

An Analysis of Long-Term Influences on Financial Markets, Uncertainty and the Sustainability of Fiscal Balances

Inaugural-Dissertation
zur Erlangung des akademischen Grades eines Doktors
der Wirtschafts- und Sozialwissenschaften
der Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

vorgelegt von
Diplom-Volkswirt Malte Rengel
aus Emden

Göttingen, 2014

Gedruckt mit Genehmigung der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der Christian-Albrechts-Universität zu Kiel

Dekan:	Prof. Dr. Achim Walter
Erstberichterstattender:	Prof. Dr. Helmut Herwartz
Zweitberichterstattender:	Prof. Dr. Kai Carstensen

Tag der Abgabe der Arbeit: 01.10.2014

Tag der mündlichen Prüfung: 11.12.2014

Acknowledgments

I have written this thesis during my time as research assistant at the Institutes of Statistics and Econometrics at the University of Kiel and the University of Göttingen. I am grateful for the comprehensive and outstanding guidance by my supervisor Helmut Herwartz. His motivation and patience in our joint projects set the right impulses while writing the thesis. Additionally, I owe gratitude to my second supervisor Kai Carstensen. Moreover, I have benefitted from a productive and enjoyable working atmosphere created by my co-authors Steffen Henzel and Fang Xu.

I am thankful for the great time with all my colleagues in Kiel and Göttingen. I want to thank Vasyl Golosnoy, Uwe Jensen, Roman Liesenfeld and all participants of the Seminar on Statistics and Econometrics at the University of Kiel and further conferences for providing insightful feedback. Moreover, helpful inspiration emerged from conversations with Matthias Hartmann, Bastian Gribisch, Markus Pape and Yabibal Walle. Particularly, I have benefited from extensive econometric and non-econometric discussions with Jan Roestel with whom I shared an office for two years. The support in computational issues by Albrecht Mengel was very helpful.

Thomas Kneib and his team at the chair of statistics in Göttingen made available their support in professional nature, but also far beyond. I thank Hauke Rennie and Simone Maxand for taking care of coffee logistics and Alexander Sohn for organizing several ‘Büro des Monats’ contests. Moreover, I gratefully acknowledge financial support from the German Research Foundation.

I want to stress my gratitude to Britta, my family and friends who encouraged and distracted me through all challenging times while writing this thesis.

Malte Rengel
Göttingen, September 2014

Contents

Introduction	1
1 A slowly evolving mean of the price-to-dividend ratio, its economic influences and predictive power for stock returns	7
1.1 Introduction	7
1.2 A state space model of the PtDR	10
1.2.1 Persistence of the PtDR	10
1.2.2 A state space approximation	13
1.2.3 Model implementation	15
1.2.4 Estimates and diagnostics	16
1.3 Forecasting performance	18
1.3.1 Predictive regressions	18
1.3.2 Forecast evaluation	20
1.4 Long-run determinants of the PtDR	26
1.4.1 The three long-term determinants	26
1.4.2 Cointegration analysis	30
1.5 Conclusions	35
Appendix to Paper 1	36
1.A Data description	36
1.B Particle filtering	38
1.C Approximation errors	39
2 Dimensions of macroeconomic uncertainty: A common factor analysis	43
2.1 Introduction	43
2.2 The dataset	45
2.3 Measuring macroeconomic uncertainty	46
2.3.1 Individual uncertainty measures	46
2.3.2 Measures of macroeconomic uncertainty	49
2.4 Estimation results	50
2.4.1 The number of common uncertainty factors	50

2.4.2	The distinct types of macroeconomic uncertainty	52
2.4.3	Indicators of macroeconomic uncertainty	58
2.4.4	Macroeconomic uncertainty and real activity	63
2.5	Conclusions	66
	Appendix to Paper 2	68
2.A	Dataset and forecast error variance decomposition for individual un- certainty measures	68
2.B	Estimation of the dynamic factor model	72
2.C	Tests for the number of factors	73
2.D	Bootstrap procedure for impulse responses of the dynamic factor model	75
2.E	Identification of the indicators of macroeconomic uncertainty	75
2.F	Comparison of macroeconomic uncertainty with alternative uncer- tainty measures	77
2.G	Dynamic relation between fundamental shocks u_t and VXO, SPREAD, JLN and EPU	80
2.H	Robustness analyses	81
3	Size corrected inference in fiscal policy reaction functions: A three country assessment	89
3.1	Introduction	89
3.2	Size distortions in policy reaction functions	93
3.2.1	Data	93
3.2.2	The policy reaction function	95
3.2.3	Unit root characteristics	96
3.2.4	Size distortions of policy response diagnostics	99
3.3	Size control for model based IBC diagnosis	101
3.3.1	Monte Carlo inference	102
3.3.2	Local to unit root asymptotics	103
3.4	Size controlled inference on fiscal sustainability	104
3.4.1	Size distortions in policy response regressions	104
3.4.2	Size corrected diagnosis of fiscal balances' sustainability . . .	106
3.4.3	The sustainability of fiscal balances in recent years	109
3.4.4	Summary of size controlled diagnostics	113
3.5	Conclusions	113
	Appendix to Paper 3	115
3.A	Data sources	115
3.B	Unit root tests for non-debt components	116

4 Sustainability of European fiscal balances – Just a statistical artifact?	117
4.1 Introduction	117
4.2 A policy reaction function for European fiscal data	121
4.2.1 The policy reaction function in the panel data framework . .	121
4.2.2 The dataset and the empirical implementation	122
4.2.3 Persistence characteristics	124
4.3 Size distortions in pooled policy reaction functions	126
4.3.1 The second-order bias	126
4.3.2 A Monte Carlo study on biases and size distortions	127
4.4 Size controlled inference on fiscal sustainability in Europe	131
4.4.1 A wild bootstrap approach	132
4.4.2 The sustainability of European fiscal balances	133
4.4.3 Results for the Eurozone and Europe excluding Greek data .	135
4.4.4 Nonlinear policy reactions	138
4.4.5 Characteristics of post-2008 fiscal figures	142
4.5 Conclusions	144
Appendix to Paper 4	145
4.A Data sources	145
4.B The second-order bias if the cross section is large	145
4.C Empirical correlation structures and wild bootstrap counterparts . .	146
4.D Recursive demeaning	147
 Concluding remarks	 151
 Bibliography	 168

List of Tables

1.1	Unit root tests for the PtDR	11
1.2	Parameter estimates and model evaluations	17
1.3	In-sample predictive regressions	21
1.4	Out-of-sample predictive regressions for stock returns	22
1.5	Unit root tests for economic determinants	31
1.6	Cointegration analysis	32
1.7	Cross validations	34
2.1	Regressions of alternative measures on macroeconomic uncertainty .	62
2.2	Forecast error variance explained by macroeconomic uncertainty . .	65
2.3	Description of the dataset and forecast error variance decomposition for individual uncertainty measures	71
2.4	Data transformations	71
2.5	Tests to determine the number of common factors	74
2.6	Tests to determine the number of fundamental shocks	74
2.7	Forecast error variance in a VAR comprising both types of uncertainty	87
2.8	FEVD using the VAR specified by Bloom (2009)	87
3.1	Unit root tests for debt-to-GDP ratios and primary surpluses	97
3.2	Pre-tests for size distortions	105
3.3	Policy response regressions for the US and the UK	108
3.4	Policy response regressions for Portugal	109
3.5	Policy response regressions for recent US and UK data	111
3.6	Policy response regressions for recent Portuguese data	112
3.7	Policy response regressions for Portugal excluding the era of circu- lating Euro banknotes	112
3.8	Unit root tests for non-debt components	116
4.1	Panel unit root tests	125
4.2	Estimation results for policy reaction functions	134
4.3	Estimation results for policy reaction functions excluding Greek data	137

4.4	Estimation results for nonlinear policy reaction functions	139
4.5	Marginal effects in nonlinear policy reaction functions	141
4.6	Estimation results for policy reaction functions until 2013	143
4.7	Cross-correlation matrix of $\hat{u}_{i,t}$ and $u_{i,t}^*$	146
4.8	Correlation of $\hat{u}_{i,t}$ and $\hat{\epsilon}_{i,t}$ and corresponding bootstrap versions . .	147

List of Figures

1.1	The price-to-dividend ratio	11
1.2	Rolling mean of the PtDR	12
1.3	Time-varying mean of the PtDR	17
1.4	The PtDR and distinct mean evaluations	20
1.5	Out-of-sample forecasting performance relative to the historical mean	24
1.6	Out-of-sample forecasting performance relative to the break adjusted model of the PtDR	25
1.7	Low frequency consumption volatility	28
1.8	Economic influences and the time-varying mean of the PtDR	29
1.9	Evaluation of approximations	41
2.1	RiskMetrics and SV measures	49
2.2	Explained variance proportions for distinct numbers of factors	51
2.3	Impulse responses of individual uncertainty measures (1)	54
2.4	Impulse responses of individual uncertainty measures (2)	55
2.5	The oil price uncertainty shock and an oil-market specific demand shock	58
2.6	Indicators of macroeconomic uncertainty	59
2.7	Indicators of macroeconomic uncertainty and the VXO	61
2.8	Impulse response functions for a shock in macroeconomic uncertainty	64
2.9	Indicators of macroeconomic uncertainty and the SPREAD	77
2.10	Indicators of macroeconomic uncertainty and the JLN measures	78
2.11	Indicators of macroeconomic uncertainty and EPU	79
2.12	Dynamic relation between fundamental shocks u_t and VXO, SPREAD, JLN and EPU	80
2.13	Impulse responses of individual uncertainty measures for an alterna- tive identification strategy (1)	82
2.14	Impulse responses of individual uncertainty measures for an alterna- tive identification strategy (2)	83
2.15	Indicators of macroeconomic uncertainty from in-sample and out-of- sample forecast errors.	84

2.16	Impulse response functions of IP using uncertainty indicators from a three-factor model	85
2.17	Impulse response functions of IP for distinct VAR specifications	86
3.1	Fiscal data for the US and the UK	94
3.2	Fiscal data for Portugal	95
3.3	Bias of policy the reaction coefficient	101
3.4	Size distortion of the robust t -test	102
4.1	Cross sectional averages of primary surpluses and debt-to-GDP ratios	124
4.2	Simulated distributions of the estimated policy reaction coefficient and the t -statistic	128
4.3	Biases and size distortions for distinct values of δ and β	129
4.4	Biases and size distortions for distinct values of N and T	131
4.5	Nonlinear policy reaction to increasing debt-to-GDP ratios	140
4.6	Simulated distributions of the policy reaction coefficient and the t -statistic for $N = 200$ and $T = 30$	145
4.7	Biases and size distortions of recursive-demeaning estimators	149

Introduction

In recent years, the world economy faced the largest economic downturn since World War II. This recession was characterized by, among others, three properties: trembling financial markets, increasing uncertainty, and, in the aftermath of the recession, exploding public debt in major industrialized economies. The joint appearance of these characteristics stresses that a profound understanding of economic dynamics and interrelations is crucial for informative policy monitoring and for the conduct of effective policy intervention.

This thesis contains four self-contained studies. Three of them have been written in cooperation with co-authors (declared below) and one is single-authored. These studies focus on historical and recent dynamics of financial and macroeconomic data. I discuss the interrelation of financial markets and the macroeconomy, the role of uncertainty and the sustainability of public debt. The informational content of time series and panel data is investigated by means of linear and nonlinear state space modeling and regression techniques. The first part of the thesis addresses the reoccurring and large upward swings of stock prices during the post-WWII period. While such patterns are often regarded as being of speculative nature, we diagnose them to relate to changes in economic fundamentals. We emphasize the role of consumption uncertainty in explaining long-term dynamics of asset prices. In a subsequent study, this thesis addresses general macroeconomic uncertainty and reveals that distinct types thereof affect real activity. The second half of this thesis focuses on the accumulation and reduction of public debt, both in a time series and an international panel data context. The results highlight that an appropriate treatment of statistical data characteristics is essential for a reliable and timely assessment of the sustainability of fiscal policies. Next, I describe the content, methods and findings of the distinct studies in more detail.

At first, *Paper 1* explores the low frequency dynamics of the price-to-dividend ratio aggregated from US stock market data during the past century, with a focus on the post-WWII period. It is joint work with Helmut Herwartz and Fang Xu. The price-to-dividend ratio is a so-called financial valuation ratio relating stock prices to fundamental values. The price-to-dividend ratio condenses information of stock

prices and dividend payments and, thus, should in theory bear predictive content for future stock returns. As documented in the empirical literature, this ratio is characterized by persistent movements and its predictive content is relatively low. In particular, in the 1990s, when the price-to-dividend ratio sharply accelerates, its predictive content diminishes (Lettau and Van Nieuwerburgh; 2008; Goyal and Welch; 2008).

To account for the persistent movements of the price-to-dividend ratio, we embed the present value model of Campbell and Shiller (1988) in a nonlinear state space framework and generalize it to allow for a latent slowly evolving mean process of the price-to-dividend ratio. Our flexible model is supported by the data compared with the model of Campbell and Shiller (1988) that relies on the assumption of a constant mean. First, we investigate if the time-varying mean process of the US price-to-dividend ratio can be explained by changes in three factors which have been proposed in the literature as determinants of the conditions on financial markets and in the aggregate economy. These factors are: consumption uncertainty (Bansal and Yaron; 2004), the demographic structure of the population (Geanakoplos, Magill and Quinzii; 2004) and the share of firms with traditional dividend payout policy (Kim and Park; 2013). We reveal that all triggers jointly determine the shape of the gradually evolving mean of the price-to-dividend ratio. Among them, consumption risk is the dominant factor in the post-WWII period. Thus, we provide comprehensive evidence that macroeconomic surroundings influence the long-term constitution of financial markets. Second, we employ so-called predictive regressions to relate the movements of transitory stock returns to variations of the lagged price-to-dividend ratio that is highly persistent. Due to the distinct persistence characteristics of left and right hand side variables, these regressions are poorly balanced and require non-standard econometric treatment to evaluate the predictive content of the price-to-dividend ratio. If the price-to-dividend ratio is adjusted for its time-varying mean process, it has ex-ante predictive content for stock returns. This predictive content is missed if applying the unadjusted measure as predictor.

So far, this thesis stresses the role of consumption uncertainty for the interrelations between the macroeconomy and financial markets. However, uncertainty is inherent in the dynamics of almost all economic quantities, not only in those of consumption. *Paper 2* examines the influence of general economic uncertainty on real activity and is joint work with Steffen Henzel. To analyze uncertainty, one has to acknowledge that uncertainty is intrinsically latent. Multiple approaches for the approximation of uncertainty have been applied in the literature. We develop

a measure of general macroeconomic uncertainty which affects the movements of key macroeconomic aggregates. First, we derive series of time-varying volatility for numerous economic and financial indicators for the last decades. Second, we summarize the joint dynamics of these measures by means of a dynamic factor model. This model allows to interpret the common factors as indicators of macroeconomic uncertainty. In related work, Jurado, Ludvigson and Ng (2014) focus on the single factor with the greatest explanatory power. In contrast, we investigate distinct dimensions of macroeconomic uncertainty. To be precise, we discuss how many factors describe the common movements of individual uncertainty measures and show that these factors have distinct economic interpretations.

From an econometric perspective, the modeling approaches in Paper 1 and Paper 2 are closely related since both are settled in the state space framework. The purpose of modeling a latent state process of the price-to-dividend ratio in a non-linear manner is to enlarge the flexibility of the present value model. The purpose of factor modeling is to condense the joint information of numerous sources in a few manageable indices. These indices are linearly linked to the data. We extract one index covering general ‘business cycle uncertainty’ and one index covering ‘oil and commodity price uncertainty’. We document that these indices peak at business cycle and oil price related events, respectively. Moreover, both indicators move countercyclically and reach their largest values during the financial crisis. Positive shocks to oil and commodity price uncertainty cause more severe economic downturns compared with shocks to business cycle uncertainty. A clear distinction of the dimensions of macroeconomic uncertainty is important from the perspective of policy monitoring.

The current European debt crisis questions if general fiscal policies are designed in a sustainable manner. *Paper 3* and *Paper 4* examine the dynamics of public debt to explore the challenges of increasing debt burdens. *Paper 3* is joint work with Helmut Herwartz and employs historical data for more than a century to test for the sustainability of fiscal balances. We compare US fiscal figures with those of the UK, currently characterized by a medium level of public debt, and those of Portugal, currently characterized by a high level of public debt. We test if fiscal balances accord with an intertemporal budget constraint (IBC). Bohn (1995) derives this IBC from a stochastic general equilibrium model.¹ Technically, the merit of the so-called model based sustainability framework is that empirical sustainability tests can be implemented in a straightforward way by regressing primary surpluses as

¹ This model draws on an endowment economy described by Lucas (1978).

share of GDP on lagged debt-to-GDP ratios (Bohn; 1998). If the primary surplus of an economy reacts positively to increases of lagged debt-to-GDP ratios, fiscal policy is classified as sustainable.

Debt-to-GDP ratios are highly persistent and already Woodford (1998) points to the correlation of primary surpluses and the innovations in debt-to-GDP ratios in the context of the ‘fiscal theory of the price level’. In the related finance literature, Stambaugh (1999) shows that estimated slope coefficients are (upward) biased and corresponding t -tests are marked by severe size distortions if lagged regressors are persistent and if their innovations are (negatively) correlated with the innovations of the depended variable. However, these data characteristics have so far been neglected in policy reaction regressions. Resulting overrated type I error probabilities of t -tests bear the risk to reject the null hypothesis of non-sustainability too often. We diagnose test sizes that are up to five times larger than the nominal reference in policy reaction regressions for the US, the UK and Portugal. We derive size controlled inferential diagnostics in policy reaction functions by means of Monte Carlo techniques and local to unit root asymptotics. US and UK post-WWII fiscal figures are spuriously indicated to be sustainable conditional on Gaussian quantiles while size corrected critical values lead to the opposite conclusion. However, US and UK fiscal balances are sustainable in samples covering more than 100 years. For these (sub)samples t -statistics are not in a critical region where test decisions change if size corrected critical values are applied instead of asymptotic critical values. Fiscal balances of Portugal are non-sustainable throughout.

The estimation of policy reaction functions in Paper 3 is closely related to the estimation of predictive regressions in Paper 1. While in the latter stock prices are measured relative to fundamental dividend payments, public debt is measured in terms of GDP. Both ratios are characterized by persistent accelerations over time. Moreover, the econometric structure of the predictive regressions and the policy reaction regressions is very similar. Mean reverting transitory processes, i.e., stock returns and primary surpluses, are regressed on lagged persistent variables, i.e., price-to-dividend and debt-to-GDP ratios. While we correct the price-to-dividend ratio in Paper 1 for its slowly evolving mean to account for its persistence, we employ Monte Carlo techniques and local to unit root asymptotics for the analysis of the sustainability of fiscal balances.

For actual policy monitoring and a timely induction of corrective policy measures, the analysis of more recent fiscal data is often advisable. This issue is addressed in *Paper 4* which is single-authored. I build upon a panel data framework to investi-

gate the sustainability of recent European fiscal policies. Such panel based policy reaction regressions, as proposed by Mendoza and Ostry (2008) and Ghosh, Kim, Mendoza, Ostry and Qureshi (2013), attract attention in the current literature. Compared to the single-equation approach in Paper 3, I enlarge the cross sectional dimension of the data and consider 14 European countries simultaneously. This allows to restrict the analysis to more recent data from 1980 to 2008. The goal of this study is to assess if panel based inference on policy reaction functions equally suffers from size distorted tests as in single-equation frameworks and if overrated type I error probabilities have possibly prevented a timely detection of the increased solvency risk of European countries before the debt crisis manifested in 2009.

Similarly, Hjalmarsson (2008, 2010) embeds predictive regressions for financial valuation ratios in the panel data framework. He argues that more severe size distortions occur in panel based predictive regressions compared with those in single-equation regressions. These larger distortions are caused by the so-called second-order bias. This bias is induced by the within-transformation of data that accounts for country specific effects if lagged financial valuation ratios are persistent and their innovations are correlated with these of the dependent variable. I reveal that size distortions accumulate over an increasing cross sectional dimension of data and that conventional t -tests spuriously indicate European fiscal figures to be sustainable in the pre-crisis period. In contrast, size corrected bootstrap based critical values would have facilitated a timely detection of non-sustainable fiscal policies.

The investigation of the sustainability of European fiscal figures and the influence of uncertainty on the macroeconomy in Paper 2 are related as they both allow timely policy monitoring. Both studies focus on economic dynamics during the last decades. From an econometric perspective, the conclusions of these two studies are inferred from panel datasets. While the indicators of macroeconomic uncertainty are condensed from a panel of numerous individual uncertainty measures, the analysis of European fiscal sustainability is based on a panel describing the dynamics of fiscal data of various countries.

Finally, this thesis concludes and provides an outlook on directions that economic and econometric research may pursue in the future in order to obtain a deeper understanding of the interrelated dynamics of key macroeconomic and financial indicators.

Paper 1

A slowly evolving mean of the price-to-dividend ratio, its economic influences and predictive power for stock returns*

Helmut Herwartz, Malte Rengel and Fang Xu

Abstract. The potential forces behind persistent variations of the log price-to-dividend ratio (PtDR) and their implications for the prediction of stock returns have attracted a lively discussion in the literature. We estimate a gradually varying mean of the PtDR in the framework of a present value model, which is then used to adjust the PtDR in predictive regressions. In real time forecasting the proposed predictor outperforms the unadjusted PtDR and an adjustment of the PtDR by means of discrete shifts. We show that during the past 60 years this slowly evolving mean process is jointly shaped by the consumption risk, the demographic structure of the population and the proportion of firms with traditional dividend payout policy. In particular, the volatility of consumption growth plays the dominant role.

1.1 Introduction

Aggregate stock prices have risen to unprecedented levels in the late 1990s relative to any fundamental values. The logarithmic price-to-dividend ratio (PtDR), for example, has increased by 26% from 1980 to 2013. Even after a substantial decline since its peak in 2000, its level is still far away from its historical values. There is evidence for structural breaks or instability in the mean of the PtDR (Lettau and Van Nieuwerburgh; 2008) and in the relation between the PtDR and future stock returns (Paye and Timmermann; 2006; Rapach and Wohar; 2006; Goyal and Welch; 2008). Empirical evidence indicates that the increasing mean of the PtDR could be

* For a different version of this paper see ‘Persistence in the price-to-dividend ratio and its macroeconomic fundamentals’, *Conference Paper – Jahrestagung des Vereins für Socialpolitik*, 2013. Another version of this paper circulated under the title ‘Stock return predictability and persistence in the price-to-dividend ratio’.

due to a persistent fall in macroeconomic risk measured as the volatility of consumption growth rates (Bansal and Yaron; 2004; Bansal, Khatchatrian and Yaron; 2005; Lettau, Ludvigson and Wachter; 2008; Bansal, Kiku and Yaron; 2010), changes in demographic structure of the population (Geanakoplos et al.; 2004; Favero, Gozluklu and Tamoni; 2011), and the dividend pay-out policy by firms (Fama and French; 2001; Robertson and Wright; 2006; Boudoukh, Michaely, Richardson and Roberts; 2007; Kim and Park; 2013). Coping with the persistence of the PtDR, Lettau and Van Nieuwerburgh (2008) suggest a regime-switching model that allows discrete mean shifts. They show that deviations from shifting means of the PtDR carry predictive power for stock returns in-sample but fail to signal stock returns ex-ante compared with the historical average return as a benchmark predictor.

In this paper we consider a continuously time-varying mean of the PtDR that not only forecasts returns out-of-sample but also enables simultaneous testing of distinct determinants of the mean of the PtDR. We propose a latent variable reflecting the slowly evolving mean of the PtDR within a generalized version of the present value model introduced by Campbell and Shiller (1988). Particle filtering (e.g., Cappé, Godsill and Moulines; 2007) is employed to estimate this nonlinear state space model. The continuously varying mean is entirely determined by the data and includes the (succession of) discrete mean shifts as a special case. The application of the state space model in this paper differs from previous applications in forecasting returns that treat the expected return and expected dividend growth as two latent processes, see among others, Binsbergen and Koijen (2010) and Rytchkov (2012). We model the mean of the PtDR as a latent process, which can be interpreted as a combination of local means of expected returns and expected dividend growth. Since we do not assume an exogenous fixed mean of the PtDR as previous studies, we use the present value model of the PtDR as an estimation equation rather than an identity restriction.¹

We find that a time-varying mean of the PtDR is strongly supported by log-likelihood diagnostics. The estimated long-term state has step-like patterns similar to mean shifts with two structural breaks as suggested by Lettau and Van Nieuwerburgh (2008). In contrast to discrete mean shifts, however, the time-varying mean evolves slowly over time. The slowly evolving process allows a simple projection towards the future, and straightforward implementation of standard predictive regressions for stock returns conditional on this information. We find that local de-

¹ In our framework, it is not straightforward to treat expected returns and expected dividend growth as latent state variables simultaneously.

variations of the PtDR from its varying mean carry out-of-sample predictive power. Using the out-of-sample degree of explanation based on the root mean squared error (RMSE) (Goyal and Welch; 2003), we confirm the significance of the out-of-sample forecasting performance in comparison with both historical average returns and PtDR adjustments by means of discrete mean shifts.

Extracting a continuously varying mean of the PtDR offers a unique opportunity to examine its potential influences and provides an economic interpretation of the out-of-sample predictive power of adjusting the PtDR by its slowly evolving mean. Following an error correction approach, we investigate the above mentioned three factors that have been documented to affect the PtDR in a long-run manner – consumption risk, the demographic structure, and the dividend payout policy of firms. We find that all three variables jointly shape the slowly evolving mean of the PtDR during the past 60 years, with consumption risk playing the most important role. A low consumption risk drives down the equity premia and pushes up the stock price (Bansal and Yaron; 2004). The decreasing volatility in the consumption growth rate has the highest contribution in explaining the increasing mean of the PtDR. A high middle-aged to young ratio, corresponding to excess demand for saving, drives up the equilibrium asset prices (Geanakoplos et al.; 2004). The significant increases in the mean of the PtDR in the 1990s are consistent with increases in the middle-aged to young ratio during this same period. In addition to the macroeconomic and demographic influences, lowered dividends can affect the long-run relationship between stock prices and dividends (Kim and Park; 2013). The fall in the proportion of firms that payout a significant fraction of their earnings in form of dividends since the 1980s is consistent with the increasing mean of the PtDR. Nevertheless, among the three factors this has the smallest contribution in explaining the variations in the mean of the PtDR.

Section 1.2 illustrates the persistence of the PtDR, sketches its implications for the standard present value model, and introduces the state space model of the PtDR incorporating a smoothly varying mean. The forecasting model, evaluation methods and forecasting performance are discussed in Section 1.3. In Section 1.4 we investigate the linkage between the time-varying mean of the PtDR and its potential influences. Section 1.5 concludes. Appendices provide detailed descriptions of the data (Appendix 1.A), the particle filtering approach (Appendix 1.B), and approximation errors involved in the derivation of the present value model (Appendix 1.C).

1.2 A state space model of the PtDR

In this section we first discuss the observed persistence of the PtDR and its implications for respective present value formulations. Then a latent time-varying mean of the PtDR is formally derived and estimated, which is in line with the diagnosed stochastic trends governing the PtDR. Log-likelihood statistics support the view that the present value model of the PtDR incorporating a continuous time-varying mean outperforms the model with a constant mean.

1.2.1 Persistence of the PtDR

The persistent increase of stock prices relative to dividends from 1980 to 2000 can be seen from Figure 1.1. We find that the PtDR can be well described by a non-stationary process, which confirms findings in previous studies, see for example Campbell (1999), Herwartz and Morales-Arias (2009) and Park (2010). Using annual CRSP data from 1926 to 2013 and S&P500 data from 1871 to 2013, Table 1.1 documents results from numerous unit root tests. The hypothesis of a non-stationary PtDR cannot be rejected with 5% significance by means of the ADF test and tests proposed by Phillips and Perron (1988, PP) and Elliott, Rothenberg and Stock (1996, DF_{GLS}).²

The PtDR is unlikely to be a stationary process even taking into account the power weakness of unit root tests under near integration. As can be seen from the last column of Table 1.1, the null hypothesis of stationarity of the PtDR is rejected by means of the KPSS statistic (Kwiatkowski, Phillips, Schmidt and Shin; 1992). Moreover, we find that the PtDR can be better described by a non-stationary process than by a stationary process with a structural break at unknown timing. Testing the unit root hypothesis and allowing for a structural break in the mean of the PtDR as proposed by Perron and Vogelsang (1992) provides the respective supporting evidence (see column 5 in Table 1.1, PV).

As noted by Campbell (2008), the persistence of the PtDR challenges the present value model in Campbell and Shiller (1988) that rests on the assumption of a

² It is worthwhile to mention that opposite to pure random walks diagnosed by common unit root tests, actual PtDR processes cannot grow to any level. Recently, bounded non-stationary processes have attracted interest in the econometric literature (Cavaliere; 2005). Cavaliere and Xu (2014) have proposed a novel ADF based approach to test for unit roots in the presence of bounds. The critical values of such tests are smaller (i.e., larger in absolute value) than those of unit root tests neglecting the bounded nature of a variable of interest. Thus, if common unit root tests hint at non-stationarity, bounded non-stationarity will be diagnosed once the bounds are taken into account.

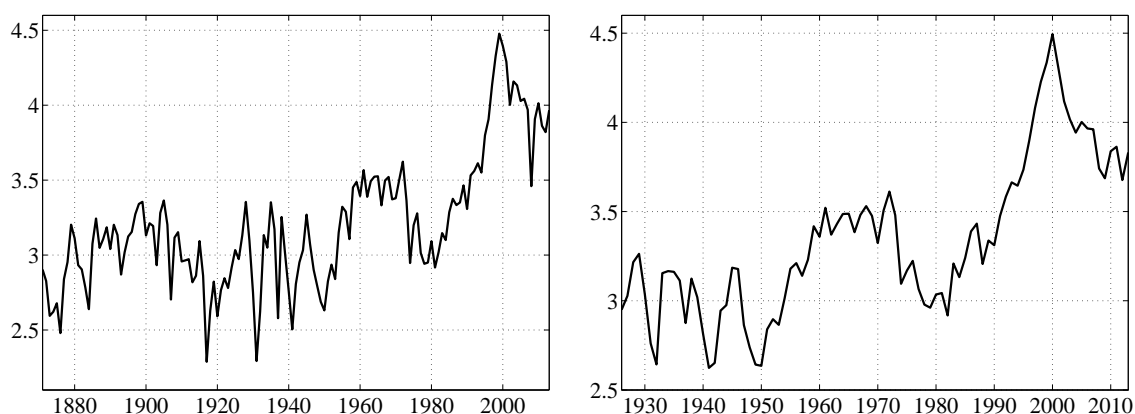


Figure 1.1: The price-to-dividend ratio The PtDR from S&P500 data from 1871 to 2013 is depicted in the left hand side panel. The right hand side panel illustrates the PtDR from CRSP data from 1926 to 2013.

	ADF	PP	DF _{GLS}	PV	KPSS
CRSP 1926-2013					
Test Statistics	-1.595	-1.604	-1.171	-3.553	0.801
Crit. val. at 5%	-2.895	-2.895	-1.945	-4.580	0.463
Crit. val. at 10%	-2.585	-2.585	-1.614	-4.240	0.347
S&P500 1871-2013					
Test Statistics	-1.783	-1.841	-1.139	-3.737	0.906
Crit. val. at 5%	-2.882	-2.882	-1.943	-4.580	0.463
Crit. val. at 10%	-2.578	-2.578	-1.615	-4.270	0.347

Table 1.1: Unit root tests for the PtDR The upper panel covers test results for CRSP data ranging from 1926 to 2013 and the lower panel those for S&P500 data ranging from 1871 to 2013. Test regressions include a constant. ADF refers to the Augmented Dickey-Fuller test where the lag selection criterion is the BIC. For PP, the test statistic considered in Phillips and Perron (1988), the spectral AR estimator is used to calculate the long-run variance. DF_{GLS} refers to the test proposed by Elliott et al. (1996) where the BIC is applied to determine the lag length. For the PV-test proposed by Perron and Vogelsang (1992) the innovation outlier model is applied and the lag length is determined by means of a t -test procedure. A Bartlett Kernel is applied in the KPSS-test of Kwiatkowski et al. (1992).

stationary PtDR. Let P_t and D_t denote stock prices and the corresponding dividends in time t , respectively. The total log-return, realized at the end of period $t + 1$, $r_{t+1} = \ln(P_{t+1} + D_{t+1}) - \ln(P_t)$, can be formulated as a nonlinear function of the PtDR, $\eta_t = \ln(P_t) - \ln(D_t)$,

$$r_{t+1} = -\eta_t + \ln(\exp(\eta_{t+1}) + 1) + \Delta d_{t+1}, \quad (1.1)$$

where $d_t = \ln(D_t)$ and Δ is shorthand for the first difference operator such that, e.g., $\Delta d_t = d_t - d_{t-1}$. A first order Taylor expansion around a fixed steady state $\bar{\eta}$

provides the linear approximation

$$r_{t+1} \simeq \kappa - \eta_t + \rho\eta_{t+1} + \Delta d_{t+1}, \quad (1.2)$$

with

$$\rho \equiv 1/(1 + \exp(-\bar{\eta})) \text{ and } \kappa \equiv -\ln(\rho) - (1 - \rho) \ln(1/\rho - 1). \quad (1.3)$$

In the empirical analysis, the constant parameter $\bar{\eta}$ is assumed to be known and commonly approximated by the sample mean (e.g., Campbell; 1999). Under persistent behaviour of the PtDR, $\bar{\eta}$ is unlikely to be constant and ρ becomes also time-varying. Figure 1.2 illustrates the time variation of sample means of the PtDR from rolling time windows covering observations from the most recent 20 years.

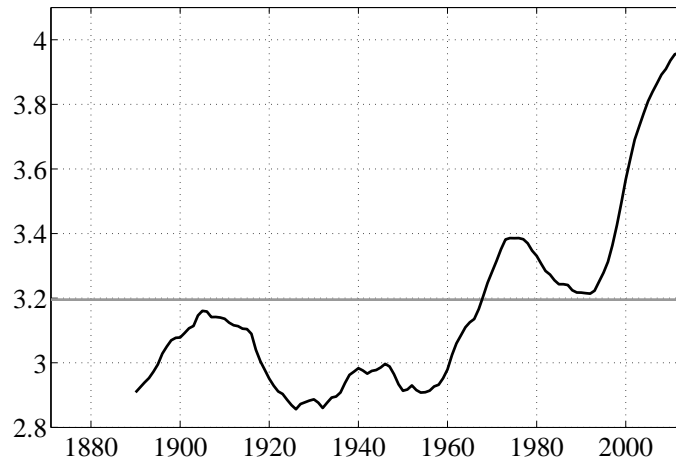


Figure 1.2: Rolling mean of the PtDR The figure illustrates the rolling window sample mean of the price-to-dividend ratio (η_t) from CRSP data that are linked to S&P500 data before 1926. The estimation windows includes the most recent 20 years of observations.

It is worthwhile to point out that a time-varying mean of the PtDR is well in line with its diagnosed non-stationarity. Even when assuming that the PtDR is a stationary but nearly integrated process, a time-varying mean could be regarded as a finite sample approximation of the local mean.

1.2.2 A state space approximation

Taking a time-varying mean of the PtDR into account, we modify the traditional present value model of the PtDR.³ Let $\tilde{\eta}_t$ denote the local mean employed to expand the Taylor approximation of the one-step-ahead stock returns in (1.2). Then, we obtain

$$r_{t+1} \simeq \kappa_t - \eta_t + \rho_t \eta_{t+1} + \Delta d_{t+1}, \quad (1.4)$$

with both parameters (κ_t and ρ_t) in (1.4) becoming time-specific, i.e.,

$$\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t)) \text{ and } \kappa_t \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(1/\rho_t - 1). \quad (1.5)$$

To derive the present value formulation of the PtDR from (1.4), the following approximations similar to those in Lettau and Van Nieuwerburgh (2008) are adopted: $E_t[\rho_{t+i}] \approx \rho_t$, $E_t[\kappa_{t+i}] \approx \kappa_t$ and $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$. Simulation studies documented in Appendix 1.C show that respective approximation errors are negligible for typical values of $\tilde{\eta}_t$. Taking the conditional expectation and iterating equation (1.4) forward provides the log-linear present value formulation of the PtDR

$$\eta_t \simeq \frac{\kappa_t}{1 - \rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} E_t[\Delta d_{t+i}^e - r_{t+i}^e] + \lim_{i \rightarrow \infty} \rho_t^i E_t[\eta_{t+i}], \quad (1.6)$$

where superscripts e symbolize the excess of dividend growth rates ($\Delta d_{t+i}^e = \Delta d_{t+i} - r_{t+i}^f$) or of returns ($r_{t+i}^e = r_{t+i} - r_{t+i}^f$) over the risk-free interest rate r_t^f . Changes in the long-term state of the PtDR affect the observed PtDR in a nonlinear fashion. A time-varying $\tilde{\eta}_t$ leads to a time-varying rather than a constant intercept term $\kappa_t/(1 - \rho_t)$, and the future return adjusted dividend growth rates are discounted at time-varying rates ρ_t rather than at a constant one. This casts a new light on the connection between stationary stock returns and dividend growth rates with the persistent PtDR. Under a constant $\bar{\eta}$, the PtDR should be stationary if stock returns and dividend growth are stationary. However, under time-varying $\tilde{\eta}_t$, the PtDR could have a time trend through the intercept term. Allowing for a time-varying $\tilde{\eta}_t$ might explain PtDR dynamics which are not fully captured by time invariant valuation of return-adjusted cash flows.

³ Herwartz and Xu (2014) follow a similar approach to model the house price-to-rent ratio.

An intuitive way to link equation (1.6) to the traditional present value model in Campbell and Shiller (1988) is to reconsider it from the perspective of an investor who can only quantify the mean of the PtDR conditional on past information. In this case, as shown in Lacerda and Santa-Clara (2010), the mean of the PtDR becomes time-varying and one can introduce directly a time index t for the parameters ρ and κ in (1.3) in the traditional present value model to derive equation (1.6) (see also Figure 1.2). Against this background the proposed model offers a structural interpretation for this approach.

We employ a state space model to estimate the latent time-varying $\tilde{\eta}_t$. Assume a random disturbance term $\epsilon_t \sim N(0, \sigma_\epsilon^2)$ to capture eventual rational bubbles, approximation errors, and other influences in $\lim_{i \rightarrow \infty} \rho_t^i E_t[\eta_{t+i}]$. Further substituting E_t in (1.6) by objective expectations conditional on the information set at the end of period t (\tilde{E}_t), equation (1.6) is transformed into the measurement equation

$$\eta_t = \frac{\kappa_t}{1 - \rho_t} + \sum_{i=1}^{\infty} \rho_t^{i-1} \tilde{E}_t[\Delta d_{t+i}^e - r_{t+i}^e] + \epsilon_t, \quad (1.7)$$

where $\rho_t = 1/(1 + \exp(-\tilde{\eta}_t))$. The state equation formalizes a dynamic pattern for the latent process $\tilde{\eta}_t$, which is consistent with the diagnosed persistence of the PtDR

$$\tilde{\eta}_t = \tilde{\eta}_{t-1} + u_t, \quad (1.8)$$

where $u_t \sim N(0, \sigma_u^2)$ and the initialization $\tilde{\eta}_0$ is treated as a model parameter.

As a particular alternative state equation we consider a stationary first order autoregressive state process, i.e.,

$$\tilde{\eta}_t = \delta + \alpha \tilde{\eta}_{t-1} + u_t, \quad (1.9)$$

where $|\alpha| < 1$, $u_t \sim N(0, \sigma_u^2)$ and $\tilde{\eta}_0 = \delta/(1 - \alpha)$.⁴

A novel feature of the state space model in (1.7) coupled with (1.8) or (1.9) is that $\tilde{\eta}_t$ can be estimated from filtered data. Compared with a framework of structural breaks, a continuously evolving steady state of the PtDR not only allows to test for its various determinants simultaneously, but might also enable a more successful out-of-sample forecasting of stock returns. In the framework of structural

⁴ We allow for many alternative specifications of the state process: a random walk with variance breaks, a non-stationary AR(2) model, a random walk with leptokurtic innovations or a moving average structure of the error term and an autoregressive process with parameter shifts. Irrespective of these alternatives the estimated state processes come very close to the one extracted from the pure random walk model in (1.8).

breaks, subsample means are used to approximate the distinct steady states. Then, timing and magnitude of the breaks have to be estimated, which may lead to a weakened real time performance of break adjusted forecasting schemes (Lettau and Van Nieuwerburgh; 2008). The state space model, however, provides a continuously evolving path of the long-term state. It can be easily extended into the future. In particular, it becomes unnecessary to locate break dates and magnitudes when using the PtDR to forecast stock returns out-of-sample. As it turns out, this distinctive characteristic is essential in improving the predictability of stock returns by means of the PtDR.

1.2.3 Model implementation

The state space model outlined in (1.7) and (1.8/1.9) is nonlinear. We use particle filtering (Cappé et al.; 2007) based on 3000 trajectories for an approximation of the models' log-likelihood, subsequent parameter and state estimation. Following Lettau and Van Nieuwerburgh (2008) we focus mainly on the annual CRSP stock market index and corresponding dividends starting in 1926 to obtain the PtDR. S&P500 data dating back until 1871 are also analysed to address the robustness of the results. A detailed description of data sources and particle filtering can be found in Appendices 1.A and 1.B, respectively.

To formulate the objective expectations about future excess dividend growth rates and excess returns in (1.7), we follow Campbell and Vuolteenaho (2004) and employ low dimensional vector autoregressions (VARs) of order one comprising the PtDR series (η_t), excess dividend growth rates (Δd_t^e), excess returns (r_t^e) and inflation π_t .⁵ The VAR based determination of ($\tilde{E}_t[\Delta d_{t+i}^e]$) and ($\tilde{E}_t[r_{t+i}^e]$) goes back to Campbell and Shiller (1988) and Campbell (1991).⁶ Including the PtDR in the VAR provides unobservable market information about the future dividends and returns. The reduced form VAR is informative and at the same time general enough to be consistent with a present value relation with a time-varying mean of the PtDR. The 3-month Treasury Bill rate is employed to approximate the risk free rate and the CPI to measure inflation.⁷

⁵ The inclusion of inflation accounts for the eventual effects of money illusion on equity prices.

⁶ In related contexts, VAR based predictions have also been used to approximate price expectations, for instance, by Sbordone (2002) and Rudd and Whelan (2006).

⁷ Note that we include a constant in the VAR model. One may argue that a VAR model with a deterministic trend might be more suitable to model a persistent PtDR. However, including a time trend results in an unbounded dynamics of the PtDR, which contradicts our observations (see Footnote 2). Also a deterministic trend is at odds with stationary stock returns and dividend growth. Estimates for the latent process ($\tilde{\eta}_t$) display similar dynamics if a time trend is included.

We take an adaptive approach to the choice of the VAR sample size such that it not only provides efficient parameter estimates under structural invariance of VAR dynamics, but also responds to structural changes in VAR dynamics. Specifically, we evaluate in each forecast origin a set of VAR models $\Omega_{t,\omega} = \{\eta_\tau, \Delta d_\tau^e, r_\tau^e, \pi_\tau, \tau = t - \omega + 1, \dots, t\}$ with alternative sample lengths ($\omega = 20, 21, \dots, 30$). From this set the particular VAR model with window size ω is employed to determine ex-ante predictions $\widehat{\Delta d}_{t+i}^e$ and \widehat{r}_{t+i}^e that minimizes the root mean squared errors for the 10 most recent in-sample observations $\{\Delta d_m^e - r_m^e\}_{m=t-9}^{m=t}$.⁸

The determination of the objective expectations in (1.7) requires an initialization period which is chosen to comprise 30 observations. Therefore, to evaluate the state space model in (1.7) from 1926 (a common starting period in the literature), we joined the CRSP data starting in 1926 with the S&P500 data before this period. Alternatively, we consider the S&P500 and estimate the state space model for the sample period 1901 to 2013.

1.2.4 Estimates and diagnostics

Turning to the evaluation of the state space model in (1.7) and (1.8), we find that the model with a time-varying mean of PtDR outperforms its constant mean counterpart. The estimated model parameters are documented in Table 1.2, which also includes estimation results for the constant state benchmark model of Campbell and Shiller (1988). Considering the random walk (RW) as the state process (1.8), the log-likelihood value conditioned on CRSP data for a time-varying state model is about 168.4 while the respective statistic for the constant state model is 23.9 (see the first row of Table 1.2). Similar log likelihood diagnostics favouring time variation in the mean of the PtDR are obtained for S&P500 data, and if the static model is contrasted against a model with $\tilde{\eta}_t$ specified as a stationary state process (1.9).

The lead of the more flexible model approach over the static benchmark present value model can be visualized by eyeballing the estimated patterns of $\tilde{\eta}_t$ provided

⁸ To assess the robustness of outcomes we consider a set of robustness tests (i) using fitted errors regarding excess returns $\{r_m^e\}_{m=t-9}^{m=t}$ instead of $\{\Delta d_m^e - r_m^e\}_{m=t-9}^{m=t}$; (ii) using the five most recent observations in $\{\Delta d_m^e - r_m^e\}_{m=t-4}^{m=t}$ to compute the RMSE; or (iii) using the mean absolute error criterion instead of RMSE. The corresponding results with regard to the evaluation of the state space model are quantitatively almost identical.

At the implementation side, non-stationarity of the PtDR may cause explosive paths in forecasts of excess returns and excess dividends growth rates. For the VAR models selected by our benchmark adaptive scheme this turns out to be the case in 1917 and periods 1993-1996 and 1998-2001. For these years we restrict the autoregressive coefficient of the PtDR in the VAR and reestimate the system by means of EGLS until all eigenvalues of the characteristic polynomial are smaller than unity in modulus.

State equation	Time-varying $\tilde{\eta}$						Constant $\tilde{\eta}$ ($\sigma_u = 0$)		
	δ	α	$\tilde{\eta}_0$	σ_u	σ_ϵ	log-lik	$\tilde{\eta}_0$	σ_ϵ	log-lik
CRSP 1926-2013									
RW	-	-	2.892	0.059	0.026	168.4	3.409	0.184	23.9
AR(1)	0.042	0.986	-	0.063	0.025	166.5	3.409	0.184	23.9
S&P500 1901-2013									
RW	-	-	3.074	0.076	0.033	196.1	3.298	0.215	13.3
AR(1)	0.046	0.985	-	0.075	0.031	195.8	3.298	0.215	13.3

Table 1.2: Parameter estimates and model evaluations In each panel the upper row documents statistics from the model with a random walk (RW) as state equation (1.8) and the lower row statistics from the model applying a stationary AR(1) process as the state equation (1.9). Out-of-sample VAR forecasts are applied to evaluate $\tilde{E}_t[\Delta d_{t+i}^e]$ and $\tilde{E}_t[r_{t+i}^e]$ in (1.7).

in Figure 1.3. The data speak against time invariance of the mean of the PtDR. In addition, estimating $\tilde{\eta}_t$ conditional on either a RW or a stationary AR(1) state equation provides very similar results. Both implied state processes can only be differentiated marginally in the early and later sample periods (see Figure 1.3). As it turns out both estimates of $\tilde{\eta}_t$ lead to qualitatively identical results for the remaining empirical analysis. We concentrate on the estimates from a RW state equation henceforth, since it is marginally in lead over a stationary AR(1) process (1.9) according to log likelihood diagnostics documented in Table 1.2.

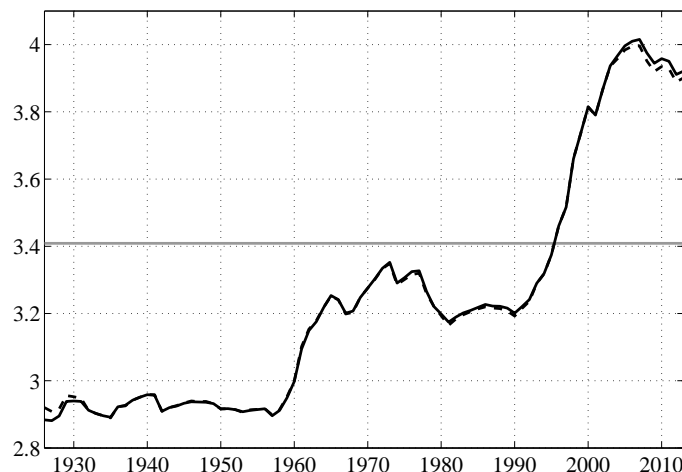


Figure 1.3: Time-varying mean of the PtDR The figure depicts the estimated time-varying mean of the PtDR $\tilde{\eta}_t$ (black solid line for the random walk and black dashed line for the stationary AR(1) state process) along with its time invariant counterpart (grey line).

Obtaining the continuous movements of the mean of the PtDR enables a regression based economic interpretation of its determinants that we provide in Section 1.4. Moreover, time variation in the mean of the PtDR is valuable for the ex-ante modeling of stock returns. In the next section, we analyse how $\tilde{\eta}_t$ exploits the informational content of the PtDR in so-called predictive regressions.

1.3 Forecasting performance

We discuss predictive regression models for stock returns conditional on CRSP data, which is also used in Lettau and Van Nieuwerburgh (2008). Results for S&P500 data are qualitatively identical. Adjusting the PtDR for its slowly evolving mean provides better out-of-sample forecasts in terms of the RMSE and the out-of-sample R^2 compared with centering the PtDR with discrete mean shifts or using the historical average of returns as the predictor. In the following we describe in-sample (IS) and out-of-sample (OOS) forecasting designs, and discuss in detail the forecasting performance of competing approaches.

1.3.1 Predictive regressions

The predictability of stock returns is evaluated by means of common predictive regressions of the following type (see, e.g., Lettau and Van Nieuwerburgh; 2008)

$$r_{t+1} = \beta_0 + \beta_1(\eta_t - s_t) + v_{t+1}, \quad (1.10)$$

where r_{t+1} denotes the total log-returns and v_{t+1} is an error term. We also use the predictive regressions to assess the predictability of dividend growth rates, substituting Δd_{t+1} for r_{t+1} in (1.10). To implement predictive regressions the PtDR (η_t) is adjusted by alternative state processes (s_t) such that ‘centered’ observations ($\eta_t - s_t$) are considered to predict stock returns.⁹ Under the null hypothesis of no predictability $\beta_1 = 0$, the corresponding equation serves as the benchmark model (see, e.g., Goyal and Welch; 2008). For IS analysis, the corresponding naive predictor is the

⁹ Taking a statistical perspective on eventually unbalanced predictive regressions, Nelson and Kim (1993) and Stambaugh (1999) invoked a discussion if in-sample ‘significant’ relations between transitory stock returns and lagged persistent valuation ratios are spurious. The persistence of predictors and the correlation of their innovations with those of the dependent variable might cause an upward bias of estimates and size distorted inference. In light of poorly balanced predictive regressions the econometric treatment of regressions involving stationary but nearly integrated predictors has been considered in various contributions (Elliott and Stock; 1994; Cavanagh, Elliott and Stock; 1995; Lewellen; 2004; Torous, Valkanov and Yan; 2004; Campbell and Yogo; 2006). We follow Lettau and Van Nieuwerburgh (2008) and adjust the PtDR for changes in its mean.

full sample mean return. For OOS analysis the naive predictor is the historical average return obtained up to the forecast origin.

In the *IS analysis* we compare forecasting specifications obtained by adjusting the PtDR by means of four alternative long-run states $s_t \in \{\bar{\eta}, \bar{\eta}_t^{(1)}, \bar{\eta}_t^{(2)}, \tilde{\eta}_t\}$. In the first specification the PtDR is centered by its (full sample) mean ($\bar{\eta}$). We refer to this setting as the ‘unadjusted’ PtDR since this model is equivalent to that of using the actual PtDR series in the predictive regressions. In the second and third specification, the PtDR is adjusted for one and two structural breaks ($\bar{\eta}_t^{(1)}$ and $\bar{\eta}_t^{(2)}$, respectively). Following Lettau and Van Nieuwerburgh (2008) the supremum F -test (Bai and Perron; 1998, 2003) is employed to determine the timing of the breaks. In case of one shift it is diagnosed to occur in 1992, and in case of two breaks the respective locations are 1955 and 1993.¹⁰ Lastly, we adjust the PtDR by means of the continuously time-varying mean $\tilde{\eta}_t$ which is extracted from the state space model in outlined in Section 1.2.

Initializing the *OOS analysis*, the first forecasting regressions use 20 years of data. Then the estimation windows are expanded recursively as in Lettau and Van Nieuwerburgh (2008). We consider three corresponding adjustments for the PtDR – $s_t \in \{\bar{\eta}, \bar{\eta}_t, \tilde{\eta}_t\}$ – all of which are recursively estimated from the respective estimation samples. In the benchmark setting, the PtDR is centered with its mean $\bar{\eta}$ from the estimation period. The second adjustment $s_t = \bar{\eta}_t$ corresponds to the case of discrete mean shifts. We apply supremum F -tests and rely on the 10% significance level to determine the mean shift processes $\bar{\eta}_t$. Lastly, the PtDR is adjusted by $\tilde{\eta}_t$ conditioning only the information from the estimation periods.

The four alternative long-run states of the PtDR entering the IS analysis, $s_t \in \{\bar{\eta}, \bar{\eta}_t^{(1)}, \bar{\eta}_t^{(2)}, \tilde{\eta}_t\}$, are displayed in Figure 1.4 over the maximum sample period (1926 to 2013). The smoothly evolving mean $\tilde{\eta}_t$ seems to be mostly close to the mean with two structural breaks $\bar{\eta}_t^{(2)}$. However, the former lags behind the latter after the diagnosed break dates (1955 and 1993). This reveals the nature of the particle filtering applied to the nonlinear state space model. Although the parameters of the state space model are estimated conditioning on the full sample information for the IS analysis, the estimated latent process is mainly based on past information. Using

¹⁰These break points are close to those diagnosed in Lettau and Van Nieuwerburgh (2008) who analyse a slightly distinct sample period (1926 to 2004). The null hypothesis of no break is rejected with 1% significance against one or two breaks ($supF(1|0) = 18.12$ and $supF(2|0) = 23.90$). The null hypothesis of one break is rejected against the alternative of two breaks ($supF(2|1) = 9.56$) with 10% significance. We consider both one and two breaks to compare prediction outcomes with those in Lettau and Van Nieuwerburgh (2008). The applied test procedure is robust to serial correlation and heteroscedasticity, the trimming is 5% of the sample.

