939)	expgdp and gdppc	1914	1915	r=0	76.82***	r=0	75.07***
				r≤1	26.69***	r=1	26.08***
	expgdp and gdp	1914	1915	r=0	75.07***	r=0	74.37***
				r≤1	26.08***	r=1	25.69***

Note: See table 1.12.

Stage II: Upper middle income

Table 1.30: Johansen cointegration test for additional functional forms (Stage II)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	expgdp and gdppc	r=0	15.982**	r=0	13.356*	r=0	28.409**	r=0	17.986*
(1886-1972)		r≤1	2.626	r=1	2.626	r≤1	10.423	r=1	10.423
	expgdp and gdp	r=0	16.013**	r=0	12.495*	r=0	27.108**	r=0	16.222
		r≤1	3.518	r=1	3.518	r≤1	10.886	r=1	10.886
Denmark	expgdp and gdppc	r=0	14.083*	r=0	14.074**	r=0	17.637	r=0	14.982
(1909-1967)		r≤1	0.009	r=1	0.009	r≤1	2.654	r=1	2.654
	expgdp and gdp	r=0	13.148	r=0	13.148*	r=0	17.961	r=0	14.774
		r≤1	0.000	r=1	0.000	r≤1	3.187	r=1	3.187
Sweden	expgdp and gdppc	r=0	10.779	r=0	10.297	r=0	20.775	r=0	12.177
(1926-1967)		r≤1	0.482	r=1	0.482	r≤1	8.598	r=1	8.598
	expgdp and gdp	r=0	11.394	r=0	10.911	r=0	21.387	r=0	12.615
		r≤1	0.483	r=1	0.483	r≤1	8.772	r=1	8.772
Finland	expgdp and gdppc	r=0	27.075***	r=0	21.514***	r=0	36.038***	r=0	21.863**
(1938-1978)		r≤1	5.562**	r=1	5.562**	r≤1	14.175**	r=1	14.175**
	expgdp and gdp	r=0	27.572***	r=0	21.628***	r=0	36.111***	r=0	21.884**
		r≤1	5.944**	r=1	5.944**	r≤1	14.227**	r=1	14.227**
Italy	expgdp and gdppc	r=0	26.952***	r=0	26.781***	r=0	46.564***	r=0	37.031***
(1940-1977)		r≤1	0.171	r=1	0.171	r≤1	9.533	r=1	9.533
	expgdp and gdp	r=0	26.702***	r=0	26.542***	r=0	46.602***	r=0	37.052***
		r≤1	0.159	r=1	0.159	r≤1	9.549	r=1	9.549

Note: See table 1.11.

Table 1.31: Johansen cointegration test with structural breaks for additional functional forms (Stage II)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	expgdp and gdppc	1915	1918	r=0	32.60**	r=0	41.32**
(1886-1972)				r≤1	9.93	r=1	10.40
	expgdp and gdp	1915	-	r=0	31.42**	r=0	42.20**
				r≤1	9.73	r=1	10.34
Denmark	expgdp and gdppc	-	1920	r=0	49.17***	r=0	36.92
(1909-1967)			1933	r≤1	16.58	r=1	7.81
	expgdp and gdp	-	1920	r=0	49.21***	r=0	33.70
			1932	r≤1	17.02*	r=1	7.89
Sweden	expgdp and gdppc	1939	1939	r=0	35.38*	r=0	48.68**
(1926-1967)			1949	r≤1	13.97	r=1	13.49
	expgdp and gdp	1939	1939	r=0	36.10*	r=0	47.27*
			1949	r≤1	13.99	r=1	13.17
Finland	expgdp and gdppc	-	1951	r=0	33.45**	r=0	32.23
(1938-1978)				r≤1	9.60	r=1	13.70
	expgdp and gdp	-	1951	r=0	33.49**	r=0	30.74
				r≤1	9.40	r=1	12.06
Italy	expgdp and gdppc	-	-	r=0	-	r=0	-
(1940-1977)				r≤1		r=1	
	expgdp and gdp	-	-	r=0	-	r=0	-
				r≤1		r=1	

Note: See table 1.12.

Stage III: High income

Table 1.32: Johansen cointegration test for additional functional forms (Stage III)

		Johansen (1)				Johansen (2)			
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
UK	expgdp and gdppc	r=0	15.216*	r=0	13.821*	r=0	21.352	r=0	14.157
(1973-2010)		r≤1	1.395	r=1	1.395	r≤1	7.195	r=1	7.195
	expgdp and gdp	r=0	16.643**	r=0	13.959*	r=0	22.980	r=0	14.517
		r≤1	2.684	r=1	2.684	r≤1	8.463	r=1	8.463
Denmark	expgdp and gdppc	r=0	15.820**	r=0	14.797**	r=0	20.660	r=0	14.904
(1968-2010)		r≤1	1.023	r=1	1.023	r≤1	5.756	r=1	5.756
	expgdp and gdp	r=0	17.192**	r=0	14.617**	r=0	19.741	r=0	14.619
		r≤1	2.575	r=1	2.575	r≤1	5.121	r=1	5.121
Sweden	expgdp and gdppc	r=0	31.294***	r=0	31.113***	r=0	43.249***	r=0	31.272***
(1968-2010)		r≤1	0.181	r=1	0.181	r≤1	11.978	r=1	11.978
	expgdp and gdp	r=0	35.721***	r=0	35.082***	r=0	47.341***	r=0	35.694***
		r≤1	0.639	r=1	0.639	r≤1	11.646	r=1	11.646
Finland	expgdp and gdppc	r=0	11.199	r=0	10.386	r=0	20.261	r=0	10.454
(1979-2010)		r≤1	0.813	r=1	0.813	r≤1	9.807	r=1	9.807
	expgdp and gdp	r=0	12.340	r=0	9.974	r=0	20.453	r=0	13.118
		r≤1	2.366	r=1	2.366	r≤1	7.335	r=1	7.335
Italy	expgdp and gdppc	r=0	6.751	r=0	5.405	r=0	13.486	r=0	8.088
(1978-2010)		r≤1	1.346	r=1	1.346	r≤1	5.398	r=1	5.398
	expgdp and gdp	r=0	16.937**	r=0	16.225**	r=0	20.676	r=0	17.204
		r≤1	0.712	r=1	0.712	r≤1	3.471	r=1	3.471

Note: See table 1.11.

Table 1.33: Johansen cointegration test with structural breaks for additional functional forms (Stage III)

		Bai-Perron Breaks		Johansen (1)		Johansen (2)	
Country	g and y	g	y		Trace		Trace
UK	expgdp and gdppc	-	1993	r=0	30.80**	r=0	31.50
(1973-2010)				r≤1	10.35	r=1	8.98
	expgdp and gdp	-	1993	r=0	30.21**	r=0	28.05
				r≤1	6.71	r=1	4.48
Denmark	expgdp and gdppc	1982	-	r=0	31.66**	r=0	25.86
(1968-2010)				r≤1	8.33	r=1	7.99
	expgdp and gdp	1982	-	r=0	29.77**	r=0	22.79
				r≤1	7.48	r=1	7.70
Sweden	expgdp and gdppc	-	-	r=0	-	r=0	-
(1968-2010)				r≤1		r=1	
	expgdp and gdp	-	-	r=0	-	r=0	-
				r≤1		r=1	
Finland	expgdp and gdppc	1995	-	r=0	31.57**	r=0	34.90*
(1979-2010)				r≤1	5.70	r=1	8.02
	expgdp and gdp	1995	-	r=0	24.24	r=0	35.41*
				r≤1	5.59	r=1	8.66
Italy	expgdp and gdppc	1993	2001	r=0	37.16**	r=0	45.00*
(1940-1977)				r≤1	8.01	r=1	15.19
	expgdp and gdp	1993	2001	r=0	40.53**	r=0	45.42*
				r≤1	11.28	r=1	15.48

Note: See table 1.12.

Long-run causality results

Table 1.34 presents the estimated error correction terms and the corresponding t-statistics for the additional functional versions of Wagner's law. Again, the estimations support the findings of the main analysis: with an advanced degree of development, public spending does not react to changes in income as sensitive as in earlier development stages. Altogether, it can be noticed that the relationship between public spending and economic growth has weakened with an advanced stage of development irrespective of the selected functional form.

Table 1.34: Long-run causality and short-run adjustment for additional functional forms

		Full Sample		Stage I		Stage II		Stage III	
Country	G and Y	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$	$Y \rightarrow G$	$G \rightarrow Y$
UK	expgdp and gdppc	-0.039**	-0.002**	-0.337**	-0.235**	-0.079***	-0.004**	0.004	-0.048***
		(-2.010)	(-2.139)	(-2.165)	(-2.216)	(-2.453)	(-1.746)	(0.341)	(-4.803)
	expgdp and gdp	-0.044**	-0.002*	-0.339***	-0.133**	-0.095***	-0.004	0.004	-0.046***
		(-2.163)	(-1.826)	(-2.875)	(-1.888)	(-2.736)	(-1.178)	(0.295)	(-4.746)
Denmark	expgdp and gdppc	-0.099***	-0.008**	-	-	-0.576***	-0.022	-0.027***	0.001
		(-2.717)	(-1.770)			(-3.659)	(-0.497)	(-2.896)	(0.057)
	expgdp and gdp	-0.099***	-0.004	-	-	-0.493***	-0.009	-0.030***	-0.000
		(-2.848)	(-0.851)			(-3.440)	(-0.287)	(-2.821)	(-0.013)
Sweden	expgdp and gdppc	-0.067*	-0.013***	-1.382***	0.085	-0.167**	-0.019***	-0.016	-0.028***
		(-1.633)	(-3.455)	(-3.502)	(1.211)	(-2.312)	(-2.589)	(-0.297)	(-3.314)
	expgdp and gdp	-0.075**	-0.013***	-1.307***	0.096	-0.168**	-0.018***	-0.014	-0.027***
		(-1.707)	(-3.378)	(-3.399)	(1.614)	(-2.308)	(-2.747)	(-0.284)	(-3.232)
Finland	expgdp and gdppc	-0.085**	-0.009***	-0.334***	-0.039**	-0.239**	-0.018***	-0.022	-0.076***
		(-1.811)	(-2.541)	(-2.964)	(-2.229)	(-2.139)	(-3.881)	(-0.334)	(-3.705)
	expgdp and gdp	-0.088**	-0.009***	-0.381***	-0.036**	-0.235**	-0.018***	-0.020	-0.074***
		(-1.839)	(-2.562)	(-3.219)	(-2.119)	(-2.118)	(-3.851)	(-0.319)	(-3.655)
Italy	expgdp and gdppc	-0.114***	-0.009***	-0.235**	-0.061***	-0.135	-0.022***	-0.042*	-0.040***
		(-2.355)	(-3.752)	(-2.089)	(-4.944)	(-1.263)	(-6.325)	(-1.411)	(-3.299)
	expgdp and gdp	-0.120**	-0.009***	-0.257**	-0.052***	-0.139	-0.021***	-0.031	-0.049***
		(-2.442)	(-3.710)	(-2.292)	(-4.794)	(-1.304)	(-6.322)	(-1.183)	(-3.733)

Note: The table displays estimated error correction terms (ect) of corresponding VECMs. The t-statistics are presented in parenthesis. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels.

2 On the size of fiscal multipliers: a counterfactual analysis[40]

2.1 Introduction

The structural VAR approach to estimating the fiscal multipliers developed by Blanchard and Perotti (2002) has been applied widely in the literature in recent years.[41] It made a substantial progress in solving the identification problem, associated with the contemporaneous correlation of shocks.[42] In the present paper we point out that while the identification of shocks has been achieved, the approach still includes the dynamic interaction among policy instruments. The derived multipliers are therefore best characterized as forecasting multipliers where governments are assumed to follow their predicted paths after an initial fiscal shock.[43] In this paper, we raise the question whether this assumption is reasonable,

[40]This chapter is based on Kuckuck and Westermann (2013, 2014).
[41]See Ramey (2011a) and Parker (2011) for an overview.
[42]Earlier approaches to identification include the military approach of Barro (1981) as well as Ramey and Shapiro (1998).
[43]As a forecasting tool, the procedure has recently been evaluated by Blanchard and Leigh (2013).

when using the results for policy advice, as a benchmark for the DSGE modeling, or for testing the Keynesian model.

We start our analysis by illustrating that there exists a significant and economically sizeable effect of a shock in expenditure on net taxes and vice versa. We find that the effect of a shock in expenditure on net taxes is positive, i.e. expenditures today tend to be financed by tax increases in the immediately following quarters. With regard to taxes, we have the opposite finding. After a standard positive shock to net taxes, there is a significant response of expenditure that is negative. Furthermore both series are autocorrelated. A fiscal policy shock will lead to further changes in fiscal policy in the subsequent quarters.

In order to isolate the effects of a pure spending and pure tax shock, we implement the following counterfactual analysis: we first estimate the model using the Blanchard and Perotti (2002) approach. When computing the impulse response functions, however, we shut down the channel that captures the discretionary dynamic interaction among policy instruments as well as each policy instruments' autocorrelation (i.e. restrict their responses to zero).[44] All other responses remain unrestricted. In particular the indirect effect that government spending has on net taxes - via automatic stabilizers - remains included in the simulation.

The main result of our analysis is that our counterfactual multiplier is substantially larger than the forecasting multiplier from standard SVAR estimates in the case of an expenditure

[44] The same argument has been made in the context of monetary policy by Ramey (1993). In her paper, she isolates the credit channel of monetary policy by shutting down the policy-velocity channel when computing impulse response functions. Our analysis translates this idea to the context of fiscal policy and the discussion on the size of fiscal multipliers. See also the working paper version of Blanchard and Perotti (2002), who already raise this issue in the extended version of their paper, as well as Perotti (2005), who used a similar method as a robustness test.

shock. The tax multiplier is initially smaller, but gets larger at longer horizons. Finally, when both spending and net taxes experience a shock at the same time, the counterfactual multiplier is close to one, as predicted by Haavelmo (1945), while the forecast multiplier is nearly zero.

We investigate the sensitivity of our findings in several robustness regressions. First, we extend the analysis to a five-variable VAR, including inflation and interest rates as additional control variables. Second, we exclude the post-financial crisis time period from our sample, and also estimate the regressions in the original Blanchard and Perotti (2002) sample. Furthermore, we add a dummy variable, capturing the 1975Q2 tax cut period. Finally, we also extend the lag length of the VAR and control for the level of public debt. Overall, the differences between the counterfactual and the forecasting multipliers remain remarkably robust across these different specifications.

Our analysis does not imply that the Blanchard and Perotti (2002) procedure is incorrect or yield biased results. We do argue however that it must be interpreted with caution whenever there is a sizeable interaction among policy instruments or the autocorrelation of policy instruments is high. If the dynamic responses are strong, the Blanchard and Perotti multiplier must be interpreted as a forecast of the future reaction of GDP that includes further future changes in spending and net taxes, which are triggered by the initial fiscal shock.

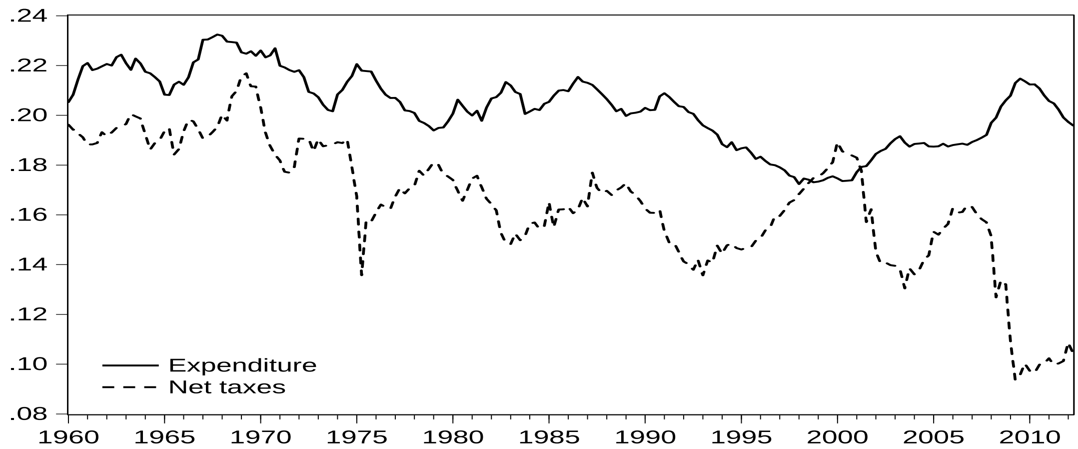
If the aim of the analysis is to use the results for policy consulting or as an input for other counterfactuals in a DSGE framework, the alternative approach suggested in this paper may be useful. In both cases, one would like to ask the question: what is the effect of an additional Dollar spent on future GDP, letting other instruments unchanged? Or put differently: what is the elasticity of GDP to a shock in government spending? To assess this question, and to move the analysis closer to the Keynesian model with its various crowding-out effects, we highlight the importance of a counterfactual analysis in our paper. We also provide an example of a fiscal program that was intended to be predominantly a reduction in net taxes, by looking at the American Recovery and Reinvestment Act (ARRA) of 2009.

2.2 Data and preliminary analysis

We start our analysis by plotting the data of expenditure and net taxes as a percentage of GDP. The solid line in figure 2.1 traces the expenditure/GDP ratio and the dotted line, net taxes/GDP.[45] The years from 1960 to 1997 are familiar from the Blanchard and Perotti (2002) article. In the past years, especially since 2007/2008, there has been a widening gap between expenditure and net taxes. This gap reflects the expansionary fiscal policy in response to the financial crisis. Initially both instruments have been used, as expenditure goes up and net taxes go down - a process that has been gradually reversed in the last four years of the sample period. In order to abstract from this exceptional period, we conduct the late analysis also in a reduced sample that stops in 2006Q4, the year before the crisis.

[45] See section 2.A.1 in the appendix for data sources and definitions.

Figure 2.1: Expenditure and net taxes to GDP ratios



Note: The graph visualizes net taxes and spending as shares of GDP. See section 2.A.1 in the appendix for data definitions and sources.

Applying the Augmented Dickey-Fuller (ADF) test as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test, we find that all variables have a unit root in levels and are stationary in first differences. The unit root test statistics are reported in section 2.A.3 in the appendix of this chapter. Furthermore, the test statistics in the appendix show that the three variables are not cointegrated. We therefore estimate the dynamic interactions between the variables in a VAR in first differences.

2.3 Results

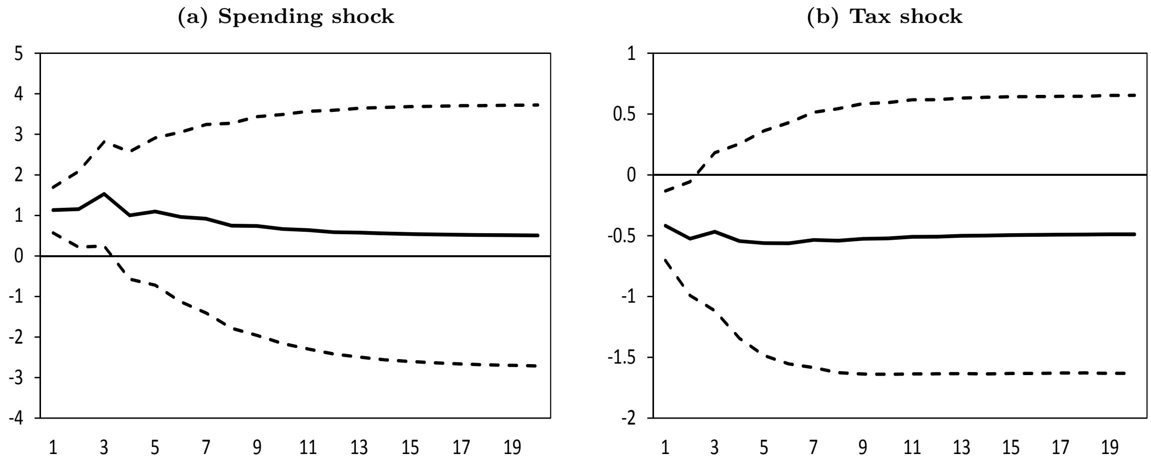
Forecast multipliers

In this section, we estimate the impulse response patterns of a shock in expenditure and net taxes on GDP, using the Blanchard and Perotti (2002) identification procedure.[46] Figure

[46] Following the methodology of Blanchard and Perotti (2002), we calculate for the updated data set a net tax elasticity to GDP of 2.756. Details on the calculation procedure can be found in section 2.A.2 in

2.2 displays the point estimates and standard errors, which contain the familiar result that spending has a positive and significant impact on GDP, while taxes have a negative impact. Table 2.1 contains information on the exact quantitative impact. The magnitude of the multipliers is comparable to those that have been reported in the literature.

Figure 2.2: Response of GDP to expenditure and tax shocks



Note: Panel (a) displays the response of GDP to a one unit increase in public spending. Panel (b) shows the response of GDP to a one unit increase in net taxes. The dashed lines show the asymptotic standard errors.

Table 2.1: Response of GDP to expenditure and tax shocks

	Q1	Q2	Q3	Q4	Q6	Q8	Q12	Q14	Q18	Q20	Peak
EXP	1.131	1.152	1.530	1.000	0.962	0.745	0.587	0.551	0.513	0.504	1.530 (3)
	(0.28)	(0.47)	(0.64)	(0.79)	(1.05)	(1.26)	(1.50)	(1.57)	(1.59)	(1.61)	(0.64)
TAX	-0.418	-0.524	-0.468	-0.544	-0.563	-0.541	-0.509	-0.499	-0.491	-0.489	-0.563 (6)
	(0.14)	(0.23)	(0.32)	(0.40)	(0.50)	(0.54)	(0.56)	(0.57)	(0.57)	(0.57)	(0.49)

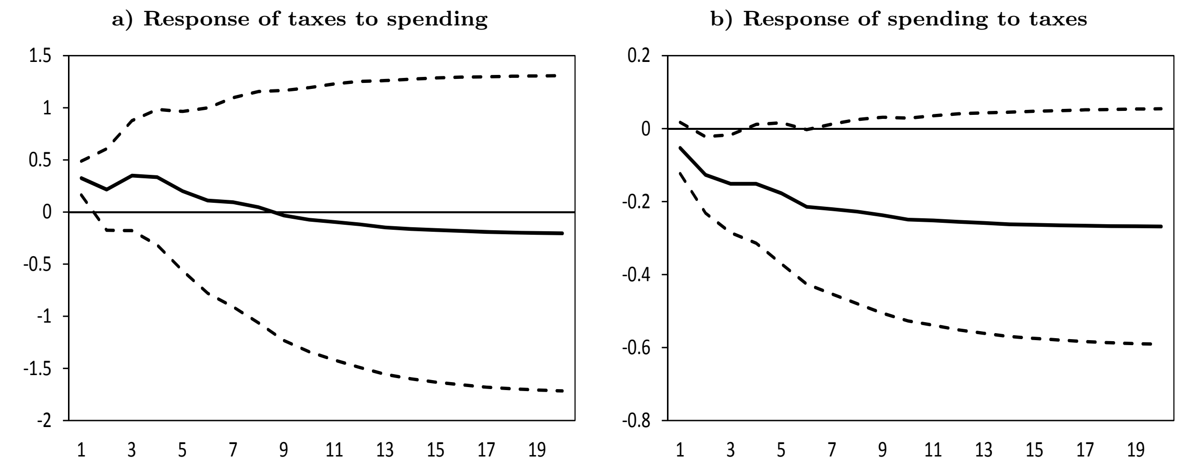
Note: The table displays output multipliers with respect to government spending (EXP) and tax (TAX) shocks. Corresponding standard errors are shown in parentheses.

the appendix.

Interaction among policy instruments

Standard estimations of the Keynesian multiplier typically include the dynamic interactions among the policy instruments, i.e. the reaction of net taxes to expenditure is included when simulating the impact of expenditure on GDP. In figure 2.3, we show that these interactions among the policy variables are economically sizeable and statistically significant. Table 2.2, again, reports the exact corresponding values of the point estimates and confidence intervals. We find that in our sample period, there has been a significant positive response of taxes to a change in expenditure, which implies that an increase in spending has been financed by a subsequent increase in net taxes.[47]

Figure 2.3: Interaction among policy instruments



Note: Panel a) displays the response of net taxes to a one unit increase in public spending. Panel b) displays the response of public spending to a one unit increase in net taxes. The dashed lines show the asymptotic standard errors.

Part of this reaction is certainly due to automatic stabilizers being at work. An increase in expenditures increases GDP, which leads to higher net taxes. The other part, however,

[47]These findings are in line with Bohn (1991) who shows by examining historical U.S. budget data that around 30 - 35% of all deficits due to higher government spending have been eliminated by subsequent tax increases.

is a discretionary response of government, the need to finance additional expenditures. In the subsequent counterfactual simulations, we only shut down the latter channel. Similarly there is also a significant negative reaction of expenditure to a shock in net taxes. This means that, on average, an increase in net taxes has been associated with a subsequent decrease in expenditure.

Table 2.2: Interaction among policy instruments

	Q1	Q2	Q3	Q4	Q6	Q8	Q12	Q14	Q18	Q20	Peak
G→T	0.326	0.216	0.349	0.333	0.110	0.047	-0.119	-0.163	-0.197	-0.204	0.349 (3)
	(0.08)	(0.20)	(0.26)	(0.33)	(0.44)	(0.55)	(0.69)	(0.72)	(0.75)	(0.76)	(0.26)
T→G	-0.053	-0.127	-0.151	-0.151	-0.214	-0.227	-0.255	-0.262	-0.267	-0.268	0.268 (20)
	(0.04)	(0.05)	(0.07)	(0.08)	(0.11)	(0.13)	(0.15)	(0.15)	(0.16)	(0.16)	(0.16)

Note: The table displays the response of net taxes and expenditures with respect to government spending (G→T) and tax shocks (T→G). Corresponding standard errors are shown in parentheses.

Counterfactual multiplier

In figure 2.4 we show that the difference between the forecast and the counterfactual multiplier is sizeable and economically important. The counterfactual multiplier is computed from a simulation where the interaction among policy variables as well as the autocorrelation of each policy variable is restricted in the following way: the reaction of spending to a shock in net taxes is restricted to zero. The reaction of net taxes to spending and the autocorrelation in net taxes are restricted to the level that would be observed if only automatic stabilizers had been at work, i.e. we shut down the discretionary response of net taxes to expenditures.[48]

[48] The automatic stabilizer is computed by multiplying the impulse response function of output to a shock in expenditure by the tax elasticity that is also used in the contemporaneous variance-covariance matrix. Additionally, we set the structural correlation between taxes and spending in both directions equal to

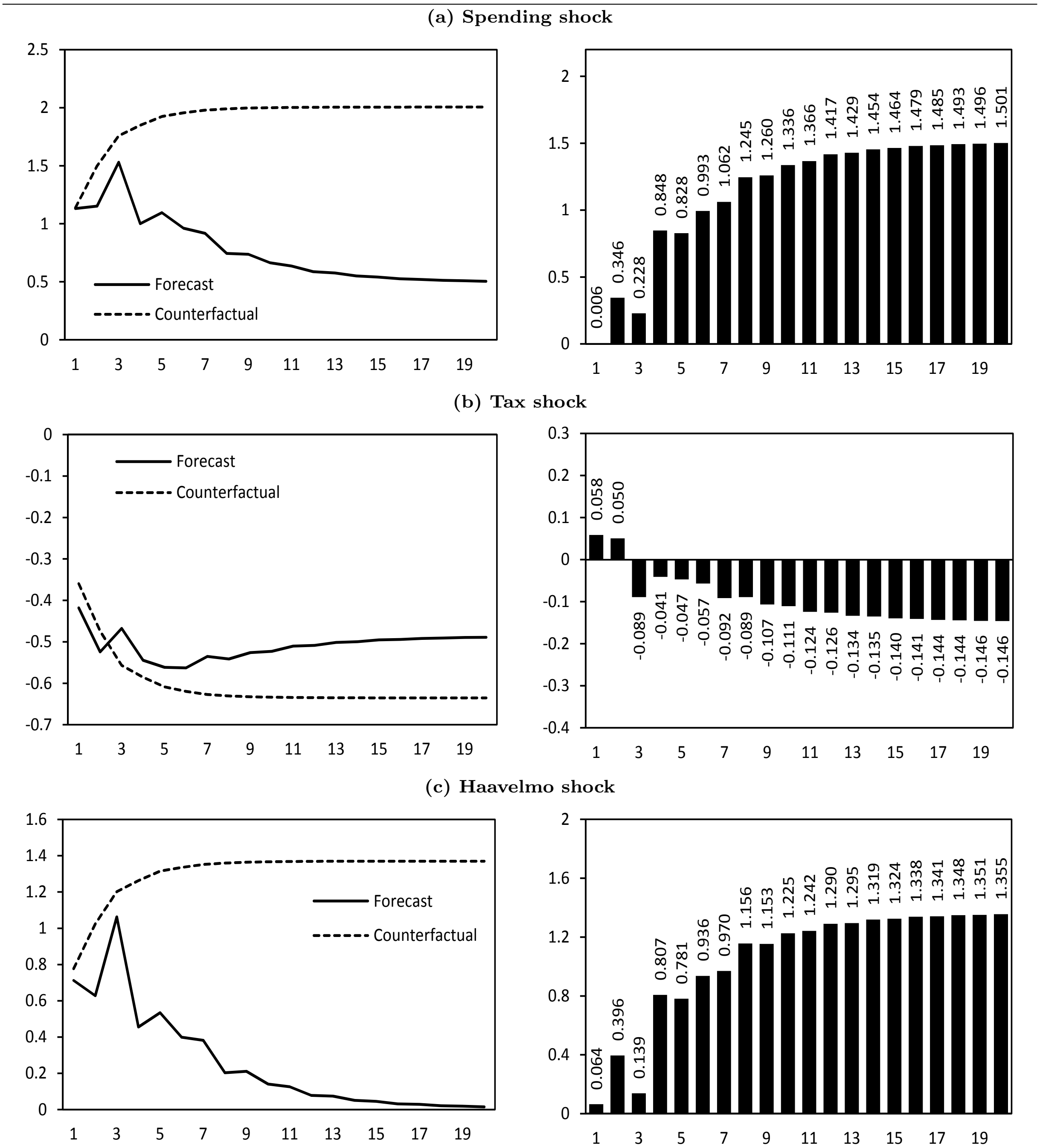
The dotted line in figure 2.4 shows the response of GDP to a shock in expenditure and net taxes, respectively. For comparison, the solid line traces the forecast multiplier from figure 2.2. We find that in the case of the expenditure multiplier, the effect on GDP is larger (panel (a)) than in the forecasting scenario. In the case of the tax multiplier (panel (b)) the counterfactual reaction is initially smaller, but at longer horizons larger than the forecasting multiplier. The total difference of the estimates is also displayed on the right hand side of figure 2.4. The peak of this difference is equal to 1.501 after five years (20 periods) for the expenditure multiplier and -0.146 after five years (20 periods) for the tax multiplier. The cumulative effect after one year (4 periods) is 0.848 larger for the expenditure multiplier and 0.041 smaller for the tax multiplier.

As a next step, we analyze the effect of a Haavelmo shock, i.e. the joint increase of expenditure and net taxes at the same time. As illustrated in panel (c) of figure 2.4, the magnitude of the shock is close to one (1.36) in the counterfactual analysis, while it is nearly zero in the forecast scenario. Abstracting from the interaction among policy instruments thus leads to results that are closer to economic theory (see Haavelmo (1945)).

In the case of the expenditure multiplier, the increase is plausible as it no longer includes the parallel increase in net taxes. The larger tax multiplier at longer horizons seems more surprising. This effect is likely to be driven by the autocorrelation in both instruments. Note however, that our sensitivity analysis shows that in shorter samples, leaving out the financial crisis, the tax multiplier is lower and not larger at longer horizons.

zero.

Figure 2.4: Difference between forecast and counterfactual multipliers



Note: The left column shows the forecast and counterfactual multipliers to a one standard deviation shock of expenditures and taxes, respectively. The right column displays the difference between the forecast and the counterfactual multiplier.

Sensitivity analysis

We investigate the robustness of our findings in a set of sensitivity tests that are summarized in table 2.3.[49] The rows in segment (1) of the table report the results of the baseline regression. In segment (2), we display the results for a VAR that is estimated with five variables, instead of three. The five-variable VAR includes the interest rate and the price level as additional variables, as for instance in Perotti (2005). For both policy instruments the importance of the counterfactual analysis remains clearly visible. The multipliers are substantially larger if the interaction of policy variables is eliminated. In segment (3) and (4) of table 2.3, we reduce the sample to exclude the last years of the global financial and economic crisis, and also replicate the original Blanchard and Perotti (2002) sample. In this reduced sample, the multipliers are lower than in the full sample, but again the effect of the counterfactual analysis remains substantial.

Further robustness tests include a current dummy as well as four lags of a dummy variable for the net tax cut 1975Q2 period in segment (5), choosing a higher lag order in segment (6) and controlling for a possible debt feedback, segment (7), by including the levels of public debt as an additional control variable (see Favero and Giavazzi (2007)). In all specifications, the differences of running a counterfactual regression display some variance with regard to the magnitude. In short samples, the effect of the tax multiplier goes in the opposite direction. But overall, the effect remains significant and economically sizeable in all specifications.

[49] A visualized representation of these findings can be found in section 2.A.4 in the appendix.

Table 2.3: Output multiplier with respect to government and tax shocks

		Q1	Q2	Q3	Q4	Q6	Q8	Q10
(1) Baseline regression								
Expenditures	FC	1.131*	1.152*	1.530*	1.000	0.962	0.745	0.664
	Δ(CF)	0.006	0.346	0.228	0.848	0.993	1.245	1.336
Net taxes	FC	-0.418*	-0.524*	-0.468	-0.544	-0.563	-0.541	0.523
	Δ(CF)	0.058	0.050	-0.089	-0.041	-0.057	-0.089	-0.111
(2) 5-VAR								
Expenditures	FC	1.066*	1.121*	1.321*	0.663	0.480	0.087	-0.114
	Δ(CF)	0.004	0.287	0.217	0.774	0.743	0.913	0.992
Net taxes	FC	-0.368*	-0.386	-0.218	-0.189	-0.057	0.025	0.055
	Δ(CF)	0.048	-0.006	-0.291	-0.354	-0.534	-0.573	-0.616
(3) Sample: 1960Q1-2006Q4								
Expenditures	FC	0.000	0.387	0.219	0.767	0.757	0.909	0.914
	Δ(CF)	-0.018	0.263	0.365	0.471	0.445	0.477	0.463
Net taxes	FC	-0.522*	-0.589*	-0.625	-0.858	-1.018	-1.041	-1.025
	Δ(CF)	0.027	-0.034	-0.093	0.096	0.191	0.190	0.165
(4) BP Sample: 1960Q1-1997Q4								
Expenditures	FC	1.226*	1.154	1.576	1.109	1.425	1.316	1.360
	Δ(CF)	0.000	0.020	-0.006	0.106	0.164	0.288	0.330
Net taxes	FC	-0.117	-0.194	-0.165	-0.412	-0.557	-0.545	-0.528
	Δ(CF)	-0.003	0.039	-0.015	0.216	0.336	0.314	0.292
(5) Dummy: 1975Q2								
Expenditures	FC	1.102*	1.148*	1.544*	1.079	1.051	0.839	0.732
	Δ(CF)	0.015	0.295	0.126	0.653	0.785	1.1031	1.149
Net taxes	FC	-0.507*	-0.550*	-0.466	-0.417	-0.414	-0.354	-0.326
	Δ(CF)	0.099	0.022	-0.144	-0.216	-0.257	-0.330	-0.362
(6) VAR(8)								
Expenditures	FC	1.138*	1.011*	1.151	0.416	0.772	1.287	0.987
	Δ(CF)	0.007	0.492	0.670	1.622	1.414	0.898	1.251
Net taxes	FC	-0.435*	-0.523*	-0.456	-0.578	-0.474	-0.536	-0.612
	Δ(CF)	0.061	0.031	-0.140	-0.089	-0.241	-0.179	-0.120

Table is continued on the next page.

Table 2.3 – continued

		Q1	Q2	Q3	Q4	Q6	Q8	Q10
(7) Debt feedback								
Expenditures	FC	1.336*	1.204*	1.421	0.878	0.977	0.796	0.757
	Δ(CF)	0.001	0.478	0.644	1.359	1.427	1.661	1.716
Net taxes	FC	-0.332*	-0.363	-0.332	-0.459	-0.434	-0.416	-0.398
	Δ(CF)	0.006	-0.048	-0.171	-0.087	-0.152	-0.183	-0.205

Note: The first row displays the response of GDP to a spending (EXP) and tax (TAX) shock in a scenario where all transmission channels are open (Forecast multiplier (FC)). The second row presents the change of these output responses in a scenario where the expenditure-tax channel is closed (Counterfactual Multiplier (CF)). The symbol * indicates that the reactions are statistically different from zero.

An application: The American Recovery and Reinvestment Act of 2009

At the beginning of 2009, in direct response to the economic crisis, the Congress of the United States passed the American Recovery and Reinvestment Act (ARRA), which involved a combination of tax cuts, unemployment benefits, federal aid to states and localities as well as increases in government spending and investment. The total approximate cost of the stimulus package was estimated by the Congressional Budget Office (CBO) to be 831 Billion US-Dollars. Although the cost was spread out between 2009 and 2019, most of the tax and spending benefits accrued in the first two years after passing the bill. In the context of our paper, it provides an interesting case study, as the fiscal stimulus was intended to be predominately a reduction in net taxes, which received only little support in the expenditure side, and thus broke from previous paths. In the following analysis, we estimate the effect of the package taking into account the actual changes that were made in 2009 to 2013, rather than the average prediction response from SVAR estimates. Table 2.4 displays the quarterly effects of ARRA on net taxes and expenditures, which are calculated from data collected by the Bureau of Economic Analysis. The largest proportion of the economic stimulus package consists of net tax cuts (including increases in transfer and social benefit payments) adding up to 632.85 billion US-Dollars until the first quarter of 2013. Increases in public consumption and gross investment expenditures have been considerably lower with an aggregate value of 57.75 billion US-Dollars.

Table 2.4: Net tax cuts and increase in expenditures as part of the ARRA stimulus package

	2009				**2010**				**2011**				**2012**				**2013**
	Q1	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**	**Q2**	**Q3**	**Q4**	**Q1**
TAX	21.45	70.7	69.5	71.78	82.53	72.65	70.6	69.8	22.35	23.13	19.58	16.55	5.75	4.65	4.13	3.3	4.43
EXP	0	0.28	3.78	3.45	4.38	5.45	6.88	5.48	4.83	4.53	4.43	3.45	2.73	2.55	2.35	1.73	1.5
ALL	21.45	70.98	73.28	75.23	86.91	78.1	77.48	75.28	27.18	27.66	24.01	20.0	8.48	7.2	6.48	5.03	5.93

Note: The table displays net tax reliefs (TAX), increases in public expenditures (EXP) as well as the overall benefits (ALL) as part of the ARRA economic stimulus package (in billion US-Dollars).

In order to analyze the impact of ARRA on GDP, we multiply the net tax reliefs and expenditure increases of each quarter by the estimated counterfactual multipliers. Accordingly, the estimated effects on GDP are summarized in figure 2.5. The analysis shows that the impact of ARRA on output peaked in the first quarter of 2010 and has since diminished. A look at the individual quarters illustrates that ARRA raised GDP at most by about 1.5 percent.[50] On aggregate, ARRA had an overall impact on GDP of 510.68 billion US-Dollars. These estimates lie within the range of official estimates published by the CBO but closer to the lower bound of their estimates. They are still somewhat higher, however, than the multipliers derived from a DSGE model by Cogan et al. (2010), who estimate a quarterly impact of ARRA on GDP of less than 0.5 percent.[51]

[50] Note that in this exercise, each quarter must be treated as a transitory shock rather than a percent one, as assumed by the VAR.

[51] Davig and Leeper (2011) demonstrate that the effect of ARRA crucially depends on the interaction of monetary and fiscal policies.

Figure 2.5: Estimated effects of ARRA on GDP

Note: The figure visualizes the estimated effects of the ARRA economic stimulus package on GDP. The solid line displays the impact of ARRA on GDP measured in billion US-Dollars; the bar graph shows the effects as a percentage of GDP.

2.4 Related literature and conclusions

The discussion of the interaction between policy instruments and other economic variables dates back to the early contributions in empirical macroeconomics of Sims (1980). Sims proposed a SVAR partly as a response to the Lucas critique. Lucas (1976) had pointed out that regression analysis up to this point did not account for agents behavioral responses to a policy change. Most of the literature in SVAR modeling has since focused on the correct identification of structural shocks. This paper illustrates for the case of fiscal policy, however, that the dynamic interactions between instruments following the shock are also important.

The issue of interaction among policy instruments has been first raised in the working paper version of Blanchard and Perotti's (2002) seminal paper and in Perotti (2005).[52] They report that multipliers are slightly higher when they set the reaction of the respective other

[52] In a recent paper, Mountford and Uhlig (2009) have proposed an alternative approach to the common Blanchard and Perotti (2002) SVAR setup that is built on long-run identifying restrictions. In this paper, the authors also close the interaction channel among policy instruments.

policy instrument to zero. Our analysis elaborates on this point and illustrates the effects in an updated data set. When we replicate our exercise in their sample period, the differences are somewhat smaller. Apparently, this is an additional identification issue that has grown in importance over time.

Note that our analysis is not entirely free of the Lucas critique. The previous literature is based on the assumption that policy instruments will continue to react to each other as they have done in the past. Our assumption is that there will be no such interaction. Both are assumptions that may be reasonable to make in different contexts. In the context of a forecasting exercise, the former is clearly better. The latter seems relevant from a theoretical perspective and as an alternative scenario when addressing particular policy questions.[53]

The size of the multipliers as well as the interaction of policy instruments has also been addressed in the context of DSGE models (see Parker (2011) for a critical overview and a comparison of SVAR and DSGE approaches). It has been shown, for instance, that parallel tax reductions dramatically reduce the impact of the expenditure multiplier (see Uhlig (2010)). Davig and Leeper (2011) furthermore illustrate that the interaction of monetary and fiscal policies is important for the size of fiscal multipliers.

Our contribution may be relevant for this line of research in the following sense. DSGE studies typically use of the SVAR estimates as an input and calibrate parameters that match the data. Using these calibrates, they then perform a much wider range of counter-

[53]Some caveats on fiscal multipliers derived from VAR models are discussed in section 2.A.5 in the appendix.

factuals that would not be feasible in the SVAR setup. With regard to this approach, it is important that the SVAR estimates do not already include the interaction among policy instruments. Our counterfactual multiplier may thus be a better benchmark for DSGE models than those from Blanchard and Perotti (2002).

More recently, the size of the multiplier has also been investigated from different angles. Ilzetzki et al. (2013) for instance perform a large cross-country analysis and show that the size of the multiplier depends on several country-specific characteristics, including the state of development, the exchange rate regime, and the level of indebtedness. Ramey (2011b) furthermore highlights the importance of the exact timing of the spending shocks. Hall (2009) and Christiano et al. (2011) point out that multipliers are larger when interest rates are close to their zero-lower band. Finally, Auerbach and Gorodnichenko (2012) show that fiscal multipliers differ in different stages of the business cycle and are substantially larger in recessions. They also document that the nature of spending matters and further control variables are needed.

Overall, there is both, renewed academic interest in the size and estimation of fiscal multipliers, as well as an increased policy relevance. While Keynesian policies have not been used for many years in most countries, the 2007/8 financial crisis has seen a revival of stabilization policies. Our contribution is intended to further refine the literature such that it gives better guidance for the optimal use of fiscal policy instruments and to illustrate the limits of conventional SVAR estimates of fiscal multipliers for concrete policy advice.

