LUDWIG MAXIMILIAN UNIVERSITY
FACULTY OF ECONOMICS

Determinants and Implications of Local Fiscal Policy

PHD THESIS

DANIEL STÖHLKER

Advisor: Clemens Fuest
Co-advisor: Panu Poutvaara

Munich, 2020

Ludwig Maximilian University

Determinants and Implications of Local Fiscal Policy

Inaugural-Dissertation
zur Erlangung des akademischen Grades
Doctor oeconomiae publicae (Dr. oec. publ.)
an der Volkswirtschaftlichen Fakultät
an der Ludwig-Maximilians-Universität München

2020

VORGELEGT VON

DANIEL STÖHLKER

Referent: Clemens Fuest
Korreferent: Panu Poutvaara
Drittprüfer: Dominik Sachs

Tag der mündlichen Prüfung: 19. Januar 2021
Promotionsabschlussberatung: 3. Februar 2021

Acknowledgements

In these four years at ifo and Ludwig Maximilian University in Munich I have learned a lot about economics, about life, and myself. First and foremost, I would like to express my deep gratitude to my main advisors: Clemens Fuest gave me important sense of direction during my early attempts at original research and much needed support and encouragement towards the end. Many insightful conversations we had have helped me a great deal to focus on intuition and relevance of my results. Panu Poutvaara was always available and interested in what I did. Illuminating discussions with him about technical aspects of my work and the political contexts have helped to flag those parts that needed more careful elaborations and, ultimately, spurred major advances. I was fortunate to be Dominik Sachs' teaching assistant for a couple of semesters. From Dominik, who is the third reader of this thesis, I learned much of the formal rigor I have. Francesca Caselli and Philippe Wingender were fantastic mentors during my stay at the International Monetary Fund and committed co-authors of what became the first chapter of this thesis.

Countless other individuals have contributed to the development of this thesis in one way or another, both in Munich and at various other places that I had the privilege to visit. Too many to mention them all, but the special debt I owe to my colleagues and friends goes beyond it. Marc Fabel, Stefan Lautenbacher, and Patrick Reich were essential parts of my life in Munich, both at work and at leisure. Lea Immel and Florian Neumeier were wonderful officemates, entertaining and always generous with their time and feedback.

I owe the greatest debt to Cornelia and Heino, my parents. Without their unconditional support, patience, and understanding since the very first day of my studies it would have been impossible to achieve this. Thanks.

Munich, 2020

Daniel Stöhlker

Contents

Acknowledgements **1**

Preface **9**

1 Individual Treatment Effects of Budget Balance Rules **15**

1.1 Introduction . . . 15
1.2 Synthetic Control Method and Model Specification . . . 16
1.3 Data . . . 19
1.4 Individual Treatment Effects . . . 20
1.5 Quality of the SCM Estimates and Robustness Exercises . . . 21
1.6 Effectiveness of Fiscal Rules . . . 24
1.7 Conclusions . . . 27

2 Tax Cuts Starve the Beast: Evidence from Germany **29**

2.1 Introduction . . . 29
2.2 Related Literature and Contribution . . . 31
2.3 Fiscal Federalism in Germany . . . 32
2.4 Constructing an Exogenous Tax Shock Series . . . 33
2.5 Conditions for Causal Inference . . . 37
2.5.1 Accuracy of Revenue Projections . . . 37
2.5.2 Exogeneity of Tax Shock Series . . . 37
2.5.3 No Anticipation Effects . . . 38
2.6 Empirical Approach . . . 38
2.7 Results . . . 40
2.8 Extensions and Robustness Checks . . . 41
2.8.1 Heterogeneous Spending Adjustments . . . 41
2.8.2 Spending on Sub-Categories . . . 45
2.9 A Note on External Validity . . . 47
2.10 Conclusions . . . 47

3 The Quality of Local Public Good Provision and Electoral Support **49**

3.1 Introduction . . . 49
3.2 Local Electoral System in England . . . 51
3.3 Data and Quality Indicators of Local Public Good Provision . . . 53