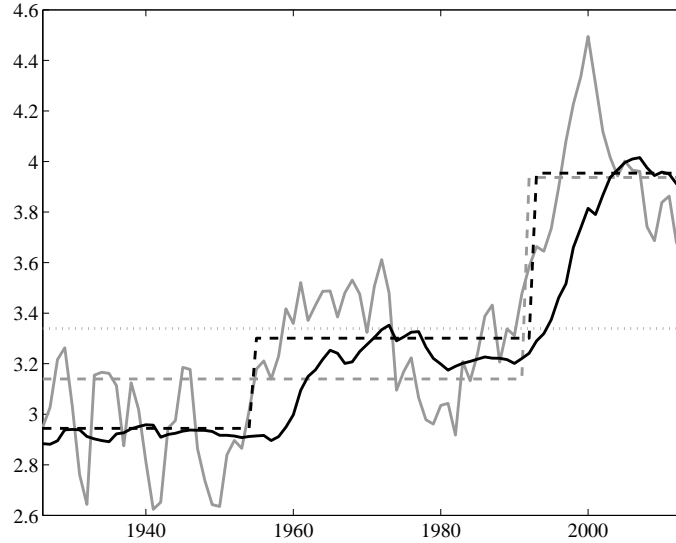


Figure 1.4: The PtDR and distinct mean evaluations The figure presents the price-to-dividend ratio η_t (grey solid line), the total sample average PtDR $\bar{\eta}$ (grey dotted line), the mean of the PtDR with one break $\bar{\eta}_t^{(1)}$ (grey dashed line), the mean of the PtDR with two breaks $\bar{\eta}_t^{(2)}$ (black dashed line), and the continuously time-varying steady state of the PtDR $\tilde{\eta}_t$ (black solid line).

the random walk state equation (1.8) as an example, each particle is equal to $\tilde{\eta}_{t-1}^{(i)}$ plus a draw from the error term with variance σ_u^2 . Thus, being a (weighted) average of particles, $\tilde{\eta}_t$ is mainly determined from the past information. This contributes to the slowly evolving nature of the estimated time-varying mean, which does not show much advantage for the IS analysis, but could be crucial for the predictive power of the PtDR in the OOS analysis. The core obstacle in using discrete break adjustments in OOS forecasting is to determine timing and magnitude of the breaks. The slowly evolving time-varying mean $\tilde{\eta}_t$ overcomes these difficulties. When there are no marked structural changes, it evolves around a relatively stable level. In response to persistent movements, it adapts and incorporates the new information gradually. Specifically, to obtain an update for $\tilde{\eta}_t$ by means of weighted averaging, particles $\tilde{\eta}_{t-1}^{(i)}$ are ranked according to the fit of the corresponding measurement equation for period $t - 1$. Particle $\tilde{\eta}_{t-1}^{(j)}$ enters $\tilde{\eta}_t$ with higher weight than particle $\tilde{\eta}_{t-1}^{(k)}$ when the error term in the measurement equation for the former is smaller than the one for the latter. Along the updating steps the fittest particles survive. Readers may consult Appendix 1.B for more details.

1.3.2 Forecast evaluation

Predicting stock returns in-sample, the unadjusted PtDR provides a small R^2 of about 0.0392 (column 2 of the first panel in Table 1.3). Adjusting the PtDR for

shifts improves the explanatory content of predictive regressions markedly. The R^2 statistics increase to 0.1027 and 0.1751 for means with one and two shifts, respectively (column 3 and 4). The magnitude and the statistical significance of the estimated predictive coefficient (β_1) increase as well. This evidence confirms findings in Lettau and Van Nieuwerburgh (2008). As expected, with an in-sample degree of explanation of about 0.0641 (column 5), adjusting the PtDR by a slowly evolving mean does not outperform adjustments for discrete shifts in the mean. As an adaptive filtering process, $\tilde{\eta}_t$ mainly depends on past information even in the in-sample setting. In contrast, the break adjustments take into account the full sample information and ex-post minimize squared approximation errors for the actual PtDR.

	$\bar{\eta}$	$\bar{\eta}_t^{(1)}$	$\bar{\eta}_t^{(2)}$	$\tilde{\eta}_t$
Stock returns				
β_0	0.0928 (4.6562)	0.0937 (4.5205)	0.0939 (6.0840)	0.1142 (5.3922)
β_1	-0.0923 (-2.5610)	-0.2516 (-4.8718)	-0.4120 (-6.0994)	-0.2089 (-2.7885)
R^2	0.0392	0.1027	0.1751	0.0641
adj R^2	0.0279	0.0922	0.1654	0.0531
Dividend growth				
β_0	0.0456 (3.3098)	0.0457 (3.5017)	0.0458 (3.3140)	0.0539 (4.0314)
β_1	-0.0013 (-0.0426)	-0.0431 (-0.9694)	-0.1157 (-1.7106)	-0.0832 (-1.8706)
R^2	0.0000	0.0057	0.0260	0.0192
adj R^2	-0.0117	-0.0060	0.0146	0.0076

Table 1.3: In-sample predictive regressions This table documents statistics from regressions (1.10) with four alternative mean processes of the PtDR using CRSP data from 1926 to 2013. In the second column forecasts for returns (dividend growth) are conditioned on the unadjusted PtDR (using the overall sample mean $\bar{\eta}$). The third and fourth column contain the estimates based on the adjusted PtDR using one ($\bar{\eta}_t^{(1)}$) or two ($\bar{\eta}_t^{(2)}$) mean shifts, respectively. In the last column forecasts are conditioned on the PtDR adjusted by means of the smooth state process $\tilde{\eta}_t$. Newey and West (1987) robust t -statistics for coefficient estimates are presented in parantheses. The bandwidth is selected by means of the procedure proposed by Newey and West (1994).

Forecasting dividend growth, the unadjusted PtDR does not have much predictive power. The predictive coefficient is not statistically significant at the 5% level and the R^2 is negligible (column 2 of lower panel). Neither break adjustments nor

smoothly varying mean adjustments improve the performance in a sizeable manner (column 3-5 of the lower panel). This evidence is in line with results from Lettau and Van Nieuwerburgh (2008) for a similar sample period (1927 to 2004).

	\bar{r}_t	$\bar{\eta}$	$\bar{\bar{\eta}}_t$	$\tilde{\eta}_t$
1946-2013				
RMSE	0.1694	0.1751	0.1793	0.1683
R_{oos}^2		-0.0691	-0.1202	0.0121***
1946-2004				
RMSE	0.1607	0.1681	0.1714	0.1589
R_{oos}^2		-0.0934	-0.1367	0.0232***

Table 1.4: Out-of-sample predictive regressions for stock returns This table documents the OOS forecasting performance of the naive prediction by means of historical average returns (\bar{r}_t) and alternative predictive regressions with the unadjusted PtDR ($s_t = \bar{\eta}$), the PtDR adjusted by mean shifts ($s_t = \bar{\bar{\eta}}_t$), and the PtDR adjusted by the smooth state ($s_t = \tilde{\eta}_t$). Root mean squared errors (RMSE) and OOS R^2 statistics (R_{oos}^2) are shown. R_{oos}^2 is constructed against the naive forecasting scheme (\bar{r}_t). Statistical significance levels of R_{oos}^2 at the 1%, 5%, 10% level denoted by ***, **, * are based on the MSE-adjusted statistic proposed by Clark and West (2006). CRSP data are considered.

Adjusting the PtDR for discrete mean shifts in real time (OOS forecasting) fails to improve upon using historical average returns as benchmark predictors. In contrast, centering the PtDR around the smoothly varying mean obtains the smallest RMSE statistic among all predictors (last column in Table 1.4). Considering the full sample period from 1946 to 2013 (first panel in Table 1.4), the naive benchmark, the unadjusted PtDR, centering with discrete mean shifts and centering around $\tilde{\eta}_t$ result in RMSE statistics of 0.1694, 0.1751, 0.1793 and 0.1683, respectively. The same ranking of the RMSEs holds if the sample period ends at 2004, as considered by Lettau and Van Nieuwerburgh (2008) (lower panel in Table 1.4). Results from mean absolute errors (MAEs) are qualitatively identical (not shown).

To evaluate the statistical significance of the forecasting performance of alternative predictors compared with the benchmark model using historical average returns, we consider an OOS degree of explanation (Goyal and Welch; 2008)

$$R_{oos}^2 = 1 - \frac{MSE_s}{MSE_{\bar{r}}}, \quad (1.11)$$

where $MSE_{\bar{r}}$ denotes the mean squared forecast error from naive forecasts and MSE_s is the corresponding statistic from alternative models (1.10) with $s_t \in$

$\{\bar{\eta}, \bar{\eta}_t, \tilde{\eta}_t\}$. Under the hypothesis of less (more) accurate forecasts from alternative model specifications compared with naive predictions, the MSE of the benchmark model is smaller (larger) than that of the alternative model, which corresponds to $R_{oos}^2 < 0$ ($R_{oos}^2 > 0$). Following Rapach, Strauss and Zhou (2010) the significance of R_{oos}^2 is evaluated by means of the MSE-adjusted statistic in Clark and West (2006).¹¹

We find that in real time adjusting the PtDR by $\tilde{\eta}_t$ outperforms the historical average return in forecasting stock returns significantly. As can be seen from the second and fourth row of Table 1.4, only adjusting the PtDR for the smoothly varying mean ($\tilde{\eta}_t$) provides positive and significant R_{oos}^2 statistics. We find the same evidence for the case of S&P500 data, forecasting excess returns instead of returns, or using different specifications to determine the window size of the VAR forecasting scheme for the state space model as already pointed out in Footnote 8. The corresponding results are available upon request.

Following Goyal and Welch (2008), we provide further insights into the OOS forecasting performance over time and depict the difference of the cumulative squared forecasting errors of naive forecasts minus those of the alternative models in Figure 1.5.¹² We find that using the continuously varying mean adjustment improves the forecasting strength of the PtDR throughout the entire sample period compared with the unadjusted PtDR or using break adjustments. The performance curve of adjusting the PtDR by its varying mean falls least during periods with structural changes, and has the longest positive trend during the relatively tranquil periods. The two ex-post identified structural changes occur in 1955 and 1993. In a real time forecasting situation, all three predictors ($s_t \in \{\bar{\eta}, \bar{\eta}_t, \tilde{\eta}_t\}$) start to underperform in comparison with the naive forecasts around 1957 and embark a negative trend. However, adjusting the PtDR by $\tilde{\eta}_t$ the respective performance curve falls least

¹¹The MSE-adjusted statistic is based on

$$f_{\tau+1} = (r_{\tau+1} - \bar{r}_\tau)^2 - ((r_{\tau+1} - \hat{r}_{\tau+1})^2 - (\bar{r}_\tau - \hat{r}_{\tau+1})^2), \tau = \tau_0, \tau_0 + 1, \dots, T - 1, \quad (1.12)$$

where \bar{r}_τ denotes forecasts from the benchmark model with historical average returns, and $\hat{r}_{\tau+1}$ those from alternative models. Predicting dividend growth rates we apply the equivalent formula replacing $r_{\tau+1}$, \bar{r}_τ and $\hat{r}_{\tau+1}$ by the corresponding counterparts. For a regression of $f_{\tau+1}$ on a constant, Clark and West (2006) show that the corresponding t -statistic is asymptotically normally distributed even in the case of nested models. Under the alternative hypothesis the mean of $f_{\tau+1}$ (the constant coefficient) is greater than zero. It is worthwhile to notice that the alternative state processes are re-estimated at each period given available sample information. Thus, the varying mean processes of the PtDR change with each forecast origin.

¹²One can look at the performance for any OOS periods by redrawing a horizontal line at the start of OOS periods. If the curve terminates at a higher (lower) point at the end of OOS periods, the alternative model has a lower (higher) RMSE over the OOS periods of interest.

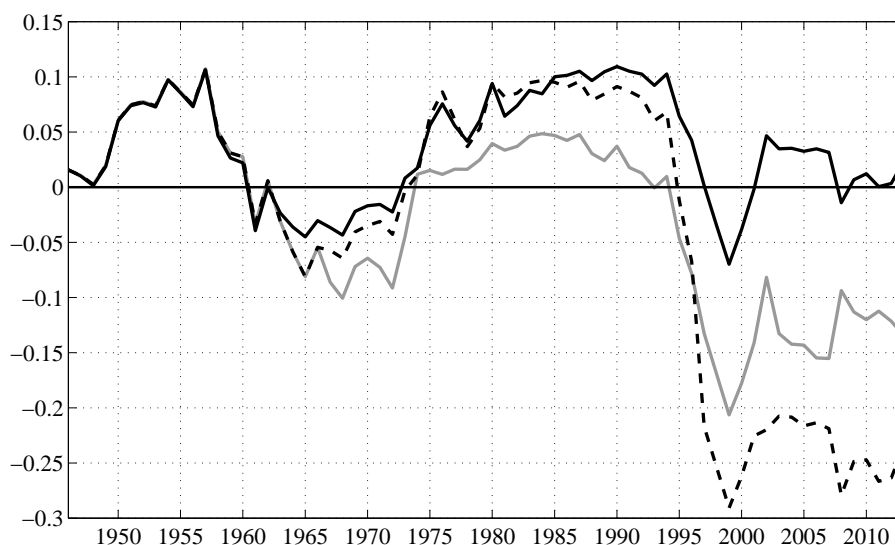


Figure 1.5: Out-of-sample forecasting performance relative to the historical mean This figure depicts the difference of the cumulative squared forecast errors of the naive prediction by means of historical average returns (\bar{r}_t) minus those of the three alternative predictive regressions. The considered predictors are the unadjusted PtDR (grey solid line with $s_t = \bar{\eta}$), the PtDR adjusted by allowing for potential mean shifts (black dashed line with $s_t = \bar{\eta}_t$), and the PtDR adjusted by means of the smooth state (black solid line with $s_t = \tilde{\eta}_t$). The first forecasting period is 1946.

(black solid line). The performance curve of the unadjusted PtDR (grey solid line) falls a bit more than the one from the break adjusted PtDR (black dashed line). The former reaches its bottom around 1968 and starts to follow the positive trend that is led by the performance lines of the two adjusted predictors. All three predictors outperform the naive forecast in periods with oil price shocks in 1973/1974. The good performance of the unadjusted PtDR during this period has also been noted by Goyal and Welch (2008). Centering the PtDR around its slowly evolving mean, however, obtains the only predictor that sustains this positive trend until 1994. The performance curves of both the unadjusted PtDR and the break adjusted PtDR reach their peaks in the early 1980s and start to fall since then. From 1994, the performance of all three predictors drops dramatically with adjusting the PtDR for its slowly varying mean dropping the least. The strongest performance deterioration (and weakest recovery since 1999) is observed when centering the PtDR around discrete shifts in mean.

In summary, reflecting its adaptive potential in both turmoil and tranquil periods adjusting the PtDR by the slowly evolving mean offers superior ex-ante signaling in particular since the beginning of the new millennium.

Taking the break adjusted PtDR as benchmark predictor, adjusting the PtDR by its varying mean has a better performance than the naive forecasts based on

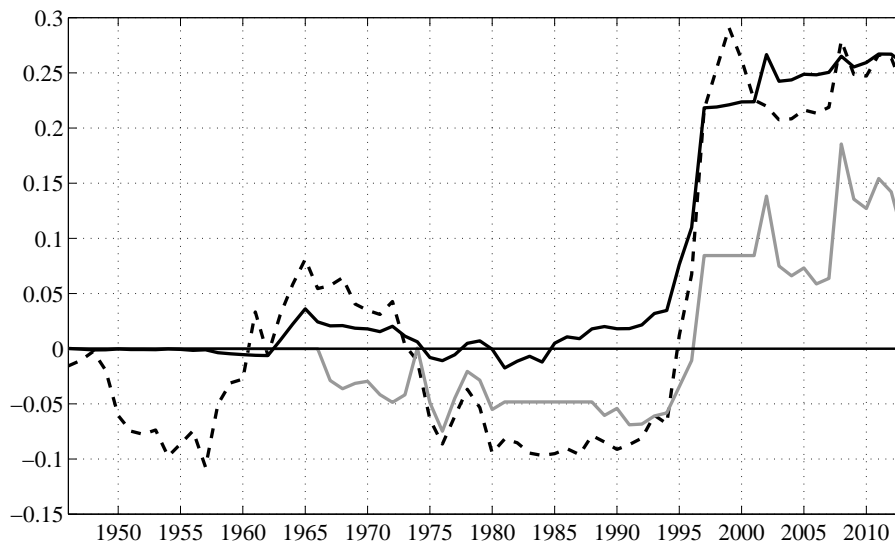


Figure 1.6: Out-of-sample forecasting performance relative to the break adjusted model of the PtDR This figure depicts the difference of the cumulative squared forecast errors of the prediction by means of the break adjusted PtDR ($s_t = \bar{\eta}_t$) minus those of the three alternative predictive regressions. The considered predictors are the historical mean (black dashed line with \bar{r}_t), the unadjusted PtDR (grey solid line with $s_t = \bar{\eta}$) and the PtDR adjusted by means of the smooth state (black solid line with $s_t = \tilde{\eta}_t$). The first forecasting period is 1946.

historical mean returns. This is visualized in Figure 1.6, which depicts OOS forecasting performance of various models through time using the PtDR adjusted for discrete shifts in mean as the benchmark. Although the break adjustment is not the best performing one (as discussed above), by using it as a benchmark one can easily compare the performance of historical mean forecasts with the one of the continuously varying mean adjusted PtDR. As can be seen from Figure 1.6, adjusting the PtDR by its varying mean (black solid line) never had the dramatic falls as those from the historical mean (black dashed line). In the turbulent periods of the 1950s and the 1970s, the performance curve of the historical mean falls dramatically, since it failed to reflect the underlying changes in the economy. In early 1990s, the varying-mean adjusted PtDR adapts to structural changes quickly and outperforms the break adjustment while the historical mean lags behind.

In summary, we conclude that centering the PtDR around its slowly evolving mean provides a superior OOS forecasting performance compared with adjusting the PtDR for discrete shifts, and is also preferable to using historical means as a predictor.

1.4 Long-run determinants of the PtDR

Extracting a continuously varying mean of the PtDR provides a unique opportunity to exam its (joint) potential influences, and offers an economic interpretation for the out-of-sample predictive power of adjusting the PtDR by the time-varying mean in predictive regressions. We investigate three important factors that have been documented to affect the PtDR in a long-run manner – consumption risk, the demographic structure of the population, and the dividend payout policy of firms. We find that all three factors jointly shape the slowly evolving mean of the PtDR, and diagnose consumption risk to be the most important influence. In the following we discuss the considered factors and provide evidence from a cointegration analysis to assess their explanatory content. A detailed description of the variables is given in Appendix 1.A.

1.4.1 The three long-term determinants

Consumption risk The influence of macroeconomic uncertainty on asset prices and equity premia has been long recognized in the asset pricing literature.¹³ More recent studies such as those in Bansal and Yaron (2004) and Lettau et al. (2008) use recursive Epstein and Zin (1989) preferences, and demonstrate that economic agents dislike economic uncertainty, and a rise in consumption volatility can raise the expected return and lower asset prices. Empirically, Lettau et al. (2008) show low frequency evidence while Bansal et al. (2005) provide higher frequency evidence on the contribution of lower consumption volatilities to higher asset prices particularly since the 1990s. Bansal et al. (2005) show that consumption volatility measures have good in-sample predictive power for the one-step ahead quarterly PtDR if historical volatilities are extracted from short time windows of one or two years of consumption data. Lettau et al. (2008) argue in favour of a regime change in consumption risk to explain a regime change in asset valuations. The estimated regime is very persistent. The lower volatility regime reached in the early 1990s is expected to last for 30 years. In this paper, we adopt the consumption risk measure used by Bansal et al. (2005) in a low frequency manner, in order to explain the continuously varying mean (low frequency movements) of the PtDR.

The consumption volatility is measured as $cr_t^W = \ln \left(\sum_{i=0}^{W-1} |co_{t-i}| \right)$, where co_t denotes the centered annual growth rate of per capita consumption and W is the size

¹³See, among others, Gennotte and Marsh (1992), Giovannini (1989) and Kandel and Stambaugh (1990).

of rolling time windows.¹⁴ We employ data on the per capita personal consumption expenditures on non-durable goods and services of the Bureau of Economic Analysis starting in 1929. To initialize time series of consumption risk we combine this series with the historical data on real per capita consumption recently collected by Barro and Ursua (2008).

To measure macroeconomic risk at low frequency one has to select W such that respective time windows carry informational content beyond short-run cycles. Figure 1.7 displays the absolute consumption growth with its Hodrick Prescott trend (the smoothing parameter is $\lambda = 100$). This trend visualizes the cyclical pattern of the consumption volatility. Counting from trough to trough, the length of the cycles are 30 (1870 to 1890), 18 (1890 to 1908), 22 (1908 to 1930), 44 (1930 to 1974), and 27 years (1974 to 2001). The 44 year cycle seems to be the odd one out, which can be regarded to contain two adjacent cycles – a 17 year cycle (1930 to 1947) around the WWI era and a 27 year cycle (1947 to 1974) around the WWII era. The durations of remaining cycles range from about 20 to 30 years. This forms our focus on alternative window lengths $W = 20, 21, \dots, 30$ to calculate time local long-run consumption volatility cr_t^W . A lower bound of 20 years is consistent with the so-called Kuznets swings in economic growth (e.g., Solomou; 2008). In addition, as argued by Geanakoplos et al. (2004) it is reasonable to assume that agents consider a 20 year horizon to incorporate demographic trends in long-term asset price expectations. A higher bound of 30 years coincides with the estimated average duration of a regime of consumption volatility in Lettau et al. (2008).

The upper right panel of Figure 1.8 depicts cr_t^W with $W = 20, 25, 30$ as examples. The shapes of all three consumption measures are similar, and become a bit smoother as the window size W increases. It appears that macroeconomic uncertainty has continuously decreased since the 1940s until the 1960s. It remained relatively stable during the 1970s and 1980s and decreases further since the 1990s. Comparing cr_t^W with $\tilde{\eta}_t$ depicted in the upper left panel of Figure 1.8, it seems that consumption risk is negatively related to movements in the varying mean of the PtDR throughout the entire sample for all considered window lengths W .

Demographics By means of an overlapping generation model, Geanakoplos et al. (2004) provide the foundation for a long-run positive relationship between the PtDR

¹⁴To calculate the volatility we use consumption growth directly instead of respective AR(1) regression residuals as considered in Bansal et al. (2005), since we do not detect any significant pattern of serial correlation in annual quotes of co_t . The p -values of respective Ljung-Box statistics including 5 and 10 lags are 0.349 or 0.476, respectively.

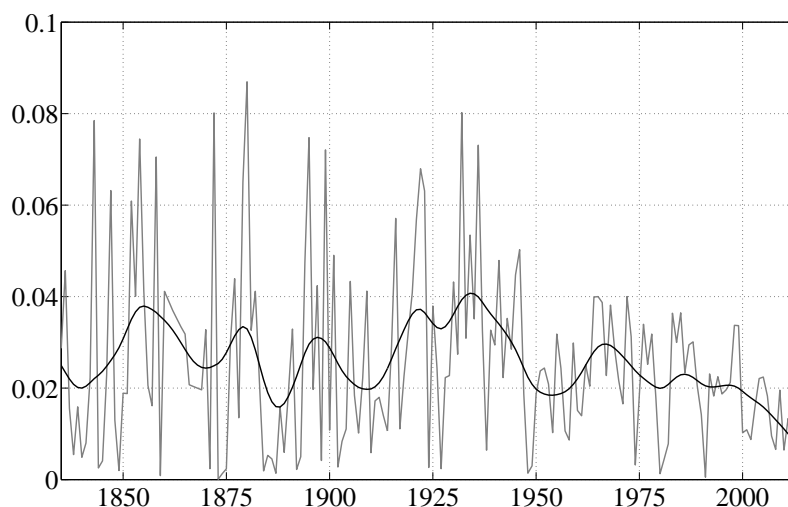


Figure 1.7: Low frequency consumption volatility This figure plots the absolute growth rate of per capita consumption with its HP trend (smoothing parameter 100).

and demographic trends. They argue that agents' incentives for holding equity vary over the life cycle. While the younger population intends to consume and willingly borrows for this purpose, the middle aged population concentrates more on saving and consumes these savings after retirement. The overall shape of the population pyramid is measured by means of the so-called middle-aged to young ratio (my_t). Geanakoplos et al. (2004) show that when my_t is large, there is excess demand for saving and equilibrium asset prices should increase to encourage consumption and to clear the market. This is consistent with price increases in the US stock market during the 1990s. Favero et al. (2011) demonstrate empirically the joint significance of my_t and the PtDR by means of long-horizon predictive regressions for stock returns, and diagnose a cointegration relationship between log dividends, log prices and my_t . These findings support the view that a slowly evolving mean of the PtDR could be driven by my_t .

Empirically my_t is defined as the ratio of the population aged 40-49 to the 20-29 year old, which is depicted in the lower left panel of Figure 1.8. Data is obtained from the US Census Bureau. The middle-aged to young ratio shows a marked U-turn since the 1960s. This is mainly under the influence of the baby boom after WWII. Beginning with the 1960s the baby boom generation entered into the statistic of the young population, reducing thereby my_t . For the same reason, the ratio has been increasing since the 1980s when the baby boom generation became middle-aged. The twin peaks around 1960 and 2000 in my_t are related with the two major increases in the PtDR. The increases in my_t in the 1950s and the 1980s are corresponding to the increases in $\tilde{\eta}_t$ in the 1960s and the 1990s, respectively.

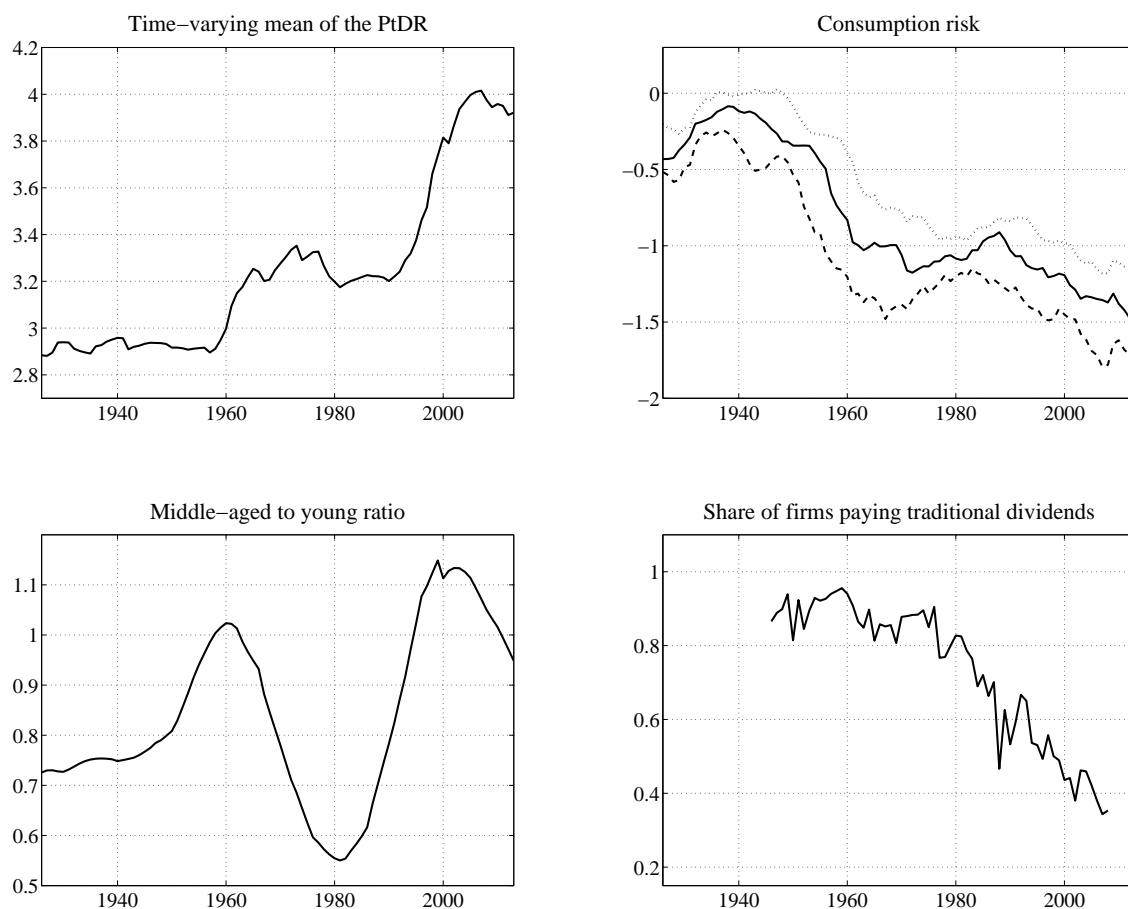


Figure 1.8: Economic influences and the time-varying mean of the PtDR The upper panels display the time-varying mean process of the PtDR and three distinct consumption risk measures (dashed line from $W = 20$, solid line from $W = 25$, dotted line from $W = 30$). The lower panels display the middle-aged to young ratio and the proportion of Type I firms.

Dividend payout policy Dynamics of the PtDR can also be influenced by changes in the dividend payout policy by firms (see Fama and French; 2001; Robertson and Wright; 2006; Boudoukh et al.; 2007). The proportion of firms paying cash dividends fell from 66.5% in 1978 to 20.8% in 1999 (Allen and Michaely; 2003). Lowered dividends may result in persistent increases of the PtDR. Kim and Park (2013) show that the changing payout policy affects the long-run relationship between stock prices and dividends: Both the proportion of firms that pay out a significant fraction of their earnings in the form of dividends and the cointegration coefficient between stock prices and dividends have followed a decreasing trend since the 1950s.

Decreasing cointegration coefficients linking log prices and log dividends are consistent with a non-stationary PtDR, which would be stationary if the cointegration coefficient were unity. If the proportion of firms with traditional payout policy re-

sults in the time-varying cointegration coefficient between prices and dividends, it should also influence the time-varying mean of the PtDR.

The firms with traditional payout policy are abbreviated as Type I firms following Kim and Park (2013). The lower right panel of Figure 1.8 depicts tp_t , the proportion of Type I firms among all firms in the (CRSP) value-weighted market portfolio. Data on tp_t for the sample period from 1946 to 2008 have been kindly provided by C.J. Kim and C. Park. The proportion of firms with traditional payout policy decreases from 87% in 1946 to 35% in 2008. The falls in the proportion of Type I firms from 1980 are consistent with the increasing trend of the $\tilde{\eta}_t$ particularly since 1990s.

1.4.2 Cointegration analysis

Unit root diagnosis First, we consider the individual characteristics of each variable by means of unit root tests. Unit root diagnostics for levels and first differences of $\tilde{\eta}_t$, cr_t^W , my_t and tp_t are documented in Table 1.5. Almost all tests indicate first order integration of $\tilde{\eta}_t$, cr_t^W and tp_t at conventional significance levels. The test regressions include a constant and a deterministic trend and refer to the periods from 1926 to 2013 in case of $\tilde{\eta}_t$, my_t and cr_t^W and from 1946 to 2008 for tp_t .¹⁵ The unit root hypothesis is rejected for all cr_t^W measures with $W = 20, \dots, 30$. Results for cr_t^W with $W = 25$ are shown in Table 1.5 as an example. Although unit root tests hint at stationarity of my_t , these results are to be taken with caution. Eyeballing my_t is hardly supportive to regard the process as mean stationary. The null hypothesis of stationarity is rejected with 10% significance by means of the KPSS test. When longer ranges of data are considered, evidence on unit roots governing my_t can be found. We follow Favero et al. (2011) and treat my_t as a first order integrated process.

Error correction model To test for a cointegration relation among all four variables and to estimate the cointegration coefficients, we employ the conditional single equation error correction model (SECM).¹⁶ With given presample values the

¹⁵ Test results are qualitatively identical if only a constant is included or the sample is reduced to the period from 1946 to 2008 for all variables.

¹⁶ As a preliminary analysis of cointegration relations, we look at the possibility of bivariate cointegration relations between the time-varying mean of the PtDR $\tilde{\eta}_t$ and each of the three long-run determinants – consumption risk, demographics and the proportion of firms with traditional dividend payout policy. We do not find evidence in support of any of the three bivariate long-run relations (not shown). This hints at the importance of taking into account all three different influences on the PtDR jointly.

	ADF	PP	DF _{GLS}	KPSS
$\tilde{\eta}_t$	-1.817	-1.837	-1.602	0.168**
$\Delta\tilde{\eta}_t$	-5.023***	-5.023***	-4.842***	0.178
cr_t^W	-2.458	-2.448	-1.542	0.127*
Δcr_t^W	-5.720***	-5.725***	-2.691***	0.153
my_t	-5.007***	-40.829***	-5.011***	0.122*
Δmy_t	-1.547	-1.690	-1.425	0.08
$\Delta^2 my_t$	-13.062***	-13.042***	-12.928***	0.097
tp_t	-2.184	-2.986	-1.438	0.262***
Δtp_t	-13.966***	-13.966***	-12.236***	0.500**

Table 1.5: Unit root tests for economic determinants This table displays results of unit root tests for the time-varying mean of the PtDR and its potential triggers. Test regressions include a constant and deterministic trend. The sample ranges from 1926 to 2013 and 1946 to 2008 in case of tp_t . Significance at 1%, 5%, 10% level is denoted by ***, **, *, respectively. For further notes see Table 1.1.

SECM reads as

$$\begin{aligned} \Delta\tilde{\eta}_t = & \delta_0 + \alpha (\tilde{\eta}_{t-1} + \beta_1 cr_{t-1}^W + \beta_2 my_{t-1} + \beta_3 tp_{t-1}) + \delta_1 \Delta cr_t^W \\ & + \delta_2 \Delta my_t + \delta_3 \Delta tp_t + \sum_{i=1}^2 \phi_i \Delta\tilde{\eta}_{t-i} + e_t, \quad t = 1, 2, \dots, T. \end{aligned} \quad (1.13)$$

The SECM specifies error correction dynamics conditional on current adjustments of weakly exogenous variables. It allows efficient inference by means of simple (non-linear) least squares estimation (see also Kremers, Ericsson and Dolado; 1992). As a particular merit, it offers a parsimonious representation that does not suffer from weakened estimator precision in comparison with full dimensional maximum likelihood estimation of a vector error correction model (Boswijk; 1995; Johansen; 1992). Model parsimony is beneficial in the present case of limited sample information. The estimation period starts in 1946 and ends in 2008 due to the nonavailability of the dividend payout ratio for earlier and later periods. To improve upon estimation uncertainty further, we apply a sequential estimation procedure eliminating in each step the short-term coefficients δ_i , $i = 1, 2, 3$, and ϕ_i , $i = 1, 2$, with the lowest t -statistic and lacking 30% significance.¹⁷ Adopting a general-to-specific model

¹⁷ Using a liberal significance level for the removal of single variables from the model we believe that joint insignificance of the removed variables is likely for common (more conservative) significance levels, 5% say.

composition, we start the model reduction from the SECM including two lags of the dependent variable which are necessary to capture patterns of serial correlations when testing for cointegration or weak exogeneity of variables below. We find that the consumption risk, the demographic factor and the proportion of Type I firms are weakly exogenous and unaffected by their cointegration relation with $\tilde{\eta}_t$ for $W \in 21, 23, \dots, 30$. This supports the use of SECM in (1.13).¹⁸

W from cr_t^W	20	21	22	23	24	25	26	27	28	29	30
α	-0.07 (-2.57)	-0.09 (-3.28)	-0.09 (-3.10)	-0.12 (-3.91)	-0.11 (-3.55)	-0.11 (-3.72)	-0.12 (-3.64)	-0.12 (-3.70)	-0.12 (-3.62)	-0.11 (-3.49)	-0.11 (-3.33)
β_1	0.50 (2.35)	0.42 (2.50)	0.58 (3.29)	0.53 (4.03)	0.62 (4.20)	0.64 (4.56)	0.64 (4.51)	0.65 (4.64)	0.66 (4.49)	0.68 (4.30)	0.67 (3.96)
β_2	-0.92 (-1.95)	-1.09 (-2.70)	-0.90 (-2.47)	-1.01 (-3.49)	-0.92 (-3.12)	-0.93 (-3.42)	-0.98 (-3.51)	-1.02 (-3.69)	-1.09 (-3.74)	-1.02 (-3.33)	-1.07 (-3.26)
β_3	0.86 (2.37)	0.85 (2.99)	0.75 (2.51)	0.74 (3.26)	0.67 (2.61)	0.61 (2.49)	0.56 (2.20)	0.49 (1.90)	0.42 (1.53)	0.46 (1.52)	0.43 (1.30)
R^2	0.45	0.51	0.48	0.53	0.51	0.52	0.51	0.51	0.51	0.51	0.50
adj R^2	0.39	0.45	0.43	0.47	0.45	0.46	0.46	0.46	0.45	0.45	0.44

Table 1.6: Cointegration analysis This table documents the estimates for the error correction model in (1.13) including consumption risk measures for distinct window sizes W . The t -statistics appear in parenthesis below the corresponding estimates. Based on critical values from surface regressions provided by Ericsson and MacKinnon (2002), adjustment coefficients (α) that are significant at 10% level are highlighted. Also significant cointegration coefficients ($\beta_1, \beta_2, \beta_3$) at the 10% level are highlighted.

We find evidence for a cointegration relation between $\tilde{\eta}_t$, cr_t^W , my_t and tp_t , where $W \in 23, \dots, 29$.¹⁹ For a significant cointegration relation, the absolute value of the t -statistic of the adjustment coefficient $\hat{\alpha}$ has to be larger than a respective non-standard critical value. For the specification in (1.13) with $W = 25$ as an example, the t -statistic of the adjustment coefficient is -3.7217 while the 10% critical value is -3.4509. As can be seen from Table 1.6, $\hat{\alpha}$ estimates for $W \in 23, \dots, 29$ are significant at 10% level. Estimating consumption risk from time windows of lengths $W = 20, 21, 22, 30$, we obtain similar degrees of explanation (see R^2 and adjusted R^2) and estimates for the cointegration coefficients, although a significant

¹⁸We test for weak exogeneity by adopting autoregressive models of order two augmented by the long-run relation between $\tilde{\eta}_{t-1}$, cr_{t-1}^W , my_{t-1} and tp_{t-1} as specified in (1.13), and looking at the significance of the corresponding adjustment coefficients. All respective adjustment coefficients are insignificant at the 10% level (not shown). Using $W = 25$ as an example, the respective p -values regarding the first differences of cr_t^W , my_t and tp_t are 0.315, 0.233 and 0.488.

¹⁹Johansen trace tests also confirm the results.

cointegration relation cannot be diagnosed within the SECM.

Significant effects from all three factors – consumption risk, the demographic structure and the dividend payout ratio – on the mean of the PtDR can be confirmed. Focusing on $W \in 23, \dots, 29$, all estimated cointegration parameters are significant (an exception is $\hat{\beta}_3$ for $W=28, 29$) and have the expected sign (see Table 1.6).²⁰ Both consumption risk (cr_t^W) and the proportion of Type I firms (tp_t) have a negative influence on $\tilde{\eta}_t$ and, thus, the signs of $\hat{\beta}_1$ and $\hat{\beta}_3$ shall be positive. For the demographic factor (my_t), it is the opposite case and the sign for $\hat{\beta}_2$ shall be negative.²¹ In addition, the variations in the estimates for the coefficients attached to cr_t^W and my_t are small – ranging from 0.53 to 0.68 and from -1.09 to -0.92 respectively. Estimates for the coefficient of tp_t exhibit some larger variation and range between 0.42 and 0.74.

Cross validations To gauge the relative importance of each long-run determinant in a systematic way within the SECM approach, we employ so-called cross-validation (CV) criteria (e.g., Picard and Cook; 1984). While augmenting (reducing) the set of explanatory variables in a regressions trivially goes along with gains (losses) in terms of in-sample model fit, CV criteria exhibit a nontrivial relation between a model’s dimensionality and predictive content. The CV statistic is calculated as the mean absolute forecast error for $\Delta\tilde{\eta}_t$, the left-hand side variable in the SECM (1.13). Specifically,

$$CV = \frac{1}{T} \sum_{t=1}^T \left| \Delta\tilde{\eta}_t - \widehat{\Delta\tilde{\eta}_t} \right|, \quad (1.14)$$

where the forecast $\widehat{\Delta\tilde{\eta}_t}$ for period t is based on a model of $\Delta\tilde{\eta}_t$ that is estimated excluding the sample information (both dependent and explanatory variables) in period t . It is also referred as the leave-one-out estimator. In this sense, $\widehat{\Delta\tilde{\eta}_t}$ is an out-of-sample forecast for $\Delta\tilde{\eta}_t$. Thus, the better the fit of the model, the smaller the CV statistic.

To unravel the relative importance of each determinant ($cr_{t-1}^W, my_{t-1}, tp_{t-1}$), we consider three different sets of models to obtain $\widehat{\Delta\tilde{\eta}_t}$. The first is the SECM (1.13),

²⁰To explore the sensitivity of these results, we also apply the dynamic least squares (DOLS) approach proposed by Stock and Watson (1993) to evaluate the sign and significance of the cointegration parameters. Test regressions include one lead and one lag of differentiated variables. DOLS estimates support significant influences of cr_t^W , my_t and tp_t of the right sign for $W = 23, \dots, 29$.

²¹It is noteworthy that we obtain estimates with correct signs for all three cointegration parameters also for $W \in (3, 20)$ (not shown).

to which we refer as the full model. The second set includes bivariate models of $\tilde{\eta}_t$ and one of the three determinants. And the third type includes trivariate models of $\tilde{\eta}_t$ including two of the three determinants. A particular determinant is regarded as more informative for the mean of the PtDR if either its CV statistic from the bivariate model is close to that of the full model, or the CV statistic from the trivariate model without this determinant indicates a deterioration of the CV statistic.

W from cr_t^W	23	24	25	26	27	28	29
Panel A: CV from the full model							
	0.0220	0.0216	0.0216	0.0215	0.0217	0.0221	0.0226
Panel B: Relative CV for the bivariate models							
With cr_t^W	1.0429	1.0458	1.0524	1.0675	1.0638	1.0557	0.9994
With my_t	1.0958	1.0946	1.0971	1.1022	1.0902	1.0717	1.0672
With tp_t	1.1346	1.1348	1.1374	1.1427	1.1303	1.1111	1.0434
Panel C: Relative CV for the trivariate models							
Without cr_t^W	1.0928	1.1018	1.1043	1.1094	1.0974	1.0788	1.0749
Without my_t	1.0801	1.0932	1.1022	1.1155	1.1128	1.1028	1.0270
Without tp_t	0.9867	0.9967	0.9956	1.0011	0.9973	0.9866	0.9832

Table 1.7: Cross validations This table documents the cross validation (CV) statistics with distinct window sizes W for consumption risk measures. Panel A displays the CVs using the full model of $\tilde{\eta}_t$ with all three determinants (cr_t^W , my_t , tp_t). Panel B shows the ratio of CVs from bivariate models of $\tilde{\eta}_t$ with one determinant and the CVs from the corresponding full model in Panel A. These quotients are referred as relative CV. Similarly, Panel C shows the relative CV from trivariate models of $\tilde{\eta}_t$ with two determinants using the CV from the corresponding full model in Panel A as the benchmark.

We find that among the three factors consumption risk is most informative for changes in the mean of the PtDR while changes in the dividend payout policy of firms appear to be least informative. Table 1.7 documents CV statistics from the full model (Panel A) along with the ratio of the CV statistics from the bivariate model (Panel B) and trivariate model (Panel C) to those from the corresponding full model. Focusing on Panel B, we can see that using cr_{t-1}^W in a bivariate model leads to markedly smaller loss than using my_{t-1} or tp_{t-1} for all different window sizes W . Using $W = 25$ as an example, the bivariate models with cr_{t-1}^W , my_{t-1} or tp_{t-1} have higher CV statistics than those of the full model by 5.2%, 9.7% and 13.7%, respectively. Conditional on the statistics documented in Panel C, cr_{t-1}^W and my_{t-1} appear to be comparably informative for the changes of the mean of the PtDR. By removing cr_{t-1}^W or my_{t-1} from the full model, the corresponding CV

statistics increase by similar proportions (around 10% for most window sizes). In contrast, the removal of tp_{t-1} shows little effect on the CV outcome.²²

1.5 Conclusions

In this paper, we consider a slowly evolving mean of the price-to-dividend ratio in the US, which is inspired by persistent dynamics of this series. We relax the assumption of a constant mean in the present value model (Campbell and Shiller; 1988) towards a time-varying mean of the PtDR, and formalize a state space model to estimate its latent path. Log-likelihood statistics support the model. Adjusting the PtDR by its slowly evolving mean is fruitful in out-of-sample forecasting of stock returns. It outperforms both adjusting the PtDR for structural mean shifts, and the historical average return as a common benchmark predictor.

We also provide evidence for economic influences on the mean of the PtDR. We find that consumption risk, the demographic structure of the population and firm's dividend payout policy all play significant roles in shaping the slowly evolving mean of the PtDR. Among these determinants, consumption risk turns out to be the dominant force.

For future research it would be interesting to compare the time-varying mean of the PtDR from different markets and to uncover potential common components in their variations. International risk sharing could be one potential (global) determinant. As Artis and Hoffmann (2008) have pointed out, international risk sharing has increased since financial markets became more integrated since the 1980s. This might have played an important role in determining variations in the long-run PtDR of different markets in this period.

²²In an in-sample framework, likelihood ratio (LR) statistics can assess the significance of distinct model fits based on residual variances. We adopt this framework to mean squared cross validation errors to underpin the strength of models' CV statistics. We apply the 95% quantile of a χ^2 -distribution with one degree of freedom as approximate critical value. All CV statistics from bivariate models are significantly different from the respective statistics of full models. In contrast, LR-type statistics for the trivariate models without cr_t^W or my_t are significant while those of models without tp_t are insignificant. Hence, it is further undepinned that tp_t is less important than cr_t^W and my_t to shape $\tilde{\eta}_t$.

Appendix to Paper 1

In the following we provide further details on the analyzed data (Appendix 1.A), the particle filtering approach to the estimation of the state space model (Appendix 1.B), and discuss the approximation errors involved when deriving the observation equation of the state space model by means of a Taylor expansion (Appendix 1.C).

1.A Data description

S&P500 stock market indices and dividends Annual series are provided by Amit Goyal and available from the internet.²³ They contain the S&P500 index based on end-of-year closing *prices* and corresponding *dividends* for the period from 1871 to 2012. Annual dividends correspond to the sum of the four quarterly paid dividends within the corresponding year. For more details see Goyal and Welch (2008). The S&P500 index in 2013 is drawn from datastream ('S&PCOMP') and the corresponding dividend is provided by Robert J. Shiller and available from the internet.²⁴

CRSP stock market indices and dividends From 1926 to 2013 we apply annual end-of-year returns based on the weighted market portfolio (NASDAQ, NYSE, AMEX) of the Center for Research in Security Prices (CRSP). We follow Lettau and Van Nieuwerburgh (2008) and calculate the *prices* from the return excluding dividend payments and the *dividends* from the dividend yield D_t/P_{t-1} . Annual dividends correspond to the sum of the four quarterly paid dividends within the corresponding year. From 1871 to 1925 we apply the end-of-year S&P500 index and corresponding dividends employed in Goyal and Welch (2008) and described above.

Interest rates and inflation Similar to Campbell and Vuolteenaho (2004), we use a *short term rate* based on 3-month US Treasury Bills of the Federal Reserve System to approximate the risk-free rate. We employ the series provided by Amit

²³<http://www.hec.unil.ch/agoyal/>

²⁴<http://www.econ.yale.edu/~shiller/data.htm>

Goyal for the period from 1871 to 2012 which is available from the internet.²⁵ More details can be found in Goyal and Welch (2008). In 2013 we update the risk free rate by means of the 3-month US Treasury Bill rate in terms of the ‘secondary market’ quote, published by the Board of Governors of the Federal Reserve System and available from the internet.²⁶

The annual *inflation* series from 1871 to 2013 is extracted from the consumer price index for all urban consumers as provided by Robert J. Shiller.²⁷ For the period before 1913 Shiller calculated the series by means of data in Warren and Pearson (1935). For more details see Shiller (1992, 2005).

Other macroeconomic variables The ratio of the 40-49 over the 20-29 year aged *population* is determined by Census annual population data collected from Datastream (period since 1950, ‘USPOP24Y’ for the 20-24 year old agents, ‘USPOP29Y’ for the 25-29, ‘USPOP44Y’ for the 40-44 and ‘USPOP49Y’ for the 45-49). Data for the period before 1950 are directly from the US Census Bureau.²⁸

Annual quotes of real per capita *consumption* (1929 to 2013) are derived from the sum of the personal consumption expenditures on non-durables and services of the Bureau of Economic Analysis which are two subgroups of the US total personal consumption expenditures (Tables 2.3.5. ‘personal consumption expenditures by major type of product’ and ‘2.3.4. price indexes for personal consumption expenditures by major type of product’).²⁹ The total US population is drawn from the sources described above (the corresponding datastream code is ‘USPOPTO.’). For periods before 1929 we use the series of real per capita total consumption collected by Barro and Ursua (2008). This series ranges from 1834 to 2009 and is available from the net.³⁰ To join the sum of non-durables and services specific consumption measures with the data of Barro and Ursua we regress the sum of both BAE series (co_t^{BEA}) on a constant and the series of Barro and Ursua (co_t^{BU}) in the overlapping sample (1929-2009) and estimate the pre-1929 data from the latter source. The estimated regression is $co_t^{BEA} = 2.932 + 1.012co_t^{BU} + \hat{v}_t$ with a R^2 of 0.998.

For information regarding the measurement of the share of firms paying traditional dividends the reader may consider Kim and Park (2013).

²⁵<http://www.hec.unil.ch/agoyal/>

²⁶<http://www.federalreserve.gov/releases/h15/data.htm>

²⁷<http://www.econ.yale.edu/~shiller/data.htm>

²⁸<http://www.census.gov/popest/archives/pre-1980/PE-11.html>

²⁹<http://www.bea.gov/itable/index.cfm>

³⁰<http://scholar.harvard.edu/barro/publications/macroeconomic-crises-1870-bpea>

1.B Particle filtering

The state space model of the price-to-dividend ratio in (1.7) and (1.8) is highly nonlinear in the parameters, and the maximization of the corresponding log-likelihood function is not tractable analytically. Using Monte Carlo approximation techniques it becomes possible to derive an approximative log-likelihood value by means of particle filtering. We apply the standard particle filter described in Cappé et al. (2007) (Algorithm 3, bootstrap filter) and an optimization technique based on the simplex search method of Lagarias, Reeds, Wright and Wright (1998) for parameter estimation that does not depend on the gradient of the log-likelihood function. The particle filtering algorithm, specific for the state space model provided in Section 1.2, involves the following steps:

Step (1): Initialization ($t=1$). Sample N particles $\tilde{\eta}_1^{(i)} \sim N(\tilde{\eta}_0, \sigma_u^2)$, $i=1, \dots, N$ and determine importance weights and normalized weights, respectively, as

$$\underline{w}_1^{(i)} = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{1}{2} \left(\underline{\epsilon}_1^{(i)}/\sigma_\epsilon\right)^2\right) \quad \text{and} \quad w_1^{(i)} = \frac{\underline{w}_1^{(i)}}{\sum_{i=1}^N \underline{w}_1^{(i)}}.$$

Step (2): Iteration ($t=2, \dots, T$).

1. *Select N particles according to weights $w_{t-1}^{(i)}$. Set accordingly $\tilde{\eta}_{t-1}^{(i)} = \tilde{\eta}_{t-1}^{(i)}$ (resampling)*
2. *For all particles draw*

$$\tilde{\eta}_t^{(i)} \sim N(\tilde{\eta}_{t-1}^{(i)}, \sigma_u^2), \quad i = 1, \dots, N,$$

and determine raw and normalized weights, respectively, as

$$\underline{w}_t^{(i)} = \frac{1}{\sqrt{2\pi\sigma_\epsilon^2}} \exp\left(-\frac{1}{2} \left(\underline{\epsilon}_t^{(i)}/\sigma_\epsilon\right)^2\right) \quad \text{and} \quad w_t^{(i)} = \frac{\underline{w}_t^{(i)}}{\sum_{i=1}^N \underline{w}_t^{(i)}}.$$

3. *go back to step '1'.*

Averaging over non-normalized weights $\underline{w}_t^{(i)}$ yields estimates of the contribution of ϵ_t to the Gaussian likelihood function, while averaging over draws $\tilde{\eta}_t^{(i)}$, i.e., $\hat{\eta}_t = \frac{1}{N} \sum_{i=1}^N \tilde{\eta}_t^{(i)}$, results in estimates of $\tilde{\eta}_t$, for $t = 1, \dots, T$.

So-called systematic resampling is used to compute uniformly distributed random numbers to implement the resampling step. This technique is described in Robert and Casella (2005). Doucet and Johansen (2009) argue that such a technique reduces the noise introduced by resampling, and it is commonly employed in the related literature.

1.C Approximation errors

- (i.) **Approximation** $E_t[\rho_{t+i}] \approx \rho_t$: Based on the empirical observations in Section 1.2 it becomes reasonable to assume η_t , and also $\tilde{\eta}_t$, to follow a random walk. To derive in (1.4), a function of returns r_{t+1} which is linear in η_t , we apply a first order Taylor approximation based on $\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t))$. In consequence, ρ_t is concave in $\tilde{\eta}_t$ and therefore $E_t(\rho_{t+i}) \leq \rho_t$ by Jensen's inequality. However, as displayed in the upper panel of Figure 1.9 the function $\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t))$ is approximately linear in the domain of $\tilde{\eta}_t \in [2.8, 4.1]$ that is the range of the estimate of $\tilde{\eta}_t$ displayed in Figure 1.3.

To evaluate the degree of concaveness of $\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t))$ and the impact of the approximation error we compute the difference between $b\rho(\tilde{\eta}_t^1) + (1 - b)\rho(\tilde{\eta}_t^2)$ and $\rho(b\tilde{\eta}_t^1 + (1 - b)\tilde{\eta}_t^2)$ for any $b \in [0, 1]$ and $\tilde{\eta}_t^1, \tilde{\eta}_t^2 \in [2.8, 4.1]$. The maximal error is 0.0061 in absolute terms and, thus, relatively small.

- (ii.) **Approximation** $E_t[\kappa_{t+i}] \approx \kappa_t$: In the first order Taylor expansion κ_t is determined as $\kappa_t \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(1/\rho_t - 1)$. Thus, κ_t is also concave in ρ_t and $E_t[\kappa_{t+i}] \leq \kappa_t$ by Jensen's inequality. In the relevant domain of $\rho_t \in [0.943, 0.984]$ which is directly implied by $\tilde{\eta}_t \in [2.8, 4.1]$ $\kappa_t = -\ln(\rho) - (1 - \rho) \ln(1/\rho - 1)$ is approximately linear as displayed in the middle panel of Figure 1.9. The maximal difference between $b\kappa(\rho_t^1) + (1 - b)\kappa(\rho_t^2)$ and $\kappa(b\rho_t^1 + (1 - b)\rho_t^2)$ for any $b \in [0, 1]$ and $\rho_t^1, \rho_t^2 \in [0.943, 0.984]$ is 0.0064 in absolute terms and relatively small. Thus, the approximation error in $E_t[\kappa_{t+i}] \approx \kappa_t$ is negligible.