2.A Appendix

In the following sections additional information and robustness checks are provided. The appendix is organized as follows: **section 2.A.1** provides essential information about data sources and definitions. Details on the construction of the net tax elasticity can be found in **section 2.A.2**. Unit root and cointegration properties of the variables are discussed in **section 2.A.3**. Subsequently, **section 2.A.4** displays fiscal multipliers derived from different model specifications and visualizes the differences between forecast and counterfactual multipliers. Finally, **2.A.5** discusses some caveats on fiscal multipliers derived from SVAR models.

2.A.1 Data sources and definitions

All the data unless otherwise noted, are taken from the National Income and Product Accounts collected by the Bureau of Economic Analysis (2013) (BEA) and cover the time period from 1960Q1 to 2012Q2. Furthermore, all data is seasonally adjusted at annual rates by the original source, applying the X-12 Autoregressive Integrated Moving Average (ARIMA) procedure. The variable definitions follow Blanchard and Perotti (2002). Net taxes consist of total revenue of federal, state, and local government less social contributions, grants paid to private households and enterprises, and interest payments. Government expenditures are defined as consumption expenditure plus gross investment at federal, state, and local stage. In order to express the variables in real terms, we use the GDP deflator. Table 2.5 provides a general overview about the data definitions and sources.

Table 2.5: Data for the calculation of government expenditures and revenues

Variable		**Definition**	**Data code**
		3-VAR	
Government expenditures		Federal government consumption expenditure and gross investment	A823RC1
	+	State and local government consumption expenditure and gross investment	A829RC1
Net taxes		Federal current receipts	W005RC1
	+	State and local current receipts	W023RC1
	-	Federal grants-in-aid to state and local governments	B089RC1
	-	Federal current transfer payments to persons	B087RC1
	+	Federal current transfer receipts from persons	B233RC1
	-	Government social benefit payments to persons	B109RC1
	-	Federal interest payments	A091RC1
	+	Federal interest receipts	B094RC1
	-	State and local interest payments	B111RC1
	+	State and local interest receipts	B112RC1
	+	Federal dividends	W053RC
	+	State and local dividends	B081RC1
Output		Gross Domestic Product	A191RC1
		5-VAR	
Price index		GDP deflator (2005=100)	B191RG3
Interest rate[54]		3-Month Treasury Bill: Secondary Market Rate	TB3MS

[54]Source: Board of Governors of the Federal Reserve System (2013)

Data for the calculation of the net tax elasticity (see section 2.A.2) is taken from the BEA as well as the U.S. Department of Labor: Bureau of Labor Statistics (2013) (BLS). In the case the series were not seasonally adjusted by the original source, we applied the X-12 ARIMA procedure. An overview about the tax categories and data sources can be found in table 2.6.

Table 2.6: Data for the calculation of the exogenous elasticities

Variable	**Definition**	**Data code**
	Category of taxes/transfer payments	
Indirect taxes	Taxes on products and imports	W056RC1
Personal income taxes	Income taxes	B245RC1
Social security taxes	Contributions for government social insurance	W782RC1
Corporate income taxes	Taxes on corporate income	W025RC1
Transfers	Federal current transfer payments + State and local current transfer payments	A063RC1
	Additional data for calculation	
Corporate profits	Corporate Profits with Inventory Valuation Adjustment (IVA) and Capital Consumption Adjustment	CPROFIT (BEA)
Earnings	Compensation of Employees: Wages & Salary Accruals	WASCUR (BEA)
Employment	All Employees: Total non-farm	PAYEMS (BLS)

2.A.2 Net tax elasticity

In order to calculate the net tax elasticity, we follow the same methodology as in Blanchard and Perotti (2002). Accordingly, we consider four categories of taxes: indirect taxes, personal income taxes, social security taxes as well as corporate income taxes. The net tax elasticity to GDP of category i is calculated as the product of the tax elasticity to its own tax base ($\tau_{tax,base}$) and the elasticity of the tax base to output ($\tau_{base,GDP}$). The overall net tax elasticity is then calculated by the sum of every tax elasticity weighted by the share of each tax component in the sum of all tax revenues. Information about the tax elasticity to its own base are taken in some parts from Girouard and André (2005) who account for individual tax code information and income distributions. Additional items are obtained from simple linear regressions: i) the elasticity of earnings to employment is estimated from a regression of the log change of the earnings on the first lead and lags 0 to 4 of the log change in employment, ii) the elasticity of employment to output is estimated from a regression of the log change of employment on the first lead and lags 0 to 4 of the log change in output, and iii) the elasticity of the base of corporate income taxes to output is estimated from a regression of quarterly changes of corporate profits on the first lead and lags 0 to 4 of changes in output.[55] Details on the calculation procedure can be found in the appendix of Blanchard and Perotti's (2002) seminal paper.

Finally, transfer payments are also supposed to be sensitive to the business cycle, e.g. an increase in output reduces unemployment benefits. As we do not have reliable data for our sample, we assume a value of -0.2 for elasticities of transfers to GDP, which is based on

[55]General information about the data sources are provided in the previous section 2.A.1.

the annual elasticity of total current expenditures taken from Girouard and André (2005). Table 2.7 provides a general overview about the calculated sub-components and the overall net tax elasticity in use.

Table 2.7: Net tax elasticities

	$\tau_{tax,base}$	$\tau_{base,GDP}$	weighted elasticity
Indirect taxes	1	1	1.05
Personal income taxes	1.50	0.11	0.05
Social security taxes	0.94	0.11	0.10
Corporate income taxes	0.85	3.93	1.23
Transfers	1	-0.2	-0.33
Net taxes elasticity			2.76

Note: Author's calculation based on Blanchard and Perotti (2002) and Girouard and André (2005). The within-quarter net tax elasticity with respect to output η_i is defined as $\sum_i \tau_{tax_i,base_i} \cdot \tau_{base_i,GDP} \cdot \frac{T_i}{T}$ where $\tau_{tax_i,base_i}$ denotes the elasticity of taxes of type i to their tax base and $\tau_{base_i,GDP}$ denotes the tax base i to GDP. The level of net taxes T can be written as $T = \sum_i T_i$, where the T_i's are positive if they correspond to taxes and negative if they correspond to transfers.

2.A.3 Unit root and cointegration analysis

This section provides some information about the unit root and cointegration properties of the variables. In fact, a VAR model estimated in first differences in the presence of cointegration relations is misspecified.[56] Table 2.8 and table 2.9 indicate that output (GDP), public expenditures (EXP) as well as net taxes (TAX) have a unit root in levels and are stationary in first differences. Further, the Johansen as well as the Engle-Granger cointegration tests do not find any cointegration relations between the variables. Hence, we estimate the dynamic interactions between the variables in a VAR in first differences.

Table 2.8: Unit root results

	ADF		**KPSS**	
	Level	1st Difference	Level	1st Difference
GDP	-1.848	-6.813***	1.829***	0.299
EXP	-1.947	-12.702***	1.775***	0.178
TAX	-2.169	-13.004***	1.305***	0.139

Note: The Augmented Dickey-Fuller (ADF) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests are calculated including a constant in the test equation. The lag length of the ADF tests is selected by SIC, while the bandwidth for the KPSS test is selected based on Newey-West using Bartlett kernel. The symbol *** indicates significance at the 1% levels.

Table 2.9: Cointegration tests

	Johansen test			**Engle-Granger test**	
	Max-Eigen.		Trace	Dependent	z-statistic
r=0	17.508	r=0	36.827	GDP	-12.216
r≤1	13.610	r≤1	19.319	TAX	-13.526
r≤2	5.708	r≤2	5.708	EXP	-11.312

Note: The Johansen as well as the Engle-Granger cointegration tests allow for a constant and a trend in the cointegration space.

[56] Following the intertemporal budget constraint of the government, an obvious candidate for cointegration is the relationship between taxes and expenditures (see Trehan and Walsh (1991)). For further information see section 4.A.2, which provides an overview of the empirical literature on debt sustainability.

2.A.4 Alternative model specifications

As fiscal multipliers derived from SVAR models are known to be very sensitive to methodological issues, we investigate the sensitivity of our findings in several robustness regressions. With regard to subsample stability, we also estimate multipliers with alternative sample definitions.[57] Figure 2.6 visualizes the response of output to fiscal shocks and displays the difference between multipliers derived from a forecast and counterfactual scenario under the assumption of alternative VAR specifications.

As far as the expenditure shock is concerned (panel (a)), it is shown that the magnitude of the impact multiplier is very robust towards different model specifications and sample sizes. Nevertheless, in the long run, the impulse response patterns expose some differences. In particular, the five-variable model (5-VAR) - including interest rate and price level as additional variables - deviates the most from the base case scenario. While the benchmark multiplier stays positive throughout time, the impulse response pattern of the 5-VAR model declines and gets negative after nine periods, which might be a consequence of additional interest rate and price related crowding-out effects.

Still, the difference between the forecast and counterfactual multiplier remains significant and economically sizeable in all specifications. In shorter samples, leaving out the financial crisis as well as replicating the original Blanchard and Perotti sample, the difference between the multipliers are smaller compared to the other estimations.

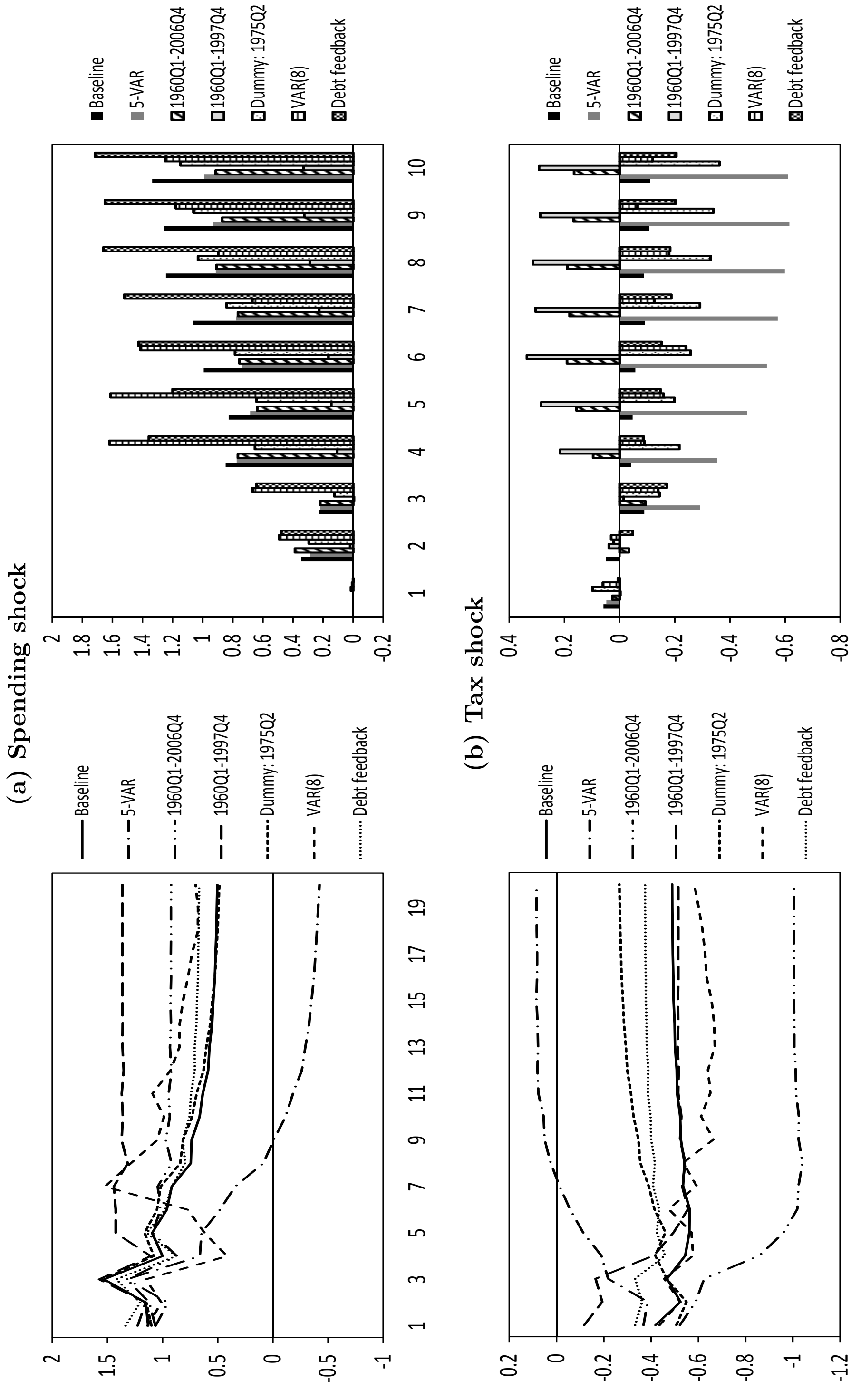
The impulse response pattern of tax shocks and the difference between the multipliers are

[57]The robustness of fiscal multipliers is also discussed in section 2.3 in the main part of this chapter.

visualized in panel (b) of table 2.6. The impact response of GDP lies between -0.117 and -0.522 depending on the model specification. Again, comparable to expenditure shocks, the five-variable model deviates the most from the benchmark estimation. After a negative and statistically significant impact, the impulse response pattern turns slightly positive after seven periods. This might be explained by possible crowding-in effects through the tax and price channel, e.g. an increase in taxes reduces interest rates, which might have a positive effect on GDP.

Regarding the difference between the forecast and the counterfactual multipliers, the importance of the counterfactual analysis remains visible. In short samples, the effect of the tax multiplier goes in the opposite direction. But overall, the effect remains significant and economically sizeable in all specifications.

Figure 2.6: Difference between forecast and counterfactual multipliers with alternative model specifications



Note: The left columns displays the response of GDP to fiscal shocks derived from standard SVAR frameworks (Forecast multiplier (FC)). The right column visualizes the difference between the forecast and counterfactual multiplier (Δ(CF)).

2.A.5 Fiscal multipliers and SVAR models: some caveats

An advantage of SVAR models over large-scale macroeconometric models is the absence of extensive assumptions about the functioning of the economy. Yet, despite its simplicity, the SVAR approach also needs some assumptions that are crucial for the identification of policy shocks. To achieve identification, the standard Blanchard and Perotti (2002) approach uses decision lags on fiscal policy as well as exogenous information about the automatic elasticity of net taxes to GDP. From this, three potential problems arise: firstly, as pointed out by Perotti (2005), while institutional decision and implementation lags help to identify fiscal shocks with high frequency data, they also assume that changes in fiscal policy are typically decided and publicized well in advance of their implementation. Consequently, financial and other macroeconomic variables might react well ahead of the initial fiscal policy change due to the anticipation of the fiscal measure by economic agents. Standard VAR frameworks, however, are only capable of analyzing unexpected fiscal policy shocks.[58] Secondly, in order to exploit the decision lag, the SVAR approach relies on the use of high frequency data (quarterly, monthly, daily etc.), which is rather scarce and subject to many conditions in order to be used. In particular, the data series have to be consistent with allocating expenditures and revenues to certain quarters, e.g. in some cases taxes are collected with lags, whereas they are already documented in the statistics (accrual versus

[58] Blanchard and Perotti (2002) address this issue and include future values of the 1975Q2 tax cut dummy in the VAR specification as a robustness test. They find a significant increase of multipliers once the anticipations are included. Conversely, Perotti (2005) uses OECD forecasts in order to account for the anticipation problem and shows that they generally do not predict fiscal shocks. In an expanded study, Ramey (2011b) illustrates that the VAR shocks are missing the timing of the news. Complementary, Mertens and Ravn (2010) propose an extended SVAR approach that is applicable when fiscal shocks are anticipated. In addition, Leeper et al. (2013) demonstrate that fiscal studies that fail to model foresight will obtain biased estimates of tax multipliers.

cash accounting). As a consequence, combining data from various sources might lead to distortions as statistical offices often use different systems of accounting. In our analysis, we follow Blanchard and Perotti (2002) and use data from one source (Bureau of Economic Analysis; see section 2.A.1) that collects the data on an accrual basis.

Thirdly, the estimated multipliers depend very much on the calculated tax elasticities that have their limits in the following way: on the one hand, a diverse range of calculation procedures provides a fair number of reasonable elasticities. On the other hand, the SVAR approach assumes stability of the elasticity over the whole sample period.[59] In the latter case, the Lucas (1976) critique applies who argues that agents' behavior change in response to policy changes. Subsequently, changes in fiscal policy should also lead to changes in the tax elasticity. Certainly, the Lucas critique is also valid for our counterfactual analysis, where we assume that even in the long-run there will be no interaction among policy instruments.

[59] See Koester and Priesmeier (2012) for a broad discussion on the calculation of tax elasticities. Further, Caldara and Kamps (2012) show in a VAR framework that different priors on elasticities generate a large dispersion in multiplier estimates. See also section 3.A.4 in the appendix of the next chapter.

3 A counterfactual SVAR analysis of fiscal policy - Evidence for European countries

3.1 Introduction

Since the beginning of the 2007/8 financial crisis and the subsequent global economic downturn, Keynesian stabilization policies have been back on the agenda of many policy makers.[60] Consequently, there has been an increasing theoretical and empirical debate about the impact of fiscal policy measures on the economy.[61] This chapter seeks to contribute to the existing empirical literature by applying a five-variable Structural Vector Autoregression (SVAR) approach to a uniform data set for four European countries (Belgium, France, Germany, and the United Kingdom). Besides studying the effects of expenditure and tax

[60] The experiences of the financial crisis and the use made of fiscal policy are discussed by Arestis and Sawyer (2010)

[61] See Ramey (2011a) and Parker (2011) for an overview.

increases on output, we additionally analyze their dynamic effects on inflation and interest rates as well as the dynamic interaction of both policy instruments. By conducting counterfactual simulations, which abstract from the dynamic response of key macroeconomic variables to the initial fiscal shock, we study the importance of these channels for the transmission of fiscal policy on output.

While a broad consensus has formed on the effects of monetary policy, the implications of fiscal policy measures on the economy remain controversial. Theoretical and empirical studies have shown that fiscal stimuli trigger a series of direct and indirect effects, whereby the magnitude and direction depend on a number of country-specific characteristics.[62] Following Keynesian theory, the direction and magnitude of the fiscal multiplier crucially depends on the response of the interest rate and price level to the initial fiscal shock, e.g. rises in the interest rate and price level as a consequence of increased government spending tend to lower the fiscal multiplier.[63] Besides this crowding-out effect, the funding of the initial fiscal policy measure in the future (e.g. increase in taxes after an increase in spending) may also have a significant effect on the size of the multiplier, especially at longer horizons (see, e.g., Kuckuck and Westermann (2014); Uhlig (2010)).

The standard VAR methodology estimates the fiscal multiplier under the assumption that following the initial policy intervention, all variables will behave as they have typically done

[62] These country-specific aspects are for instance the degree of development, the exchange rate regime, the level of indebtedness (Ilzetzki et al. (2013); Corsetti et al. (2012)), the level of interest rates (Christiano et al. (2011)), the degree of openness (Favero et al. (2011)) as well as the stage of the business cycle (Auerbach and Gorodnichenko (2012)) to name the principle ones.

[63] The efficiency of fiscal policy measures also depends on the consumption behavior of households and the employment level of the economy. A theoretical discussion on the effects of fiscal policy and the Keynesian multiplier is provided by Blinder and Solow (1976).

in the past. In the counterfactual simulations, we assume that price levels, interest rates as well as net taxes (expenditures) are not affected by the initial spending (tax) shock. A comparison between the standard and counterfactual fiscal multipliers allows us to analyze the importance of the interest and price channels as well as the government's fiscal policy reaction for the transmission of fiscal policy on output.[64]

In order to identify the fiscal shocks we follow the methodology proposed by Blanchard and Perotti (2002). As this approach considers decision lags in fiscal policy making and institutional information about elasticities, it relies on the existence of reliable and non-interpolated quarterly data. Hence, our analysis benefits from the adoption of a common statistical standard in the EMU, namely the European System of Accounts (ESA95) that collects and classifies accrual fiscal data at quarterly frequencies. This uniform data set from 1991Q1 to 2011Q4 allows us to analyze the efficiency of fiscal policy in recent decades and to compare the effects between various countries.[65] In order to analyze the effects of fiscal policy, we adopt a five-variable VAR setup as in Perotti (2005) that includes the variables GDP, public expenditures, net taxes, inflation, and interest rate. In addition, we investigate the sensitivity of our findings in several robustness regressions including changes

[64]The discussion of the interaction between policy instruments and other economic variables dates back to the early contribution of Sims (1980). Ramey (1993) uses a similar approach by shutting down the policy-velocity channel in order to study the importance of the credit channel in the transmission of monetary policy. In the context of fiscal policy, the issue of dynamic interaction among fiscal policy instruments in VAR models has been initially raised in the working paper version of Blanchard and Perotti (2002) and in Perotti (2005). Kuckuck and Westermann (2014) elaborate on this point and discuss fiscal multipliers derived from standard VAR models regarding policy advice and DSGE modeling.

[65]Empirical studies analyzing fiscal multipliers of more than one country typically derive the data from various sources that do not follow a uniform classification system (see, e.g., Afonso and Sousa (2012, 2011); Perotti (2005)). This aggravates the direct comparison between the countries as fiscal variables might include different items across countries.

in identification assumptions and variable definitions.

Our findings show that the effects of fiscal shocks are limited and rather different across countries. Only expenditure shocks in France and Germany as well as tax shocks in Germany have a statistically significant but short-lived effect on output. A one Euro increase in government spending raises GDP by around 1.05 Euros in France and 1.24 Euros in Germany. In this respect, public expenditure increases tend to be more efficient than tax cuts. In Germany, a one Euro increase in net taxes drops GDP by around 0.58 Euros. Further, we find that fiscal multipliers do not always have the sign as predicted by economic theory, e.g. interest rates significantly decrease in response to an expenditure shock and increase in response to a tax shock across countries.

The results of the counterfactual exercise vary from country to country. On the one hand, the marginal impact of the price and interest channel is not economically significant in France and Germany. On the other hand, Belgium and the UK display significant differences between the counterfactual and base case impulse response functions. While shutting down the interest channel eliminates the impact of expenditure in Belgium, it slightly strengthens the effect on GDP in the UK. Regarding the dynamic interaction of policy instruments, the effects of the counterfactual analysis are rather limited and without economic importance.

The remainder of the chapter is organized as follows. Section 3.2 provides a brief overview of the related literature and presents a survey of the short-run and long-run effects of fiscal policy on key macroeconomic variables in European countries. Subsequently, section 3.3 outlines the data and analyzes the unit root and cointegration properties of the utilized variables. Section 3.4 presents the empirical framework, discusses the identification assumptions, and provides information on the calculation of the exogenous net tax elasticities. The contemporaneous and dynamic effects of expenditure and tax shocks are analyzed in section 3.5. The subsequent section 3.6 uses counterfactual simulations in order to study the importance of the price and interest rate channel for the transmission of fiscal policy. Finally, section 3.7 analyzes fiscal multipliers that abstract from the dynamic interaction of policy instruments. Concluding remarks are offered in section 3.8.

3.2 Related literature

The empirical literature on fiscal multipliers provides two main approaches in order to analyze the effects of fiscal policy on the economy: DSGE models and VAR studies. The former, DSGE methodology, attempts to explain the impact of fiscal policy measures on output on the basis of fitted macroeconomic models derived from microeconomic principles, while, the latter VAR-based studies, investigate the impact of fiscal policy on the basis of the evolution of the economy.[66] On these grounds, the empirical articles provide mixed and partially contradictory results. On the one hand, DSGE simulations, which are typically derived from real business cycle (see, e.g. Rebelo (2005); Uhlig (2010)) or New Keynesian models (see, e.g., Linnemann and Schabert (2003); Christiano et al. (2011)), differ substantially in its outcome due to different assumptions about the functioning of the economy.[67] On the other hand, VAR-based studies yield different results due to alternative

[66]See Parker (2011) who critically reviews and compares the two approaches.
[67]See Coenen et al. (2012a) and Hebous (2011) who provide an comprehensive overview.

identification strategies of fiscal policy shocks, different definitions of fiscal variables, and different sample sizes (see Ramey (2011b)).

As much of the research has focused on the United States (see, e.g., Blanchard and Perotti (2002), Leeper et al. (2010)), the availability of studies on European countries is rather limited and the outcomes are often ambiguous. As our subsequent analysis adopts a SVAR approach, table 3.1 provides a comprehensive and country-specific overview of the short-run and long-run effects of fiscal policy on key macroeconomic variables (output, inflation, and interest rate) derived from VAR models for different European countries.[68]

The majority of country-specific studies finds a positive linkage between public expenditures and economic activity as predicted by Keynesian theory. Regarding revenue shocks, the empirical literature yields inconsistent results. On the one hand some studies find that output responses statistically significant and negative to a (positive) tax shock, on the other hand some studies obtain positive effects on output.[69] Nevertheless, there is an agreement in the literature that spending shocks typically have a more pronounced effect on economic activity than tax shocks.

[68]Details on the quantitative size of the fiscal multipliers, the construction of fiscal variables as well as the exact empirical model can be found in the corresponding studies. Note that table 3.1 does not include any DSGE simulations as in Coenen et al. (2012b), Cwik and Wieland (2011), or Forni et al. (2009).

[69]Caldara and Kamps (2012) link this discussion to the chosen identification approach and the corresponding calibrated size of the automatic stabilizers. With regard to our analysis, we address this issue in section 3.A.4 in the appendix.

Table 3.1: Empirical evidence on the effects of fiscal policy in Europe

Author (Date)	Identification strategy	Countries (Period)	Policy instrument	Output		Inflation		Interest rate	
				≤2y	>2y	≤2y	>2y	≤2y	>2y
Afonso and Sousa (2012)	Recursive	DE(1980Q3-2006Q4)	EXP	(-)	(-)	(-)	(-)		
	(BVAR)		TAX	(+)	0	0	0		
		IT(1986Q2-2004Q4)	EXP	(+)	(-)	(+)	0		
			TAX	(-)	(+)	(-)	-		
		UK1964:Q2-2007Q4	EXP	(+)	(+)	0	+		
			TAX	(+)	(+)	(-)	(-)		
Afonso and Sousa (2011a)	BP approach (BVAR)	PT(1978Q1-2007Q4)	EXP	(-)	(0)	(+)	0		
			TAX	(-)	(-)	(+)	(+)		
Bergman and Hutchison (2010)	BP approach (SVAR)	DK(1971Q1-2004Q4)	EXP	(+)	(+)				
			TAX	(-)	(-)				
Burriel et al. (2010)	BP approach	EMU(1981Q1-2007Q4)	EXP	(+)	-	(+)	0	(+)	0
	(Panel SVAR)		TAX	(-)	0	(-)	-	(-)	0
Monacelli and Perotti (2010)	BP approach (SVAR)	UK(1980Q1-2006Q4)	EXP	(+)	+				
Tenhofen et al. (2010)	BP approach	DE(1974Q1-2008Q4)	EXP	(+)	-	(+)	+	+	0
	(SVAR)		TAX	(0)	0	0	0	+	+
Beetsma et al. (2008)	Recursive (Panel VAR)	EU-14(1970-2004)	EXP	(+)	0				
Perotti (2008)	BP approach (SVAR)	UK(1963Q1-2006Q2)	EXP	+	+				
Giordano et al. (2007)	BP approach	IT(1982Q1-2004Q4)	EXP	(+)	0	(+)	0	(±)	0
	(SVAR)		TAX	(+)	+	(±)	0	(+)	0
Badinger (2006)	BP approach	AT(1983Q1-2002Q4)	EXP	(+)	(+)				
	(SVAR)		TAX	(-)	(-)				
De Castro (2006)	BP approach	ES(1980Q1-2001Q2)	EXP	(+)	(-)	(±)	(+)	(+)	(+)
	(SVAR)		TAX	(+)	(-)	(-)	0	(+)	(+)
Marcellino (2006)	Recursive	DE(1981S1-2001S2)	EXP	(+)	0	-	(-)	-	(-)
	(SVAR)		TAX	-	0	0	0	0	0
		ES(1981S1-2001S2)	EXP	0	0	0	-	+	0
			TAX	+	0	-	0	(+)	+
		FR(1981S1-2001S2)	EXP	0	-	(+)	-	(+)	(-)
			TAX	0	0	+	0	(-)	0
		IT(1981S1-2001S2)	EXP	+	0	-	0	0	+
			TAX	(+)	-	(+)	-	+	-
Biau and Girard (2005)	BP approach	FR(1978Q1-2003Q4)	EXP	(+)	0	(+)	(+)	(+)	(+)
	(SVAR)		TAX	(0)	0	0	0	(-)	0
Perotti (2005)	BP approach	DE(1960Q1-1974Q4)	EXP	(+)	(+)	(+)	(-)	(+)	0
	(SVAR)		TAX	(+)	0	0	0	0	(+)
		DE(1975Q1-1989Q4)	EXP	(+)	(+)	(-)	(+)	(+)	(+)
			TAX	0	+	(-)	(-)	(-)	(-)
		UK(1960Q1-1974Q4)	EXP	(+)	(+)	(+)	(+)	(+)	(+)
			TAX	(-)	(-)	+	(-)	0	(-)
		UK(1975Q1-1989Q4)	EXP	(+)	(+)	0	(-)	0	(+)
			TAX	(+)	(+)	(-)	(+)	-	(-)
Plötscher et al. (2005)	BP approach	DE(?)	EXP	(+)	0				
	(SVAR)		TAX	(-)	-				

Note: The table reviews published VAR-based studies on expenditure (EXP) and tax (TAX) shocks in European countries. The column ≤2y displays the average effect of fiscal policy in less or equal 2 years and the column >2y shows the average effect of fiscal policy in more than two years. The symbol + indicates a positive accumulated response, the symbol − denotes a negative accumulated response, the sign 0 illustrates no response (economically insignificant) and ± displays an ambiguous accumulated response of the corresponding variable. The parenthesis indicate that the reactions are statistically different from zero. All entries represent the benchmark findings of the corresponding studies.

The response of interest and inflation rates to fiscal stimulus has long been a matter of controversy in the literature (see, e.g., Eaton (1981); Evans (1985); Plosser (1987)). While most standard text book models agree that an increase in spending should be associated with a rise in interest rates, the direction and the magnitude of the response of inflation depends very much on the type of fiscal stimulus and the corresponding crowding-out effects. Table 3.1 displays that previous VAR investigations on the effects of fiscal policy on inflation and interest rates provide rather mixed results. Although, expenditure shocks have a significant and positive effect on the price and the interest rate level in the majority of the studies, the findings for tax shocks are rather ambiguous.[70]

Overall, the empirical SVAR findings on fiscal policy measures do not present a homogeneous picture. This has been attributed to differences in identification strategies, investigated time periods, and various country-specific characteristics. Favero et al. (2011) argue that every country has many fiscal multipliers that depend on the debt dynamics, the degree of openness, and the fiscal reaction function of each country. Further, Ilzetzki et al. (2013) point out that the size of the multiplier depends on the state of development, the exchange rate regime, and the level of indebtedness. Additionally, Auerbach and Gorodnichenko (2012) show that fiscal multipliers differ in different stages of the business cycle and are substantially larger in recessions.

[70] Perotti (2005) and Caldara and Kamps (2012) show that the price elasticity of taxes is a crucial parameter in estimating the effects of taxes on prices and can have a significant impact on the direction of the effect.

3.3 Data description and preliminary analysis

Data sources and definitions

To analyze the effects of fiscal policy, we propose a five-variable setup including the macroeconomic variables in terms of output, inflation rate, interest rate, government spending, and government net taxes for the four Western European countries Belgium, France, Germany, and the United Kingdom. All data is collected from the quarterly non-financial accounts for general government (govq_ggnfa) as well as the quarterly national accounts (namqgdp) collected by the European Commission (see Eurostat (2013b)) ranging from 1991Q1 to 2011Q4 that are fully consistent with the accrual accounting concept of ESA95. Thus, all data is based on direct information available from basic sources and allows for a direct comparison between the countries. Whenever necessary, the time series were seasonally adjusted by applying the X12-ARIMA procedure.[71]

Following Blanchard and Perotti (2002), we define government spending as the sum of final consumption expenditure and gross fixed capital formation. Net taxes are calculated as the sum of net tax receipts including contributions for social security less net property income, net current transfers, and net subsidies.[72] Output is captured by real GDP, the consumer price index (CPI) is used as an indicator for the price level, and the 3-month London In-

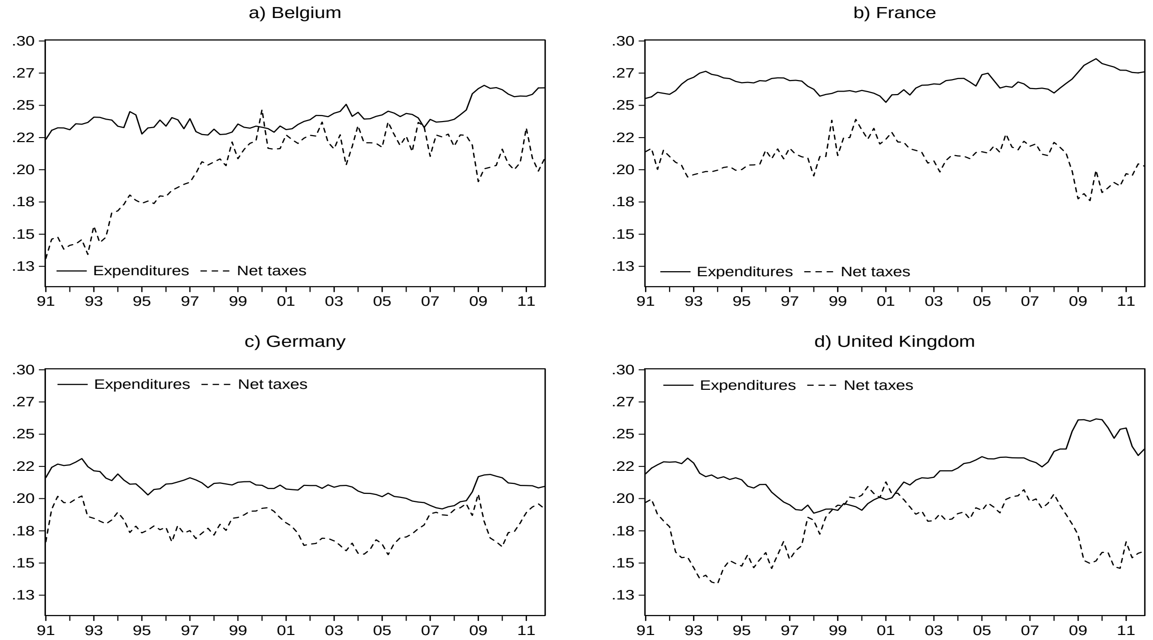
[71]See data sources and definitions in section 3.A.1 in the appendix.

[72]In our base case scenario, net taxes are calculated without capital transfers that are often influenced by economic restructuring and reallocations associated with debt rescheduling from the private to the public sector. In order to obtain results that are not affected by these reschedulings, we excluded this category from our baseline regression. In 2011, capital transfers account for our set of countries only around 5 to 8 percentage of total expenditures. Nevertheless, in section 3.A.5 in the appendix, we estimate the effects of fiscal policy with alternative net tax definitions (also including capital transfers) and provide evidence that adding this item does not qualitatively affect the empirical outcome.

terbank offered Rate (LIBOR) is used as a measure for the interest rate. More information on the data sources and data calculation can be found in the appendix in section 3.A.1.

Figure 3.1 displays the ratio of government expenditure and of net taxes to GDP over the

Figure 3.1: Government expenditure and net revenue as a percentage of GDP

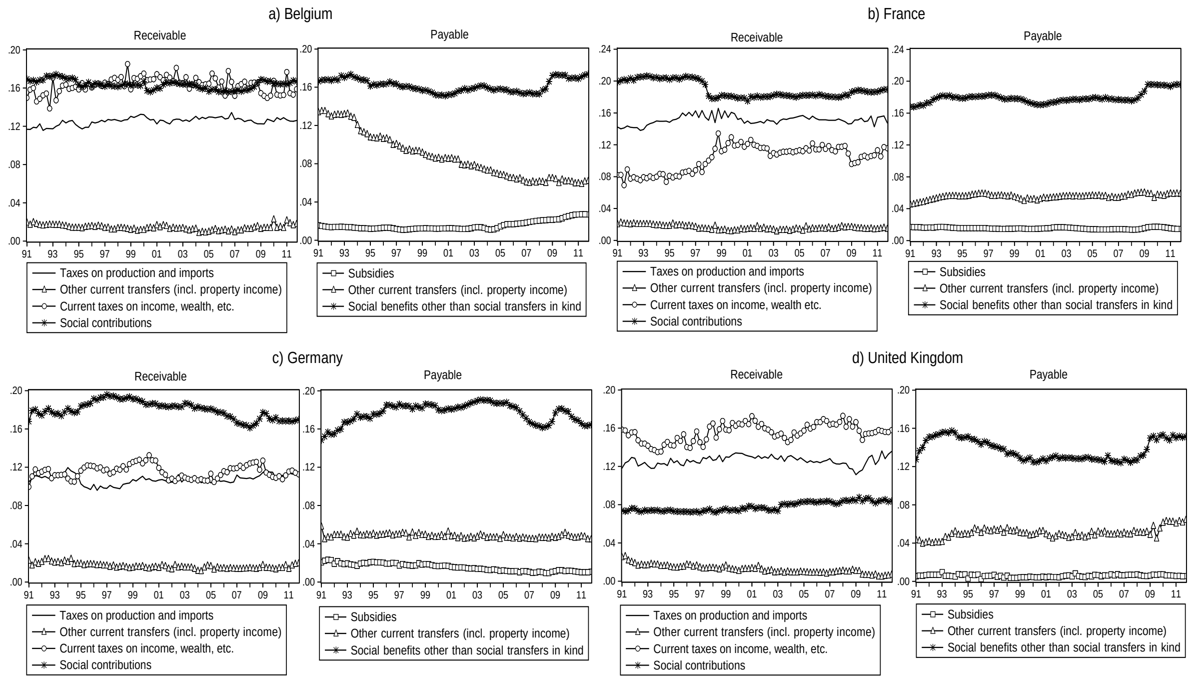


Note: The graph displays government expenditures and net taxes as a percentage of GDP for a) Belgium, b) France, c) Germany, and d) the United Kingdom. See section 3.A.1 in the appendix for data definitions and sources.

sample period from 1991Q1 to 2011Q4 for the countries Belgium, France, Germany, and the United Kingdom. At the end of 2011, government spending at a percentage of GDP ranges around 26.4% in Belgium, 27.6% in France, 20.9% in Germany, and 23.8% in the UK . Net taxes have been slightly lower taking a value of 19.8% of GDP in Belgium and France, 18.5% of GDP in Germany and 15.7% of GDP in the UK. In fact, there has been a steady gap between expenditures and net taxes throughout the whole sample period, however, especially since the beginning of the financial crisis in 2007, all countries reveal that this gap has widened. As we analyze government expenditure excluding transfer payments as well taxes net of transfer payments, it is not surprising that both series do not follow a trend pattern. Thus, the values for France, Germany, and the UK remain - with the exception of some periods - at a stable niveau throughout the whole sample period. Only the net taxes series of Belgium steadily increases from 1991 to 1999 due to decreasing expenditures on property income.

To gain a better insight into the composition of net taxes, figure 3.2 visualizes the devel-

Figure 3.2: Composition of net tax revenues as a percentage of GDP



Note: The graph displays the composition of net taxes as a percentage of GDP for a) Belgium, b) France, c) Germany, and d) the United Kingdom. See section 3.A.1 in the appendix for data definitions and sources.

opment and the proportion of the various net tax components as a percentage to GDP. A closer look into the graphs suggest that the receipts as well expenditures of the net taxes remain - with few exceptions - at a stable niveau throughout the sample period. Furthermore, it becomes apparent that in Belgium, France, and Germany social contributions comprise the largest part of net taxes ranging around 16 to 20 percentage of GDP on the revenue and expenditure side. In the United Kingdom, however, the proportion of social contributions is much lower as citizens are responsible to make private provisions for their own welfare. Further, it can be seen that expenditures on social benefits have increased as a percentage of GDP in the wake of the financial crisis in 2007/2008 in all countries.

Overall, public expenditures and net taxes exhibit a stable development throughout the last two decades. Thus, subsample stability should not be an important concern in the following analysis. Nevertheless, the 2007/8 financial crisis has brought expansionary fiscal policy back on the agenda as expenditure goes up and net taxes go down.[73]

Unit root and cointegration analysis

As we estimate fiscal multipliers by using a VAR approach, it is important to gain some insights into the properties of the five variables implemented in our benchmark model. All variables except interest rates are expressed in logarithms, whereas all variables except for prices and interest rates are expressed in real as well as per capita terms. In table 3.2, we report the unit root test statistics as well as cointegration results for government expenditures (EXP), net taxes (TAX), gross domestic product (GDP), consumer price index (INF), and interest rates (INT). Applying the Phillips-Perron (PP) test, we find that all the variables may be treated as integrated of order one. Furthermore, the test statistics of the Johansen and Juselius (1990) cointegration test indicate that the five

[73]In order to abstract from this exceptional period, we conduct the later analysis also in a reduced sample that stops in 2006Q4. This has no effect on the subsequent results.

variables utilized in our analysis are not cointegrated. For all countries up to a maximum of three cointegration equations could be found. In the following, we therefore estimate the dynamic interactions between the variables in a VAR in first differences.

Table 3.2: Unit root and cointegration results

	Belgium		France		Germany		United Kingdom	
Unit root results								
	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
EXP	-1.778	-11.893***	-2.749*	-7.789***	-0.497	-9.636***	-1.149	-6.618***
TAX	-1.810	-13.655***	-1.702	-12.845***	-1.055	-11.496***	-1.405	-8.426***
GDP	-1.540	-6.573***	-2.391	-5.992***	-0.892	-7.082***	-1.393	-5.908***
INF	-0.744	-5.836***	0.298	-6.678***	-0.689	-7.076***	-0.511	-10.328***
INT	-2.393	-9.926***	-2.080	-7.439***	-2.147	-5.978***	-2.059	-6.768***
Cointegration results								
	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
r=0	31.442*	37.591**	28.261	28.752	42.899***	49.309***	30.895	38.231**
r≤1	25.731*	30.974*	24.483	25.258	35.736***	40.572***	28.567**	30.862*
r≤2	17.389	22.091	21.075**	21.359	12.546	16.988	9.070	23.171*
r≤3	9.556	11.638	14.249**	15.358	5.291	9.023	7.515	7.614
r≤4	0.039	9.108	0.893	3.522	0.044	3.302	0.011	6.061

Note: *Unit root results*: The Phillips-Perron (PP) test for the levels are calculated including a constant and a trend in the test equation. The test statistics for the first differences are calculated including a constant. The bandwidth for the PP test is selected based on Newey-West using Bartlett kernel. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from MacKinnon (1996). *Cointegration results*: Model (1) allows for a constant in the cointegration space and for a linear trend in the level data. Model (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

3.4 Empirical framework

VAR specification and estimation

The VAR methodology distinguishes between three different approaches to estimate fiscal multipliers. All methods provide an alternative way on how to address the identification problem associated with the possible directions of causation between government spending and output. The narrative approach first introduced by Ramey and Shapiro (1998) identifies large episodes of military buildups and uses them to disentangle the exogenous component of fiscal policy changes.[74] The SVAR approach by Blanchard and Perotti (2002) uses exogenous institutional information about elasticities in order to identify fiscal shocks from the system. Finally, the sign restriction approach by Mountford and Uhlig (2009) builts on long-run identifying restrictions that imposes a positive reaction of the impulse response of the appropriate fiscal variable.