3.4 Estimation Strategy and Baseline Results 59
3.5 Robustness Tests 60
3.6 Conclusion 62

4 The Tax-Elasticity of Corporate Profits under Formula Apportionment **65**
4.1 Introduction 65
4.2 Related Literature 67
4.3 Business Taxation in Germany 68
4.4 Data 69
4.5 Empirical Approach 70
4.6 Results 71
4.7 Profit Shifting Responses by Industries 72
4.8 Conclusion 74

Appendices **77**

A Details: Chapter 1 **79**
A.1 Details about National Fiscal Rules of Treated Countries 79
A.2 Matrix of Weights from Synthetic Control Approach and Additional Results 81

B Details: Chapter 4 **89**
B.1 Robustness Check for Trimmed Data 89
B.2 Industry-Specific Results 90

Bibliography **91**

List of Figures

1.1 Individual Treatment Effects of Fiscal Rules by Level of Counter-factual Government Balance . . . 21
1.2 Assessment of the Quality of the Pre-Treatment Fit in the Baseline SC Specification . . 22
1.3 Assessment of the Quality of the Pre-Treatment Fit in the Robustness SC Specification without Income Group Conditioning . . . 23

2.1 The Series of Exogenous Tax Revenue Shocks at the State-Level . . . 36

3.1 Electoral Structure in England: Local Authorities and Wards . . . 51
3.2 Example Complaint Posted Online on FixMyStreet.com . . . 55
3.3 Geographic Distribution of Complaints between May 2007 and May 2015 . . . 56
3.4 Quality Indicators at the Ward-Level: Fraction of Complaints solved in Different Time Horizons . . . 58
3.5 The Impact of the Quality of Services of Sub-Categories on Re-election Chances . . . 61

4.1 Local Scaling Factors in Germany . . . 69
4.2 The Extent of Tax Base and Payroll Cost Responses to Tax Rate Differentials for the Largest Industries . . . 74

A.1 Matrix of Donor Weights for Baseline SC Specification . . . 82
A.2 Average Treatment Effect of the Fiscal Rule on Governments' Balances . . . 83
A.3 Time Paths of the 10th and the 90th Percentiles of the Treated and Counter-factual Balances 84
A.4 Individual Treatment Effects by Level of Counter-factual Government Balance from Estimation based on Covariates . . . 85
A.5 Average Treatment Effects of Fiscal Rule Introduction from Estimation based on Covariates . . . 86
A.6 Individual Treatment Effects by Level of Counter-factual Government Balance with Shortened Pre- & Post-Treatment Horizon . . . 87

B.1 The Extent of Tax Base and Payroll Cost Responses to Tax Rate Differentials for all Industries . . . 90

List of Tables

1.1 Summary Statistics on Main Economic Variables for Countries with and without Fiscal Rule . . . 20
1.2 Summary Statistics on Fiscal Rule Designs . . . 25
1.3 The Impact of Rule Complexity on the Effect of Fiscal Rules . . . 26
1.4 The Impact of Specific Rule Features on the Effect of Fiscal Rules . . . 27

2.1 Accuracy of Revenue Projections from the Annual Fiscal Report of the Ministry of Finance 37
2.2 Reverse Causality Check: The Impact of Spending Adjustments on the Tax Revenue Shock Series . . . 39
2.3 Baseline Regressions: The Impact of Revenue Shocks on Aggregate Spending . . . 41
2.4 The Impact of Exogenous Tax Revenue Cuts on Aggregate Spending . . . 42
2.5 The Impact of Revenue Shocks on Aggregate Spending for States with High and Low Debt Levels . . . 43
2.6 The Impact of Revenue Shocks on Aggregate Spending for States with and without Coalition Governments . . . 44
2.7 The Impact of Revenue Shocks on Aggregate Spending for States for Left- and Right-Wing Governments . . . 45
2.8 The Impact of Revenue Shocks on Spending on Specific Sub-Categories . . . 46

3.1 Data Cleaning for Removal of Duplicate Complaints . . . 56
3.2 Summary of Descriptive Statistics for Major Complaint Categories . . . 57
3.3 Pair-wise Correlations of Quality Indicators with Alternative Measures . . . 63
3.4 Baseline Results: The Impact of the Quality of Public Services on Re-election Chances 63
3.5 Robustness Checks: The Impact of the Quality of Public Services on Re-election Chances 64