- (iii.) **Approximation** $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$: To evaluate the magnitude of the error implied by this approximation we perform a simulation study. The parameter estimations $\tilde{\eta}_0 = 2.892$ and $\sigma_u = 0.059$ from Ta-

ble 1.2 are applied to simulate the process $\tilde{\eta}_t = \tilde{\eta}_{t-1} + u_t$ as a random walk, for $t = 1, \dots, T$. To reflect the range of the empirical PtDR we bound the random walk by the minimum (2.288) and the maximum of the this measure (4.495). The innovations u_t are generated as $N(0, \sigma_u)$. The process of $\rho_t = 1/(1 + \exp(-\tilde{\eta}_t))$ is simulated subsequently. Moreover, we separate the dataset into $t = 1, \dots, T_1, T_1 + 1, \dots, T$ and neglect the first T_1 observations as initialization period. To simulate η_t we add a first order moving average process, obtaining

$$\eta_t = \tilde{\eta}_t + \gamma \xi_{t-1} + \xi_t. \quad (1.15)$$

The moving average specification for ξ_t accounts for correlation of leads and lags. The parameter γ and the standard deviation of ξ_t are estimated based on our empirical data ($\gamma = 0.726$ and $\sigma_\xi = 0.188$). We draw innovations ξ_t from $N(0, \sigma_\xi)$.

To approximate $E_t[\rho_{t+i}\eta_{t+i+1}]$ we define $\psi_t = \rho_t\eta_{t+1}$ and forecast at each point in time t $\hat{\psi}_{t+i}$ by means of estimated AR(10) processes, for $i = 1, \dots, H$. With regard to $E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ we estimate AR(10) processes for each series separately and determine at each point in time t the corresponding forecasts $\hat{\rho}_{t+i}$ and $\hat{\eta}_{t+i+1}$, for $i = 1, \dots, H$. To estimate all AR(10) models we separate the dataset as $t = T_1 + 1, \dots, T_2, T_2 + 1, \dots, T$ and apply a recursive window starting with the first $T_2 - T_1$ observations. Thus, in total we are left with $T - T_1 - T_2$ periods for which we compute H forecasts. If the approximation error implied by setting $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ is small, the product of the two separate forecasts $\hat{\rho}_{t+i}$ and $\hat{\eta}_{t+i+1}$ should come close to the forecast $\hat{\psi}_{t+i}$.

This procedure is repeated R times and calculations are stored at each point in time t as $\hat{\psi}_{r,t+i}$, $\hat{\rho}_{r,t+i}$ and $\hat{\eta}_{r,t+i+1}$, for $r = 1, \dots, R$ and $i = 1, \dots, H$. To determine the approximation error of interest we use the following statistic

$$\bar{\Omega}_i = \frac{1}{R(T - T_1 - T_2)} \sum_{r=1}^R \sum_{t=T_2+1}^T \Omega_{r,t+i}, \quad \text{for } i = 1, 2, \dots, H, \quad (1.16)$$

where

$$\Omega_{r,t+i} = \frac{|df_{r,t+i}|}{\hat{\psi}_{r,t+i}}, \quad (1.17)$$

with

$$|df_{r,t+i}| = |\hat{\psi}_{r,t+i} - \hat{\rho}_{r,t+i}\hat{\eta}_{r,t+i+1}|. \quad (1.18)$$

We set $R = 1000$, $T = 2000$, $T_1 = 500$, $T_2 = 500$ and $H = 100$. The lower panel of Figure 1.9 displays $\bar{\Omega}_i$, for $i = 1, \dots, 100$. With increasing forecasting horizons the average approximation error converges. It reaches not more than 1% for the 100-step ahead forecasting. As a result, the magnitude of this error is rather small.

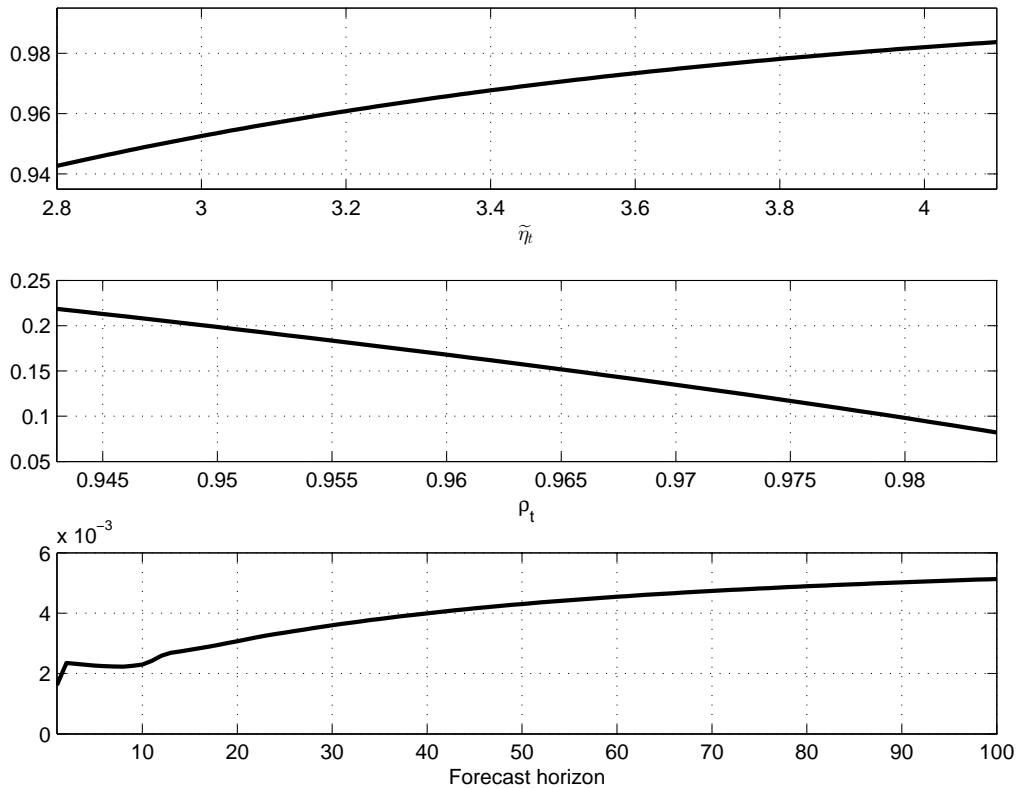


Figure 1.9: Evaluation of approximations The upper panel displays $\rho_t \equiv 1/(1 + \exp(-\tilde{\eta}_t))$ for $\tilde{\eta}_t \in [2.8, 4.1]$ and the middle panel $\kappa(\rho_t) \equiv -\ln(\rho_t) - (1 - \rho_t) \ln(1/\rho_t - 1)$ for $\rho_t \in [0.943, 0.984]$. The lower panel indicates the approximation error ($\bar{\Omega}_i$) from $E_t[\rho_{t+i}\eta_{t+i+1}] \approx E_t[\rho_{t+i}]E_t[\eta_{t+i+1}]$ for alternative forecast horizons $i = 1, \dots, 100$.

Paper 2

Dimensions of macroeconomic uncertainty: A common factor analysis*

Steffen Henzel and Malte Rengel

Abstract. Uncertainty about the future course of the economy is a potential driver of aggregate fluctuations. To identify the distinct dimensions of uncertainty in the macroeconomy, we construct a large dataset covering all types of economic uncertainty. We then identify two fundamental factors which account for the common dynamics in this dataset. These factors are interpreted as macroeconomic uncertainty. The first factor captures business cycle uncertainty, while the second factor is identified as oil and commodity price uncertainty. While both types of uncertainty generate a decline in output, surprise increases in oil and commodity price uncertainty are more important for fluctuations in real activity.

2.1 Introduction

Uncertainty about the future course of the economy is identified as a potential driving force of business cycle fluctuations by recent research (see Bloom; 2009; Alexopoulos and Cohen; 2009; Knotek II and Khan; 2011; Bachmann, Elstner and Sims; 2013, among others). However, many different dimensions of uncertainty are considered in the economic literature.¹ In fact, time-varying uncertainty surrounds the movements of almost all macroeconomic variables (Gonçalves and Kilian; 2004). Thus, policy makers and researchers face the challenge of monitoring and disentangling the distinct types of uncertainty in the macroeconomy.

* Different versions of this paper are available as ‘Dimensions of macroeconomic uncertainty: A common factor analysis’, *Ifo Working Paper No. 167*, 2013 and *CESifo Working Paper No. 4991*, 2014. The first version of the paper circulated under the title ‘Common uncertainty factors’.

¹ Among many others, Friedman (1977) and Ball (1992) discuss the effects of inflation uncertainty. Furthermore, Ramey and Ramey (1995) and Grier and Perry (2000) are concerned with the effects of production uncertainty. Born and Pfeifer (2014), Fernández-Villaverde, Guerrón-Quintana, Kuester and Rubio-Ramírez (2013) and Baker, Bloom and Davis (2013) model policy uncertainty. Oil price uncertainty is analyzed by, for instance, Elder and Serletis (2010).

In this paper, we provide an empirical justification for the number of fundamental second moment shocks that affect macroeconomic dynamics and interpret these shocks economically. We analyze a large dataset as in Giannone, Reichlin and Sala (2004) which consists of 164 individual uncertainty measures covering all types of economic uncertainty. Using a dynamic factor model, we identify the common fundamental factors that have an impact on these uncertainty measures and interpret them as distinct types of macroeconomic uncertainty.² This definition of macroeconomic uncertainty is similar to Jurado et al. (2014) who independently proposed a factor based approach to measure uncertainty. However, Jurado et al. (2014) focus on the factor with the largest explanatory power, whereas the identification and disentanglement of distinct fundamental types of macroeconomic uncertainty is at the core of our study.

A large body of literature on factor models documents that there are only a handful of fundamental first moment shocks which are relevant for explaining the business cycle (see Sargent and Sims; 1977; Forni and Reichlin; 1998; Stock and Watson; 2002; Giannone et al.; 2004, among others). Since the number of fundamental shocks contained in a factor model is typically much smaller than the number of variables in the dataset, factor models provide a justification for the modeling strategy underlying modern business cycle models. Recent studies enhance business cycle models with shocks to second moments and analyze their role for business cycle fluctuations, for instance, during the recent global financial crisis (Gilchrist, Sim and Zakrajšek; 2014; Born and Pfeifer; 2014; Arellano, Bai and Kehoe; 2012; Fernández-Villaverde, Guerrón-Quintana, Rubio-Ramírez and Uribe; 2011; Basu and Bundick; 2012; Bloom, Floetotto, Jaimovich, Saporta-Eksten and Terry; 2012; Schaal; 2012; Bachmann and Bayer; 2013). Our study contributes by providing guidance how many shocks to the second moments are important for the fluctuations of uncertainty at the macroeconomic level.

Our results provide evidence that macroeconomic uncertainty in the US economy is driven by two fundamental shocks. We document that the first fundamental shock triggers uncertainty associated with demand related variables such as aggregate production and capacity utilization. Thus, we identify the corresponding factor as ‘business cycle uncertainty’. The second fundamental shock mainly affects uncertainty associated with supply related variables such as oil and commodity prices, and we interpret the related factor as ‘oil and commodity price uncertainty’. Moreover, we show that the indicators of business cycle uncertainty and oil and

² It is worthwhile to notice that macroeconomic uncertainty relates to aggregate shocks, as opposed to idiosyncratic (e.g., firm-specific) uncertainty which results from the dispersion of idiosyncratic shocks, which themselves average out over many units (e.g., firms).

commodity price uncertainty, which we obtain from the factor approach, peak at business cycle and oil-market related events, respectively. We document that the factor-based measure of Jurado et al. (2014) combines both types of macroeconomic uncertainty. Moreover, we diagnose non-negligible differences between our measures of macroeconomic uncertainty and measures of financial market uncertainty which are often used as a proxy for aggregate economic uncertainty (Bloom; 2009; Basu and Bundick; 2012; Gilchrist et al.; 2014, among others). We demonstrate that uncertainty about economic policy, as put forward by Baker et al. (2013), is closely related to business cycle uncertainty, but not to oil and commodity price uncertainty.

Impulse response analyses reveal that both types of macroeconomic uncertainty generate a drop in real activity, i.e., both types of macroeconomic uncertainty are countercyclical. However, both types of uncertainty have distinct quantitative implications for real activity. While the decline following a surprise increase in business cycle uncertainty is not very pronounced, we document a larger and more persistent effect for oil and commodity price uncertainty. Such a finding suggests that unforeseen shifts in oil and commodity price uncertainty are more important for business cycle fluctuations than those in business cycle uncertainty.

The remainder of the paper is structured as follows. In Section 2.2, we introduce the dataset and Section 2.3 explains how we measure macroeconomic uncertainty by means of a dynamic factor model. The number of fundamental uncertainty factors is analyzed in Section 2.4, where we also give an interpretation of the factors and provide a comparison to alternative uncertainty measures. Subsequently, we analyze how macroeconomic uncertainty relates to economic activity. The paper concludes in Section 2.5. An appendix contains technical descriptions of the estimation procedures and several robustness analyses.

2.2 The dataset

We use a large-scale monthly dataset as in Giannone et al. (2004) which is commonly used to describe the development of the US economy over the business cycle (see also Stock and Watson; 2002). The data is split into 14 categories: industrial production, capacity utilization, employment, sales and consumption, housing and construction, inventories, new and unfilled orders, financial variables, interest rates, monetary variables, prices, wages, merchandise ex- and imports and business outlook. Our analysis is based on data covering the period 1970M1 to 2011M4. During this period, the Federal Reserve has undertaken a number of unconventional policy

measures ('quantitative easing') leading to severe outliers in monetary aggregates distorting the estimation of uncertainty later on. Hence, we exclude the monetary base (series 117, Table 2.3 in Appendix 2.A), the depository institutions reserves (series 118-119) and the loans and securities at all commercial banks (series 125). Moreover, consistent data for the commercial paper outstanding (series 104), the delinquency rate on bank-held consumer installment loans (series 126), and the index of sensitivity materials prices (series 132) are not available. This leaves us with $N = 164$ variables which are transformed to achieve stationarity. A detailed list of single series and a description of the transformations are provided in Table 2.3 in Appendix 2.A.

2.3 Measuring macroeconomic uncertainty

2.3.1 Individual uncertainty measures

In the following, we define macroeconomic uncertainty as the common movement of time-varying uncertainty surrounding the dynamics of a broad range of variables. To measure macroeconomic uncertainty, it is necessary to measure latent variable specific uncertainty first. For this purpose, we rely on the time series dimension of individual economic variables since a direct measure of uncertainty is not available. We consider a data driven filter to approximate the conditional volatility which is known as the RiskMetrics procedure (Morgan; 1996).³

An alternative time series based measure of uncertainty is the conditional volatility of GARCH models which can be interpreted as (in-sample) forecast uncertainty (see Baillie, Chung and Tieslau; 1996; Grier and Perry; 1998; Lahiri and Sheng; 2010; Grimme, Henzel and Wieland; 2014, among others). Moreover, stochastic volatility (SV) models are frequently used since they involve an additional independent shock to the time-varying second moment of a process which itself has an economic interpretation (see Fernández-Villaverde and Rubio-Ramírez; 2010; Born and Pfeifer; 2014; Jurado et al.; 2014, for a discussion). However, the calculation

³ Measures of conditional volatility have the advantage that they are not affected by time-varying risk aversion as opposed to observable indicators of financial market risk such as the option-based CBOE implied volatility index (VXO) (see Bekaert, Hoerova and Lo Duca; 2013, for a discussion). Alternatively, the literature discusses several measures of uncertainty relying on the cross sectional dimension of data. Examples are the dispersion of forecasts collected in surveys or the average standard deviation of subjective probability distributions, for instance, polled in the Survey of Professional Forecasters (see Zarnowitz and Lambros; 1987; Giordani and Söderlind; 2003; Rich and Tracy; 2010, among others). However, surveys are limited to selected variables such as GDP, GDP deflator, and CPI inflation and, thus, these approaches are not suited to measure general macroeconomic uncertainty.

of GARCH and SV measures necessitates an explicit formulation of a (parametric) data generating process for the conditional volatility. To meet the specific requirements of a certain variable, innumerable extensions of the basic GARCH and SV model would have to be considered. To provide a transparent measurement for our multitude of time series, we consider the data driven RiskMetrics filter to approximate the conditional volatility which is a robust alternative to GARCH and SV models. The objective is to calculate an exponentially weighted moving average of the squared (in-sample) forecast errors of an appropriately defined forecast model.

We follow Stock and Watson (2002), who forecast each variable of interest y_t by means of a single-equation autoregressive model augmented with predictions of a vector of common factors $\hat{g}_{t+1|t}$ which summarizes the dynamics of a large cross section of variables. This forecast equation is formalized as⁴

$$y_{t+1} = \alpha_0 + \sum_{j=1}^M \alpha'_j \hat{g}_{t-j+2|t-j+1} + \sum_{j=1}^P \beta_j y_{t-j+1} + e_{t+1}. \quad (2.1)$$

The (in-sample) forecast error is given by e_{t+1} . Adding factors ensures that predictions are based on a large information set which prevents an overestimation of the uncertainty surrounding the forecast (see Jurado et al.; 2014, for a similar argument). Estimates of the factors are obtained from a parametric dynamic factor model. The specification of the dynamic factor model follows common practice in the forecasting literature (see, for instance, Kose, Otrok and Whiteman; 2003; Bańbura, Giannone and Reichlin; 2011) and reads as

$$Y_t = \Lambda g_t + \zeta_t \quad (2.2)$$

$$g_t = \Psi(L)g_{t-1} + \varepsilon_t, \quad (2.3)$$

where Y_t is of dimension $N \times 1$ and collects the variables in our information set. The factors are collected in the vector g_t , and Λ is a conformable matrix of factor loadings. Idiosyncratic movements are captured by ζ_t . The factors follow a simple vector autoregression with innovations ε_t . We estimate the model with the quasi maximum likelihood procedure proposed by Doz, Giannone and Reichlin (2012), i.e., we apply an expectation maximization (EM) algorithm combined with Kalman filtering and smoothing.⁵ Once the model parameters are estimated using full sample information, the filtered factors $\hat{g}_{t+1|t}$ employed in (2.1) are obtained from the

⁴ To simplify the notation, we suppress the variable specific index i , for $i = 1, \dots, N$, in univariate equations in the following. In (2.1), examples are variables $y_{i,t}$, parameters $\alpha_{0,i}$ and lag orders M_i .

⁵ A detailed description of the estimation procedure is provided in Appendix 2.B.

Kalman prediction step which processes only information available up to period t .⁶

During estimation we set the number of factors in (2.2) to 12 as indicated by the information criteria of Bai and Ng (2002). The lag polynomial $\Psi(L)$ contains four lags. The number of factors r subsequently employed in the forecast equation (2.1) and the lag orders M and P are determined for each variable by means of the BIC in the range of $0 \leq r \leq 12$, $1 \leq M \leq 4$ and $1 \leq P \leq 6$, respectively. Based on an initial sample beginning in 1968M6, we obtain $T = 496$ (in-sample) forecast errors for each of the 164 variables in our dataset covering the period 1970M1 to 2011M4.

The time-varying uncertainty σ_t^2 for each individual series follows

$$\sigma_t^2 = \phi\sigma_{t-1}^2 + (1 - \phi)e_{t-1}^2. \quad (2.4)$$

As before, e_t denotes the one-step ahead forecast error from (2.1). We implement the RiskMetrics procedure by rewriting (2.4) as an infinite sum $\sigma_t^2 = (1 - \phi) \sum_{h=1}^{\infty} \phi^{h-1} e_{t-h}^2$, which, in practice, is truncated at $t - 1$. To avoid distortions of the uncertainty measure at the beginning of the sample, we apply the adjustment factor $1/(1 - \phi^t)$ as proposed by Andersen, Bollerslev, Christoffersen and Diebold (2006). The decay parameter ϕ in (2.4) governs the smoothness of the variance process.⁷ The thin black lines in Figure 2.1 show the resulting uncertainty measures associated with industrial production (IP) and consumer price inflation (CPI), respectively.

Before we proceed with the measurement of macroeconomic uncertainty, we verify whether the uncertainty measures for IP and CPI from the RiskMetrics approach are consistent with a measure obtained from the estimated parametric SV model

$$e_t = \exp(\log(\sigma_{t,SV}^2)/2)\vartheta_t, \quad (2.5)$$

$$\log(\sigma_{t,SV}^2) = \gamma_0 + \gamma_1 \log(\sigma_{t-1,SV}^2) + \omega\nu_t. \quad (2.6)$$

The logarithmic conditional variance in (2.6) is modeled as a first-order autoregressive process. The innovations ϑ_t and ν_t are both $N(0, 1)$ distributed and mutually

⁶ In this in-sample framework, the parameters of the factor model and the forecast equation (2.1) are estimated using full sample information and, thus, the parameter uncertainty tends to be underestimated. Recursive estimation would solve the problem but a substantial part of the sample has to be reserved to initialize the estimation of the factor model in (2.2) and (2.3). In a robustness analysis in Appendix 2.H.2, we apply observations from the first five years to initialize recursive model estimates, and demonstrate that, for the remaining years, the measure of macroeconomic uncertainty derived below is almost unaffected by using either in-sample or out-of-sample forecast errors.

⁷ We set $\phi = 0.7$ implying a half-life period of the innovation e_t of approximately two months. As proposed by Morgan (1996), we use the same value for ϕ for all series in the considered dataset.

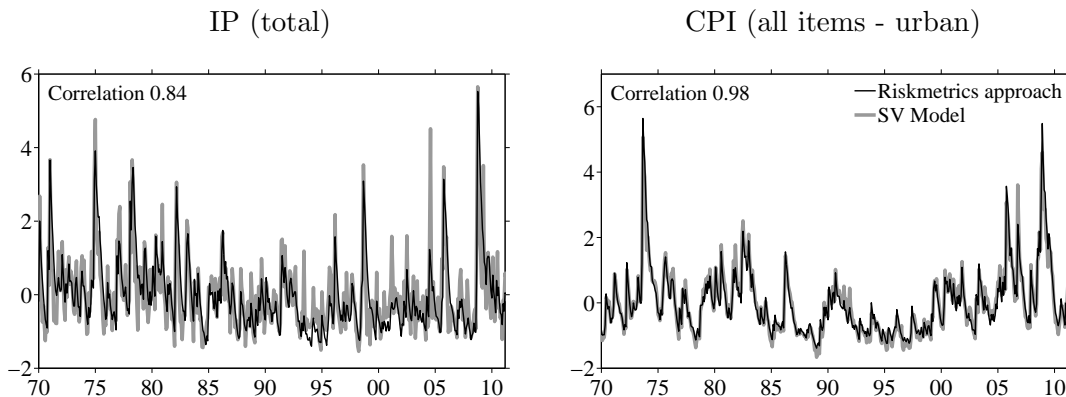


Figure 2.1: RiskMetrics and SV measures Panels display SV measures (bold gray lines) and RiskMetrics measures (thin black lines). To facilitate comparability both measures are demeaned and standardized.

independent. We estimate the model using a maximum likelihood approach. Due to the non-linearity of the model, we simulate the likelihood function by means of the efficient importance sampler (EIS) (Liesenfeld and Richard; 2003). Given the parameter estimates, we use efficient importance sampling to obtain the filtered variance estimates. The bold gray lines in Figure 2.1 display the resulting series. SV and RiskMetrics measures co-move closely and their correlation amounts to 0.84 (IP), and 0.98 (CPI), respectively. In summary, individual uncertainty measures are largely unaffected by the procedure used to filter the variance. The RiskMetrics approach appears to be a simple but admissible method to evaluate the uncertainty surrounding each of the 164 variables in a data-driven, transparent and robust manner.

2.3.2 Measures of macroeconomic uncertainty

To identify the common dynamics of the individual uncertainty measures, which we interpret as macroeconomic uncertainty, we estimate a factor model similar to that used in Section 2.3.1

$$X_t = \lambda f_t + \xi_t, \quad (2.7)$$

$$f_t = \psi(L)f_{t-1} + \epsilon_t. \quad (2.8)$$

The standardized individual uncertainty measures are collected in the data vector $X_t = (\log(\sigma_{1,t}), \dots, \log(\sigma_{N,t}))'$. We take the logarithmic square root of $\sigma_{i,t}^2$ to allow factors to adopt negative values. Macroeconomic uncertainty is assessed by means of the k -dimensional vector of fundamental factors f_t which is responsible for the common dynamics of the individual uncertainty measures. The latter is summarized

by means of the common component $\lambda_t = \lambda f_t$. The remaining unexplained part of X_t is governed by the idiosyncratic component ξ_t which is allowed to be weakly cross sectionally and serially correlated ('approximate dynamic factor model'). Doz et al. (2012) show that the unobserved factors f_t of an approximate dynamic factor model are estimated consistently if N and T become large even if cross sectional and serial correlation of the idiosyncratic components is not modeled explicitly.⁸ The factors in (2.8) follow an autoregressive process where the innovations ϵ_t can be decomposed into $\epsilon_t = Ru_t$. The k -dimensional vector u_t contains orthogonal shocks and R is a $k \times k$ conformable matrix. We assume that the number of shocks u_t is identical to the number of common factors k , i.e., f_t follows a vector autoregression (e.g., Kose et al.; 2003). Later on we make use of this model structure to identify and interpret each of the orthogonal shocks and the common factors separately. Since $k \ll N$, the number of fundamental shocks k identifies the reduced dimension of the dataset.

As in Section 2.3.1, we estimate the model (2.7) to (2.8) using the method of Doz et al. (2012). The lag order of $\psi(L)$ is set to six which is sufficient to capture the dynamics of the data.

2.4 Estimation results

2.4.1 The number of common uncertainty factors

In the following, we aim at disclosing the number of fundamental factors which have an impact on a broad range of the individual uncertainty measures. First, we consider the explained variance proportion (R^2) of the estimated factor model in (2.7) and (2.8) for the aggregate dataset of uncertainty measures which results for distinct numbers of dynamic factors k . The model with one factor explains 18% of the variation. The gain in explained variance by adding a second factor is 8% while the third factor adds about 6%, and the fourth factor adds 4%. In summary, the gain of an additional factor decreases relatively fast suggesting not more than four common factors.

In Figure 2.2, we display the explained variance proportion (R^2) for each individual uncertainty measure. In the upper panel, gray bars indicate the R^2 when $k = 1$. In this specification the common component has explanatory power mainly

⁸ As a robustness analysis, we allow the idiosyncratic processes to follow a first order autoregressive process (Bańbura and Modugno; 2012). The pairwise correlations of the factors from this generalized model and from our baseline model are all greater than 0.98.

for the uncertainty associated with most production variables, capacity utilization and some employment variables. However, the explanatory power for other uncertainty measures, especially those of most price variables, is relatively low. One factor appears to be insufficient to capture the common dynamics of uncertainty. Consequently, we introduce a second fundamental factor and calculate the individual R^2 for this model specification which is represented by the crossed line in the upper panel of Figure 2.2. As a result, the variance proportion which is explained by the common component improves for a broad range of uncertainty variables. In particular, the R^2 increases for many price related uncertainty measures. In

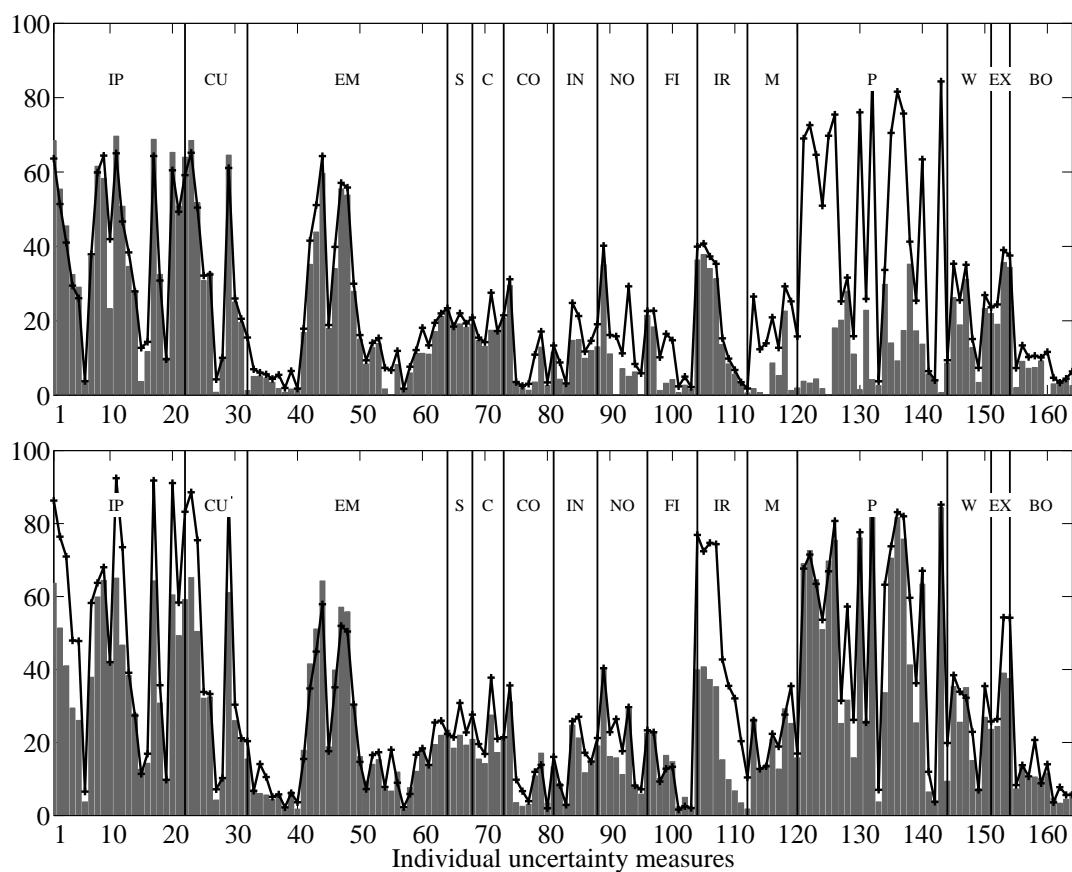


Figure 2.2: Explained variance proportions for distinct numbers of factors The upper panel displays on the y-axis the respective proportion of the explained variance for $k = 1$ (gray bars) and $k = 2$ (crossed line). The lower panel documents the explained variance shares for $k = 2$ (gray bars) and $k = 3$ (crossed line). The x-axis corresponds to individual uncertainty measures in the same order as in Table 2.3 in Appendix 2.A. They are grouped into the following categories: *IP* (1-21, industrial production), *CU* (22-31, capacity utilization), *EM* (32-63, employment), *S* (64-67, sales), *C* (68-72, consumption), *CO* (73-80, housing and construction), *IN* (81-87, inventories), *NO* (88-95, new and unfilled orders), *FI* (96-104, financial variables), *IR* (105-113, interest rates), *M* (114-126, monetary variables), *P* (127-151, prices), *W* (152-158, wages), *EX* (159-161, merchandise ex- and imports), *BO* (162-167, business outlook).

the lower panel of Figure 2.2, we display the effect of adding a third factor to the model. An improvement in terms of R^2 appears only for a rather narrow group of individual measures of uncertainty surrounding short-term interest rates (variables 105 to 108). Since the third factor does not add much to the common dynamics, we conclude that it does not represent macroeconomic uncertainty. The number of common factors appears to be two.

Two common factors explain a large fraction of the variation in uncertainty associated with important business cycle variables such as industrial production ($R^2 = 0.64$), industrial production in manufacturing ($R^2 = 0.65$), capacity utilization ($R^2 = 0.59$), employment in manufacturing ($R^2 = 0.57$), and consumer prices ($R^2 = 0.75$).⁹ To put these numbers in perspective, note that our data are monthly and the noise component accounts for roughly one third of the total variance.¹⁰

In Appendix 2.C, we provide more formal tests for the number of common factors. Some caution is however warranted since the factors are themselves subject to estimation error which makes the testing problem non-standard. As a result conventional tests might have problems in correctly determining the number of factors in our dataset. In summary, information criteria do not provide a unique indication of the precise number of factors, however they support the notion that this number ranges around two. Thus, we proceed the analysis using two common factors.¹¹

2.4.2 The distinct types of macroeconomic uncertainty

To interpret the two fundamental shocks u_t that govern the common dynamics of economic uncertainty, we study the response of the individual uncertainty measures to each of these shocks. The factors have the moving average representation

$$f_t = [I_k - \psi(L)L]^{-1}Ru_t. \quad (2.9)$$

It follows that the impulse response function of the common component $\chi_t = \lambda f_t$ is

$$\chi_t = \lambda[I_k - \psi(L)L]^{-1}Ru_t, \quad (2.10)$$

$$= B(L)u_t. \quad (2.11)$$

⁹ A complete list of the results is provided in Table 2.3 in Appendix 2.A.

¹⁰ To obtain this number, we estimate the spectral density of each series and calculate the average contribution to the total variance of the frequencies smaller than 18 months. These frequencies are usually attributed to the noise component of a time series in the business cycle literature (see, for instance, the seminal work of Burns and Mitchell; 1946).

¹¹ In a robustness exercise in Appendix 2.H.3, we show that our main results do not change if a factor model with three common factors is applied.

Given a rotation matrix H , with $HH' = I_k$, the common component can be rewritten as $\chi_t = C(L)v_t$, where $C(L) = B(L)H$ and $v_t = H'u_t$. However, there is an infinite number of rotation matrices H and the impulse response function is, in general, not identified. As proposed by Giannone et al. (2004), we solve the identification problem by selecting the rotation matrix H such that the target function

$$g(H) = \frac{\sum_{i \in D} \sum_{l=0}^{\infty} (c_{i1}^l)^2}{\sum_{i \in D} \sum_{l=0}^{\infty} (c_{i1}^l)^2 + \sum_{i \in D} \sum_{l=0}^{\infty} (c_{i2}^l)^2} \quad (2.12)$$

is maximized. This target function is based on a forecast error variance decomposition. The item c_{ij}^l denotes the response of the individual uncertainty measure i to innovation j (for $j = 1, 2$) after l periods. The selection vector D determines a subset of uncertainty measures which enters the target function. The denominator is the forecast error variance of the selected uncertainty measures which is explained by the two shocks jointly. In the baseline specification, we identify the first shock such that its explanatory power for uncertainty surrounding all production related variables is maximized, i.e., D determines all uncertainty measures related to output (series 1 to 31). The identification strategy reflects the idea that the model should at least be able to capture fluctuations in production uncertainty. The second shock is unrestricted.

To quantify the relative importance of each fundamental shock for individual uncertainty measures, we perform a forecast error variance decomposition. Results for the 12-month horizon are reported in the columns labeled shock 1 and 2 in Table 2.3 in Appendix 2.A. Moreover, we use impulse responses to assess the sign (and the significance) of the response of individual uncertainty measures to a positive one unit shock. Dynamic responses to the first fundamental shock are displayed in the left hand sided column of Figures 2.3 and 2.4, respectively. Responses to the second fundamental shock are depicted in the corresponding right hand sided columns.¹²

¹²Confidence intervals are obtained from a bootstrap procedure which is outlined in Appendix 2.D.

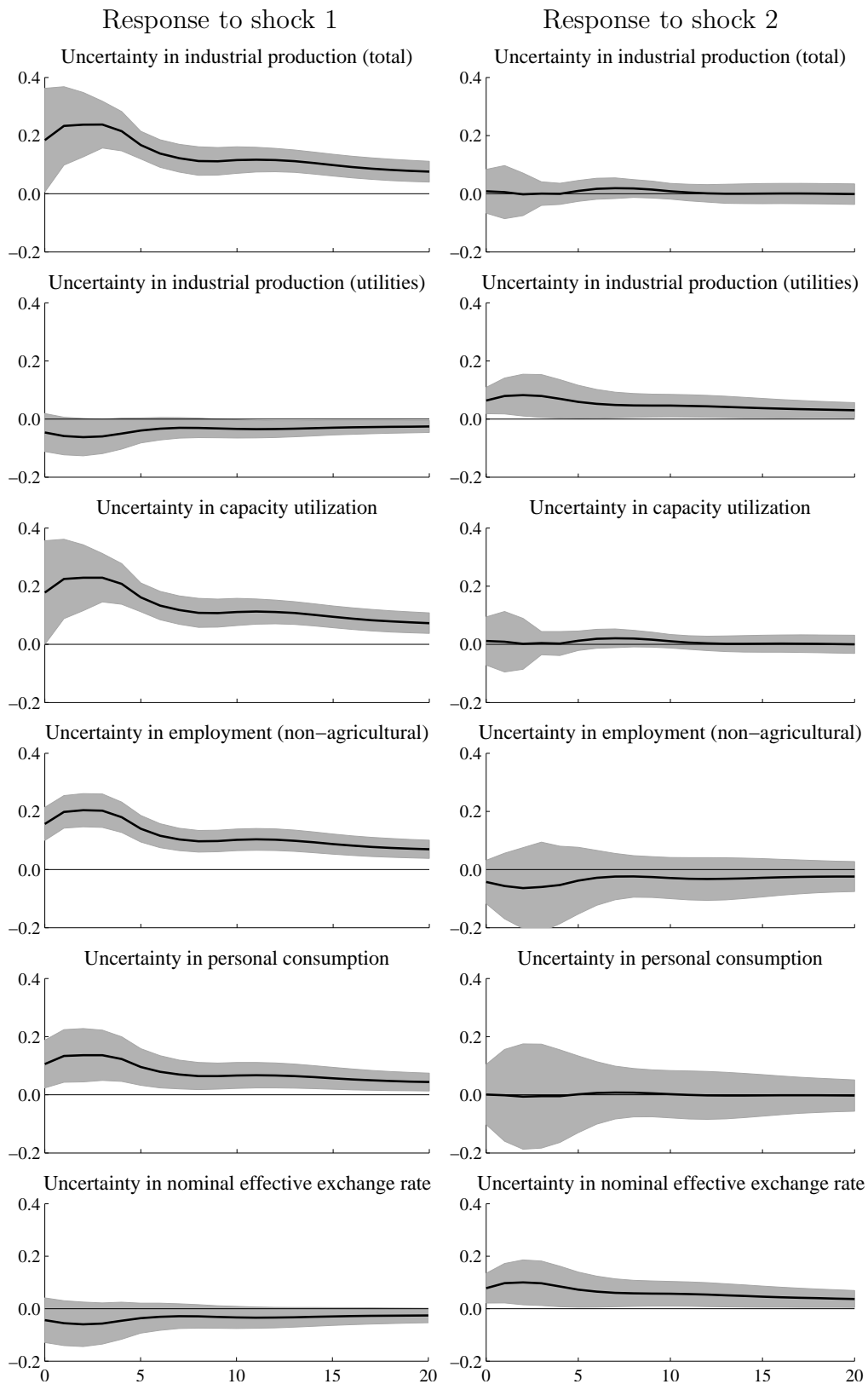


Figure 2.3: Impulse responses of individual uncertainty measures (1) Responses to the first fundamental shock ($u_{1,t}$, left column) and to the second fundamental shock ($u_{2,t}$, right column) are indicated by the bold lines. Shaded areas display bootstrapped 90% confidence intervals.

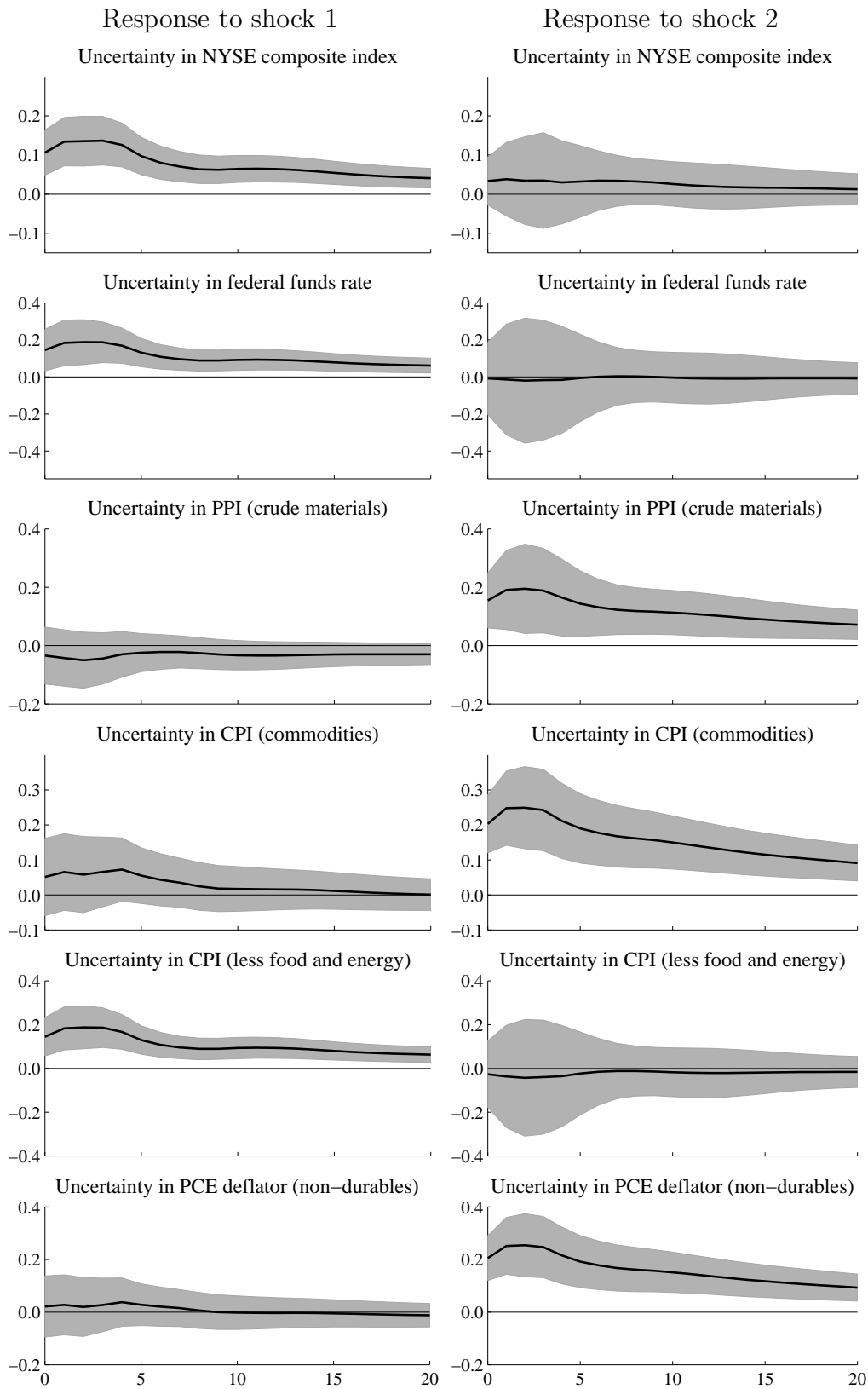


Figure 2.4: Impulse responses of individual uncertainty measures (2) For further notes see Figure 2.3.

From the forecast error variance decomposition, we confirm that the first fundamental shock is important for fluctuations of uncertainty associated with aggregate production, capacity utilization, and aggregate employment. Moreover, uncertainty related to important business cycle indicators such as sales, consumption, inventories, and new orders as well as exports and imports is captured in most cases by shock 1. This surprise innovation also explains a major fraction of fluctuations in stock market uncertainty. Moreover, this shock is an important driver of uncertainty related to the federal funds rate as well as for uncertainty surrounding assets which in general carry little risk such as treasuries. The results for the distinct inflation uncertainty measures are less unambiguous. Shock 1 explains uncertainty about core inflation, which is also closely linked to aggregate demand, while it does not contribute to uncertainty about other price variables related to (unprocessed) input goods such as commodities. The first shock also governs uncertainty about hourly earnings.

The impulse responses in Figure 2.3 and 2.4 show that the first shock precedes a significant increase of uncertainty surrounding variables which bear a close relation to the business cycle, whereas it is not followed by an increase in uncertainty related to utility production and the nominal effective exchange rate. Moreover, Figure 2.4 confirms that shock 1 is followed by a rise in uncertainty related to core inflation (CPI less food and energy), while there is no significant increase in inflation uncertainty measures for non-processed goods and non-durable consumption goods, which comprises mainly gasoline.¹³ In summary, our results suggest that shock 1 governs uncertainty surrounding variables which determine aggregate demand conditions in an economy and bear a close relation to the business cycle. Thus, we identify the first fundamental shock as ‘business cycle uncertainty shock’.

In Table 2.3 (Appendix 2.A), we observe that the second fundamental shock is important for uncertainty surrounding the production of utilities. Shock 2 is also quantitatively important for uncertainty of all producer prices. Other price variables which are explained by this shock are consumer prices which reflect energy price movements such as the CPI for transportation, for commodities, or the PCE deflator for non-durables. Moreover, shock 2 explains a major fraction of uncertainty related to the exchange rate which itself is an important determinant of domestic energy

¹³There are also a number of individual uncertainty measures for which both fundamental shocks play a role. Notably, 73% (76%) of the common variation in uncertainty surrounding CPI (PCE deflator) headline inflation is attributable to the second shock, and 27% (24%) are explained by the first shock. Furthermore, uncertainty related to total loans and securities is governed by both shocks.

and import prices.

The impulse responses in the right column of Figure 2.3 confirm that there is no significant increase of uncertainty associated with most business cycle variables. The second shock precedes a rise in uncertainty associated with the production of utilities, and it is followed by higher exchange rate uncertainty. Moreover, Figure 2.4 reveals that the second shock precedes an increase in uncertainty about prices of unprocessed goods such as oil and commodities. We identify this shock as ‘oil and commodity price uncertainty shock’.¹⁴

In the following, we compare the historical movements of shock 2 to the precautionary oil-market specific demand shock put forward by Kilian (2009). The latter shock may occur on the oil market when agents develop a precautionary demand for oil which serves as an insurance against a potential shortfall of oil supply. If shock 2 governs oil and commodity price uncertainty it should be related to the movements of Kilian’s (2009) precautionary oil demand shock, which is identified using a VAR framework.¹⁵ Figure 2.5 shows that the peaks of both series are highly synchronized which suggests that they reflect the same oil-market specific events. However, both innovations do not operate in the same direction throughout.¹⁶ The economic interpretation of the oil-market specific demand shock and a shock in oil and commodity price uncertainty is different. While an increase in precautionary demand likely reflects the fear of rising oil prices, uncertainty may also increase when oil and commodity prices drop. In summary, these results provide further support for our interpretation of the second shock as oil and commodity price uncertainty shock.

¹⁴In Appendix 2.H.1, we perform an alternative rotation where we identify the second shock as driver of oil and commodity price uncertainty while the first shock is left unrestricted. The interpretation of the two fundamental shocks is robust to this change in the identification strategy.

¹⁵We have updated the respective shock series until 2011 using the material provided on Lutz Kilian’s website: <http://www-personal.umich.edu/~lkilian/>.

¹⁶While, for instance, both series co-move during the Gulf Wars, a large drop in oil-specific demand is accompanied by a rise in oil and commodity price uncertainty during the global financial crisis. The sign of the oil-market specific demand shock is, for instance, negative during a recession when there are expectations about an oil production surplus. During such episodes oil and commodity price uncertainty typically increases.

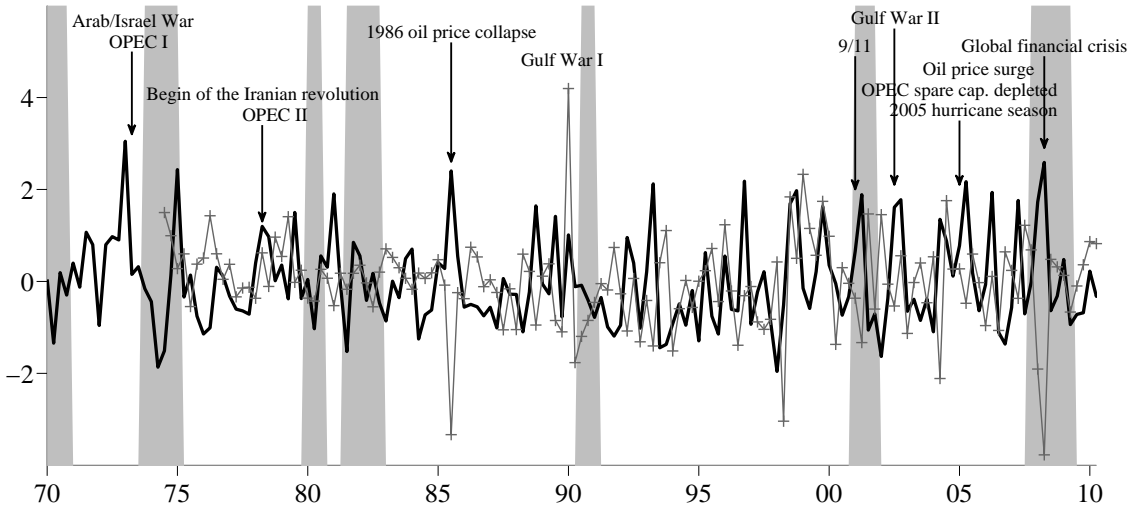


Figure 2.5: The oil price uncertainty shock and an oil-market specific demand shock The thick black line represents innovations to oil and commodity price uncertainty, and the crossed thin gray line depicts the oil-market specific demand shock of Kilian (2009). The correlation among both series is 0.23 and significant at the one percent level. Shaded areas represent NBER recession dates. To improve the readability, the figure displays quarterly averages, and the series are standardized.

2.4.3 Indicators of macroeconomic uncertainty

In the following, we use the parametric structure of the dynamic factor model in (2.7) and (2.8) to derive two indicators of macroeconomic uncertainty. For this purpose we separately identify the $k = 2$ common factors in $f_t = (f_{1,t}, f_{2,t})'$. Since the factors are latent, we encounter a further identification problem which we solve by means of a matrix rotation procedure such that the proportion of variance explained by the first factor is maximized for uncertainty measures associated with output (series 1 to 31) (see also Kose, Otrok and Prasad; 2012). This procedure is similar to the rotation approach used in Section 2.4.2, but focuses on the unconditional variances of uncertainty measures instead of their forecast error variances.¹⁷ We refer to f_t as the estimates which are identified by this rotation strategy. Note that while the fundamental shocks are orthogonal, this is not necessarily the case for the factors since they follow the VAR process in (2.8). However, when we calculate the forecast error variance decomposition for the VAR in the rotated factors with respect to the elements in u_t , the first fundamental shock explains 98.4% of the forecast error variance of the first common factor $f_{1,t}$ twelve months ahead, while the second fundamental shock explains 94.4% of the variation in $f_{2,t}$. Thus, our interpretation of the fundamental shocks carries over to the common factors, and $f_{1,t}$ is a measure of business cycle uncertainty whereas $f_{2,t}$ provides a measure of oil

¹⁷In Appendix 2.E, we present the details of the rotation strategy that we use to separately identify the two factors.

and commodity price uncertainty.

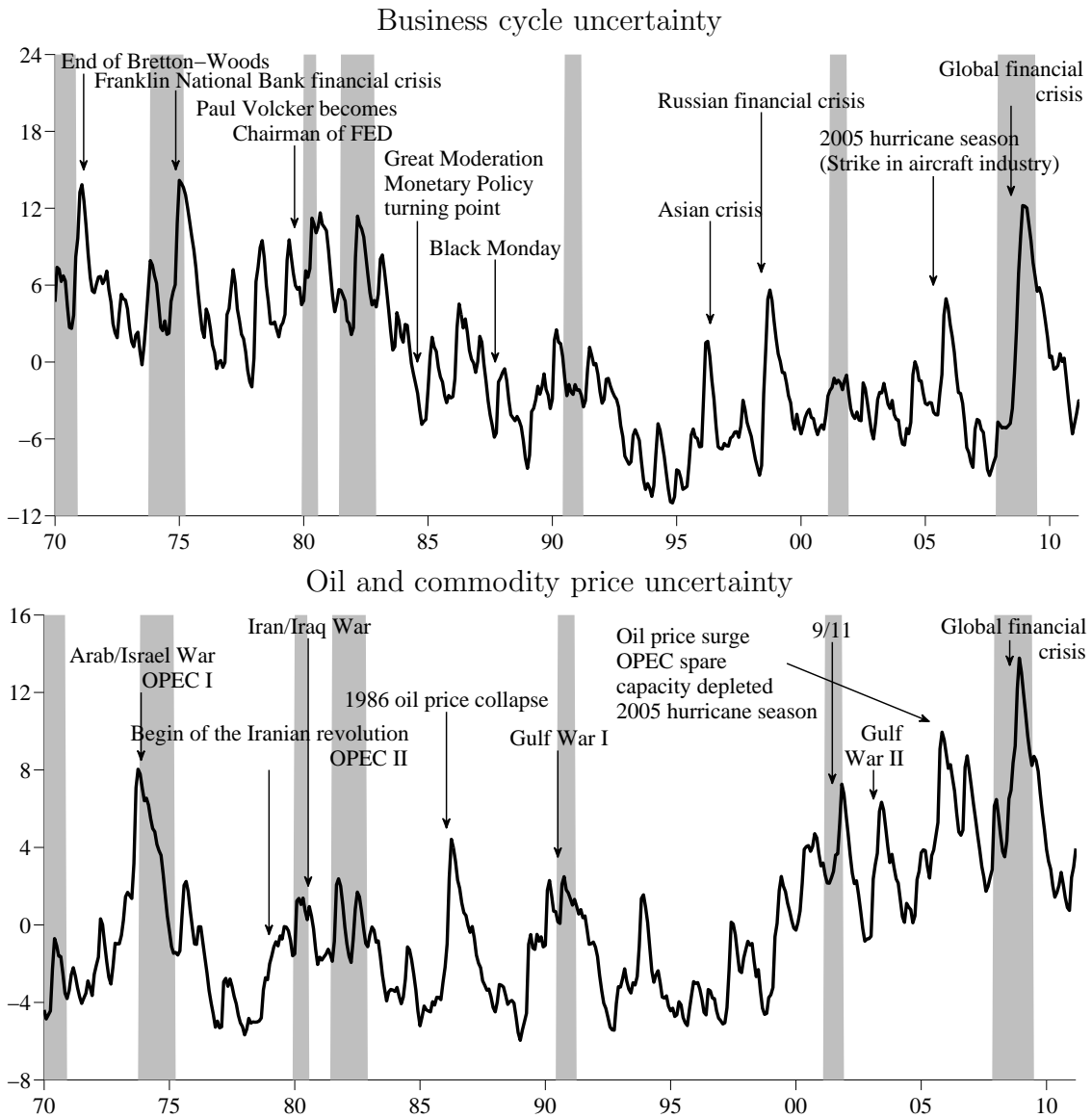


Figure 2.6: Indicators of macroeconomic uncertainty The upper panel displays business cycle uncertainty and the lower panel depicts oil and commodity price uncertainty. NBER recession dates are indicated by the shaded areas.