As the narrative approach is not practical for cross-country comparison and appears less suitable for limited sample sizes and while the sign restriction approach rules out by assumption a whole set of "non-keynesian" output responses to fiscal shocks, this chapter focuses on the SVAR approach as in Blanchard and Perotti (2002).[75]

Our analysis starts with the following reduced form VAR specification

$$X_t = \theta(L)X_{t-1} + U_t, \tag{3.1}$$

where $X_t \equiv [y_t, p_t, i_t, g_t, r_t]$ is a five-dimensional vector of first differences of real per capita GDP (y), GDP deflator (p), interest rate (i), real per capita government expenditure (g) and real per capita net taxes (r). θ(L) denotes a lag polynomial matrix and $U_T \equiv [u_t^y, u_t^p, u_t^i, u_t^g, u_t^r]$ is the corresponding vector of reduced form innovations. Our

[74] See Edelberg et al. (1999) as well as Burnside et al. (2004) for different versions of this approach.

[75] A general discussion on the different methodologies can be found in Perotti (2005, 2008) as well as in Fontana (2009). In section 3.A.6 in the appendix, we estimate the effects of fiscal policy measures on output by applying a recursive identification strategy (see Fatás and Mihov (2001)) and finally compare the results to the Blanchard and Perotti (2002) identification scheme.

benchmark estimations also include a drift parameter, which we omit from the notation for convenience.[76]

In order to track the dynamic response of GDP to a unit increase of public expenditure, we compute accumulated impulse responses with analytically calculated standard errors using the structural moving average representation of equation (3.1).[77] Afterward, the accumulated impulse responses are converted in order to measure the output reaction caused by a one unit increase in the fiscal variable. To identify expenditure and tax shocks in the system and to obtain interpretative impulse response functions, we use restrictions imposed by exogenous information and economic theory.

Identification of fiscal shocks

Following Amisano and Giannini (1997) the relation between the reduced form innovations U_t and the structural shocks E_t can be described by the linear relationship $A \cdot U_t = B \cdot E_t$. As we are interested in estimating the structural shocks u_t^r and u_t^g and the response of GDP, we achieve identification of the model by exploiting the existence of decision lags in fiscal policy and institutional information about the automatic elasticity of fiscal variables to real GDP and price levels (see Blanchard and Perotti (2002)).

Following Perotti (2005) the reduced form residuals of the fiscal variables in equation (3.1) can be written as a linear combination of three types of shocks: i) the automatic response

[76]The basic lag length of the VAR is determined by using Schwarz information criterion. In the presence of autocorrelation the lag length of the VAR is successively enhanced to remove all serial correlation from the data. This leads in our benchmark regressions to a lag length of 2 for all countries.

[77]These cumulative multipliers are defined at some horizon N as $\frac{\sum_{t=0}^{N} \Delta y_t}{\sum_{t=0}^{N} \Delta g_t}$.

of taxes and government spending to innovations in output, prices and interest rates; ii) the systematic discretionary response of policymakers to output, price, and interest rate innovations, and iii) random discretionary shocks to fiscal policies. This leads to the following formal representation:

$$u_t^g = a_y^g u_t^y + a_p^g u_t^p + a_i^g u_t^i + b_r^g e_t^r + e_t^g \tag{3.2}$$

$$u_t^r = a_y^r u_t^y + a_p^r u_t^p + a_i^r u_t^i + b_g^r e_t^g + e_t^r, \tag{3.3}$$

where e_t^g and e_t^r are the structural shocks of government expenditure and net taxes, respectively. Thus, in the present case of a five-variable VAR, the A and B matrices can be written as follows:

$$\begin{bmatrix} 1 & -a_p^y & -a_i^y & -a_g^y & -a_r^y \\ -a_y^p & 1 & -a_i^p & -a_g^p & -a_r^p \\ -a_y^i & -a_p^i & 1 & -a_g^i & -a_r^i \\ -a_y^g & -a_p^g & -a_i^g & 1 & 0 \\ -a_y^r & -a_p^r & -a_i^r & 0 & 1 \end{bmatrix} \times \begin{bmatrix} u_t^y \\ u_t^p \\ u_t^i \\ u_t^g \\ u_t^r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & b_r^g \\ 0 & 0 & 0 & b_g^r & 1 \end{bmatrix} \times \begin{bmatrix} e_t^y \\ e_t^p \\ e_t^i \\ e_t^g \\ e_t^r \end{bmatrix}. \tag{3.4}$$

Since the reduced form innovations U_t and the structural shocks E_t are correlated, we use exogenous elasticities and ordering assumptions in order to identify the structural shocks E_t from the reduced VAR residuals U_t.[78]

[78]To identify the structural model from an estimated VAR, it is necessary to impose - in the case of a five-variable VAR model - ten restrictions ($(n^2 - n)/2$) on the structural model in order to identify the the 25 (n^2) unknowns from the known 15 independent elements ($(n^2 + n)/2$) of the variance/covariance matrix of the regression residuals. Altogether, we estimate 9 parameters of the matrix A and 6 elements (including the 5 diagonal elements) of the matrix B. The remaining parameters are either set to zero or one, or equal to the respective exogenous elasticities. In the present analysis we adopt the identification strategy proposed by Tenhofen et al. (2010).

The identification procedure begins by constructing the parameters $-a_y^g$ and $-a_y^r$ that show the contemporaneous response of government direct expenditure and net revenue to changes in GDP, respectively. The use of quarterly data allows us to set $-a_y^g$ equal to zero as the government is not able to adjust its expenditures to changes in GDP within the same quarter due to decision and implementation lags. The direct within-quarter elasticity of net taxes to GDP ($-a_y^r$) is calculated with the help of exogenous information about the tax and transfer system.[79] Based on the assumption that nominal prices do not influence real GDP ($a_p^y = 0$), the net revenue as well as the public spending elasticity to inflation ($-a_p^r$ and $-a_p^g$) is simply the real GDP elasticity of the fiscal variable minus 1 (see Tenhofen et al. (2010)). Further, we assume that changes in GDP, price level, public spending, and net taxes do not respond to unexpected movements in the interest rate within the same quarter ($a_i^y = a_i^p = a_i^r = a_i^g = 0$).

In the last step, it is necessary to make an assumption concerning the ordering of the fiscal decisions. In the benchmark case, we assume that spending decisions come first, setting b_g^r equal to zero. As later shown in section 3.5, we find that the correlations between spending and taxes are very low for every country, yielding relatively low estimated values of b_g^r. As a consequence, the ordering of spending and taxes makes only very little difference to the subsequent analysis.

[79]Those exogenous net tax elasticities will be discussed in more detail in the next subsection.

Net tax elasticities

In order to compute the exogenous tax elasticities, we follow the same procedure as in Perotti (2005). We consider four main categories of taxes for every country: indirect taxes, personal income taxes, corporate income taxes, and social security taxes. The net tax elasticity to GDP of category i is calculated as the product of the tax elasticity to its own tax base and the elasticity of the tax base to output. The overall net tax elasticity is then calculated by the sum of every single tax elasticity weighted by the share of each tax component in the sum of all tax revenues. Information about the tax elasticity to its own base are taken in some parts from Girouard and André (2005), who account for individual tax code information and income distributions. Missing items are calculated with the help of simple linear regressions. A detailed description about the calculation of the net tax elasticities and the used data can be found in section 3.A.2 in the appendix.

In our benchmark model, we use net tax elasticities that are calculated including transfer payments and we assume for every country that corporate income taxes are not collected within the same quarter. In the base case scenario, we measure a net tax elasticity of 0.97 for Belgium, 0.53 for France, 0.95 for Germany, and 0.42 for the United Kingdom.[80] In section 3.A.4 in the appendix, we investigate the robustness of our results in a couple of sensitivity tests considering different specifications of tax elasticities.

[80] These values are comparable to those that have been reported in the literature. Bode et al. (2009) calculates for Germany in a comparable sample size a net tax elasticity of 0.46. Although for a different time horizon Perotti (2005) estimates a net tax elasticity of 0.76 for the United Kingdom and Biau and Girard (2005) a net tax elasticity of 0.8 for France.

3.5 The effects of fiscal shocks

Contemporaneous effects

We start our analysis by investigating the estimated coefficients of the contemporaneous relations between fiscal shocks (expenditure as well as net taxes) and output, inflation as well as interest rates (see table 3.3). In order to make a statement about the structural correlation between taxes and spending, the table also displays the estimated contemporaneous effects of the policy instruments among each other. For convenience of interpretation, all estimates are expressed in terms of elasticities, i.e. a one percentage change in one variable leads to a one percentage change in another.

As far as a spending shock is concerned, table 3.3 reveals that an increase in public expenditures has a contemporaneous positive effect on output in all countries. Eventually this response is only statistically significant in France and Germany. Inflation also reacts positive to a government spending shock in France, Germany, and the UK, which is especially pronounced in the latter. In contrast, the response in Belgium is slightly negative and nearly zero. Furthermore, it can be seen that interest rates respond negative to spending in all countries. Nevertheless, this effect is only statistically significant in Germany.

In reverse, the signs of the contemporaneous effects of net taxes show some disparities across countries. Surprisingly, output reacts positive to a tax shock in Belgium, France, and the UK. Only Germany exhibits a negative and statistically significant effect as predicted by economic theory. In addition, it can be seen that the interest rate responds positive to a tax shock in Belgium, Germany, and the UK. The contemporaneous effects on inflation

Table 3.3: Estimated contemporaneous coefficients

	Spending shock			**Tax shock**			**Fiscal ordering**	
	a_g^y	a_g^p	a_g^i	a_r^y	a_r^p	a_r^i	b_r^g	b_g^r
Belgium	0.083	-0.005	-0.947	0.013	0.009	0.107	-0.002	-0.007
	(0.224)	(0.836)	(0.292)	(0.543)	(0.249)	(0.705)	(0.179)	(0.179)
France	0.277	0.046	-1.709	0.039	0.027	-0.049	0.897	0.897
	(0.001)	(0.153)	(0.189)	(0.056)	(0.002)	(0.896)	(0.369)	(0.369)
Germany	0.442	0.139	-2.674	-0.117	-0.011	0.334	0.489	(0.512)
	(0.002)	(0.042)	(0.024)	(0.019)	(0.616)	(0.407)	(0.143)	(0.147)
United Kingdom	0.173	3.840	-0.039	0.013	-0.103	0.289	0.597	0.597
	(0.202)	(0.007)	(0.970)	(0.519)	(0.551)	(0.377)	(0.551)	(0.551)

Note: The table displays the estimated contemporaneous effects of a spending and tax shock to output ($a_{g,t}^y$), inflation ($a_{g,t}^p$) and interest rate ($a_{g,t}^i$). Corresponding p-values are shown in parentheses. All effects are expressed as elasticities.

are positive in Belgium and France and negative in Germany and the United Kingdom. Overall, Table 3.3 yields three conclusions. First, the signs of the contemporaneous effects are not always those one would expect from Keynesian theory. On the one side government spending lowers interest rates, on the other side a positive tax shock tends to raise interest rates and output. Nevertheless, the lack of statistical and economical significance casts serious doubts about the efficiency of fiscal policy measures, which certainly need to be explored more thoroughly in the next section. Secondly, the contemporaneous effects of fiscal policy are rather different across countries. While the direction of responses to spending shocks are similar across countries, although with some differences in the magnitude of the effects, the directions of the effects of tax shocks clearly differ in the individual countries. As shown by Caldara and Kamps (2012) the effects depend very much on the inserted tax and price elasticities and raises the issue of robustness.[81] In section 3.A.4 in the appendix, we discuss the results in a couple of sensitivity tests, where we account for different speci-

[81] The significance of the inserted contemporaneous elasticities is also discussed by Blanchard and Perotti (2002) as well as Perotti (2005).

fications of tax and price elasticities. The third conclusion is that under either of the two alternative identification assumptions, the estimated values of b_r^g and b_g^r are relatively low and statistically insignificant. This implies that the ordering of taxes and spending makes little difference to the subsequent results. In the further analysis we assume that spending is ordered first.

Dynamic multipliers

In the next step, we analyze the dynamic effects of fiscal policy on output, inflation, and interest rates as well as the dynamic interaction of the policy instruments. For convenience of interpretation and comparability across countries, we transform the original impulse responses into dynamic multipliers that display a X Euros response in one variable to a one Euro increase in one of the fiscal variables. Figure 3.3 shows the point estimates (solid line) and two standard error bands (dotted line) of a shock in spending (top row) and taxes (bottom row) on output, inflation, interest rate as well as the respective fiscal counterpart.[82]

Starting with the effects of spending shocks, it can be seen that in all countries economic growth is positively stimulated by an increase in public expenditures. Nevertheless, this effect is only statistically significant in the first two quarters for France and Germany. On impact, a one Euro increase in public spending leads to a 1.05 Euros increase in GDP in France as well as to a 1.24 Euros increase in GDP in Germany. Furthermore, spending

[82] Section 3.A.3 in the appendix provides a complete picture of the SVAR estimations and briefly discusses the effects of price, interest rate, and output shocks on the economy.

has a positive and statistically significant effect on inflation in Germany and the United Kingdom.[83] The long-run response of the interest rate to a fiscal shock is negative in all countries. A statistically significant effect, however, was only found in Belgium in the third quarter after the initial shock.[84]

Finally, the reaction of net taxes to an increase in government spending is shown in last column of figure 3.3. For France and Germany, we find a very limited but positive response of taxes to a change in expenditure. Part of this reaction is certainly due to automatic stabilizers being at work. An increase in expenditures increases GDP, which leads to higher net taxes. The other part, however, might be a discretionary response of government, the need to finance additional expenditures. In the United Kingdom, however, we find that taxes significantly decrease in response to a spending shock, which reflects an expansionary fiscal policy path.

[83] In our estimations, we use a spending elasticity to inflation of -1. Perotti (2005) shows that the inserted spending elasticity is a crucial assumption in estimating the effects of public spending on prices (see also section 3.A.4 and 3.A.6 in the appendix).

[84] The response of interest rates to fiscal policy has been broadly discussed in the literature (see, e.g., Evans (1985); Plosser (1987)). The standard Keynesian text book model would assume that an increase in spending is associated with an increase in interest rates. The empirical SVAR literature provides rather mixed results on this issue (see section 3.2). An extensive discussion on the relationship between fiscal policy and interest rates is provided by Faini (2006).

Figure 3.3: Response of macroeconomic variables to fiscal shocks

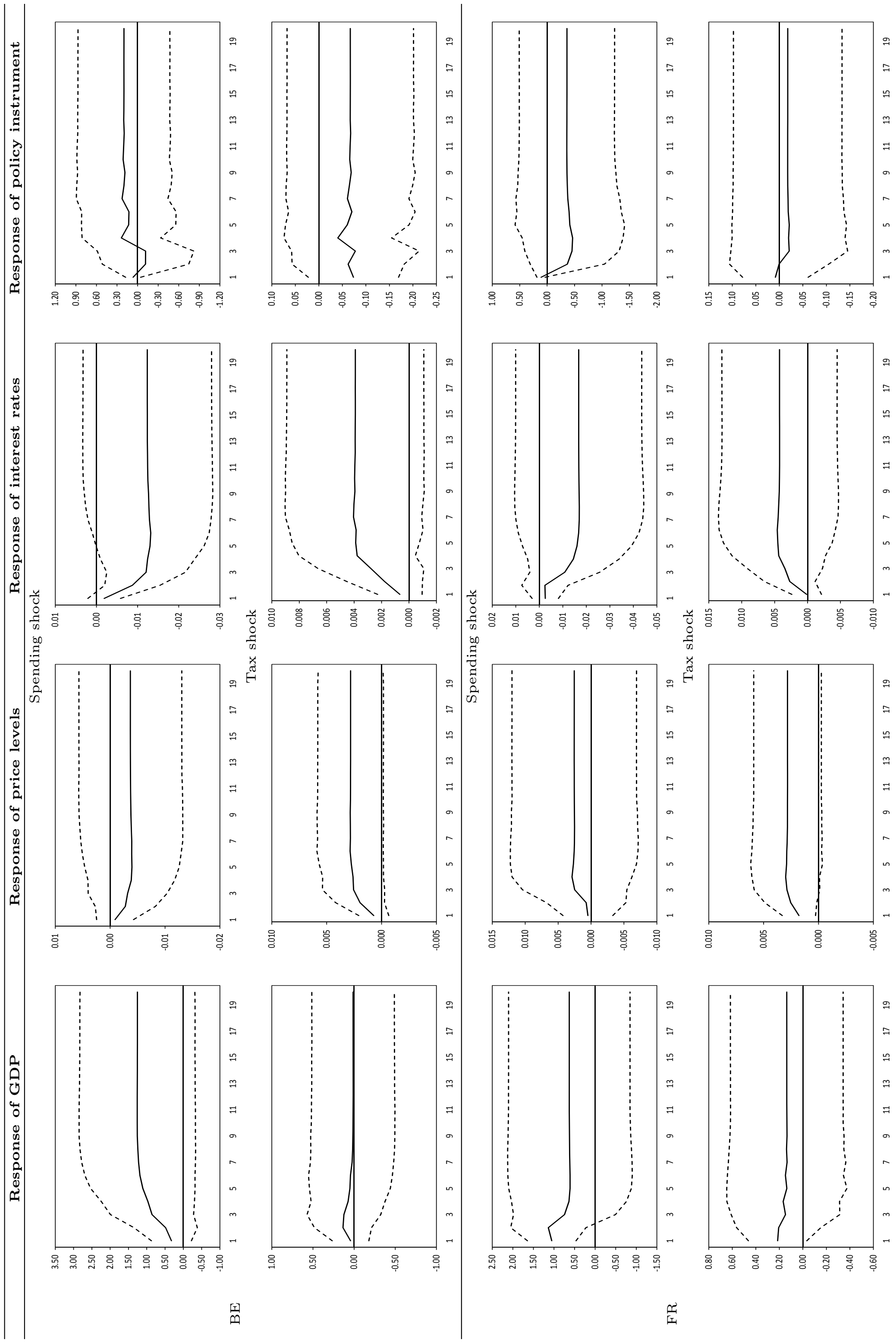
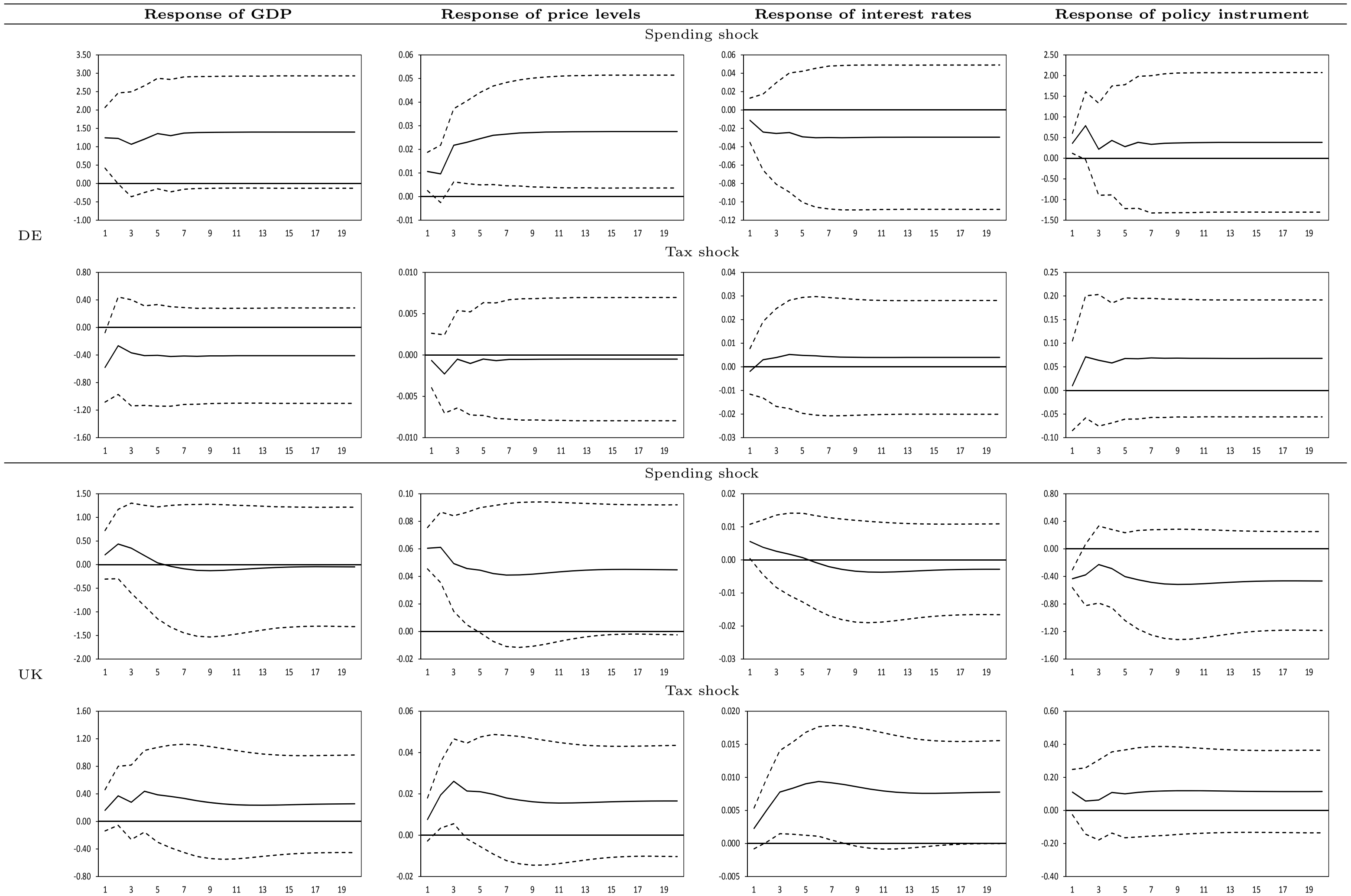


Figure is continued on the next page.

Figure 3.3: Response of macroeconomic variables to fiscal shocks - continued



Note: The figure displays the response of macroeconomic variables to a one unit increase in public spending and net taxes in Belgium (BE), France (FR), Germany (DE), and the United Kingdom (UK). The figure contains the responses (solid line) and a confidence interval (dashed lines) representing two times the standard deviation.

The effects of an increase in net taxes is visualized in the bottom row of each country segment. The point estimates of Belgium, France, and the United Kingdom display a slightly positive but statistically insignificant response of GDP.[85] Only in Germany a positive tax shock has a negative and statistically significant effect on output amounting to -0.58 Euros on impact.

Further, it can be seen that the consequences of tax shocks on inflation are generally very limited but tend to have a positive impact, which is statistically significant in France (first quarter) and the United Kingdom (third quarter). The response of the interest rate to a tax shock has the exact opposite sign in comparison to the expenditure shocks. In all countries, the inflation rate increases subsequent to a tax shock, which is even statistically significant in the United Kingdom. Finally, the responses of expenditures to a tax increase are close to zero and statistically insignificant in all countries.

Overall, figure 3.3 yields three main conclusions. First, the impulse response functions show that the effects of fiscal policy on output are very limited. Only expenditure shocks in France and Germany as well as tax shocks in Germany have a statistically significant effect on output. In this respect, public expenditure increases tend to be more efficient than tax cuts.[86] Secondly, the impulse responses do not always have the sign one would

[85] It is not uncommon for country-specific SVAR studies on fiscal multipliers to find limited but positive and statistically significant effects of tax increases on GDP (see, e.g, Perotti (2005); Marcellino (2006); Giordano et al. (2007)). This positive response could be associated with the improvement in government deficits and fiscal solvency, respectively. Additionally, Marcellino (2006) points out that since a tax shock is actually a revenue shock, it can be due either to an increase in the tax rate or to an increase in the tax base, and the latter is positively correlated with the output gap. Further, Caldara and Kamps (2012) show that the response of GDP to tax shocks crucially depends on the automatic stabilizers. Section 3.A.4 in the appendix provides multipliers with different calibrated elasticities. In Belgium, France, and the UK, the response of GDP to tax shocks generally stays insignificant and positive irrespective of the inserted net tax elasticity.

[86] The literature strongly disagrees whether tax cuts or spending increases are more expansionary. While

expect from economic theory, e.g. interest rates significantly decrease in response to an expenditure shock and increase in response to a tax shock. Thirdly, the effects of fiscal policy are rather different across countries, likely reflecting the different institutional frameworks. This becomes particular evident with regard to tax shocks.

The derived multipliers in figure 3.3 include the dynamic and indirect interactions of the macroeconomic variables. The response of output to a spending (tax) shock might be either explained by the direct effect of expenditures (net taxes) on output (right shift of the IS-curve) or by an indirect effect through the inflation and interest rate channel. In order to analyze the importance of these channels for the transmission of fiscal policy, we conduct in the following section some counterfactual simulations, which abstract from the dynamic response of some macroeconomic variables to the initial fiscal shock.

3.6 The importance of the price and interest rate channel for the transmission of fiscal policy

In this section, the following question is asked: how would the response of output to a fiscal shock differ if the fiscal variables do not have a direct effect on the price level and the inflation rate. To answer this question, in a first step, we estimate the VAR model setting the contemporaneous effects of the respective variables equal to zero. Subsequently, when computing the impulse response functions, we shut down the channel that captures

cross-sectional studies (see, e.g., Alesina and Ardagna (2010)) provide empirical evidence that tax cuts are more likely to stimulate output, VAR-based studies (see, e.g., Marcellino (2006)) generally find public spending increases to be more efficient.

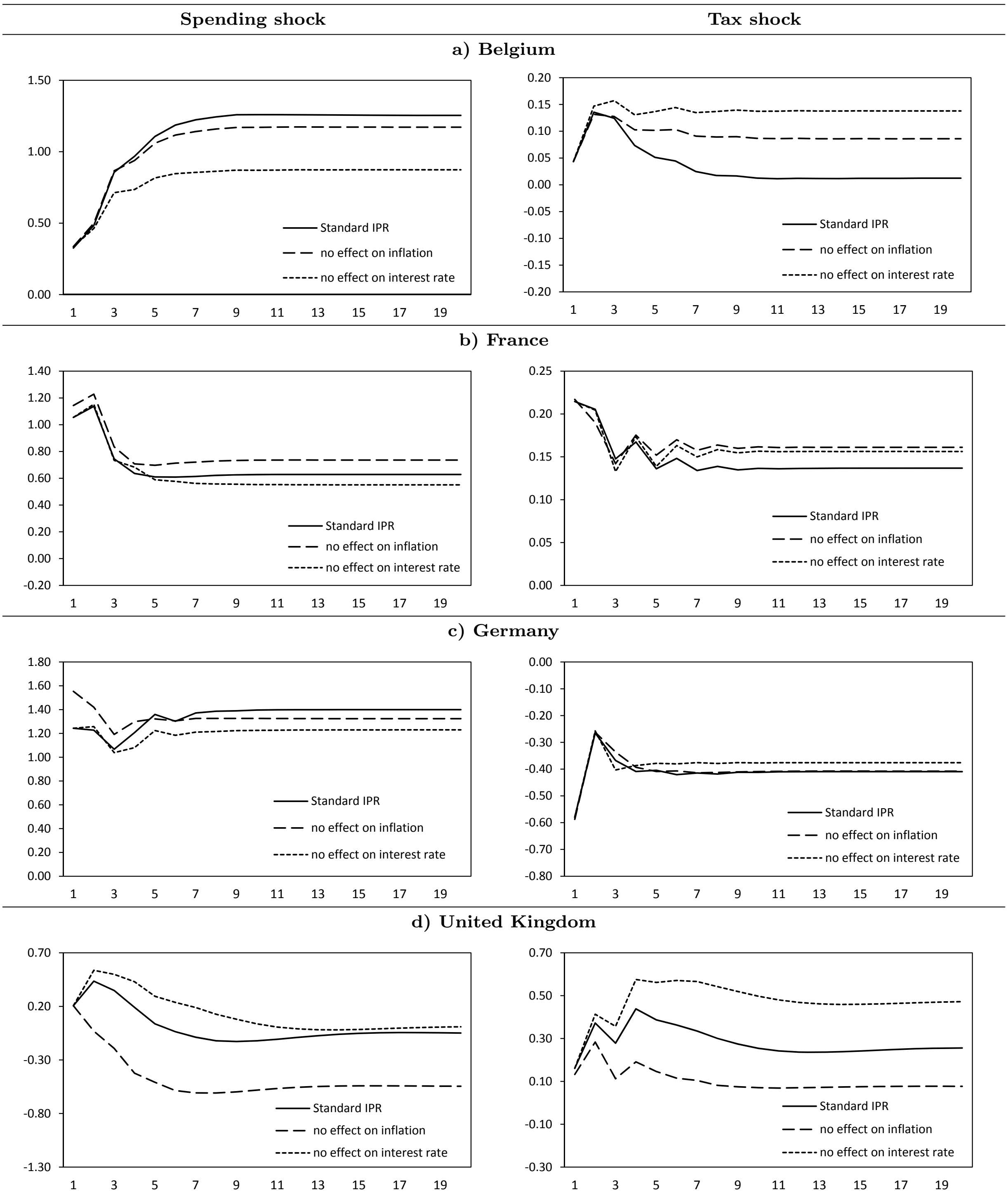
the dynamic response between the variables and restrict their responses to zero. All other responses in the model remain unrestricted.

Figure 3.4 shows the response of GDP to a spending (left column) and tax (right column) shock for three different cases. The impulse response labeled "Standard IPR" is derived from the unrestricted benchmark model and is identical to the one in figure 3.3. The two other graphs display the response of output when the respective policy instrument does not have an effect on the price level ("no effect on inflation") or on the interest rate ("no effect on the interest rate").

Considering that the response of expenditures and taxes cannot have an effect on inflation, figure 3.4 shows that only in the United Kingdom the differences between the standard and the counterfactual impulse responses are visual significant. This implies that in the United Kingdom the price channel has a positive effect for the transmission of fiscal policy as the counterfactual responses are lower than the benchmark response. In all other countries, the counterfactual impulse responses only differ little from the base case scenario. Thus, the marginal impact of the price channel is not economically significant in Belgium, France, and Germany.

A quite similar picture emerges when assessing the interest rate channel. Especially for France and Germany, the differences between the standard and the counterfactual responses are very small and economically insignificant. For Belgium, however, shutting down the interest channel essentially eliminates the impact of expenditure innovations on output. In contrast, tax innovations have an even more pronounced effect on GDP at least in Belgium and the United Kingdom.

Figure 3.4: Response of GDP when transmission channels are closed



Note: The figure displays the response of GDP to a one unit increase in public spending and net taxes when the inflation and interest rate channel is closed.

Overall, the counterfactual simulations provide mixed results on the importance of the price and interest rate channels in the transmission of fiscal policy. On the one hand, the marginal impact of the price and interest channel is not economically significant in France and Germany. On the other hand, Belgium and the United Kingdom display economically significant differences between the counterfactual and base case impulse response functions. But even here the direction of the marginal impact is different across countries. While shutting down the interest channel eliminates the impact of expenditure in Belgium, it slightly strengthens the effect on GDP in the United Kingdom.

3.7 Fiscal multipliers and the dynamic interaction of policy instruments

In order to satisfy the governmental intertemporal budget constraint, an increase in government expenditures requires either an adjustment of revenues or a reversal of expenditures in future periods, i.e. expenditures today might be financed by tax increases in the immediately following quarters.[87] The derived multipliers in section 3.5 and 3.6 still include these dynamic interaction of policy instruments. A spending multiplier eventually becomes smaller if expenditure shocks are accompanied by tax increases in the subsequent

[87] The issue of budgetary imbalances has been broadly discussed by Hoover and Sheffrin (1992) and Garcia and Henin (1999). A survey of the international empirical evidence on the tax-spend debate can be found by Payne (2003).

quarters.[88]

In order to analyze fiscal multipliers, which abstract from this dynamic interaction, we compute counterfactual multipliers with the same procedure as described in the previous chapter. Note, however, that in the case of the reaction of net taxes to spending, the response of net taxes are restricted to the level that would be observed if only automatic stabilizers had been at work, i.e. we shut down the discretionary response of net taxes to expenditures. The automatic stabilizer is computed by multiplying the impulse response function of output to a shock in expenditure by the tax elasticity that is also used in the contemporaneous variance-covariance matrix.[89]

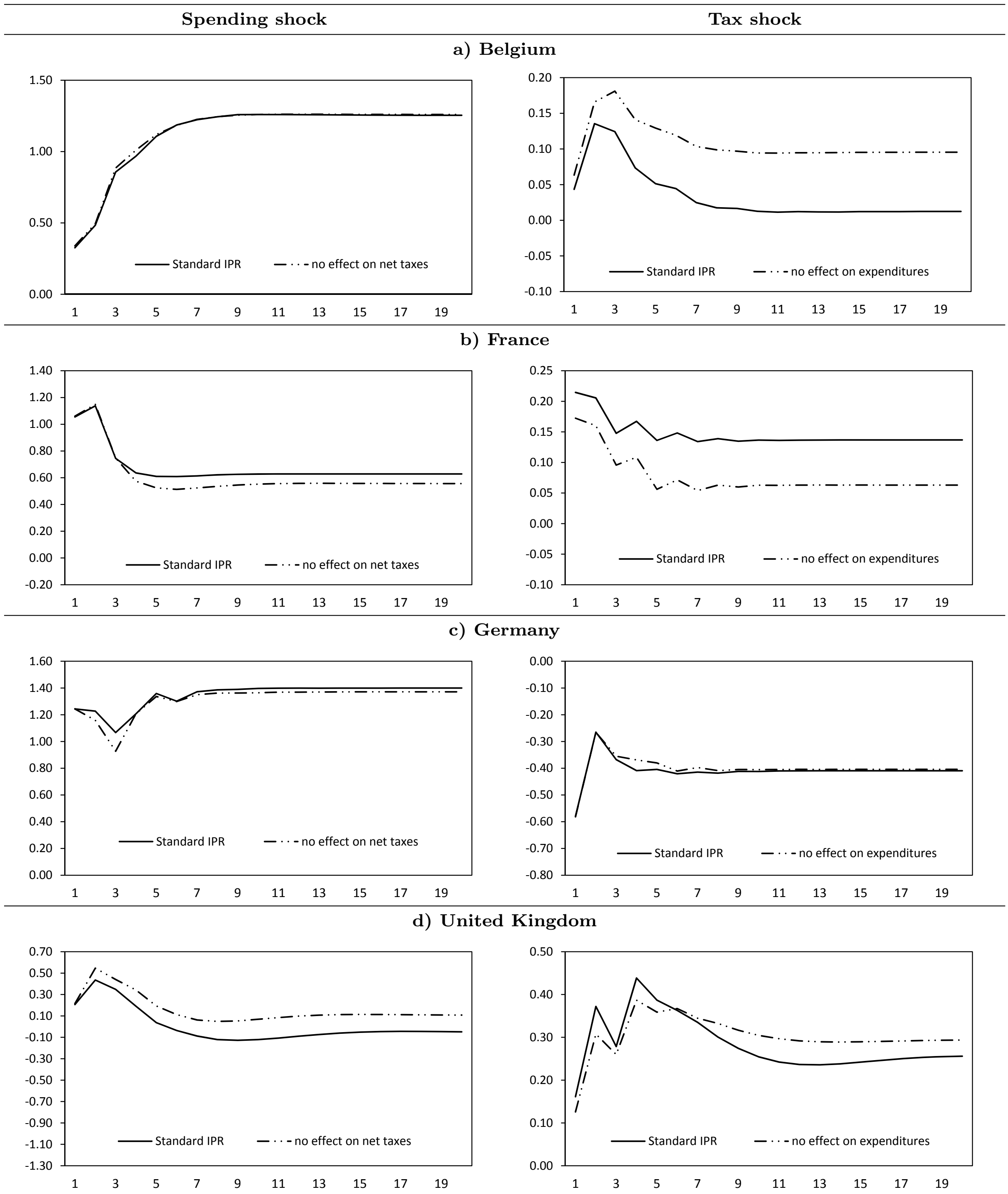
Figure 3.5 visualizes the difference between the base case response of output to a standard fiscal shock and the response of output to a "pure" fiscal shock, which abstracts from the discretionary dynamic interaction among policy instruments. Again, it can be seen that the results of the counterfactual exercise vary widely from country to country. While in Germany and the United Kingdom, the fiscal interaction has almost no economically significant effect on the estimated multipliers, the tax multipliers in Belgium and France differ from the unrestricted base case scenario. In the counterfactual scenario, Belgium's output response increases, while France's tax multiplier decreases. Still, the peak of these differences are rather small and equal to 0.07 after three and a half years (14 periods) in Belgium and -0.08 after one and three-quarter years (7 periods) in France.

The direction of the difference between the restricted and the unrestricted fiscal multiplier

[88] Fiscal multipliers derived from comparable policy experiments can be found in some VAR based studies that primarily focus on the United States (see Kuckuck and Westermann (2014), Ramey (2011b), Mountford and Uhlig (2009), Caldara and Kamps (2008), Perotti (2005)).

[89] The same calculation methodology has been applied in the previous chapter.

Figure 3.5: Response of GDP with no interaction of policy instruments



Note: The figure displays the response of GDP to a one unit increase in public spending and net taxes with no discretionary interaction of policy instruments. The reaction of net taxes to spending is restricted to the level that would be observed if only automatic stabilizers had been at work.

depends on two factors. On the one hand, it is essential in which direction taxes (expenditures) automatically respond to an increase in expenditures (taxes). On the other hand, it is essential in which way output reacts to a change of fiscal policy measures. In the case that expenditures today are financed by tax increases in the immediately following quarters, one would assume - in the estimation of the spending multiplier - a dampening effect on GDP. Thus, in such a scenario, the counterfactual multiplier should be substantially larger than the forecasting multiplier from standard SVAR estimates. If, conversely, expenditure increases are accompanied by tax reductions, the counterfactual multiplier should be substantially lower. However, in our sample the direction of these differences has to be interpreted with caution. As the benchmark regressions provide evidence that tax increases do not necessarily reduce output, a tax financed expenditure increase does not necessarily have a weaker effect on GDP than an expenditure increase without any tax adjustments. Nevertheless, the analysis still reveals that the direct interaction among policy instruments has only a limited impact on the fiscal multipliers, which is in contrast with the results of Kuckuck and Westermann (2014) for the United States.

3.8 Conclusion

This chapter contributes to the existing empirical literature by applying a five-variable SVAR approach to a uniform data set for the European countries Belgium, France, Germany, and the United Kingdom. Besides studying the effects of fiscal policy on output, we additionally analyze its dynamic effects on inflation and interest rates. Further, we investigate if the magnitude of fiscal multipliers depends on the discretionary dynamic interaction of policy instruments. By conducting counterfactual simulations, which abstract from the dynamic response of key macroeconomic variables to the initial fiscal shock, we study the importance of these channels for the transmission of fiscal policy on output.

The findings show that the effects of fiscal shocks are limited and rather different across countries. Only expenditure shocks in France and Germany as well as tax shocks in Germany have a statistically significant but short-lived effect on output. A one Euro increase in government spending raises GDP by around 1.05 Euros in France and 1.24 Euros in Germany. In this respect, public expenditure increases tend to be more efficient than net tax cuts. Furthermore, we find that the response of inflation and interest rate variables to fiscal shocks do not always have the sign as predicted by economic theory, e.g. interest rates significantly decrease in response to an expenditure shock and increase in response to a tax shock, which is robust across specifications and countries.

Subsequently, in order to study the importance of the price and interest rate channel in the transmission of fiscal policy, we conduct counterfactual simulations, which estimate the response of output to a fiscal shock, assuming that fiscal variables do not have a direct effect on the price level and the inflation rate. Across countries the results are inconclusive: on the one hand, the marginal impact of the price and interest channel is not economically significant in France and Germany. On the other hand, Belgium and the United Kingdom display economically significant differences between the counterfactual and base case impulse response functions. But even here the direction of the marginal impact is different across countries. While shutting down the interest channel eliminates the impact of expenditure in Belgium, it slightly strengthens the effect on GDP in the United Kingdom. Finally, we provide fiscal multipliers which abstract from the dynamic discretionary interaction of policy instruments. Again, the results of the counterfactual exercise vary from country to country. While in Germany and the UK, the fiscal interaction has almost no economically significant effect on the estimated expenditure multipliers, the tax multipliers in Belgium and France differ from the unrestricted base case scenario. In the counterfactual scenario, Belgium's output response increases, while France's tax multiplier decreases. Nevertheless, the picture emerges that the reciprocal financing of fiscal measures has no economically significant effect on the estimated multipliers.

3.A Appendix

In the following sections, further background information and robustness checks related to the main part of this chapter are provided. **Section 3.A.1** defines the variables and gives an overview of the data sources used in this chapter. Subsequently, **section 3.A.2** describes in detail the calculation methodology for the net tax elasticities and displays estimation results for different sub-categories of net taxes. **Section 3.A.3** presents impulse response functions of the base case VAR estimations and briefly discusses the effects of price, interest rates, and output shocks on the economy. **Section 3.A.4** displays several within-quarter net tax elasticities calculated under alternative assumptions and illustrates the correlation between inserted elasticity and estimated multipliers. The subsequent **section 3.A.5** provides fiscal multipliers with alternative definitions of net taxes as well as alternative definitions of price and interest rates. Finally, **section 3.A.6** compares the effects of fiscal policy measures derived from a recursive identification strategy and the Blanchard and Perotti (2002) identification scheme.

3.A.1 Data sources and definitions

VAR model

The data for the benchmark VAR model - unless otherwise noted - are taken from the quarterly non-financial accounts for general government (gov_q_ggnfa) as well as the quarterly national accounts (namq_gdp) collected by the European Commission (Eurostat (2013b)). The data covers the time period from 1991Q1 to 2011Q4 (see table 3.4). Whenever necessary, the time series were seasonally adjusted by applying the X12-ARIMA procedure.

Table 3.4: Data for the calculation of macroeconomic variables

Variable		**Definition**	**Data code**
Government expenditures		Final consumption expenditure	P3
	+	Gross fixed capital formation	P51
Net taxes		Taxes on production and imports, receivable	D2REC
	+	Current taxes on income, wealth, etc., receivable	D5REC
	-	Current taxes on income, wealth, etc., payable	D5PAY
	+	Social contributions, receivable	D61REC
	-	Social benefits other than social transfers in kind, payable	D62PAY
	+	Property income, receivable	D4REC
	-	Property income, payable	D4PAY
	+	Other current transfers, receivable	D7REC
	-	Other current transfers, payable	D7PAY
	-	Subsidies, payable	D3PAY
	-	Adjustment for the change in net equity of households in pension funds reserves	D8
Output		Gross Domestic Product	B1GM
Price index[90]		Consumer price index (2005=100)	043_PCPI
Interest rate[91]		3-Month London Interbank Offered Rate (LIBOR)	LDN:BBA

[90]Source: IMF (2012b)
[91]Source: Thomson Reuters (2013)

Exogenous elasticities and sensitivity analysis

Data for the calculation of the exogenous net tax elasticities (see appendix 3.A.2) and the additional sensitivity analysis are taken from various sources (see table 3.5). Again, whenever necessary, the series were seasonally adjusted by applying the X12-ARIMA procedure.