4.1 Descriptive Statistics for Multi-Jurisdictional Entities . . . 70
4.2 The Effect of Tax Rate Differentials on Tax Base and Payroll Costs of Affiliates of MJEs 72
4.3 Differences in the Prevalence of Payment by Hours, Mini-Jobs, and Paid Overtime by Industry . . . 73

A.1 Countries included in the Sample and their National Fiscal Rules . . . 80

B.1 The Effect of Tax Rate Differentials on Tax Base and Payroll Costs of Affiliates of MJEs Based on Trimmed Data . . . 89

Preface

It is well understood that fiscal policy has important implications for many of the most fundamental economic outcome variables, including firms' main short- and long-run business decisions, the shape of the labor market, and household income in more general (see for example Kneller et al., 1999, among many others in this context). While not always in the spotlight of the public attention, fiscal policy at the sub-national and local level is by no means less important than decisions made at the national level. In fact, through local tax policies and spending decisions local governments do not only manage the provision of numerous services that are nearest to our day-to-day lives – including the quality of local road infrastructure and other services, the degree of public safety, educational opportunities, and many more – but also shape the local business environment for firms to invest and hire, or not. Navigating the increasingly complex crosscurrents of local governments' budgets has become an ever more challenging, yet essential, duty of local policy makers. Local administrations find themselves between the sometimes conflicting priorities of political, economic, social, and legal considerations, to name just a few. For example, political and institutional constraints, such as budget balance rules, might require governments to take unpopular decisions such as tax increases in order to increase their fiscal space in the longer run, but can easily stoke discontent and fury among affected citizens.

This thesis studies the determinants and implications of fiscal policies with a particular focus on sub-national and local tax and spending decisions. It consists of four self-contained chapters that cover different aspects in this context. The first two chapters are related to the determinants of fiscal policy. While Chapter 1 is investigating the (sometimes unintended) consequences of national and supra-national fiscal rules for government budget deficits, Chaper 2 is looking at exogenous revenue shocks, both positive and negative ones, and how German states adjust their budgets in response. The last two chapters focus on the implications of local fiscal policies for citizens and firms. Chapter 3 analyzes how the quality of local public good provision varies across English neighborhoods, how it compares with local spending levels and to what extent differences in the quality affect re-election chances of incumbent neighborhood councilors. Lastly, Chapter 4 is turning to the question how tax policy at the municipality-level affects firms' decisions, and in particular their willingness to report profits and pay taxes there.

Chapter 1 of this thesis, which is joint work with Francesca Caselli and Philippe Wingender, is investigating the effects of budget balance rules on governments' deficits. Despite being a frequently used instrument at various levels of government in the EU, in African and Asian countries, and elsewhere, the empirical evidence on its effectiveness in constraining the budget balance is scarce and surprisingly ambiguous. While a couple of authors claim to see a positive effect of rule adoption, Heinemann et al. (2018) in a meta-study analysis point to "strong evidence for the presence of endogeneity and the relevance of accounting for it". For example, countries that have adopted a deficit rule could have a stronger preference for fiscal prudence anyway, even in the absence of the rule. The authors conclude on a more pessimistic note, arguing that properly addressing endogeneity concerns "significantly weakens the seemingly optimistic message with respect to the effectiveness of fiscal rules". Apart from endogeneity issues, existing studies have also made little effort to uncover differences in the size of the treatment effect across countries and over time, and identify potentially important layers of heterogeneity which could also be, at least partially, responsible for the ambiguous empirical findings so far.

In this study, we look behind and around the average treatment effect of budget balance rules. By carefully applying the *Synthetic Control Method* of Abadie et al. (2010), we identify the country-specific impact of the rule and follow thereby closely the advances in the program evaluation literature which has been emphasizing the importance of tracing out the whole distribution of treatment effects (see in particular Heckman et al., 1997, on this issue). Our results do lend support to the hypothesis that countries respond very differently to the introduction of budget balance rules. Two margins of differences across countries are particularly important: first, the counter-factual fiscal balance in absence of the fiscal rule. We show that those countries which would have had large deficits in absence of the rule see their balance improve most while countries with smaller deficits, or even budget surpluses, reduce their fiscal efforts. Thus, taken together, the introduction of the fiscal rule exerts a 'magnet effect' on budget balances in the sense that the tails of the counter-factual deficit distribution are pulled towards the center of it which is determined by the numerical value of the fiscal rule. The second important determinant of the success of budget balance rules is their specific design and their interaction with the overall rule environment of the country. Most importantly, we find that maintaining a smaller number of fiscal rules is essential. We also see that the existence of a credible monitoring mechanism at the supra-national level strengthens the treatment effects of fiscal rules while the existence of an escape clause undermines it completely. This work makes helpful contributions to the policy debate about the effectiveness of fiscal rules and their fiscal consequences in more general. Despite their overall positive effect on deficits, the country-specific responses to budget balance rules can be surprising at times when not properly designed.