Time paths of both indicators of macroeconomic uncertainty are linked to major events that are commonly associated with economic turmoil.¹⁸ In Figure 2.6 the upper panel displays our indicator of business cycle uncertainty along with important business cycle events. NBER dated recessions are indicated by shaded areas. It is apparent that business cycle uncertainty is countercyclical since it rises dur-

¹⁸As discussed in Section 2.3.1, we calculate uncertainty indicators based on real time forecast errors in Appendix 2.H.2 as a sensitivity analysis. The indicators of macroeconomic uncertainty are almost unaffected by using either in-sample or out-of-sample forecast errors.

ing recessions. The indicator of business cycle uncertainty first increases when the Bretton-Woods system collapsed in 1973. The next extraordinary peak is related to the bankruptcy of the Franklin National Bank in 1974 which led to a banking crisis and a period of increased macroeconomic turmoil. Moreover, the indicator peaks in 1979 when Paul Volcker became chairman of the FED and changed the conduct of monetary policy. In the period extending from the mid-1980s, business cycle uncertainty retraces the decline of macroeconomic volatility known as the ‘Great Moderation’. Business cycle uncertainty increases only moderately around the Black Monday in 1987, while a more pronounced increase is observed during the Asian crisis, the Russian financial crisis, the hurricane season 2005 (‘Hurricane Katrina’), and after the bankruptcy of Lehman Brothers in 2008 which preceded the global financial crisis.

Oil and commodity price uncertainty is displayed in the lower panel of Figure 2.6 along with oil-market specific events and NBER dated recessions. As business cycle uncertainty, oil and commodity price uncertainty rises during recessions. Peaks occur in 1973 during the Arabian/Israel War, the Iranian Revolution in 1978/79, and the Iran/Iraq War in 1980, as well as in 1986 when the oil price dropped heavily, and during the first Gulf War in 1990/91. Beginning in the mid-1990s, oil and commodity price uncertainty is characterized by an upward trending pattern which coincides with a time when the OPEC repeatedly adjusted its quotas. The indicator rises sharply around 9/11 when changes in the oil production quota were no longer sufficient to stabilize the oil price. Afterwards, it increases moderately during the second Gulf War and peaks in the mid-2000s reflecting the decrease in oil spare capacity in the aftermath of the second Gulf War. At the beginning of the global financial crisis, the oil price first increased until its sudden decline accompanied by a peak in oil and commodity price uncertainty.

Several alternative measures have been proposed to investigate the dynamics of uncertainty in the macroeconomy and its relation to economic activity. A number of studies use the option-based CBOE implied volatility index (VXO) to measure uncertainty perceived on financial markets (see, for instance, Bloom; 2009; Basu and Bundick; 2012). Figure 2.7 compares the VXO and the indicator of business cycle uncertainty.¹⁹ Both series increase around, for instance, the Franklin National Bank crisis, during the Russian financial crisis, and the recent global financial crisis. However, the tightness of the relation among both measures varies over time. In

¹⁹To obtain a long history of implied stock market volatility, we use Bloom’s (2009) measure which dates back to 1962.

particular, the VXO peaks around the Black Monday and the stock market crash of 2002, while there is virtually no reaction of business cycle uncertainty. The comparison of the VXO with the indicator of oil and commodity price uncertainty is provided in the lower panel of Figure 2.7. We also observe that these two measures are synchronized during certain episodes.

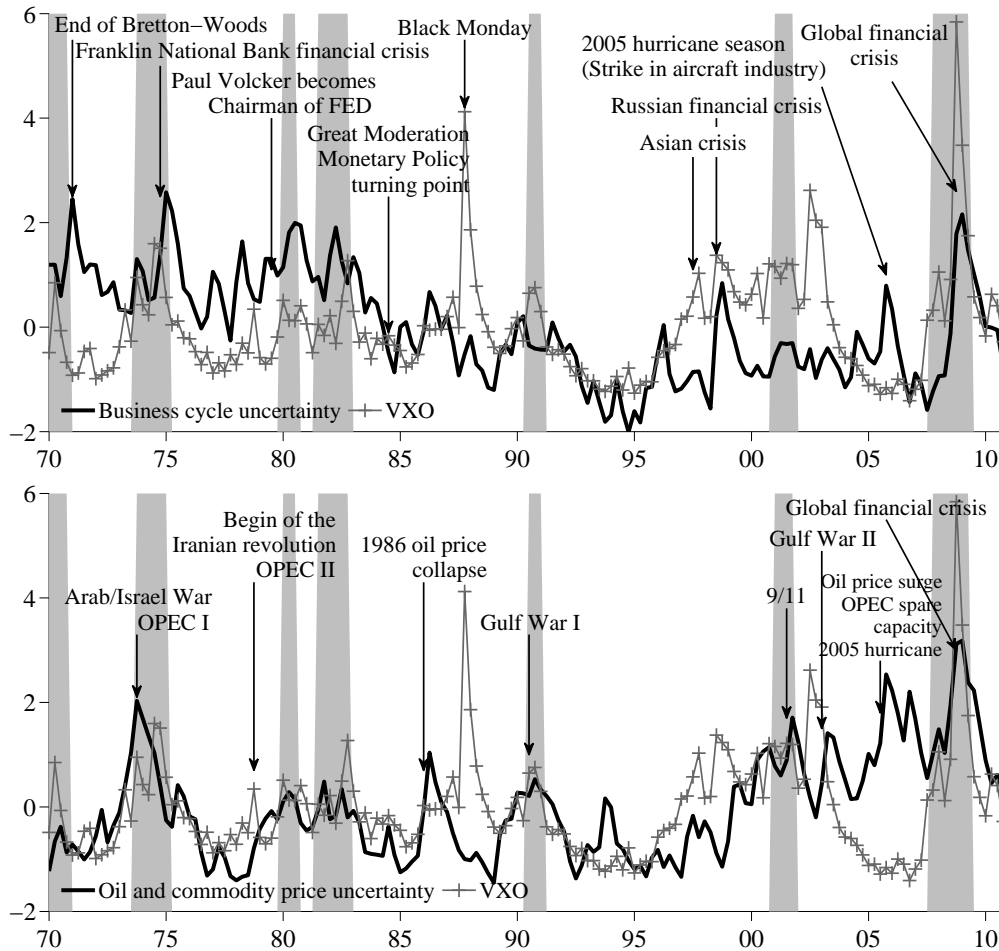


Figure 2.7: Indicators of macroeconomic uncertainty and the VXO The upper panel depicts business cycle uncertainty and the lower panel oil and commodity price uncertainty. The crossed (gray) line represents stock market volatility (VXO). To facilitate the interpretation of the graph, we calculate quarterly averages of the monthly observations and standardize all series. NBER recession dates are indicated by the shaded areas.

The results of a regression of the VXO on contemporaneous values of business cycle uncertainty and oil and commodity price uncertainty are reported in the first row of Table 2.1. Both indicators of macroeconomic uncertainty are significant at the 5% level which suggests that the VXO reflects, both, business cycle uncertainty and oil and commodity price uncertainty.

Z	α_1	α_2	R^2
<i>VXO</i>	0.17**	0.32***	0.14
<i>SPREAD</i>	0.44***	0.32***	0.32
<i>JLN</i>	0.59***	0.48***	0.62
<i>EPU</i>	0.48***	0.01	0.23

Table 2.1: Regressions of alternative measures on macroeconomic uncertainty For each alternative measure we run the following regression: $Z_t = \alpha_1 f_{1,t} + \alpha_2 f_{2,t} + d_{\text{Oct87}} + \eta_t$, where $Z_t \in \{VXO, SPREAD, JLN, EPU\}$. Where *VXO* denotes the CBOE stock market volatility index (Bloom; 2009), *SPREAD* the difference between the 30-year Baa-corporate bond yield and the 30-year Treasury bond yield, *JLN* the measure used in Jurado et al. (2014) and *EPU* the index of economic policy uncertainty proposed by Baker et al. (2013). All series are quarterly averages and standardized. Significance at the 1%, 5%, 10% level is denoted by ***, **, *. The regressions contain a dummy for the Black Monday d_{Oct87} .

Similarly, credit spreads (*SPREAD*) have been used to measure the perception of risk of financial investors (Gilchrist, Yankov and Zakrajšek; 2009; Gilchrist and Zakrajšek; 2012; Bachmann et al.; 2013).²⁰ Table 2.1 documents that the regression result is similar to that for the *VXO*. Although financial market uncertainty appears to be related to macroeconomic uncertainty, it differs in a non-negligible way. Turmoil on financial markets (e.g., during the ‘Black Monday’) does not necessarily root in bad economic fundamentals. Another explanation for the differences may be that the *VXO* is to a large extent influenced by time-varying risk aversion of financial investors which is unrelated to economic uncertainty (Bekaert et al.; 2013). In summary, employing financial market uncertainty as a measure for macroeconomic uncertainty may be misleading (Jurado et al.; 2014, reach a similar conclusion).

The measure of Jurado et al. (2014) is conceptually close to ours since it defines macroeconomic uncertainty as the single common factor with the largest explanatory power for a large number of individual uncertainty measures. We obtain a positive regression coefficient with 1% significance in Table 2.1 for both measures of macroeconomic uncertainty. Our identification strategy allows us to discriminate two types of uncertainty which are otherwise combined in the factor with the largest explanatory power.

Baker et al. (2013) propose a measure of economic policy uncertainty (*EPU*) which is derived from newspaper references to policy, tax rate and fiscal budget uncertainty. *EPU* reflects a specific type of uncertainty relating exclusively to future policy decisions. The regression results in Table 2.1 reveal a significantly positive contemporaneous relation to business cycle uncertainty but not to oil and commod-

²⁰The time profiles of alternative uncertainty measures along with the indicators of macroeconomic uncertainty are provided in Appendix 2.F.

ity price uncertainty. Given that EPU seems to reflect business cycle uncertainty only, the comparison confirms our interpretation of the uncertainty factors.²¹

2.4.4 Macroeconomic uncertainty and real activity

We now turn our attention to the dynamic relation between macroeconomic uncertainty and economic activity. Economic activity and macroeconomic uncertainty might be highly endogenous (see, for instance, Bachmann and Moscarini; 2012). To meet this concern we estimate a number of bivariate VARs which consist of one uncertainty factor $f_{j,t}$, for $j = 1, 2$, and one additional variable representing economic activity. The use of bivariate VARs is a parsimonious way to model the joint dynamics of both measures, while controlling for the interrelations among them. To study the response of real activity, we consider industrial production (IP), capacity utilization (CU), and average hours worked (Hours).²² These variables enter in log-levels. We include a linear trend and set the lag length for each VAR to 12 which is sufficient to control for the dynamic history of the variables. The VAR innovations are orthogonalized by means of a Cholesky decomposition where the respective uncertainty factor is ordered last. This ordering implies that uncertainty reacts instantaneously while macroeconomic quantities – which are comparatively slow-moving – react only after one month to a shock in uncertainty. Moreover, it ensures that we control for first-moment shocks when we analyze the result of a surprise increase in macroeconomic uncertainty. The size of the shock to uncertainty is set to obtain an initial increase of macroeconomic uncertainty equal to its observed average monthly increase during the global financial crisis in the second half of 2008. Such a target entails that the size of the shock to business cycle uncertainty is 2.8 standard deviations, and the shock to oil and commodity price uncertainty has a size of 1.6 standard deviations.²³

The upper panel of Figure 2.8 collects the impulse responses for a shock to business cycle uncertainty. The average monthly increase in business cycle uncertainty during the financial crisis leads to a drop in production by 0.6% after six months which is followed by a rebound. A similar path is obtained for capacity utilization and hours worked which decline by 0.9% and 0.2%, respectively. Our results con-

²¹ As a robustness exercise we also calculate the dynamic response of VXO, SPREAD, JLN, and EPU to the two fundamental shocks $u_{1,t}$ and $u_{2,t}$ in Appendix 2.G. Since the dynamic relations are very similar to the contemporaneous ones this exercise further strengthens our results.

²² As in Bloom (2009), Bachmann et al. (2013), and Jurado et al. (2014), we focus on the manufacturing sector.

²³ Commonly, the uncertainty shock in VARs is calibrated for rare events like the global financial crisis, Bloom (2009) and Jurado et al. (2014) consider a four standard deviation shock.

firm the findings documented in previous studies which argue that macroeconomic uncertainty generates economic downturns (see, for instance, Bloom; 2009; Bloom et al.; 2012; Bachmann et al.; 2013).

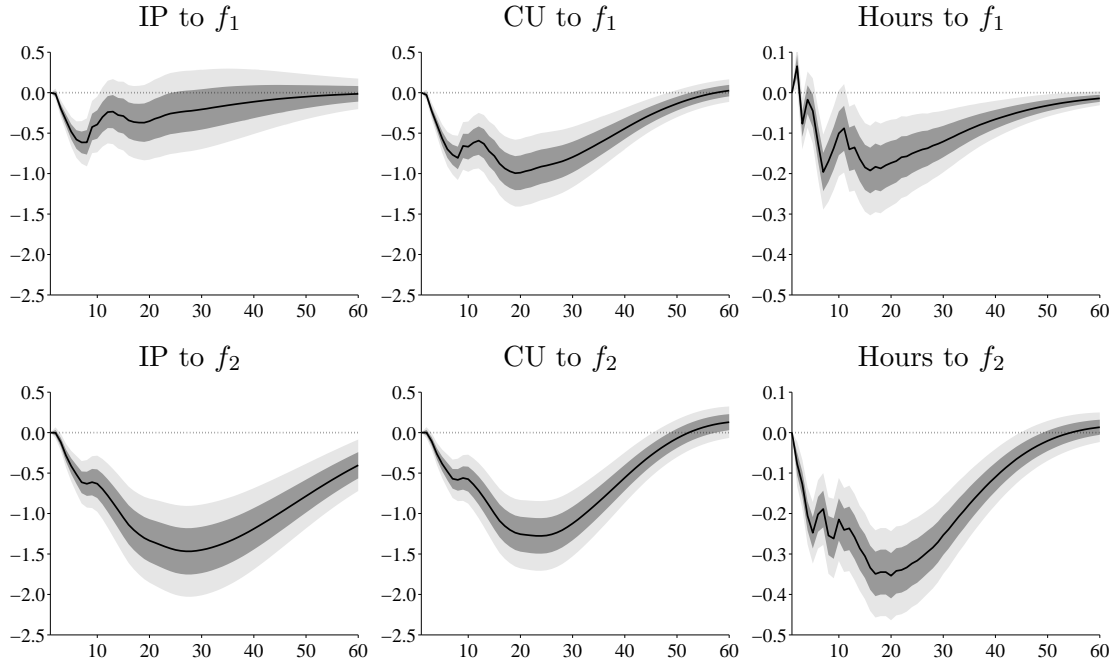


Figure 2.8: Impulse response functions for a shock in macroeconomic uncertainty
 The upper panel displays the response of industrial production (IP), capacity utilization (CU), and hours worked (Hours) to a surprise innovation in business cycle uncertainty. The lower panel displays the response of the same variables to a shock to oil and commodity price uncertainty. The shock size is chosen to obtain an initial increase of macroeconomic uncertainty equal to its observed average monthly increase in the second half of 2008. Confidence intervals are derived from the bias adjusted bootstrap procedure proposed by Kilian (1998). Dark shaded areas indicate \pm one standard deviation confidence intervals, light shaded areas indicate \pm two standard deviation confidence intervals.

The lower panel of Figure 2.8 displays the responses to a surprise innovation in oil and commodity price uncertainty as observed during the global financial crisis. In this case, we observe a long-lasting decline of IP which is much more pronounced than for a shock to business cycle uncertainty, IP drops by 1.5% after about two years. Similarly, capacity utilization drops by 1.3%, and hours worked drop by -0.4%. Since we calibrate the shock to the average increase during the global financial crisis, the size of the shock to oil and commodity price uncertainty is smaller than the shocks to business cycle uncertainty. Nevertheless, the former precedes a stronger decline in real activity.²⁴

²⁴In Appendix 2.H.3, we demonstrate that this result also holds when we allow for $k = 3$ factors

We conduct several robustness analyses with respect to VAR specifications that are documented in Appendix 2.H.4. We calculate the impulse response functions of IP for VAR systems where we choose the lag length for the bivariate VARs by means of the BIC. We also estimate a version without trend. Moreover, we explore whether the ordering of the variables has an impact on the impulse responses. Following the previous literature, we estimate larger VARs consisting of eight variables (Bloom; 2009; Bachmann et al.; 2013; Jurado et al.; 2014). The ordering of the variables is: IP, employment, Hours, CPI, hourly wages, federal funds rate, macroeconomic uncertainty, S&P500 stock prices. All measures except for the interest rate and macroeconomic uncertainty enter in log-levels. In summary, respective results are qualitatively identical.

We document the shares of the forecast error variance attributable to each type of macroeconomic uncertainty for distinct forecast horizons in Table 2.2. Increases in business cycle uncertainty explain at most 1.2% of the fluctuations of IP, for capacity utilization the largest share is 10.7%, and for hours worked it is 4.6%. In contrast, up to 21.5% of the fluctuations of IP are attributable to a sudden increase in oil and commodity price uncertainty, while for CU the ratio is 25.6% and for hours worked it is 27.2%. In Appendix 2.H.5 we demonstrate that this result also holds in a VAR containing both types of macroeconomic uncertainty simultaneously. Moreover, Appendix 2.H.6 documents that using a larger VAR as in Bloom (2009) does not change these results.

Horizon	Business cycle uncertainty			Oil and commodity price uncertainty		
	IP	CU	Hours	IP	CU	Hours
12	1.20	3.28	1.22	3.22	3.82	7.22
24	0.90	7.13	3.33	10.18	15.93	20.29
36	0.84	10.33	4.35	16.70	24.75	26.49
60	0.83	10.71	4.65	21.54	25.64	27.23

Table 2.2: Forecast error variance explained by macroeconomic uncertainty This table presents the share of the forecast error variance (in %) due to business cycle uncertainty (left hand side panel) and oil and commodity price uncertainty (right hand side panel). The shares are obtained from bivariate VARs with 12 lags containing one type of macroeconomic uncertainty and one measure of economic activity.

In summary, our results are in line with the notion that uncertainty generates an economic downturn. However, our findings suggest that unforeseen increases in the

when estimating the factor model in (2.7) and (2.8).

factor which we identify as indicator of oil and commodity price uncertainty are more important for real activity than unforeseen increases in the factor which we interpret as in the indicator of business cycle uncertainty.

2.5 Conclusions

Economic uncertainty has recently been identified as a potential source of business cycle fluctuations. However, the movements of almost all measures in an economy are subject to time-varying uncertainty. As a consequence, numerous individual uncertainty measures exist. Policy makers and researchers face the challenge of monitoring and disentangling the distinct types of uncertainty. In this paper, we disclose the fundamental factors governing the uncertainty about macroeconomic outcomes by means of a dynamic factor model. We interpret the common dynamics underlying individual uncertainty measures as macroeconomic uncertainty.

First, we document that only a small number of fundamental factors accounts for the movements of macroeconomic uncertainty. Our results suggest that two shocks are responsible for the common dynamics of the individual uncertainty measures, i.e., the dimension of macroeconomic uncertainty is two. Second, we demonstrate that the first factor is related to uncertainty surrounding the dynamics of common business cycle variables and, thus, is interpreted as indicator of ‘business cycle uncertainty’ while the second factor reflects ‘oil and commodity price uncertainty’. Third, we analyze how these types of macroeconomic uncertainty relate to alternative measures applied in the literature. Increases of both types of macroeconomic uncertainty are accompanied by an increase in financial market uncertainty measured by means of the VXO or the spread for Baa rated corporate bonds (SPREAD). However, measures such as the VXO or SPREAD should be interpreted carefully since they may indicate uncertainty which is unrelated to macroeconomic fundamentals. Fourth, economic policy uncertainty as investigated by Baker et al. (2013) is closely related to business cycle uncertainty but not to oil and commodity price uncertainty.

In line with the previous literature, we demonstrate that uncertainty is counter-cyclical. However, both types of macroeconomic uncertainty have distinct quantitative implications for real activity. Business cycle uncertainty generates a short-lived and not very pronounced decline in output. Our results suggest that the large swings in business cycle uncertainty are more likely a by-product of recessions than a cause (see Bachmann and Moscarini; 2012, for a theoretical argument). In con-

trast, oil and commodity price uncertainty generates a large and persistent drop in real activity. In summary, time-varying oil and commodity price uncertainty is more important for fluctuations of real activity than business cycle uncertainty.

Appendix to Paper 2

2.A Dataset and forecast error variance decomposition for individual uncertainty measures

		Statistics for <i>uncertainty</i> measures			
		Transformation	R^2	FEVD	
Series				Shock 1	Shock 2
<i>Industrial production</i>					
1	Index of IP: Total	3	0.64	1.00	0.00
2	Index of IP: Final products and nonindustrial supplies	3	0.51	1.00	0.00
3	Index of IP: Final products	3	0.41	1.00	0.00
4	Index of IP: Consumer goods	3	0.29	1.00	0.00
5	Index of IP: Durable consumer goods	3	0.26	0.98	0.02
6	Index of IP: Nondurable consumer goods	3	0.04	0.90	0.10
7	Index of IP: Business equipment	3	0.38	0.87	0.13
8	Index of IP: Materials	3	0.60	0.98	0.02
9	Index of IP: Materials, nonenergy, durables	3	0.64	0.90	0.10
10	Index of IP: Materials, nonenergy, nondurables	3	0.42	0.54	0.46
11	Index of IP: Mfg	3	0.65	1.00	0.00
12	Index of IP: Mfg, durables	3	0.47	0.99	0.01
13	Index of IP: Mfg, nondurables	3	0.38	0.88	0.12
14	Index of IP: Mining	3	0.28	0.99	0.01
15	Index of IP: Utilities	3	0.13	0.34	0.66
16	Index of IP: Energy, total	3	0.14	0.72	0.28
17	Index of IP: Nonenergy, total	3	0.64	1.00	0.00
18	Index of IP: Motor vehicles and parts (MVP)	3	0.31	1.00	0.00
19	Index of IP: Computers, comm. equip. and semiconductors (CCS)	3	0.10	0.98	0.02
20	Index of IP: Nonenergy excl CCS	3	0.60	1.00	0.00
21	Index of IP: Nonenergy excl CCS and MVP	3	0.49	0.99	0.01
<i>Capacity utilization</i>					
22	Capacity utilization: Total	2	0.59	0.99	0.01
23	Capacity utilization: Mfg	2	0.65	0.99	0.01
24	Capacity utilization: Mfg, durables	2	0.50	0.95	0.05
25	Capacity utilization: Mfg, nondurables	2	0.32	0.92	0.08
26	Capacity utilization: Mining	2	0.32	0.97	0.03
27	Capacity utilization: Utilities	2	0.04	0.25	0.75
28	Capacity utilization: Computers, comm. equip. and semiconductors	2	0.10	0.94	0.06
29	Capacity utilization: Mfg excl CCS	2	0.61	0.99	0.01
30	Purchasing managers index (PMI)	3	0.26	1.00	0.00
31	ISM mfg index: Production	3	0.21	0.94	0.06
<i>Employment</i>					
32	Index of help-wanted advertising	3	0.15	0.12	0.88
33	No. of unemployed in the civ. Labor force (CLF)	3	0.07	0.69	0.31
34	CLF employed: Total	3	0.06	1.00	0.00
35	CLF employed: Nonagricultural industries	3	0.06	0.93	0.07
36	Mean duration of unemployment	3	0.04	0.73	0.27
37	Persons unemployed less than 5 weeks	3	0.05	0.29	0.71
38	Persons unemployed 5 to 14 weeks	3	0.02	0.42	0.58
39	Persons unemployed 15 to 26 weeks	3	0.06	0.28	0.72
40	Persons unemployed 15+ weeks	3	0.02	0.99	0.01
41	Avg weekly initial claims	3	0.18	0.98	0.02
42	Employment on nonag. payrolls: Total	3	0.42	0.92	0.08
43	Employment on nonag. payrolls: Total private	3	0.51	0.92	0.08
44	Employment on nonag. payrolls: Goods-producing	3	0.64	0.97	0.03
45	Employment on nonag. payrolls: Mining	3	0.19	0.99	0.01

2.A Dataset and forecast error variance decomposition

		Statistics for <i>uncertainty</i> measures			
Series	Transformation	R^2	FEVD		
			Shock 1	Shock 2	
46	Employment on nonag. payrolls: Construction	3	0.40	0.92	0.08
47	Employment on nonag. payrolls: Manufacturing	3	0.57	0.99	0.01
48	Employment on nonag. payrolls: Manufacturing, durables	3	0.56	0.99	0.01
49	Employment on nonag. payrolls: Manufacturing, nondurables	3	0.30	0.99	0.01
50	Employment on nonag. payrolls: Service-producing	3	0.16	0.99	0.01
51	Employment on nonag. payrolls: Utilities	3	0.09	0.99	0.01
52	Employment on nonag. payrolls: Retail trade	3	0.14	0.97	0.03
53	Employment on nonag. payrolls: Wholesale trade	3	0.15	0.95	0.05
54	Employment on nonag. payrolls: Financial activities	3	0.07	0.22	0.78
55	Employment on nonag. payrolls: Professional and bus. services	3	0.07	0.01	0.99
56	Employment on nonag. payrolls: Education and health services	3	0.12	0.79	0.21
57	Employment on nonag. payrolls: Leisure and hospitality	3	0.02	0.89	0.11
58	Employment on nonag. payrolls: Other services	3	0.08	0.81	0.19
59	Employment on nonag. payrolls: Government	3	0.12	0.99	0.01
60	Avg weekly hrs. of production or nonsupervisory workers	3	0.18	0.67	0.33
61	Avg weekly hrs. of PNW: Mfg	3	0.13	0.85	0.15
62	Avg weekly overtime hrs. of PNW: Mfg	3	0.19	0.97	0.03
63	ISM mfg index: Employment	3	0.22	0.95	0.05
<i>Sales</i>					
64	Sales: Mfg and trade-total (mil of chained 96\$)	3	0.23	0.98	0.02
65	Sales: Mfg and trade-mfg, total (mil of chained 96\$)	3	0.18	0.98	0.02
66	Sales: Mfg and trade-merchant wholesale (mil of chained 96\$)	3	0.22	0.95	0.05
67	Sales: Mfg and trade-retail trade (mil of chained 96\$)	3	0.19	0.96	0.04
<i>Consumption</i>					
68	Personal cons. expenditure: Total (bil chained 96\$)	3	0.21	1.00	0.00
69	Personal cons. expenditure: Durables (bil of chained 96\$)	3	0.15	0.98	0.02
70	Personal cons. expenditure: Nondurables (bil of chained 96\$)	3	0.14	1.00	0.00
71	Personal cons. expenditure: Services (bil of chained 96\$)	3	0.28	0.73	0.27
72	Personal cons. expenditure: Durables (bil of chained 96\$)	3	0.17	0.99	0.01
<i>Housing and construction</i>					
73	Privately-owned housing, started: Total (thous)	3	0.21	0.97	0.03
74	New privately-owned housing authorized: Total (thous)	3	0.31	0.97	0.03
75	New 1-family houses sold: Total (thous)	3	0.04	1.00	0.00
76	New 1-family houses-months supply at current rate	3	0.03	0.88	0.12
77	New 1-family houses for sale at end of period (thous)	3	0.03	0.52	0.48
78	Mobile homes-mfg shipments (thous)	3	0.11	0.33	0.67
79	Construction put in place: Total (mil of 96\$)	3	0.17	0.82	0.18
80	Construction put in place: Private (mil of 96\$)	3	0.03	0.86	0.14
<i>Inventories</i>					
81	Inventories: Mfg, Total (mil of chained 96\$)	3	0.13	0.90	0.10
82	Inventories: Mfg (mil of chained 96\$)	3	0.09	0.46	0.54
83	Inventories: Mfg, durables (mil of chained 96\$)	3	0.03	0.88	0.12
84	Inventories: Mfg, nondurables (mil of chained 96\$)	3	0.25	0.58	0.42
85	Inventories: Merchant wholesale (mil of chained 96\$)	3	0.21	0.80	0.20
86	Inventories: Retail trade (mil of chained 96\$)	3	0.12	0.93	0.07
87	ISM mfg index: Inventories	3	0.15	0.88	0.12
<i>New and unfilled orders</i>					
88	ISM mfg index: New orders	3	0.19	0.66	0.34
89	ISM mfg index: Suppliers deliveries	3	0.40	0.96	0.04
90	Mfg new orders: All mfg industries (mil of current \$)	3	0.16	0.58	0.42
91	Mfg new orders: Mfg ind. with unfilled orders (mil of current \$)	3	0.16	0.01	0.99
92	Mfg new orders: Durables (mil of current \$)	3	0.11	0.52	0.48
93	Mfg new orders: Nondurables (mil of current \$)	3	0.29	0.16	0.84
94	Mfg new orders: Nondefense capital goods (mil of current \$)	3	0.08	0.71	0.29
95	Mfg unfilled orders: All mfg industries (mil of current \$)	3	0.06	0.00	1.00

2 Dimensions of macroeconomic uncertainty

		Statistics for <i>uncertainty</i> measures			
Series	Transformation	R^2	FEVD		
			Shock 1	Shock 2	
<i>Financial variables</i>					
96	NYSE composite index	3	0.23	0.91	0.09
97	S&P composite index	3	0.23	0.76	0.24
98	S&P P/E ratio	3	0.10	0.10	0.90
99	Nominal effective exchange rate	3	0.16	0.24	0.76
100	Spot Euro/US	3	0.15	0.32	0.68
101	Spot SZ/US	3	0.02	0.38	0.62
102	Spot Japan/US	3	0.05	0.67	0.33
103	Spot UK/US	3	0.02	0.82	0.18
104	Commercial paper outstanding (mil of current \$)*	-	-	-	-
<i>Interest rates</i>					
105	Interest rate: Federal funds rate	2	0.40	0.99	0.01
106	Interest rate: U.S. 3-mo Treasury (sec. market)	2	0.41	0.99	0.01
107	Interest rate: U.S. 6-mo Treasury (sec. market)	2	0.37	0.99	0.01
108	Interest rate: 1-year Treasury	2	0.35	0.99	0.01
109	Interest rate: 5-year Treasury (constant maturity)	2	0.15	0.98	0.02
110	Interest rate: 7-year Treasury (constant maturity)	2	0.10	0.96	0.04
111	Interest rate: 10-year Treasury (constant maturity)	2	0.07	0.95	0.05
112	Bond yield: Moodys AAA corporate	2	0.03	0.94	0.06
113	Bond yield: Moodys BAA corporate	2	0.02	0.32	0.68
<i>Monetary variables</i>					
114	M1 (bil of current \$)	3	0.27	0.06	0.94
115	M2 (bil of current \$)	3	0.12	0.06	0.94
116	M3 (bil of current \$)	3	0.14	0.02	0.98
117	Monetary base, adj. for reserve requirement (rr) changes (bil of \$)*	-	-	-	-
118	Depository institutions reserves: total (adj. for rr changes)*	-	-	-	-
119	Depository institutions: nonborrowed (adj. for rr changes)*	-	-	-	-
120	Loans and securities at all comm banks: Total (mil of current \$)	3	0.21	0.37	0.63
121	Loans and securities at all comm banks: Securities, total (mil of \$)	3	0.13	0.37	0.63
122	Loans and securities at all comm banks: Securities, U.S. govt (mil of \$)	3	0.29	0.76	0.24
123	Loans and securities at all comm banks: Real estate loans (mil of \$)	3	0.25	0.08	0.92
124	Loans and securities at all comm banks: Comm and Ind loans (mil of \$)	3	0.16	0.12	0.88
125	Loans and securities comm banks: Consumer loans (mil of \$)*	-	-	-	-
126	Delinquency rate on bank-held consumer installment loans*	-	-	-	-
<i>Prices</i>					
127	PPI: Finished goods (1982 = 100 for all PPI data)	3	0.69	0.06	0.94
128	PPI: Finished consumer goods	3	0.73	0.05	0.95
129	PPI: Intermediate materials	3	0.65	0.06	0.94
130	PPI: Crude materials	3	0.51	0.05	0.95
131	PPI: Finished goods excl food	3	0.70	0.00	1.00
132	Index of sensitive materials prices*	-	-	-	-
133	CPI: All items (urban)	3	0.75	0.27	0.73
134	CPI: Food and beverages	3	0.25	0.93	0.07
135	CPI: Housing	3	0.32	0.99	0.01
136	CPI: Apparel	3	0.16	0.84	0.16
137	CPI: Transportation	3	0.76	0.02	0.98
138	CPI: Medical care	3	0.26	0.95	0.05
139	CPI: Commodities	3	0.85	0.06	0.94
140	CPI: Commodities, durables	3	0.04	0.67	0.33
141	CPI: Services	3	0.34	0.98	0.02
142	CPI: All items less food	3	0.71	0.22	0.78
143	CPI: All items less shelter	3	0.82	0.13	0.87
144	CPI: All items less medical care	3	0.76	0.26	0.74
145	CPI: All items less food and energy	3	0.41	0.96	0.04

2.A Dataset and forecast error variance decomposition

		Statistics for <i>uncertainty</i> measures			
Series	Transformation	R^2	FEVD		
			Shock 1	Shock 2	
146	Price of gold (\$/oz) on the London market (recorded in the p.m.)	3	0.25	0.71	0.29
147	PCE chain weight price index: Total	3	0.63	0.24	0.76
148	PCE prices: Total excl. food and energy	3	0.06	0.97	0.03
149	PCE prices: Durables	3	0.04	1.00	0.00
150	PCE prices: Nondurables	3	0.84	0.01	0.99
151	PCE prices: Services	3	0.09	0.99	0.01
<i>Wages</i>					
152	Avg hourly earnings: Total nonagricultura (current \$)	3	0.35	0.83	0.17
153	Avg hourly earnings: Construction (current \$)	3	0.26	0.84	0.16
154	Avg hourly earnings: Mfg (current \$)	3	0.35	0.93	0.07
155	Avg hourly earnings: Finance, insurance, and real estate (current \$)	3	0.15	0.96	0.04
156	Avg hourly earnings: Professional and business services (current \$)	3	0.07	0.44	0.56
157	Avg hourly earnings: Education and health services (current \$)	3	0.27	0.96	0.04
158	Avg hourly earnings: Other services (current \$)	3	0.24	1.00	0.00
<i>Merchandise ex- and imports</i>					
159	Total merchandise exports (FAS value) (mil of \$)	3	0.24	0.87	0.13
160	Total merchandise imports (CIF value) (mil of \$) (NSA)	3	0.39	1.00	0.00
161	Total merchandise imports (customs value) (mil of \$)	3	0.38	1.00	0.00
<i>Business outlook</i>					
162	Philadelphia Fed business outlook: General activity	2	0.07	0.28	0.72
163	Outlook: New orders	2	0.13	0.68	0.32
164	Outlook: Shipments	2	0.10	0.70	0.30
165	Outlook: Inventories	2	0.11	0.83	0.17
166	Outlook: Unfilled orders	2	0.10	0.97	0.03
167	Outlook: Prices paid	2	0.12	0.01	0.99
168	Outlook: Prices received	2	0.05	0.69	0.31
169	Outlook: Employment	2	0.03	0.95	0.05
170	Outlook: Work hours	2	0.04	0.94	0.06
171	Federal govt deficit or surplus (mil of current \$)	2	0.06	0.46	0.54

Table 2.3: Description of the dataset and forecast error variance decomposition for individual uncertainty measures Variables marked with a ‘*’ are not available for the full sample period and therefore excluded from the analysis. The transformations applied to the data and determined in the column ‘Transformation’ are explained in detail in Table 2.4. The three remaining right hand sided columns correspond to the individual uncertainty measures of the variables. The column labeled ‘ R^2 ’ reports the degree of explanation of the factor model in (2.7) and (2.8) based on two fundamental factors. Columns labeled ‘Shock 1’ and ‘Shock 2’ display forecast error variances (12-month-ahead) explained by the first or the second fundamental uncertainty shock, respectively.

Transformations applied to the data	
0	X_t
1	$\ln(X_t)$
2	$(1 - L)X_t$, L denotes the lag-operator
3	$(1 - L)\ln(X_t)$

Table 2.4: Data transformations

Remark 1: Whenever a series has not been available in NAICS classification for the entire sample period, missing values have been linked with data based on the SIC classification.

Remark 2: Series 32 has been published only until 2010M7. It has been linked with the Help Wanted Online Index published by the Conference Board.

Remark 3: Whenever a series denoted in mil. of chained 2005 dollar has not been available for the entire sample period, missing values have been linked with data published in mil. of chained 1996 dollar.

Remark 4: Series 116 has been replaced by the monetary aggregates index published by the Federal Reserve Bank of St. Louis (monetary aggregate (all), sum, comparable to old index M3).

2.B Estimation of the dynamic factor model

Doz et al. (2012) introduce a quasi maximum likelihood estimation procedure for the so-called ‘approximate dynamic factor model’ where the idiosyncratic components ξ_t in (2.7) are allowed to be weakly cross sectionally correlated (e.g., Chamberlain and Rothschild; 1983) and also weakly serially correlated. Doz et al. (2012) show that the factors are estimated consistently by means of the quasi maximum likelihood approach if N and T are large even if cross sectional and serial correlation of the idiosyncratic components is omitted. The estimation procedure is based on the EM-algorithm which we initialize using an estimate of \hat{f}_t obtained from principal components. Given this initialization we iterate between the following steps.

M-step:

- Obtain the loadings matrix $\hat{\lambda}$ by regressing X_t on \hat{f}_t .
- Estimate the idiosyncratic components from $\hat{\xi}_t = X_t - \hat{\lambda}\hat{f}_t$. Further, evaluate the corresponding covariance matrix of the idiosyncratic components denoted by \hat{S}_ξ . Restrict the off-diagonal elements to zero as proposed by Doz et al. (2012).
- Estimate the h -th order VAR in f_t using \hat{f}_t .
- Estimate the covariance of the residuals \hat{S}_ϵ .

E-step:

- ξ_t and ϵ_t are both assumed to be normally distributed and independent from each other.
- Apply the Kalman filter given the parameter estimates from the M-step and update the estimated factors \hat{f}_t .
- The likelihood function is evaluated by means of the Kalman filter.

Iterate forward between the E-step and the M-step until convergence is reached.

We apply an eigenvalue decomposition to the final estimate of the covariance matrix $\hat{\mathcal{S}}_\epsilon$. Let Q be a matrix containing the eigenvalues on the main diagonal and zeros elsewhere. The corresponding eigenvectors are stored in the matrix V and R is given by

$$\hat{R} = MV^{\frac{1}{2}}, \quad (2.13)$$

and

$$\hat{u}_t = \hat{\epsilon}_t MV^{\frac{1}{2}}. \quad (2.14)$$

2.C Tests for the number of factors

In Table 2.5, we summarize the number of common uncertainty factors k indicated by distinct information criteria. Conditional on our previous results, it appears reasonable to fix the maximum number of common factors to $k_{max} = 4$. The criteria of Bai and Ng (2002) suggest $k = 4$ (IC_{p1} , IC_{p2} , IC_{p3}). These criteria overestimate the number of factors when idiosyncratic components are correlated which is likely for our dataset since it contains subcategories of a number of variables (see also Onatski; 2010; Jurado et al.; 2014). Bai and Ng (2008) suggest the criterion IC_{BN08} which works relatively well when idiosyncratic components are cross correlated. According to IC_{BN08} there are three factors. The test of Onatski (2010, IC_{ON10}) and the criteria of Ahn and Horenstein (2013, ER , GR) uniformly indicate one common factor although our previous analysis suggests that one factor does not represent the dataset in a satisfactory manner. While the former criterion is more appropriate when idiosyncratic components feature autocorrelation, the latter two are less sensitive to the choice of k_{max} .

IC_{p1}	IC_{p2}	IC_{p3}	IC_{BN08}	IC_{ON10}	ER	GR
4	4	4	3	1	1	1

Table 2.5: Tests to determine the number of common factors This table displays the number of fundamental factors determined by means of the information criteria proposed by Bai and Ng (2002, $IC_{p\bullet}$) and Bai and Ng (2008, IC_{BN08}). Results are conditioned on the maximum number of factors $k_{max} = 4$. The column labeled IC_{ON10} reports the number of factors indicated by the test procedure proposed by Onatski (2010), while columns ER and GR document the results of the procedures described in Ahn and Horenstein (2013).

Moreover, we employ tests to determine the number of fundamental shocks u_t in a general dynamic factor model where this number is allowed to differ from the quantity of factors f_t . In model (2.7) and (2.8), we assume that the quantities of fundamental shocks and factors are identical (e.g., Kose et al.; 2003). In the more general dynamic factor model, tests for the number of fundamental shocks depend on a prespecified quantity of factors. Table 2.6 reveals that the criterion proposed in Bai and Ng (2007) points at two to three fundamental shocks while the procedure of Hallin and Liška (2007) indicates up to two fundamental shocks conditional on distinct numbers of factors. Notably, the results are largely independent from this condition, i.e., adding more common factors to the model hardly affects the estimated number of fundamental shocks.

Prespecified number of factors	2	4	6	8	10	12	14
Number of fundamental shocks							
Bai and Ng (2007)	2	2	3	3	3	2	2
Hallin and Liška (2007)	1	2	2	2	2	2	2

Table 2.6: Tests to determine the number of fundamental shocks The upper panel reports the number of fundamental shocks determined by means of the criterion of Bai and Ng (2007), the lower panel displays the results for the criterion of Hallin and Liška (2007).

2.D Bootstrap procedure for impulse responses of the dynamic factor model

We employ a bootstrap procedure similar to that proposed by Giannone, Reichlin and Sala (2002) and Forni, Giannone, Lippi and Reichlin (2007) to compute confidence bounds for responses of individual uncertainty measures to fundamental shocks u_t (Figures 2.3 and 2.4). We make use of the estimates for the dynamic factor model in (2.7) and (2.8) – $\hat{\lambda}$, $\hat{\xi}_t$, $\hat{\psi}(L)$ and $\hat{\epsilon}_t$.

- To simulate the factors, denoted by \tilde{f}_t , we apply the residual based bootstrap and draw with replacement from centered residuals of the VAR in the factors $\hat{\epsilon}_t - \bar{\epsilon}$, where $\bar{\epsilon}$ is a vector of sample averages.
- The estimated parameters in $\hat{\psi}(L)$ are employed to simulate a new dataset from $\tilde{X}_t = \hat{\lambda}\tilde{f}_t + \hat{\xi}_t$. We do not permute $\hat{\xi}_t$ because the major part of estimation uncertainty is related to the estimation of the VAR in the factors (compare Giannone et al.; 2002, on this point).
- Use \tilde{X}_t and reestimate the factor model in (2.7) and (2.8) to obtain a new set of parameters and impulse responses.

We obtain 500 draws of each impulse response function by repeating of this bootstrap procedure.

2.E Identification of the indicators of macroeconomic uncertainty

In the following, we present the rotation strategy that we employ to separately identify the two factors. The rotation matrix G is used to rewrite the factor model in (2.7) such that

$$X_t = \lambda^* G G' f_t^* + \xi_t, \quad (2.15)$$

with $G G' = I$ and where λ^* and f_t^* represent arbitrary rotated quantities. To derive the covariance matrix of the factors, we rewrite the dynamic process of the factors as first order VAR

$$F_t^* = \mathcal{A} F_{t-1}^* + U_t, \quad (2.16)$$

where $F_t^* = (f_t^{*'}, f_{t-1}^{*'}, \dots, f_{t-h+1}^{*'})'$ denotes a vector containing the stacked factors. It follows that the parameter matrix \mathcal{A} and the residual vector U_t have the usual companion form. The covariance matrix of U_t is denoted by Ω . The vectorized covariance matrix \mathcal{S}_F^* of the stacked factors F_t^* can be expressed as

$$vec(\mathcal{S}_F^*) = (I_{(kh)^2} - [\mathcal{A} \otimes \mathcal{A}])^{-1} vec(\Omega), \quad (2.17)$$

where the covariance matrix \mathcal{S}_f^* of the factors f_t^* corresponds to the upper left $k \times k$ sub matrix of \mathcal{S}_F^* . In Section 2.4.3, we consider the rotated factors $f_t = G' f_t^*$. The corresponding loadings are $\lambda = \lambda^* G$. The covariance matrix of the rotated factors is defined as

$$\mathcal{S}_f = G' \mathcal{S}_f^* G. \quad (2.18)$$

Further, we introduce the matrix S_{f_1} (S_{f_2}) which is a zero matrix except for the upper left (lower right) hand side element which equals the corresponding element of \mathcal{S}_f . The i -th diagonal element of the covariance matrix of the idiosyncratic processes \mathcal{S}_ξ is denoted as $\mathcal{S}_{\xi,i}$ and the i -th row of loadings matrix λ^* as λ_i^* . To determine the rotation matrix G we maximize the target function

$$k(G) = \frac{\sum_{i \in D} \lambda_i^* G \mathcal{S}_{f_1} G' \lambda_i^{*'}}{\sum_{i \in D} (\lambda_i^* G \mathcal{S}_{f_1} G' \lambda_i^{*'} + \lambda_i^* G \mathcal{S}_{f_2} G' \lambda_i^{*'} + \mathcal{S}_{\xi,i})}. \quad (2.19)$$

This rotation procedure maximizes the variance proportions of individual uncertainty measures selected in D which are explained by the first factor (see also Kose et al.; 2012). We chose D such that it selects the uncertainty measures associated with output (series 1 to 31). Since the factors are not restricted to be orthogonal, the denominator in (2.19) consists of the sum of the variance proportions individually explained by the two factors which can deviate from the jointly explained variance proportion $\lambda_i^* G \mathcal{S}_f G' \lambda_i^{*'}$.

2.F Comparison of macroeconomic uncertainty with alternative uncertainty measures

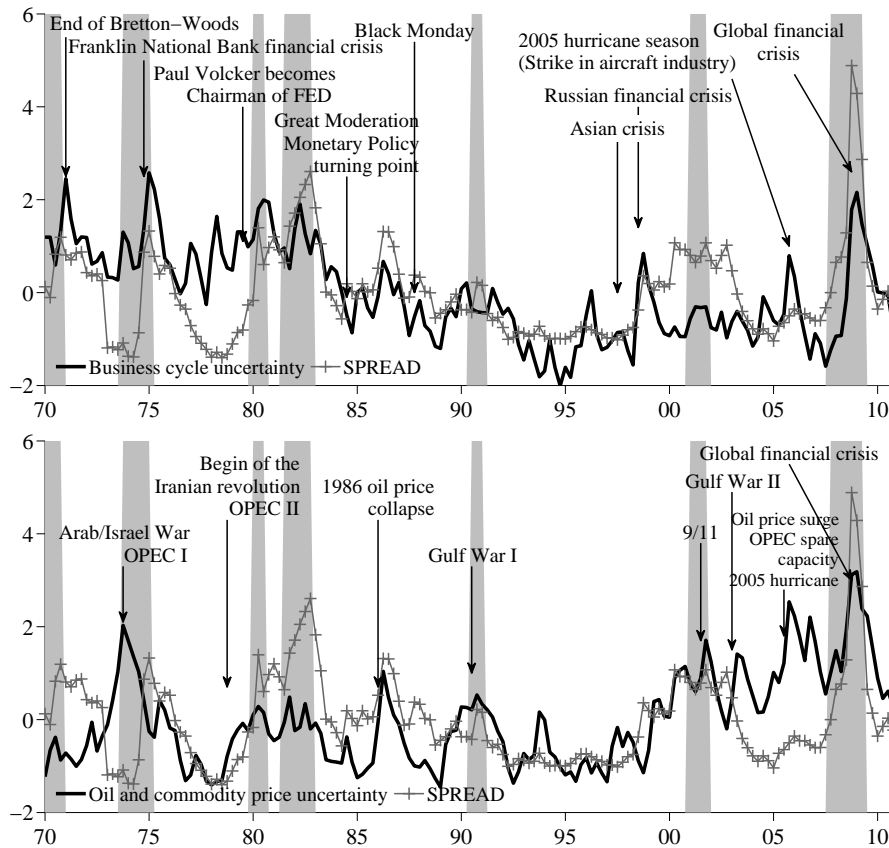


Figure 2.9: Indicators of macroeconomic uncertainty and SPREAD The upper panel depicts business cycle uncertainty and the lower panel oil and commodity price uncertainty. The crossed gray line represents the difference between the 30-year Baa-corporate bond yield and the 30-year Treasury bond yield (SPREAD). To facilitate the interpretation of the graph, we calculate quarterly averages of the monthly observations and standardize all series. NBER recession dates are indicated by the shaded areas.

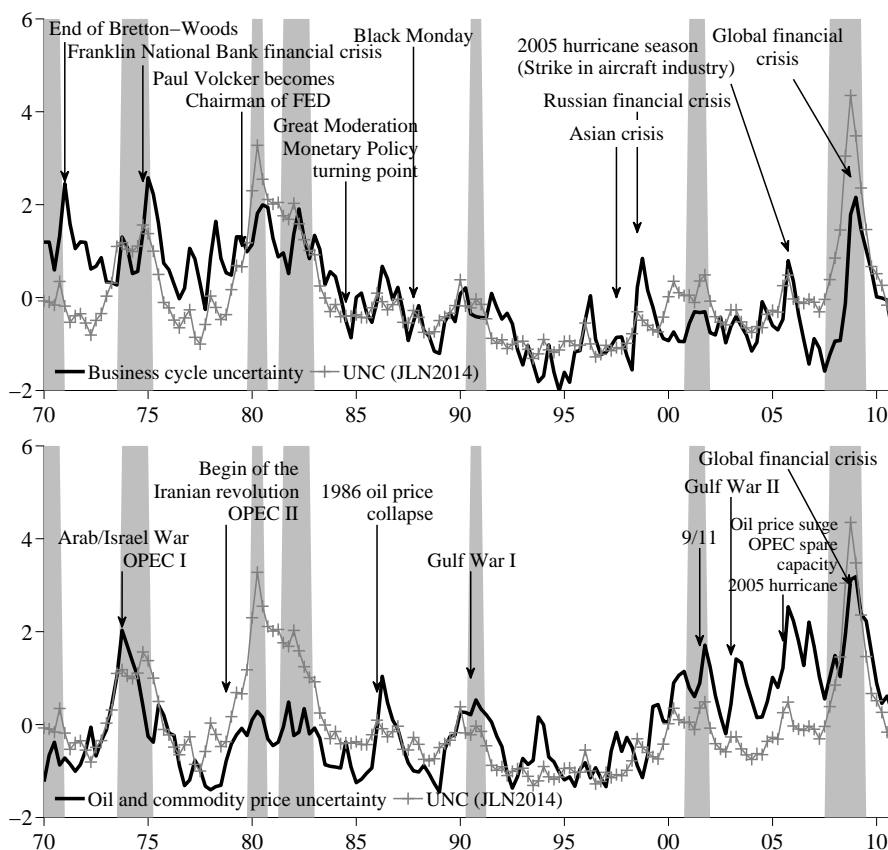


Figure 2.10: Indicators of macroeconomic uncertainty and the JLN measures The upper panel depicts business cycle uncertainty and the lower panel oil and commodity price uncertainty. The crossed gray line represents the measure of macroeconomic uncertainty proposed by Jurado et al. (2014). This measure is the factor with the largest explanatory power for in comparable dataset of individual uncertainty measures. To facilitate the interpretation of the graph, we calculate quarterly averages of the monthly observations and standardize all series. NBER recession dates are indicated by the shaded areas.

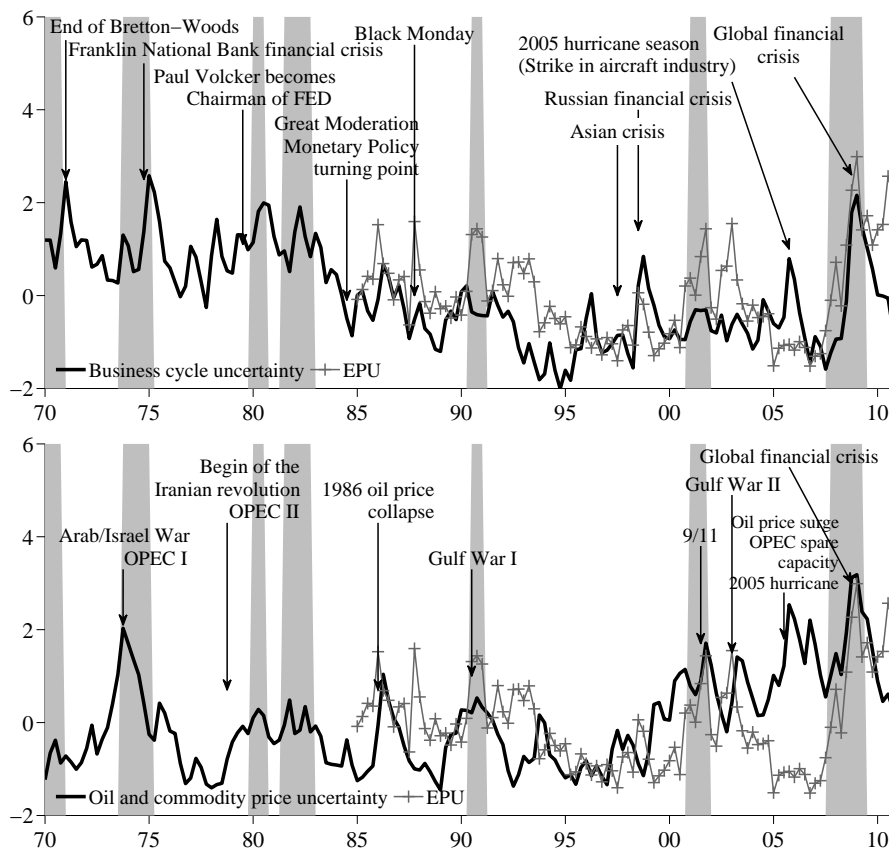


Figure 2.11: Indicators of macroeconomic uncertainty EPU The upper panel depicts the business cycle uncertainty and the lower panel the oil and commodity price uncertainty. The crossed gray line represents the measures of economic policy uncertainty (EPU) proposed by Baker et al. (2013). To facilitate the interpretation of the graph, we calculate quarterly averages of the monthly observations and standardize all series. NBER recession dates are indicated by the shaded areas.