Table 3.5: Data for the calculation of net tax elasticities

Country	Variable	Data source	Data code
Personal income taxes			
Belgium	Gross wages and salaries	Eurostat (2013a)	tec00014_BE
	Employment	Banque Nationale de Belgique (2013)	-
France	Gross wages and salaries	Eurostat (2013a)	tec00014_FR
	Employment	National Institute of Statistics and Economic Studies (2013)	-
Germany	Gross wages and salaries	Eurostat (2013a)	tec00014_DE
	Working population	Federal Statistical Office (2013)	13321
United Kingdom	Gross wages and salaries	Eurostat (2013a)	tec00014_UK
	Workforce jobs	Office for National Statistics (2013)	DYDC
Corporate income taxes			
Belgium	Company profits	Thomson Reuters (2013)	BGXCOGT.A
France	Company profits	Thomson Reuters (2013)	FRXCOGT.B
Germany	Company profits	Thomson Reuters (2013)	BDXCOGT.A
United Kingdom	Company profits	Thomson Reuters (2013)	UKXCOGT.B
Sensitivity analysis			
Belgium	Capital transfers	Eurostat (2013b)	D9PAY/D9REC
	GDP deflator	IMF (2012b)	N_GDP_D
	Government bond yields	IMF (2012b)	FIGB_S
France	Capital transfers	Eurostat (2013b)	D9PAY/D9REC
	GDP deflator	Eurostat (2013b)	B1GM_CPI00_EUR
	Government bond yields	IMF (2012b)	FIGB_S
Germany	Capital transfers	Eurostat (2013b)	D9PAY/D9REC
	GDP deflator	Eurostat (2013b)	B1GM_CPI00_EUR
	Government bond yields	IMF (2012b)	FIGB_S
United Kingdom	Capital transfers	Eurostat (2013b)	D9PAY/D9REC
	GDP deflator	Eurostat (2013b)	B1GM_CPI00_EUR
	Government bond yields	IMF (2012b)	FIGB_S

3.A.2 Calculation of exogenous elasticities

Indirect taxes

Following Perotti (2005), we take the tax base to be GDP. Based on estimations by Girouard and André (2005), we use for every country a value of 1.0 for the elasticity of the indirect taxes to the tax base. Altogether, the revenue of indirect taxes are proportional to output.

Personal income taxes

Personal income taxes (PIT) are decomposed into tax rate t and tax base $W * E$ as follows:

$$PIT = t(WP) \cdot W(E) \cdot E(Y) \tag{3.5}$$

where t is the tax rate, W the real wage, P the price level, E the employment, and Y the output. After taking the logs (lowercase letters) and total differentiating the expression for total personal income taxes, one obtains:

$$\begin{aligned} d(pit_t) &= \tfrac{\partial t_t}{\partial w_t} dw_t + \tfrac{\partial w_t}{\partial e_t} de_t + \tfrac{\partial e_t}{\partial y_t} dy_t + \tfrac{\partial t_t}{\partial p_t} dp_t \\ &= \tfrac{\partial t_t}{\partial w_t}\tfrac{\partial e_t}{\partial e_t}\tfrac{\partial y_t}{\partial y_t} dw_t + \tfrac{\partial w_t}{\partial e_t}\tfrac{\partial e_t}{\partial e_t}\tfrac{\partial y_t}{\partial y_t} de_t + \tfrac{\partial e_t}{\partial y_t} dy_t + \tfrac{\partial t_t}{\partial p_t} dp_t \\ &= (\tfrac{\partial t_t}{\partial w_t}\tfrac{\partial w_t}{\partial e_t} + \tfrac{\partial w_t}{\partial e_t} + 1)\tfrac{\partial e_t}{\partial y_t} dy_t + \tfrac{\partial t_t}{\partial p_t} dp_t \\ &= ((\tfrac{\partial t_t}{\partial w_t} + 1)\tfrac{\partial w_t}{\partial e_t} + 1)\tfrac{\partial e_t}{\partial y_t} dy_t + \tfrac{\partial t_t}{\partial p_t} dp_t \end{aligned} \tag{3.6}$$

The term $\frac{\partial t_t}{\partial w_t} + 1$ denotes the elasticity of tax revenues per person to average real earnings and is taken from Girouard and André (2005). The elasticity of real wages to employment

($\frac{\partial w_t}{\partial e_t}$) is estimated from a regression of the log change of the gross wages and salaries on the first lead and four lags of the log change in employment. Finally, the elasticity of employment to GDP ($\frac{\partial e_t}{\partial y_t}$) is estimated from a regression of the employment on the first lead and four lags of the log change in GDP. Table 3.6 displays a general overview of the partial elasticities needed to calculate the overall personal income tax elasticity.

Table 3.6: Elasticities of personal income taxes

	$\frac{\partial t_t}{\partial w_t}+1$	$\frac{\partial w_t}{\partial e_t}$	$\frac{\partial e_t}{\partial y_t}$	$\frac{d(pit_t)}{dy_t}$
Belgium	1.6	3.7426	0.1324	0.9252
France	1.7	-0.3103	0.0719	0.0339
Germany	2.3	0.7929	0.2482	0.7008
United Kingdom	1.7	-0.2656	0.0173	0.0094

Note: The expression $\frac{\partial t_t}{\partial w_t}+1$ is taken from Girouard and André (2005). The term $\frac{\partial w_t}{\partial e_t}$ is estimated from a regression of the log change of the gross wages and salaries on the first lead and four lags of the log change in employment and $\frac{\partial e_t}{\partial y_t}$ is estimated from a regression of the employment on the first lead and four lags of the log change in GDP.

Social security taxes

The procedure to calculate the within-quarter elasticities for social security contributions (SSC) is the same as for the elasticities of personal income taxes. However, here the elasticity of tax revenues per person to average real earnings is substituted by the elasticity of social security contributions relative to earnings ($\frac{\partial s_t}{\partial w_t}+1$), which is taken from Girouard and André (2005). Table 3.7 provides a general overview of the subcomponents needed to calculate the overall elasticities.

Table 3.7: Elasticities of social security taxes

	$\frac{\partial s_t}{\partial w_t} + 1$	$\frac{\partial w_t}{\partial e_t}$	$\frac{\partial e_t}{\partial y_t}$	$\frac{d(ssc_t)}{dy_t}$
Belgium	1.1	3.7426	0.1324	0.6774
France	1.1	-0.3103	0.0719	0.0474
Germany	0.8	0.7929	0.2482	0.4056
United Kingdom	1.3	-0.2656	0.0173	0.0113

Note: The expression $\frac{\partial s_t}{\partial w_t}+1$ is taken from Girouard and André (2005). The term $\frac{\partial w_t}{\partial e_t}$ is estimated from a regression of the log change of the gross wages and salaries on the first lead and four lags of the log change in employment and $\frac{\partial e_t}{\partial y_t}$ is estimated from a regression of the employment on the first lead and four lags of the log change in GDP.

Corporate income taxes

We assume that corporate income taxes are proportional to company profits in all countries. This relationship is estimated from a regression of the log change of the company profits on the first lead and four lags of the log change in GDP where the results are shown in table 3.8. However, following Perotti (2005) this elasticity is only relevant when corporate income taxes are collected in the same quarter. In our benchmark estimation, we assume for every country that output has no contemporaneous effect on corporate income taxes due to collection lags. For robustness, we also calculate elasticities assuming no collection lag.

Table 3.8: Elasticities of corporate income taxes

	Belgium	France	Germany	United Kingdom
no collection lag	0.8273	0.9940	1.3902	0.9499
collection lag	0	0	0	0

Note: In the case of no collection lag the elasticity of corporate income taxes to GDP is estimated from a regression of the log change of the company profits on the first lead and four lags of the log change in GDP.

Transfer payments

In the case of the within-quarter elasticity of transfer payments to GDP, we use the elasticity of current primary expenditures to the economic cycle estimated by Girouard and André (2005) (see table 3.9). As these elasticities are based on annual data, it is just an approximation of the within-quarter elasticity of transfers. This procedure is comparable to Blanchard and Perotti (2002) as well as Perotti (2005). As only a small portion of transfer payments is sensitive to the business cycle within a quarter (i.e. unemployment benefits account only for 3 to 5 percentage of total expenditures) and disaggregated data is not available, we calculate overall net tax elasticities excluding current transfer payments for robustness.[92]

Table 3.9: Elasticities of transfer payments

Belgium	France	Germany	United Kingdom
-0.14	-0.11	-0.18	-0.05

Note: Values are taken from Girouard and André (2005).

[92] An overview of within-quarter net tax elasticities calculated under alternative assumptions is provided in table 3.10 in section 3.A.4.

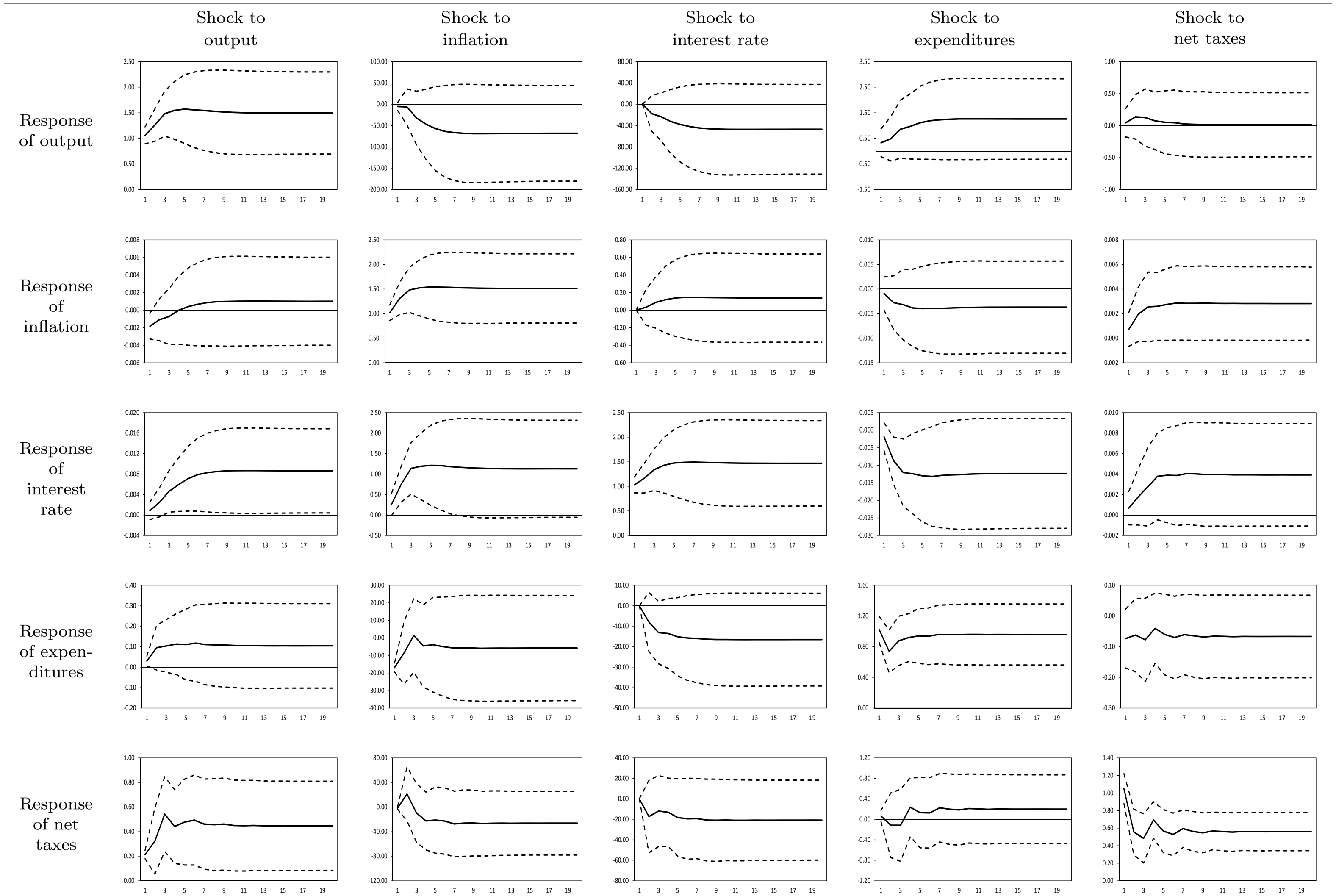
3.A.3 Base case SVAR estimations

This section provides a complete picture of the country-specific SVAR estimations and the corresponding cumulative impulse response functions. In addition to the effects of fiscal policy, figure 3.6 to 3.9 display the effects of price, interest rate, and output shocks on the economy. The impulse responses in the last two columns of each figure are identical to the ones discussed in section 3.5.

Similar to the findings on fiscal policy, the whole set of impulse responses shows that the effects of output, inflation, and interest rate shocks are quite different across countries. Although an increase in interest rates leads to a negative reaction of output in all countries, this effect is only statistically significant in the United Kingdom. Moreover, an increase in interest rates has no significant effect on fiscal variables. Only the net taxes in France and the United Kingdom seem to be significantly affected by an interest rate shock. Overall, we find that responses of output and price levels to an interest rate shock are in line with standard monetary VAR findings (see, e.g., Christiano et al. (1999)).

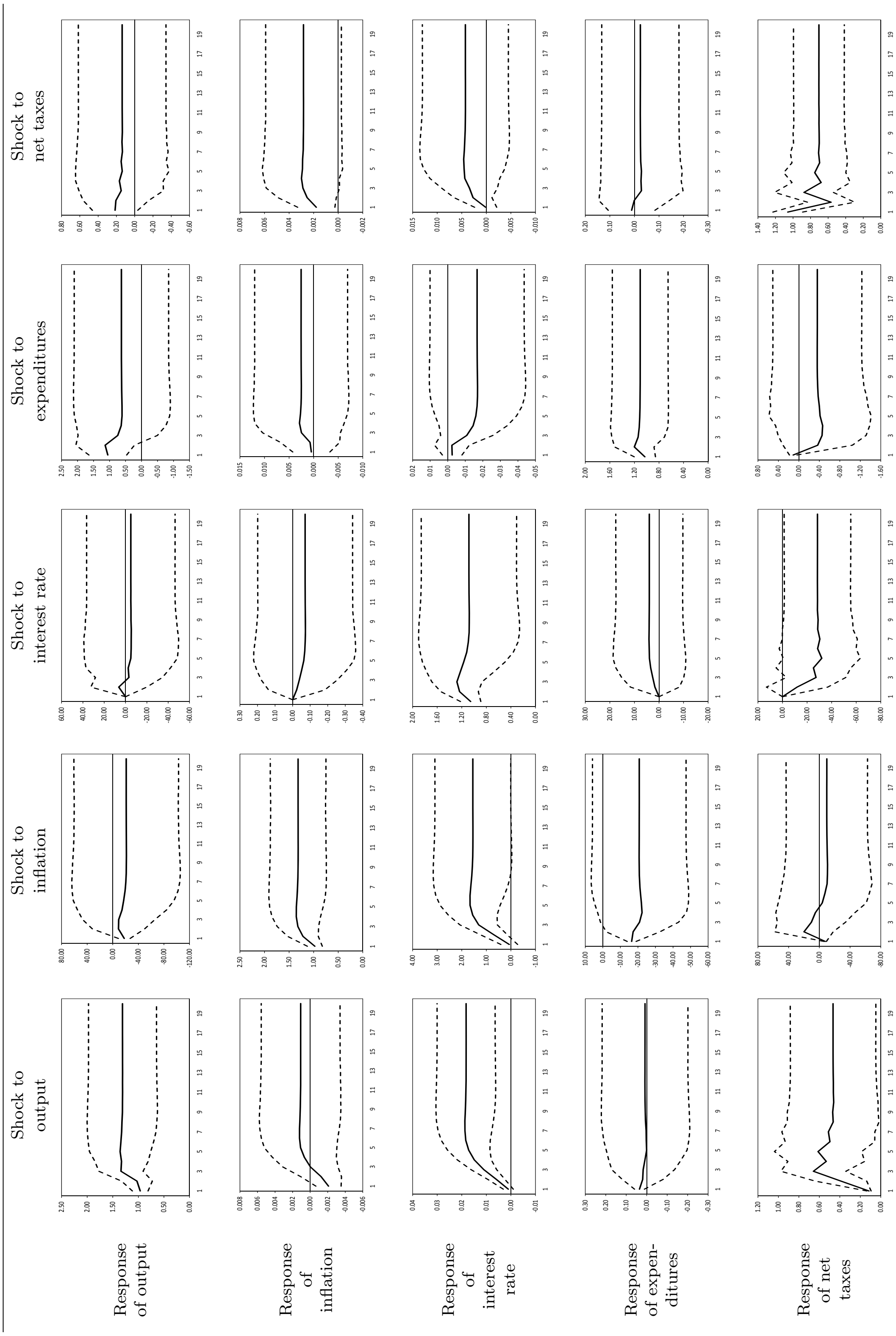
Furthermore, the figures reveal the general absence of countercyclical spending behavior of governments. An increase in output has a positive effect on government expenditures in Belgium, France, and the United Kingdom. Only in the case of Germany, government spending displays a significant and negative reaction to an increase in output. Additionally, we find that net taxes respond persistently positive to an output shock.

Figure 3.6: SVAR estimations for Belgium



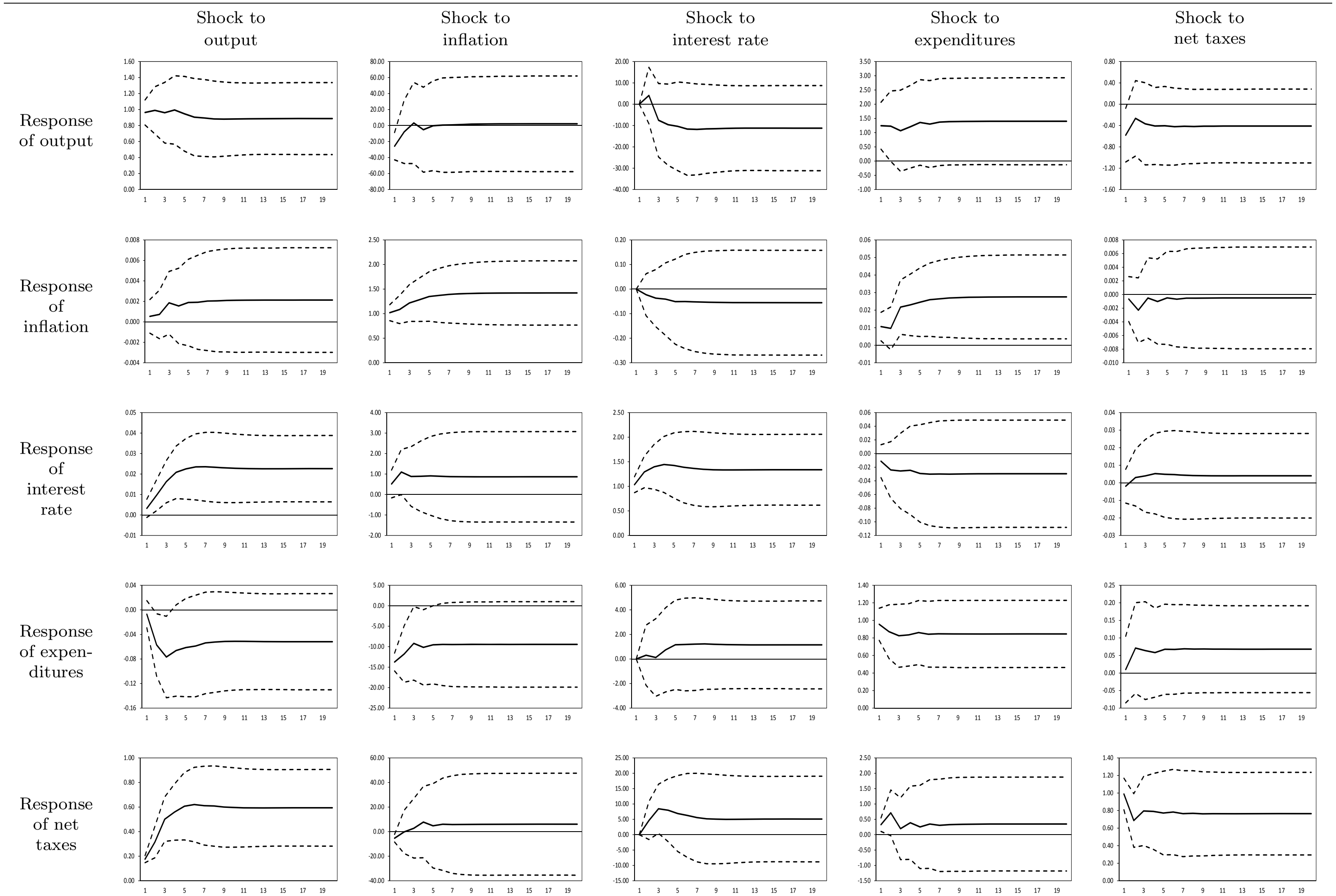
Note: SVAR model contains five variables (output, price level, interest rate, expenditures, and net taxes). The solid line traces the accumulated impulse response of one variable to a one unit shock in another variable. The dotted lines visualize the 95% confidence bands.

Figure 3.7: SVAR estimations for France



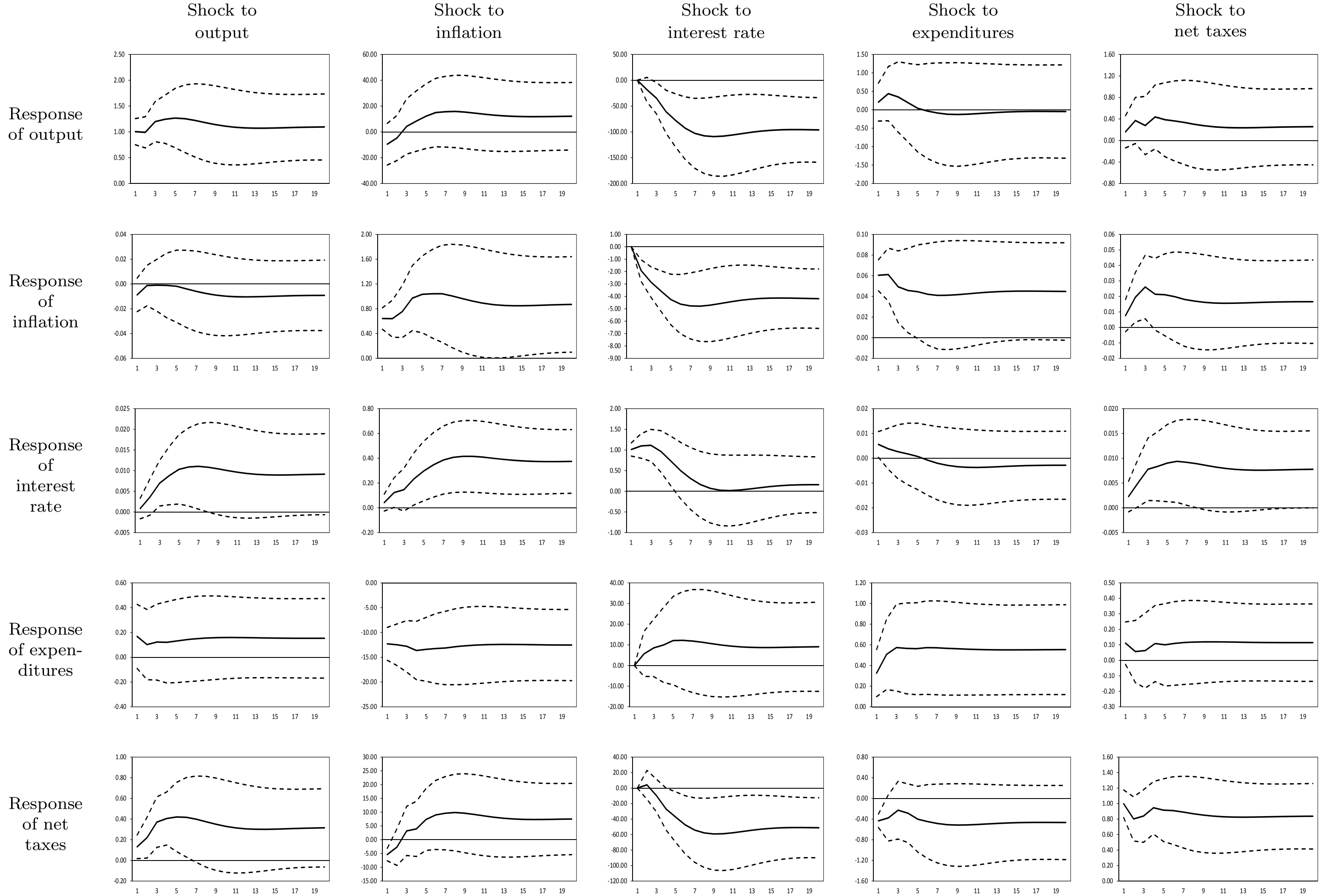
Note: SVAR model contains five variables (output, price level, interest rate, expenditures, and net taxes). The solid line traces the accumulated impulse response of one variable to a one unit shock in another variable. The dotted lines visualize the 95% confidence bands.

Figure 3.8: SVAR estimations for Germany



Note: SVAR model contains five variables (output, price level, interest rate, expenditures, and net taxes). The solid line traces the accumulated impulse response of one variable to a one unit shock in another variable. The dotted lines visualize the 95% confidence bands.

Figure 3.9: SVAR estimations for the United Kingdom



Note: SVAR model contains five variables (output, price level, interest rate, expenditures, and net taxes). The solid line traces the accumulated impulse response of one variable to a one unit shock in another variable. The dotted lines visualize the 95% confidence bands.

3.A.4 Alternative exogenous elasticities

Table 3.10: Within-quarter net tax elasticity to GDP

	η_i	η_i^*	η_i^{**}	η_i^{***}	η_i^{****}
Belgium	0.9652	0.8580	0.8014	0	1.6028
France	0.5304	0.4455	0.3893	0	0.7786
Germany	0.9504	0.7009	0.6094	0	1.2188
United Kingdom	0.4243	0.4765	0.3891	0	0.7782

Note: $\eta_i \,\hat{=}$ Net tax elasticity calculated with collection lag for corporate income taxes and including transfer payments; $\eta_i^* \,\hat{=}$ Net tax elasticity calculated with no collection lag for corporate income taxes and excluding transfer payments; $\eta_i^{**} \,\hat{=}$ Net tax elasticity calculated with collection lag for corporate income taxes and excluding transfer payments; $\eta_i^{***} \,\hat{=}$ Zero lower bound net tax elasticity; $\eta_i^{****} \,\hat{=}$ Upper bound net tax elasticity ($\eta_i^{****} = 2 \times \eta_i$).

Caldara and Kamps (2012) point out that standard identification schemes imply different priors on elasticities, generating a large dispersion in multiplier estimates. On that account, this section provides some additional estimations of fiscal multipliers with altering tax and price elasticities.[93]

Table 3.10 displays several within-quarter net tax elasticities calculated under alternative assumptions. In our analysis we consider i) net tax elasticities calculated with collection lags for corporate income taxes and including transfer payments (benchmark regression), ii) net tax elasticities calculated with no collection lags for corporate income taxes and excluding transfer payments, iii) net tax elasticities calculated with collection lag for corporate income taxes and excluding transfer payments as well as a iv) zero lower bound net tax elasticity and a v) upper bound net tax elasticity, which is twice the benchmark elasticity.[94] The latter two elasticities serve the purpose of emphasizing the linkage be-

[93] Remember that the price elasticity is merely the real output elasticity of the fiscal variable less one. See also section 3.A.6 in which we discuss a recursive identification strategy with a zero price elasticity.

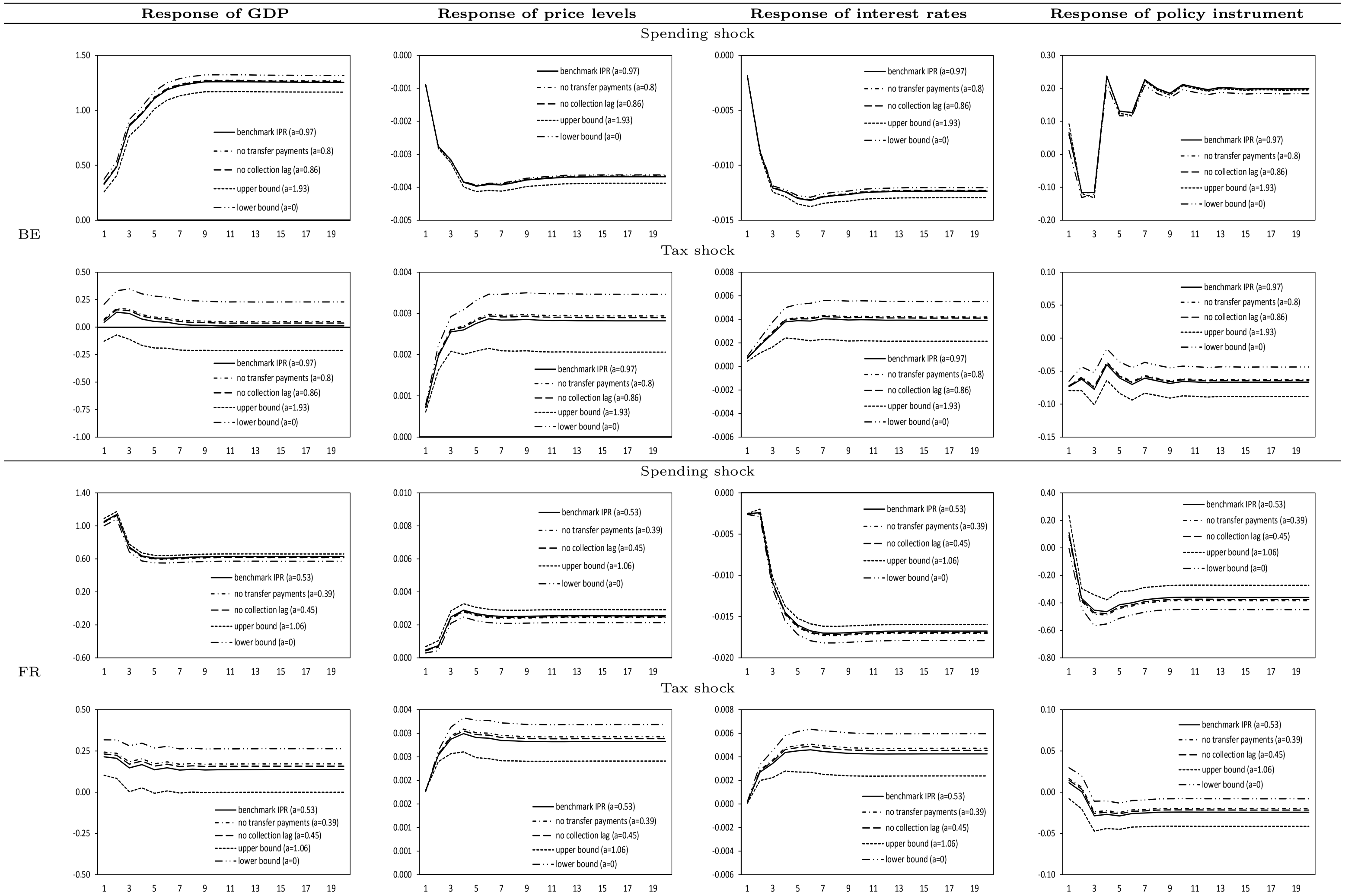
[94] The item 'other current transfers' comprise net non-life insurance premiums, non-life insurance claims, current transfers within general government as well as transfer payments for current international cooperation. It is questionable whether all of these categories are sensitive to the cycle. Thus, we estimate elasticities excluding the sensitivity of 'other current transfers'; see section 3.A.2.

tween inserted elasticity and fiscal multipliers and should cover the relevant range for net taxes. Further, it is noticeable that there are only marginal differences between net tax elasticities that are calculated excluding transfer payments as well as no corporate income collection lag. Details on the calculation procedure can be found in section 3.A.2. Figure 3.10 visualizes the response of macroeconomic variables to fiscal shocks under alternative tax and price elasticities. In general, it can be seen that expenditure multipliers are only affected little by the choice of the net tax elasticity. This applies in principle for all variables and countries. Only Germany and the United Kingdom display some variance with regard to the magnitude of the multipliers, which particular stands out in the case of a zero tax elasticity.

On the contrary, figure 3.10 shows that tax multipliers crucially depend on the inserted output elasticity. Yet, the difference between the effects of tax shocks calculated with elasticities excluding and including transfer payments and with or without collection lag are very limited. It seems to be a plausible assumption that this range of elasticities accurately capture the working procedure of the automatic stabilizers.

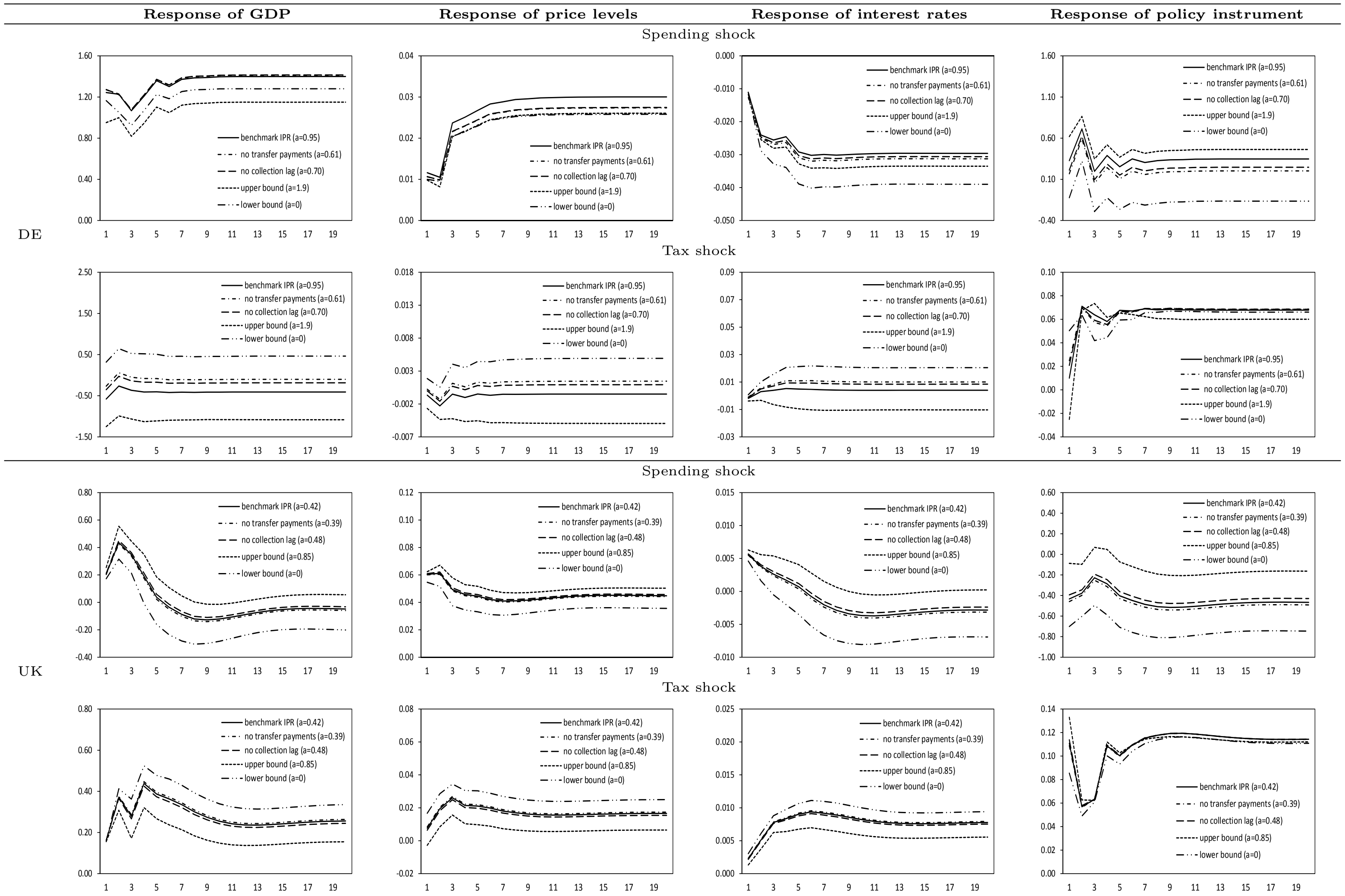
Finally, the lower and upper bound elasticities exhibit that a change from the bottom to the top elasticity generally leads to a down shift of the impulse response pattern. Overall, it can be seen that increasing automatic stabilizers in terms of net taxes lead to a down shift of the impulse response pattern. However, the response of GDP to tax shock stays statistically insignificant irrespective of the inserted elasticity in Belgium, France and the United Kingdom.

Figure 3.10: Response of macroeconomic variables to fiscal shocks under alternative tax and price elasticities



The figure is continued on the next page.

Figure 3.10: Response of macroeconomic variables to fiscal shocks under alternative tax and price elasticities - continued



Note: The figure displays the response of macroeconomic variables to a one unit increase in public spending and net taxes in Belgium (BE), France (FR), Germany (DE), and the United Kingdom (UK) under alternative tax and price elasticities. Details on the inserted within-quarter net tax elasticities can be found in table 3.10.

3.A.5 Alternative definitions of variables

There has been a broad discussion about the definition of fiscal variables when analyzing the effects of fiscal policy on economic activity. In our benchmark analysis we define net taxes as the sum of net tax receipts including contributions for social security less net property income, net current transfers and net subsidies (see section 3.A.1). As mentioned in the main part, capital transfers have been omitted from the analysis due to statistical inconsistencies as a consequence of economic restructurings and reallocations from the private to the public sector. As shown in figure 3.11, these debt reschedulings have led to extreme values of government spending and revenue in some periods, which have no economic importance for the analysis of fiscal policy. Nevertheless, capital transfers also cover inheritance taxes, death duties, taxes on gifts as well as investment grants. Items that should not be excluded when analyzing the effects of fiscal policy. Hence, this section provides fiscal multipliers with a definition of net taxes that include adjusted capital transfers. Further, we analyze if the estimated multipliers are sensitive to transfer expenditures and reestimate our model with net taxes disregarding transfer payments.[95]

Instead of using the CPI as a measure for the price level, we also deflated the variables with the GDP deflator and analyzed whether this has any effect on the estimated multipliers. Additionally, we performed robustness checks regarding the interest rates by substituting the LIBOR with long-term government bond yields and we also reestimated the multipliers with absolute instead of per capita values.[96]

[95]In the latter case, we use the tax elasticity calculated excluding transfer payments, which is 0.80 for Belgium, 0.39 for France, 0.61 for Germany and 0.39 for the United Kingdom.

[96]See section 3.A.1 for data sources and definitions.

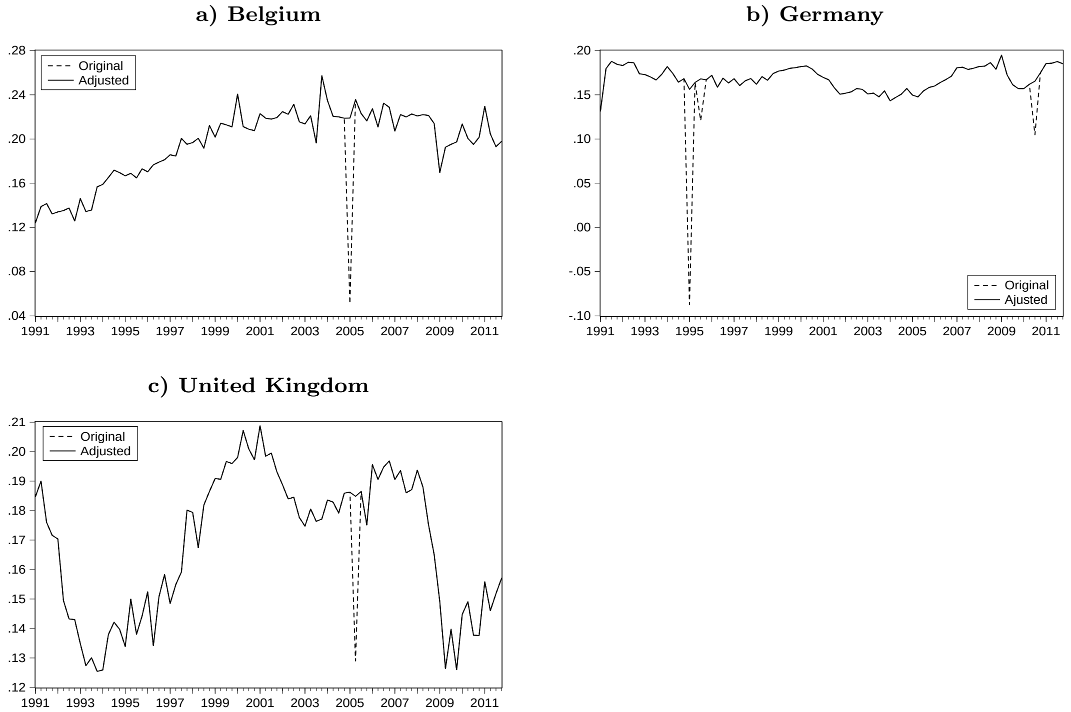
Data adjustment of capital transfers

In order to account for the statistical inconsistencies of the capital transfer series, we adjust the data - with the help of Eurostat and national statistical offices - for periods with extraordinary debt rescheduling. In our sample this affects Belgium, Germany, and the United Kingdom. Details about this data cleansing process can be found in table 3.11. In addition, figure 3.11 displays the difference between the original net tax series and the adjusted series.

Table 3.11: Data adjustment of capital transfers

Country	**Period**	**Time series**	**Adjustment**
Belgium	2005Q1	Capital transfer payments	Debt assumption by government for the railway company SNCB in the amount of 7.5 billion Euros.
Germany	1995Q1	Capital transfer payments	Debt assumption by government for the Treuhandanstalt (Trust agency) in the amount of 205 billion Deutsche Mark (=104.8 billion Euros).
	1995Q3	Capital transfer payments	Transfer payments for the Eastern German housing industry in the amount of 31 billion Deutsche Mark (=15.8 billion Euros).
	2010Q3	Capital transfer payments	Transfer of assets and liabilities from the Hypo Real Estate Holding AG to the FMS Wertmanagement (bad bank) in the amount of 30 billion Euros (estimated).
United Kingdom	2005Q2	Capital transfer payments	Transfer payments for nuclear reactors from British Nuclear Fuels Limited to the Nuclear Decommissioning Authority in the amount of 15 billion Euros (estimated).

Figure 3.11: Data adjustment of capital transfers



Note: The figure displays the original and adjusted capital transfer series. Details about the data adjustment process can be found in table 3.11.

Fiscal multipliers with alternative definitions of net taxes

The results of fiscal multipliers with alternative definitions of net taxes are summarized in table 3.12. Segment (1) displays expenditure and tax multipliers from the baseline model, segment (2) fiscal multipliers estimated with net taxes including capital transfers, and segment (3) fiscal multipliers estimated with net taxes excluding transfer payments. It can be seen that a change of the tax definition has only a marginal effect on the estimated multipliers. In fact, multipliers derived from a model including capital transfers are slightly lower than the benchmark estimations, while multipliers excluding transfer payments generally remain unchanged.

Fiscal multipliers with alternative definitions of variables

The outcome of fiscal multipliers with alternative variable definitions are displayed in table 3.13. Segment (1) shows expenditure and tax multipliers derived from a model with the GDP deflator as price measure, segment (2) displays results with government bond yields as a long-run measure for interest rates, and segment (3) presents multipliers derived from a model with absolute values.