Chapter 2 is concerned with the question of how additional revenues impact spending decisions of governments, both in terms of overall spending levels and, more specifically, for what exactly the additional funds are used or where spending is cut. Being closely intertwined with each other, credibly identifying the causal effect of one on the other is empirically challenging and has led, not surprisingly, to contradicting results so far. Together with Clemens Fuest and Florian Neumeier, we use the institutional setting of German fiscal federalism to its advantage in order to explore how fiscal policy reacts to exogenous tax revenue shocks: while states in Germany enjoy full spending autonomy, virtually all taxes are set and collected at the federal level, thus providing an interesting testing ground to assess whether exogenous changes in tax revenues affect aggregate public expenditure as well as specific sub-categories of local government spending.

In order to construct a series of 'truly' exogenous tax revenue shocks at the state-level, we apply the so-called narrative approach, pioneered by Romer and Romer (2009), and exclude changes in tax legislation which are related to the current or expected future economic or fiscal situation. Examples of obviously endogenous tax changes include those that are implemented as part of a broader stimulus package during an economic downturn or changes whose revenues are earmarked to future spending requirements. With this series of exogenous tax shocks at hand, we document that cuts to available tax revenues lead to subsequent reductions in states' public spending of roughly the same amount, following a delay of two to three years. When looking at more specific spending items of local governments, it is found that a revenue decline of one Euro reduces public spending on administration and, with a larger delay, social security, by 30 to 45 cents in each case. Spending on infrastructure declines by ten cents. Interestingly, we find no significant effects on spending on education, legal protection and public safety, or culture.

The results of this analysis feed into the discussion concerning the welfare consequences of ever growing public spending and its effectiveness. More broadly, our results confirm the claim that lower tax revenues restrain government spending at the sub-national level – a hypothesis that has found prominent support among scholars but lacked convincing empirical grounding until now. Nevertheless, the question whether revenue cuts and resulting spending cuts are actually efficiency-enhancing, as it is believed by the supporters of this idea, remains to be discussed. While it is certainly comforting that a large share of the spending cut is made by reducing (supposedly unproductive) administrative spending with no (statistically significant) changes to expenditures on education and science, it needs to be borne in mind that, according to our estimation results, government spending on infrastructure projects as well as health and social security related issues declines as well, something which is likely to not let overall productivity unaffected in the mid- and longer run if not offset by private investment activity.

Chapter 3 of this thesis turns to the implications of local tax and expenditure policies and raises an old question: do voters value the quality of local public goods in their neighborhood, such as whether pavements are clean, whether benches in the park need repair, or whether local roads are in good shape? The question is not easy to answer, particularly so because scarce data on the quality of public services at the local level, e.g. from surveys, has been a limiting factor in the existing literature. Not surprisingly, most studies are restricted to city-specific case studies or cross-sectional variation therein, obviously raising questions concerning the internal and external validity of the available results (Burnett and Kogan, 2017, and Arnold and Carnes, 2012, are recent examples).

This study breaks new ground by using posts from the online complaint website FixMyStreet.com, a platform that enjoys increasing popularity in the United Kingdom with more than two million entries to date since its existence in early 2007. With an application on their cell-phone installed, users can easily send complaints to the platform in order to alert their local authorities about broken street lamps, potholes, fly-tipping, and other issues. Using around 550,000 geo-located complaints that were posted on the website between May 2007 and May 2015, I compute quality indicators for all 7,500 local wards that comprise the whole of England based on how quickly complaints are solved. For example, a relatively large share of complaints that is left unsolved for a year or even longer speaks to a rather poorly functioning complaint management of the local district. As would be expected, the share of complaints not solved after twelve months shows a large negative correlation with targeted district-level spending of councils on the provision of certain services, suggesting that the level of funding of public administration is relevant for the quality of service provision.