2.G Dynamic relation between fundamental shocks u_t and VXO, SPREAD, JLN and EPU

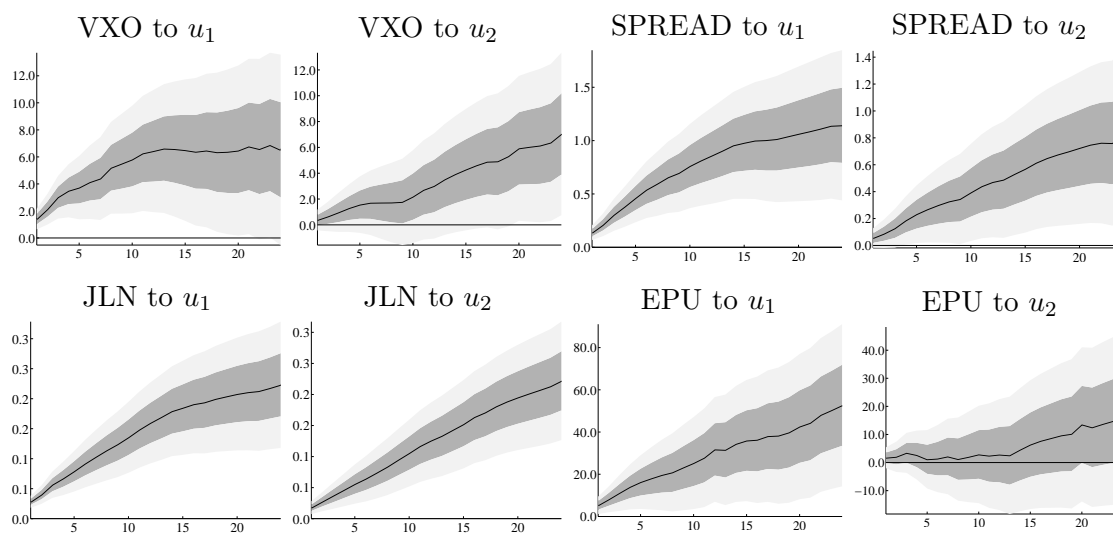


Figure 2.12: Dynamic relation between fundamental shocks u_t and VXO, SPREAD, JLN and EPU The upper panel depicts the response of the VXO and the SPREAD to a surprise innovation in business cycle uncertainty $u_{1,t}$ and oil and commodity price uncertainty $u_{2,t}$. The lower panel displays the responses for the JLN measure and the EPU to the same shocks. The response is calculated from a regression of the respective uncertainty measure on 24 lags of the two shock series $u_{1,t}$ and $u_{2,t}$. Confidence intervals are derived from a block bootstrap using block size 10 and 20,000 replications. The dark shaded area indicates the \pm one standard deviation confidence interval while light gray represents the \pm two standard deviations confidence interval.

2.H Robustness analyses

2.H.1 Alternative rotation of fundamental shocks

In Section 2.4.2 we identify the fundamental shocks u_t by means of a rotation procedure such that the first fundamental shock explains most of the forecast error variance of individual uncertainty measures associated with output variables. In contrast to this baseline rotation, we now maximize the variance share explained by shock 2 for variables representing oil and commodity price uncertainty. To be specific, D in the target function (2.12) selects individual uncertainty measures associated with the following variables: total energy production (series 16), PPI of crude materials (series 130), CPI of commodities (series 139) and CPI of durable commodities (series 140). In the following we compare the impulse response functions obtained from this alternative rotation with those from the baseline rotation. Since the impulse responses are almost identical the interpretation of the two fundamental shocks is robust to the change of the identification strategy.

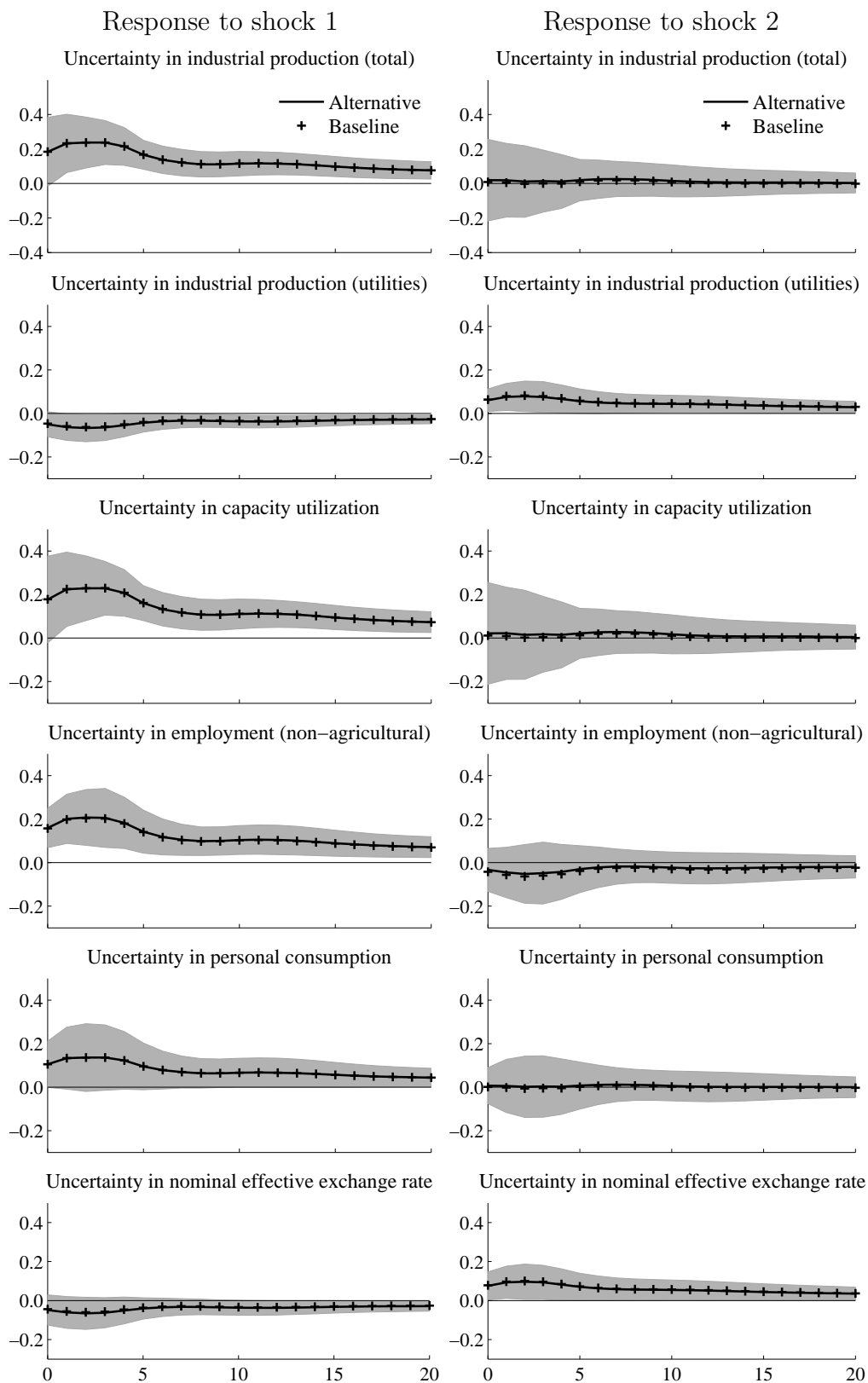


Figure 2.13: Impulse responses of individual uncertainty measures for an alternative identification strategy (1) Impulse responses to the first fundamental shock (left column) and to the second fundamental shock (right column) identified with the alternative rotation are indicated by the bold lines. Impulse responses from the baseline rotation are marked with crosses. Shaded areas indicate bootstrapped 90% confidence intervals.

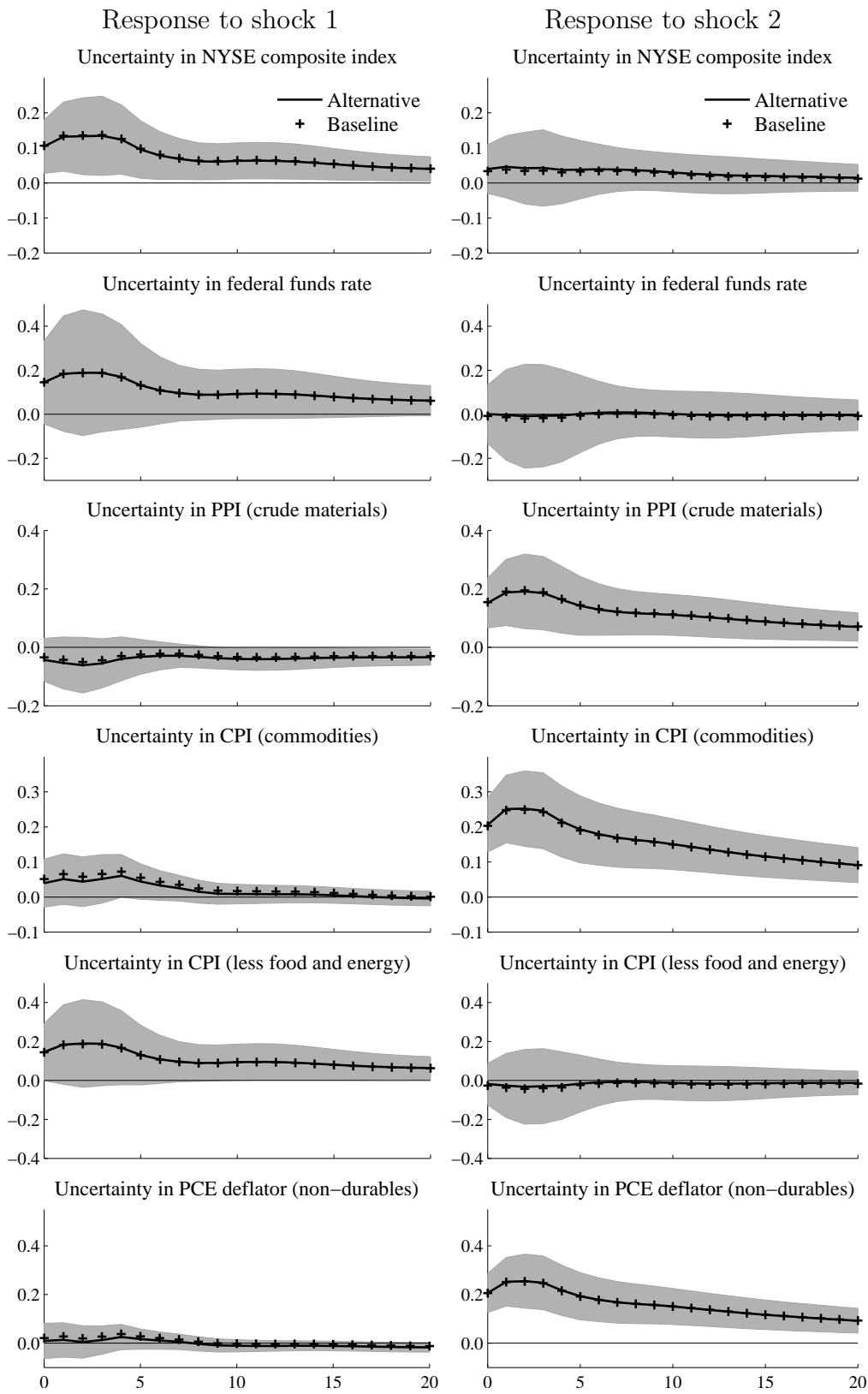


Figure 2.14: Impulse responses of individual uncertainty measures for an alternative identification strategy (2) For additional notes see Table 2.13.

2.H.2 Indicators of macroeconomic uncertainty derived from out-of-sample forecast errors

The indicators of macroeconomic uncertainty discussed in Section 2.4.3 are estimated from in-sample forecast errors of (2.1). Likewise, the predictive factors employed in this equation are estimated from full sample information. In this section we consider indicators of macroeconomic uncertainty that are conditioned on out-of-sample forecast errors. We estimate the predictive factors using recursive windows initialized with observations from the first five years. Out-of-sample forecast errors are generated by recursive estimation of (2.1). Individual uncertainty measures and the indicators of macroeconomic uncertainty are evaluated as described in Section 2.3.1 and 2.3.2. Figure 2.15 contrasts the in-sample and out-of-sample based indicators. Even though the precision of out-of-sample forecasts for early periods might suffer from small sample information the indicators co-move closely. The correlations are 0.95 and 0.90, respectively.

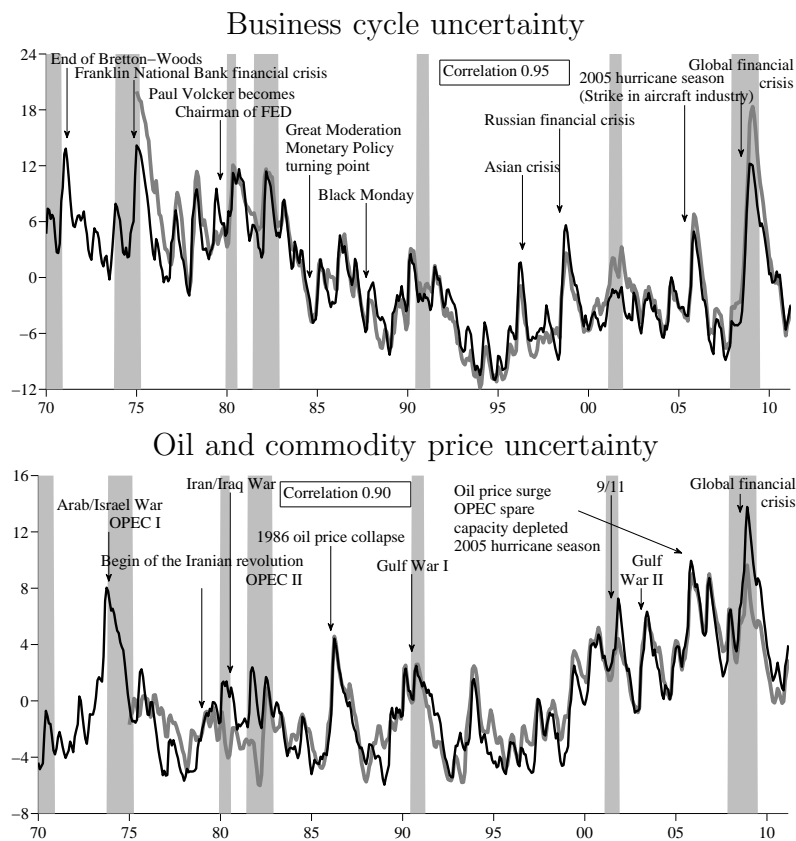


Figure 2.15: Indicators of macroeconomic uncertainty from in-sample and out-of-sample forecast errors. The upper panel depicts business cycle uncertainty and the lower panel depicts oil and commodity price uncertainty. NBER recession dates are indicated by the shaded areas. In both panels the black line represents the baseline estimates while the bold gray line represents estimates based on out-of-sample forecast errors.

2.H.3 The VAR if a three-factor model is considered

We review the robustness of the VAR impulse responses for industrial production using uncertainty indicators originating from the factor model in (2.7) and (2.8) with $k = 3$ common factors.

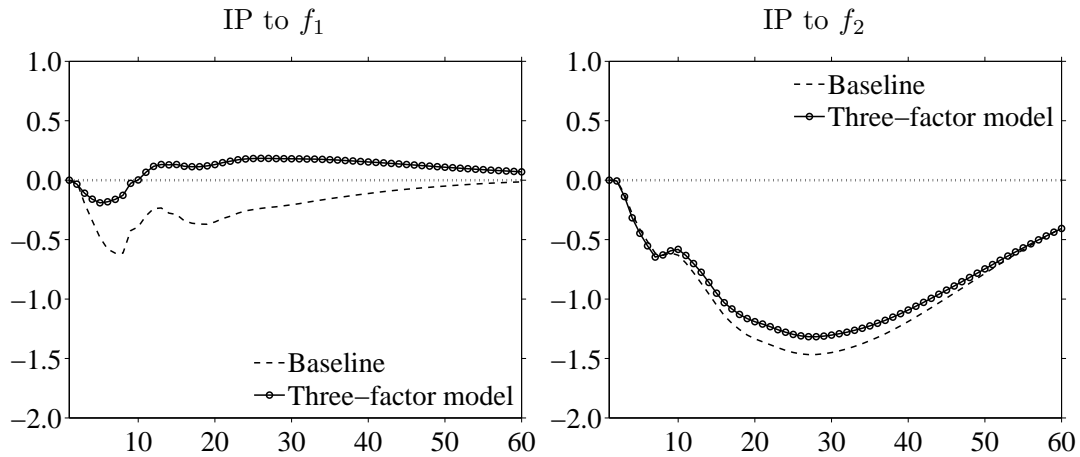


Figure 2.16: Impulse response functions of IP using uncertainty indicators from a three-factor model This figure compares the baseline VAR to a VAR where uncertainty indicators are derived from a three factor model. The left panel displays the response of the IP following a sudden increase in $f_{1,t}$, and the right panel depicts the results for a sudden increase in $f_{2,t}$. The shock size is chosen to obtain an initial increase of macroeconomic uncertainty equal to its observed average monthly increase in the second half of 2008.

2.H.4 Impulse response functions from alternative VAR specifications

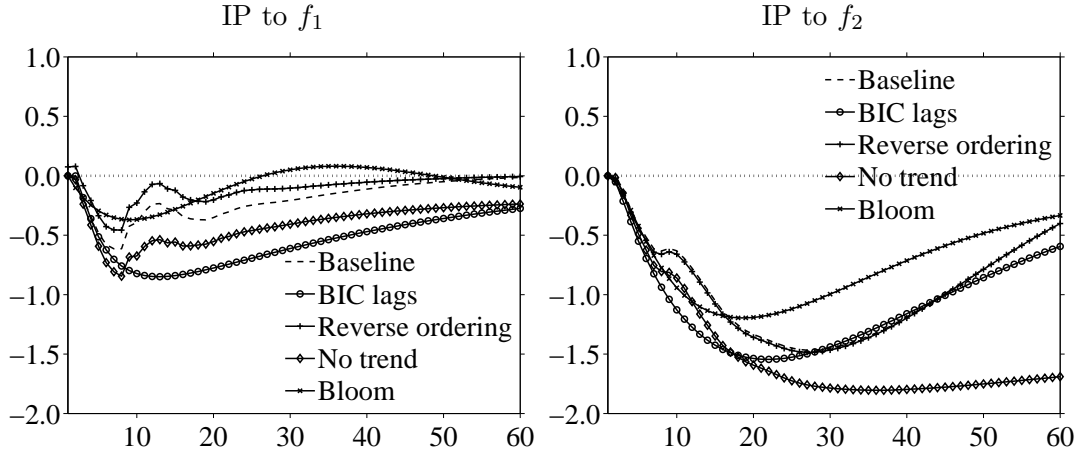


Figure 2.17: Impulse response functions of IP for distinct VAR specifications
 The left hand side panel displays the response of the IP following a sudden increase in business cycle uncertainty for distinct VAR systems. The right hand side panel displays the corresponding response of the IP to a sudden increase in oil and commodity price uncertainty. Along with our baseline results, a number of robustness checks are shown. First, we choose the lag-length by means of the BIC (BIC lags). Second, we revert the ordering of the variables in the VAR (Reverse ordering). Third, we remove the linear trend from the specification (No trend). Fourth, we consider a eight-variable VAR as estimated by Bloom (2009) which includes IP, E, Hours, CPI, hourly wages, the federal funds rate, $f_{\bullet,t}$ and the S&P500 index (Bloom). All variables except for the federal funds rate enter in logs. The shock size is chosen to obtain an initial increase of macroeconomic uncertainty equal to its observed average monthly increase in the second half of 2008.

2.H.5 Both types of macroeconomic uncertainty in the VAR

In Section 2.4.4, the VARs contain only one type of macroeconomic uncertainty, i.e., we do not control for the effects of the other type during the estimation. Table 2.7 provides a forecast error variance decomposition for VARs which contain both uncertainty indicators simultaneously. Since theory does not provide guidance for the ordering of the uncertainty indicators in such a VAR, we present results for two distinct orderings which both rely on the assumption that real activity reacts with a lag to surprise increases in macroeconomic uncertainty. The first ordering assumes that business cycle uncertainty reacts with a lag to oil and commodity price uncertainty while business cycle uncertainty affects oil and commodity price uncertainty instantaneously, i.e., we control for innovations in business cycle uncertainty when analyzing the effects of oil and commodity price uncertainty on real activity. Those results are given in the left hand side panel of Table 2.7. The second ordering

presented in the right hand side panel of Table 2.7 reverses the ordering of the two types of uncertainty, i.e., we control for innovations in oil and commodity price uncertainty when analyzing the effects of business cycle uncertainty. Table 2.7 reveals that surprise increases in oil and commodity price uncertainty are more important for real activity than those to business cycle uncertainty even if we use this larger VAR.

Horizon	Ordering: $(Y_{\bullet}, f_1, f_2)'$						Ordering: $(Y_{\bullet}, f_2, f_1)'$					
	IP		CU		Hours		IP		CU		Hours	
	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2	f_1	f_2
12	1.42	3.65	2.92	2.47	1.16	6.47	0.52	4.54	1.46	3.93	0.37	7.27
24	0.96	13.79	5.99	11.12	3.06	17.39	0.44	14.30	2.04	15.07	0.44	20.01
36	1.00	21.81	9.24	16.42	4.72	22.03	0.53	22.28	3.16	22.50	0.79	25.96
60	1.28	26.28	10.26	16.42	5.28	22.36	0.52	27.05	3.93	22.75	1.08	26.56

Table 2.7: Forecast error variance in a VAR comprising both types of uncertainty

This table presents the share of the forecast error variance (in %) that is explained by innovations in business cycle uncertainty (f_1) and oil and commodity price uncertainty (f_2). The shares are obtained from trivariate VARs with 12 lags consisting of one measure of economic activity and both types of macroeconomic uncertainty. Shocks are orthogonalized using a recursive ordering.

2.H.6 FEVD for the VAR specification proposed by Bloom (2009)

Horizon	Business cycle uncertainty			Oil and commodity price uncertainty		
	IP	Employment	Hours	IP	Employment	Hours
12	2.72	4.94	5.79	10.31	7.00	5.02
24	1.57	5.74	6.27	15.90	13.87	6.39
36	1.05	5.35	6.04	16.33	14.99	6.05
60	0.83	6.21	6.01	15.91	14.23	6.19

Table 2.8: FEVD using the VAR specified by Bloom (2009) This table presents the share of the forecast error variance (in %) that is explained by business cycle uncertainty (left hand side panel) and oil and commodity price uncertainty (right hand side panel). The shares are obtained from VARs including eight variables as in Bloom (2009) with lag length selected by means of the BIC. The ordering of the variables is: IP, employment, Hours, CPI, hourly wages, federal funds rate, macroeconomic uncertainty, S&P500 index. All variables except for the interest rate enter in log-levels. For reasons of comparability, results are documented only for variables describing the real activity.

Paper 3

Size corrected inference in fiscal policy reaction functions: A three country assessment*

Helmut Herwartz and Malte Rengel

Abstract. In this paper we reevaluate the test for the so-called model based sustainability of fiscal policies in the US, the UK and Portugal. As a reflection of both strong persistence of debt-to-GDP ratios and correlation of respective innovations with governments' primary surpluses, conventional t -tests in policy reaction functions show actual significance levels that are up to five times larger than their nominal reference. Adopting size controlled inference by means of Monte Carlo based and asymptotic Bonferroni critical values, US and UK fiscal policies are diagnosed as sustainable in samples covering more than 100 years. Conditioning on post-WWII subsamples and 5% nominal significance, conventional t -tests signal fiscal sustainability for these countries. In sharp contrast, size corrected inference hints at a lack of fiscal sustainability and, thus, is recommended for the 'real time' monitoring of public debt. The fiscal policy of Portugal is found to lack sustainability irrespective of the considered sample period.

3.1 Introduction

The investigation of the sustainability of fiscal policy is of central interest for economists and policy makers since decades. In the guidelines for public debt management of the International Monetary Fund and the World Bank it is argued that “*in a broader macroeconomic context for public policy, governments should seek to ensure that both the level and rate of growth in their public debt is fundamentally sustainable, and can be serviced under a wide range of circumstances while meeting cost and risk objectives*” (IMF and WB; 2001). In recent years public debt has accumulated to rather high levels especially in European countries. For instance, the Portuguese debt-to-GDP ratio nearly doubled from 0.7 in 2007 to 1.2 in 2012. In managing high levels of public debt reliable inference on sustainability of fiscal balances becomes indispensable for policy authorities. To infer if fiscal balances are

sustainable economists developed several approaches, which are natural candidates to extract indications of solvency risks.

In this paper we focus on evaluating the sustainability of fiscal policies in the US, the UK and Portugal conditional on historical time series covering more than a century. We diagnose that conventional statistical inference on so-called model based sustainability may suffer from marked size distortions that bear severe risks of rejecting the null-hypothesis of non-sustainability too often. Against this background, we advocate two procedures of size corrected inference to assess the sustainability of public debt in a reliable manner.

In the economic literature one may distinguish, in principle, two approaches to test for the sustainability of fiscal balances (see, Bohn; 2008, for an overview). The first divides itself into two attempts exploring if the fiscal policy is in accord with an intertemporal budget constraint (IBC). Either it is tested against a unit root in debt (e.g., Hamilton and Flavin; 1986; Kremers; 1988; Wilcox; 1989), or for a cointegration relation between government revenues and expenditures including interest payments (e.g., Trehan and Walsh; 1988; Ahmed and Rogers; 1995; Quintos; 1995). These quantities are commonly measured in real terms or as shares of GDP to account for growth effects. The second approach proposed by Bohn (1995, 1998) focuses on an alternative IBC which is derived from a stochastic general equilibrium model.¹ The model implied IBC differs from the deterministic IBC by accounting for agents' optimizing behavior under uncertainty.²

As a particular merit of this approach model based sustainability can be tested by means of a stylized policy reaction function (Bohn; 1998). For a sustainable fiscal policy an increasing debt-to-GDP ratio must invoke a positive reaction of the primary (non-interest) budget surplus measured as a share of GDP. Bohn (2007, 2008) argues that stationarity and cointegration conditions are sufficient but not necessary for the sustainability of fiscal policy. In consequence, such conditions are overly strong to assess fiscal balances' sustainability, and the evaluation of policy reaction functions might be seen as a more suited framework.³

Many studies applying unit root and cointegration tests to assess the sustain-

¹ This model draws on an endowment economy described by Lucas (1978).

² It is worthwhile to notice that the quoted literature focuses on long-run sustainability implied by IBCs. Alternatively, a related strand of the literature considers the so-called short-term sustainability (e.g., Abeyasinghe and Jayawickrama; 2013).

³ Quintos (1995) supports the conclusion that cointegration conditions are sufficient but not necessary, and introduces a so-called 'weak' condition for deficit sustainability that accords with the deterministic IBC. This condition requires that the with-interest deficit has to be difference stationary.

ability of fiscal balances arrive at ambiguous conclusions, and argue in favor of at least ‘near’ to unit root processes of debt-to-GDP ratios.⁴ For instance, Antonini, Lee and Pires (2013) provide a detailed description of the persistent movements of debt-to-GDP ratios of 10 EU countries. Generally, the diagnosis of sustainability vs. non-sustainability depends on test specifications as, for instance, the choice of respective lag lengths (Greiner and Semmler; 1999). In contrast, applying policy reaction functions is more often in favor of the IBC. Bohn (1998, 2008) supports sustainability for the US. Greiner, Köller and Semmler (2006) find a significantly positive reaction of the policy parameter for the case of Germany. Focusing on US, German, French, Italian and Portuguese fiscal figures starting in the 60s and ending in 2003, Greiner, Köller and Semmler (2007) diagnose sustainability for all these countries by means of single country regressions. Applying a panel approach, Mendoza and Ostry (2008) argue in favor of sustainability for a set of emerging and industrial countries. For a panel of 23 advanced economies Ghosh et al. (2013) provide evidence for a positive reaction parameter which starts to decrease if the debt-to-GDP ratio approaches unity. Lukkezen and Rojas-Romagosa (2012) diagnose a positive and significant response of primary surpluses to increases in debt-to-GDP ratios for the US, the UK, the Netherlands and Belgium.

The significance of policy response coefficients is commonly evaluated by means of standard t -tests or robust versions applying general covariance estimates (e.g., Newey and West; 1987). However, debt-to-GDP ratios of many countries may be best described by (‘near’ to) unit root processes. In contrast, primary surpluses measured as shares of GDP are typically found to be stationary (Bohn; 1998, 2008; Marinheiro; 2006; Lukkezen and Rojas-Romagosa; 2012). Moreover, Woodford (1998) argues in the context of the ‘fiscal theory of the price level’ that innovations of debt-to-GDP ratios are likely to be correlated with primary surpluses. At a first glance, a poorly balanced test regression on the one hand, and the correlation of debt-to-GDP innovations and primary surpluses on the other hand might exacerbate reliable inference on movements of primary surpluses governed by accelerations of debt-to-GDP ratios. However, non-standard econometric treatment of the policy rule allows for size controlled inference on model based sustainability in finite samples.

⁴ Among many others, Hamilton and Flavin (1986), Trehan and Walsh (1988), Wilcox (1989) and Hakkio and Rush (1991) analyze US fiscal figures. Ahmed and Rogers (1995) and MacDonald and Speight (1990) analyze the sustainability of the public debt of the UK and Greiner and Semmler (1999) that of Germany. Marinheiro (2006) focuses on Portuguese observations while Afonso (2005) considers data of 15 EU countries. The reader may consult Afonso (2005) for a more detailed overview of empirical studies on sustainability of fiscal policy employing unit root and cointegration tests.

Elliott and Stock (1994) argue that strong persistence of regressors joint with correlation of their innovations with a mean reverting dependent variable invoke marked size distortions of t -tests in finite samples. Moreover, Stambaugh (1999) demonstrates that estimated slope coefficients are biased.⁵ Such characteristics have been observed in distinct fields of economic research. For instance, analyzing the permanent income hypothesis Mankiw and Shapiro (1985, 1986) hint at slow convergence of the t -ratio to the common asymptotic distribution. Studying the dynamics of the US external position, news of future returns and trade flows, Evans and Fuertes (2010) mention potential size distortions of respective t -tests. Especially, the finance literature explicitly accounts in so-called predictive regressions for these characteristics, when relating equity returns to persistent valuation ratios as, for instance, the price-to-dividend ratio (e.g., Stambaugh; 1999; Campbell and Yogo; 2006; Goyal and Welch; 2008; Lettau and Van Nieuwerburgh; 2008; Kim and Park; 2013). In the model based approach to test for the IBC, eventual biases of estimated slope coefficients and size distortions of the t -test have not been considered yet. Thus, it is unclear if the more supportive evidence for sustainability gathered from policy reaction regressions reflects weaker necessary conditions subjected to testing, or if it is due to spurious significance under size distortions of conventional diagnostics. Since such distortions bear the risk to reject the null hypothesis of non-sustainability more frequently as suggested by the nominal significance level, spuriously diagnosed fiscal soundness may postpone the adoption of suitable fiscal measures to re-install sustainability of public debt.

We test for sustainability of the US, the UK and Portuguese public debt at the common 5% nominal significance level. Fiscal data for these countries cover more than a century. To account for heterogeneous data quality and the huge interest in recent accelerations of public debt-to-GDP ratios, we consider additionally post-1900 and post-WWII subsamples. We provide evidence that policy reaction coefficients estimated for the US, the UK and Portugal are upward biased and corresponding t -tests suffer from considerable size distortions. Coping with such size distortions of up to 20%, we apply Monte Carlo critical values and Bonferroni confidence intervals to infer on model based sustainability. These approaches have been fruitfully applied in predictive regressions by Goyal and Welch (2008) and Campbell and Yogo (2006). As it turns out, after size adjustment the US and UK

⁵ As an illustration consider the simple regression $y_t = \theta_0 + \theta_1 z_{t-1} + \epsilon_t$, where the regressor follows a persistent process $z_t = \gamma_0 + \gamma_1 z_{t-1} + \nu_t$ with γ_1 close to unity. The stronger the (negative) correlation of innovations ϵ_t with ν_t and the stronger the persistence of z_t the larger is the (upward) bias of $\hat{\theta}_1$ and the size distortion of a t -test for this parameter in finite samples.

fiscal policies accord with the IBC except for the post-WWII period. Several post-WWII subsamples are contemplated to assess if the diagnosed non-sustainability of US and UK post-WWII fiscal balances mirrors expansive fiscal policies during the recent Great Recession, and if oversized t -tests prevent a timely detection of potential imbalances. For a number of post-WWII subsamples ending in periods of fiscal expansion around 2008, we diagnose US and UK fiscal policies to be sustainable by means of (biased) Gaussian critical values. Owing to size distortions this diagnosis turns out to be spurious, and factually rests on an unknown or uncontrolled significance level which could markedly exceed the nominal 5% benchmark. In contrast, at the common 5% level size corrected inference signals a lack of sustainability. Thus, size corrected inference is of core relevance for a timely policy monitoring. Fiscal policy of Portugal is indicated to be non-sustainable throughout.

The US, the UK and Portuguese data are introduced subsequently followed by a sketch of the policy reaction function in Section 3.2. There we also discuss unit root characteristics and consider size distortions of t -tests regarding the policy response. Size controlled inference is sketched in Section 3.3 and applied to US, UK and Portuguese data in Section 3.4. Section 3.5 concludes. If not mentioned otherwise the empirical discussion of fiscal sustainability refers to the 5% nominal significance level.

3.2 Size distortions in policy reaction functions

Unit root and cointegration diagnostics accurately control the probability of falsely confirming the IBC, however, both approaches have been argued to target at overly strong sufficient conditions for the IBC. While policy reaction regressions address more directly weaker necessary IBC conditions, they may suffer from excess type I error probabilities. In this section, we first introduce the data and state the policy reaction function. Then, we discuss the time series properties of involved variables. In light of a weakly balanced policy response regression, the magnitude of eventual size distortions of conventional t -ratios are uncovered by means of a Monte Carlo approach.

3.2.1 Data

We investigate annual fiscal data for the US, the UK and Portugal over periods from 1792-2011, 1830-2012 and 1852-2012, respectively. US data have been widely analyzed in the literature. In light of the current European debt crisis, we select

Portugal and the UK to include countries with currently high and medium debt-to-GDP ratios. Long-term time series are available for all three countries. Each dataset consists of a series of debt, revenues, expenditures, interest payments and the nominal and real GDP. For the US and the UK data also comprise military expenditures. A detailed description of respective data sources is provided in Appendix 3.A.⁶ The series of debt-to-GDP ratios and primary surpluses for the US and the UK are depicted in Figure 3.1.

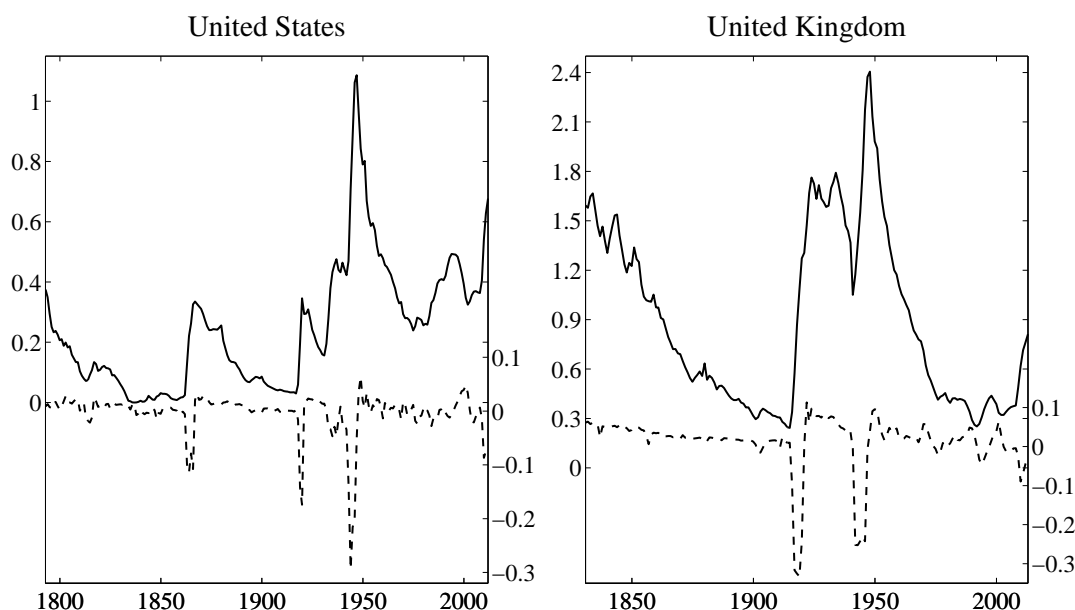


Figure 3.1: Fiscal data for the US and the UK Both panels display respectively a debt-to-GDP ratio (solid line and left hand side axis) along with the corresponding primary surplus relative to GDP (dashed lines and right hand sided axis).

The debt-to-GDP ratio of the UK decreases during the 19th century while US figures decline after the Civil War from 1861-1865 (Bohn; 2008). Both countries show an outstanding acceleration of debt-to-GDP ratios during the two World Wars. The debt-to-GDP series of Portugal is displayed in Figure 3.2 and characterized by relatively low quotes after the military coup in 1926. Since the mid of the 1970s and especially since the circulation of the Euro banknotes in 2002 the Portuguese debt-to-GDP ratio accelerates (Marinheiro; 2006).

⁶ We thank Henning Bohn for providing us with the series of US military expenditures and Carlos Marinheiro for updating the Portuguese dataset and adding a GDP deflator.

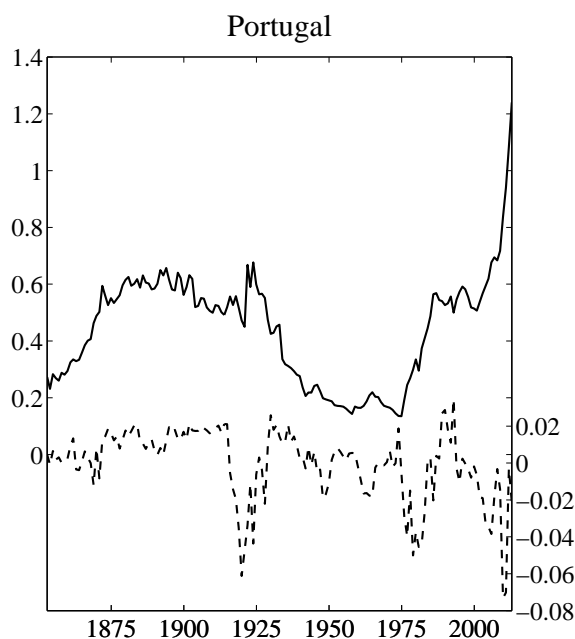


Figure 3.2: Fiscal data for Portugal For further notes see Figure 3.1.

3.2.2 The policy reaction function

To assess the so-called model based sustainability empirically, Bohn (1998) introduces the policy reaction regression

$$s_t = \alpha_0 + \rho d_{t-1} + \alpha_1 \tilde{g}_t + \alpha_2 \tilde{y}_t + u_t. \quad (3.1)$$

In (3.1) s_t is the primary surplus as share of GDP which adjusts to movements of the lagged end-of-period debt-to-GDP ratio denoted by d_{t-1} .⁷ A positive and significant reaction parameter ρ indicates that fiscal policy is in line with the model based IBC. The two additional control variables, temporary government spending (\tilde{g}_t) and business cycle variation (\tilde{y}_t), are motivated from the tax-smoothing model of Barro (1979) and considered as non-debt determinants of the primary surplus (Bohn; 1998). The measure of business cycle variation is determined as

⁷ In contrast to the with-interest deficit/surplus, the primary surplus excludes interest payments. Similar to Ghosh et al. (2013), we employ the lagged end-of-period debt in (3.1). As an alternative Bohn (1998) uses the initial debt at the beginning of period t . This is defined as the lagged end-of-period debt plus the interest payments of the current period. Greiner et al. (2006) and Greiner et al. (2007) argue that using the lagged debt is appropriate, since the government budget plans are determined in each year for the next year. Moreover, a related strand of literature addressing the ‘fiscal theory of the price level’ questions the causal direction between the surpluses and debt (Woodford; 1998; Canzoneri, Cumby and Diba; 2001). We also employ the lagged debt-to-GDP ratio to take account of this critique (see also Potrafke and Reischmann; 2014, on this point).

the difference between the log real GDP and a trend measured by means of the Hodrick-Prescott (HP) filter with smoothing parameter 100 (see, e.g., Bohn; 2008; Mendoza and Ostry; 2008; Ghosh et al.; 2013). Since military expenditures constitute the dominant source of temporary government spending in the US and the UK during the past 200 years, we employ these processes measured as shares of GDP (Bohn; 2008; Lukkezen and Rojas-Romagosa; 2012). The Portuguese temporary government spending is not dominated by military expenditures. As common in the literature we measure the temporary government spending by means of the gap between the logarithmic real expenditures and their HP trend again calculated with smoothing parameter 100 (see, e.g., Mendoza and Ostry; 2008; Lukkezen and Rojas-Romagosa; 2012; Ghosh et al.; 2013).⁸

3.2.3 Unit root characteristics

To assess the persistent movements of debt-to-GDP ratios of the US, the UK and Portugal displayed in Figures 3.1 and 3.2 and of the remaining variables involved in policy reaction functions we test for unit roots. A constant is included in the test regressions. For each country the full sample period and two subsamples covering post-1900 data and post-WWII data are considered.⁹

Debt-to-GDP ratios According to results displayed in the upper panel of Table 3.1 the null hypothesis of a unit root in the US debt-to-GDP ratio from 1792 to 2011 cannot be rejected at common significance levels applying the ADF and PP test. However, the unit root hypothesis can be rejected with 10% significance by means of the DF_{GLS} test. Moreover, the conclusion that the US debt-to-GDP ratio is stochastically trending is confirmed by means of the KPSS test statistic for which the null hypothesis of stationarity can be rejected with 1% significance. In general, comparable figures are obtained for the subsample covering the period from 1900 to 2011. In this case also the DF_{GLS} statistic indicates a unit root in the debt-to-GDP ratio. For the post-WWII period ranging until 2011 test results are rather heterogeneous. Applying the ADF and DF_{GLS} test a unit root for the debt-to-GDP ratio cannot be rejected at common significance levels, while the PP and KPSS statistics indicate stationarity of debt-to-GDP ratios.¹⁰

⁸ Results are qualitatively identical if we employ the real expenditures without interest payments in the HP filter as proposed by Potrafke and Reischmann (2014).

⁹ Initial years of post-WWII subsamples depend on country specific aftermaths of WWII and are chosen as in Lukkezen and Rojas-Romagosa (2012).

¹⁰ Unit root diagnostics are robust to several variations in test designs. Especially in the post-WWII samples or considering the real debt instead of debt as a share of the GDP, they become even

3.2 Size distortions in policy reaction functions

Series	Sample	ADF	PP	DF _{GLS}	KPSS
United States					
Debt	1792-2011	-1.846	-1.693	-1.831*	0.954***
	1900-2011	-1.931	-1.742	-1.103	0.376*
	1948-2011	-2.237	-3.289**	-0.746	0.283
Primary surplus	1792-2011	-6.945***	-4.996***	-6.744***	0.245
	1900-2011	-5.267***	-3.578***	-5.251***	0.092
	1948-2011	-2.484	-2.555	-1.746*	0.346
United Kingdom					
Debt	1830-2012	-2.312	-1.980	-1.419	0.227
	1900-2012	-2.101	-1.648	-1.652*	0.381*
	1946-2012	-4.877***	-3.376**	-1.695*	0.775***
Primary surplus	1830-2012	-4.857***	-4.769***	-3.659***	0.211
	1900-2012	-4.872***	-3.910***	-4.852***	0.065
	1946-2012	-6.117***	-6.868***	-4.588***	0.692**
Portugal					
Debt	1852-2012	1.718	0.704	2.368	0.218
	1900-2012	2.247	1.251	1.039	0.317
	1945-2012	2.367	2.771	2.265	0.969***
Primary surplus	1852-2012	-4.651***	-4.787***	-4.586***	0.601**
	1900-2012	-4.127***	-4.177***	-3.332***	0.324
	1945-2012	-3.763***	-3.807***	-3.453***	0.192

Table 3.1: Unit root tests for debt-to-GDP ratios and primary surpluses All series are measured as shares of GDP. A constant is included in the test regressions. The Augmented Dickey-Fuller test is denoted by ADF where the lag length is determined by means of the BIC. The column labeled PP shows results of the Phillips and Perron (1988) test. The Bartlett kernel is applied to calculate the long-run variance. The column DF_{GLS} refers to the Elliott et al. (1996) test where the BIC is applied to determine the lag length. The KPSS test of Kwiatkowski et al. (1992) is based on a Bartlett kernel. Significance at 1%, 5%, 10% level is indicated by ***, **, *.

Focusing on the UK (see the middle panel of Table 3.1) and the period between 1830 and 2012 one cannot reject a unit root in the debt-to-GDP ratio by means of the ADF, PP and DF_{GLS} test, while stationarity is indicated by the KPSS statistic. For the sample period from 1900 to 2012 the result for the KPSS test implies with 10% significance non-stationarity of public debt, while the null hypothesis of a unit root can be rejected by means of the DF_{GLS} test with 10% significance. For the post-WWII subsample of the UK, the debt-to-GDP ratio is diagnosed to be stationary, however, the KPSS statistic indicates non-stationarity.

more in favor of a unit root if a linear trend is added as proposed by Bohn (2008).

The lower panel of Table 3.1 documents diagnostic results for the Portuguese data. By means of the ADF, PP and DF_{GLS} statistics, we cannot reject the null hypothesis of a unit root for debt-to-GDP ratios throughout. For the KPSS statistic the null hypothesis of stationarity cannot be rejected conditional on the sample ranging from 1852 to 2012 and the sample covering the post-1900 period. For the post-WWII period, however, non-stationarity is also indicated by this statistic.

In summary, the findings for the US, the UK and Portugal are largely comparable to those in the related literature (compare, e.g., Wilcox; 1989; MacDonald and Speight; 1990; Marinheiro; 2006; Antonini et al.; 2013).¹¹ However, test results might mirror effects of small sample sizes or low power of unit root diagnostics for ‘near’ to unit root processes. Bohn (1998) argues that common unit root tests fail to detect mean reversion of debt-to-GDP ratios due to omission of the influence of temporary government spending and business cycle variations. Nevertheless, from an econometric perspective (‘near’ to) unit root processes might be best suited to describe the movements of the debt-to-GDP ratio, and we find strong evidence for integrated or at least highly persistent debt-to-GDP ratios of the US, the UK and Portugal in most subsamples. The further analysis can be interpreted as being embedded in a local to unit root framework where autoregressive coefficients are allowed to be close or identical to unity.

Primary surpluses For the US samples covering the periods 1792-2011 and 1900-2011 (see the upper panel of Table 3.1), unit root tests indicate throughout that primary surpluses follow a stationary process. For the post-WWII subsample stationarity of the primary surplus is indicated by means of DF_{GLS} and KPSS diagnostics, while the null hypothesis of a unit root cannot be rejected at common significance levels if the ADF and the PP test are applied. Results documented in the middle panel of Table 3.1 support almost uniformly the view that the primary surplus of the UK is stationary. Indicating non-stationarity, a notable exception is the KPSS statistic conditioning on the post-WWII period. For Portugal the results in the lower panel of Table 3.1 are similar to those for the UK, except for the KPSS statistic conditioning on the period 1852-2012. Thus, primary surpluses may be

¹¹These unit root diagnostics and tests for cointegration between revenues and expenditures as shares of GDP can be employed to test if fiscal policies are in accordance with an alternative deterministic IBC. In sum, we cannot unambiguously diagnose if fiscal policies are in accordance with the deterministic IBC from the documented unit root and cointegration tests employed in a former version of this work.

best described by means of stationary processes.¹²

3.2.4 Size distortions of policy response diagnostics

In distinct fields of economic research it has been argued that the persistence of regressors may invoke biased estimated slope coefficients and size distortions of conventional t -tests if the dependent variable in a regression is stationary. For instance, in the finance literature so-called predictive regressions are employed to evaluate the informational content of valuation ratios (e.g., the price-to-dividend ratio) for future equity returns. In principle, such predictive regressions have the same structure as the policy reaction function in (3.1). The left hand side variable of a predictive regression is a stationary measure of equity returns and the right hand side variable is a persistent valuation ratio relating equity prices to fundamentals (e.g., dividend payments).¹³ In case of the policy reaction function in (3.1), the left hand side variable is the stationary primary surplus, while the persistent debt-to-GDP ratio is among the explanatory variables. Stambaugh (1999) argues that negative correlations of equity returns with innovations of valuation ratios and regressors' persistence jointly invoke upward biases of estimated slope coefficients and size distortions of t -tests in predictive regressions. In light of a weakly balanced policy response regression, the magnitude of eventual biases and size distortions of conventional t -ratios are next uncovered by means of a Monte Carlo approach.

Since debt-to-GDP ratios are characterized by persistent movements, we expect the estimate of the policy reaction coefficient ρ in (3.1) to be upward biased and both the standard t -ratio and its heteroskedasticity and autocorrelation robust variant (e.g., Newey and West; 1987) to be oversized in finite samples if debt innovations are negatively correlated with primary surpluses. To illustrate such properties for distinct degrees of persistence and correlation, we simulate from the model

$$s_t = \alpha_0 + \rho d_{t-1} + \alpha_1 \tilde{g}_t + \alpha_2 \tilde{y}_t + u_t, \quad (3.2)$$

$$d_t = \beta_0 + \beta_1 d_{t-1} + e_t, \quad (3.3)$$

¹²Unit root diagnostics for temporary government spending (\tilde{g}_t) and business cycle variations (\tilde{y}_t) are summarized in Table 3.8 in Appendix 3.B. Measures of business cycle variations of all countries and the measure of temporary government spending of Portugal are by design transitory processes. The series of military expenditures appear stationary for all UK (sub)samples and the full US sample. For the US all unit root diagnostics except the PP statistic indicate stationarity of this series for the period 1900 to 2011. During the post-WWII period unit root diagnostics for US military expenditures are rather mixed.

¹³There is an extensive literature discussing the ('near' to) unit root characteristics of financial valuation ratios and their implications for the predictability of equity returns. The reader may consult Kojien and Nieuwerburgh (2011) for an overview.

under the null hypothesis of no sustainability, i.e., $\rho = 0$ (Campbell and Yogo; 2006). Innovations u_t and e_t are drawn from a bivariate normal distribution with mean zero and covariance matrix

$$\text{Cov}(u_t, e_t) = \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{pmatrix}. \quad (3.4)$$

The parameters $\alpha_0 = 0.002$, $\alpha_1 = -0.734$, $\alpha_2 = 0.094$, $\sigma_u = 0.025$ and $\sigma_e = 0.040$ are set to the estimates for the full US sample (1792-2011). As explanatory variables \tilde{g}_t and \tilde{y}_t we take the actual series of temporary military outlays and business cycle variation, the sample size is $T = 220$.¹⁴ The persistence parameter β_1 varies on a grid covering the interval $[0, 1]$ and the correlation between u_t and e_t , denoted by $\delta = \sigma_{ue}/(\sigma_u\sigma_e)$, ranges on the interval $[-1, 0]$.¹⁵ Without loss of generality the initial condition d_0 and coefficient β_0 are set to zero. Otherwise simulated debt-to-GDP processes would become random walks with drifts in case of $\beta_1 = 1$ and, thus, lack comparability to processes simulated for $\beta_1 < 1$.

Figure 3.3 displays the estimated policy reaction coefficient $\hat{\rho}$ averaged over 10,000 Monte Carlo replications where the true value is $\rho = 0$. The bias is an increasing function in the absolute correlation of primary surpluses' and debt-to-GDP ratios' innovations and the persistence of debt-to-GDP ratios. For the empirical estimates of $\delta = -0.582$ and $\beta_1 = 0.990$ indicated by the black dashed lines, the bias is about 0.008 units which is an essential proportion of the empirically small estimated slope coefficients shown below. Figure 3.4 displays rejection frequencies of a right hand sided Newey and West (1987) heteroskedasticity and autocorrelation robust t -test with 5% nominal significance.¹⁶ Noting that the Monte Carlo experiment is run under the null hypothesis, the actual rejection frequencies should be contrasted against the 5% benchmark. Apparently, the actual type I error probabilities of sustainability diagnostics exceed the nominal benchmark, underlining slow convergence of the t -ratio to its asymptotic Gaussian limit distribution. Size distortions

¹⁴To address potential endogeneity of the variables in (3.2), we specify a four dimensional VAR model comprising $(s_t, d_{t-1}, \tilde{g}_t, \tilde{y}_t)'$ and, alternatively, a two dimensional VAR model for $(s_t, d_{t-1})'$ where \tilde{g}_t and \tilde{y}_t enter as exogenous variables. Model selection is according to the BIC criterion. With regard to the marginal equation for s_t considering \tilde{g}_t and \tilde{y}_t as exogenous is supported by the BIC criterion. This result holds uniformly for all considered (sub)samples, if a constant or additionally a trend is incorporated in VAR specifications.

¹⁵The negative range of δ is chosen since empirical correlations are typically negative. In contrast, a positive correlation would imply a downward bias of $\hat{\rho}$ (Stambaugh; 1999).

¹⁶Since we detect significant heteroskedasticity and autocorrelation in the empirical residuals, we apply the robust t -ratio in all remaining sections following the related literature. The empirical rejection frequencies of a conventional one sided t -test are similar.

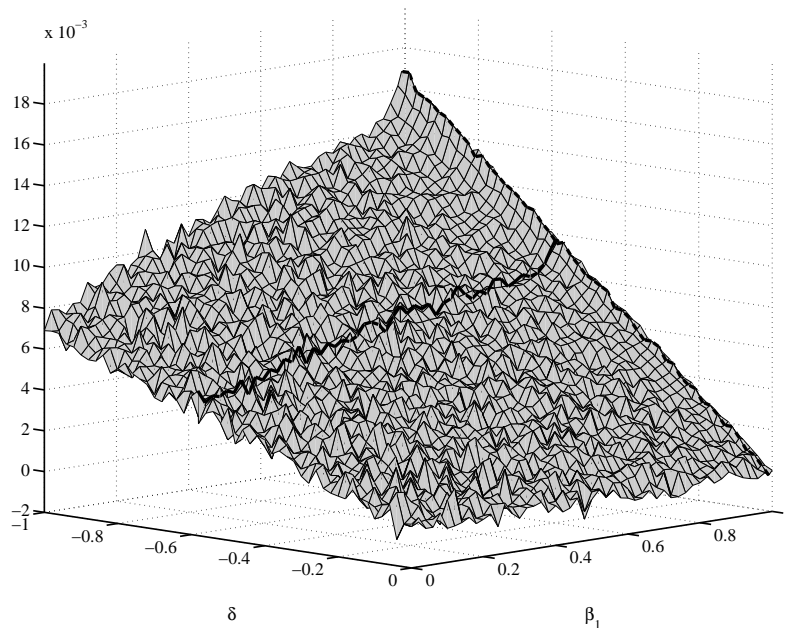


Figure 3.3: Bias of policy the reaction coefficient The figure displays the average estimated policy reaction coefficient when the true value is $\rho = 0$. The underlying data cover US fiscal figures for the period 1792-2011. For each point on the grid 10,000 simulations are run. Black dashed lines indicate empirical biases if $\delta = -0.582$ or respectively $\beta_1 = 0.990$ are fixed at their empirical estimates.

increase with β_1 and δ in absolute terms. For the combination of the estimates $\hat{\delta} = -0.582$ and $\hat{\beta}_1 = 0.990$, the actual size of the right hand sided t -test with 5% nominal significance is 20.8%. Not alone the persistence, but also innovation correlations in the US data are responsible for the strongly distorted significance level of the test. We provide the correlation figures for all datasets in Section 3.4 explicitly. At this point it is worthwhile noting that the likelihood to falsely reject the null hypothesis of no sustainability is sizeable for the empirically relevant range of estimates $\hat{\delta}$ and $\hat{\beta}_1$.