Using the GDP deflator as an alternative measure of price development, we find significant differences in comparison to the base case scenario in Belgium, France, and the UK. Only in the case of Germany the results are similar to the benchmark estimations. Besides the statistically significant and positive effect of spending shocks on GDP in Germany, we now also find a significant effect in the UK. Again, we find evidence that public expenditure increases tend to be more efficient than tax cuts. The responses of interest rate and price levels to fiscal shocks remain unchanged. The same also applies for multipliers derived from a model with government bond yields as well as absolute values. In both cases the qualitative outcome does not change.[97]

[97] We also conduct the analysis in a reduced sample, which excludes the financial crisis and stops in 2006Q4 (results are not shown). Generally, this has no effect on the results.

Table 3.12: Fiscal multipliers with alternative definitions of net taxes

		Spending shock						**Tax shock**					
Country		Q1	Q2	Q3	Q4	Q6	Q8	Q1	Q2	Q3	Q4	Q6	Q8
(1) Baseline regression													
BE	GDP	0.326	0.484	0.856	0.968	1.186	1.244	0.043	0.135	0.124	0.073	0.044	0.017
	INF	-0.001	-0.003	-0.003	-0.004	-0.004	-0.004	0.001	0.002	0.003	0.003	0.003	0.003
	INT	-0.002	-0.009*	-0.012*	-0.012*	-0.013	-0.013	0.001	0.002	0.003	0.004	0.004	0.004
	POL	0.066	-0.116	-0.116	0.237	0.126	0.199	-0.074	-0.062	-0.078	-0.040	-0.070	-0.065
FR	GDP	1.055*	1.138*	0.745	0.636	0.609	0.622	0.215	0.206	0.148	0.167	0.148	0.139
	INF	0.000	0.001	0.002	0.003	0.003	0.003	0.002*	0.003*	0.003	0.003	0.003	0.003
	INT	-0.003	-0.002	-0.011	-0.014	-0.017	-0.017	0.000	0.003	0.003	0.004	0.004	0.004
	POL	0.111*	-0.368	-0.453	-0.465	-0.400	-0.368	0.011	0.001	-0.029	-0.027	-0.026	-0.024
DE	GDP	1.243*	1.227	1.067	1.207	1.302	1.386	-0.580*	-0.265	-0.368	-0.409	-0.421	-0.418
	INF	0.011*	0.010	0.022*	0.023*	0.026*	0.027*	-0.001	-0.002	-0.001	-0.001	-0.001	-0.001
	INT	-0.011	-0.024	-0.026	-0.025	-0.030	-0.030	-0.002	0.003	0.004	0.005	0.005	0.004
	POL	0.327*	0.712	0.197	0.389	0.348	0.327	0.010	0.071	0.064	0.058	0.067	0.068
UK	GDP	0.207	0.436	0.348	0.191	-0.036	-0.121	0.161	0.372	0.278	0.439	0.363	0.301
	INF	0.060*	0.061*	0.049*	0.046*	0.042	0.041	0.008	0.019*	0.026*	0.021	0.020	0.017
	INT	0.006*	0.004	0.003	0.002	-0.001	-0.003	0.002	0.005*	0.008*	0.008*	0.009	0.009
	POL	-0.433*	-0.378	-0.228	-0.286	-0.448	-0.509	0.111	0.057	0.063	0.109	0.109	0.118
(2) Net taxes including adjusted capital transfers													
BE	GDP	0.299	0.502	0.866	0.976	1.163	1.218	0.006	0.113	0.058	0.020	-0.024	-0.043
	INF	-0.001	-0.002	-0.003	-0.003	-0.003	-0.003	0.001	0.002	0.003	0.003	0.003	0.003
	INT	-0.002	-0.008*	-0.012*	-0.012*	-0.012	-0.012	0.001	0.002	0.003	0.004*	0.004	0.004
	POL	0.060	0.060	0.054	0.462	0.348	0.376	-0.079	-0.065	-0.085	-0.049	-0.076	-0.073
FR	GDP	1.060*	1.131*	0.737	0.635	0.612	0.619	0.184	0.181	0.123	0.155	0.147	0.138
	INF	0.001	0.001	0.003	0.003	0.003	0.003	0.002*	0.002*	0.003	0.003	0.003	0.003
	INT	-0.003	-0.003	-0.011	-0.015	-0.017	-0.017	0.000	0.002	0.003	0.004	0.005	0.005
	POL	0.111*	-0.429	-0.493	-0.523	-0.426	-0.384	0.001	-0.029	-0.073	-0.070	-0.069	-0.066
DE	GDP	1.081*	0.960	0.674	0.957	1.130	1.180	-0.335*	-0.549*	-0.642*	-0.620*	-0.494*	-0.442*
	INF	0.007*	0.008	0.016*	0.016*	0.019*	0.020*	0.000	0.000	-0.001	-0.001	-0.001	0.000
	INT	-0.009	-0.019	-0.024	-0.024	-0.025	-0.025	-0.003	-0.007	-0.013*	-0.017*	-0.018*	-0.016*
	POL	0.194*	-0.088	-0.038	-0.308	-0.352	-0.292	-0.035	-0.006	0.002	0.013	0.005	-0.003
UK	GDP	0.200	0.497	0.424	0.287	0.066	-0.017	0.138	0.208	0.162	0.220	0.102	0.038
	INF	0.061*	0.064*	0.055*	0.051*	0.046	0.045	0.007	0.015*	0.017*	0.009	0.007	0.005
	INT	0.005	0.003	0.002	0.001	-0.001	-0.003	0.003*	0.005*	0.007*	0.007*	0.007*	0.006
	POL	-0.435*	-0.303	-0.346	-0.384	-0.553	-0.607	0.074	0.033	0.044	0.088	0.087	0.090
(3) Net taxes excluding transfer payments													
BE	GDP	0.329	0.483	0.854	0.970	1.195	1.255	0.053	0.163	0.176	0.120	0.097	0.066
	INF	-0.001	-0.003	-0.003	-0.004	-0.004	-0.004	0.001	0.002	0.003	0.003	0.003	0.003
	INT	-0.002	-0.009*	-0.012*	-0.013	-0.013	-0.013	0.001	0.002	0.002	0.004	0.004	0.004
	POL	0.057	-0.113	-0.069	0.203	0.135	0.192	-0.070	-0.065	-0.071	-0.031	-0.062	-0.057
FR	GDP	1.043*	1.133*	0.743	0.633	0.605	0.619	0.257	0.251	0.179	0.202	0.180	0.169
	INF	0.000	0.001	0.002	0.003	0.002	0.002	0.002*	0.003*	0.003	0.003	0.003	0.003
	INT	-0.002	-0.002	-0.011	-0.014	-0.017	-0.017	0.000	0.002	0.003	0.004	0.004	0.004
	POL	0.080*	-0.283	-0.365	-0.364	-0.314	-0.289	0.027	0.026	-0.007	-0.004	-0.005	-0.003
DE	GDP	1.239*	1.198	1.033	1.218	1.304	1.381	-0.508	-0.083	-0.191	-0.293	-0.296	-0.305
	INF	0.008*	0.007	0.016*	0.016*	0.019*	0.020*	0.000	-0.001	0.000	0.000	0.000	0.000
	INT	-0.008	-0.018	-0.019	-0.018	-0.021	-0.021	0.002	0.009	0.011	0.013	0.013	0.012
	POL	0.108	0.252	-0.020	0.068	0.060	0.059	0.031	0.115	0.089	0.084	0.098	0.098
UK	GDP	0.219	0.412	0.275	0.089	-0.184	-0.292	0.136	0.381	0.343	0.537	0.513	0.478
	INF	0.056*	0.052*	0.038*	0.032	0.028	0.026	0.015*	0.031*	0.040*	0.037*	0.037*	0.035
	INT	0.004*	0.003	0.002	0.002	-0.001	-0.003	0.002	0.005	0.007	0.008	0.010	0.010
	POL	-0.425*	-0.422*	-0.275	-0.346	-0.530	-0.592	0.083	0.067	0.080	0.120	0.120	0.128

Note: The table displays the responses of output (GDP), price level (INF), interest rate (INT), and the fiscal counterpart (POL) to government spending and tax shocks. The symbol * indicates that the reactions are statistically different from zero.

Table 3.13: Fiscal multipliers with alternative definitions of variables

		Spending shock						Tax shock					
Country		Q1	Q2	Q3	Q4	Q6	Q8	Q1	Q2	Q3	Q4	Q6	Q8
(1) GDP deflator													
BE	GDP	-0.024	-0.056	-0.047	0.180	0.294	0.332	0.147	0.365*	0.358	0.350	0.354	0.347
	INF	0.005	0.004	0.010*	0.008	0.010*	0.011*	-0.001	-0.001	0.000	0.000	0.000	0.000
	INT	0.000	-0.006	-0.010	-0.011	-0.012	-0.012	0.001	0.002	0.003	0.004	0.004	0.004
	POL	-0.007	-0.330	-0.422	-0.032	-0.124	-0.092	-0.050	-0.008	-0.049	-0.020	-0.032	-0.026
FR	GDP	0.347	-0.138	-0.724	-0.904	-0.806	-0.712	0.260*	0.316*	0.373	0.398	0.401	0.384
	INF	0.023*	0.035*	0.041*	0.041*	0.037*	0.036*	0.001	0.000	-0.002	-0.001	-0.001	-0.001
	INT	-0.002	0.002	-0.006	-0.013	-0.021	-0.023	0.000	0.003	0.003	0.004	0.005	0.005
	POL	-0.074	-0.758	-0.855	-0.989	-0.886	-0.763	0.044	0.059	0.057	0.038	0.035	0.034
DE	GDP	1.475*	0.934	0.566	0.454	0.248	0.259	-0.300	0.019	0.082	0.015	-0.008	-0.038
	INF	0.007	0.019*	0.028*	0.035*	0.044*	0.049	-0.004*	-0.006	-0.006	-0.006	-0.006	-0.006
	INT	-0.002	-0.003	-0.003	-0.004	-0.008	-0.010	-0.002	-0.001	0.000	0.000	0.001	0.000
	POL	0.266*	0.620	0.153	0.144	-0.057	-0.132	0.078	0.155*	0.169*	0.141	0.156	0.157
UK	GDP	1.303*	1.342*	1.000*	0.927	0.895	0.863	0.225	0.595*	0.770*	0.770*	0.726	0.636
	INF	0.000	-0.004	-0.005	-0.006	-0.007	-0.007	0.001	0.001	0.000	-0.001	-0.002	-0.002
	INT	0.006*	0.006	0.005	0.004	0.003	0.002	0.003	0.008*	0.013*	0.015*	0.017*	0.017*
	POL	0.216*	0.258	0.265	0.241	0.179	0.160	0.198*	0.299*	0.438*	0.436*	0.437	0.408
(2) Government bond yields													
BE	GDP	0.320	0.446	0.767	0.784	0.916	0.938	0.022	0.097	0.070	0.019	0.020	-0.002
	INF	-0.001	-0.003	-0.003	-0.004	-0.003	-0.003	0.001	0.002	0.003	0.003	0.003	0.003
	INT	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000	0.000
	POL	0.065	-0.160	-0.196	0.078	0.012	0.053	-0.083	-0.065	-0.087	-0.044	-0.060	-0.057
FR	GDP	0.975*	1.071*	0.664	0.583	0.511	0.505	0.241*	0.240	0.201	0.236	0.234	0.235
	INF	0.000	0.001	0.003	0.003	0.003	0.003	0.002*	0.002*	0.003	0.003	0.003	0.003
	INT	0.000	-0.001	-0.002	-0.003	-0.003	-0.003	0.000	0.001	0.000	0.000	0.000	0.000
	POL	0.102*	-0.378	-0.491	-0.531	-0.615	-0.633	0.011	-0.002	-0.027	-0.027	-0.032	-0.032
DE	GDP	1.332*	1.302*	1.146	1.207	1.198	1.279	-0.605*	-0.393	-0.542	-0.553	-0.540	-0.538
	INF	0.011*	0.009	0.021*	0.022*	0.024*	0.025	-0.001	-0.003	-0.001	-0.002	-0.001	-0.001
	INT	0.001	0.002	0.002	0.002	0.001	0.002	0.000	0.000	0.000	0.000	0.000	0.000
	POL	0.352*	0.779*	0.269	0.552	0.463	0.411	0.006	0.068	0.071	0.072	0.075	0.074
UK	GDP	0.404	0.802*	1.016	1.038	1.033	1.001	0.194	0.442	0.395	0.612	0.613	0.638
	INF	0.072*	0.081*	0.075*	0.074*	0.074*	0.074	0.014*	0.030*	0.035*	0.032*	0.035*	0.036
	INT	0.003*	0.004	0.006*	0.006*	0.005*	0.006*	0.000	0.001	0.001	0.001	0.001	0.001
	POL	-0.502*	-0.308	-0.009	0.048	-0.001	-0.011	0.117	0.062	0.073	0.105	0.097	0.093
(3) Absolute values													
BE	GDP	0.314	0.453	0.812	0.915	1.113	1.162	0.039	0.125	0.110	0.060	0.029	0.004
	INF	0.000	-0.023	-0.026	-0.032	-0.032	-0.031	0.000	0.018	0.023	0.024	0.026	0.026
	INT	0.000	-0.078*	-0.108*	-0.110*	-0.118	-0.114	0.000	0.016	0.025	0.035	0.035	0.036
	POL	0.063	-0.132	-0.147	0.209	0.084	0.153	-0.059	-0.052	-0.065	-0.035	-0.059	-0.054
FR	GDP	1.060*	1.150*	0.763	0.653	0.618	0.628	0.216	0.207	0.149	0.170	0.151	0.142
	INF	0.000	0.001	0.004	0.005	0.004	0.004	0.000	0.004	0.004	0.005	0.004	0.004
	INT	0.000	-0.003	-0.016	-0.021	-0.025	-0.025	0.000	0.004	0.005	0.007	0.007	0.007
	POL	0.111*	-0.368	-0.457	-0.470	-0.411	-0.382	0.012	0.001	-0.029	-0.027	-0.025	-0.024
DE	GDP	1.264*	1.257	1.147	1.442*	1.633*	1.713	-0.602*	-0.230	-0.329	-0.448	-0.439	-0.448
	INF	0.000*	0.014	0.029*	0.030*	0.035*	0.036*	0.000	-0.004	-0.001	-0.002	-0.001	-0.001
	INT	0.000	-0.011	-0.014	-0.014	-0.014	-0.012	0.000	0.003	0.004	0.004	0.004	0.003
	POL	0.335*	0.691	0.176	0.347	0.380	0.424	0.015	0.082	0.072	0.054	0.073	0.073
UK	GDP	0.192	0.407	0.300	0.138	-0.088	-0.170	0.139	0.319	0.246	0.375	0.315	0.267
	INF	0.001	0.098	0.079	0.072	0.066	0.065	0.000	0.021*	0.028*	0.023	0.022	0.019
	INT	0.000	0.006	0.005	0.003	-0.001	-0.005	0.000	0.005*	0.008*	0.009*	0.010*	0.010*
	POL	-0.438*	-0.380	-0.232	-0.296	-0.466	-0.527	0.077	0.041	0.045	0.075	0.075	0.082

Note: The table displays the responses of output (GDP), price level (INF), interest rate (INT), and the fiscal counterpart (POL) to government spending and tax shocks. The symbol * indicates that the reactions are statistically different from zero.

3.A.6 Recursive approach

In this section we analyze the effects of fiscal policy shocks by applying a recursive identification strategy (see Fatás and Mihov (2001)).[98] Subsequently, we compare the obtained multipliers with the baseline estimations from the Balanchard and Perotti (2002) approach. The recursive approach implies a causal ordering of the variables. On these grounds, we assume the following relation between the reduced form innovations U_t and the structural shocks E_t:

$$\begin{bmatrix} 1 & 0 & 0 & -a_g^y & 0 \\ -a_y^p & 1 & 0 & -a_g^p & 0 \\ -a_y^i & -a_p^i & 1 & -a_g^i & -a_r^i \\ 0 & 0 & 0 & 1 & 0 \\ -a_y^r & -a_p^r & 0 & -a_g^r & 1 \end{bmatrix} \times \begin{bmatrix} u_t^y \\ u_t^p \\ u_t^i \\ u_t^g \\ u_t^r \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} e_t^y \\ e_t^p \\ e_t^i \\ e_t^g \\ e_t^r \end{bmatrix}. \quad (3.7)$$

This ordering scheme may be justified as follows: government expenditures do not respond to output changes within the same quarter due to implementation and decision lags. In addition, output and inflation have an immediate impact on the tax base and influence net taxes within the same quarter. Hence, these variables are ordered before net taxes. On the downside, this ordering rules out that net taxes do have a contemporaneous effect on output and inflation, respectively. Finally, interest rates are ordered last which implies that they are contemporaneously effected by all variables in the system.[99] A comparison of

[98] See also Almunia et al. (2010) who rely on comparable recursive identification strategies in order to study the effects of fiscal policy on economic activity in a panel VAR setup.

[99] This ordering scheme implies that net taxes are defined net of interest payments so that taxes do not respond to changes in the interest rate within the same quarter. Please note that in our estimations net taxes are calculated including property incomes, which include interest payments. However, property incomes account for only around 6% of total expenditures (see figure 3.2 in section 3.3) and should not

fiscal multipliers obtained from different identification strategies can be found in Caldara and Kamps (2008).

Figure 3.12 illustrates the impulse response pattern of GDP, price level, interest rate as well as the fiscal counterpart to a one unit increase in expenditures and taxes, respectively. Applying the recursive approach, it can be seen that in Belgium, France, and the United Kingdom expenditure shocks have a more pronounced effect on GDP. In the case of the United Kingdom, the impact expenditure multiplier even gets statistically significant taking a value of 1.48. Only in Germany, we find a lower fiscal multiplier at longer horizons that is statistically significant in the first two periods. Further, we find that the response of price levels and interest rates is mitigated in all countries emphasizing the negative reaction of interest rates to expenditure increases.

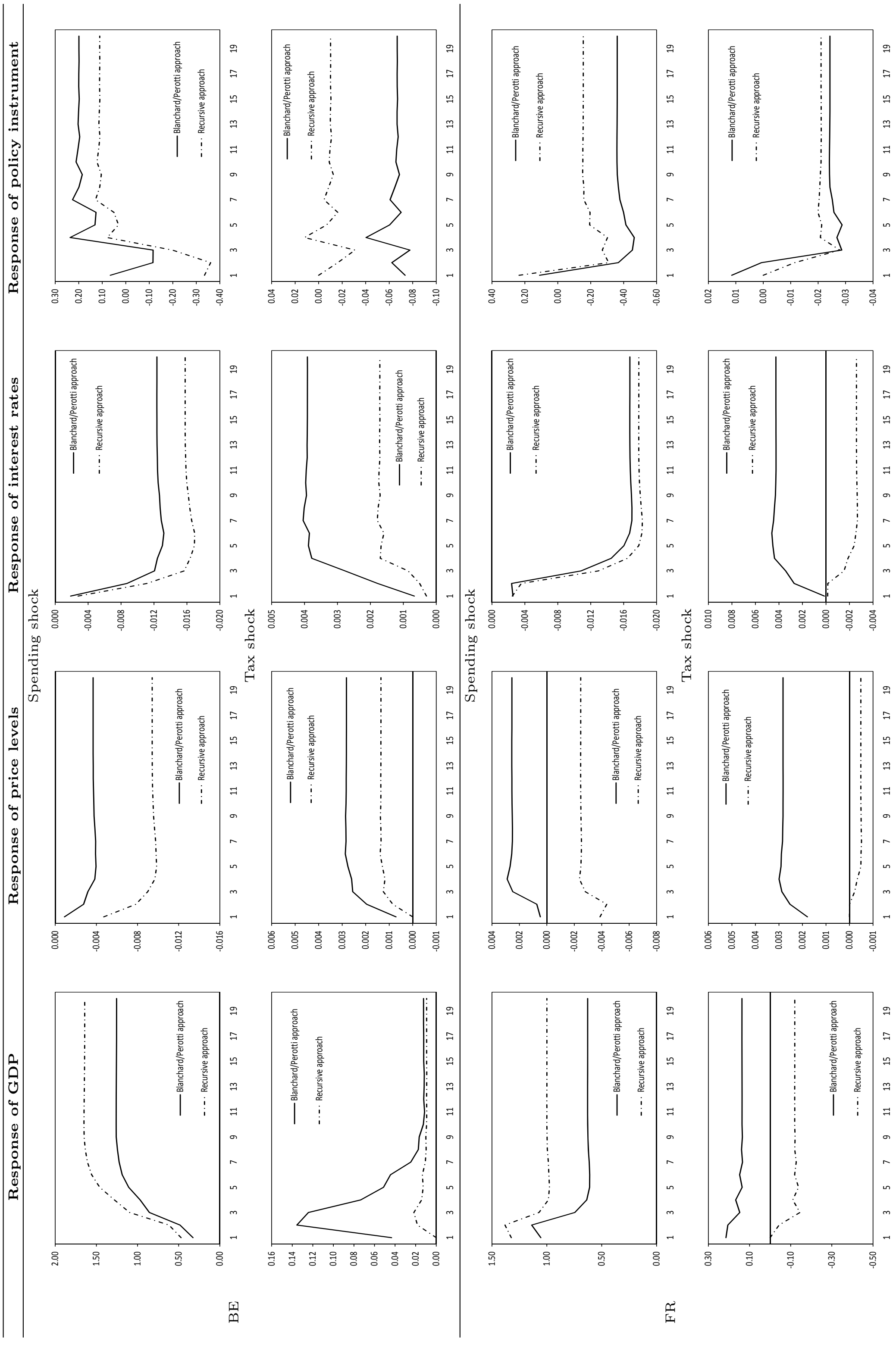
Regarding the contemporaneous effects of expenditure shocks, the only difference between the two identification strategies is that the Blanchard and Perotti (2002) approach assumes that inflation has a within-quarter effect on spending, while the recursive approach sets this contemporaneous effect to zero (see equation (3.4) and (3.7) in section 3.4 and 3.A.6). Hence, the impulse responses obtained from the recursive system might be interpreted as expenditure multipliers with zero inflation elasticity.

Although the response of output to tax increases remains unchanged regarding statistical significance, it can be seen that the impulse response pattern of Belgium, France, and the UK shift downwards in comparison to the ones obtained from the Blanchard and Perotti

substantially affect the results. The additional section 3.A.5 shows that the definition of net taxes has only a very little effect on the estimated fiscal multipliers.

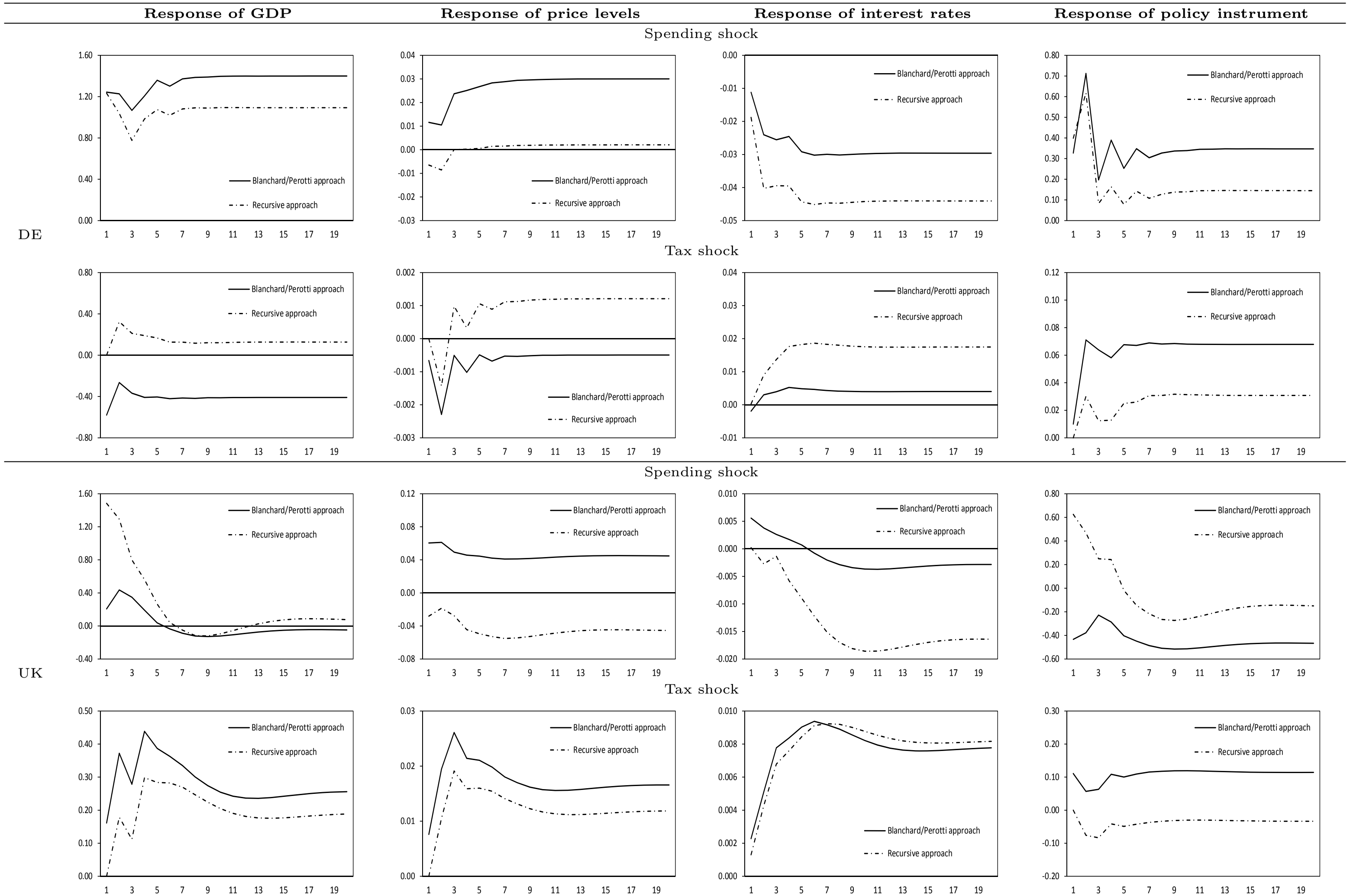
(2002) approach. Only in Germany we observe a decreasing efficiency of tax shocks. Overall the results provide evidence that the contemporaneous response of GDP to taxes is a crucial assumption for the level and sign of the estimated tax multipliers.

Figure 3.12: Response of macroeconomic variables to fiscal shocks under alternative identification strategies



Note: Table is continued on the next page.

3.12: Response of macroeconomic variables to fiscal shocks under alternative identification strategies - continued



Note: The figure displays the responses of macroeconomic variables to a one unit increase in public spending and net taxes in Belgium (BE), France (FR), Germany (DE), and the United Kingdom (UK) under alternative identification strategies.

4 On debt sustainability in the EMU: an analysis of asymmetric persistence in public deficits

4.1 Introduction

Using time-series techniques, sustainability of public debt has been studied in recent years by many authors.[100] The prevailing empirical literature employs unit root and cointegration tests to verify if time paths of deficits are consistent with the government's intertemporal budget constraint. However, common studies testing for debt sustainability treat the persistence of positive and negative deficit changes with equal measure. We illustrate that these symmetric testing approaches can be adapted more effectively in an asymmetric unit root setting, reducing the likelihood of drawing false inferences on debt sustainability and improving the power of the test. Therefore, we introduce a modified asymmetric testing

[100] See table 4.6 in section 4.A.2 for an overview.

procedure giving the possibility to assess fiscal sustainability more precisely.

Recent empirical studies apply miscellaneous unit root and cointegration tests on different data sets to substantiate fiscal sustainability for various groups of countries.[101] Bohn (2007) shows that such commonly used tests are incapable of rejecting fiscal sustainability. Therefore, only a rejection of the unit root and no cointegration null leads to the inference that countries do stay within their intertemporal budget constraint, concluding that the expected present value of future primary surpluses does equal the initial debt. We argue that even in the case of a non-rejection of unit root behavior, fiscal policy can be confirmed as being consistent with the intertemporal budget constraint as long as positive deficit changes are transitory.[102]

In order to capture the asymmetric persistence in a series, we propose - based on Enders' and Granger's (1998) momentum threshold autoregressive model (MTAR) - an asymmetric unit root testing approach, analyzing positive and negative deficit changes separately. This asymmetric approach gives the possibility of testing fiscal sustainability more precisely by separating between the persistence of positive and negative deficit ratio changes. Now, fiscal policy is considered to be sustainable as long as the deficit series is either global stationary or positive deficit changes follow a mean reverting pattern regardless whether the deficit series has a global unit root.

[101] See for instance Westerlund and Prohl (2010), Afonso and Rault (2010), Holmes (2010), Holmes et al. (2010), Prohl and Westerlund (2009), Ehrhart and Llorca (2008), Llorca and Redzepagic (2008) as well as Ghatak and Sanchez-Fung (2007). A detailed overview about recent studies on European countries is provided in table 4.6 in section 4.A.2 in the appendix

[102] The cointegration approach also exhibits an equivalent inaccuracy when the null of cointegration and accordingly the sustainability hypothesis is rejected due to higher revenues than expenditures. In this case, the transversality condition would also hold despite the imbalance between outlays and revenues. A cointegration analysis of public spending and revenue can be found in section 4.A.4 in the appendix.

The empirical framework of testing fiscal sustainability is based on Hamilton and Flavin (1986) and was improved by several authors during the late eighties and mid-nineties.[103] This time period was characterized by continuous global rising deficit ratios as a result of the energy crisis in the late eighties. Therefore, a non-rejection of the unit root null was always by default equated with an increase in the deficit ratio time series and thus a possible violation of the intertemporal budget constraint. Nevertheless, this inference seems to be debatable because during the 2000s, periods of overall declining deficits were gaining in importance. Studies from Hamori and Hamori (2009) and Greiner et al. (2007) show that fiscal rules and years of high deficits led to a decay of high and continuous deficit spending in the following years. Especially - in the case of Europe - the Treaty of Maastricht and the subsequent introduction of a common currency should have contributed to lower deficit levels.

Overall, our empirical results show that at least 10 out of 14 member countries of the EMU adhere to a long-run sustainable fiscal policy. These results can be explained by the successive cutback of deficit ratios until the turn of the century. This downward sloping curve of deficit series might give an alternative explanation why prior studies on fiscal sustainability employing unit root or cointegration approaches commonly do not verify sustainability for most of the European countries even though this does not appear to be compatible with the introduction of fiscal policy rules, the Maastricht Treaty, and the Stability and Growth Pact.

[103] See amongst others Trehan and Walsh (1988, 1991), Wilcox (1989), and Quintos (1995). A brief discussion on the evolution of empirical literature on assessing debt sustainability can be found in section 4.A.2 in the appendix.

The remainder of this chapter is organized as follows. Section 4.2 briefly outlines the theoretical background of testing debt sustainability. Subsequently, based on a simple unit root analysis of 14 countries in the EMU, section 4.3 shows that symmetric linear unit root tests might lead to incorrect inferences. Section 4.4 introduces a MTAR model and develops a new two-step approach of testing fiscal sustainability. In order to evaluate the power of the new testing procedure, a Monte Carlo experiment is conducted. In section 4.5 the asymmetric two-step approach is used to analyze the sustainability of 14 countries in the EMU. Section 4.6 concludes.

4.2 Debt sustainability and the intertemporal government budget constraint

The literature provides several concepts for defining and valuating debt sustainability. While early views focused on public debt and its general effects on the economy and generational distribution, the current discussion is driven by conceptual and methodological similarities to the sustainability analyses in resource economics.[104] The intuition behind debt sustainability is straightforward: "a borrower is expected to be able to continue servicing its debts without an unrealistic large future correction to the balance of income and expenditure" (IMF (2002), p. 4). Hence, the basic requirement for debt sustainability is that governments should stay within their intertemporal budget constraint without getting

[104] For a detailed look on the conceptual and formal development of debt sustainability see Neck and Sturm (2008).

bog down in a Ponzi finance scheme.

The theoretical concept of fiscal sustainability can be derived by the starting point of the budget identity. The with-interest deficit DEF_t is defined as the difference between outlays and revenues plus an interest charge. It also equals the change in public debt. Thus, in algebraic terms, the budget identities are

$$DEF_t = G_t - T_t + i_t \cdot D_{t-1} \tag{4.1}$$

and

$$D_t = D_{t-1} + DEF_t, \tag{4.2}$$

where D_t is the public debt, T_t the total revenues, G_t non-interest spending, Y_t the nominal GDP, and i_t the interest charge at the end of period t. From (4.1) and (4.2) the nominal budget equation can be rewritten as a GDP-ratio version:

$$\frac{D_t}{Y_t} = \frac{G_t - T_t}{Y_t} + (\frac{1 + i_t}{1 + g_t})\frac{D_{t-1}}{Y_{t-1}}, \tag{4.3}$$

where g_t denominates the nominal GDP-growth rate. Rearranging equation (4.3) in a scaled version the public debt can be described as

$$d_t = (1 + \eta_t)d_{t-1} - s_t, \tag{4.4}$$

with d_t denoting the debt-to-gdp ratio $\frac{D_t}{Y_t}$, s_t corresponds to the primary surplus $\frac{G_t - T_t}{Y_t}$, and η_t denotes the appropriate version on the "return" of the debt with $\eta_t = \frac{1+i_t}{1+g_t} - 1 \approx i_t - g_t$. From equation (4.4) sequences of primary surpluses and interest charges compute the path of public debt and provide the basis for sustainability:

$$d_{t+n} = [\prod_{k=0}^{n}(1+\eta_{t+n})]d_{t-1} - \sum_{j=0}^{n}[\prod_{k=j+1}^{n}(1+\eta_{t+n})s_{t+j}]. \tag{4.5}$$

Assuming a constant return and reformulate equation (4.5) by taking expectations,

$$E_t[d_{t+n}] = (1+\eta)^n d_t^* - \sum_{j=0}^{n}(1+\eta)^{n-j}E_t[s_{t+j}], \tag{4.6}$$

where $d_t^* = (1+\eta)d_{t-1}$ denominates the debt-to-gdp ratio of period t. Finally rearrange for d_t^* and generate the limiting value for $n \to \infty$:

$$\lim_{n\to\infty} d_t^* = \sum_{j=0}^{\infty}\frac{1}{(1+\eta)^j}E_t[s_{t+j}] + \lim_{n\to\infty}\frac{1}{(1+\eta)^n}E_t[d_{t+n}]. \tag{4.7}$$

Under the assumption that the discounted future debt (transversality condition) converges to zero, the initial debt equals the expected present value of future primary surpluses and the governmental intertemporal budget constraint (IBC) holds. Empirical investigations focus on showing that budget deficit ratios follow a stationary process to satisfy "ad hoc sustainability" (Bohn (2008), p. 22). Only if the deficit ratio is stationary, the debt ratio increases at most straight proportional and the expected future debt-ratio converges to zero. The empirical investigations address the question whether the government's creditors could rationally expect that the government budget would be balanced in present-value terms. This broaden definition implies that running substantial deficits over a long period of time can be in accordance with sustainability as long as these deficits can be repaid by adequately high future surpluses.[105]

4.3 Symmetric deficit adjustment and the sustainability hypothesis

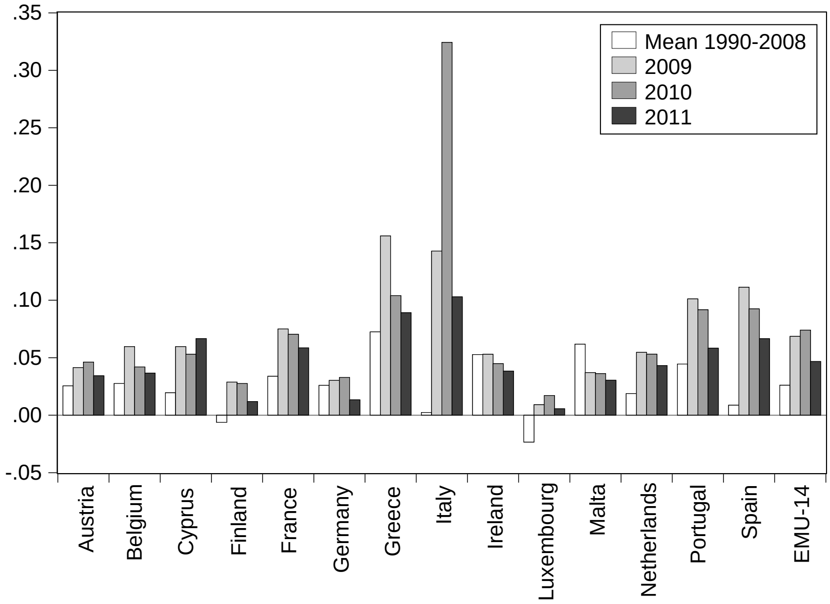
The breakout of the global financial crisis in 2008 aggravated the financial position of many governments in the last three years dramatically. Especially within the EMU the fears of a sovereign debt crisis rose continuously and as illustrated in figure 4.1, the years 2009, 2010, and 2011 are distinctly characterized by very high deficit-to-gdp ratios.[106] Malta is the only country where the deficit ratios in 2009 and 2010 are below the average of the last twenty years. All other countries exhibit significantly higher deficits, whereby Ireland, Greece, Spain, Portugal, and France show the highest annual new indebtedness.

As shown by Trehan and Walsh (1988, 1991) the governmental intertemporal budget balance holds, if the with-interest deficit follows a stationary process. Therefore, a battery of unit root tests is used in an attempt to validate the sufficient sustainability condition,

[105] In order to account for this shortcoming, we analyze in section 4.A.5 in the appendix the long-run steady state values of public debt for various time periods across Euro member countries, based on the model by Domar (1944). The results show that in the majority of European countries debt ratios of more than 100% would arise, if the government spending behavior would not change at the current GDP growth rates.

[106] Slovak Republic and Slovenia are the only countries left out in the EMU due to few data points. Slovak Republic and Slovenia were newly-established in 1993 and 1991, respectively.

Figure 4.1: Deficit-to-gdp ratio in 14 countries of the EMU



Note: The graph displays the annual with-interest deficit as a share of GDP for 14 member countries of the EMU. Data is taken from European Commission (2013).

using public deficit as a percentage of GDP. The largest part of the annual data comes from the European Commission AMECO (Annual Macro-Economic Data) database. On account of short time series, missing data points were calculated by the use of IMF's Government Financial Statistic (GFS) database and the 2001 manual (GFSM 2001).[107] Overall, our analysis covers a time period from 1972 to 2011.

Table 4.1 reports the Phillips-Perron (PP), the Augmented Dickey-Fuller (ADF), and the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test results for a pre-crisis sample (1972 to 2008) and for a crisis sample (1972 to 2011). All test statistics are calculated including a constant in the test equation. Akaike information criterion (AIC) is used for choosing the lag structure of the ADF test equation, while the bandwidth for the PP and KPSS test is

[107]More information on the data sources and data calculation can be found in section 4.A.1 in the appendix.

selected based on Newey-West using Bartlett kernel.[108]

Table 4.1: Results of unit root tests on deficit-to-gdp ratio

	Pre-crisis sample					**Crisis Sample**				
	PP test		ADF test		KPSS test	PP test		ADF test		KPSS test
Country	t-stat.	Prob.	t-stat.	Prob.	LM-stat.	t-stat.	Prob.	t-stat.	Prob.	LM-stat.
Austria	-3.417	0.017	-3.494	0.014	0.195	-3.714	0.008	-3.827	0.006	0.170
Belgium	-1.141	0.689	-0.794	0.809	0.408*	-1.412	0.567	-1.183	0.672	0.396*
Cyprus	**-2.345**	**0.164**	**-2.345**	**0.164**	**0.421***	**-3.034**	**0.040**	**-3.263**	**0.024**	**0.165**
Finland	**-2.024**	**0.276**	**-2.511**	**0.122**	**0.387***	**-2.779**	**0.071**	**-2.779**	**0.071**	**0.192**
France	-2.538	0.115	-2.673	0.089	0.445*	-2.227	0.200	-2.409	0.146	0.525**
Germany	-4.029	0.004	-4.116	0.003	0.061	-4.370	0.001	-4.554	0.008	0.055
Greece	-1.886	0.335	-2.580	0.108	0.195	-2.313	0.173	-2.313	0.014	0.255
Ireland	-1.258	0.638	-1.697	0.424	0.455*	-2.502	0.123	-1.595	0.475	0.172
Italy	-1.204	0.662	-1.157	0.682	0.289	-1.351	0.596	-1.365	0.589	0.314
Luxembourg	-2.855	0.061	-4.244	0.002	0.065	-2.638	0.094	-4.353	0.001	0.139
Malta	-2.158	0.225	-2.117	0.239	0.594**	-2.266	0.188	-2.239	0.197	0.504**
Netherlands	**-2.449**	**0.136**	**-2.567**	**0.109**	**0.348***	**-3.019**	**0.042**	**-3.070**	**0.037**	**0.167**
Portugal	-3.606	0.011	-3.682	0.009	0.733**	-3.953	0.004	-3.738	0.007	0.342
Spain	-3.546	0.012	-4.228	0.002	0.169	-2.982	0.046	-4.054	0.003	0.119

Note: The Phillips-Perron (PP), the Augmented Dickey-Fuller (ADF) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test are calculated including a constant in the test equation. The bandwidth for the PP and KPSS test is selected based on Newey-West using Bartlett kernel, while the lag length of the ADF test is selected based on AIC. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from Kwiatkowski et al. (1992).

The test results - including the high deficit years 2009, 2010, and 2011 - show that the deficit series of Austria, Cyprus, Finland, Germany, Luxembourg, Netherlands, Portugal, and Spain can be treated as stationary. Consequently, these eight member countries of the EMU follow a sustainable fiscal policy path, while for the remaining six countries (Belgium, France, Greece, Ireland, Italy, and Malta) a sustainable fiscal budget management cannot be confirmed. However, when comparing the results of the full sample (including the crisis periods) and the pre-crisis sample, the empirical outcome appears contradictory. Surprisingly, the test statistics in table 4.1 indicate that in a debt crisis sample (1972-2011) more countries confirm fiscal sustainability than in a time period excluding the crisis (1972-

[108] The unit root results stay qualitatively unchanged if we additionally include a trend in the test specification. Furthermore, the results are robust to alternative lag specifications.

2008). Unit root behavior is rejected for the countries Cyprus, Finland, and Netherlands in a sample period from 1972 to 2011 but rejects sustainability in the pre-crisis sample from 1972 to 2008. Generally speaking, the empirical findings suggest that in 2011 more countries follow a sustainable fiscal policy path than in 2008, although the above average deficit-to-gdp ratios in figure 4.1 indicate that the last three years certainly did not contribute to a budgetary relaxation.[109]

An explanation for the observed contradictory results might be seen in the brief characterization of government's deficit-to-gdp ratio pictured in figure 4.3 in section 4.A.1 in the appendix. Visual inspections of the deficit time series for each country suggest that deficit-to-gdp ratios might not follow a symmetric adjustment process over the last 40 years. As a response to the potential financial and economic instability stemming from the collapse of the Bretton Woods system in the wake of the oil crisis, deficit ratios were rising all over the Euro area during the seventies. After a short budgetary relaxation phase until the early eighties, deficit ratios followed again a rising process triggered by the second oil crisis. As a consequence, a majority of countries exhibited very high deficit ratios during the eighties and at the early stage of the nineties. Until the burst of the dotcom bubble in 2000, a phase of consistent decreasing - partially sharply falling - deficit ratios can be observed. Recently, the impact of the financial crisis in 2008 increased the level of deficit ratios for some countries to an even higher level than in the eighties and early nineties, respectively. In summary, it can be stated that the contradictory empirical results of this section can

[109] These results are robust to alternative pre-crisis sample sizes. In each case the empirical results including the high deficit years 2009, 2010, and 2011 confirm sustainability for the largest number of countries. Test results and further interpretations can be found in section 4.A.3 in the appendix.

be explained by the fact that in the sample from 1972 to 2008 in some European countries periods of decreasing deficit ratios dominated the data-generating process. Triggered by the financial crisis in 2008, deficit ratios increased dramatically in 2009, 2010, and 2011. As a result, the unit root test did not reject unit root behavior in the sample excluding the financial crisis - despite decreasing new indebtedness - and rejected unit root behavior in the sample including the crisis. The empirical sustainability literature commonly does not distinguish between the asymmetric persistence of positive and negative adjustment paths. In fact, deficit time series with positive or negative stochastic trends are treated equally. However, from an empirical point of view this specification might lead to erroneous inferences in terms of a balanced intertemporal budget. If deficit series follow a negative stochastic trend, the exponential rising discount rate still implies a zero limit of the transversality condition.

4.4 Threshold autoregressive model and partial unit root regime

Empirical framework

As seen in the previous section, it is important to distinguish between negative and positive changes in deficit-to-gdp ratio when assessing debt sustainability. The regression specification to test this hypothesis follows a Threshold Autoregressive (TAR) time-series approach, where the switch from one regime to another depends on the past values of the series (see

Tong and Lim (1980)).

A conventional time series test for persistence would regress the changes in a given variable on the lagged levels. Starting from a regression with the change of deficit, Δd_t, on the left hand side and the levels of deficit, d_{t-1}, on the right hand side:

$$\Delta d_t = c + \rho_1 d_{t-1} + \sum_{p=1}^{P} \theta_p \Delta d_{t-p} + \epsilon_t, \tag{4.8}$$

a $\rho_1 < 0$ would imply that in the long run the persistence of the shock would actually be equal to zero. This is consistent with the view that positive as well as negative deficit ratio changes are transitory. Furthermore, if ρ_1 is equal to zero, positive or negative deficit ratio changes would be persistent. To address the issue of asymmetric persistence of negative and positive changes in deficit-to-gdp ratios, we follow Enders and Granger (1998) and extend the simplified regression (4.8) by adding an interaction term, I_{t-1}, as a Heaviside step function that splits the regression into two regimes:

$$\Delta d_t = (c_1 + \rho_1 d_{t-1}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1}) \cdot (1 - I_{t-1}) + \sum_{p=1}^{P} \theta_p \Delta d_{t-p} + \epsilon_t. \tag{4.9}$$

An interaction term in the form of $I_{t-1} = 1$ if $\Delta d_{t-1} \geq \tau$ and $I_{t-1} = 0$ if $\Delta d_{t-1} < \tau$ corresponds to a momentum threshold autoregressive model and indicates that a change in the debt ratio above a defined attractor τ captures asymmetrically "sharp" movements in the time series on condition that $\rho_1 \neq \rho_2$. Setting this attractor $\tau = 0$ gives us the possibility to analyze positive and negative deficit ratio changes separately and gives information about the decay of positive respectively negative first differences of deficit ratios. If, for example, $|\rho_1| < |\rho_2|$, regression (4.9) exhibits little decay for positive deficit changes but substantial decay for negative changes. Equation (4.9) shows that positive changes in budget deficits would be persistent if $\rho_1 = 0$. Negative deficits such that expenditures exceed revenues are transitory if $\rho_2 < 0$.

Franses and van Dijk (2000) show that especially relatively simple non-linear time-series models may give rise to rather complicated autocorrelation structures that only can be captured by choosing lag orders p_1 and p_2 separately for each regime:

$$\Delta d_t = (c_1 + \rho_1 d_{t-1} + \sum_{j=1}^{p_1} \theta_j \Delta d_{t-j}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1} + \sum_{k=1}^{p_2} \theta_k \Delta d_{t-k}) \cdot (1 - I_{t-1}) + \epsilon_t. \quad (4.10)$$

The interpretation of the coefficients ρ_1 and ρ_2 remains unchanged compared to the regression in equation (4.9). The problem of choosing the right lag dynamics in each regime can be solved by using different information criteria. Tong (1990) defines an alternative AIC for a two-regime model as the sum of the AICs for the AR models in the two regimes.[110] Enders and Granger (1998) use the F-test to show that within a reasonable range of adjustment parameters, the power of the test $H_0 : \rho_1 = \rho_2 = 0$ for equation (4.9) is greater than that of the corresponding ADF test $H_0 : \rho_1 = 0$ for equation (4.8). However, Caner and Hansen (2001) deduce from equation (4.9) another case of interest. This is the intermediate

[110] $AIC(p_1, p_2) = n_1 \cdot ln(\hat{\sigma}_1^2) + n_2 \cdot ln(\hat{\sigma}_2^2) + 2(p_1 + 1) + 2(p_2 + 1)$ where n_i, i = 1, 2, is the number of observations in the ith regime, and $\hat{\sigma}_i^2$, i = 1, 2, is the variance of the residuals in the ith regime. Analogously the SIC for the two-regime model is defined as $SIC(p_1, p_2) = n_1 \cdot ln(\hat{\sigma}_1^2) + n_2 \cdot ln(\hat{\sigma}_2^2) + (p_1 + 1) \cdot ln(n_1) + (p_2 + 1) \cdot ln(n_2)$.

case of a partial unit root regime:

$$H_1 = \begin{cases} \rho_1 < 0 \quad and \quad \rho_2 = 0 \\ \\ \rho_1 = 0 \quad and \quad \rho_2 < 0. \end{cases} \tag{4.11}$$

If equation (4.11) holds, then the deficit series will behave like a unit root process in one regime, but will behave like a stationary process in the other one. While Enders and Granger (1998) report only critical values for the F-test, we propose the evaluation of the t-statistic within each regime to address the issue of a partial unit root regime. In fact, Shin and Lee (2001) show that in MTAR models, t-type tests for $H_0 : \rho_1 = \rho_2$ based on ordinary least square estimators have still the highest power despite a general downward bias of the estimated coefficients, especially in cases of near unit roots.

Monte Carlo experiment

To expose the properties of the asymmetric unit root test presented above, a Monte Carlo experiment is conducted to show the power of the test varying sample size and regime parameters. We start by drawing 10 000 normally distributed and uncorrelated pseudorandom numbers with standard deviation equal to unity to represent the ϵ_t. Setting the initial value of X_t equal to zero, 10 000 of time series with a switching regime were generated following Enders' and Granger's (1998) MTAR model:

$$\Delta X_t = \rho_1 X_{t-1} \cdot I_t + \rho_2 X_{t-1} \cdot (1 - I_t) + \epsilon_t, \tag{4.12}$$

where I_t represents the Heaviside step function such that

$$I_t = \begin{cases} 1 & if \quad \Delta X_{t-1} \geq 0 \\ \\ 0 & if \quad \Delta X_{t-1} < 0. \end{cases} \tag{4.13}$$

For each artificial MTAR model an auxiliary regression given by equations (4.12) and (4.13) was estimated by using OLS and two different test statistics were tabulated. The t-statistics for the null hypothesis $\rho_1 = 0$ and for the null hypothesis $\rho_2 = 0$ were recorded. In both cases the null of a unit root should be rejected as long as positive and negative first differences are transitory and the parameters ρ_1 and ρ_2 are smaller than zero. Table 4.2 reports results for the single threshold momentum model with altering adjustment parameters and sample sizes.

Table 4.2: Power of asymmetric unit root test

MTAR model		1000 Obs.		500 Obs.		100 Obs.		50 Obs.		30 Obs.	
ρ_1	ρ_2	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.	pos.	neg.
-0.9	0	100	87.5	100	89.0	100	90.1	100	91.6	99.5	93.9
-0.7	0	100	89.9	100	91.3	100	91.1	99.9	91.9	98.2	92.7
-0.5	0	100	91.9	100	90.9	100	91.5	99.3	92.1	93.2	92.0
-0.3	0	100	89.9	100	90.3	99.7	90.5	92.2	91.3	75.2	92.0
-0.1	0	100	91.4	100	90.6	90.3	81.2	62.3	89.1	46.2	89.7
0	0	89.2	89.0	88.5	88.7	89.6	88.8	89.6	89.5	89.2	89.8
-0.5	-0.5	100	100	100	100	99.8	99.8	91.9	91.7	73.5	73.5
-0.5	-0.9	100	100	100	100	99.8	100	91.1	99.8	70.8	96.9

Note: The entry is the percentage of instances for which the null hypothesis of a unit root is correctly rejected for $\rho_i < 0$ and for which the null hypothesis of a unit root is not correctly rejected for $\rho_i = 0$ at a significance level of 10% using critical values from the t-distribution. The positive regime (pos.) corresponds to $\Delta X_t = \rho_1 X_{t-1} + \epsilon_t$ for all $\Delta X_{t-1} \geq 0$ and the negative regime (neg.) relates to $\Delta X_t = \rho_2 X_{t-1} + \epsilon_t$ for all $\Delta X_{t-1} < 0$.

It can be seen that the asymmetric unit root test performs very well using the t-distribution even in a setting with very small sample sizes. For example, in a sharp asymmetric adjustment setting with $\rho_1 = -0.9$, $\rho_2 = 0$ and only 30 observations the test correctly indicates a stationary process in the positive regime in 99.5% and a unit root in the negative regime in 93.9%. Even in the case of a near-unit root in one regime, e.g. $\rho_1 = -0.1$, the test identifies stationary behavior at a sample size of 50 observations in 62.3%.[111]

Partial unit root regime, stationarity and the sustainability hypothesis

Even in the case of a partial unit root with $\rho_i = 0$, the global series is always stationary as long as ϵ_t is i.i.d. and the distribution of ϵ_t is absolutely continuous.[112] Despite the partial unit root, all shocks - positive as well as negative - are transitory in the long run. The unit root regime merely leads to a delayed decay of the initial shock. Hence, the emphasis of the unit root regime on the pattern of the global time series is considerably stronger than the stationary regime. Consequently, in the short-run, shocks are persistent as long as the first differences of deficit ratios belong to the unit root regime and the algebraic sign of deficit changes stays unchanged. However, the local stationarity within one regime always leads to the mean reversion of all shocks in the long run. This interaction of both regimes and their effect on the global pattern of the deficit series is crucial for the further assessment of debt sustainability.

In terms of the intertemporal budget constraint, sustainability of public debt can be con-

[111] For comparison the standard Dickey-Fuller test correctly rejects the null hypothesis in a symmetric set up with $\rho_1 = \rho_2 = -0.1$ and 100 observations only in 53.56% (see Enders and Granger (1998)).

[112] Further assumptions are that $E[\epsilon_t] < \infty$ and ϵ_t is not degenerate at 0. For a detailed look see Lee and Shin (2001).

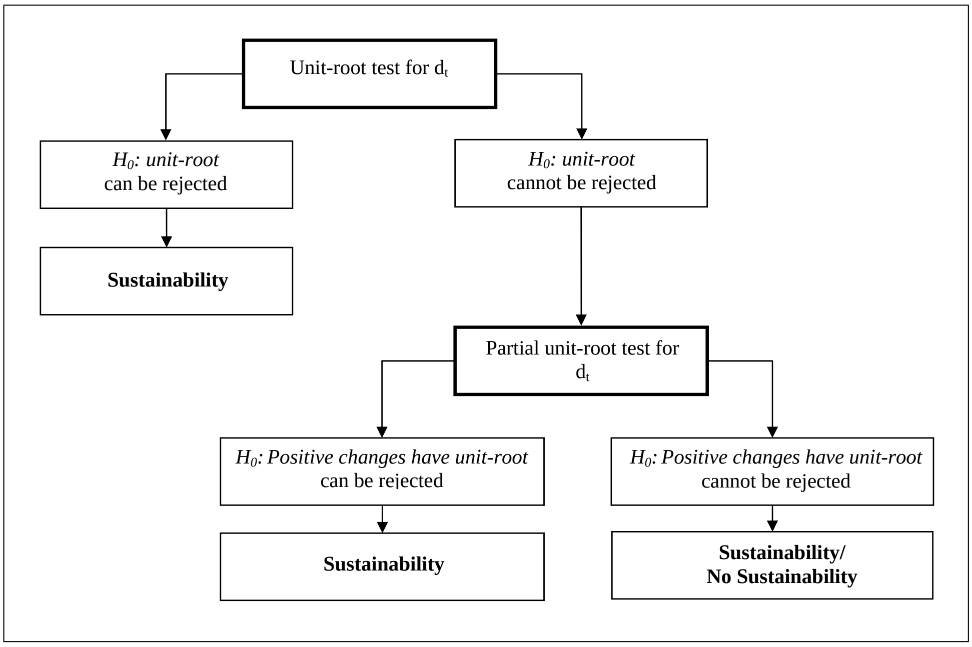
firmed if the deficit series is either global stationary or has a negative stochastic trend. In consideration of the partial unit root test, the government's fiscal budget is always balanced - even if the deficit series is not globally stationary - as long as positive deficit ratio changes are transitory. The regime of transitory positive deficit changes within a global non-stationary time series displays a global negative stochastic trend or at least a mean reversion of positive outliers and hence a zero limit of the transversality condition. On the contrary, persistent positive movements do not automatically indicate a violation of the intertemporal budget constraint, because on the one hand a stationary negative regime might still lead to a decay of all shocks in the long run, on the other hand a persistent negative regime could also induce to an overall declining deficit series.[113]

Therefore the procedure to assess the sustainability of the intertemporal government budget constraint involves a modified stepwise testing approach. In a first step, governments deficit ratios are tested on stationarity as before. A rejection of unit root behavior implies debt sustainability. In a second step, all non-stationary deficit series are tested on partial unit roots by using the MTAR approach as described above. Mean reverting behavior of positive deficit changes evince fiscal sustainability although the deficit series are not globally stationary. However, a non-rejection of unit root behavior in the positive regime does not induce by default a breach of the intertemporal budget constraint. In this case the deficit series can either have a declining or increasing character. In such cases, the sustainability hypothesis can neither be rejected nor confirmed. The modified procedure

[113]Complementary, Bohn (2007) shows that the transversality condition even holds if the deficit series follows a positive polynomial growth path of any order. The exponential growth of the discount factor will asymptotically dominate the growth of the expected future debt, whenever the with-interest deficit is integrated of any arbitrary order. Unit root test are not valid to find evidence against sustainability.

to test the sustainability of fiscal policy is summarized in figure 4.2.

Figure 4.2: Two-step approach for testing fiscal sustainability



4.5 A two-step approach for testing sustainability in the Euro area

In this section, we apply the two-step approach to the deficit ratios from section 4.3. The first step of testing the series for level stationarity is already done. As a reminder, the countries Belgium, France, Greece, Ireland, Italy, and Malta did not allow to reject unit root behavior. Although the deficit ratios of this countries are not global stationary, fiscal sustainability cannot per se be ruled out. Yet, in order to determine which country satisfies the intertemporal budget constraint, in a second step, the asymmetric unit root approach from the previous section is used.

Table 4.3: Results of asymmetric unit root test on deficit-to-gdp ratio (crisis sample)

		Model (1)			Model (2)			
Country	Period	p	t-stat.	Prob.	p_1	p_2	t-stat.	Prob.
Belgium	1972-2011	1	-0.870	0.391	1	0	-0.636	0.529
France	1972-2011	0	-0.754	0.456	0	0	-0.754	0.456
Greece	1972-2011	0	-1.981	0.056	0	0	-1.981	0.056
Ireland	1972-2011	1	0.606	0.549	0	0	1.433	0.161
Italy	1973-2011	2	-0.462	0.648	2	0	-0.775	0.445
Malta	1972-2011	0	-2.626	0.014	0	0	-2.626	0.014

Note: Model (1): t-statistics of OLS estimation of ρ_1 : $\Delta d_t = (c_1 + \rho_1 d_{t-1}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1}) \cdot (1 - I_{t-1}) + \sum_{p=1}^{P} \theta_p \Delta d_{t-p} + \epsilon_t$. Model (2): t-statistics of OLS estimation of ρ_1 : $\Delta d_t = (c_1 + \rho_1 d_{t-1} + \sum_{j=1}^{p_1} \theta_j \Delta d_{t-j}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1} + \sum_{k=1}^{p_2} \theta_k \Delta d_{t-k}) \cdot (1 - I_{t-1}) + \epsilon_t$. Definition of variables: Δd_t, change in deficit-to-gdp ratio. $\rho_1 = 0$, increasing deficits at the ratio of GDP would be persistent. $I_{t-1} = 1$ if $\Delta d_{t-1} \geq 0$; $= 0$ otherwise.

Table 4.3 shows the estimation results of the asymmetric unit root test using regression (4.9) and (4.10). The asymmetric approach allows to reject the null hypothesis that positive deficit changes are persistent for the countries Greece and Malta at least at a 10% significance level. The visual inspection of Greece's deficit ratio data demonstrates that increasing deficits have a mean reverting trend. In particular, the negative deficit change from 2009 to 2011 leads to the non-rejection of the sustainability hypothesis. By applying the asymmetric unit root test, it can be seen in figure 4.3 in section 4.A.1 in the appendix that the declining deficit ratios from the mid-nineties lead to a significant rejection of unit root behavior in the positive regime in a time interval from 1972 to 2006 and 1972 to 2007, respectively. The subsequent drastic increase of deficit spending in 2008 and 2009 leads to a non-rejection of unit root behavior in the regime of positive deficit changes and puts Greece's fiscal policy on an unstable path. Triggered by extensive government austerity programs, the deficit series sharply declines in 2010/2011 and consequently induces a rejection of non-stationarity and a return to a sustainable fiscal policy.[114] Malta's deficit series also displays a mean reverting trend since the beginning of the early nineties. Additionally, Malta is the only country within the EMU that has a below average deficit during the crisis years 2009, 2010, and 2011. Belgium, France, Ireland, and Italy do not allow to reject unit root behavior in the positive regime. Thus, the sustainability hypothesis cannot be confirmed. A violation of the intertemporal budget constraint, however, cannot be derived from these results because persistent positive deficit changes do not necessarily lead to an upwards tendency of the deficit series.

After all, the modified two-step approach strongly hints that in a time period from 1972 to 2011 government budgets are balanced in present-value terms for the countries Austria, Cyprus, Finland, Germany, Greece, Luxembourg, Malta, Netherlands, Portugal, and Spain. Only for the countries Belgium, France, Ireland, and Italy no sustainability statement can be made. The analysis in this chapter shows that despite the financial crisis in 2008 and the corresponding presumable debt crisis in the Euro area for at least 10 out of 14 countries sustainability of fiscal policy can be confirmed. Using the asymmetric two-step approach, table 4.4 shows a comparable pre-crisis analysis for the years 1972 to 2008. Now, the sustainability hypothesis can be confirmed for the countries Austria, France, Germany, Luxembourg, Malta, Netherlands, Portugal, and Spain. A final statement about the budget management of the remaining states cannot be derived from these asymmetric

[114] The symmetric testing procedure indicates a sustainable fiscal policy for Greece in the sample from 1972 to 2004 but rejects sustainability in samples including the subsequent years. These results seem to be elusive because the years 2005, 2006, and 2007 are characterized by below average and declining deficit ratios. In contrast, the asymmetric test provides more traceable results.

Table 4.4: Results of asymmetric unit root test on deficit-to-gdp ratio (pre-crisis sample)

			Model (1)				Model (2)	
Country	Period	p	t-stat.	Prob.	p_1	p_2	t-stat.	Prob.
Belgium	1972-2008	1	-0.310	0.759	1	0	0.044	0.965
Cyprus	1972-2008	1	-1.562	0.128	0	0	-1.563	0.128
Finland	1972-2008	4	-0.767	0.449	1	0	-2.011	0.053
Greece	1972-2008	0	-1.675	0.104	0	0	-1.675	0.104
Ireland	1973-2008	0	-1.470	0.152	0	0	-1.470	0.152
Italy	1972-2008	2	-0.046	0.964	2	0	-0.505	0.617
Malta	1973-2008	0	-2.540	0.017	0	0	-2.540	0.017
Netherlands	1972-2008	1	-3.251	0.003	1	0	-2.506	0.018

Note: Model (1): t-statistics of OLS estimation of $\rho_1 : \Delta d_t = (c_1 + \rho_1 d_{t-1}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1}) \cdot (1 - I_{t-1}) + \sum_{p=1}^{P} \theta_p \Delta d_{t-p} + \epsilon_t$. Model (2): t-statistics of OLS estimation of $\rho_1 : \Delta d_t = (c_1 + \rho_1 d_{t-1} + \sum_{j=1}^{p_1} \theta_j \Delta d_{t-j}) \cdot I_{t-1} + (c_2 + \rho_2 d_{t-1} + \sum_{k=1}^{p_2} \theta_k \Delta d_{t-k}) \cdot (1 - I_{t-1}) + \epsilon_t$. Definition of variables: Δd_t, change in deficit-to-gdp ratio. $\rho_1 = 0$, increasing deficits at the ratio of GDP would be persistent. $I_{t-1} = 1$ if $\Delta d_{t-1} \geq 0$; $= 0$ otherwise.

unit root test results. While the symmetric testing procedure validates sustainability only for six countries, the two-step approach approves budget sustainability for eight countries without equating a rejection of the unit root null with a violation of the intertemporal budget constraint. In comparison to the previous sustainability results including the high deficit years 2009 and 2010, the asymmetric two-step approach obtains much more traceable results for a pre-crisis analysis than the symmetric unit root test does. However, the empirical findings do not prove the fact that the current financial crisis and the associated high deficit ratios over the last years impaired the government's budgetary positions within a violation of the intertemporal budget constraint. This seems to be the cause of decreasing deficit ratios since the introduction of the Maastricht Criteria in 1992.

Generally speaking, these findings confirm with other empirical results from Greiner et al. (2007), who suggest that fiscal policies in several European countries are sustainable in a time period from around 1960 to 2003 due to a decline in public spending, which generates primary surpluses in the subsequent years of high deficits. Additionally, Hamori and Hamori (2009) support the notion that fiscal discipline rules are functioning and individual countries' fiscal deficits are steadily shrinking in the Euro area. Governments take corrective actions as a result of rising deficits over the last 30 years. Especially, the introduction of the Maastricht convergence criteria seems to have a positive effect on the response of governments to years of high deficit ratios.

The asymmetric testing procedure provides an possible explanation why prior studies on fiscal sustainability including samples at the turn of the century commonly do not corroborate sustainability for most of the European countries, whereas more recent studies provide evidence for a sustainable budget management.[115] Uctum and Wickens (2000), Bravo and Silvestre (2002) as well as Afonso (2005) tend to find evidence against sustainable fiscal policy for most of the countries, while on the other hand more recent studies from Prohl and Westerlund (2009), Holmes et al. (2010), and Afonso and Rault (2010) provide evidence in favor of sustainability. Certainly, the evolutionary process of the econometric methodology - especially the introduction of panel applications - seems to be a reasonable cause for these mixed findings, however, the development of new indebtedness indicates that the last ten years are generally characterized by rising deficits (see figure 4.3). With this in mind the empirical findings in the literature seem to be somehow inconsistent and the asymmetric approach renders a method of resolution against these contradictory results.

[115] An detailed overview of studies on European countries using unit root or assimilable cointegration approaches can be seen in table 4.6 in section 4.A.2 in the appendix.

4.6 Conclusion

Symmetric testing approaches on debt sustainability can be adapted more effectively in an asymmetric unit root setting. Therefore, an asymmetric stepwise testing procedure was developed to assess debt sustainability more precisely. This modified empirical approach has the advantage of reducing the likelihood of drawing false sustainability inferences and improving the power of the test. Now fiscal policy is considered to be sustainable as long as the deficit series is either global stationary or positive deficit changes follow a mean reverting trend regardless whether the series has a global unit root. The use of the test is exemplified by 14 countries in the Euro area, finding more consistency with the intertemporal budget constraint throughout the EMU compared to the earlier literature. These results can be explained by the successive cutback of deficit ratios since the introduction of the Maastricht criteria and the missing feature of symmetric unit root tests to identify this decay of deficits as sustainable fiscal policy. Hence, the allowance for asymmetric persistence in public deficits sheds additional light on reasons why prior and recent empirical studies covering European countries might come to divergent conclusions.

4.A Appendix

The following sections provide complementary information and robustness checks. At the outset, **section 4.A.1** provides further information about the data sources and the definition of fiscal deficits. Additionally, the development of the deficit-to-gdp ratios from 1972 to 2011 for various countries of the EMU is visualized. Subsequently, **section 4.A.2** outlines the empirical methodologies on assessing debt sustainability and discusses existing empirical evidence for European countries. **Section 4.A.3** presents recursive unit root estimations and shows that the unit root results presented in the main part of this chapter are robust to alternative pre-crisis sample sizes. Afterward, **section 4.A.4** comprises a supplementary cointegration analysis between government expenditures and revenues and discusses the results with regard to weak and strong sustainability. Finally, **section 4.A.5** provides estimation on the long-run steady state values of public debt for various time periods across Euro member countries based on the model by Domar (1944).

4.A.1 Data sources and definitions

Most of the deficit-to-gdp ratio data is taken from the European Commission (2013) AMECO database. Net lending/net borrowing (AMECO code: UBLG) of general government is defined following the Government Financial Statistics Manual (GFSM) 2001 classification.

As shown in table 4.5, missing data points are calculated by using IMF's Government Finance Statistics (GFS) (see IMF (2012a)). The main GFSM 1986 categories are classified to the main GFSM 2001 aggregates as follows:

Table 4.5: Calculation of deficit-to-gdp ratios

GFSM 2001 category	**GFSM 1986 category**
1 Revenue	Total revenue and grants, minus sales of fixed capital assets, sales of stocks, and sales of land and intangible assets
2 Expense	Current expenditure plus capital transfer
3 Net acquisition of non-financial assets	Acquisition of fixed capital assets plus purchases of stocks purchases of land and intangible assets minus sales of fixed capital assets minus sales of stocks minus sales of land and intangible assets

Figure 4.3: Deficit-to-gdp ratios in the Euro Area

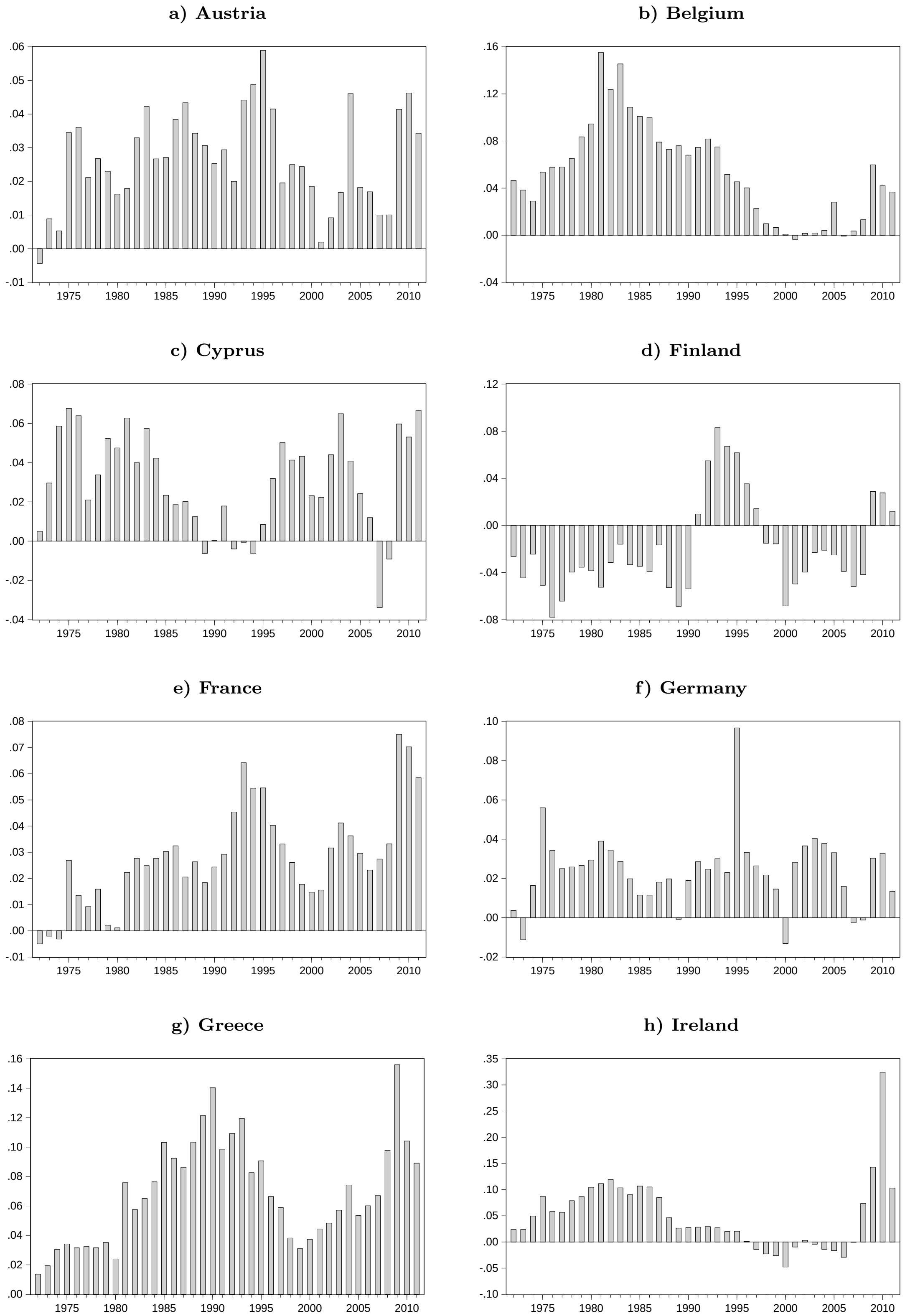
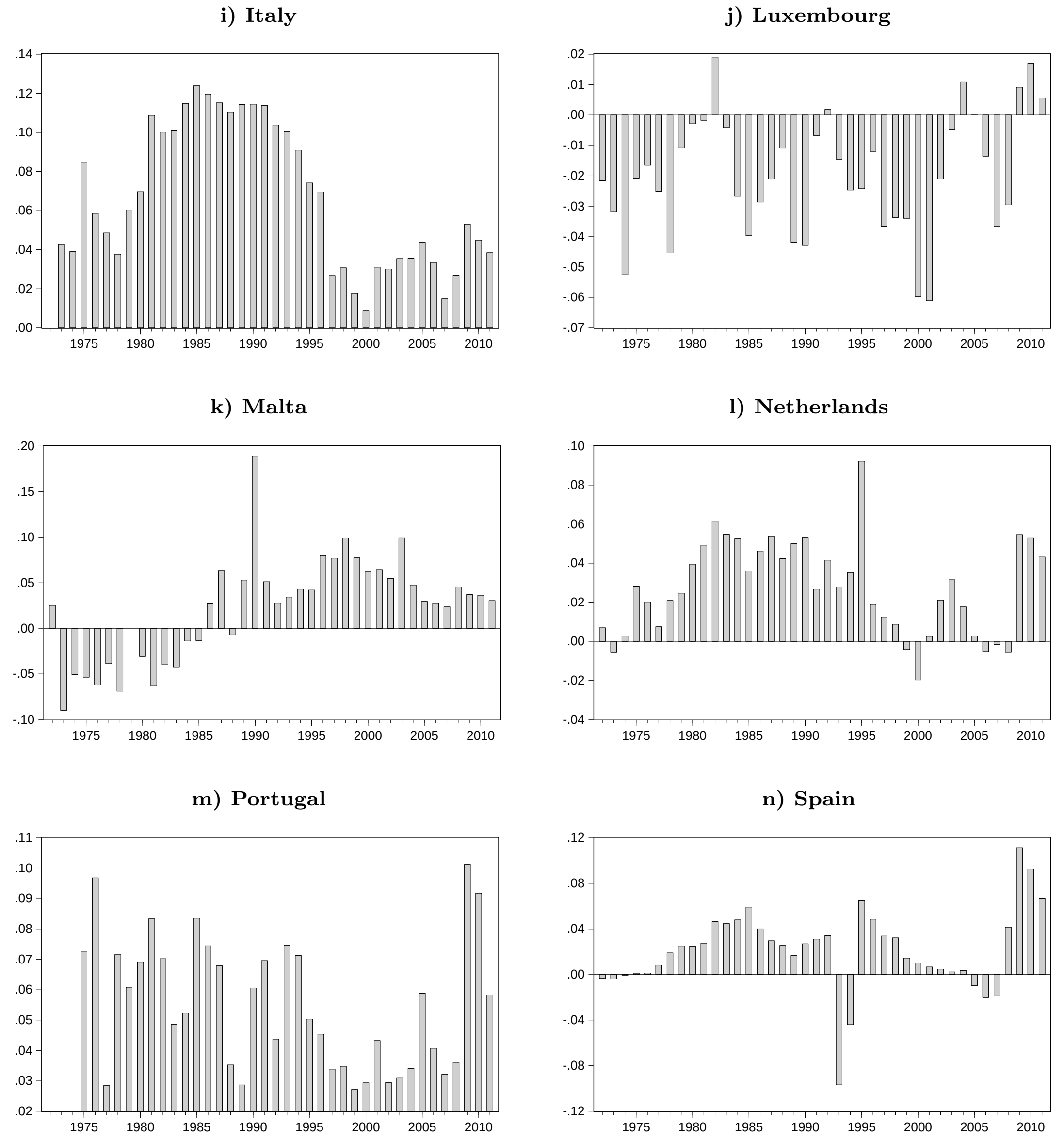


Figure is continued on the next page.

Figure 4.3: Deficit-to-gdp ratios in the Euro Area - continued



Note: The figure visualizes the with-interest deficit series for 14 member countries of the European Monetary Union (EMU).

4.A.2 Existing empirical evidence on debt sustainability in Europe

The empirical literature on assessing debt sustainability can be divided into three major approaches: i) stationarity tests, ii) cointegration tests, and iii) fiscal reaction function tests. The pioneering work was done by Hamilton and Flavin (1986), who showed that under the assumption of constant interest rates, stationary primary deficits as well as stationary debts are a necessary condition for a balanced government budget in present-value terms. Using annual data for the United States, Hamilton and Flavin (1986) examined the order of integration of real government debt and real primary deficit and found that the intertemporal budget constraint of the U.S. government was satisfied in a time period from 1962 to 1984.[116] In a continuative study, Trehan and Walsh (1988) argue that even in the case of integrated debt and deficit series, the intertemporal budget constraint might still hold as long as debt and primary balances are cointegrated. Assuming a constant interest rate i this can be derived from a slightly modified equation (4.2):

$$D_{t+1} - D_t = i \cdot D_t + S_t. \tag{4.14}$$

If D_t follows an I(1) process, then $D_{t+1} - D_t$ is stationary by definition, which implies that the with-interest deficit $DEF = i \cdot D_t + S_t$ is stationary and D_t and S_t are cointegrated with a cointegrating vector (1, -r). While Trehan and Walsh (1988) measure debt and the primary deficit in real terms, Hakkio and Rush (1991) employ the test expressing

[116] Wilcox (1989) allows for non-stationarity in the non-interest surplus and extends this testing procedure by allowing for a stochastic real interest rate.

the variables as a percentage of GDP. Trehan and Walsh (1988) were unable to reject the hypothesis of intertemporal budget balance in the United States in a time period from 1890 to 1960 and Hakkio and Rush (1991) conclude that the postwar U.S. data are inconsistent with this hypothesis.

Trehan and Walsh (1991) relax in a further seminal work the assumption of a constant interest rate and show that stationarity of the inclusive-of-interest deficit is a sufficient condition for intertemporal budget balance, as long as the expected real rate is positive. In this context, one could also test whether public expenditures including interest payments are cointegrated with public revenues (see, e.g., Trehan and Walsh (1991); Hakkio and Rush (1991)). The stationarity of a deficit measure and the cointegration of its non-stationary components are conceptually equivalent. Accordingly, Quintos (1995) distinguishes between strong and weak sustainability. The former requires that revenues and expenditures are cointegrated with a unit slope, while the latter only requires the slope to lie between zero and one. With this weaker condition, cointegration is only a sufficient condition with the necessary and sufficient condition being that the debt process grows slower than the growth rate of the mean interest rate.

Bohn (1998) argues that tests on sustainability should be independent of future discount rates applied in computing the present value of public debt and should be more linked to practical politics.[117] Unit root and cointegration analysis are an adequate way to test for "ad hoc sustainability" (Bohn (2008), p. 22) but do not provide sufficient information about

[117] Bohn (1995) shows that in a stochastic environment the discounting of future government debt in the intertemporal budget constraint and in the transversality condition depends on the probability distribution and generally cannot be written in terms of expected fiscal variables discounted at a fixed interest rate.

model-based conditions, as, for example tax smoothing or general political responsiveness to public debt. As a consequence, Bohn (1998, 2008) suggests estimating the following fiscal reaction function:

$$s_t = \rho \cdot d_t + \nu \cdot Z_t + \epsilon, \tag{4.15}$$

where Z_t is a set of additional explanatory variables. A positive reaction coefficient ρ indicates that fiscal policy makers respond to increases in the debt-GDP ratio by raising the primary surplus, or equivalently, by reducing the primary deficit. As pointed out by Bohn (1998), a positive reaction of primary surpluses to changes in the debt-to-GDP ratio is in line with a balanced government budget in present-value terms, regardless of the development of interest rates and growth rates.[118]

Table 4.6 provides a summary of some existing empirical evidence on fiscal sustainability in member countries of the EMU. Overall, it can be seen that earlier studies on fiscal sustainability in Europe (including samples at the turn of the century) find inconclusive results. More recent studies, however, present - for the majority of countries - results in line with a sustainable fiscal policy path.[119]

[118] Within this framework, Potrafke and Reischmann (2012) highlight the importance of fiscal equalization transfers when assessing debt sustainability. Furthermore, Collignon (2012) proposes a different approach to assessing Europe's debt sustainability by looking at governments' fiscal policy reaction patterns within the given context of economic growth and interest rates.

[119] Neck and Sturm (2008) report several contributions that give an evaluation of fiscal policy for various European countries such as Austria, Denmark, Italy, Netherlands, Switzerland, and the United Kingdom. In general, these results provide a more optimistic picture than for instance the findings of Afonso (2005).

Table 4.6: Empirical evidence on fiscal sustainability for European countries

Author and date	**Countries (Period)**	**Empirical method**	**Major findings**
Heinemann (1993)	BE, FR, DE, IT, IE, NL (ca. 1952-1989)	Unit root tests for debt and primary surplus; cointegration analysis between G and T; Maastricht criteria analysis	Overall, the results are inconclusive. Time series analyses reject sustainability for BE and FR.
Vanhorebeek and van Rompuy (1995)	BE, DK, FR, DE, IE, IT, NL, UK (1970-1994)	Unit root tests for debt and deficit ratios	Evidence for sustainability in FR and DE. Results for most other countries are inconclusive.
Payne (1997)	G7 countries (1949-1994)	Cointegration analysis between G and T	For FR, IT, and JP expenditures and revenues (real GDP) are not cointegrated. Weak sustainability could be confirmed for CA, IT, UK, and the US (real levels and real per capita); strong sustainability was found for DE (all measures).
Greiner and Semmler (1999)	DE (1955-1994)	Unit root tests for debt and (primary) surplus	Most of the tests lead to the conclusion that the IBC of the government is not met.
Papadopoulos and Sidiropoulos (1999)	BE, ES, GR, IT, PT (1961-1994)	Cointegration analysis between G and T	Sustainable budget deficits are found in GR, ES, and PT. On the contrary, the intertemporal budget balance does not hold for BE and IT.
Feve and Henin (2000)	G7 countries (semi-annual data on the last three decades)	Unit root tests for debt-to-GDP ratio	Robust evidence for fiscal sustainability only in the US and JP.
Uctum and Wickens (2000)	AT, BE, DK, ES, FR, DE, IE, IT, NL, PT, UK, US (1965-1994)	Unit root tests for debt-to-GDP ratio	Evidence for fiscal sustainability only for DK, NL, and IE. Sustainability is improved for all countries by extending the sample into the year 2000 by using forecast data.
Bravo and Silvestre (2002)	AT, BE, DK, ES, FR, DE, IE, IT, NL, PT, UK (1960-2000)	Cointegration analysis between G and T	Evidence for weak fiscal sustainability in AT, DE, FR, NL, the UK.
Hatemi-J (2002)	SE (1963-2000)	Cointegration analysis between G and T	Evidence for fiscal sustainability.
Afonso (2005)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, LU, PT, SE, UK (1970-2003)	Unit root tests for ΔD_t; cointegration analysis between G and T	Fiscal policy is sustainable only for few expectations (AT, DE, FI, NL, PO, UK). Still, the estimated coefficients for expenditures in the cointegration equations for those countries are less than one.
Haber and Neck (2006)	AT (1960-2003)	Reaction function of primary surplus ratio to changes in debt ratio	Austrian fiscal policies were sustainable in the period from 1960 to 1974, while from 1975 on, public debt grew much more rapidly.
Claeys (2007)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, PT, SE, UK (1970-2001)	Panel cointegration analysis between G, T and real interest payments	Results show that European fiscal policy has been sustainable overall, yet national experiences differ considerably.

Table 4.6 – continued

Author and date	Countries (Period)	Empirical method	Major findings
Greiner et al. (2007)	FR, DE, IT, PT, US (1960-2003)	Reaction function of primary surplus ratio to changes in debt ratio	Fiscal policies in the countries under consideration seem to be sustainable.
Landolfo (2008)	EMU (1966 - 2004); US (1977-2003)	Cointegration analysis between primary deficits, public debt, and interest rates	EMU and US on a sustainable fiscal policy path.
Hamori and Hamori (2009)	AT, BE, DK, ES, FI, FR, DE, IE, IT, NL, PT (1991-2005)	Panel unit root tests for government budget deficits	Evidence against fiscal sustainability in a sample from 1991 to 2005; empirical evidence seems to be most favorable to the sustainability hypothesis in a sample from 1997 to 2005.
Prohl and Westerlund (2009)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, LU, PT, SE, UK (1970-2004)	Panel cointegration analysis between G and T	Sustainability hypothesis cannot be rejected for the panel; hypothesis of a unit slope between G and T cannot be rejected.
Afonso and Rault (2010)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, LU, PT, SE, UK (1970-2006)	(Panel) unit root tests for ΔD_t; (panel) cointegration analysis between G and T	Panel unit root as well as cointegration analysis confirms sustainable fiscal policy for the panel set and within sub-periods.
Holmes et al. (2010)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, SE, UK (1971-2006)	Panel unit root tests for budget deficits	Evidence in favor of sustainable fiscal policy throughout the whole sample as well as pre- and post-Maastricht criteria subsamples.
Westerlund and Prohl (2010)	CA, FI, FR, IE, JP, SE, UK, US (1977Q1-2005Q4)	Panel cointegration analysis between G and T	Sustainability hypothesis cannot be rejected for the panel, hypothesis of a unit slope between G and T cannot be rejected.
Collignon (2012)	AT, BE, DK, ES, FI, FR, DE, GR, IE, IT, NL, PT, SE, UK (1978-2009)	Fiscal policy reaction patterns within the given context of economic growth and interest rates	Empirical evidence in favor of sustainable fiscal policy.

Note: The table reviews all published empirical studies since 1990 that largely cover European countries and apply time-series methods (updated until December 2012). ΔD_t denotes the first difference of the stock of public debt, while G and T indicate public expenditures and revenues.

4.A.3 Unit root tests for alternative pre-crisis sample sizes

The empirical findings in section 4.3 of this chapter suggest that in a sample period until 2011 (post-crisis) more deficit series follow a sustainable path than in a period until 2008 (pre-crisis). From these findings, we deduce a major shortcoming of symmetric unit root tests when it comes to assessing fiscal sustainability: the general testing procedure does not distinguish between negative or positive stochastic trends. This section shows that the unit root results presented in the main part of this chapter are robust to alternative pre-crisis sample sizes. In each case the empirical results including the high deficit years 2009, 2010, and 2011 confirm sustainability for the largest number of countries.