3.3 Size control for model based IBC diagnosis

In this Section we describe techniques that correct for size distortions of t -tests and biases of estimated slope coefficients in policy reaction functions. We explain (i) the Monte Carlo based procedure employed by Goyal and Welch (2008), and (ii) the local to unit root asymptotic framework introduced by Campbell and Yogo (2006). Then, Monte Carlo based inference and the local to unit root asymptotic analysis are adapted to provide size controlled inference on the sustainability of the fiscal balances.

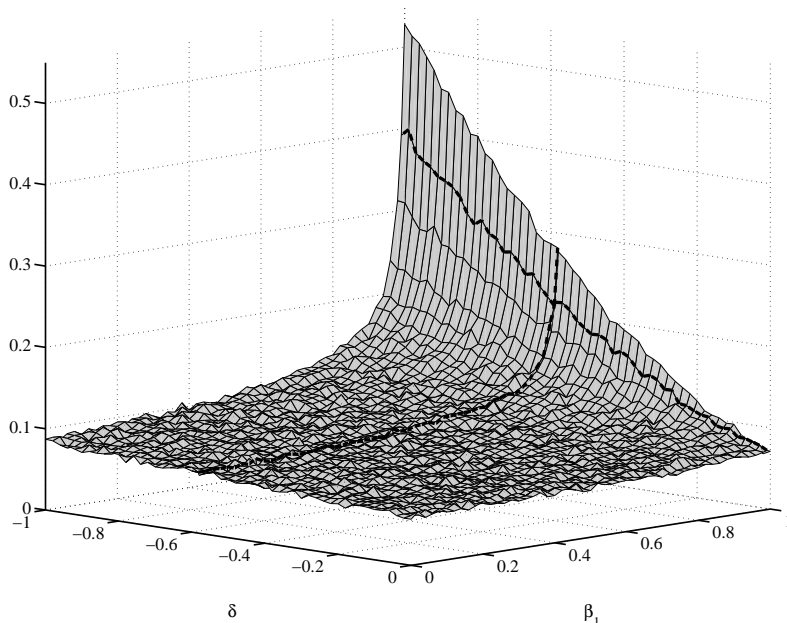


Figure 3.4: Size distortion of the robust t -test The figure displays the empirical rejection frequencies of a 5% right hand sided heteroskedasticity and autocorrelation robust t -test applying the general covariance estimate of Newey and West (1987) when $\rho = 0$. The underlying data cover US fiscal figures for the period 1792-2011. The bandwidth in the covariance estimate is automatically selected as proposed by Newey and West (1994). For each point on the grid 10,000 simulations are run. Black dashed lines indicate empirical rejection frequencies if $\delta = -0.582$ or respectively $\beta_1 = 0.990$ are fixed at their empirical estimates.

3.3.1 Monte Carlo inference

To account for regressors' persistence and correlation of its innovations with equity returns, Goyal and Welch (2008) suggest a Monte Carlo approach for inference in predictive regressions.¹⁷ In small samples Monte Carlo procedures have often size advantages compared with first order asymptotic approximations. This should be helpful when evaluating the sustainability of fiscal policies conditional on post-WWII subsamples that cover about 70 annual observations. In the context of policy reaction functions such a Monte Carlo procedure is based on the model in (3.2) to (3.4). The regression specific simulation follows the steps already described in Section 3.2.4. Given parameter estimates we draw innovations from the bivariate normal distribution with zero mean and the covariance matrix in (3.4), and simulate 10,000 processes of surpluses and debt-to-GDP ratios under the null hypothesis of non-sustainability ($\rho = 0$).¹⁸ The measures of temporary government spending (\tilde{g}_t)

¹⁷ A related approach is adopted by Mark (1995) and Kilian (1999) in the context of exchange rate predictions.

¹⁸ Alternatively, we apply the bootstrap procedure proposed by Goyal and Welch (2008) drawing with replacement from estimated residuals. In this framework the cross-correlation structure

and business cycle variation (\tilde{y}_t) are set to their actual realizations. The initial values d_0 are randomly selected from the empirical range of the debt-to-GDP series. Monte Carlo critical values are determined by estimating (3.2) for each replication and evaluating the empirical 95% quantile of the simulated distribution of robust t -ratios for parameter estimates $\hat{\rho}$. Moreover, this approach can easily be adopted for a bias correction of estimates. The average over the 10,000 estimated policy reaction coefficients where the true value is $\rho = 0$ can be interpreted as an estimate of the bias in the empirical $\hat{\rho}$, which is used for bias correction (Mark; 1995).¹⁹

3.3.2 Local to unit root asymptotics

As an alternative to the Monte Carlo approach, Campbell and Yogo (2006) propose to evaluate the predictability of equity returns by means of local to unit root asymptotic theory. In the following we briefly sketch the intuition of the approach and adapt this framework to test for sustainability of fiscal policy.²⁰

If the parameter β_1 in (3.3) is known, it is possible to apply a Neyman-Pearson uniformly most powerful test which accounts for residuals' correlation and the persistence of debt-to-GDP ratios. Campbell and Yogo (2006) show that the corresponding statistic can be calculated by replacing the primary surplus s_t in (3.2) with $s_t - \sigma_{ue}/\sigma_e^2(d_t - \beta_1 d_{t-1})$. Then, the t -statistic for the reaction parameter $\hat{\rho}$, denoted Q , is conditionally uniformly most powerful. Treating β_1 as unknown, a feasible version of the test procedure is based on the local to unit root framework, expressing the persistence parameter as $\beta_1 = 1 + c/T$, where T is the sample size. Conditional on a given significance level, the DF_{GLS} statistic and the correlation between residuals u_t and e_t in (3.2) and (3.3), Campbell and Yogo (2005) document confidence intervals for c (or implicitly β_1). Given this interval Campbell and Yogo (2006) derive a refined Bonferroni confidence interval for the Q -statistic.²¹ Moreover, by means of a pre-test it can be evaluated if the persistence of the debt-to-GDP ratio joint with correlation of the debt and surplus innovations in (3.2) and (3.3) are likely to invalidate the nominal significance level of conventional t -tests. The

is captured by means of the so-called pairs bootstrap of drawing with replacement from tuples $(\hat{u}_t \hat{e}_t)'$. Results from pairs bootstrap and Monte Carlo simulation are qualitatively identical.

¹⁹Such bias correction is related to the approach described by Stambaugh (1999) who similarly evaluates the limiting bias of $\hat{\rho}$ based on estimates of δ and β_1 . It is worthwhile to notice that the magnitude of the bias is independent from the value of true parameter ρ (Stambaugh; 1999).

²⁰A detailed derivation of the test procedure is provided by Campbell and Yogo (2005, 2006).

²¹This asymptotic framework is also valid under heteroskedasticity of residuals (Campbell and Yogo; 2006). It is worthwhile to notice that the estimate of ρ can be close to the upper bound of the confidence interval calculated by means of the Bonferroni procedure or even larger due to the biased estimate of ρ in finite samples (Campbell and Yogo; 2006).

intuition behind the pre-test is to assess the degree of the size distortion. This is comparable to contrasting the actual size of a test with 5% nominal significance as illustrated in Figure 3.4 with the 5% benchmark. Campbell and Yogo (2006, Table 1) document regions (c_{min}, c_{max}) for which the size of a 5% right hand sided t -test is in excess of 7.5%. These intervals depend on $\hat{\delta}$. If the confidence interval for c is strictly below or above the region (c_{min}, c_{max}) the null hypothesis of an empirical size in excess of 7.5% can be rejected.

It is worthwhile to notice that compared with the standard specification of a predictive regression, the model in (3.2) includes additional regressors, namely temporary government spending (\tilde{g}_t) and business cycle variation (\tilde{y}_t) . Both respective nuisance parameters, α_1 and α_2 , are estimated consistently.

3.4 Size controlled inference on fiscal sustainability

In this section we first highlight the potential and magnitude of size distortions in policy response regressions for the US, the UK and Portugal. In the second place, we draw bias and size controlled conclusions on the sustainability of fiscal balances in these three countries. A brief summary offers a condensed view at the empirical results and further remarks on the need for size control in empirical assessments of model based sustainability.

3.4.1 Size distortions in policy response regressions

As a first result, Table 3.2 documents the estimated residual correlations $\hat{\delta} = \widehat{\text{Corr}}(u_t, e_t)$ and autoregressive parameters $\hat{\beta}_1$ for the US, the UK and Portugal. Given Figure 3.4 these two quantities offer an impression if the t -test suffers from notable size distortions. The correlation of the debt and surplus innovations is around -0.6 for the US data. Compared with the corresponding subsamples for the UK and Portugal, the correlation in the US data is the largest in absolute terms. While the correlation estimates remain almost invariant over the three US subsamples, correlations increase in absolute value for the two remaining countries over the subsamples. The autocorrelation parameter $\hat{\beta}_1$ is close to unity for all countries and (sub)samples except for the post-WWII period and US and UK data. For the case of Portugal $\hat{\beta}_1$ is even in excess of unity and increases over the subsamples, while it decreases with the size of the US and UK (sub)samples. These findings confirm former diagnostics for unit roots in debt-to-GDP ratios (see Section 3.2.3).

Column 4 in Table 3.2 documents the simulated size of a 5% right hand sided

Sample	$\hat{\delta}$	$\hat{\beta}_1$	sim. size 5%	95% CI: β_1	95% CI: c
United States					
1792-2011	-0.582	0.990	0.208	[0.971, 1.018]	[- 6.335 , 3.984]
1900-2011	-0.637	0.973	0.225	[0.971, 1.040]	[- 3.259 , 4.467]
1948-2011	-0.690	0.917	0.276	[0.910, 1.065]	[- 5.671 , 4.091]
United Kingdom					
1830-2012	-0.163	0.983	0.097	[0.975, 1.024]	[- 4.536 , 4.282]
1900-2012	-0.163	0.982	0.116	[0.953, 1.037]	[- 5.270 , 4.171]
1946-2012	-0.590	0.926	0.203	[0.949, 1.067]	[- 3.398 , 4.445]
Portugal					
1852-2012	-0.308	1.031	0.115	[1.003, 1.034]	[0.485 , 5.375]
1900-2012	-0.319	1.048	0.137	[0.993, 1.044]	[- 0.753 , 4.954]
1945-2012	-0.459	1.090	0.161	[1.020, 1.087]	[1.335 , 5.858]

Table 3.2: Pre-tests for size distortions The table documents in column 2 the estimated correlations of innovations u_t and e_t in (3.2) and (3.3), respectively. The estimated autoregressive parameters β_1 are provided in column 3. Column 4 displays the simulated size of a 5% right hand sided heteroskedasticity and autocorrelation robust t -test (Newey and West; 1987). The bandwidth for the robust t -statistic is automatically selected as proposed by Newey and West (1994). The 95% confidence intervals for β_1 in column 5 are followed by the 95% confidence intervals for c . Bold figures indicate non-rejection of the null hypothesis of a size in excess of 7.5% by means of the pre-test proposed in Campbell and Yogo (2006).

robust t -test for the policy reaction parameter $\hat{\rho}$ in (3.2). In consequence, for the US data the actual size of a robust t -test with 5% nominal significance is 20.8% and 22.5% for the full sample and the period 1900 to 2011, respectively. For the post-WWII period it is 27.5% and, thus, five times larger than the nominal reference. The rise in excess type I error probabilities over the three subsamples is related to the (in absolute terms) strengthened residual correlation. In consequence, a conventional t -test is most likely inappropriate to evaluate the sustainability of the US fiscal policy with 5% nominal significance.²² This conclusion is supported by the pre-test proposed by Campbell and Yogo (2006). The 95% confidence bounds of β_1 and c in $\beta_1 = 1 - T/c$ are shown in columns 5 and 6 of Table 3.2, respectively. These intervals are computed by means of the distribution of the DF_{GLS} statistic under the alternative hypothesis of stationarity. The confidence interval for β_1 includes the case of a unit root for all US (sub)samples. Moreover, the bolded entries in column 6 of Table 3.2 indicate that the interval for the parameter c is not strictly below or above the critical region tabulated in Campbell and Yogo (2006,

²²If we consider alternatively standard instead of robust t -statistics results are qualitatively identical.

Table 1). Thus, the null hypothesis of a size in excess of 7.5% cannot be rejected with 5% significance.²³

Examining UK data the simulated size of the right hand sided t -test comes closer to the 5% benchmark in each (sub)sample in comparison with corresponding results for US data. However, it still amounts to 9.7%, 11.6% and 20.3% for the respective (sub)periods. The pre-test indicates with 5% significance an actual size which is in excess of 7.5% regardless of the sample period. Again the strengthened correlation of innovations in debt-to-GDP ratios and primary surpluses over the three (sub)samples appears to trigger the size distortions. Similarly, we diagnose that t -tests are size distorted conditional on all (sub)samples of Portuguese figures. It is noteworthy that the simulated size of the t -test for the Portuguese post-WWII period is below respective estimates for the US and the UK, although $\hat{\beta}_1$ indicates an explosive path of Portuguese debt-to-GDP ratios. Such a weaker size distortion may reflect the relatively small residual correlation of $\widehat{\text{Corr}}(u_t, e_t) = -0.459$. The pre-test indicates throughout significant size distortions, even if the lower bounds of the confidence interval of c become positive for the 1852-2012 and 1945-2012 sample periods.

In summary, the null hypothesis that the actual size of a right hand sided t -test with 5% nominal significance is in excess of 7.5% cannot be rejected for all countries and (sub)samples. The actual significance levels exceed the nominal benchmark in a sizable manner throughout, and especially for the case of the US and the post-WWII UK fiscal data. As a consequence, the t -statistic may be invalid to assess model based sustainability and, thus, unsuitable for policy monitoring.

3.4.2 Size corrected diagnosis of fiscal balances' sustainability

Accounting for persistence of regressors and correlation of innovations, Table 3.3 documents bias and size adjusted inferential diagnostics on the sustainability of US and UK budget balances. On the one hand we employ Monte Carlo based 5% critical values, and on the other hand asymptotic 90% Bonferroni confidence intervals.²⁴ The OLS estimate of the policy response parameter ρ is displayed in the first row and the corresponding robust t -statistic in parenthesis below. Bias adjusted policy reaction coefficients summarized in the row labeled $\hat{\rho}_{adj}$ are uniformly smaller

²³ As an example, consider the US sample period from 1792 to 2011. For a given $\hat{\delta} = -0.582$ the 95% confidence interval of c ($-6.335, 3.984$) is not strictly below or above the critical region ($c_{min} = -33.899, c_{max} = 6.557$) (Campbell and Yogo; 2006, Table 1).

²⁴ The lower bound of these two-sided 90% confidence intervals is identical to that of a 95% 'one sided interval'.

than the respective OLS estimates. Applying as critical value the 95% quantile of the standard normal distribution, 1.645, it appears that the US fiscal policy was sustainable for the entire sample from 1792 to 2011 and the period from 1900 to 2011. For the post-WWII period the bias adjusted slope coefficient becomes negative and the US fiscal policy lacks sustainability. Simulation based critical values in the third row of Table 3.3 are around 2.7 and increase if the size of the three US (sub)samples decreases. Employing these critical values, the conclusions derived by means of the Gaussian quantile do not change. Moreover, the 90% Bonferroni confidence interval for ρ does not include zero (indicated with bold entries) for the first two subsamples, but indicates non-sustainability of fiscal balances during the post-WWII period.²⁵

For the full UK sample (1830-2012), the simulated critical value regarding the policy reaction to debt-to-GDP movements comes closer to the Gaussian quantile compared with the respective statistic for the US. This corresponds to somewhat weaker size distortions of the t -test for this sample (see Table 3.2). The estimated policy reaction parameter is positive and significant for the entire sample period and the post-1900 period given Gaussian, Monte Carlo and Bonferroni critical values. However, we find evidence against a sustainable public debt for the post-WWII period from size corrected Monte Carlo and Bonferroni critical values. In contrast, the conventional t -test indicates that the UK fiscal policy accords with the IBC for this subsample. In summary, the fiscal balance of the UK might be in accordance with the IBC for the sample periods covering more than a century. It is worthwhile to mention that although we arrive at similar conclusions on non-sustainability of post-WWII fiscal policy in the US and the UK, for the latter this diagnosis relies on the application of size corrected critical values.

For the case of Portugal policy response estimates $\hat{\rho}$ displayed in Table 3.4 are relatively small and bias adjusted complements are negative in two (sub)samples. The t -statistics do not indicate a significantly positive response parameter. These conclusions hold for the standard asymptotic and the Monte Carlo critical values. All Bonferroni confidence intervals include zero. The Portuguese fiscal policy does not appear sustainable regardless of the sample choice. Next, we analyze post-WWII subsamples in more detail in light of currently high debt burdens.²⁶

²⁵ As argued by Campbell and Yogo (2006), the Bonferroni test has an upward biased rejection frequency if the sample contains less than 100 observations. Thus, the Monte Carlo critical values might be more appropriate for the analysis of the post-WWII samples.

²⁶ Noteworthy, Bohn (1998, 2008) and Ghosh et al. (2013) add quadratic and cubic terms of the debt-to-GDP ratio to the policy reaction function to cover nonlinear effects. Along these lines we augment the linear policy rule in (3.2) by means of the squared lagged debt-to-GDP ratio

United States			
	1792-2011	1900-2011	1948-2011
$\hat{\rho}$	0.100	0.109	0.033
t -stat	(4.355)	(3.807)	(0.946)
sim crit 5%	2.592	2.752	3.056
Q -test 90% CI: ρ	[0.040, 0.078]	[0.016, 0.079]	[-0.034, 0.057]
$\hat{\rho}_{adj}$	0.091	0.090	-0.012
$\hat{\alpha}_0$	0.002 (0.535)	-0.002 (-0.319)	-0.015 (-1.612)
$\hat{\alpha}_1$	-0.734 (-6.784)	-0.731 (-5.112)	0.006 (0.030)
$\hat{\alpha}_2$	0.094 (1.328)	0.112 (1.183)	0.402 (2.343)
R^2 /adj. R^2	0.632/0.627	0.579/0.567	0.173/0.131
United Kingdom			
	1830-2012	1900-2012	1946-2012
$\hat{\rho}$	0.049	0.053	0.031
t -stat	(9.185)	(7.965)	(1.729)
sim crit 5%	2.051	2.297	2.831
Q -test 90% CI: ρ	[0.040, 0.053]	[0.040, 0.060]	[-0.022, 0.024]
$\hat{\rho}_{adj}$	0.048	0.051	0.007
$\hat{\alpha}_0$	0.006 (0.787)	0.004 (0.377)	-0.016 (-0.891)
$\hat{\alpha}_1$	-0.663 (-9.836)	-0.666 (-9.436)	0.191 (0.340)
$\hat{\alpha}_2$	-0.256 (-2.183)	-0.290 (-2.080)	0.389 (1.351)
R^2 /adj. R^2	0.870/0.868	0.869/0.865	0.364/0.333

Table 3.3: Policy response regressions for the US and the UK For each country the first row displays the estimates of ρ followed by the heteroskedasticity and autocorrelation robust t -statistics (Newey and West; 1987) in parenthesis. The bandwidth for the robust t -statistic is automatically selected as proposed by Newey and West (1994). The simulated 5% Monte Carlo based critical values of a right hand sided test are provided in the third row. The 90%-confidence intervals of the Bonferroni Q -statistics are shown in the fourth row. Bold intervals exclude zero. Bias adjusted estimates of ρ are presented in the row labels $\hat{\rho}_{adj}$. The lower panel displays the remaining estimates, corresponding heteroskedasticity and autocorrelation robust t -statistic in parenthesis and the (adjusted) degree of explanation.

and consider marginal effects at distinct levels of debt-to-GDP ratios. In sum, we do not detect notable variation of the policy reaction.

	Portugal		
	1852-2012	1900-2012	1945-2012
$\hat{\rho}$	0.004	-0.010	-0.018
t -stat	(0.218)	(-0.548)	(-1.135)
sim crit 5%	2.343	2.706	3.093
Q -test 90% CI: ρ	[-0.006, 0.019]	[-0.019, 0.010]	[-0.023, 0.011]
$\hat{\rho}_{adj}$	0.004	-0.010	-0.018
$\hat{\alpha}_0$	-0.002 (-0.299)	-0.001 (-0.082)	-0.003 (-0.585)
$\hat{\alpha}_1$	-0.097 (-5.380)	-0.121 (-4.682)	-0.136 (-2.852)
$\hat{\alpha}_2$	0.145 (3.979)	0.188 (6.099)	0.119 (1.615)
R^2 /adj. R^2	0.233/0.219	0.312/0.293	0.268/0.234

Table 3.4: Policy response regressions for Portugal For further notes see Table 3.3

3.4.3 The sustainability of fiscal balances in recent years

The costly policy interventions during the recent financial crisis, the Great Recession and the European debt crisis as interrelated economic challenges disclosure the importance of a detailed analysis of recent fiscal figures. Figure 3.1 and 3.2 emphasize the strong accelerations of debt burdens for all three countries since the financial crisis in 2008. In Portugal an upward trend already occurs around the introduction of Euro banknotes in 2002. Thus, data of most recent years might substantially influence the test decisions regarding countries' model based sustainability. A natural question which arises is if size distortions in policy reaction regressions prevent a timely detection of unbalanced fiscal figures.

To investigate this issue, we mimic the perspective of a 'fiscal monitor' and condition the IBC diagnosis on post-WWII subsamples ranging until 2006 and extend these subsamples recursively until the last sample quote date in 2011 (US) or 2012 (UK/PT). Respective diagnostics for US and UK fiscal figures are documented in Table 3.5. As described above the US fiscal policy is diagnosed as non-sustainable for the 1948-2011 subsample independent of applied critical values. In contrast, conventional and size corrected inference arrive at opposite conclusions for the 1948-2010 US subsample. The fiscal balance appears to accord with the IBC with nominal 5% significance if test decisions are conditioned on Gaussian critical values. In contrast, however, conditioning the test decision on Monte Carlo or the Bonfer-

roni critical values arrives at the opposite conclusion of fiscal non-sustainability. Thus, size corrected inferential diagnosis issues warning signals of fiscal imbalances which are hidden by size distortions of conventional test approaches. For subsamples excluding post-2009 observations, the policy reaction coefficient is positive with 5% significance and test decisions from spurious and size controlled p -values do not contradict at the 5% nominal significance level.

Investigating the fiscal policy of the UK, we diagnose three further subsamples where spurious p -values based on Gaussian quantiles spuriously indicate fiscal sustainability with nominal 5% significance. The test decision conditional on conventional quantiles and size corrected versions firstly contradict for the post WWII subsample ending 2009. While the null hypothesis of an non-sustainable fiscal policy is rejected with nominal 5% significance applying Gaussian critical values, it cannot be rejected for Bonferroni critical values. In contrast to the US, such opposing test decisions are diagnosed for all remaining UK subsamples including post-2008 data. The analysis of future data has to show if the lack of fiscal sustainability obtained from recent data is a temporary or persistent phenomenon.

In contrast, policy reaction coefficients estimated for Portuguese subsamples documented in Table 3.6 are throughout insignificant regardless of the applied test procedure. In a further analysis summarized in Table 3.7, we exclude observations from the era with circulating Euro banknotes and consider Portuguese subsamples starting in 1852, 1900 and 1945 and ending in 2001. Estimated policy reaction coefficients and their bias adjusted versions become positive. However, fiscal policies appear to accord with the IBC only for the largest sample ranging from 1852 to 2001 conditional on conventional and size corrected critical values.²⁷ In sum, we diagnose strong evidence for non-sustainable fiscal policies for Portugal.²⁸

²⁷In the 1900-2001 period, Bonferroni critical values indicate sustainability. However, Monte Carlo based and also distorted asymptotic critical values hint at non-sustainability.

²⁸With respect to the influence of the two control variables, temporary government spending and business cycle variation, results in Table 3.3 to 3.7 can be summarized as follows. In line with Bohn (1998, 2008), the influence of temporary government spending on the primary surplus is in general negative, while that of business cycle variation is positive. Focusing on the post-WWII subsamples, in the US only the effect of the business cycle variation is significant at the 1% level and for the UK in most subsamples neither temporary government spending nor business cycle variation is significant at the 10% level. Notably, the primary surplus of the US and the UK accelerates with an increasing government expenditure in those subsamples where the influence of the lagged debt-to-GDP ratio is insignificant. In samples covering more than 100 years, the effect of temporary government spending is negative with 1% significance for the US and the UK. Moreover, an increase of the business cycle variation generates a decrease of the primary surplus of the UK. For Portugal temporary government spending has throughout a negative and the business cycle variation a positive effect with 5% significance.

3.4 Size controlled inference on fiscal sustainability

United States							
	1948 -2006	1948 -2007	1948 -2008	1948 -2009	1948 -2010	1948 -2011	1948 -2012
$\hat{\rho}$	0.066	0.066	0.066	0.065	0.052	0.033	–
t -stat	(4.248)	(4.240)	(4.173)	(3.802)	(2.330)	(0.946)	
sim crit 5%	2.906	2.892	2.880	2.898	2.983	3.056	–
Q -test 90% CI: ρ	[0.011, 0.069]	[0.010, 0.068]	[0.011, 0.070]	[0.003, 0.077]	[-0.014, 0.071]	[-0.034, 0.057]	–
$\hat{\rho}_{adj}$	0.038	0.038	0.038	0.033	0.013	-0.012	–
$\hat{\alpha}_0$	-0.013 (-1.644)	-0.013 (-1.691)	-0.013 (-1.826)	-0.017 (-1.995)	-0.017 (-1.973)	-0.015 (-1.612)	–
$\hat{\alpha}_1$	-0.144 (-1.199)	-0.145 (-1.234)	-0.137 (-1.147)	-0.106 (-0.771)	-0.056 (-0.339)	0.006 (0.030)	–
$\hat{\alpha}_2$	0.290 (2.775)	0.283 (2.759)	0.283 (2.811)	0.385 (2.459)	0.408 (2.299)	0.402 (2.343)	–
R^2	0.374	0.367	0.362	0.328	0.244	0.173	–
adj. R^2	0.339	0.332	0.327	0.292	0.205	0.131	
United Kingdom							
	1946 -2006	1946 -2007	1946 -2008	1946 -2009	1946 -2010	1946 -2011	1946 -2012
$\hat{\rho}$	0.042	0.042	0.042	0.039	0.036	0.033	0.031
t -stat	(4.204)	(4.102)	(3.756)	(2.853)	(2.249)	(1.883)	(1.729)
sim crit 5%	2.661	2.680	2.629	2.733	2.789	2.784	2.831
Q -test 90% CI: ρ	[0.009, 0.040]	[0.008, 0.039]	[0.006, 0.038]	[-0.005, 0.033]	[-0.014, 0.029]	[-0.019, 0.026]	[-0.022, 0.024]
$\hat{\rho}_{adj}$	0.033	0.032	0.030	0.023	0.016	0.011	0.007
$\hat{\alpha}_0$	0.005 (0.507)	0.004 (0.402)	0.002 (0.158)	-0.005 (-0.376)	-0.010 (-0.616)	-0.014 (-0.752)	-0.016 (-0.891)
$\hat{\alpha}_1$	-0.347 (-1.406)	-0.325 (-1.285)	-0.282 (-0.963)	-0.118 (-0.285)	0.017 (0.033)	0.118 (0.208)	0.191 (0.340)
$\hat{\alpha}_2$	0.226 (0.870)	0.228 (0.878)	0.237 (0.928)	0.396 (1.334)	0.428 (1.353)	0.416 (1.339)	0.389 (1.351)
R^2	0.452	0.453	0.453	0.431	0.403	0.380	0.364
adj. R^2	0.423	0.425	0.425	0.402	0.373	0.350	0.333

Table 3.5: Policy response regressions for recent US and UK data For further notes see Table 3.3

	Portugal						
	1945 -2006	1945 2007	1945 -2008	1945 -2009	1945 -2010	1945 -2011	1945 -2012
$\hat{\rho}$	0.001	0.001	0.000	-0.010	-0.021	-0.017	-0.018
t -stat	(0.041)	(0.066)	(0.006)	(-0.480)	(-0.884)	(-0.908)	(-1.135)
sim crit 5%	2.726	2.578	2.626	2.874	2.945	3.000	3.093
Q -test 90% CI: ρ	[-0.021, 0.021]	[-0.020, 0.021]	[-0.020, 0.019]	[-0.029, 0.012]	[-0.036, 0.005]	[-0.028, 0.010]	[-0.023, 0.011]
$\hat{\rho}_{adj}$	-0.005	-0.004	-0.005	-0.015	-0.024	-0.018	-0.018
$\hat{\alpha}_0$	-0.008 (-1.286)	-0.008 (-1.389)	-0.007 (-1.387)	-0.005 (-0.853)	-0.002 (-0.299)	-0.003 (-0.539)	-0.003 (-0.585)
$\hat{\alpha}_1$	-0.120 (-2.921)	-0.123 (-2.960)	-0.123 (-2.962)	-0.128 (-2.848)	-0.131 (-2.755)	-0.134 (-2.823)	-0.136 (-2.852)
$\hat{\alpha}_2$	0.123 (1.766)	0.120 (1.732)	0.117 (1.698)	0.139 (1.812)	0.141 (1.744)	0.126 (1.672)	0.119 (1.615)
R^2	0.260	0.269	0.269	0.256	0.261	0.257	0.268
adj. R^2	0.221	0.231	0.232	0.219	0.225	0.221	0.234

Table 3.6: Policy response regressions for recent Portuguese data For further notes see Table 3.3

	Portugal		
	1852-2001	1900-2001	1945-2001
$\hat{\rho}$	0.028	0.018	0.021
t -stat	(2.289)	(1.177)	(1.219)
sim crit 5%	1.998	2.201	2.883
Q -test 90% CI: ρ	[0.017, 0.039]	[0.002, 0.033]	[-0.002, 0.042]
$\hat{\rho}_{adj}$	0.027	0.016	0.010
$\hat{\alpha}_0$	-0.010 (-1.955)	-0.008 (-1.587)	-0.012 (-2.275)
$\hat{\alpha}_1$	-0.091 (-5.498)	-0.112 (-4.743)	-0.118 (-2.978)
$\hat{\alpha}_2$	0.142 (3.555)	0.186 (5.153)	0.129 (1.864)
R^2 /adj. R^2	0.368/0.355	0.401/0.383	0.338/0.299

Table 3.7: Policy response regressions for Portugal excluding the era of circulating Euro banknotes For further notes see Table 3.3

3.4.4 Summary of size controlled diagnostics

In summary, model based assessments of the IBC by means of common t -ratios suffer from marked size distortions. Pretesting for size violations highlights that respective test decisions are not taken with the supposed significance level. For the sustainability diagnosis in long time series covering more than a century t -statistics are not in a critical region. Thus, the application of standard or size corrected critical values does not arrive at contradictory conclusions on the sustainability of fiscal balances. The fact that the diagnosis of sustainability appears invariant with regard to the distinct inferential procedures, however, should not be taken to support common inferential techniques for the monitoring of fiscal policy for two reasons. First, the concordance of test decisions gathered from biased vs. unbiased critical values depends generally on the presumed nominal level. Second, only unbiased critical values respect the nominal significance chosen by the analyst and is therefore in line with her risk attitudes.

Moreover, we identify several subsamples for which conventional and size corrected inference induces contradicting conclusions regarding the sustainability of fiscal policies. Focusing on post-WWII data of the US and the UK, oversizing of conventional test statistics is sufficiently severe in small samples to spuriously diagnose fiscal balances to accord with the IBC. Thus, for the most recent periods of international integration and globalization monitoring of fiscal sustainability by means of policy response regressions deserves size controlled econometric diagnostics.

3.5 Conclusions

For the cases of the US, the UK and Portugal, historic debt-to-GDP ratios might be best represented by means of ('near' to) unit root processes, while primary surpluses are stationary. Pre-tests indicate that the persistence of debt-to-GDP ratios joint with the correlation of debt and surplus innovations invoke marked size distortions of commonly applied t -tests. By means of Monte Carlo simulations, we estimate actual significance levels of up to 27.6% for t -tests with a 5% nominal level. Overall, our results reflect that diagnosing an IBC by means of policy response regressions is less conservative in comparison with unit root tests, but deserves careful treatment of type I error probabilities. Conditional on a 5% nominal significance level, decisions on sustainability of fiscal policies in the US, the UK and Portugal do not change for historical data covering more than a century, if Monte Carlo based critical values or

Bonferroni confidence bounds or Gaussian critical values are employed. While the former diagnostics respect the nominal significance level, Gaussian quantiles turn out to correspond with actual significance levels that may exceed their nominal counterpart by up to five times. Focusing on post-WWII subsamples we discover several cases where size distortions prevent a timely detection of US and UK fiscal imbalances if test decisions are conditioned on Gaussian critical values.

Size control in model based diagnosis of the IBC is essential for at least two reasons. First, eventual agreement of size distorted and size controlled inferential outcomes is always conditional on a specific level of nominal significance that has to be selected by the analyst. Second, in light of both intensified international economic integration and the recent debt crises in Europe, policy authorities have a natural interest in IBC diagnosis that conditions on more time local sample information covering only a few decades of fiscal data. Given slow convergence to asymptotic limits under regressor persistence, size controlled inferential tools are of utmost importance for the monitoring of fiscal sustainability. Recently, panel data approaches have been suggested to provide a suitable number of observations of recent fiscal figures due to an enlarged cross sectional dimension of data (Mendoza and Ostry; 2008; Ghosh et al.; 2013). Apart from pooled test regressions, in the panel context, Fisher type aggregation of p -values from single test regressions could be seen as a viable alternative for summarizing country specific evidence (Fisher; 1934). As a particular direction of future research, the implementation of Fisher tests deserves conditioning on correct rather than spurious, i.e., oversized, p -values.

Appendix to Paper 3

3.A Data sources

United States US data on public net debt are provided by Henning Bohn and can be downloaded from the web.²⁹ They are also used in Bohn (2008) and range from 1792 till 2011. We thank Henning Bohn for sending us the series of military outlays per GDP.

United Kingdom We construct a dataset on net public debt for the UK similar to Lukkezen and Rojas-Romagosa (2012). The sample ranges from 1830 to 2012. The public net debt for the period 1830 to 1979 has been collected from Mitchell (1988). More recent quotes have been drawn from the IMF database.³⁰ Revenues, expenditures and interest payments for the period 1830 to 1945 have been collected from Mitchell (1988). Respective quotes for periods after 1945 are from the Office for National Statistics. The time series of military expenditures from 1981 to 2012 is downloaded from the Stockholm International Peace Research Institute (SIPRI) database,³¹ and collected from Mitchell (1988) for the period 1830-1980. The nominal and real GDP are calculated by merging data of Mitchell (1988) from 1830 to 1947 (1954, respectively) and from the UK Office for Budget Responsibility (quotes until 2012).

Portugal The Portuguese data on gross public debt are downloaded from the website of Carlos Marinheiro.³² The sample ranges from 1852 to 2012 and has been analyzed in Marinheiro (2006). We thank Carlos Marinheiro for updating the dataset and adding a GDP deflator upon our request.

²⁹<http://econ.ucsb.edu/~bohn/>

³⁰<http://www.imf.org/external/pubs/ft/weo/2013/02/weodata/index.aspx>

³¹<http://www.sipri.org/databases>

³²http://www4.fe.uc.pt/carlosm/index_en.htm

3.B Unit root tests for non-debt components

Series	Sample	ADF	PP	DF _{GLS}	KPSS
United States					
\tilde{g}_t	1792-2011	-5.261***	-3.227**	-4.655***	0.833***
	1900-2011	-4.415***	-2.524	-3.780***	0.270
	1948-2011	-1.884	-1.449	-2.275**	0.843***
\tilde{y}_t	1792-2011	-8.691***	-5.872***	-8.593***	0.009
	1900-2011	-8.223***	-2.562	-6.895***	0.024
	1948-2011	-5.246***	-3.036**	-2.929***	0.196
United Kingdom					
\tilde{g}_t	1830-2012	-5.505***	-4.063***	-5.349***	0.216
	1900-2012	-4.543***	-3.326**	-4.275***	0.242
	1946-2012	-14.571***	-13.289***	-15.140***	0.877***
\tilde{y}_t	1830-2012	-8.325***	-5.041***	-7.984***	0.010
	1900-2012	-6.594***	-3.333**	-5.394***	0.017
	1946-2012	-4.792***	-2.908**	-3.148***	0.182
Portugal					
\tilde{g}_t	1852-2012	-7.937***	-11.542***	-4.086***	0.035
	1900-2012	-7.544***	-7.751***	-7.250***	0.016
	1945-2012	-4.981***	-4.025***	-5.018***	0.029
\tilde{y}_t	1852-2012	-8.431***	-8.625***	-1.468	0.015
	1900-2012	-6.392***	-5.932***	-4.711***	0.021
	1945-2012	-6.347***	-4.061***	-5.523***	0.069

Table 3.8: Unit root tests for non-debt components For further notes see Table 3.1.

Paper 4

Sustainability of European fiscal balances – Just a statistical artifact?

Malte Rengel

Abstract. The recent European debt crisis has emphasized the importance of a reliable assessment of the sustainability of fiscal balances. The literature on policy reaction functions interprets debt as sustainable if a rising debt-to-GDP ratio invokes a positive reaction in the primary surplus. In this study, I show in a panel framework that the persistence of debt-to-GDP ratios and the correlation of respective innovations with primary surpluses generate size distortions of t -tests for common policy reaction coefficients. An analysis of European data prior to 2009 reveals that these size distortions make the diagnosis of non-sustainable fiscal balances by means of conventional inference highly unlikely. In contrast, size corrected bootstrap based critical values allow for a timely detection of increased solvency risk.

4.1 Introduction

The current European debt crisis reinforces the academic, political and public attention to the solvency risk of nations. As established by the European Stability and Growth Pact, annual deficits of member countries should not exceed 3% of GDP and public debt should amount to 60% of GDP at maximum. However, present debt levels of European countries strongly exceed the targets and raise serious concerns about the resilience of the monetary union. For instance, the debt burden of Greece reached 170% of GDP in 2013. Aizenman, Hutchison and Lothian (2013) argue that the European debt crisis is one of the strongest risks for a downturn of the world economy after the global financial crisis.¹

In this study, I analyze the so-called model based sustainability of fiscal balances of 14 European countries before the European debt crisis manifested in 2009. I reveal that type I error probabilities of conventional sustainability tests exceed the

¹ For a detailed discussion on the European debt crisis and a literature overview the reader may consult Lane (2012).

nominal benchmark and that fiscal policies are spuriously classified as sustainable. This ex-post analysis cautions against potential misinterpretations in assessing the fiscal soundness of countries and provides methods for a more reliable policy monitoring.

Bohn (1998) establishes inference on policy reaction functions to diagnose if fiscal policy is designed in a sustainable manner. These functions relate the primary (non-interest) surplus to lagged public debt (both measured as shares of GDP). By means of a stochastic general equilibrium model, Bohn (1995, 1998) derives an intertemporal budget constraint (IBC) which holds if a government raises the primary surplus in response to an increase in the lagged debt-to-GDP ratio.² Estimated policy reaction coefficients are typically found to be significantly positive. For instance, Bohn (1998, 2008) argues in favor of a positive and significant reaction coefficient for historical US fiscal data starting in 1916 or 1792, respectively. Focusing on European economies, Greiner et al. (2007) diagnose sustainability of German, French, Italian and Portuguese fiscal balances. While these studies are settled in a single-equation framework, panel data approaches have recently been employed to jointly assess fiscal soundness of distinct countries. Adding cross sectional information allows to focus on more recent data. Thus, panel based inference allows for active policy monitoring. Mendoza and Ostry (2008) diagnose positive and significant common reaction coefficients for a pooled set of 22 industrial and a pooled set of 34 emerging economies observed from 1980 to 2005. Likewise, joining the two subsets yields a significantly positive coefficient. Conditional on a panel of 23 advanced economies observed from 1980 to 2007, Ghosh et al. (2013) find nonlinear reactions of fiscal policies to increased debt-to-GDP ratios. The policy response exhibits an inverted U-shape and is positive for debt-to-GDP ratios between 0.4 and 1.5.³

While the panel framework is appropriate to test for pre-crisis sustainability of European fiscal policies, it is worthwhile to consider at first particular statistical

² A deterministic IBC can be addressed alternatively to the model based IBC. Whether fiscal figures accord with the deterministic IBC can be investigated by testing against a unit root in the debt-to-GDP ratio, or for a cointegration relation between government revenues and expenditures. Conclusions from these tests are often ambiguous. The reader may consult Bohn (2008) for a discussion. These approaches are employed to investigate long-run sustainability of fiscal balances. In comparison, for instance, Abeysinghe and Jayawickrama (2013) consider short-run sustainability.

³ Another strand of literature focuses on the sustainability of fiscal policy on the state level. Claeys, Ramos and Suriñach (2008) find a positive response of states primary surpluses to increasing debt-to-GDP ratios in the US, but not for Germany. Potrafke and Reischmann (2014) emphasize the influence of fiscal transfers on the assessment of states' fiscal soundness.

characteristics of debt-to-GDP ratios and primary surpluses. It is extensively discussed in the literature that debt-to-GDP ratios may be characterized as (‘near’ to) unit root processes (Hamilton and Flavin; 1986; Kremers; 1988; Wilcox; 1989). Antonini et al. (2013) argue that debt-to-GDP ratios of 10 European countries exhibit stochastic trends. They apply single-equation and panel unit root tests. In contrast to debt-to-GDP ratios, primary surpluses are commonly diagnosed to be stationary processes of low persistence (Bohn; 1998, 2008; Marinheiro; 2006). In general, regressions that are characterized by transitory left hand side variables and persistent right hand side variables are poorly balanced. So far, eventual effects of this property of test regressions have been neglected in the context of policy reaction functions. In addition, the innovations of primary surpluses and debt-to-GDP ratios are in general negatively correlated. If the primary surplus of a period is large, the debt at the end of this period tends to be low. Woodford (1998) points to these correlations in the ‘fiscal theory of the price level’.

These two characteristics are well recognized in the related finance literature on predictive regressions. This literature relates future returns to lagged persistent valuation ratios, e.g., the price-to-dividend ratio.⁴ In single-equation predictive regressions, Stambaugh (1999), Lewellen (2004) and Campbell and Yogo (2006) point out that estimates of slope coefficients are upward biased and the size of corresponding t -tests is overrated.⁵ Using Bonferroni and Monte Carlo techniques, Herwartz and Rengel (2014) diagnose that the size of t -tests in single-equation policy reaction functions is up to five times larger than its nominal reference for US, UK and Portuguese data covering almost two centuries. These size distortions bear the risk of classifying fiscal figures falsely as sustainable more often than prespecified by the type I error probability of the test.

In the panel framework, such biases have the potential to equally invalidate inference on predictive regressions and policy reaction functions. In the finance context, Hjalmarsson (2008, 2010) demonstrates that the within-transformation employed in predictive regressions to account for individual cross sectional characteristics induces a so-called second-order bias. This bias enforces the bias of slope estimates and the excess of type I error probabilities of t -tests observed in single-equation

⁴ Mankiw and Shapiro (1985, 1986) hint at similar patterns in measures which are applied to analyze the permanent income hypothesis.

⁵ In general, predictive regressions have the form $y_t = \theta_0 + \theta_1 z_{t-1} + \epsilon_t$, where the regressor follows the dynamic process $z_t = \gamma_0 + \gamma_1 z_{t-1} + \nu_t$ and is nearly integrated if γ_1 is close to unity. The (upward) bias of $\hat{\theta}_1$ and the size distortion of the corresponding t -test become more pronounced the stronger the (negative) correlation between ϵ_t and ν_t is and the more persistent the movements of z_t are.

models. He diagnoses by means of a simulation study an actual size of up to 60% of a t -test with nominal 5% reference and provides empirical examples of spuriously inferred predictive evidence of valuation ratios by pooling country specific financial market data. Similarly, Pastor, Stambaugh and Taylor (2014) find a second-order bias in evaluating the returns to scale of active mutual fund management. Baglan and Yoldas (2013) account for the persistence of the debt-to-GDP ratio when analyzing its predictive content for macroeconomic activity. A significant negative relation between economic activity and the debt-to-GDP ratio cannot be established when relying on robust estimation techniques and data of 20 advanced economies ranging until 2008.

In sum, whether the frequent non-rejection of model based sustainability is just a statistical artifact caused by oversized t -tests is a question of general interest. I test for the sustainability of fiscal balances of 14 European countries in a panel ranging from 1980 until the beginning of the European debt crisis in 2009. First, I demonstrate by means of a Monte Carlo study that a severe upward bias of the estimated joint policy reaction coefficient and a size distortion of the corresponding t -test arise for empirically reasonable degrees of persistence of debt-to-GDP ratios and correlation of respective innovations with primary surpluses. In general, the bias and excess of type I error probabilities reinforce if the cross sectional dimension of the panel increases. Second, I provide bias and size corrected inference on the sustainability of pre-crisis European fiscal policies by means of wild bootstrap techniques. Common policy reaction coefficients are spuriously found to be positive with 5% nominal significance if conventional asymptotic critical values are applied. In contrast, based on size corrected bootstrap based critical values, tests strongly support the conclusion of non-sustainability. Excluding Greek data, type I error probabilities are still overrated, but there is weak evidence in favor of sustainable fiscal policies. In a nonlinear version of the policy reaction function, I reveal a significant upward correction of European primary surpluses only at high levels of debt-to-GDP ratios. In consequence, conventional estimators spuriously indicate sustainable fiscal balances in Europe until 2009, while appropriate inferential techniques allow a timely disclosure of the increased solvency risk. The proposed size corrected inference is helpful to assess the fiscal soundness of future fiscal balances.

Section 4.2 briefly sketches the policy reaction function in a panel framework, introduces the dataset and considers persistence properties of European primary surpluses and debt-to-GDP ratios. A Monte Carlo study in Section 4.3 investigates the bias of the conventional fixed effects estimator and size distortions of t -tests.

In Section 4.4, I describe a wild bootstrap approach providing size controlled inference and test for fiscal sustainability in distinct specifications of the policy reaction function. Section 4.5 concludes. An appendix comprises the data sources and an examination of the wild bootstrap approach and alternative size corrected estimation techniques proposed in the literature.

4.2 A policy reaction function for European fiscal data

In this section, I first describe the panel data framework for policy reaction functions. Then, the fiscal data of 14 European countries are introduced and their persistence characteristics are discussed.

4.2.1 The policy reaction function in the panel data framework

Mendoza and Ostry (2008) and Ghosh et al. (2013) pool fiscal data of several countries and employ panel regressions to test for fiscal soundness. While single-equation policy reaction regressions are commonly based on post-WWII data or historical series covering up to 200 years (e.g., Bohn; 2008), debt dynamics in more recent and shorter time spans are investigated by means of panel data techniques. Due to the enlarged cross sectional information it is possible to investigate if fiscal policies in more recent years accord with the IBC.⁶ Thus, the panel framework offers the possibility of timely policy monitoring.

A potential shortcoming of the panel data approach is the relative restrictive assumption of a common policy reaction parameter.⁷ On the one hand, the common coefficient is estimated more precisely than time-series estimates of country specific parameters if this assumption holds. On the other hand, the panel estimator provides a biased measure of heterogeneous parameters if the assumption does not hold. Hjalmarrsson (2010) puts forward several arguments for inferring common coefficients even if their assumption is restrictive. Accordingly, common parameters can be interpreted as capturing the average influence of regressors even if they may be biased estimates of heterogeneous influences. From the perspective of policy monitoring, such an interpretation is of immediate interest since it allows to analyze the ‘average’ relation between primary surpluses and lagged debt-to-GDP ratios of several economies simultaneously. Conditional on this joint analysis, a

⁶ This advantage of panel data is similarly described by Frankel and Rose (1996) when analyzing, e.g., the implications of Purchasing Power Parity.

⁷ It is worthwhile to notice that the parameters capturing the effects of additional control variables are likewise treated as homogeneous.

detailed characterization of country specific data can be conducted subsequently. Pure time series based estimates of slope coefficients tend to be relatively noisy if the sample size is small. Thus, the interpretation of a common coefficient may be advantageous in light of the natural variance-bias trade-off.⁸ The panel policy reaction function reads as

$$s_{i,t} = \rho d_{i,t-1} + \psi' x_{i,t} + \alpha_i + \gamma_t + u_{i,t}, \quad (4.1)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T_i$. The primary surplus measured as share of GDP is denoted by $s_{i,t}$ and the lagged debt-to-GDP ratio by $d_{i,t-1}$. The policy reaction coefficient ρ measures the strength of the adjustment of the primary surplus to variations of the debt-to-GDP ratio and, thus, is of major interest for the assessment of model based sustainability. If it is significantly positive, economies accord with the IBC (Bohn; 1998).

Country specific fixed effects are captured by α_i and distinct control variables are comprised in the $m \times 1$ dimensional vector $x_{i,t}$. Bohn (1998) highlights the importance of these control variables for the estimation of ρ . Year dummies γ_t absorb the effects of time specific events influencing European countries equally as, for instance, the breakdown of the former Soviet Union in 1991 or the introduction of the European Stability and Growth Pact in 1997.⁹

4.2.2 The dataset and the empirical implementation

The panel dataset includes 14 European economies: Austria, Belgium, Denmark, Finland, France, Germany, Greece, Ireland, Italy, the Netherlands, Portugal, Spain, Sweden and the United Kingdom.¹⁰ Although Denmark, Sweden and the UK are not members of the Eurozone, they are subject to the requirements of the European Stability and Growth Pact. Thus, I refrain from restricting the sample to members of the Eurozone in the baseline specification to enlarge the information set. Economies that have become members of the European Union during the 2000s are excluded from the analysis. The annual sample is unbalanced and generally

⁸ To investigate the sensitivity of the results presented below towards country heterogeneity, I restrict the analysis to fiscal figures of Eurozone countries or exclude Greek data in Section 4.4.3.

⁹ The literature on macroeconomic panels considers more general forms of cross sectional dependence. For instance, Pesaran (2006) and Hjalmarrsson (2010) propose a factor structure for innovations. However, these extensions require the estimation of several latent factors and cross sectional specific loadings which is infeasible if N is small (Pesaran; 2006; Hjalmarrsson; 2010). Thus, I restrict the analysis to year dummies and employ a wild bootstrap to account for potential heteroscedasticity and cross sectional dependence.

¹⁰ A detailed description of the data sources is provided in Appendix 4.A.

ranges from 1980 to 2008. As discussed by Lane (2012), the European debt crisis manifested in 2009 when the newly elected Greek government corrected the expected deficit from 6% of GDP to 12.7%. Moreover, large spreads between Irish, Portuguese, Spanish and Italian ten year sovereign bonds and the German benchmark have demonstrated the enlarged solvency risk from this year onward. The focus of this work is to investigate if European fiscal policies have to be classified as non-sustainable in the pre-crisis period. Estimates of policy reaction coefficients from data ranging until 2013 are briefly described in Section 4.4.5.

The dataset contains series of the public debt and the primary balances as shares of GDP. Additional non-debt determinants of primary surpluses collected in $x_{i,t}$ are identical to those suggested by Mendoza and Ostry (2008): business cycle variation, temporary government spending, CPI inflation and current account balances as shares of GDP. The former two measures are also employed by Bohn (1998, 2008) and motivated in light of the tax smoothing model of Barro (1979). As common in the literature, business cycle variation is measured as deviation of log real GDP from its Hodrick-Prescott (HP) trend calculated with smoothing parameter 100 (e.g., Bohn; 2008; Mendoza and Ostry; 2008; Ghosh et al.; 2013). Similarly, temporary government spending is measured as the difference between log real expenditures and the HP trend.¹¹ Inflation possibly influences decisions of debtors and creditors. By employing series of current account balances, I account for so-called twin-deficits. External debt tends to be high, when public debt is large (Mendoza and Ostry; 2008; Hartmann and Herwartz; 2012).

Figure 4.1 displays the primary surplus and the debt-to-GDP ratio as cross sectional average over the 14 European countries between 1980 and 2013. The average primary surplus varies around zero and peaks in 2000 at the maximum of 4% of GDP, while it reaches its minimum around -5% of GDP in 2010. In 2009, the surplus is characterized by a strong downward correction. In contrast, the average debt-to-GDP ratio is a persistent series increasing from around 40% of GDP in 1980 to 62% in 2008 and to over 90% in 2013.

¹¹ Bohn (1998) applies military expenditures as a measure of temporary government spending in single-equation regressions for US data. Mendoza and Ostry (2008) discuss that military expenditures are not the dominant source of temporary government spending during the last 30 years and propose the use of the HP-detrended series. Series of GDP and expenditures are at maximum available since 1960 and used to calculation HP trend measures.

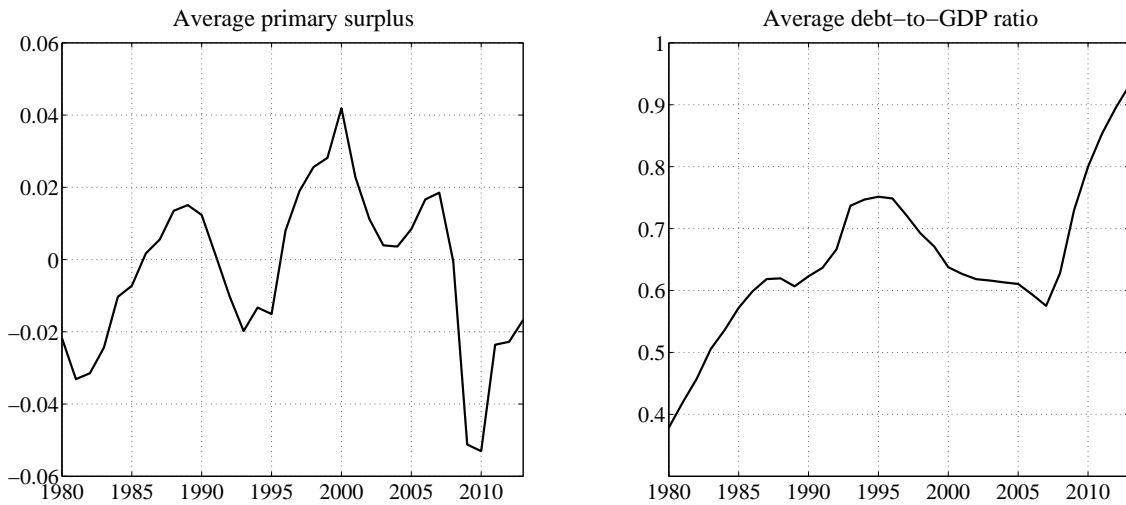


Figure 4.1: Cross sectional averages of primary surpluses and debt-to-GDP ratios
 The left hand side panel displays for each period the average primary surplus relative to GDP calculated over the corresponding series of the 14 European countries. The right hand side panel displays the cross sectional average of debt-to-GDP ratios.