Figure 4.4 displays the development of ADF test results for different sample sizes. Starting from the year 2000 we added one year in each step and visualized every corresponding AR coefficient from the Augmented Dickey-Fuller test equation including 90 (light gray) and 95 (dark gray) percent confidence bands. This first order AR coefficient reflects the dynamics of the transitory components in the deficit series. The recursive unit root results provide evidence that for some countries the estimated AR coefficients turn more negative in a sample including the 2007/8 financial crisis.[120] This behavior is particularly visible in Cyprus and the Netherlands. The unit root results generally suggest that in a sample including the financial crisis more deficit series follow a sustainable path than in a time period without the crisis.

[120] A first order AR coefficient in the ADF test equation equal to zero implies that random shocks in the deficit series are persistent. With a declining AR parameter (less than zero) shocks become more transitory. If the coefficient is statistically significant and negative, the ADF test rejects the unit root hypothesis.

Figure 4.4: Unit root results for different sample sizes

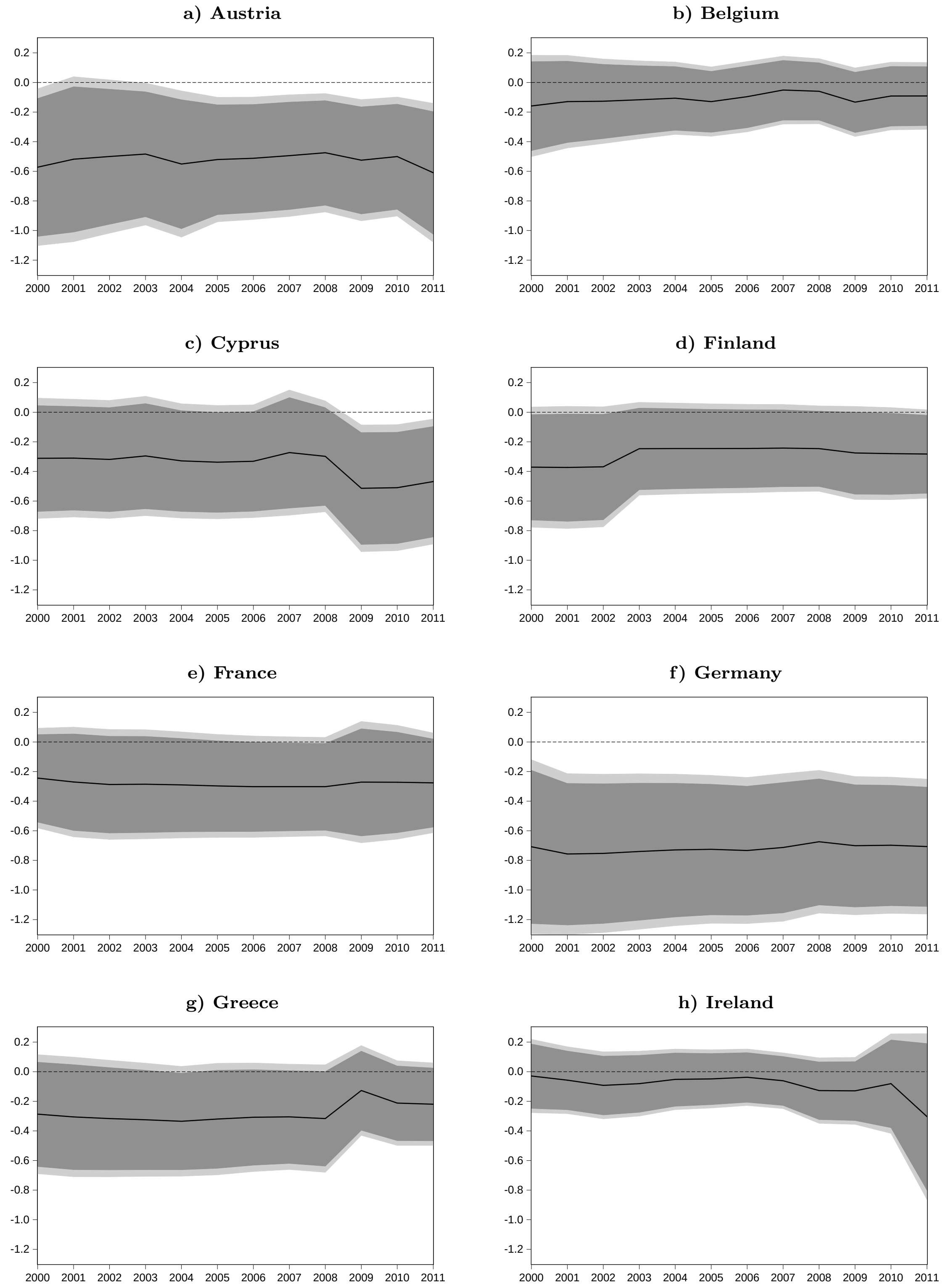
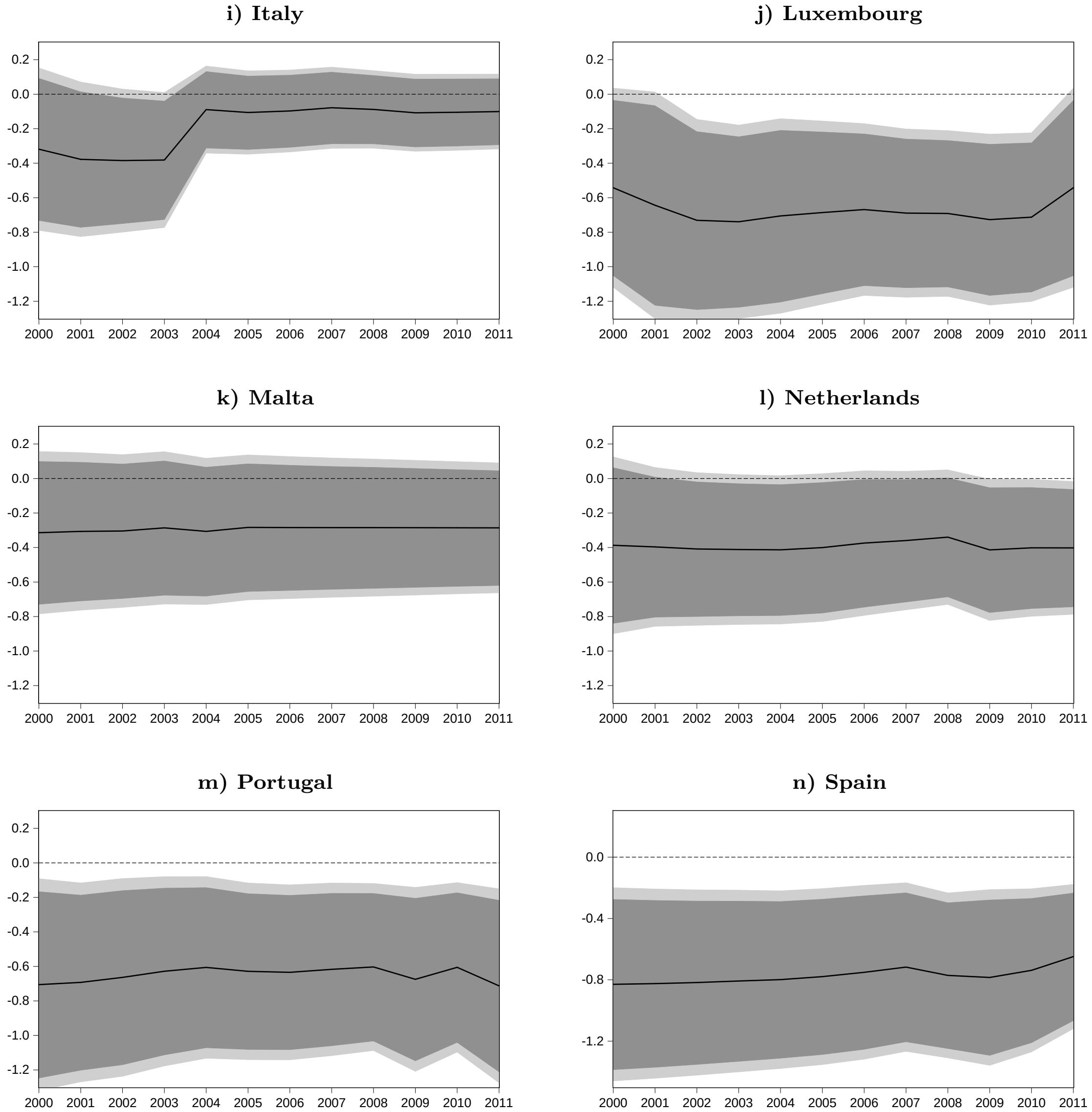


Figure is continued on the next page.

Figure 4.4: Unit root results for different sample sizes - continued



Note: The solid line visualizes the point estimate of the coefficient ρ_1 obtained from the ADF test equation $\Delta d_t = c + \rho_1 d_{t-1} + \sum_{p=1}^{P} \theta_p \Delta d_{t-p} + \epsilon_t$ for different sample sizes. The light gray shaded area visualizes the 90% confidence band and the dark gray shaded area the 95% confidence band.

4.A.4 Cointegration analysis between public expenditures and revenues

As mentioned in section 4.A.2, there are several empirical approaches on assessing debt sustainability. While the main chapter provides evidence on the stationarity properties of government deficits, this section focuses on the cointegration relationship between government expenditures and revenues for the largest eleven Euro member countries, using data from the AMECO database collected by the European Commission (2013). In theory, the stationarity of fiscal deficits and the cointegration of its non-stationary components should be conceptually equivalent. Nevertheless, the advantage of the cointegration approach is that it allows to distinguish between weak and strong sustainability (see Quintos (1995)). Strong sustainability requires that revenues and expenditures be cointegrated with a unit slope, while weak sustainability only requires the slope to lie between zero and one. An overview of the procedures to test fiscal sustainability and the resulting conclusions is provided by Afonso (2005).

Data and preliminary analysis

We start our analysis by visualizing the data of public spending and revenue expressed in nominal values and as a percentage of GDP (see figure 4.5). Data for the Netherlands cover the longest time period, namely from 1969 to 2011, while data for Greece, Ireland, Italy, and Spain encompass the shortest time horizon from 1980 to 2011. The data period for all other countries is somewhere in between. Overall, figure 4.5 shows that public expenditures and revenues at current market prices as well as a share of GDP have increased considerably over the whole sample period. In fact, only in Finland the expenditure/GDP ratio does not permanently exceed the revenue/GDP ratio. At the end of 2011, government spending at a percentage of GDP ranges on average around 45% to 55% depending on the country. In contrast, revenues have been slightly lower, e.g 33% in Ireland (lowest value) and 54% in Finland (highest value).

Figure 4.5: Government expenditures and revenues

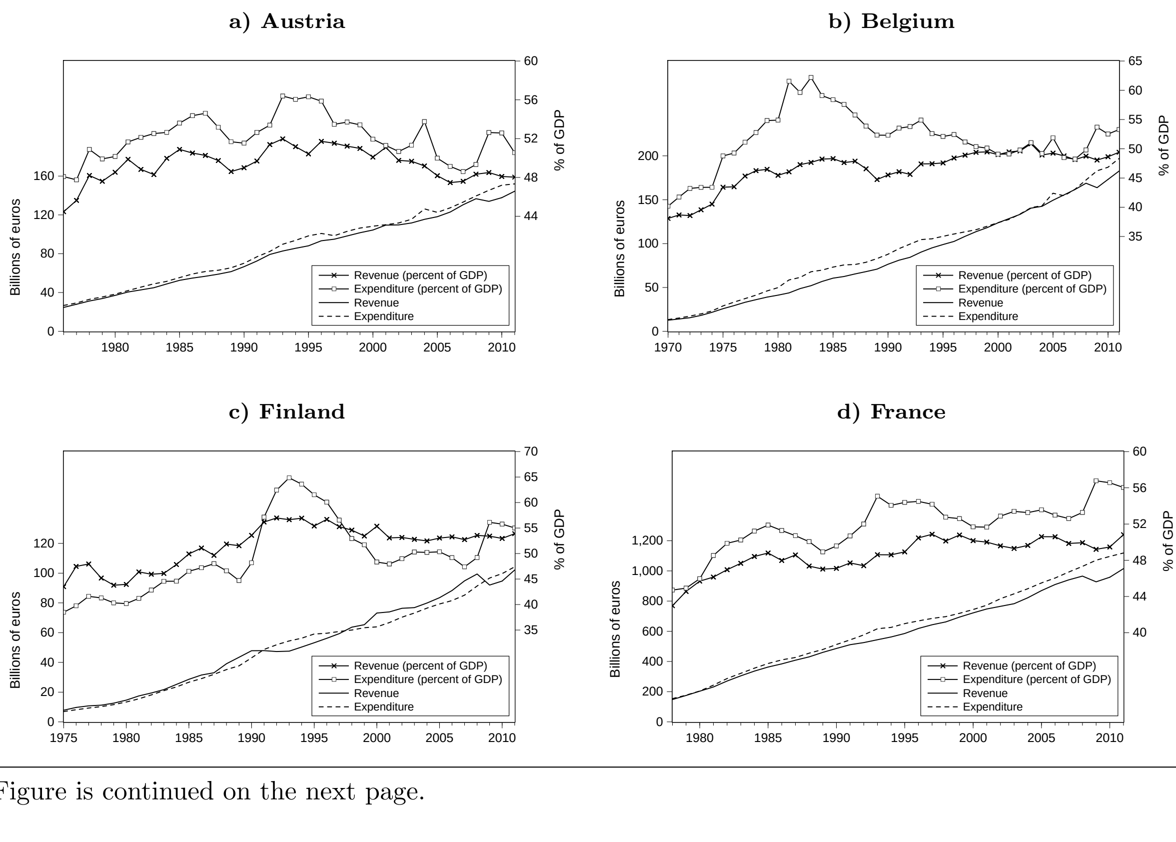
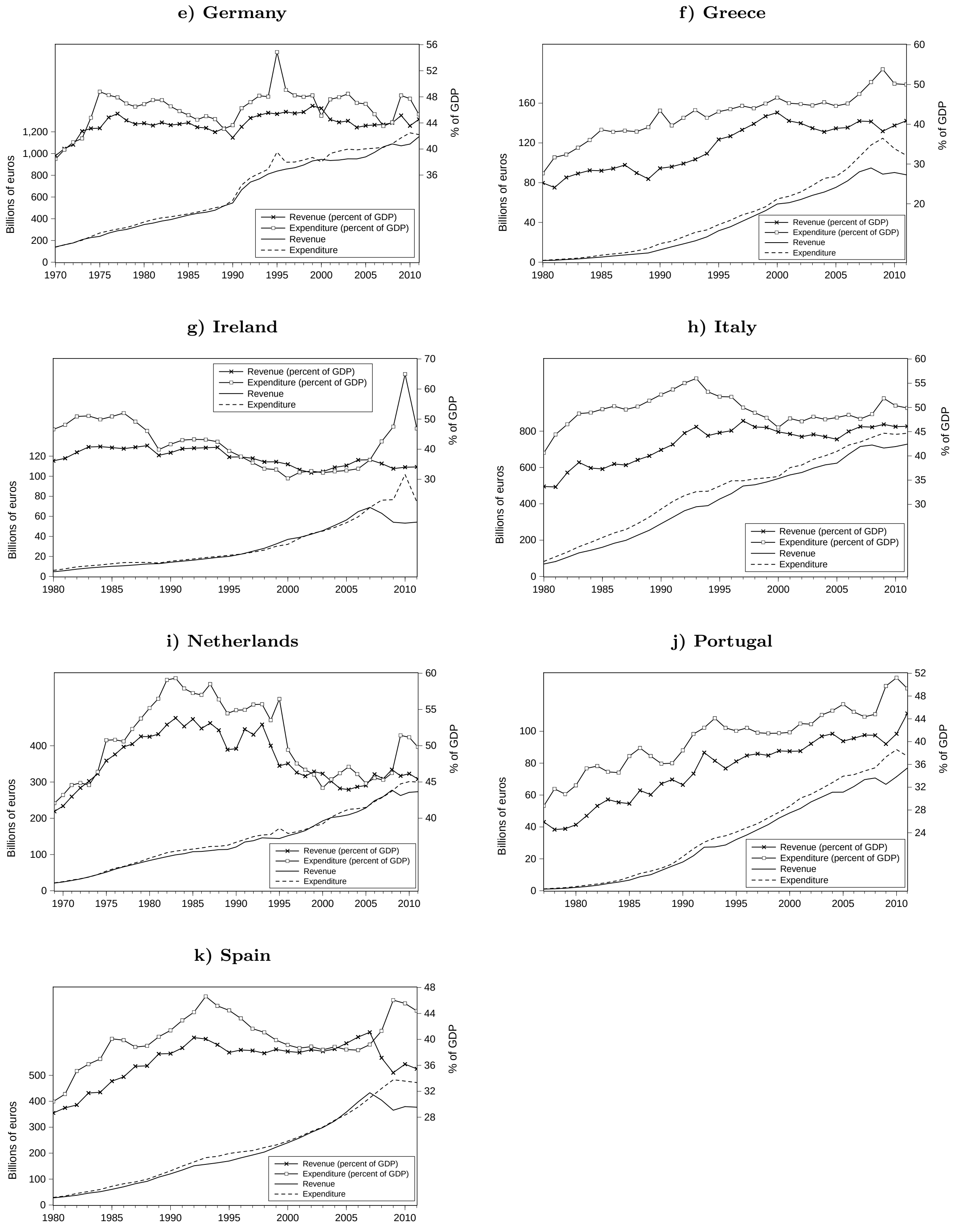


Figure is continued on the next page.

Figure 4.5: Government expenditures and revenues - continued



Note: The figure displays government expenditures and revenues at current market prices (left axis) and government expenditures and revenues as a percentage of GDP (right axis).

Unit root results

Visual inspections of the time series might suggest that revenue and expenditures follow a long-run relationship. To formally analyze the existence of a cointegration relationship between government expenditure and revenue, we apply the cointegration procedure developed by Johansen and Juselius (1990). Because this testing procedure requires the use of difference stationary variables, we start our empirical analysis by testing the unit root properties applying the PP test, the ADF test, and the KPSS test in levels and first differences of the variables. The test statistics in table 4.7 generally provide coinciding results, concluding that most data series have a unit root in level and are stationary in first differences. In fact, at least one test statistic of the various testing procedures allows to assume unit root behavior in levels and stationarity in first differences. Consequently, the data series can be treated as integrated of order one.[121]

Table 4.7: Stationarity of government revenues and expenditures

		PP test		**ADF test**		**KPSS test**	
Country	Variable	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
Austria	REV	-2.527	-5.026***	-2.538	-5.188***	0.084	0.175
	EXP	-2.418	-5.869***	-2.416	-5.861***	0.069	0.083
	REVGDP	-2.898	-5.276***	-2.881	-5.288***	0.198**	0.437*
	EXPGDP	-2.230	-5.396***	-1.784	-5.402***	0.178**	0.283
Belgium	REV	-0.716	-6.323***	-0.896	-1.588	0.215**	0.889***
	EXP	0.106	-6.652***	0.026	-6.584***	0.146	0.433*
	REVGDP	-2.271	-6.099***	-2.271	-6.034***	0.135*	0.283
	EXPGDP	-2.358	-6.942***	-2.438	-6.964***	0.147**	0.307

Table is continued on the next page.

[121] Greece is the only country where revenue and expenditures seem to be integrated of an order above one. Here, in fact, the PP as well as the ADF test do not allow to reject the unit root hypothesis in levels and first differences. Also the KPSS test rejects stationarity in levels and first differences. It should be noted that in the case of Greece the following Johansen cointegration results have to be interpreted with caution.

Table 4.7 – continued

		PP test		ADF test		KPSS test	
Country	Variable	Level	1st Diff.	Level	1st Diff.	Level	1st Diff.
Finland	REV	-2.933	-6.214***	-3.907**	-5.263***	0.108	0.334
	EXP	-1.746	-6.329***	-1.986	-6.331***	0.432***	0.244
	REVGDP	-1.681	-6.472***	-1.641	-6.472***	0.161**	0.243
	EXPGDP	-1.704	-6.902***	-2.503	-3.230**	0.143*	0.116
France	REV	-2.505	-4.800***	-3.774**	-4.494***	0.083	0.248
	EXP	-1.583	-3.578**	-3.113	-3.450**	0.306***	0.115
	REVGDP	-3.429*	-5.033***	-3.446*	-5.048***	0.122*	0.219
	EXPGDP	-2.597	-4.281***	-3.510*	-4.267***	0.133*	0.152
Germany	REV	-1.858	-3.641***	-2.198	-3.666***	0.228***	0.095
	EXP	-1.869	-6.329***	-1.705	-6.331***	0.229***	0.106
	REVGDP	-3.269*	-6.258***	-3.271*	-6.258***	0.130*	0.229
	EXPGDP	-3.156	-6.902***	-3.233*	-6.852***	0.097	0.224
Greece	REV	-1.587	-2.389	-1.624	-2.510	0.182**	0.565**
	EXP	1.010	-1.381	0.834	-1.658	0.192**	0.755***
	REVGDP	-1.558	-4.668***	-1.350	-4.648***	0.156**	0.120
	EXPGDP	-3.862**	-7.366***	-3.849**	-6.632***	0.145*	0.253
Ireland	REV	-1.179	-1.942	-2.712	-2.971*	0.180**	0.329
	EXP	4.813	-0.592	4.288	-0.938	0.183**	0.563**
	REVGDP	-2.877	-4.372***	-2.877	-4.423***	0.111	0.208
	EXPGDP	-1.645	-5.477***	-0.272	-2.065	0.149**	0.147
Italy	REV	-2.166	-4.575***	-2.166	-4.624***	0.123*	0.097
	EXP	-1.607	-4.204***	-1.446	-4.213***	0.138*	0.172
	REVGDP	-1.764	-4.913***	-1.795	-4.929***	0.175**	0.300
	EXPGDP	-3.134	-4.701***	-1.835	-4.703***	0.135*	0.312
Netherlands	REV	1.239	-2.927*	1.728	-2.927*	0.184**	0.446*
	EXP	-0.337	-5.480***	-0.316	-5.485***	0.140*	0.321
	REVGDP	-2.506	-5.666***	-2.496	-5.605***	0.179**	0.412*
	EXPGDP	-2.062	-5.808***	-2.056	-5.809***	0.168**	0.307
Portugal	REV	-3.142	-3.637**	-3.142	-3.742***	0.147**	0.407*
	EXP	-3.199	-3.483**	-2.897	-3.602**	0.174**	0.300
	REVGDP	-2.784	-6.879***	-2.807	-6.605***	0.180**	0.039
	EXPGDP	-3.281*	-4.954***	-3.784**	-4.857***	0.129*	0.066
Spain	REV	-1.505	-2.484	-3.066	-3.286**	0.168**	0.318
	EXP	1.823	-0.134	1.164	0.805	0.151**	0.468**
	REVGDP	-0.946	-4.373***	-1.221	-4.373***	0.181**	0.506**
	EXPGDP	-2.290	-3.595**	-2.705	-3.595**	0.129*	0.209

Note: The Phillips-Perron (PP), the Augmented Dickey Fuller (ADF) as well as the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test for the levels are calculated including a constant and a trend in the test equation. The test statistics for the first differences are calculated including a constant. The bandwidth for the PP and KPSS test is selected based on Newey-West using Bartlett kernel and the lag length for the ADF test is selected based on SIC. The symbols *, **, and *** indicate significance at the 10%, 5%, and 1% levels using critical values from MacKinnon (1996) and Kwiatkowski et al. (1992).

Cointegration results

After determining the unit root properties of the data series, we apply the Johansen cointegration procedure in order to test for a long-run relationship between government expenditures and revenues. Cointegration tests are conducted for all countries, irrespective of the unit root results displayed in table 4.7. As seen in table 4.8, a cointegration relationship between the variables expressed at current market prices as well as a share of GDP could be detected for Austria, Germany, Netherlands, Portugal, and Spain.[122] Furthermore, it can be seen that in Belgium, France, and Ireland a cointegration relation only exists between the nominal values. No cointegration could be found for Finland, Greece, and Italy. Overall, the cointegration results for the variables expressed at a percentage of GDP are consistent with the unit root results of the deficit-to-gdp ratio presented in table 4.1 in the main part of this chapter.[123]

Table 4.8: Cointegration of government revenues and expenditures

			Johansen (1)				Johansen (2)		
Country	Variable		Trace		Max-Eigen.		Trace		Max-Eigen.
Austria	at market prices	r=0	16.191**	r=0	15.634**	r=0	25.346**	r=0	19.741**
(1976-2011)		r≤1	0.558	r=1	0.558	r≤1	5.605	r=1	5.605
	percent of GDP	r=0	18.658**	r=0	15.949**	r=0	20.648	r=0	15.951
		r≤1	2.708	r=1	2.708	r≤1	4.697	r=1	4.697
Belgium	at market prices	r=0	33.168***	r=0	30.471***	r=0	43.794***	r=0	36.378***
(1970-2011)		r≤1	2.697	r=1	2.697	r≤1	7.415	r=1	7.415
	percent of GDP	r=0	11.823	r=0	10.281	r=0	14.229	r=0	10.347
		r≤1	1.541	r=1	1.541	r≤1	3.882	r=1	3.882

Table is continued on the next page.

[122] Details about the test specifications and lag selection of the cointegration analysis can be found in the note below the table.

[123] Only the unit root and cointegration results for Finland lead to different inferences. While the unit root results are significant and allow to reject unit root behavior for the deficit ratio series at the 10 percent level, the cointegration results do not allow to reject the no cointegration hypothesis. These counter-intuitive results might be explained by differences in the investigated sample period and econometric inaccuracies.

Table 4.8 – continued

Country	Variable		Johansen (1) Trace		Max-Eigen.		Johansen (2) Trace		Max-Eigen.
Finland	at market prices	r=0	9.926	r=0	8.013	r=0	22.575	r=0	16.734
(1975-2011)		r≤1	1.913	r=1	1.913	r≤1	5.651	r=1	5.651
	percent of GDP	r=0	11.612	r=0	10.104	r=0	12.891	r=0	11.218
		r≤1	1.508	r=1	1.508	r≤1	1.673	r=1	1.673
France	at market prices	r=0	15.042*	r=0	14.838**	r=0	24.154*	r=0	16.095
(1978-2011)		r≤1	0.204	r=1	0.204	r≤1	8.058	r=1	8.058
	percent of GDP	r=0	9.283	r=0	6.219	r=0	19.878	r=0	13.755
		r≤1	3.064	r=1	3.064	r≤1	6.123	r=1	6.123
Germany	at market prices	r=0	17.001**	r=0	16.990**	r=0	25.912**	r=0	21.139**
(1970-2011)		r≤1	0.011	r=1	0.011	r≤1	4.773	r=1	4.773
	percent of GDP	r=0	30.153***	r=0	19.240***	r=0	31.139***	r=0	20.225**
		r≤1	10.913***	r=1	10.913***	r≤1	10.914	r=1	10.914
Greece	at market prices	r=0	4.994	r=0	2.965	r=0	15.832	r=0	13.331
(1980-2011)		r≤1	2.029	r=1	2.029	r≤1	2.500	r=1	2.500
	percent of GDP	r=0	7.495	r=0	4.871	r=0	18.752	r=0	14.912
		r≤1	2.625	r=1	2.625	r≤1	3.839	r=1	3.839
Ireland	at market prices	r=0	26.211***	r=0	25.288***	r=0	34.629***	r=0	30.089***
(1980-2011)		r≤1	0.923	r=1	0.923	r≤1	4.539	r=1	4.539
	percent of GDP	r=0	8.401	r=0	5.774	r=0	17.036	r=0	11.263
		r≤1	2.628	r=1	2.628	r≤1	5.773	r=1	5.773
Italy	at market prices	r=0	9.229	r=0	5.909	r=0	13.302	r=0	7.406
(1980-2011)		r≤1	3.321	r=1	3.321	r≤1	5.896	r=1	5.896
	percent of GDP	r=0	12.758	r=0	9.117	r=0	15.826	r=0	11.689
		r≤1	3.641	r=1	3.641	r≤1	4.137	r=1	4.137
Netherlands	at market prices	r=0	36.793***	r=0	36.631***	r=0	42.826***	r=0	38.299***
(1969-2011)		r≤1	0.162	r=1	0.162	r≤1	4.527	r=1	4.527
	percent of GDP	r=0	21.254***	r=0	18.305***	r=0	26.876**	r=0	19.941**
		r≤1	2.949	r=1	2.949	r≤1	6.936	r=1	6.936
Portugal	at market prices	r=0	8.682	r=0	8.306	r=0	29.866**	r=0	26.453***
(1977-2011)		r≤1	0.377	r=1	0.377	r≤1	3.413	r=1	3.413
	percent of GDP	r=0	19.798**	r=0	18.998***	r=0	28.439**	r=0	19.000**
		r≤1	0.801	r=1	0.801	r≤1	9.439	r=1	9.439
Spain	at market prices	r=0	8.832	r=0	8.775	r=0	24.549*	r=0	17.017*
(1980-2011)		r≤1	0.057	r=1	0.057	r≤1	7.533	r=1	7.533
	percent of GDP	r=0	18.974**	r=0	14.255**	r=0	22.393	r=0	17.575*
		r≤1	4.719**	r=1	4.719**	r≤1	4.818	r=1	4.818

Note: Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The lag length in the test specifications is chosen by SIC. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels by employing critical values from Osterwald-Lenum (1992).

Weak versus strong sustainability

Finally, we conclude our analysis by examining the estimated cointegrating vectors. Quintos (1995) distinguishes between strong and weak sustainability. The former requires that revenues and expenditures are cointegrated with a unit slope, while the latter only requires the slope to lie between zero and one. Table 4.9 shows that in the case of an existing significant cointegration vector, the estimated coefficient for public spending - with revenues being the dependent variable - is typically less than one. These findings demonstrate that government expenditures have been rising more pronounced since the 1980s than public revenues. In summary, it can be stated that Austria, Belgium, France, Germany, Ireland, Netherlands, and Spain satisfy the weaker form of sustainability at least for one specification or pair of variables. Still, fairly robust results are only obtained for Austria and Germany. Only in the case of Portugal, the estimated cointegration vectors with revenue as percentage of GDP being the dependent variable are slightly higher than one. In this case, the strong form of debt sustainability might be satisfied.

Table 4.9: Cointegrating vector for government expenditures and revenues

		Dependent	**at market prices**		**percent of GDP**	
Country	Sample	variable	Johansen (1)	Johansen (2)	Johansen (1)	Johansen (2)
Austria	1976-2011	REV	$[1, -0.959]**$	$[1, -0.606]***$	$[1, -0.693]*$	–
		EXP	$[1, -1.042]***$	$[1, -1.649]*$	$[1, -1.444]**$	–
Belgium	1970-2011	REV	$[1, -0.749]$	$[1, -2.280]$	–	–
		EXP	$[1, -1.334]***$	$[1, -0.439]***$	–	–
Finland	1975-2011	REV	–	–	–	–
		EXP	–	–	–	–
France	1978-2011	REV	$[1, -0.926]*$	$[1, -0.671]***$	–	–
		EXP	$[1, -1.080]***$	$[1, -1.490]***$	–	–
Germany	1970-2011	REV	$[1, -0.949]$	$[1, -0.772]**$	$[1, -0.575]**$	$[1, -0.544]**$
		EXP	$[1, -1.053]***$	$[1, -1.295]***$	$[1, -1.739]***$	$[1, -1.838]***$
Greece	1980-2011	REV	–	–	–	–
		EXP	–	–	–	–
Ireland	1980-2011	REV	$[1, -0.682]$	$[1, -0.663]$	–	–
		EXP	$[1, -1.466]***$	$[1, -1.507]***$	–	–
Italy	1980-2011	REV	–	–	–	–
		EXP	–	–	–	–
Netherlands	1969-2011	REV	$[1, -0.936]$	$[1, -0.855]$	$[1, -0.672]**$	$[1, -0.637]***$
		EXP	$[1, -1.069]***$	$[1, -1.169]***$	$[1, -1.488]***$	$[1, -1.569]***$
Portugal	1977-2011	REV	–	$[1, -1.195]$	$[1, -1.090]***$	$[1, -1.112]***$
		EXP	–	$[1, -0.836]***$	$[1, -0.917]***$	$[1, -0.899]***$
Spain	1980-2011	REV	–	$[1, -1.428]$	–	$[1, -3.381]$
		EXP	–	$[1, -0.700]*$	–	$[1, -0.296]***$

Note: Cointegrating vectors are only reported for variables with a significant Trace or Maximum Eigenvalue test statistic according to table 4.8. Johansen (1) allows for a constant in the cointegration space and for a linear trend in the level data: $H_1(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0) + \alpha_\perp \gamma_0$; Johansen (2) allows for a constant and a trend in the cointegration space and for a linear trend in the level data: $H^*(r) = \Pi y_{t-1} + Bx_t = \alpha(\beta' y_{t-1} + \rho_0 + \rho_1 t) + \alpha_\perp \gamma_0$. The symbols *, **, and *** indicate significance at 10%, 5%, and 1% levels.

4.A.5 Long-run steady state values of public debt

A common criticism of time series based tests of debt sustainability is that they neglect the actual level of debt-to-GDP ratios. The empirical investigations are typically interested in the question whether the government's creditors could rationally expect that the government budget would be balanced in present-value terms. This broaden definition implies that running substantial deficits over a long period of time can be in accordance with sustainability as long as these deficits can be repaid by adequately high future surpluses. In other words, an annual deficit ratio of 100 percent would be considered sustainable as long as the deficit rate is continuously 100 percent every year. In order to account for this shortcoming, we analyze in this section the long-run steady state values of public debt for various time periods across Euro member countries based on the model by Domar (1944). A debt-to-GDP ratio of no more than 60 percent is one of five Euro convergence criteria that have been outlined in article 109j(1) of the Maastricht treaty.[124] Our findings provide evidence that only for very few countries the long-run steady state value of public debt stays within this 60 percent boundary. Further, we find for most countries debt-to-GDP ratios that exceed by far the 100 percent value. In addition, it can be seen that the time period just before and after the Euro introduction seems to display a more sustainable debt development in comparison to earlier and later time periods.

[124] In general, the Euro convergence criteria (also known as Maastricht criteria) contain five points that must be met by European countries if they wish to adopt the European Union's single currency: i) inflation of no more than 1.5 percentage points above the average rate of the three EU member states with the lowest inflation, ii) deficits at or below three percent of gross domestic product, iii) debt to GDP ratios below 60 percent, iv) long-term interest rates (ten year government bond yields) should not be than 2.0 percentage points above the three EU member states with the lowest inflation v) applicant countries should have joined the exchange-rate mechanism under the European Monetary System for two consecutive years. A detailed discussion on this issue is provided by Garrett (1993).

Domar's model of debt convergence

Domar (1944) shows that even in the case of consecutive new indebtedness, the public debt burden does not increase illimitably as long as the interest rate for government loans does not exceed the economic growth rate. Given the assumptions that debt and GDP follow a constant relation, interest rates are exogenous, and the economic growth rate is independent from the level of public debt, Domar illustrates that the debt-to-GDP ratios head for a constant limiting value.

As a starting point, Domar uses the expression that every deficit B in period t leads to a new level of debt in period $t+1$:

$$D_{t+1} = D_t + B_t. \tag{4.16}$$

The deficit B in every period is defined as a constant portion of gross domestic product Y that can be described by the following linear function:

$$B_t = \alpha \cdot Y_t. \tag{4.17}$$

Given the assumption that GDP grows with a constant rate g in every period, one obtains from equation (4.16) and (4.17) the following debt development path:

$$D_n = D_0 + \alpha \cdot Y_0 \cdot \sum_{t=0}^{n-1} (1+g)^t = D_0 + \alpha \cdot Y_0 \cdot \frac{1-(1+g)^n}{1-(1+g)}. \tag{4.18}$$

In order to express equation (4.18) as a ratio of GDP, we divide the whole expression by $Y_n = (1+g)^n \cdot Y_0$. Consequently, it follows that

$$\frac{D_n}{Y_n} = \frac{D_0}{(1+g)^n \cdot Y_0} + (1 - \frac{1}{(1+g)^n})\frac{\alpha}{g}, \tag{4.19}$$

where the debt/GDP ratio converges towards a finite value:

$$d^{\infty} = \lim_{n\to\infty} \frac{D_n}{Y_n} = \frac{\alpha}{g}, \tag{4.20}$$

with g denominating the average nominal GDP-growth rate and α corresponding to the average deficit ratio. By multiplication of the interest rate i to equation (4.20), it can be seen that the interest payment ratio also converges towards a finite value r^{∞}:

$$r^{\infty} = \lim_{n\to\infty} \frac{D_n}{Y_n} \cdot i = \frac{\alpha}{g} \cdot i. \tag{4.21}$$

Finally, it can be concluded that in case $g > i$, namely if the economic growth rate exceeds the interest rate, the interest payment ratio decreases in the long run. Thus, even in the case of consecutive new indebtedness, the public debt burden does not increase illimitably as long as the interest rate for government loans does not exceed the economic growth rate.

Empirical results

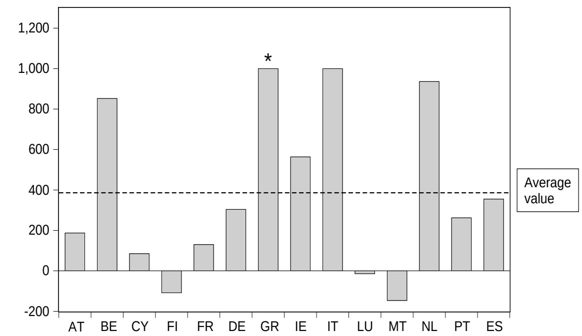
Table 4.10 displays the calculated steady state values of public debt for several five year time intervals beginning at 1980. Overall, it can be seen that the public debt ratio of almost every country converges towards a value that exceeds the Maastricht criteria of 60 percent. Especially, countries like Greece and Italy converge continuously to very high debt ratios that even exceed the 1 000 percent threshold in some cases. But also Austria, Belgium, France, and Portugal continuously display constant steady-states values of more than 100 percent. Only for Finland and Luxembourg, we find reasonably low values of debt-to-GDP ratios that are in line with the Euro convergence criteria. Furthermore, the time period around the Euro introduction display the lowest steady state values on average compared to earlier and later time periods. Still, the average value of the 14 countries exceeds the 100 percent boundary in every period. A visualization of these findings can be found in figure 4.6.

Table 4.10: Steady state value of public debt (5-year windows)

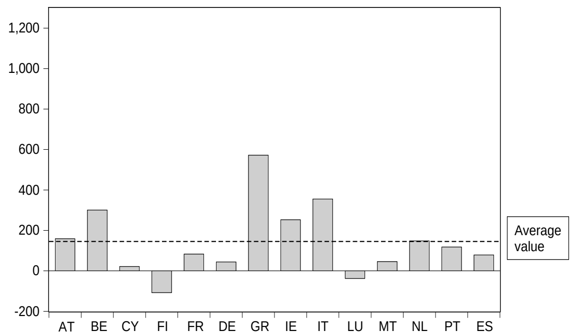
	1980-1984	**1985-1989**	**1990-1994**	**1995-1999**	**2000-2005**	**2006-2011**
Austria (AT)	187.18	158.44	132.01	114.33	90.57	162.10
Belgium (AT)	852.05	300.45	407.78	37.23	27.68	192.90
Cyprus (CY)	84.96	21.41	2.98	73.44	103.93	116.82
Finland (FI)	-108.51	-107.55	∞	35.49	-123.18	-84.40
France (FR)	129.92	82.65	329.79	143.28	145.70	575.15
Germany (DE)	303.89	43.75	88.63	240.89	266.92	87.01
Greece (GR)	∞	571.88	1309.96	191.42	132.86	∞
Ireland (IE)	563.46	252.33	62.47	-8.29	-24.84	1614.52
Italy (IT)	1184.51	354.98	925.85	248.78	212.06	∞
Luxembourg (LU)	-14.03	-37.62	-36.51	-59.15	-51.32	-42.41
Malta (MT)	-146.39	45.55	125.21	181.09	374.00	162.90
Netherlands (NL)	936.09	148.36	146.54	66.05	53.78	173.33
Portugal (PT)	262.15	117.89	210.85	97.68	275.39	3001.64
Spain (ES)	354.69	78.49	-56.42	94.03	8.20	556.18
EMU-14	386.10	145.07	309.94	104.02	106.55	421.40

Note: Steady state values of public debt are calculated as the ratio of the average deficit-to-GDP ratio and the average nominal growth rate of GDP for different time periods. Debt ratios converge towards infinity (∞) if the average GDP growth rate is negative.

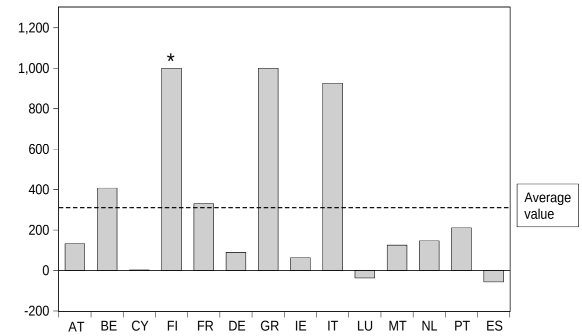
Figure 4.6: Steady state value of public debt (5-year windows)



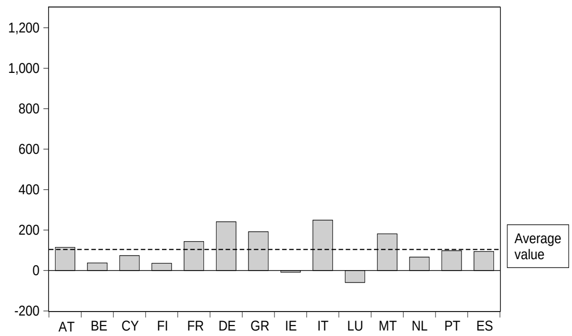
(a) 1980-1984



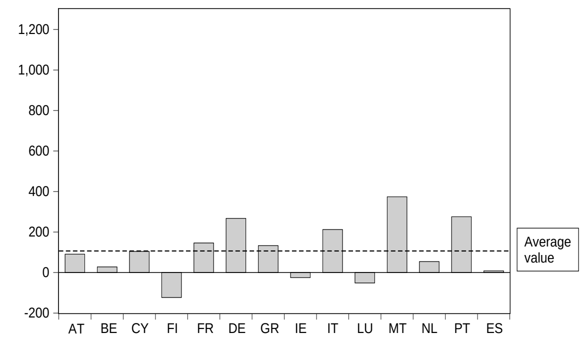
(b) 1985-1989



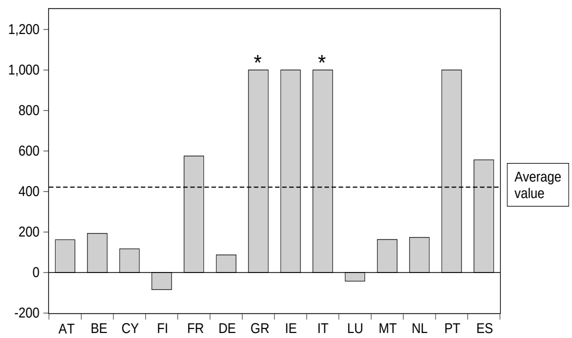
(c) 1990-1994



(d) 1995-1999



(e) 2000-2005



(f) 2006-2011

Note: See table 4.10. If the average annual growth rate is negative, the long-run steady state value is set to 1 000 percent and marked with the symbol *.