4.2.3 Persistence characteristics

I consider several panel unit roots tests (PURT) to explore the persistence characteristics of the debt-to-GDP ratios, the primary surpluses and the additional control variables. Test results are summarized in Table 4.1. Except for the LLC statistic (Levin, Lin and Chu; 2002), all tests hint at a panel unit root in the debt-to-GDP ratios. In general, PURTs prevent multiple testing and are more powerful than univariate tests for nearly integrated series (Herwartz, Siedenburg and Walle; 2014). However, also PURTs suffer from low power for selected small samples. For instance, Herwartz and Siedenburg (2008) show by means of a Monte Carlo study that PURTs have low power when $N = 10$ and $T = 30$. Given that these dimensions are comparable to those of the dataset on European fiscal figures, the non-rejection of the null hypothesis of stochastically trending debt-to-GDP ratios may be caused by the low power of PURTs. In consequence, debt-to-GDP ratios might be best represented by ('near' to) unit root processes. Hence, the following discussion can be interpreted as being settled in a local to unit root framework that allows for both unit root and 'near' to unit root processes (Campbell and Yogo; 2006).

In contrast to debt-to-GDP ratios, primary surpluses can be classified as stationary processes of low persistence. The null hypothesis of a panel unit root can be rejected by means of all considered tests at conventional significance levels. In the baseline specification, I condition the policy reaction in (4.1) on the two control variables employed by Bohn (1998). These are business cycle variation and temporary

	LLC	IPS	Fisher-ADF	BR	HS
Debt	-2.275**	-0.248	23.866	1.540	1.460
Primary surplus	-2.931***	-3.122***	70.140***	-1.810**	-1.299*
Business cycle variation	-4.894***	-8.420***	160.444***	-3.704***	-1.375*
Temporary gov. spending	-3.312***	-5.817***	104.728***	-2.189**	-1.984**
Inflation	-8.561***	-7.785***	143.680***	0.018	0.904
Current account balance	0.499	-0.193	35.731	-0.938	-0.705

Table 4.1: Panel unit root tests Series of debt, primary surpluses and current account balances are measured in terms of GDP. For the test proposed by Levin et al. (2002), which is denoted by LLC, the Bartlett kernel with a Newey-West bandwidth selection procedure is applied. Results of the Im, Pesaran and Shin (2003) test are provided in the column labeled IPS. In the Fisher-type test based on ADF statistics denoted by Fisher-ADF the inverse Chi-square statistic proposed by Choi (2001) is employed. The column denoted by BR comprises a version of the test proposed by Breitung (2000) and Breitung and Das (2005) which is robust to cross sectional dependence. To account for serial correlation, I apply a prewhitening with one lag for this statistic. The column denoted by HS documents the results from the test proposed by Herwartz and Siedenburg (2008) which is robust to heteroskedasticity and cross sectional dependence as shown in Herwartz et al. (2014). I apply a prewhitening with one lag for the calculation of the statistic. All tests include individual intercepts and the lag length is set to unity. For debt-to-GDP ratios, test regressions include deterministic intercepts and trends. Significance at the 1%, 5%, 10% level is indicated by ***, **, *.

government spending. Tests indicate both measures to be stationary. For inflation rates, the unit root null hypothesis is rejected with 1% significance by means of the LLC, the IPS (Im et al.; 2003), or the Fisher-ADF (Choi; 2001) statistic. In contrast, conditioning on tests proposed by Breitung (2000)/Breitung and Das (2005, BR) and Herwartz and Siedenburg (2008, HS), this null hypothesis cannot be rejected with 10% significance. Series of current account balances are indicated to be non-stationary conditional on all tests. In sum, panel unit root tests hint at a poorly balanced policy reaction regression which implies that transitory primary surpluses are related to persistent debt-to-GDP ratios.¹² Next, it is discussed how these properties affect the outcomes of panel regressions.¹³

¹²Regarding the poor balance of policy reaction functions, I test for a cointegration relation between (lagged) debt-to-GDP ratios and current account balances as two potentially non-stationary right hand side variables. I employ the panel cointegration test proposed by Westerlund (2007) which is robust to heteroscedasticity and cross sectional dependence. Fixed effects, one lead and one lag of differentiated variables are employed in test regressions. Critical values are based on a bootstrap procedure with 5,000 replications. The null hypothesis of no cointegration cannot be rejected at common significance levels. Moreover, no cointegration relation is found for (lagged) debt-to-GDP ratios and inflation and in a trivariate system with these two measures and current account balances.

¹³In case of adding a deterministic trend to the test regressions or excluding a deterministic trend for debt-to-GDP ratios, the test results remain qualitatively identical. Moreover, the test results are not sensitive to restricting the sample to economies of the Eurozone or excluding Greek series.

4.3 Size distortions in pooled policy reaction functions

This section describes the rationale for the bias of the estimated policy reaction coefficients and the size distortions of conventional t -tests in the panel framework. The magnitude of this bias and the size distortion is evaluated by means of a Monte Carlo study.

4.3.1 The second-order bias

In the finance literature, it has been well established that slope coefficients in single-equation predictive regressions, that relate transitory returns to lagged persistent valuation ratios, are upward biased and corresponding t -tests are oversized if the innovations of returns and regressors are negatively correlated (e.g. Stambaugh; 1999; Campbell and Yogo; 2006). Recently, Herwartz and Rengel (2014) argue that such distortions equally occur in single-equation policy reaction regressions. Intuitively, the debt level tends to be low if the primary surplus is large and, thus, the innovations of both measures are likely to be negatively correlated.

Regarding the policy reaction function in (4.1), the estimate of the common slope coefficient $\hat{\rho}$ can be interpreted as the weighted average of country specific reaction parameters and, thus, carries biases from single-equation estimates (Pesaran and Smith; 1995; Pastor et al.; 2014). Additionally, there is a methodological cause for the bias of the conventional panel estimator. This estimator is commonly based on within-transformed data to avoid an omitted variable bias. As emphasized by Hjalmarsson (2008, 2010) and Pastor et al. (2014), this within-transformation prevents the omitted variable bias, but induces a so-called second-order bias if the lagged regressor is persistent and its respective innovations are correlated with those of the left hand side variable. This second-order bias generates even more severe upward biases of slope coefficients and overrated type I error probabilities in panel based inferential diagnostics on predictive regressions than those derived from single-equation regressions. To illustrate this argument, it is helpful to consider the policy reaction function in (4.1) and to formalize an autoregressive process for the debt-to-GDP ratio

$$s_{i,t} = \rho d_{i,t-1} + \psi' x_{i,t} + \alpha_i + \gamma_t + u_{i,t}, \quad (4.2)$$

$$d_{i,t} = \beta d_{i,t-1} + \phi_i + e_{i,t}, \quad (4.3)$$

for $i = 1, \dots, N$ and $t = 1, \dots, T_i$, where the correlation structure of primary

surpluses' and debt-to-GDP ratios' innovations reads as

$$\text{Cov}(u_{i,t}, e_{i,t}) = \Sigma = \begin{pmatrix} \sigma_u^2 & \sigma_{ue} \\ \sigma_{ue} & \sigma_e^2 \end{pmatrix}. \quad (4.4)$$

If the within-transformation is applied to estimate (4.2), the transformed lagged debt-to-GDP-ratio $\tilde{d}_{i,t-1} = d_{i,t-1} - 1/T_i \sum_{t=1}^{T_i} d_{i,t-1}$ in period $t-1$ is correlated with $u_{i,\tau}$, for $\tau = t, \dots, T_i$. This correlation causes the second-order bias.¹⁴

4.3.2 A Monte Carlo study on biases and size distortions

I employ a Monte Carlo study to demonstrate the extend of the upward bias of the estimated common policy reaction parameters for European fiscal figures and the size distortions of the corresponding t -tests. It is revealed how these distortions depend on the persistence of the debt-to-GDP ratios captured by β , the correlation between $u_{i,t}$ and $e_{i,t}$ denoted by $\delta = \sigma_{ue}/(\sigma_u\sigma_e)$, and the cross sectional (N) and time dimension (T) of the data. Series of primary surpluses and debt-to-GDP ratios are simulated from the model in (4.2) and (4.3) under the null hypothesis of non-sustainability with $\rho = 0$. In the baseline specification of (4.2), business cycle variation and temporary government spending are employed as control variables $x_{i,t}$ (Bohn; 1998). These measures are treated as given and set to the empirical observations. Simulation parameters α_i , γ_t , $\psi' = (0.493, -0.568)$ and $\sigma_u = 0.017$ are set to empirical values which I estimate for (4.2) conditional on pre-2009 observations of European fiscal data.

To determine the coefficients necessary to simulate from (4.3), I employ country wise single-equation regressions and estimate a common slope coefficient β by averaging over the respective individual estimates. This mean group procedure has been proposed by Pesaran and Smith (1995) and avoids the bias of fixed effects estimators in dynamic panel models (Nickell; 1981).¹⁵ The parameters $\beta = 0.926$, $\sigma_e = 0.039$ and country specific fixed effects ϕ_i estimated for the European fiscal data are employed in the simulation. The common correlation coefficient $\delta = -0.294$ captures the relation between the innovations in primary surpluses and debt-to-GDP ratios.

¹⁴For an extensive explanation, the reader may consult Hjalmarrsson (2010) and Pastor et al. (2014). Noteworthy, the cause of the second order-bias is closely related to the so-called Nickell (1981) bias in dynamic panel models.

¹⁵The conventional mean group estimator avoids by design the inclusion of time dummies. Moreover, the single-equation estimates of autoregressive coefficients are potentially downward biased in small samples (Stambaugh; 1999). Thus, the simulated bias of $\hat{\rho}$ might be understated. An alternative estimation procedure is, for instance, the post-estimation bias correction proposed by Kiviet (1995).

Artificial realizations of residuals $u_{i,t}$ and $e_{i,t}$ are drawn from a joint normal distribution with zero mean and the covariance matrix Σ that contains the described simulation parameters. Based on these innovations, $N = 14$ series of debt-to-GDP ratios of length $t = 1, \dots, 29$ are generated from (4.3). Respective processes are initialized at random draws from the country specific empirical series. The simulated panels include missing values where empirical European fiscal figures are lacking. Simulations results are based on 10,000 replications.

The left hand side panel of Figure 4.2 displays the simulated distribution of $\hat{\rho}$. The figure highlights a severe upward bias of the estimated joint policy reaction coefficient. The average $\hat{\rho}$ is 0.014 units larger than the reference $\rho = 0$. Even though this number may appear small, it represents a substantial fraction of empirically estimated coefficients documented in the next section. The right hand side panel of Figure 4.2 displays the simulated distribution of the corresponding right hand sided heteroskedasticity and autocorrelation robust t -statistic along with the theoretical standard Gaussian asymptotic distribution (solid black line).¹⁶ The actual size of the right hand sided t -test with 5% nominal reference amounts to 28.1%.

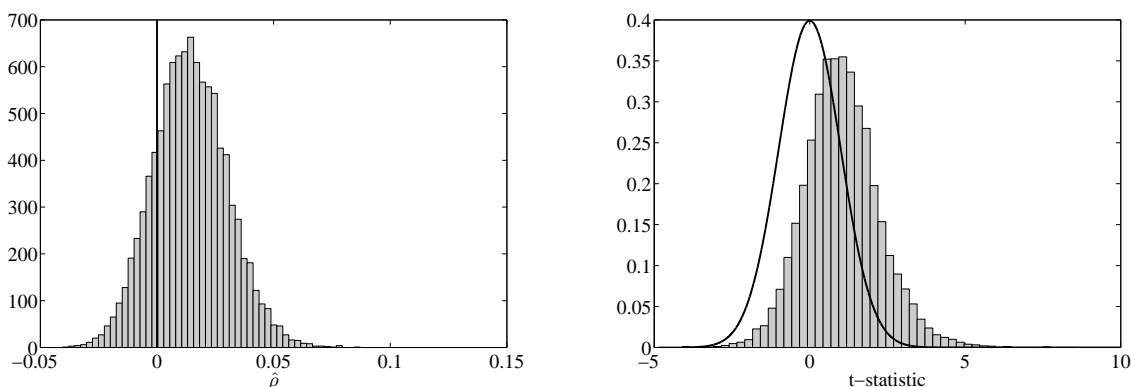


Figure 4.2: Simulated distributions of the estimated policy reaction coefficient and the t -statistic The left hand side panel displays the histogram of $\hat{\rho}$. The black solid line indicates $\rho = 0$. The histogram of the simulated right hand sided heteroskedasticity and autocorrelation robust t -statistics displayed in the right hand side panel is accompanied by the standard Gaussian asymptotic distribution (the black solid line). The histogram is rescaled to facilitate the comparison with the asymptotic distribution. Figures are based on 10,000 Monte Carlo replications with $\delta = -0.294$ and $\beta = 0.926$. Business cycle variation and temporary government spending are employed as control variables. The simulated panel lacks observations where its empirical counterpart includes missing values with $N = 14$ and $T_i = 28$ at maximum.

Figure 4.3 illustrates how the bias of the estimated common policy reaction pa-

¹⁶Testing for the sustainability of European fiscal policies heteroscedasticity and autocorrelation robust standard errors are applied throughout in Section 4.4 to account for empirically diagnosed serial correlation. Empirical rejection frequencies of non-robust t -tests are similar.

parameter and the excess of the type I error probability vary for distinct values of $\beta \in [0, 1]$ and $\delta \in [-1, 0]$, where the negative range of δ is motivated from empirical estimates. Without loss of generality, debt-to-GDP ratios are initialized at zero and country specific fixed effects β_i are set to zero. Otherwise, the process of the debt-to-GDP ratio is a random walk with drift for $\beta = 1$, which would not be comparable to processes conditioned on $\beta < 1$. The upper panel of Figure 4.3 displays the average of the estimated coefficients $\hat{\rho}$ across simulations for distinct values of β and δ , where the true reference is $\rho = 0$. The lower panel documents the simulated size of the right hand sided t -test with 5% nominal reference. Both, the upward bias and the size of the t -test are positive functions in β and δ in absolute terms.

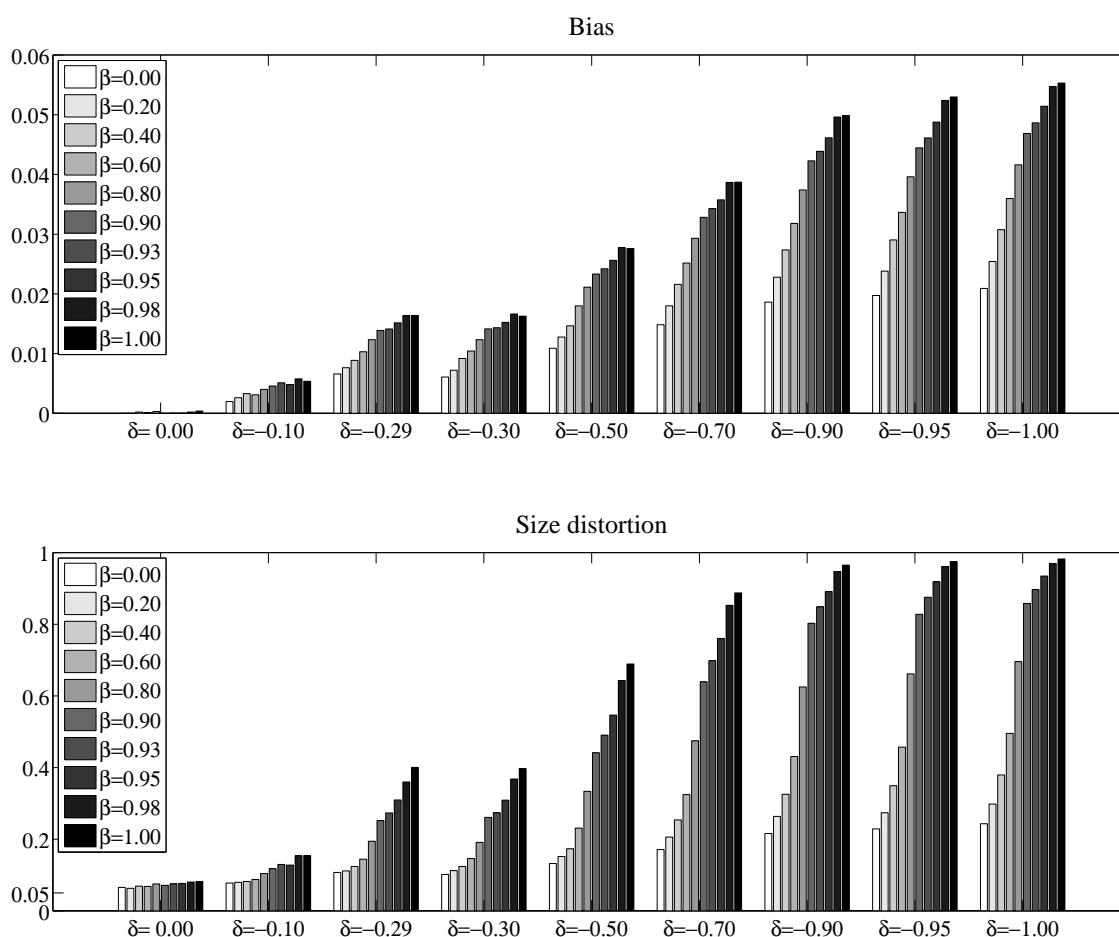


Figure 4.3: Biases and size distortions for distinct values of δ and β The upper panel displays the average of coefficient $\hat{\rho}$ estimated from 10,000 panel datasets which are simulated by means of the Monte Carlo procedure under the null hypothesis of $\rho = 0$. The lower panel documents the actual size of the right hand sided heteroskedasticity and autocorrelation robust t -test regarding $\hat{\rho}$ with 5% nominal reference. Business cycle variation and temporary government spending are employed as control variables. The simulated panel lacks observations where its empirical counterpart includes missing values with $N = 14$ and $T_i = 28$ at maximum.

I further investigate the dependence of the bias and the size distortion on the cross sectional dimension N and the time series dimension T of the panel dataset by means of a modified simulation study. The cross sectional dimension N can be larger than 14 and, thus, the effects of business cycle variation and temporary government spending which are otherwise set to their actual realizations are neglected. Similarly, time dummies are not considered since the time dimension T can be larger than that of the European data. Moreover, panels do not include missing values. Coefficients α_i are randomly drawn from a uniform distribution ranging on the interval determined by empirical estimates. Remaining simulation parameters correspond to those presented before. The persistence parameter of debt-to-GDP ratios is set to $\beta = 0.926$. Given that this value is considerably distinct from unity, coefficients ϕ_i are not set to zero but are drawn from a uniform distribution on the interval determined by the empirical estimates. Moreover, processes of debt-to-GDP ratios are initialized at randomly selected observations of the pooled empirical series. The upper panel of Figure 4.4 displays the average estimate of ρ for distinct values of N and T , while the lower panel documents the estimated size of the corresponding right hand sided t -test with 5% nominal reference.

If T increases, the bias and the excess of the type I error probabilities decrease. This observation has been equally found in single-equation frameworks (Stambaugh; 1999). The value of N leaves the upward bias of the estimated policy reaction coefficient unaffected. Intuitively, this coefficient represents the average of country specific parameters and, as a result, is unaffected by N , since simulation parameters for cross sections are identical. In contrast, the excess of the test size accumulates with rising N . If the set of countries becomes larger, the distribution of $\hat{\rho}$ tightens, at a value to the right of $\rho = 0$. In consequence, the probability mass to the left of $\rho = 0$ shrinks and the size of the right hand sided t -test increases.¹⁷ Even if macroeconomic panels are commonly characterized by a rather small cross sectional dimension, simulation results for $N = 200$ can be interpreted as a cautionary note. The actual size of the t -test with 5% nominal reference approaches 100%.¹⁸

In sum, conventional inferential diagnostics turn out to be unsuitable for policy monitoring of fiscal soundness due to overrated type I error probabilities. In the case of the European debt crisis, oversized t -tests have potentially prevented an early detection of the increased solvency risk.

¹⁷In Appendix 4.B, Figure 4.6 illustrates this argument for the case of $N = 200$ and $T = 30$.

¹⁸Pastor et al. (2014) observe similar magnitudes of type I errors probabilities in a panel with $N = 300$ and $T = 100$ in the context of funds performance and returns to scale.

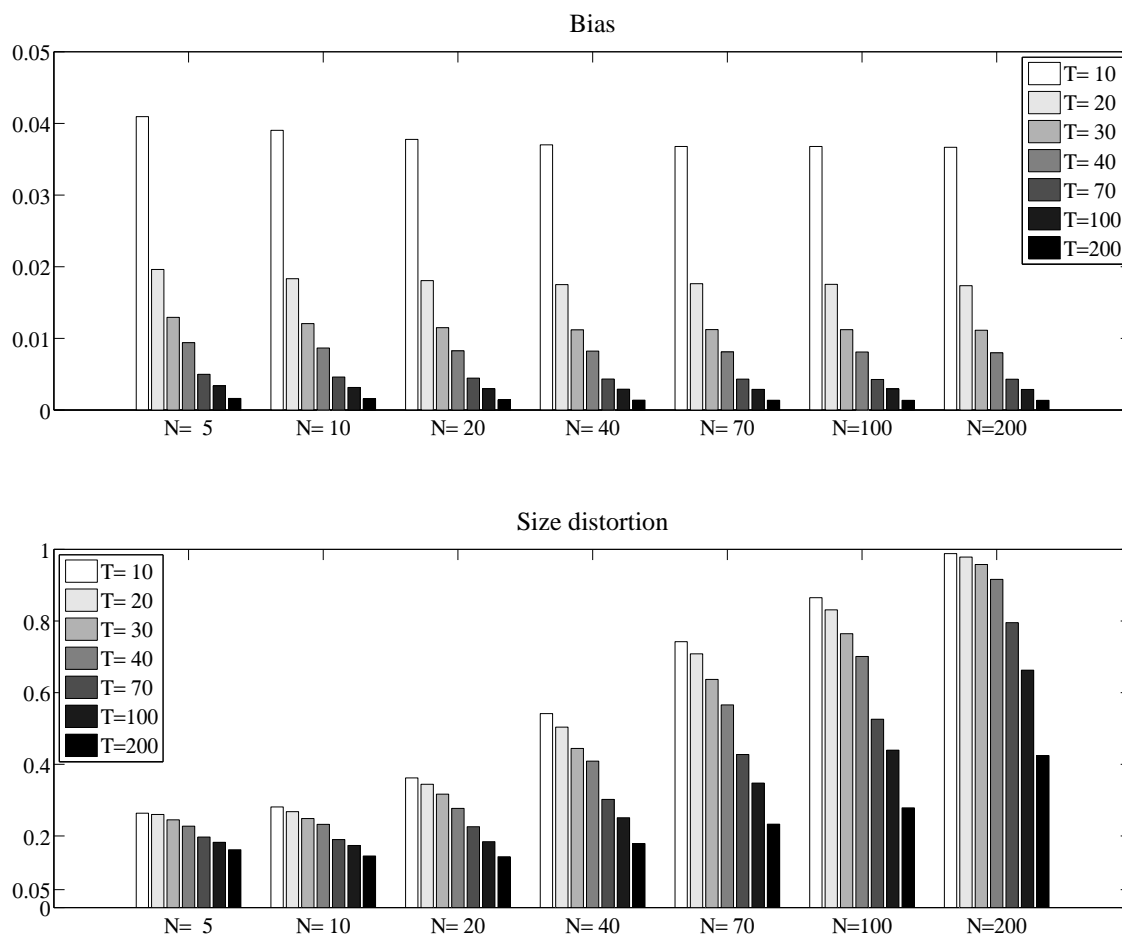


Figure 4.4: Biases and size distortions for distinct values of N and T The upper panel displays the average of coefficient $\hat{\rho}$ estimated from 10,000 panel datasets which are simulated by means of the Monte Carlo procedure under the null hypothesis of $\rho = 0$. The lower panel documents the actual size of the right hand sided heteroskedasticity and autocorrelation robust t -test regarding $\hat{\rho}$ with 5% nominal reference. Simulation parameters are $\delta = -0.294$ and $\beta = 0.926$. Since the cross sectional sample size N and the time dimension T varies additional control variables are neglected. Simulated samples do not include missing values.

4.4 Size controlled inference on fiscal sustainability in Europe

Next, I propose a bootstrap technique that allows for size controlled inference. This procedure is used to test for the sustainability of European fiscal balances before 2009 in linear policy reaction functions. Subsequently, I consider nonlinear extensions. The section concludes with a brief analysis of data until 2013.

4.4.1 A wild bootstrap approach

I perform a wild bootstrap approach to provide size controlled inference on fiscal figures from 14 European countries in the panel data framework.¹⁹ The Monte Carlo study in Section 4.3.2 was helpful to assess the magnitude of the bias of estimated policy reaction parameters and the size distortions of corresponding t -tests for distinct degrees of persistence and correlation of respective innovations. However, it neglects a typical characteristic of cross-country panel data. Open economies are affected by common international shocks. Thus, innovations in cross country panel regressions are cross sectionally dependent (Gengenbach, Palm and Urbain; 2010). While, the time dummies formalized in the policy reaction function (4.2) capture a part of the common influences, the wild bootstrap allows to account for heteroscedasticity and more general forms of cross sectional dependence (Herwartz and Siedenburg; 2008). I extend this approach to simultaneously account for the correlation between $u_{i,t}$ and $e_{i,t}$. Artificial innovations for $i = 1, \dots, N$ and $t = 1, \dots, T_i$ are generated from

$$u_{i,t}^* = \eta_t \hat{u}_{i,t}, \quad (4.5)$$

$$e_{i,t}^* = \eta_t \hat{e}_{i,t}, \quad (4.6)$$

where $\eta_t \sim N(0,1)$ and independent from the data. Concentrating exclusively on (4.5), multiplying the estimated innovations $\hat{u}_{i,t}$ with draws from the Gaussian distribution η_t reproduces patterns of heteroscedasticity in bootstrap innovations $u_{i,t}^*$ (Wu; 1986; Liu; 1988). Likewise, $u_{i,t}^*$ reflects the cross sectional dependence by using the same draw η_t for all cross sectional units (Herwartz and Siedenburg; 2008). Additionally, I reproduce the correlation between $u_{i,t}$ and $e_{i,t}$ by employing η_t in both, $u_{i,t}^*$ and $e_{i,t}^*$.²⁰ Conditional on bootstrap innovations, series of primary surpluses are generated from (4.2) where simulation parameters are set to

¹⁹In single-equation predictive regressions and policy reaction regressions alike, Monte Carlo and bootstrap based critical values turned out to be helpful to assess the significance of slope coefficients (Goyal and Welch; 2008; Herwartz and Rengel; 2014).

²⁰To investigate if the wild bootstrap procedure is able to replicate the cross sectional correlation of empirical innovations, I contrast the cross-correlation matrix of $\hat{u}_{i,t}$ with the average cross-correlation matrix of bootstrap innovations $u_{i,t}^*$ in Appendix 4.C. Moreover, I display the country specific correlations between $\hat{u}_{i,t}$ and $\hat{e}_{i,t}$ and compare them with the corresponding average quotas for $u_{i,t}^*$ and $e_{i,t}^*$. In sum, the bootstrap versions of these quantities are very close to the empirical counterparts. Alternatively to the wild bootstrap technique, I employ the Monte Carlo approach sketched in Section 4.3.2 and a residual based bootstrap drawing pairs of pooled series of $\hat{u}_{i,t}$ and $\hat{e}_{i,t}$ with replacement. These two techniques lead to qualitatively identical results for the remaining empirical analyses compared with those that are derived by means of the wild bootstrap.

the empirical estimates as described in Section 4.3.2. These are documented in the respective regression tables displayed in the next section when analyzing the sustainability of European fiscal balances. If control variables are used, they are treated as given. Moreover, primary surpluses are simulated under the null hypothesis of non-sustainability ($\rho = 0$). Country specific debt-to-GDP processes are simulated from (4.3) and initialized at random draws from empirical series. The procedure replicates the missing values where European fiscal data are not available.

Statistics are based on 10,000 bootstrap replications. In contrast to the Monte Carlo approach, an estimate of $\hat{\delta}$ is not required for the wild bootstrap. The correlation between $u_{i,t}^*$ and $e_{i,t}^*$ varies with $i = 1, \dots, N$ and $t = 1, \dots, T_i$. However, estimated parameters $\hat{\delta}$ are displayed in the regression tables to provide an impression of the strength of the ‘average’ correlation. The upward bias of the policy reaction coefficient is estimated by averaging over slope parameters estimated from the artificial data simulated under $\rho = 0$. A bias correction of the policy reaction coefficient is achieved by subtracting the evaluated bias from the empirically estimated $\hat{\rho}$ (Mark; 1995).²¹ Quantiles of bootstrap distributions of t -statistics determine the critical values.²²

4.4.2 The sustainability of European fiscal balances

Table 4.2 displays in columns I to V parameters estimated for the policy reaction function conditional on distinct control variables. The estimated policy reaction coefficients ($\hat{\rho}$) are positive and fluctuate around 0.05. Applying right-hand sided heteroskedasticity and autocorrelation robust t -tests and conditioning on asymptotic Gaussian critical values – 1.282 (10%), 1.645 (5%) or 2.326 (1%) – it appears that all reaction coefficients are positive with at least 5% nominal significance.²³

²¹ It is worthwhile to notice that the bias of $\hat{\rho}$ depends on β (and δ), but not on the true value of ρ (Stambaugh; 1999; Hjalmarsson; 2008).

²² As alternative to the proposed bootstrap approach, Hjalmarsson (2010) employs recursive demeaned series to prevent the second-order bias of pooled estimators. As explained in Appendix 4.D, this procedure offers a straightforward and unbiased estimation, but adds noise to the estimation process and is infeasible for the relatively small sample of European fiscal data.

²³ I employ heteroskedasticity and autocorrelation robust covariance estimates to account for patterns of serial correlation in the residuals. Cross-sectional dependence is considered by means of the wild bootstrap approach. Rejection frequencies of conventional t -tests are similar. Moreover, I review the robustness of the results employing covariance estimates proposed by Thompson (2011) which are robust to both serial and cross sectional dependence. Even though the estimation of these so-called double-clustered standard errors adds noise to the test procedure in small samples, test conclusions which I draw in the remaining sections are qualitatively identical.

	I	II	III	IV	V
<i>Policy reaction coefficients</i>					
$\hat{\rho}$	0.039 (1.901)	0.051 (2.273)	0.057 (2.555)	0.043 (1.800)	0.047 (1.989)
$\hat{\rho}_{adj}$	0.000	0.030	0.036	0.025	0.030
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.583	0.413	0.393	0.343	0.342
sim size 5%	0.441	0.286	0.273	0.226	0.236
sim size 1%	0.228	0.119	0.118	0.084	0.087
sim crit 10%	3.027	2.469	2.396	2.223	2.245
sim crit 5%	3.599	2.957	2.920	2.694	2.684
sim crit 1%	4.732	3.970	3.961	3.538	3.560
<i>Control variables</i>					
Business cycle variation		0.493 (2.660)	0.486 (2.613)	0.558 (3.579)	0.547 (3.456)
Temporary gov. spending		-0.568 (-9.812)	-0.559 (-9.563)	-0.522 (-8.488)	-0.522 (-8.517)
Inflation			0.144 (1.971)		0.082 (1.278)
Current account balance				0.188 (1.970)	0.168 (1.845)
<i>Information statistics</i>					
R^2	0.330	0.530	0.492	0.620	0.604
$\hat{\sigma}_u$	0.022	0.017	0.017	0.017	0.017
No. of countries	14	14	14	14	14
No. of obs	334	334	334	334	334
<i>Simulation parameters</i>					
$\hat{\beta}$	0.926	0.926	0.926	0.926	0.926
$\hat{\sigma}_e$	0.039	0.039	0.039	0.039	0.039
$\hat{\delta}$	-0.418	-0.294	-0.278	-0.262	-0.256

Table 4.2: Estimation results for policy reaction functions The upper panel displays estimated policy reaction coefficients in columns I to V conditional on several combinations of control variables along with heteroskedasticity and autocorrelation robust t -statistics in parenthesis. Time dummies are included. In the row denoted $\hat{\rho}_{adj}$ the estimated bias is subtracted from policy reaction coefficients. The next panel documents actual test sizes and critical values of right hand sided heteroskedasticity and autocorrelation robust t -tests derived by means of the wild bootstrap approach. The tests have 10%, 5% and 1% nominal reference, respectively. The bootstrap procedure is repeated 10,000 times. Coefficient estimates for different control variables are documented underneath along with robust t -statistics in parenthesis. Information statistics are followed by simulation parameters which are displayed in the lowest panel.

As displayed in the lower panel of Table 4.2, the estimated common autoregressive parameter of debt-to-GDP ratios $\hat{\beta}$ is 0.926 and the pooled estimate of the correlation of the innovations in primary surpluses and debt-to-GDP ratios ranges from -0.418 to -0.256 for model specifications I to V. For right hand sided t -tests with 5% nominal reference, the actual size is up to 28.6% for model specifications II to V and increases even further if neglecting the influence of control variables (specification I). Thus, test decisions on the relation between European primary surpluses and lagged debt-to-GDP ratios by means of conventional t -tests do not refer to the significance level specified by the analyst. Conditional on size controlled bootstrap based critical values almost non of the estimated adjustment coefficients is significant, not even at the 10% level. Conventional t -test spuriously indicate fiscal balances to accord with the IBC due to the statistical characteristics of primary surpluses and debt-to-GDP ratios. In contrast, bootstrap critical values that account for the excess type I error probability allow the timely detection of increased solvency risks.

In Table 4.2, the row labeled $\hat{\rho}_{adj}$ documents bias adjusted policy reaction coefficients. These coefficients range from 0.025 to 0.036 for specifications II to V and, thus, are about one third smaller compared with their unadjusted counterparts. The parameters measuring the influence of control variables have the expected signs independent of the model specification. While above trend GDP values cause a rise of primary surpluses, above trend values of temporary government spending cause a decline of primary surpluses (columns II-V). Employing conventional one hand sided t -tests, both effects are positive with 1% significance. Adding either the inflation rate (III) or the current account balance (IV) to the list of control variables, the parameters measuring their effect are positive with 5% significance. If both regressors are incorporated simultaneously (regression V), inflation becomes insignificant while the current account balance still has a positive effect with 5% significance. As described in Section 4.2.3, measures of current account balances are persistent processes and, thus, the estimated coefficients displayed in columns IV and V have to be interpreted with caution.

4.4.3 Results for the Eurozone and Europe excluding Greek data

Countries that are members of the Eurozone are possibly more homogeneous compared with the larger group of European countries having distinct currencies. Potentially sustainable fiscal balances of the Eurozone countries could be masked by fluctuations of non Eurozone countries' data comprised in the baseline sample. To

investigate the sensibility of the results with respect to the heterogeneity of the data, I neglect the information for Denmark, Sweden and the United Kingdom and restrict the analysis to the subsample of the remaining 11 economies which are members of the Eurozone. Results are nevertheless similar compared with those of the baseline specification. Estimated policy reaction coefficients over specifications I to V do not deviate more than 0.002 units from those discussed in Table 4.2.²⁴ The size of right hand sided t -tests with 5% nominal reference ranges between 18.7% and 23.7% for specifications II to V and is again larger for model I. The policy reaction coefficients are insignificant at the 10% level conditional on bootstrap based critical values. Considering asymptotic Gaussian quantiles as reference values, these coefficients are spuriously indicated to be positive with 5% nominal significance. Estimates governing the influence of control variables are also similar to those shown in Table 4.2.

Similarly, data for Greece as the country with the highest debt-to-GDP ratio in the sample from 2006 to 2013 might reduce the estimated joint policy reaction coefficients for the 14 European countries. As documented in Table 4.3, the estimate of ρ increases by about 0.01 units in all specifications if Greek data are excluded (taking Table 4.2 as reference). However, the estimate of β still amounts to 0.924 and the correlation between the innovations of primary surpluses and debt-to-GDP ratios δ ranges between -0.312 and -0.279 for specifications II to V. These correlations are slightly larger in absolute terms than the correlations displayed in Table 4.2. In consequence, the excesses of type I error probabilities and the bootstrap critical values rise marginally. Estimated coefficients $\hat{\rho}$ are spuriously indicated to be positive with 1% nominal significance in specifications I to V conditional on asymptotic critical values. Conditional on size corrected critical values, $\hat{\rho}$ is insignificant at the 10% level in specification I and IV and at the 5% level in specifications II and V. Controlling for the influence of inflation, $\hat{\rho}$ is positive with 5% significance (columns III). In sum, the size distortions of t -tests for the common policy reaction coefficient are not exclusively caused by the incorporation of Greek data. Gaussian quantiles still do not refer to the specified significance level. However, empirical t -ratios range in a region where test decisions are less sensitive to underlying critical values if Greek data are excluded. There is weak evidence that fiscal balances of European countries apart from Greece before 2009 were sustainable.²⁵

²⁴ Due to space limitations I refrain from presenting the respective table. It is available upon request.

²⁵ Results for the Eurozone subsample excluding Greek data are qualitatively identical.

4.4 Size controlled inference on fiscal sustainability in Europe

	I	II	III	IV	V
<i>Policy reaction coefficients</i>					
$\hat{\rho}$	0.047 (2.478)	0.062 (3.034)	0.067 (3.221)	0.053 (2.352)	0.057 (2.528)
$\hat{\rho}_{adj}$	0.006	0.040	0.045	0.034	0.037
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.581	0.452	0.436	0.386	0.390
sim size 5%	0.441	0.310	0.304	0.268	0.271
sim size 1%	0.222	0.132	0.129	0.114	0.111
sim crit 10%	3.002	2.533	2.519	2.359	2.405
sim crit 5%	3.598	3.073	2.950	2.873	2.818
sim crit 1%	4.781	4.145	4.035	3.908	3.847
<i>Control variables</i>					
Business cycle variation		0.609 (3.747)	0.606 (3.769)	0.658 (4.607)	0.650 (4.556)
Temporary gov. spending		-0.570 (-11.841)	-0.564 (-11.605)	-0.537 (-10.177)	-0.537 (-10.226)
Inflation			0.126 (1.549)		0.077 (1.110)
Current account balance				0.160 (1.606)	0.142 (1.529)
<i>Information statistics</i>					
R^2	0.358	0.558	0.531	0.619	0.604
$\hat{\sigma}_u$	0.022	0.017	0.016	0.016	0.016
No. of countries	13	13	13	13	13
No. of obs	320	320	320	320	320
<i>Simulation parameters</i>					
$\hat{\beta}$	0.924	0.924	0.924	0.924	0.924
$\hat{\sigma}_e$	0.038	0.038	0.038	0.038	0.038
$\hat{\delta}$	-0.426	-0.312	-0.298	-0.284	-0.279

Table 4.3: Estimation results for policy reaction functions excluding Greek data
For further notes see Table 4.2.

4.4.4 Nonlinear policy reactions

In the literature on model based sustainability, not only linear regression frameworks are considered. As argued by Bohn (1998, 2008) in single-equation settings and by Ghosh et al. (2013) in panel data settings, a likely pattern is that the policy reaction strengthens at higher debt levels and may start to decrease again at some turning point. I follow Ghosh et al. (2013) and augment the policy reaction function in (4.2) with the lagged squared debt-to-GDP ratio and the lagged debt-to-GDP ratio to the power of three. This function reads as

$$s_{i,t} = \rho_1 d_{i,t-1} + \rho_2 d_{i,t-1}^2 + \rho_3 d_{i,t-1}^3 + \psi' x_{i,t} + \alpha_i + \gamma_t + u_{i,t}. \quad (4.7)$$

Estimation results for the nonlinear policy reaction function are summarized in Table 4.4. Linear responses measured by $\hat{\rho}_1$ are negative in all specifications. I find severe size distortions of t -tests and an upward bias of parameter estimates by means of the bootstrap approach. In model specifications I to III, $\hat{\rho}_1$ is negative with 10% significance conditional on bootstrap based critical values, while Gaussian quantiles spuriously imply insignificance at the nominal 10% level.²⁶ In columns IV and V $\hat{\rho}_1$ is significant at the 5% level.

The estimates of coefficient ρ_2 that captures the effect of the squared debt-to-GDP ratio are positive. The actual size of the corresponding t -tests tends to be smaller than the nominal reference. Estimates of ρ_2 displayed in columns II and III are significant at the 10% level and in columns IV and V at the 5% level. Estimates of parameter ρ_3 that captures the cubic pattern of primary surpluses' reaction to increasing debt-to-GDP ratios are negative. For specifications II and III, estimates of ρ_3 are significant at the 10% level and for specifications IV and V at the 5% level. These conclusions do not change if asymptotic critical values are applied instead of bootstrap versions.

In sum and comparable to Ghosh et al. (2013), estimated parameters imply that governments borrow at low levels of debt-to-GDP ratios and, thus, enlarge the debt burden. However, expenditure driven fiscal policies are corrected if debt-to-GDP ratios increase. This pattern flattens out at high levels of $d_{i,t}$.

²⁶It is worthwhile to notice that the upward bias of coefficient $\hat{\rho}_1$ causes lower tail quantiles of the simulated distribution to be smaller in absolute terms than those of the asymptotic distribution.

	I	II	III	IV	V
<i>Policy reaction coefficients</i>					
$\hat{\rho}_1$	-0.120 (-0.972)	-0.121 (-1.121)	-0.129 (-1.166)	-0.243 (-1.660)	-0.237 (-1.605)
$\hat{\rho}_{1,adj}$	-0.172	-0.160	-0.166	-0.273	-0.266
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.452	0.411	0.390	0.355	0.347
sim size 5%	0.346	0.311	0.296	0.274	0.280
sim crit 5%	-0.592	-0.797	-0.821	-0.806	-0.802
sim crit 2.5%	-1.058	-1.185	-1.262	-1.185	-1.203
$\hat{\rho}_2$	0.150 (0.906)	0.211 (1.403)	0.231 (1.465)	0.382 (1.938)	0.378 (1.893)
$\hat{\rho}_{2,adj}$	0.155	0.235	0.256	0.402	0.399
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.143	0.071	0.066	0.062	0.065
sim size 5%	0.088	0.040	0.045	0.036	0.035
sim crit 10%	1.544	1.006	0.994	0.952	0.928
sim crit 5%	2.041	1.529	1.559	1.428	1.408
$\hat{\rho}_3$	-0.025 (-0.355)	-0.071 (-1.052)	-0.079 (-1.125)	-0.146 (-1.741)	-0.144 (-1.695)
$\hat{\rho}_{3,adj}$	-0.023	-0.078	-0.087	-0.153	-0.151
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.166	0.071	0.066	0.067	0.067
sim size 5%	0.105	0.041	0.044	0.039	0.039
sim crit 10%	-1.666	-1.038	-1.021	-1.029	-1.018
sim crit 5%	-2.147	-1.513	-1.542	-1.494	-1.504
<i>Control variables</i>					
Business cycle variation		0.450 (2.513)	0.442 (2.455)	0.541 (3.871)	0.530 (3.711)
Temporary gov. spending		-0.545 (-9.822)	-0.535 (-9.667)	-0.489 (-8.361)	-0.489 (-8.353)
Inflation			0.158 (2.265)		0.080 (1.317)
Current account balance				0.246 (2.123)	0.226 (1.977)
<i>Information statistics</i>					
R^2	0.363	0.532	0.490	0.632	0.618
$\hat{\sigma}_u$	0.021	0.017	0.017	0.016	0.016
No. of countries	14	14	14	14	14
No. of obs	334	334	334	334	334
<i>Simulation parameters</i>					
$\hat{\beta}$	0.926	0.926	0.926	0.926	0.926
$\hat{\sigma}_e$	0.039	0.039	0.039	0.039	0.039
$\hat{\delta}$	-0.429	-0.306	-0.289	-0.260	-0.254

Table 4.4: Estimation results for nonlinear policy reaction functions In case of $\hat{\rho}_1$ heteroskedasticity and autocorrelation robust t -tests are two-sided since positive and negative values of ρ_1 are economically reasonable. Size corrected critical values correspond to the $\zeta/2$ -quantile of bootstrap distributions of test statistics, where $\zeta = \{0.1, 0.05\}$. I test by means of right hand sided robust tests for significance of $\hat{\rho}_2$ and size corrected quantiles correspond to the $(1 - \zeta)$ -quantile of bootstrap distributions of t -statistics. In contrast, left hand sided robust t -tests are evaluated in case of $\hat{\rho}_3$. Critical values are the ζ -quantile of bootstrap distributions of t -statistics. Results for t -tests with 1% nominal reference are disregarded for brevity. Primary surpluses are simulated under the null hypothesis of non-sustainability ($\rho_1 = \rho_2 = \rho_3 = 0$). For further notes see Table 4.2.

To visualize the inverted U-shape of the policy reaction, Figure 4.5 displays the marginal effect at distinct levels of the debt-to-GDP ratio. The black curve represents conventional estimates, while the gray curve shows bias adjusted estimates of policy responses. The turning point of the policy response is close to a debt-to-GDP ratio of unity which is equally found by Ghosh et al. (2013) in a larger sample containing also non European economies. Conventionally estimated marginal effects are upward biased at all evaluation points.

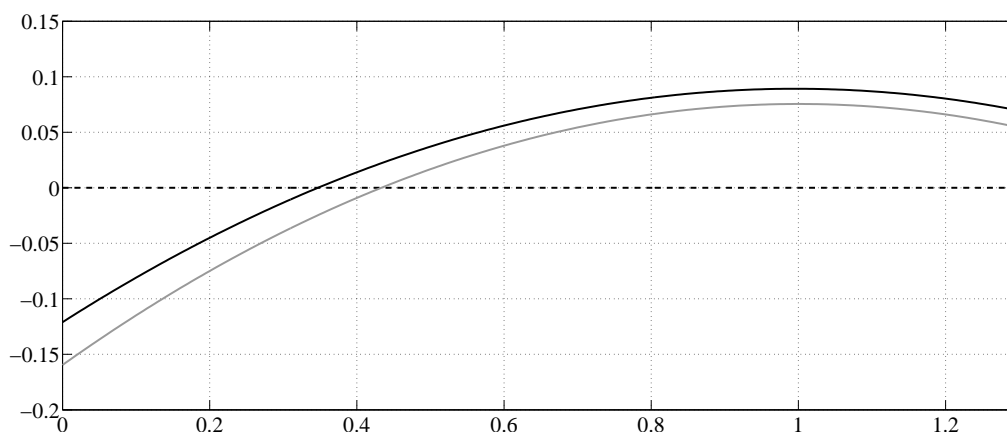


Figure 4.5: Nonlinear policy reaction to increasing debt-to-GDP ratios This figure displays the marginal policy reaction effects at distinct evaluation points. The black curve represents the conventionally estimated policy reaction and the gray curve represents the bias adjusted policy reaction. Employed parameter values are those of the baseline model specification documented in column II of Table 4.4.

To assess the significance of the marginal effects, I focus on the marginal effects evaluated at the moderate $d_{i,t-1} = 0.5$ and the high level $d_{i,t-1} = 1$. Additionally, $d_{i,t-1} = 0.63$, the average European debt-to-GDP ratio at the end of 2008, is considered as evaluation point to explore if oversized conventional t -tests have prevented the early detection of the increasing fiscal risk at the beginning of the crisis. Estimated marginal effects are displayed in Table 4.5. At the relative moderate level of $d_{i,t-1} = 0.5$, marginal effects of the distinct model specifications are positive, but indicated to be insignificant at the 5% nominal level regardless whether test decisions are conditioned on asymptotic or bootstrap based critical values. At the debt-to-GDP level of 0.63, estimated marginal effects double compared with those evaluated at $d_{i,t-1} = 0.5$. However, they are spuriously indicated to be significant at the nominal 1% level conditional on Gaussian quantiles (specifications II to V). Considering size corrected critical values, the reaction in column I turns out to be insignificant at the 10% level. Estimates in columns II and III are positive with 10% significance and these in columns IV and V with 5% significance.

4.4 Size controlled inference on fiscal sustainability in Europe

	I	II	III	IV	V
<i>Marginal policy reaction</i>					
$\partial s_{i,t}/\partial d_{i,t-1} _{d_{i,t-1}=0.5}$	0.011 (0.429)	0.037 (1.354)	0.043 (1.562)	0.029 (1.127)	0.033 (1.234)
$\partial s_{i,t}/\partial d_{i,t-1} _{d_{i,t-1}=0.5,adj}$	-0.035	0.017	0.024	0.015	0.019
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.560	0.342	0.320	0.277	0.269
sim size 5%	0.429	0.233	0.223	0.183	0.173
sim size 1%	0.239	0.101	0.086	0.062	0.062
sim crit 10%	3.199	2.374	2.240	2.073	2.021
sim crit 5%	3.801	2.827	2.765	2.530	2.501
sim crit 1%	5.304	3.991	3.760	3.363	3.360
<i>Marginal policy reaction</i>					
$\partial s_{i,t}/\partial d_{i,t-1} _{d_{i,t-1}=0.63}$	0.038 (1.662)	0.061 (2.326)	0.068 (2.506)	0.064 (2.803)	0.067 (2.814)
$\partial s_{i,t}/\partial d_{i,t-1} _{d_{i,t-1}=0.63,adj}$	-0.005	0.043	0.052	0.052	0.056
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.537	0.298	0.277	0.241	0.229
sim size 5%	0.401	0.198	0.186	0.152	0.143
sim size 1%	0.220	0.081	0.069	0.050	0.048
sim crit 10%	3.077	2.181	2.093	1.938	1.877
sim crit 5%	3.673	2.638	2.599	2.370	2.326
sim crit 1%	5.135	3.785	3.548	3.174	3.173
<i>Marginal policy reaction</i>					
$\partial s_{i,t}/\partial d_{i,t-1} _{d_{i,t-1}=1}$	0.105 (3.796)	0.089 (3.657)	0.096 (3.941)	0.082 (3.756)	0.086 (3.965)
$\partial s_t/\partial d_{i,t-1} _{d_{i,t-1}=1,adj}$	0.067	0.075	0.084	0.073	0.078
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.502	0.254	0.240	0.214	0.196
sim size 5%	0.369	0.167	0.154	0.126	0.115
sim size 1%	0.185	0.061	0.056	0.039	0.038
sim crit 10%	2.863	2.005	1.900	1.797	1.741
sim crit 5%	3.429	2.483	2.387	2.226	2.169
sim crit 1%	4.584	3.505	3.319	3.048	3.055

Table 4.5: Marginal effects in nonlinear policy reaction functions This table displays responses of primary surpluses to accelerations of debt-to-GDP ratios evaluated at distinct levels of the latter. For further notes see Table 4.2.

The bias adjusted estimates are considerably smaller than unadjusted versions. For instance, in the baseline specification II, the bias adjusted marginal effect at

$d_{i,t} = 0.63$ is 30% smaller than the unadjusted version. This finding is worthwhile to notice since the European Stability and Growth Pact establishes a maximum level of the debt-to-GDP ratio of 0.6. At higher rates, governments are supposed to reduce their debt burdens. Significant policy reactions are revealed at $d_{i,t-1} = 1$. At this point, bias adjusted marginal reaction coefficients range around 0.75 for specifications II to V and, thus, are relatively large. The marginal effects are statistically significant at the 1% level conditional on size corrected critical values.

In sum, the investigation of the potential non-linearity of the European policy reaction function reveals that fiscal policies do not accord with the model based IBC at moderate levels of debt-to-GDP ratios. However, fiscal authorities start to adjust primary surpluses upward at high levels of debt-to-GDP ratios which are considerable larger than the turning points established by the European Stability and Growth Pact and indicated by conventional inference.

4.4.5 Characteristics of post-2008 fiscal figures

The main focus of this work is to analyze pre-crisis European fiscal balances. However, interest in the current development of European fiscal figures naturally arises. Results for the baseline linear policy reaction function (4.2) estimated with data until 2013 are summarized in Table 4.6.

The estimated coefficient ρ decreases in all specifications compared with those conditioned on pre-2009 observations (Table 4.2). The common autoregressive parameter of debt-to-GDP ratios $\hat{\beta}$ and the correlation $\hat{\delta}$ of the innovations in primary surpluses and in debt-to-GDP ratios increase in absolute terms. This pattern causes larger size violations and consequently tighter bootstrap based critical values. However, regardless whether asymptotic or size controlled quantiles are applied, the coefficient $\hat{\rho}$ is indicated to be insignificant at the 10% nominal level for almost all regression specifications. Only in specifications I and III, Gaussian critical values spuriously indicate positiveness with 10% nominal significance. Bias adjusted policy reaction coefficients are nearly zero and support the conclusion that European fiscal policies do not accord with the model based IBC in recent years of economic turmoil.²⁷ The estimated coefficients for the control variables are relatively close to the estimates documented in Table 4.2. It remains an open question if the measures initiated by the European authorities are effective to rebalance European fiscal

²⁷This conclusion holds focusing on the Eurozone subsample. Excluding Greek data, policy reaction coefficients are slightly larger, but are still insignificant at common levels conditional on bootstrap based critical values.