Table 4.11 displays Domar steady state values for some longer time periods. Besides a 11 year time interval (1980 - 1990), we also calculated steady state values for a 21 (1980

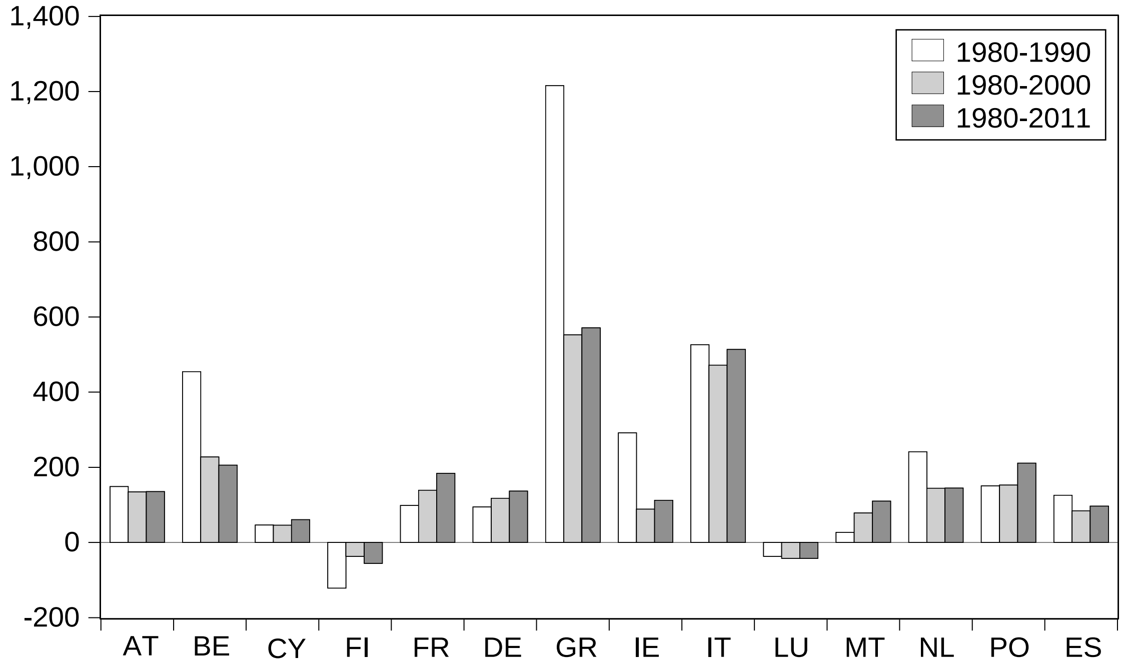
- 2000) and 32 (1980 - 2011) time window. Again, a visualization of these findings can be found in figure 4.7. In general, it can be seen that most countries exhibit steady state values that exceed by far a debt-to-GDP of 100 percent. In a time period from 1980 to 1990 Greece debt ratio even converges towards a debt-to-GDP ratio of 1 216 percent. Regarding the 32 year time window, only Malta, Finland, Luxembourg, and Spain head for a debt ratio below 100 percent. In the case of Finland and Luxembourg the long-run steady state value is even negative. Comparing the steady state values of 1980 to 2000 and 1980 to 2011, it should be noted that the debt values slightly increased. Apparently this is the result of high fiscal deficits and economic downturns as a consequence of the 2007/8 financial crisis.

Table 4.11: Steady state value of public debt (10-year, 20-year and 30-year windows)

	1980-1990	**1980-2000**	**1980-2011**
Austria (AT)	148.69	134.55	135.23
Belgium (BE)	454.52	227.47	205.89
Cyprus (CY)	46.35	45.82	60.68
Finland (FI)	-121.42	-36.70	-55.45
France (FR)	98.22	138.97	183.72
Germany (DE)	94.25	117.46	136.65
Greece (GR)	1215.49	552.32	571.23
Ireland (IE)	291.75	88.57	112.01
Italy (IT)	526.12	471.80	513.44
Luxembourg (LU)	-36.79	-42.28	-42.27
Malta (MT)	26.70	78.20	109.88
Netherlands (NL)	241.28	144.15	144.92
Portugal (PT)	150.48	152.65	210.94
Spain (ES)	125.58	84.26	96.43
EMU-14	232.94	154.09	170.24

Note: Steady state values of public debt are calculated as the ratio of the average deficit-to-GDP ratio and the average nominal growth rate of GDP for different time periods.

Figure 4.7: Steady state value of public debt (10-year, 20-year and 30-year windows)



Note: Steady state values of public debt are calculated as the ratio of the average deficit-to-GDP ratio and the average nominal growth rate of GDP for different time periods.

References

Abizadeh, S. and Yousefi, M. (1988). An empirical re-examination of Wagner's law. *Economics Letters*, 26(2):169–173.

Afonso, A. (2005). Fiscal Sustainability: The Unpleasant European Case. *Finanz-Archiv/Public Finance Analysis*, 61(1):19–44.

Afonso, A. and Rault, C. (2010). What do we really know about fiscal sustainability in the EU? A panel data diagnostic. *Review of World Economics (Weltwirtschaftliches Archiv)*, 145(4):731–755.

Afonso, A. and Sousa, R. (2011a). The macroeconomic effects of fiscal policy in Portugal: a Bayesian SVAR analysis. *Portuguese Economic Journal*, 10(1):61–82.

Afonso, A. and Sousa, R. (2011b). What are the effects of fiscal policy on asset markets? *Economic Modelling*, 28(4):1871–1890.

Afonso, A. and Sousa, R. (2012). The macroeconomic effects of fiscal policy. *Applied Economics*, 44(34):4439–4454.

Ahsan, S., Kwan, A., and Sahni, B. (1996). Cointegration and Wagner's hypothesis: time series evidence for Canada. *Applied Economics*, 28(8):1055–1058.

Akitoby, B., Clements, B., Gupta, S., and Inchauste, G. (2006). Public spending, voracity, and Wagner's law in developing countries. *European Journal of Political Economy*, 22(4):908–924.

Alesina, A. and Ardagna, S. (2010). Large Changes in Fiscal Policy: Taxes versus Spending. In Brown, J. R., editor, *NBER Book Series Tax Policy and the Economy*, volume 24, chapter 2, pages 35–68. University of Chicago Press, Chicago.

Alesina, A. and Wacziarg, R. (1998). Openness, country size and government. *Journal of Public Economics*, 69(3):305–321.

Almunia, M., Bénétrix, A., Eichengreen, B., O'Rourke, K. H., and Rua, G. (2010). From Great Depression to Great Credit Crisis: similarities, differences and lessons. *Economic Policy*, 25(62):219–265.

Amisano, G. and Giannini, C. (1997). *Topics in structural VAR econometrics*. Springer, Berlin, 2nd edition.

Andres, J., Domenech, R., and Fatas, A. (2008). The stabilizing role of government size. *Journal of Economic Dynamics and Control*, 32(2):571–593.

Arestis, P. and Sawyer, M. (2010). The return of fiscal policy. *Journal of Post Keynesian Economics*, 32(3):327–346.

Auerbach, A. J. and Gorodnichenko, Y. (2012). Measuring the Output Responses to Fiscal Policy. *American Economic Journal: Economic Policy*, 4(2):1–27.

Babatunde, M. A. (2011). A bound testing analysis of Wagner's law in Nigeria: 1970-2006. *Applied Economics*, 43(21):2843–2850.

Badinger, H. (2006). Fiscal shocks, output dynamics and macroeconomic stability: an empirical assessment for Austria (1983-2002). *Empirica*, 33(5):267–284.

Bai, J. and Perron, P. (1998). Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica*, 66(1):47–78.

Bai, J. and Perron, P. (2003). Computation and analysis of multiple structural change models. *Journal of Applied Econometrics*, 18(1):1–22.

Banque Nationale de Belgique (2013). Population and labour market. Retrieved from http://www.nbb.be/pub/stats.

Barro, R. J. (1981). Output Effects of Government Purchases. *Journal of Political Economy*, 89(6):1086–1121.

Barro, R. J. (1990). Government Spending in a Simple Model of Endogenous Growth. *Journal of Political Economy*, 98(5):103–126.

Baumol, W. J. (1967). Macroeconomics of Unbalanced Growth: The Anatomy of Urban Crisis. *American Economic Review*, 57(3):415–426.

Beetsma, R., Giuliodori, M., and Klaassen, F. (2008). The Effects of Public Spending Shocks on Trade Balances and Budget Deficits in the European Union. *Journal of the European Economic Association*, 6(2-3):414–423.

Bergh, A. and Henrekson, M. (2011). Government Size and Growth: A Survey and Interpretation of the Evidence. *Journal of Economic Surveys*, 25(5):872–897.

Bergman, U. M. and Hutchison, M. (2010). Expansionary Fiscal Contractions: Re-evaluating the Danish Case. *International Economic Journal*, 24(1):71–93.

Biau, O. and Girard, E. (2005). Politique budgétaire et dynamique économique en France. L'approche var structurel. *Revue économique*, 56(3):755–764.

Biswal, B., Dhawan, U., and Lee, H.-Y. (1999). Testing Wagner versus Keynes using disaggregated public expenditure data for Canada. *Applied Economics*, 31(10):1283–1291.

Blanchard, O. J. and Leigh, D. (2013). Growth Forecast Errors and Fiscal Multipliers. *American Economic Review*, 103(3):117–20.

Blanchard, O. J. and Perotti, R. (2002). An Empirical Characterization Of The Dynamic Effects Of Changes In Government Spending And Taxes On Output. *Quarterly Journal of Economics*, 117(4):1329–1368.

Blinder, A. S. and Solow, R. M. (1976). Does fiscal policy still matter?: A reply. *Journal of Monetary Economics*, 2(4):501–510.

Board of Governors of the Federal Reserve System (2013). Economic Research & Data. Retrieved from http://www.federalreserve.gov/releases/h15/data.htm.

Bode, O., Gerke, R., and Schellhorn, H. (2009). Die Wirkung fiskalischer Schocks auf das Bruttoinlandsprodukt. Arbeitspapier 01/2006, Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung.

Bohl, M. T. (1996). Some International Evidence on Wagner's Law. *Public Finance/Finances publiques*, 51(2):185–200.

Bohn, H. (1991). Budget balance through revenue or spending adjustments? Some historical evidence for the United States. *Journal of Monetary Economics*, 27(3):333–359.

Bohn, H. (1995). The Sustainability of Budget Deficits in a Stochastic Economy. *Journal of Money, Credit and Banking*, 27(1):257–271.

Bohn, H. (1998). The Behavior Of U.S. Public Debt And Deficits. *Quarterly Journal of Economics*, 113(3):949–963.

Bohn, H. (2007). Are stationarity and cointegration restrictions really necessary for the intertemporal budget constraint? *Journal of Monetary Economics*, 54(7):1837–1847.

Bohn, H. (2008). The Sustainability of Fiscal Policy in the United States. In Neck, R. and Sturm, J.-E., editors, *Sustainability of Public Debt*, chapter 2, pages 15–49. The MIT Press, Cambridge.

Bolt, J. and van Zanden, J. (2013). The First Update of the Maddison Project; Re-Estimating Growth Before 1820. Maddison Project Working Paper 4, The Maddison Project.

Bravo, A. B. S. and Silvestre, A. L. (2002). Intertemporal sustainability of fiscal policies: some tests for European countries. *European Journal of Political Economy*, 18(3):517–528.

Brennan, G. and Buchanan, J. (1980). *The Power to Tax: Analytic Foundations of a Fiscal Constitution*. Cambridge University Press, Cambridge.

Brückner, M., Chong, A., and Gradstein, M. (2012). Estimating the permanent income elasticity of government expenditures: Evidence on Wagner's law based on oil price shocks. *Journal of Public Economics*, 96(11):1025–1035.

Bureau of Economic Analysis (2013). National Income and Product Accounts Tables. Retrieved from http://www.bea.gov/iTable/index.cfm.

Bureau of Labor Statistics (2013). Quarterly Census of Employment and Wages. Retrieved from http://www.bls.gov/data/#employment.

Burnside, C., Eichenbaum, M., and Fisher, J. D. (2004). Fiscal shocks and their consequences. *Journal of Economic Theory*, 115(1):89–117.

Burriel, P., de Castro Fernandez, F., Garrote, D., Gordo, E., Paredes, J., and Garcia, J. J. P. (2010). Fiscal Multipliers in the Euro Area. *Revista de Economia y Estadistica*, XLVIII(2):7–27.

Caldara, D. and Kamps, C. (2008). What are the effects of fiscal shocks? A VAR-based comparative analysis. Working Paper Series 877, European Central Bank.

Caldara, D. and Kamps, C. (2012). The analytics of SVARs: a unified framework to measure fiscal multipliers. Finance and Economics Discussion Series 2012-20, Board of Governors of the Federal Reserve System.

Candelon, B. and Lütkepohl, H. (2001). On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models. *Economics Letters*, 73(2):155–160.

Caner, M. and Hansen, B. E. (2001). Threshold Autoregression with a Unit Root. *Econometrica*, 69(6):1555–1596.

Chang, T. (2002). An econometric test of Wagner's law for six countries based on cointegration and error-correction modelling techniques. *Applied Economics*, 34(9):1157–1169.

Chang, T., Liu, W., and Caudill, S. (2004). A re-examination of Wagner's law for ten countries based on cointegration and error-correction modelling techniques. *Applied Financial Economics*, 14(8):577–589.

Chletsos, M. and Kollias, C. (1997). Testing Wagner's law using disaggregated public expenditure data in the case of Greece: 1958-93. *Applied Economics*, 29(3):371–377.

Chow, Y.-F., Cotsomitis, J. A., and Kwan, A. C. C. (2002). Multivariate cointegration and causality tests of Wagner's hypothesis: evidence from the UK. *Applied Economics*, 34(13):1671–1677.

Christiano, L. J., Eichenbaum, M., and Evans, C. L. (1999). Monetary policy shocks: What have we learned and to what end? In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics Vol. 1*, chapter 2, pages 65–148. Elsevier, Amsterdam.

Christiano, L. J., Eichenbaum, M., and Rebelo, S. (2011). When Is the Government Spending Multiplier Large? *Journal of Political Economy*, 119(1):78–121.

Claeys, P. (2007). Sustainability of EU Fiscal Policies: a Panel Test. *Journal of Economic Integration*, 22:112–127.

Coenen, G., Erceg, C. J., Freedman, C., Furceri, D., Kumhof, M., Lalonde, R., Laxton, D., Lindé, J., Mourougane, A., Muir, D., Mursula, S., de Resende, C., Roberts, J., Roeger, W., Snudden, S., Trabandt, M., and in't Veld, J. (2012a). Effects of Fiscal Stimulus in Structural Models. *American Economic Journal: Macroeconomics*, 4(1):22–68.

Coenen, G., Straub, R., and Trabandt, M. (2012b). Fiscal Policy and the Great Recession in the Euro Area. *American Economic Review*, 102(3):71–76.

Cogan, J. F., Cwik, T., Taylor, J. B., and Wieland, V. (2010). New Keynesian versus old Keynesian government spending multipliers. *Journal of Economic Dynamics and Control*, 34(3):281–295.

Collignon, S. (2012). Fiscal Policy Rules And The Sustainability Of Public Debt In Europe. *International Economic Review*, 53(2):539–567.

Corsetti, G., Meier, A., and Müller, G. J. (2012). What determines government spending multipliers? *Economic Policy*, 27(72):521–565.

Courakis, A. S., Moura-Roque, F., and Tridimas, G. (1993). Public expenditure growth in Greece and Portugal: Wagner's law and beyond. *Applied Economics*, 25(1):125–134.

Cwik, T. and Wieland, V. (2011). Keynesian government spending multipliers and spillovers in the euro area. *Economic Policy*, 26(67):493–549.

Dalena, M. and Magazzino, C. (2012). Public Expenditure and Revenue in Italy, 1862-1993. *Economic Notes*, 41(3):145–172.

Dao, M. Q. (1995). Determinants of Government Expenditures: New Evidence from Disaggregative Data. *Oxford Bulletin of Economics and Statistics*, 57(1):67–76.

Davig, T. and Leeper, E. M. (2011). Monetary-fiscal policy interactions and fiscal stimulus. *European Economic Review*, 55(2):211–227.

De Castro, F. (2006). The macroeconomic effects of fiscal policy in Spain. *Applied Economics*, 38(8):913–924.

Devarajan, S., Swaroop, V., and Zou, H. (1996). The composition of public expenditure and economic growth. *Journal of Monetary Economics*, 37(2):313–344.

Domar, E. D. (1944). The "Burden of the Debt" and the National Income. *American Economic Review*, 34(4):798–827.

Dritsakis, N. and Adamopoulos, A. (2004). A causal relationship between government spending and economic development: an empirical examination of the Greek economy. *Applied Economics*, 36(5):457–464.

Durevall, D. and Henrekson, M. (2011). The futile quest for a grand explanation of long-run government expenditure. *Journal of Public Economics*, 95(7-8):708–722.

Easterly, W. (2007). Inequality does cause underdevelopment: Insights from a new instrument. *Journal of Development Economics*, 84(2):755–776.

Eaton, J. (1981). Fiscal Policy, Inflation and the Accumulation of Risky Capital. *Review of Economic Studies*, 48(3):435–45.

Edelberg, W., Eichenbaum, M., and Fisher, J. D. (1999). Understanding the Effects of a Shock to Government Purchases. *Review of Economic Dynamics*, 2(1):166–206.

Ehrhart, C. and Llorca, M. (2008). The sustainability of fiscal policy: evidence from a panel of six South-Mediterranean countries. *Applied Economics Letters*, 15(10):797–803.

Eloranta, J. (2007). From the great illusion to the Great War: Military spending behaviour of the Great Powers, 1870-1913. *European Review of Economic History*, 11(2):255–283.

Enders, W. and Granger, C. W. J. (1998). Unit-Root Tests and Asymmetric Adjustment with an Example Using the Term Structure of Interest Rates. *Journal of Business & Economic Statistics*, 16(3):304–11.

European Commission (2012). The 2012 Ageing Report: Economic and budgetary projections for the EU27 Member States (2010-2060). Joint report prepared by the european commission (dg ecfin) and the economic policy committee (awg).

European Commission (2013). Annual macro-economic database (AMECO). Retrieved from http://ec.europa.eu/economy_finance/ameco.

Eurostat (2012). Annual government finance statistics. Retrieved from http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.

Eurostat (2013a). National Accounts aggregates and employment by branch. Retrieved from http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.

Eurostat (2013b). Quarterly national accounts and government finance statistics. Retrieved from http://epp.eurostat.ec.europa.eu/portal/page/portal/statistics/search_database.

Evans, P. (1985). Do Large Deficits Produce High Interest Rates? *American Economic Review*, 75(1):68–87.

Faini, R. (2006). Fiscal policy and interest rates in Europe. *Economic Policy*, 21(47):443–489.

Fatás, A. and Mihov, I. (2001). The Effects of Fiscal Policy on Consumption and Employment: Theory and Evidence. CEPR Discussion Papers 2760, Centre for Economic Policy Research.

Favero, C. and Giavazzi, F. (2007). Debt and the Effects of Fiscal Policy. NBER Working Papers 12822, National Bureau of Economic Research.

Favero, C., Giavazzi, F., and Perego, J. (2011). Country Heterogeneity and the International Evidence on the Effects of Fiscal Policy. *IMF Economic Review*, 59(4):652–682.

Federal Statistical Office (2013). Destatis: Arbeitsmarkt. Retrieved from https://www-genesis.destatis.de/genesis/online.

Ferrera, M. (1997). The uncertain future of the Italian welfare state. *West European Politics*, 20(1):231–249.

Feve, P. and Henin, P.-Y. (2000). Assessing Effective Sustainability of Fiscal Policy within the G-7. *Oxford Bulletin of Economics and Statistics*, 62(2):175–195.

Fontana, G. (2009). The transmission mechanism of fiscal policy: a critical assessment of current theories and empirical methodologies. *Journal of Post Keynesian Economics*, 31(4):587–604.

Forni, L., Monteforte, L., and Sessa, L. (2009). The general equilibrium effects of fiscal policy: Estimates for the Euro area. *Journal of Public Economics*, 93(3-4):559–585.

Forte, F. and Magazzino, C. (2011). Optimal Size Government and Economic Growth in EU Countries. *Economia Politica*, XXVIII(3).

Franses, P. H. (2001). How to deal with intercept and trend in practical cointegration analysis? *Applied Economics*, 33(5):577–579.

Franses, P. H. and van Dijk, D. (2000). *Non-linear time series models in empirical finance.* Cambridge University Press, Cambridge.

Gandhi, V. P. (1971). Wagner's Law of Public Expenditure: Do Recent Cross-Section Studies Confirm it? *Public Finance/Finances publiques*, 26(1):44–56.

Garcia, S. and Henin, P.-Y. (1999). Balancing budget through tax increases or expenditure cuts: is it neutral? *Economic Modelling*, 16(4):591–612.

Garrett, G. (1993). The Politics of Maastricht. *Economics & Politics*, 5(2):105–123.

Ghatak, S. and Sanchez-Fung, J. R. (2007). Is Fiscal Policy Sustainable in Developing Economies? *Review of Development Economics*, 11(3):518–530.

Ghate, C. and Zak, P. J. (2002). Growth of government and the politics of fiscal policy. *Structural Change and Economic Dynamics*, 13(4):435–455.

Giordano, R., Momigliano, S., Neri, S., and Perotti, R. (2007). The effects of fiscal policy in Italy: Evidence from a VAR model. *European Journal of Political Economy*, 23(3):707–733.

Girouard, N. and André, C. (2005). Measuring Cyclically-adjusted Budget Balances for OECD Countries. Economics Department Working Papers 434, Organisation for Economic Co-operation and Development.

Goffman, I. (1968). On the empirical testing of Wagner's law: A technical note. *Public Finance/Finances publiques*, 23(3):359–366.

Gollin, D., Parente, S. L., and Rogerson, R. (2004). Farm Work, Home Work, and International Productivity Differences. *Review of Economic Dynamics*, 7(4):827–850.

Granger, C. W. J. (1969). Investigating Causal Relations by Econometric Models and Cross-Spectral Methods. *Econometrica*, 37(3):424–438.

Granger, C. W. J. (1988). Some recent development in a concept of causality. *Journal of Econometrics*, 39(1-2):199–211.

Greiner, A., Köller, U., and Semmler, W. (2007). Debt sustainability in the European Monetary Union: Theory and empirical evidence for selected countries. *Oxford Economic Papers*, 59(2):194–218.

Greiner, A. and Semmler, W. (1999). An Inquiry Into the Sustainability of German Fiscal Policy: Some Time-Series Tests. *Public Finance Review*, 27(2):194–220.

Groningen Growth & Development Centre (2012). Maddison Historical Statistics. Retrieved from http://www.rug.nl/research/ggdc/databases.

Gupta, S. P. (1967). Public expenditure and economic growth: a time series analysis. *Public Finance/Finances publiques*, 22(4):423–461.

Gupta, S. P. (1969). Public expenditure and economic development - a cross-section analysis. *FinanzArchiv/Public Finance Analysis*, 28(1):26–41.

Gyles, A. F. (1991). A time-domain transfer function model of Wagner's Law: the case of the United Kingdom economy. *Applied Economics*, 23(2):327–330.

Haavelmo, T. (1945). Multiplier Effects of a Balanced Budget. *Econometrica*, 13(4):311–318.

Haber, G. and Neck, R. (2006). Sustainability of Austrian public debt: a political economy perspective. *Empirica*, 33(2):141–154.

Hakkio, C. S. and Rush, M. (1991). Is the Budget Deficit “Too Large?”. *Economic Inquiry*, 29(3):429–45.

Hall, R. E. (2009). By How Much Does GDP Rise If the Government Buys More Output? *Brookings Papers on Economic Activity*, 40(2):183–249.

Hamilton, J. D. and Flavin, M. A. (1986). On the Limitations of Government Borrowing: A Framework for Empirical Testing. *American Economic Review*, 76(4):808–819.

Hamori, S. and Hamori, N. (2009). On the Sustainability of Budget Deficits in the Euro Area. *Economics Bulletin*, 29(1):56–66.

Hatemi-J, A. (2002). Fiscal policy in Sweden: effects of EMU criteria convergence. *Economic Modelling*, 19(1):121–136.

Hayo, B. (1996). Testing Wagner’s Law for Germany: An Exercise in Applied Time Series Analysis. *Jahrbuch für Nationalökonomie und Statistik/Journal of Economics and Statistics*, 215(3):328–343.

Hebous, S. (2011). The Effects Of Discretionary Fiscal Policy On Macroeconomic Aggregates: A Reappraisal. *Journal of Economic Surveys*, 25(4):674–707.

Heinemann, F. (1993). Sustainability of national debt in Europe. *Intereconomics*, 28(2):61–68.

Henrekson, M. (1993). Wagner’s Law-A Spurious Relationship? *Public Finance/Finances publiques*, 48(3):406–415.

Henry, O. and Olekalns, N. (2010). Peacock and Wiseman's displacement hypothesis: some new long-run evidence for the UK. *Applied Economics*, 42(11):1455–1460.

Holcombe, R. G. (2005). Government Growth in the Twenty-First Century. *Public Choice*, 124(1):95–114.

Holmes, M. J. (2010). Nonlinearities, co-trending and budget balance sustainability. *Economic Analysis and Policy*, 40(3):369–376.

Holmes, M. J., Otero, J., and Panagiotidis, T. (2010). Are EU budget deficits stationary? *Empirical Economics*, 38(3):767–778.

Hondroyiannis, G. and Papapetrou, E. (1995). An Examination of Wagner's Law for Greece: A Cointegration Analysis. *Public Finance/Finances publiques*, 50(1):67–79.

Hoover, K. D. and Sheffrin, S. M. (1992). Causation, Spending, and Taxes: Sand in the Sandbox or Tax Collector for the Welfare State? *American Economic Review*, 82(1):225–248.

Ilzetzki, E., Mendoza, E. G., and Végh, C. A. (2013). How big (small?) are fiscal multipliers? *Journal of Monetary Economics*, 60(2):239–254.

IMF (2002). Assessing Sustainability. *Policy Development and Review Department*. Retrieved from http://www.imf.org/external/np/pdr/sus/2002/eng/052802.pdf.

IMF (2012a). Government Finance Statistics. Retrieved from http://elibrary-data.imf.org/.

IMF (2012b). International Financial Statistics. Retrieved from http://elibrary-data.imf.org/.

Iniguez-Montiel, A. (2010). Government expenditure and national income in Mexico: Keynes versus Wagner. *Applied Economics Letters*, 17(9):887–893.

Islam, A. (2001). Wagner's law revisited: cointegration and exogeneity tests for the USA. *Applied Economics Letters*, 8(8):509–515.

Jaen-Garcia, M. (2011). Empirical Analysis of Wagner's Law for the Spain's Regions. *International Journal of Academic Research in Accounting, Finance and Management Sciences*, 1(1):1–17.

Johansen, S. (1988). Statistical analysis of cointegration vectors. *Journal of Economic Dynamics and Control*, 12(2-3):231–254.

Johansen, S. and Juselius, K. (1990). Maximum Likelihood Estimation and Inference on Cointegration–With Applications to the Demand for Money. *Oxford Bulletin of Economics and Statistics*, 52(2):169–210.

Johansen, S., Mosconi, R., and Nielsen, B. (2000). Cointegration analysis in the presence of structural breaks in the deterministic trend. *Econometrics Journal*, 3(2):216–249.

Joyeux, R. (2007). How to deal with structural breaks in practical cointegration analysis? In Rao, B. B., editor, *Cointegration for the Applied Economist*, chapter 7, pages 195–221. Palgrave Macmillan, New York, 2nd edition.

Karagianni, S. and Pempetzoglou, M. (2009). Evidence For Non-Linear Causality Between Public Spending And Income in the European Union Countries. *Journal of Applied Business Research*, 25(1):69–82.

Karagianni, S., Pempetzoglou, M., and Strikou, S. (2002). Testing Wagner's Law For The European Union Economies. *Journal of Applied Business Research*, 18(4):107–114.

Koester, G. B. and Priesmeier, C. (2012). Estimating dynamic tax revenue elasticities for Germany. Discussion Papers 23/2012, Deutsche Bundesbank.

Koester, G. B. and Priesmeier, C. (2013). Does Wagnerï£¡s Law Ruin the Sustainability of German Public Finances? *FinanzArchiv/Public Finance Analysis*, 69(3):256–288.

Kolluri, B., Panik, M., and Wahab, M. (2000). Government expenditure and economic growth: evidence from G7 countries. *Applied Economics*, 32(8):1059–1068.

Kuckuck, J. (2012). Testing Wagner's Law at Different Stages of Economic Development - A Historical Analysis of Five Western European Countries. Working Papers 91, Institute of Empirical Economic Research, Osnabrueck University.

Kuckuck, J. (2014). Testing Wagner's Law at Different Stages of Economic Development. *FinanzArchiv/Public Finance Analysis*, 70(1):128–168.

Kuckuck, J. and Westermann, F. (2013). On the Size of Fiscal Multipliers: A Counterfactual Analysis. Working Papers 96, Institute of Empirical Economic Research, Osnabrueck University.

Kuckuck, J. and Westermann, F. (2014). On the Size of Fiscal Multipliers: A Counterfactual Analysis. *Economics Letters*, 123(2):26–32.

Kumar, S., Webber, D. J., and Fargher, S. (2012). Wagner's Law revisited: cointegration and causality tests for New Zealand. *Applied Economics*, 44(5):607–616.

Kuznets, S. (1958). Underdeveloped countries and the pre-industrial phase in the advanced countries. In Agarwala, A. N. and Singh, S. P., editors, *The Economics of Underdevelopment*, chapter 2, pages 135–153. Oxford University Press, Oxford.

Kwiatkowski, D., Phillips, P. C., Schmidt, P., and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? *Journal of Econometrics*, 54(1-3):159–178.

Lamartina, S. and Zaghini, A. (2011). Increasing Public Expenditure: Wagner's Law in OECD Countries. *German Economic Review*, 12(2):149–164.

Landolfo, L. (2008). Assessing the sustainability of fiscal policies: Empirical evidence from the Euro Area and the United States. *Journal of Applied Economics*, 11(2):305–326.

Lanne, M., Lütkepohl, H., and Saikkonen, P. (2002). Comparison of unit root tests for time series with level shifts. *Journal of Time Series Analysis*, 23(6):667–685.

Lee, C.-C. and Chang, C.-P. (2006). Social security expenditure and GDP in OECD countries: A cointegrated panel analysis. *International Economic Journal*, 20(3):303–320.

Lee, O. and Shin, D. W. (2001). A note on stationarity of the MTAR process on the boundary of the stationarity region. *Economics Letters*, 73(3):263–268.

Leeper, E. M., Plante, M., and Traum, N. (2010). Dynamics of fiscal financing in the United States. *Journal of Econometrics*, 156(2):304–321.

Leeper, E. M., Walker, T. B., and Yang, S. S. (2013). Fiscal Foresight and Information Flows. *Econometrica*, 81(3):1115–1145.

Legrenzi, G. and Milas, C. (2002a). A Multivariate Approach to the Growth of Governments. *Public Finance Review*, 30(1):56–76.

Legrenzi, G. and Milas, C. (2002b). The Role of Omitted Variables in Identifying a Long-run Equilibrium Relationship for the Italian Government Growth. *International Tax and Public Finance*, 9(4):435–449.

Lindert, P. H. (1996). What Limits Social Spending? *Explorations in Economic History*, 33(1):1–34.

Linnemann, L. and Schabert, A. (2003). Fiscal Policy in the New Neoclassical Synthesis. *Journal of Money, Credit and Banking*, 35(6):911–29.

Llorca, M. and Redzepagic, S. (2008). Debt sustainability in the EU New Member States: empirical evidence from a panel of eight Central and East European countries. *Post-Communist Economies*, 20(2):159–172.

Loizides, J. and Vamvoukas, G. (2005). Government expenditure and economic growth: Evidence from trivariate causality testing. *Journal of Applied Economics*, 8(1):125–152.

Lott, J. R. and Kenny, L. W. (1999). Did Women's Suffrage Change the Size and Scope of Government? *Journal of Political Economy*, 107(6):1163–1198.

Lucas, R. J. (1976). Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(1):19–46.

Lumsdaine, R. L. and Papell, D. H. (1997). Multiple Trend Breaks and the Unit-Root Hypothesis. *Review of Economics and Statistics*, 79(2):212–218.

Lütkepohl, H., Saikkonen, P., and Trenkler, C. (2001). Maximum eigenvalue versus trace tests for the cointegrating rank of a VAR process. *Econometrics Journal*, 4(2):287–310.

MacKinnon, J. G. (1996). Numerical Distribution Functions for Unit Root and Cointegration Tests. *Journal of Applied Econometrics*, 11(6):601–618.

Magazzino, C. (2011). Disaggregated Public Spending, GDP and Money Supply: Evidence for Italy. *European Journal of Economics, Finance and Administrative Sciences*, 41:118–131.

Magazzino, C. (2012a). Wagner versus Keynes: Public spending and national income in Italy. *Journal of Policy Modeling*, 34(6):890–905.

Magazzino, C. (2012b). Wagner's Law and Augmented Wagner's Law in EU-27. A Time-

Series Analysis on Stationarity, Cointegration and Causality. *International Research Journal of Finance and Economics*, 59:205–220.

Mann, A. (1980). Wagner's law: An econometric test for Mexico, 1925-1970. *National Tax Journal*, 33(2):189–201.

Marcellino, M. (2006). Some stylized facts on non-systematic fiscal policy in the Euro area. *Journal of Macroeconomics*, 28(3):461–479.

Meltzer, A. H. and Richard, S. F. (1981). A Rational Theory of the Size of Government. *Journal of Political Economy*, 89(5):914–927.

Mertens, K. and Ravn, M. O. (2010). Measuring the Impact of Fiscal Policy in the Face of Anticipation: A Structural VAR Approach. *Economic Journal*, 120(544):393–413.

Milesi-Ferretti, G. M., Perotti, R., and Rostagno, M. (2002). Electoral Systems And Public Spending. *Quarterly Journal of Economics*, 117(2):609–657.

Mitchell, B. (2007). *International Historical Statistics (Europe 1750 - 2005).* Palgrave Macmillan, London, 3rd edition.

Monacelli, T. and Perotti, R. (2010). Fiscal Policy, the Real Exchange Rate and Traded Goods. *Economic Journal*, 120(544):437–461.

Morris, C. T. and Adelman, I. (1989). Nineteenth-century development experience and lessons for today. *World Development*, 17(9):1417–1432.

Mountford, A. and Uhlig, H. (2009). What are the effects of fiscal policy shocks? *Journal of Applied Econometrics*, 24(6):960–992.

Murthy, N. R. (1994). Wagner's Law, Spurious in Mexico or Misspecification: A Reply. *Public Finance/Finances publiques*, 49(2):295–303.

Musgrave, R. (1969). *Fiscal systems*. Yale University Press, New Haven.

National Institute of Statistics and Economic Studies (2013). Macro-economic database. Retrieved from http://www.bdm.insee.fr/.

Neck, R. and Sturm, J.-E. (2008). *Sustainability of Public Debt*. The MIT Press, Cambridge.

Office for National Statistics (2013). LMS Labour Market Statistics. Retrieved from http://www.ons.gov.uk/ons/rel/lms/labour-market-statistics/index.html.

Osterwald-Lenum, M. (1992). A Note with Quantiles of the Asymptotic Distribution of the Maximum Likelihood Cointegration Rank Test Statistics. *Oxford Bulletin of Economics and Statistics*, 54(3):461–472.

Oxley, L. (1994). Cointegration, Causality and Wagner's Law: A Test for Britain 1870-1913. *Scottish Journal of Political Economy*, 41(3):286–298.

Papadopoulos, A. and Sidiropoulos, M. (1999). The Sustainability of Fiscal Policies in the European Union. *International Advances in Economic Research*, 5(3):289–307.

Parker, J. A. (2011). On Measuring the Effects of Fiscal Policy in Recessions. *Journal of Economic Literature*, 49(3):703–18.

Payne, J. E. (1997). International evidence on the sustainability of budget deficits. *Applied Economics Letters*, 4(12):775–779.

Payne, J. E. (2003). A Survey of the International Empirical Evidence on the Tax-Spend Debate. *Public Finance Review*, 31(3):302–324.

Payne, J. E. and Ewing, B. T. (1996). International Evidence on Wagner's Hypothesis: A Cointegration Analysis. *Public Finance/Finances publiques*, 51(2):258–274.

Peacock, A. T. and Scott, A. (2000). The Curious Attraction of Wagner's Law. *Public Choice*, 102(1-2):1–17.

Peacock, A. T. and Wiseman, J. (1961). *The Growth of Public Expenditure in the United Kingdom*. Princeton University Press, Princeton.

Peltzman, S. (1980). The Growth of Government. *Journal of Law and Economics*, 23(2):209–287.

Perotti, R. (2005). Estimating the effects of fiscal policy in OECD countries. *Proceedings*.

Perotti, R. (2008). In Search of the Transmission Mechanism of Fiscal Policy. In Acemoglu, D., Rogoff, K., and Woodford, M., editors, *NBER Macroeconomics Annual 2007*, volume 22, chapter 3, pages 169–226. University of Chicago Press, Chicago.

Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

Plosser, C. I. (1987). Fiscal policy and the term structure. *Journal of Monetary Economics*, 20(2):343–367.

Plötscher, M., Seidel, T., and Westermann, F. (2005). Fiskalpolitik in Deutschland: Eine empirische Analyse am Beispiel des Vorziehens der Steuerreform. *Kredit und Kapital*, 38(1):23–51.

Potrafke, N. and Reischmann, M. (2012). Fiscal Equalization Schemes and Fiscal Sustainability. CESifo Working Paper Series 3948, CESifo GmbH.

Prohl, S. and Westerlund, J. (2009). Using Panel Data to Test for Fiscal Sustainability within the European Union. *FinanzArchiv/Public Finance Analysis*, 65(2):246–269.

Pryor, F. L. (1968). *Public expenditures in communist and capitalist nations.* Allen & Unwin, London.

Przywara, B. (2010). Projecting future health care expenditure at European level: drivers, methodology and main results. European Economy - Economic Papers 417, Directorate General Economic and Monetary Affairs (DG ECFIN), European Commission.

Quintos, C. E. (1995). Sustainability of the Deficit Process with Structural Shifts. *Journal of Business & Economic Statistics*, 13(4):409–417.

Ram, R. (1987). Wagner's Hypothesis in Time-Series and Cross-section Perspectives:. *Review of Economics and Statistics*, 69(2):194–204.

Ramey, V. A. (1993). How important is the credit channel in the transmission of monetary policy? *Carnegie-Rochester Conference Series on Public Policy*, 39(1):1–45.

Ramey, V. A. (2011a). Can Government Purchases Stimulate the Economy? *Journal of Economic Literature*, 49(3):673–685.

Ramey, V. A. (2011b). Identifying Government Spending Shocks: It's all in the Timing. *Quarterly Journal of Economics*, 126(1):1–50.

Ramey, V. A. and Shapiro, M. D. (1998). Costly capital reallocation and the effects of government spending. *Carnegie-Rochester Conference Series on Public Policy*, 48(1):145–194.

Rebelo, S. (2005). Real Business Cycle Models: Past, Present and Future. *Scandinavian Journal of Economics*, 107(2):217–238.

Reinhart, C. M. and Rogoff, K. S. (2009). The Aftermath of Financial Crises. *American Economic Review*, 99(2):466–472.

Rousseau, P. L. and Wachtel, P. (1998). Financial Intermediation and Economic Performance: Historical Evidence from Five Industrialized Countries. *Journal of Money, Credit and Banking*, 30(4):657–678.

Shelton, C. A. (2007). The size and composition of government expenditure. *Journal of Public Economics*, 91(11-12):2230–2260.

Shin, D. W. and Lee, O. (2001). Tests for Asymmetry in Possibly Nonstationary Time Series Data. *Journal of Business & Economic Statistics*, 19(2):233–244.

Sims, C. A. (1980). Macroeconomics and Reality. *Econometrica*, 48(1):1–48.

Sobhee, S. K. and Joysuree, V. (2004). The Implications of Openness for WAGNER'S Law. An International Comparison of 20 Countries, 1971-2000. *Applied Econometrics and International Development*, 4(3):47–58.

Tanzi, V. and Schuknecht, L. (2000). *Public Spending in the 20th Century. A Global Perspective*. Cambridge University Press, Cambridge.

Tenhofen, J., Wolff, G. B., and Heppke-Falk, K. H. (2010). The Macroeconomic Effects of Exogenous Fiscal Policy Shocks in Germany: A Disaggregated SVAR Analysis. *Jahrbuch für Nationalökonomie und Statistik/Journal of Economics and Statistics*, 230(3):328–355.

Thomson Reuters (2013). Datastream 5.0.

Thornton, J. (1999). Cointegration, causality and Wagner's Law in 19th century Europe. *Applied Economics Letters*, 6(7):413–416.

Timm, H. (1961). Das Gesetz der wachsenden Staatsausgaben. *FinanzArchiv/Public Finance Analysis*, 21(2):201–247.

Tong, H. (1990). *Non-Linear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford.

Tong, H. and Lim, K. (1980). Threshold Autoregression, Limit Cycles and Cyclical Data. *Journal of the Royal Statistical Society Series B*, 42(3):245–292.

Trehan, B. and Walsh, C. E. (1988). Common trends, the government's budget constraint, and revenue smoothing. *Journal of Economic Dynamics and Control*, 12(2-3):425–444.

Trehan, B. and Walsh, C. E. (1991). Testing Intertemporal Budget Constraints: Theory and Applications to U.S. Federal Budget and Current Account Deficits. *Journal of Money, Credit and Banking*, 23(2):206–223.

Uctum, M. and Wickens, M. (2000). Debt and Deficit Ceilings, and Sustainability of Fiscal Policies: An Intertemporal Analysis. *Oxford Bulletin of Economics and Statistics*, 62(2):197–222.

Uhlig, H. (2010). Some Fiscal Calculus. *American Economic Review: Papers and Proceedings*, 100(2):30–34.

Vanhorebeek, F. and van Rompuy, P. (1995). Solvency and Sustainability of Fiscal Policies in the EU. *De Economist*, 143(4):457–453.

Wagner, A. (1893). *Grundlegung der Politischen Ökonomie. Teil I: Grundlagen der Volkswirtschaft*. C.F. Winter'sche Verlagshandlung, Leipzig, 3rd edition.

Wagner, A. (1911). Staat (in nationalökonomischer Hinsicht). In Conrad, J., Elster, L., Lexis, W., and Loening, E., editors, *Handwörterbuch der Staatswissenschaften*, volume 7, pages 727–739. Gustav Fisher Verlag, Jena, 3rd edition.

Westerlund, J. and Prohl, S. (2010). Panel cointegration tests of the sustainability hypothesis in rich OECD countries. *Applied Economics*, 42(11):1355–1364.

Wilcox, D. W. (1989). The Sustainability of Government Deficits: Implications of the Present-Value Borrowing Constraint. *Journal of Money, Credit and Banking*, 21(3):291–306.

Wu, S.-Y., Tang, J.-H., and Lin, E. S. (2010). The impact of government expenditure on economic growth: How sensitive to the level of development? *Journal of Policy Modeling*, 32(6):804–817.

Zivot, E. and Andrews, D. W. K. (1992). Further Evidence on the Great Crash, the Oil-Price Shock, and the Unit-Root Hypothesis. *Journal of Business & Economic Statistics*, 10(3):251–270.