4.4 Size controlled inference on fiscal sustainability in Europe

	I	II	III	IV	V
<i>Policy reaction coefficients</i>					
$\hat{\rho}$	0.022 (1.317)	0.020 (1.059)	0.026 (1.413)	0.015 (0.802)	0.019 (1.039)
$\hat{\rho}_{adj}$	-0.020	-0.003	0.002	-0.006	-0.002
<i>Actual size and bootstrap based critical values</i>					
sim size 10%	0.696	0.567	0.570	0.483	0.512
sim size 5%	0.571	0.457	0.467	0.377	0.388
sim size 1%	0.338	0.247	0.269	0.198	0.213
sim crit 10%	3.547	3.320	3.333	2.925	3.065
sim crit 5%	4.278	3.997	4.118	3.655	3.777
sim crit 1%	5.603	5.717	5.855	5.122	5.428
<i>Control variables</i>					
Business cycle variation		0.568 (4.743)	0.557 (4.417)	0.651 (6.702)	0.637 (6.076)
Temporary gov. spending		-0.542 (-16.704)	-0.530 (-15.419)	-0.515 (-16.635)	-0.510 (-16.060)
Inflation			0.153 (1.850)		0.091 (1.109)
Current account balance				0.198 (2.817)	0.180 (2.527)
<i>Information statistics</i>					
R^2	0.386	0.608	0.585	0.687	0.682
$\hat{\sigma}_u$	0.027	0.021	0.020	0.020	0.020
No. of countries	14	14	14	14	14
No. of obs	404	404	404	404	404
<i>Simulation parameters</i>					
$\hat{\beta}$	0.979	0.979	0.979	0.979	0.979
$\hat{\sigma}_e$	0.046	0.046	0.046	0.046	0.046
$\hat{\delta}$	-0.462	-0.336	-0.333	-0.313	-0.313

Table 4.6: Estimation results for policy reaction functions until 2013 For further notes see Table 4.2.

policies. However, to answer this question in upcoming years, it is indispensable to adopt size controlled econometric tools.

4.5 Conclusions

This study investigates if the strong persistence of debt-to-GDP ratios and the negative correlation of the innovations in primary surpluses and debt-to-GDP ratios invalidate the conclusions on model based sustainability of pre-2009 European fiscal balances in panel data analyses. I reveal that a second-order bias, known from the related finance literature, equally occurs in panel based policy reaction regressions. This bias is induced by the within-transformation of panel data. Estimated policy reaction coefficients for pre-crisis European fiscal figures are upward biased and conventional t -tests suffer from size distortions.

Overrated test sizes bear the risk to reject the null hypothesis of non-sustainability more frequently than specified by the analyst. Test decisions conditional on asymptotic critical values spuriously indicate model based sustainability of pre-2009 European fiscal policies. Excluding Greek data, weak evidence hints at a significantly positive policy reaction coefficient. Conditioning inference in nonlinear versions of the policy reaction function on bias adjustments and size controlled bootstrap based critical values, strong upward corrections of primary surpluses only arise if debt-to-GDP ratios approach unity. Thus, fiscal authorities begin to effectively reduce the debt burden at higher debt-to-GDP ratios than supposed by the European Stability and Growth Pact. For academics, policy makers and the European public, it is of essential interest if corrective measures taken since 2009 are able to return European fiscal balances on a sustainable path. To evaluate the effectiveness of these measures, size corrected inference is essential.

Appendix to Paper 4

4.A Data sources

Data of gross public debt as share of GDP and CPI inflation are published in the IMF World Economic Outlook.²⁸ If series of debt-to-GDP ratios have missing values in the beginning of the sample period, I replaced them by the corresponding quotas of the IMF Historical Public Debt Database.²⁹ Series of gross domestic products, primary balances as share of GDP, government expenditures and current account balances as share of GDP are downloaded from the OECD Economic Outlook: Statistics and Projections.³⁰ Real variables are calculated by means of the respective GDP deflator provided by the same OECD source.

4.B The second-order bias if the cross section is large

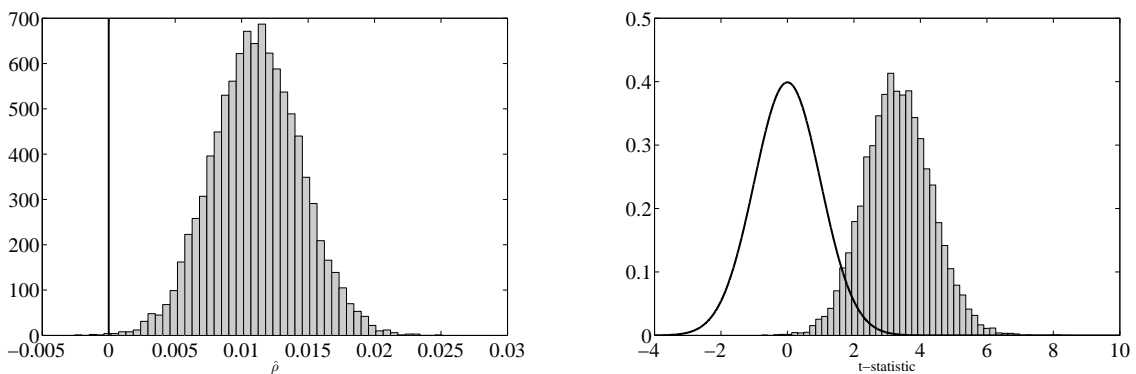


Figure 4.6: Simulated distributions of the policy reaction coefficient and the t -statistic for $N = 200$ and $T = 30$ For further notes see Figure 4.2 and Figure 4.4

²⁸<http://www.imf.org/external/ns/cs.aspx?id=28>

²⁹<http://www.imf.org/external/pubs/cat/longres.cfm?sk=24332.0>

³⁰<http://stats.oecd.org/>

4.C Empirical correlation structures and wild bootstrap counterparts

	AT	BE	DK	FI	FR	DE	GR	IE	IT	NL	PT	ES	SE	UK
Cross-correlation matrix of empirical estimates $\hat{u}_{i,t}$														
AT	1.00													
BE	-0.31	1.00												
DK	-0.38	0.11	1.00											
FI	0.33	-0.46	0.70	1.00										
FR	0.41	-0.76	-0.24	0.62	1.00									
DE	0.19	-0.04	-0.22	-0.25	0.10	1.00								
GR	-0.01	-0.09	-0.83	-0.68	0.43	0.27	1.00							
IE	0.01	-0.39	-0.60	-0.79	0.33	0.13	0.72	1.00						
IT	-0.48	0.45	-0.60	-0.82	-0.69	0.27	0.28	0.48	1.00					
NL	-0.07	0.24	0.11	-0.17	-0.43	0.16	-0.47	-0.24	0.29	1.00				
PT	0.16	-0.65	-0.29	0.09	0.38	-0.28	-0.00	0.35	-0.21	-0.41	1.00			
ES	-0.44	0.59	0.63	-0.40	-0.75	-0.25	-0.70	-0.30	0.39	0.30	-0.31	1.00		
SE	-0.18	0.15	0.64	0.84	-0.59	-0.40	-0.68	-0.88	-0.37	0.51	-0.70	0.43	1.00	
UK	0.48	-0.58	-0.08	0.59	0.84	0.10	0.27	0.01	-0.69	-0.48	0.29	-0.74	-0.41	1.00
Average cross-correlation matrix of bootstrap innovations $u_{i,t}^*$														
AT	1.00													
BE	-0.31	1.00												
DK	-0.28	0.02	1.00											
FI	0.33	-0.46	0.64	1.00										
FR	0.41	-0.76	-0.08	0.61	1.00									
DE	0.15	-0.04	-0.21	-0.19	0.06	1.00								
GR	-0.04	-0.03	-0.84	-0.65	0.20	0.25	1.00							
IE	0.00	-0.28	-0.62	-0.70	0.18	0.13	0.70	1.00						
IT	-0.46	0.45	-0.46	-0.82	-0.68	0.20	0.22	0.38	1.00					
NL	-0.07	0.24	0.11	-0.16	-0.43	0.12	-0.44	-0.23	0.29	1.00				
PT	0.15	-0.65	-0.17	0.10	0.40	-0.24	-0.00	0.34	-0.21	-0.40	1.00			
ES	-0.44	0.58	0.50	-0.39	-0.75	-0.23	-0.61	-0.27	0.39	0.31	-0.30	1.00		
SE	-0.16	0.09	0.65	0.74	-0.29	-0.34	-0.66	-0.86	-0.29	0.49	-0.40	0.36	1.00	
UK	0.48	-0.59	-0.08	0.57	0.84	0.08	0.25	0.03	-0.68	-0.47	0.30	-0.73	-0.32	1.00

Table 4.7: Cross-correlation matrix of $\hat{u}_{i,t}$ and $u_{i,t}^*$. The empirical innovations $\hat{u}_{i,t}$ considered in the upper panel are estimated from the baseline specification of (4.2) comprising series of business cycle variation and temporary government spending as control variables. Parameter estimates are documented in Table 4.2. These estimates are used to generate innovations $u_{i,t}^*$. The lower panel displays the average cross-correlation matrix for $u_{i,t}^*$ based on 10,000 bootstrap replications.

	AT	BE	DK	FI	FR	DE	GR	IE	IT	NL	PT	ES	SE	UK
$\hat{u}_{i,t}$	0.16	-0.83	-0.43	-0.55	-0.20	0.55	-0.08	-0.49	0.33	0.02	-0.06	-0.31	-0.44	-0.32
$u_{i,t}^*$	0.17	-0.82	-0.37	-0.52	-0.21	0.46	-0.13	-0.40	0.30	-0.04	-0.05	-0.33	-0.45	-0.32

Table 4.8: Correlation of $\hat{u}_{i,t}$ and $\hat{e}_{i,t}$ and corresponding bootstrap versions The empirical innovations $\hat{u}_{i,t}$ considered in the upper panel are estimated from the baseline specification of (4.2) comprising series of business cycle variation and temporary government spending as control variables. Parameter estimates are documented in Table 4.2. Innovations $\hat{e}_{i,t}$ are estimated from (4.3). These estimates are employed to generate bootstrap innovations $u_{i,t}^*$ and $e_{i,t}^*$. The upper panel displays country specific correlations between $\hat{u}_{i,t}$ and $\hat{e}_{i,t}$. The lower panel documents the country specific correlations between $u_{i,t}^*$ and $e_{i,t}^*$ averaged over 10,000 bootstrap replications.

4.D Recursive demeaning

In the literature emerging around Hjalmarsson (2008, 2010), so-called recursive demeaning techniques have been established to deal with the second-order bias of the fixed effects estimator when the regressor is persistent and its innovations correlate with those of the dependent variable.³¹ This approach simultaneously allows to control for the appropriateness of the size of t -tests and is an alternative to the bootstrap techniques described in Section 4.4.1.

The within-transformed regressor $\tilde{d}_{i,t-1} = d_{i,t-1} - 1/T_i \sum_{t=1}^{T_i} d_{i,t-1}$ absorbs post $t - 1$ information about $e_{i,\tau}$ and is by design correlated with the innovations of country specific primary surpluses $u_{i,\tau}$, for $\tau = t, \dots, T_i$ (Hjalmarsson; 2010). This correlation induces the so-called second-order bias and motivates the use of recursive demeaning techniques. Following the lines of Pastor et al. (2014) and abstracting from control variables $x_{i,t}$ and time dummies γ_t in (4.2), the policy reaction function for forward demeaned data formalizes as

$$\bar{s}_{i,t} = \alpha_i + \rho \bar{d}_{i,t-1} + \bar{u}_{i,t}, \quad (4.8)$$

where

$$\bar{s}_{i,t} = s_{i,t} - \frac{1}{T_i - t + 1} \sum_{h=t}^{T_i} s_{i,h}, \quad (4.9)$$

$$\bar{d}_{i,t-1} = d_{i,t-1} - \frac{1}{T_i - t + 1} \sum_{h=t}^{T_i} d_{i,h-1}. \quad (4.10)$$

³¹Recursive demeaning goes back to Moon and Phillips (2000) and Sul, Phillips and Choi (2005) and is used in distinct applications as, for instance, unit root testing (Shin and So; 2001; Broda, Carstensen and Paoletta; 2009).

The error term $\bar{u}_{i,t}$ is defined similarly. The evolving recursive-demeaning pooled estimator is

$$\hat{\rho}_{RD} = \left(\sum_{i=1}^N \sum_{t=2}^{T_i} \bar{d}_{i,t-1} \underline{d}_{i,t-1} \right) \left(\sum_{i=1}^N \sum_{t=2}^{T_i} \bar{s}_{i,t} \underline{d}_{i,t-1} \right), \quad (4.11)$$

where the backward demeaned debt-to-GDP ratio

$$\underline{d}_{i,t-1} = d_{i,t-1} - \frac{1}{t-1} \sum_{h=1}^{t-1} d_{i,h-1}, \quad (4.12)$$

is employed as instrument for $\bar{d}_{i,t-1}$. This instrument is by design highly correlated with $\bar{d}_{i,t-1}$ and does not incorporate information about $\bar{e}_{i,\tau}$ and $\bar{u}_{i,\tau}$, for $\tau = t, \dots, T_i$. Backward demeaning of instruments is necessary since the debt-to-GDP ratio is a non-zero mean regressor. Zero mean regressors can be instrumented by themselves (Hjalmarsson; 2010; Pastor et al.; 2014).³² A robust covariance estimator of $(\hat{\rho}_{RD})$ is provided by Hjalmarsson (2010). Hjalmarsson (2010) and Pastor et al. (2014) show by means of extensive simulation studies that the RD estimator is unbiased and that the corresponding t -test has appropriate size and power.

However, forward and backward demeaning of data adds noise to the estimation procedure. I employ a simulation study under the null hypothesis of $\rho = 0$ to investigate the influence of datasets' dimensions on the properties of the RD estimator. The setup of the simulation is identical to that discussed in the context of Figure 4.4. The panels in the left column of Figure 4.7 display the histograms of $\hat{\rho}_{RD}$ from 10,000 replications for distinct sample sizes. The panels in the right column of Figure 4.4 present the histograms of corresponding robust right hand sided t -statistics along with the asymptotic Gaussian distribution. In the upper panel ($N = 14$ and $T = 28$), the estimator appears to be unbiased but has a large variance compared with that of the biased fixed effects estimator considered in Figure 4.2. However, the t -statistic displayed in the right hand sided column is well-shaped. As can be reasoned from the histograms presented in the middle and lower left hand sided panels of Figure 4.7 the standard deviation of $\hat{\rho}_{RD}$ reduces for increasing N or T . While Hjalmarsson (2010) and Pastor et al. (2014) are concerned with questions of financial economics where datasets are typically large the number of observations on the fiscal balances of European economies is relatively small ($N = 14$ and $T = 28$). The application of the recursive demeaning procedure

³²The vector of regressors in (4.11) becomes $\bar{z}_{i,t} = (\bar{d}_{i,t-1}, \bar{x}_{i,t}')$ if control variables are considered. The forward demeaned control variables $\bar{x}_{i,t}$ can be instrumented by themselves.

necessitates that the dimensions of the dataset are larger than $N = 14$ and $T = 28$.

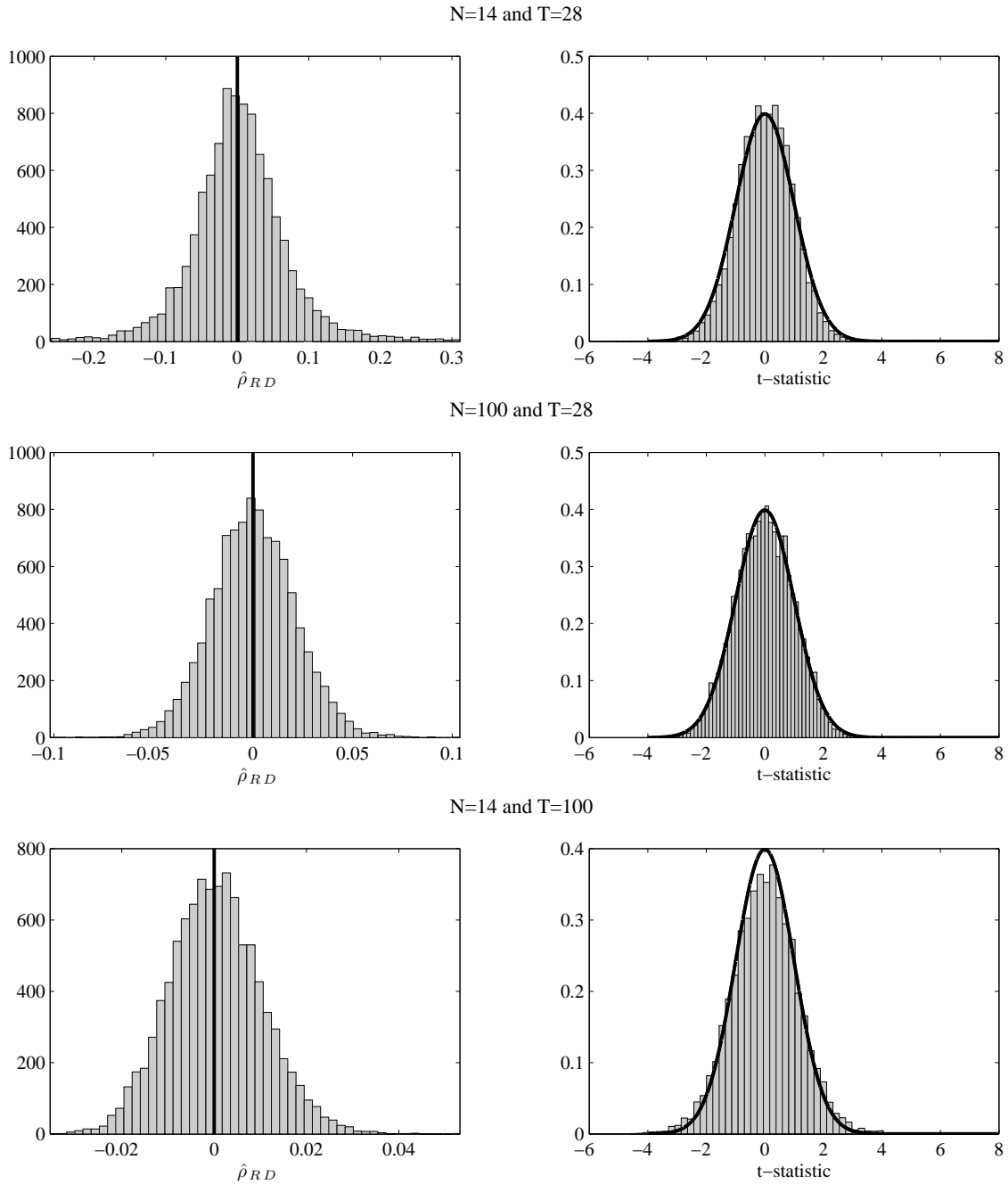


Figure 4.7: Biases and size distortions of recursive-demeaning estimators The left hand side panels display histograms of $\hat{\rho}_{RD}$ while the right hand sided panels display the empirical distributions of the corresponding robust t -statistics. Primary surpluses are simulated by means of the Monte Carlo procedure described in Section 4.3.2 under the null hypothesis of $\rho = 0$. Figures are based on 10,000 Monte Carlo replications of balanced panel datasets. The remaining simulation parameters are equal to those described in the context of Figure 4.4. Since the RD estimator delivers severe outliers for small samples the left hand side panel displays the histogram for $\hat{\rho}_{RD}$ where the lowest 1% and the largest 1% estimations are disregarded when $N = 14$ and $T = 28$.

Concluding remarks

This thesis contributes to the literature in several directions. It provides a step towards a better understanding of the interrelation between financial markets and the macroeconomy. The role of uncertainty as an explanatory factor emerges in two dimensions. First, we reveal that consumption uncertainty is a determinant of the long-term movements of the US price-to-dividend ratio. Second, we provide evidence that macroeconomic uncertainty affects real activity. Moreover, this thesis highlights the importance of reliable statistical approaches to assess the sustainability of fiscal balances. The analysis is throughout based on modern statistical techniques and modeling approaches as, for instance, nonlinear state space and dynamic factor modeling. The informational content of distinct data categories is employed to derive a profound understanding of economic dynamics and interrelations. The topics of this thesis are related to three recent interrelated challenges that the world economy faced: the global financial crisis, the Great Recession and the European debt crisis.

In Paper 1, we reveal three factors that jointly influence the low frequency dynamics of the US price-to-dividend ratio in the post-WWII period. To this end, we generalize the present value model of Campbell and Shiller (1988) towards a flexible nonlinear state space model that allows for a time-varying mean process of the price-to-dividend ratio. Explanatory factors for the variations of this mean process are consumption uncertainty, the demographic structure of the population and the share of firms with traditional dividend payout policy. We find that consumption uncertainty is the dominant factor in shaping the slowly evolving mean of the price-to-dividend ratio. Moreover, this study provides evidence that the price-to-dividend ratio has ex-ante predictive content for stock returns when it is adjusted for its time-varying mean.

The effects of uncertainty on economic activity are highlighted in Paper 2. By means of a dynamic factor model, we summarize the common dynamics of several individual uncertainty measures from distinct areas of the economy. This condensation of information is useful for policy authorities to monitor economic dynamics. We conclude that two fundamental factors summarize macroeconomic uncertainty.

We distinguish between two types thereof – ‘business cycle uncertainty’ and ‘oil and commodity price uncertainty’. The distinction between both types of macroeconomic uncertainty is crucial since each type causes a decline of real activity of different magnitude. The effect of a shock to oil and commodity price uncertainty is more pronounced.

Paper 3 addresses the question if past and current fiscal balances have been sustainable. Taking US, UK and Portuguese data as example, it is revealed that policy reaction regressions, that are employed to test if fiscal policy accords with an intertemporal budget constraint (IBC), are of similar structure as the predictive regressions employed in Paper 1. We diagnose upward biased estimates of policy reaction coefficients and overrated type I error probabilities of t -tests. The actual size of a t -test is up to five times larger than its 5% nominal reference. This size distortion bears the risk to reject the null hypothesis of non-sustainability too often. We provide Monte Carlo techniques and local to unit root asymptotics which allow for size corrected inference. UK and US fiscal policy is found to be sustainable in samples covering more than 100 years, but not for the post-WWII period. We explore several post-WWII subsamples, where conventional t -tests spuriously indicate sustainability of fiscal policy in these two countries. Thus, only size corrected inference allows for conclusive policy monitoring to assesses if fiscal balances accord with the IBC.

Similarly, Paper 4 focuses on sustainability tests for fiscal balances. However, this paper relies on a panel dataset that is useful to concentrate on the dynamics of European fiscal figures during the most recent decades. As in the single-equation framework of Paper 3, inference on panel based policy reaction functions suffers from size distortions. However, the within-transformation of the panel data, commonly applied to account for country specific effects, compounds the bias and the size distortion. The wild bootstrap procedure provided in Paper 4 is able to allow for size corrected inference. I reveal that the increased solvency risk of European countries could have been detected by means of this method before the European debt crisis manifested in 2009. For policy makers, panel based inference is useful when assessing the sustainability of most recent fiscal policy. Conditional on the diagnosis of an insignificant common policy reaction coefficient for the entire sample, fiscal balances of individual countries can be investigated in detail in a subsequent analysis (see Paper 3).

In light of these findings, economic and econometric research can follow several fruitful directions in the future. First, the empirical results of Paper 1 call for a

structural modeling of the interactions between financial markets and the macroeconomy while accounting for slow moving mean processes of key financial indicators (for a related literature see, e.g., Bansal et al.; 2010). Moreover, the international integration of capital markets makes it worth to investigate if and how the low frequency dynamics of key financial indicators of distinct markets are connected. Second, the investigation of the impact of macroeconomic uncertainty on real activity in Paper 2, motivates the incorporation of uncertainty shocks in structural economic models (e.g., Basu and Bundick; 2012) and the account for distinct types thereof. Third, it is attempting to set out policy reaction functions into a real time forecasting setting to assess if actual primary surpluses strongly deviate from ex-ante predicted values. In each period the primary surplus that is predicted by means of the policy rule can be contrasted with its realization to obtain an early indicator of changes in the conduct of fiscal policy.

Bibliography

- Abeyasinghe, T. and Jayawickrama, A. (2013). A segmented trend model to assess fiscal sustainability: The US experience 1929-2009, *Empirical Economics* **44**(3): 1129–1141.
- Afonso, A. (2005). Fiscal sustainability: The unpleasant European case, *FinanzArchiv: Public Finance Analysis* **61**(1): 19–44.
- Ahmed, S. and Rogers, J. H. (1995). Government budget deficits and trade deficits are present value constraints satisfied in long-term data?, *Journal of Monetary Economics* **36**(2): 351–374.
- Ahn, S. C. and Horenstein, A. R. (2013). Eigenvalue ratio test for the number of factors, *Econometrica* **81**(3): 1203–1227.
- Aizenman, J., Hutchison, M. and Lothian, J. (2013). The European sovereign debt crisis: Background and perspectives, overview of the special issue, *Journal of International Money and Finance* **34**(0): 1 – 5.
- Alexopoulos, M. and Cohen, J. (2009). Uncertain times, uncertain measures. University of Toronto Working Paper 352.
- Allen, F. and Michaely, R. (2003). Payout policy, in G. Constantinides, M. Harris and R. Stulz (eds), *North-Holland Handbook of Economics*, Elsevier, pp. 337–429.
- Andersen, T. G., Bollerslev, T., Christoffersen, P. F. and Diebold, F. X. (2006). *Volatility and correlation forecasting*, Vol. 1 of *Handbook of Economic Forecasting*, Elsevier, chapter 15, pp. 777–878.
- Antonini, M., Lee, K. and Pires, J. (2013). Public sector debt dynamics: The persistence and sources of shocks to debt in 10 EU countries, *Journal of Money, Credit and Banking* **45**(2-3): 277–298.
- Arellano, C., Bai, Y. and Kehoe, P. J. (2012). Financial markets and fluctuations in uncertainty. Federal Reserve Bank of Minneapolis, March 2012, Staff Report.
- Artis, M. J. and Hoffmann, M. (2008). Financial globalization, international business cycles and consumption risk sharing, *Scandinavian Journal of Economics* **110**(3): 447–471.
- Bachmann, R. and Bayer, C. (2013). ‘Wait-and-see’ business cycles?, *Journal of Monetary Economics* **60**(6): 704–719.

- Bachmann, R., Elstner, S. and Sims, E. R. (2013). Uncertainty and economic activity: Evidence from business survey data, *American Economic Journal: Macroeconomics* **5**(2): 217–49.
- Bachmann, R. and Moscarini, G. (2012). Business cycles and endogenous uncertainty. Yale University, Manuscript.
- Baglan, D. and Yoldas, E. (2013). Government debt and macroeconomic activity: A predictive analysis for advanced economies, *Finance and Economics Discussion Series 05*, Board of Governors of the Federal Reserve System.
- Bai, J. and Ng, S. (2002). Determining the number of factors in approximate factor models, *Econometrica* **70**(1): 191–221.
- Bai, J. and Ng, S. (2007). Determining the number of primitive shocks in factor models, *Journal of Business and Economic Statistics* **25**(1): 52–60.
- Bai, J. and Ng, S. (2008). Large dimensional factor analysis, *Foundations and Trends in Econometrics* **3**(2): 89–163.
- 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.
- Baillie, R. T., Chung, C.-F. and Tieslau, M. A. (1996). Analysing inflation by the fractionally integrated ARFIMA-GARCH model, *Journal of Applied Econometrics* **11**(1): 23–40.
- Baker, S. R., Bloom, N. and Davis, S. J. (2013). Measuring economic policy uncertainty. Stanford University, Manuscript.
- Ball, L. (1992). Why does high inflation raise inflation uncertainty?, *Journal of Monetary Economics* **29**: 371–388.
- Bañbura, M., Giannone, D. and Reichlin, L. (2011). Nowcasting, in M. P. Clements and D. F. Hendry (eds), *Oxford Handbook on Economic Forecasting*, Oxford University Press, Oxford.
- Bañbura, M. and Modugno, M. (2012). Maximum likelihood estimation of factor models on datasets with arbitrary pattern of missing data, *Journal of Applied Econometrics* pp. 1–27.
- Bansal, R., Khatchatrian, V. and Yaron, A. (2005). Interpretable asset markets?, *European Economic Review* **49**(3): 531–560.
- Bansal, R., Kiku, D. and Yaron, A. (2010). Long run risks, the macroeconomy, and asset prices, *American Economic Review* **100**(2): 542–46.

- Bansal, R. and Yaron, A. (2004). Risks for the long run: A potential resolution of asset pricing puzzles, *Journal of Finance* **59**(4): 1481–1509.
- Barro, R. J. (1979). On the determination of the public debt, *Journal of Political Economy* **87**(5): 940–71.
- Barro, R. J. and Ursua, J. F. (2008). Macroeconomic crises since 1870, *Brookings Papers on Economic Activity* **39**(1 (Spring)): 255–350.
- Basu, S. and Bundick, B. (2012). Uncertainty shocks in a model of effective demand, *NBER Working Paper 18420*, National Bureau of Economic Research, Inc.
- Bekaert, G., Hoerova, M. and Lo Duca, M. (2013). Risk, uncertainty and monetary policy, *Journal of Monetary Economics* **60**(7): 771–788.
- Binsbergen, J. H. v. and Koijen, R. S. J. (2010). Predictive regressions: A present-value approach, *Journal of Finance* **65**(4): 1439–1471.
- Bloom, N. (2009). The impact of uncertainty shocks, *Econometrica* **77**(3): 623–686.
- Bloom, N., Floetotto, M., Jaimovich, N., Saporta-Eksten, I. and Terry, S. J. (2012). Really uncertain business cycles, *NBER Working Paper 18245*, National Bureau of Economic Research, Inc.
- Bohn, H. (1995). The sustainability of budget deficits in a stochastic economy, *Journal of Money, Credit and Banking* **27**(1): 257–71.
- Bohn, H. (1998). The behavior of U.S. public debt and deficits, *The 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 R. Neck and J.-E. Sturm (eds), *Sustainability of Public Debt*, MIT Press, chapter 2, pp. 15–49.
- Born, B. and Pfeifer, J. (2014). Policy risk and the business cycle, *Journal of Monetary Economics* (forthcoming).
- Boswijk, H. P. (1995). Efficient inference on cointegration parameters in structural error correction models, *Journal of Econometrics* **69**(1): 133–158.
- Boudoukh, J., Michaely, R., Richardson, M. and Roberts, M. R. (2007). On the importance of measuring payout yield: Implications for empirical asset pricing, *Journal of Finance* **62**: 877–915.

- Breitung, J. (2000). The local power of some unit root tests for panel data, in B. H. Baltagi, T. B. Fomby and R. C. Hill (eds), *Nonstationary Panels, Panel Cointegration, and Dynamic Panels (Advances in Econometrics, Vol. 15*, Emerald Group Publishing Limited, pp. 161–177.
- Breitung, J. and Das, S. (2005). Panel unit root tests under cross-sectional dependence, *Statistica Neerlandica* **59**(4): 414–433.
- Broda, S., Carstensen, K. and Paoletta, M. (2009). Assessing and improving the performance of nearly efficient unit root tests in small samples, *Econometric Reviews* **28**(5): 468–494.
- Burns, A. F. and Mitchell, W. C. (1946). *Measuring business cycles*, NBER.
- Campbell, J. Y. (1991). A variance decomposition for stock returns, *The Economic Journal* **101**(405): 157–179.
- Campbell, J. Y. (1999). Asset prices, consumption, and the business cycle, in J. B. Taylor and M. Woodford (eds), *Handbook of Macroeconomics*, Vol. 1, part 3, Elsevier, chapter 19, pp. 1231 – 1303.
- Campbell, J. Y. (2008). Viewpoint: Estimating the equity premium, *Canadian Journal of Economics* **41**(1): 1–21.
- Campbell, J. Y. and Shiller, R. J. (1988). The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* **1**(3): 195–228.
- Campbell, J. Y. and Vuolteenaho, T. (2004). Inflation illusion and stock prices, *The American Economic Review* **94**(2): 19–23.
- Campbell, J. Y. and Yogo, M. (2005). Implementing the econometric methods in ‘Efficient tests of stock return predictability’, *Technical appendix*, University of Pennsylvania.
- Campbell, J. Y. and Yogo, M. (2006). Efficient tests of stock return predictability, *Journal of Financial Economics* **81**(1): 27–60.
- Canzoneri, M. B., Cumby, R. E. and Diba, B. T. (2001). Is the price level determined by the needs of fiscal solvency?, *American Economic Review* **91**(5): 1221–1238.
- Cappé, O., Godsill, S. and Moulines, E. (2007). An overview of existing methods and recent advances in Sequential Monte Carlo, *Proceedings of the IEEE* **95**(5).
- Cavaliere, G. (2005). Limited time series with a unit root, *Econometric Theory* **21**(05): 907–945.
- Cavaliere, G. and Xu, F. (2014). Testing for unit roots in bounded time series, *Journal of Econometrics* **178**(Part 2): 259–272.

-
- Cavanagh, C. L., Elliott, G. and Stock, J. H. (1995). Inference in models with nearly integrated regressors, *Econometric Theory* **11**: 1131–1147.
- Chamberlain, G. and Rothschild, M. (1983). Arbitrage, factor structure, and mean-variance analysis on large asset markets, *Econometrica* **51**(5): 1281–1304.
- Choi, I. (2001). Unit root tests for panel data, *Journal of International Money and Finance* **20**(2): 249–272.
- Claeys, P., Ramos, R. and Suriñach, J. (2008). Fiscal sustainability across government tiers, *International Economics and Economic Policy* **5**(1): 139–163.
- Clark, T. E. and West, K. D. (2006). Using out-of-sample mean squared prediction errors to test the martingale difference hypothesis, *Journal of Econometrics* **135**(1-2): 155–186.
- Doucet, A. and Johansen, A. M. (2009). A tutorial on particle filtering and smoothing: Fifteen years later, in D. Crisan and B. Rozovsky (eds), *Handbook of Non-linear Filtering*, Oxford University Press.
- Doz, C., Giannone, D. and Reichlin, L. (2012). A quasi-maximum likelihood approach for large, approximate dynamic factor models, *The Review of Economics and Statistics* **94**(4): 1014–1024.
- Elder, J. and Serletis, A. (2010). Oil price uncertainty, *Journal of Money, Credit and Banking* **42**(6): 1137–1159.
- Elliott, G., Rothenberg, T. J. and Stock, J. H. (1996). Efficient tests for an autoregressive unit root, *Econometrica* **64**(4): 813–36.
- Elliott, G. and Stock, J. H. (1994). Inference in time series regression when the order of integration of a regressor is unknown, *Econometric Theory* **10**(3-4): 672–700.
- Epstein, L. G. and Zin, S. E. (1989). Substitution, risk aversion, and the temporal behavior of consumption and asset returns: A theoretical framework, *Econometrica* **57**(4): 937–69.
- Ericsson, N. R. and MacKinnon, J. G. (2002). Distributions of error correction tests for cointegration, *Econometrics Journal* **5**(2): 285–318.
- Evans, M. and Fuertes, A. (2010). Understanding the dynamics of the US external position, *Working Papers gueconwpa 10-10-05*, Georgetown University, Department of Economics.
- Fama, E. F. and French, K. R. (2001). Disappearing dividends: Changing firm characteristics or lower propensity to pay, *Journal of Financial Economics Letters* **60**: 3–43.

- Favero, C. A., Gozluklu, A. E. and Tamoni, A. (2011). Demographic trends, the dividend-price ratio, and the predictability of long-run stock market returns, *Journal of Financial and Quantitative Analysis* **46**(05): 1493–1520.
- Fernández-Villaverde, J., Guerrón-Quintana, P., Kuester, K. and Rubio-Ramírez, J. (2013). Fiscal volatility shocks and economic activity. University of Pennsylvania, Manuscript.
- Fernández-Villaverde, J., Guerrón-Quintana, P., Rubio-Ramírez, J. F. and Uribe, M. (2011). Risk matters: The real effects of volatility shocks, *American Economic Review* **101**(6): 2530–61.
- Fernández-Villaverde, J. and Rubio-Ramírez, J. F. (2010). Macroeconomics and volatility: Data, models, and estimation, *NBER Working Paper 16618*, National Bureau of Economic Research, Inc.
- Fisher, R. (1934). *Statistical Methods for Research Workers*, 5th edn, Oliver and Boyd, Edinburgh.
- Forni, M., Giannone, D., Lippi, M. and Reichlin, L. (2007). Opening the black box: Structural factor models with large cross-sections. European Central Bank, Working Paper 702.
- Forni, M. and Reichlin, L. (1998). Let's get real: A factor analytical approach to disaggregated business cycle dynamics, *Review of Economic Studies* **65**(3): 453–73.
- Frankel, J. A. and Rose, A. K. (1996). A panel project on Purchasing Power Parity: Mean reversion within and between countries, *Journal of International Economics* **40**(1-2): 209–224.
- Friedman, M. (1977). Nobel lecture: Inflation and unemployment, *The Journal of Political Economy* **85**(3): 451–472.
- Geanakoplos, J., Magill, M. and Quinzii, M. (2004). Demography and the long-run predictability of the stock market, *Brookings Papers on Economic Activity* **2004**(1): 241–307.
- Gengenbach, C., Palm, F. and Urbain, J.-P. (2010). Panel unit root tests in the presence of cross-sectional dependencies: Comparison and implications for modelling, *Econometric Reviews* **29**(2): 111–145.
- Gennotte, G. and Marsh, T. (1992). Variations in economic uncertainty and risk premium on capital assets, *European Economic Review* **37**: 1021–1041.
- Ghosh, A. R., Kim, J. I., Mendoza, E. G., Ostry, J. D. and Qureshi, M. S. (2013). Fiscal fatigue, fiscal space and debt sustainability in advanced economies, *Economic Journal* **123**(566): F4–F30.

- Giannone, D., Reichlin, L. and Sala, L. (2002). Tracking Greenspan: Systematic and unsystematic monetary policy revisited, *CEPR Discussion Paper 3550*, Centre for Economic Policy Research.
- Giannone, D., Reichlin, L. and Sala, L. (2004). Monetary policy in real time, *NBER Macroeconomics Annual* **19**: 161–224.
- Gilchrist, S., Sim, J. W. and Zakrajšek, E. (2014). Uncertainty, financial frictions, and investment dynamics, *NBER Working Paper 20038*, National Bureau of Economic Research, Inc.
- Gilchrist, S., Yankov, V. and Zakrajšek, E. (2009). Credit market shocks and economic fluctuations: Evidence from corporate bond and stock markets, *Journal of Monetary Economics* **56**(4): 471–492.
- Gilchrist, S. and Zakrajšek, E. (2012). Credit spreads and business cycle fluctuations, *American Economic Review* **102**(4): 1692–1720.
- Giordani, P. and Söderlind, P. (2003). Inflation forecast uncertainty, *European Economic Review* **47**: 1037–1059.
- Giovannini, A. (1989). Uncertainty and liquidity, *Journal of Monetary Economics* **23**: 239–258.
- Gonçalves, S. and Kilian, L. (2004). Bootstrapping autoregressions with conditional heteroskedasticity of unknown form, *Journal of Econometrics* **123**: 89–120.
- Goyal, A. and Welch, I. (2003). Predicting the equity premium with dividend ratios, *Management Science* **49**(5): 639–654.
- Goyal, A. and Welch, I. (2008). A comprehensive look at the empirical performance of equity premium prediction, *Review of Financial Studies* **21**(4): 1455–1508.
- Greiner, A., Köller, U. and Semmler, W. (2006). Testing the sustainability of German fiscal policy: Evidence for the period 1960–2003, *Empirica* **33**(2): 127–140.
- 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): 220–236.
- Grier, K. B. and Perry, M. J. (1998). On inflation and inflation uncertainty in the G7 countries, *Journal of International Money and Finance* **17**(4): 671–689.
- Grier, K. B. and Perry, M. J. (2000). The effects of real and nominal uncertainty on inflation and output growth: Some GARCH-M evidence, *Journal of Applied Econometrics* **15**: 45–58.

- Grimme, C., Henzel, S. R. and Wieland, E. (2014). Inflation uncertainty revisited: A proposal for robust measurement, *Empirical Economics* (forthcoming).
- Hakkio, C. S. and Rush, M. (1991). Is the budget deficit “Too Large?”, *Economic Inquiry* **29**(3): 429–45.
- Hallin, M. and Liška, R. (2007). Determining the number of factors in the general dynamic factor model, *Journal of the American Statistical Association* **102**: 603–617.
- Hamilton, J. D. and Flavin, M. A. (1986). On the limitations of government borrowing: A framework for empirical testing, *American Economic Review* **76**(4): 808–19.
- Hartmann, M. and Herwartz, H. (2012). Consolidation first - About twin deficits and the causal relation between fiscal budget and current account imbalances, *Economics Bulletin* **32**(4): 3313–3319.
- Henzel, S. and Rengel, M. (2013). Dimensions of macroeconomic uncertainty: A common factor analysis, *Ifo Working Paper No. 167*, Ifo Institute for Economic Research at the University of Munich.
- Henzel, S. and Rengel, M. (2014). Dimensions of macroeconomic uncertainty: A common factor analysis, *CESifo Working Paper No. 4991*, CESifo Group Munich.
- Herwartz, H. and Morales-Arias, L. (2009). In-sample and out-of-sample properties of international stock return dynamics conditional on equilibrium pricing factors, *European Journal of Finance* **15**(1): 1–28.
- Herwartz, H. and Rengel, M. (2014). Size corrected inference in fiscal policy reaction functions: A three country assessment. Georg-August-University Göttingen, Department of Economics, Manuscript, See Paper 3 of this thesis.
- Herwartz, H., Rengel, M., and Xu, F. (2013). Persistence in the price-to-dividend ratio and its macroeconomic fundamentals, *Conference Paper – Jahrestagung des Vereins für Socialpolitik 2013 79860*, Verein für Socialpolitik.
- Herwartz, H. and Siedenburg, F. (2008). Homogenous panel unit root tests under cross sectional dependence: Finite sample modifications and the wild bootstrap, *Computational Statistics & Data Analysis* **53**(1): 137–150.
- Herwartz, H., Siedenburg, F. and Walle, Y. M. (2014). Heteroskedasticity robust panel unit root testing under variance breaks in pooled regressions, *Econometric Reviews* (forthcoming).
- Herwartz, H. and Xu, F. (2014). On housing booms and credit market conditions. Georg-August-University Göttingen, Department of Economics, Manuscript.
- Hjalmarsson, E. (2008). The Stambaugh bias in panel predictive regressions, *Finance Research Letters* **5**(1): 47–58.

-
- Hjalmarsson, E. (2010). Predicting global stock returns, *Journal of Financial and Quantitative Analysis* **45**(01): 49–80.
- Im, K. S., Pesaran, M. H. and Shin, Y. (2003). Testing for unit roots in heterogeneous panels, *Journal of Econometrics* **115**(1): 53–74.
- IMF and WB (2001). Guidelines for public debt management, *Online Publication*, The International Monetary Fund and the World Bank.
URL: <https://www.imf.org/external/np/mae/pdebt/2000/eng/Index.htm>
- Johansen, S. (1992). Cointegration in partial systems and the efficiency of single-equation analysis, *Journal of Econometrics* **52**(3): 389–402.
- Jurado, K., Ludvigson, S. C. and Ng, S. (2014). Measuring uncertainty, *American Economic Review* (forthcoming).
- Kandel, S. and Stambaugh, R. (1990). Expectations and volatility of consumption and asset returns, *Review of Financial Studies* **3**: 207–232.
- Kilian, L. (1998). Small-sample confidence intervals for impulse response functions, *The Review of Economics and Statistics* **80**: 218–230.
- Kilian, L. (1999). Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?, *Journal of Applied Econometrics* **14**(5): 491–510.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the Crude Oil market, *American Economic Review* **99**(3): 1053–1069.
- Kim, C.-J. and Park, C. (2013). Disappearing dividends: Implications for the dividend-price ratio and return predictability, *Journal of Money, Credit and Banking* **45**: 933–952.
- Kiviet, J. F. (1995). On bias, inconsistency, and efficiency of various estimators in dynamic panel data models, *Journal of Econometrics* **68**(1): 53–78.
- Knotek II, E. S. and Khan, S. (2011). How do households respond to uncertainty shocks?, *Economic Review* (Q II). Federal Reserve Bank of Kansas City.
- Koijen, R. S. and Nieuwerburgh, S. V. (2011). Predictability of returns and cash flows, *Annual Review of Financial Economics* **3**(1): 467–491.
- Kose, M. A., Otrok, C. and Prasad, E. (2012). Global business cycles: Convergence or decoupling?, *International Economic Review* **53**(2): 511–538.
- Kose, M. A., Otrok, C. and Whiteman, C. H. (2003). International business cycles: World, region, and country-specific factors, *American Economic Review* **93**(4): 1216–1239.

- Kremers, J. J. M. (1988). Long-run limits on the US federal debt, *Economics Letters* **28**(3): 259–262.
- Kremers, J. J. M., Ericsson, N. R. and Dolado, J. J. (1992). The power of cointegration tests, *Oxford Bulletin of Economics and Statistics* **54**(3): 325–48.
- Kwiatkowski, D., Phillips, P. C. B., Schmidt, P. and Shin, Y. (1992). Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root?, *Journal of Econometrics* **54**(1-3): 159–178.
- Lacerda, F. and Santa-Clara, P. (2010). Forecasting dividend growth to better predict returns. Universidade Nova de Lisboa, Manuscript.
- Lagarias, J. C., Reeds, J. A., Wright, M. H. and Wright, P. E. (1998). Convergence properties of the nelder-mead simplex method in low dimensions, *SIAM Journal of Optimization* **9**: 112–147.
- Lahiri, K. and Sheng, X. (2010). Measuring forecast uncertainty by disagreement: The missing link, *Journal of Applied Econometrics* **25**(4): 514–538.
- Lane, P. R. (2012). The European sovereign debt crisis, *Journal of Economic Perspectives* **26**(3): 49–68.
- Lettau, M., Ludvigson, S. C. and Wachter, J. A. (2008). The declining equity premium: What role does macroeconomic risk play?, *Review of Financial Studies* **21**: 1653–1687.
- Lettau, M. and Van Nieuwerburgh, S. G. (2008). Reconciling the return predictability evidence, *Review of Financial Studies* **21**: 1607–1652.
- Levin, A., Lin, C.-F. and Chu, C.-S. J. (2002). Unit root tests in panel data: Asymptotic and finite-sample properties, *Journal of Econometrics* **108**(1): 1–24.
- Lewellen, J. (2004). Predicting returns with financial ratios, *Journal of Financial Economics* **74**(2): 209–235.
- Liesenfeld, R. and Richard, J.-F. (2003). Univariate and multivariate stochastic volatility models: Estimation and diagnostics, *Journal of Empirical Finance* **10**(4): 505–531.
- Liu, R. Y. (1988). Bootstrap procedures under some non-i.i.d. models, *The Annals of Statistics* **16**(4): 1696–1708.
- Lucas, Robert E, J. (1978). Asset prices in an exchange economy, *Econometrica* **46**(6): 1429–45.
- Lukkezen, J. and Rojas-Romagosa, H. (2012). When is debt sustainable?, *CPB Discussion Paper 212*, CPB Netherlands Bureau for Economic Policy Analysis.

- MacDonald, R. and Speight, A. E. H. (1990). The intertemporal government budget constraint in the U.K., 1961-1986, *The Manchester School* **58**(4): 329–347.
- Mankiw, G. N. and Shapiro, M. D. (1985). Trends, random walks, and tests of the permanent income hypothesis, *Journal of Monetary Economics* **16**(2): 165–174.
- Mankiw, G. N. and Shapiro, M. D. (1986). Do we reject too often? Small sample properties of tests of rational expectations models, *Economics Letters* **20**(2): 139–145.
- Marinheiro, C. (2006). The sustainability of Portuguese fiscal policy from a historical perspective, *Empirica* **33**(2): 155–179.
- Mark, N. C. (1995). Exchange rates and fundamentals: Evidence on long-horizon predictability, *American Economic Review* **85**(1): 201–18.
- Mendoza, E. G. and Ostry, J. D. (2008). International evidence on fiscal solvency: Is fiscal policy “responsible”?, *Journal of Monetary Economics* **55**(6): 1081–1093.
- Mitchell, B. R. (1988). *British historical statistics*, Cambridge University Press, Cambridge.
- Moon, H. R. and Phillips, P. C. (2000). Estimation of autoregressive roots near unity using panel data, *Econometric Theory* **16**(06): 927–997.
- Morgan, J. (1996). *Riskmetrics*, 4th edn, Technical Documents, New York.
- Nelson, C. and Kim, M. (1993). Predictable stock returns: The role of small sample bias, *Journal of Finance* **48**: 641–661.
- Newey, W. K. and West, K. D. (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix, *Econometrica* **55**(3): 703–08.
- Newey, W. K. and West, K. D. (1994). Automatic lag selection in covariance matrix estimation, *Review of Economic Studies* **61**(4): 631–53.
- Nickell, S. J. (1981). Biases in dynamic models with fixed effects, *Econometrica* **49**(6): 1417–26.
- Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues, *The Review of Economics and Statistics* **92**(4): 1004–1016.
- Park, C. (2010). When does the dividend-price ratio predict stock returns?, *Journal of Empirical Finance* **17**(1): 81–101.
- Pastor, L., Stambaugh, R. F. and Taylor, L. A. (2014). Scale and skill in active management, *Journal of Financial Economics* (forthcoming).
- Paye, B. S. and Timmermann, A. (2006). Instability of return prediction models, *Journal of Empirical Finance* **13**: 274–315.

- Perron, P. and Vogelsang, T. J. (1992). Nonstationarity and level shifts with an application to Purchasing Power Parity, *Journal of Business & Economic Statistics* **10**(3): 301–20.
- Pesaran, M. H. (2006). Estimation and inference in large heterogeneous panels with a multifactor error structure, *Econometrica* **74**(4): 967–1012.
- Pesaran, M. H. and Smith, R. (1995). Estimating long-run relationships from dynamic heterogeneous panels, *Journal of Econometrics* **68**(1): 79–113.
- Phillips, P. C. B. and Perron, P. (1988). Testing for a unit root in time series regression, *Biometrika* **75**(2): 335–346.
- Picard, R. R. and Cook, R. D. (1984). Cross-validation of regression models, *Journal of the American Statistical Association* **79**(387): 575–583.
- Potrafke, N. and Reischmann, M. (2014). Fiscal transfers and fiscal sustainability, *Journal of Money, Credit and Banking* (forthcoming).
- Quintos, C. E. (1995). Sustainability of the deficit process with structural shifts, *Journal of Business & Economic Statistics* **13**(4): 409–17.
- Ramey, G. and Ramey, V. A. (1995). Cross-country evidence on the link between volatility and growth, *The American Economic Review* **85**(5): 1138–1151.
- Rapach, D. E., Strauss, J. K. and Zhou, G. (2010). Out-of-sample equity premium prediction: Combination forecasts and links to the real economy, *Review of Financial Studies* **23**(2): 821–862.
- Rapach, D. E. and Wohar, M. E. (2006). Structural breaks and predictive regression models of aggregate US stock returns, *Journal of Financial Econometrics* **4**: 238–274.
- Rich, R. and Tracy, J. (2010). The relationship among expected inflation, disagreement, and uncertainty: Evidence from matched point and density forecasts, *The Review of Economics and Statistics* **92**(1): 200–207.
- Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*, 2nd edn, Springer Texts in Statistics.
- Robertson, D. and Wright, S. (2006). Dividends, total cash flow to shareholders, and predictive return regressions, *The Review of Economics and Statistics* **88**(1): 91–99.
- Rudd, J. and Whelan, K. (2006). Can rational expectations sticky-price models explain inflation dynamics?, *American Economic Review* **96**(1): 303–320.
- Rytchkov, O. (2012). Filtering out expected dividends and expected returns, *Quarterly Journal of Finance* **2**(3): 1250012:1–1250012:56.

- Sargent, T. and Sims, C. (1977). Business cycle modelling without pretending to have too much a priori economic theory, in C. Sims (ed.), *New Methods in Business Cycle Research*, Federal Reserve Bank of Minneapolis, Minneapolis.
- Sbordone, A. M. (2002). Prices and unit labor costs: A new test of price stickiness, *Journal of Monetary Economics* **49**(2): 265–292.
- Schaal, E. (2012). Uncertainty, productivity and unemployment in the great recession. Federal Reserve Bank of Minneapolis, Manuscript.
- Shiller, R. J. (1992). *Market Volatility*, MIT Press Books, The MIT Press.
- Shiller, R. J. (2005). *Irrational Exuberance*, Books, 2nd edn, Princeton University Press, Princeton.
- Shin, D. W. and So, B. S. (2001). Recursive mean adjustment for unit root tests, *Journal of Time Series Analysis* **22**(5): 595–612.
- Solomou, S. N. (2008). Kuznets Swings, in S. N. Durlauf and L. E. Blume (eds), *New Palgrave Dictionary of Economics*, 2nd edn, Vol. 4, Palgrave Macmillan.
- Stambaugh, R. F. (1999). Predictive regressions, *Journal of Finance* **54**: 375–421.
- Stock, J. H. and Watson, M. W. (1993). A simple estimator of cointegrating vectors in higher order integrated systems, *Econometrica* **61**(4): 783–820.
- Stock, J. H. and Watson, M. W. (2002). Macroeconomic forecasting using diffusion indexes, *Journal of Business & Economic Statistics* **20**(2): 147–162.
- Sul, D., Phillips, P. C. B. and Choi, C.-Y. (2005). Prewhitening bias in HAC estimation, *Oxford Bulletin of Economics and Statistics* **67**(4): 517–546.
- Thompson, S. B. (2011). Simple formulas for standard errors that cluster by both firm and time, *Journal of Financial Economics* **99**(1): 1–10.
- Torous, W., Valkanov, R. and Yan, S. (2004). On predicting stock returns with nearly integrated explanatory variables, *Journal of Business* **77**: 937–966.
- Trehan, B. and Walsh, C. E. (1988). Common trends, the government’s budget constraint, and revenue smoothing, *Journal of Economic Dynamics and Control* **12**: 425 – 444.
- Warren, G. and Pearson, F. (1935). *Gold and Prices*, Books, John Wiley and Sons, New York.
- Westerlund, J. (2007). Testing for error correction in panel data, *Oxford Bulletin of Economics and Statistics* **69**(6): 709–748.
- 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.

- Woodford, M. (1998). *Comment on Cochrane's 'A Frictionless View of U.S. Inflation'*, MIT Press, pp. 390–418.
- Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis, *The Annals of Statistics* **14**(4): 1261–1295.
- Zarnowitz, V. and Lambros, L. A. (1987). Consensus and uncertainty in economic predictions, *The Journal of Political Economy* **95**(3): 591–621.

Lebenslauf

- Seit 09/2012 Wissenschaftlicher Mitarbeiter am Lehrstuhl für Ökonometrie der Georg-August Universität Göttingen (Prof. Dr. Helmut Herwartz)
- 09/2010–08/2012 Wissenschaftlicher Mitarbeiter am Institut für Statistik und Ökonometrie der Christian-Albrechts-Universität zu Kiel (Prof. Dr. Helmut Herwartz)
- Seit 10/2010 Promotionsstudium in *Quantitative Economics* an der Christian-Albrecht-Universität zu Kiel
- 10/2005–09/2010 Studium der Volkswirtschaftslehre an der Christian-Albrechts-Universität zu Kiel
Abschluss: Diplom-Volkswirt mit dem Schwerpunkt *Quantitative Wirtschaftsforschung*
- 06/2005 Abitur am Gymnasium am Treckfahrtstief in Emden

Göttingen, Dezember 2014