I leverage within-neighborhood variation with respect to the quality of local public good provision over time to test whether increases in the share of complaints unsolved for at least a year have an impact on the re-election probability of the local councilor's party or the councilor itself. Councilors are ordinary citizens of the neighborhood, elected as representatives of the ward and the people living in it and play an important role in planning, running, monitoring, and developing local council activities.

Competencies of local district councils are far-reaching and include, for example, rubbish collection, recycling practices, housing development as well as other planning applications such as repair work or local infrastructure maintenance and development. These are services for which most requests are sent to the online complaint platform. The results provide compelling evidence for a strong punishment effect in local council elections: an increase in the fraction of complaints that are solved only after twelve months or never raises the probability that the incumbent party is voted out of office by up to nine percentage points. The effect is also visible when moving to the individual politician-level. The overall effect is driven by reports concerning incidents of fly-tipping and dog fouling on public grounds. Contrary to previous results in this literature, I find that the number of complaints *per se* is not relevant for incumbents' re-election chances.

To some extent, the results of this chapter are resembling the findings of the academic literature on best management practices and effective complaint management of firms: it is often argued that a complaint is rather seen as a chance for firms to restore and strengthen trust and loyalty among customers if the firm's service in response to the complaint "exceeds expectations", e.g. quickly delivering on the promise, potentially providing a personal touch to the interaction and, most importantly, showing the willingness to go the extra mile in order to fix things (Stone, 2011). On the other hand, a poor service quality that does not deliver what was promised or that reveals that no effort is being made to work on the situation, makes it very likely that the customer is switching to other companies – just as in the case of complaints about local public goods and how it is dealt with it.

Finally, Chapter 4, based on joint work with Clemens Fuest and Florian Neumeier, studies the sensitivity of corporate profits to changes in the local corporate income tax under a system of 'Formula Apportionment', an idea that has taken center stage in ongoing debates about a reform of the corporate tax system. The current tax scheme at the national level in the EU is based on the principle of 'Separate Accounting': each affiliate is taxed based on the profits generated within the borders of each jurisdiction. Given the changing nature of business practices, this is not considered suitable anymore. Increasing usage on intangible assets in the production process and an accelerating trend towards digital business models have made it easy for multi-national companies to shift their profits to those jurisdictions that offer the most favorable tax conditions in order to minimize their overall tax obligations. Beyond this background, the European Commission has proposed to replace the current system by a 'Consolidated Common Corporate Tax Base' (CCCTB), according to which profits of all company's affiliates are aggregated to generate a 'consolidated tax base' before being apportioned to the individual jurisdictions again according to an apportionment factor, such as the payroll cost share. It is widely believed, and hoped, that the system of 'Formula Apportionment' makes it more difficult for firms to shift their profits across jurisdictions by manipulating apportionment keys.

The local corporate profit taxation in Germany provides an ideal testing ground for this hypothesis. Profits of firms that are present in more than one local municipality are consolidated at the national level and then apportioned across all jurisdictions according to their relative payroll share in each municipality. Municipalities in Germany – more than twelve thousand in total – can independently decide about their local corporate tax rate via a tax rate scaling factor which varies substantially across municipalities and over time. In order to address the question how firms' profit shifting activities respond to tax rate differentials across jurisdictions, we employ administrative tax return data on all multi-jurisdictional enterprises. An advantage of such data in this context, which is worth to be pointed out, is the fact that it is not limited by usual empirical difficulties, such as incomplete and erroneous information on firms' activities in 'tax havens', as highlighted for example in Tørsløv et al. (2018).

Contrary to popular belief, a range of robust panel-regression results suggests that a system of 'Formula Apportionment' does not mean the end to profit shifting activities of multi-jurisdictional firms. On average, an increase in the local tax rate by ten percentage points leads to a reduction of the local tax base of firms by around twelve per cent. Importantly, we find that adjustments of firms across industries varies greatly: sectors in which workers are predominantly paid by the hours worked, where overtime work is paid, or in which so-called Mini-Jobs are more prevalent find it easier to adjust local payroll costs and respond strongly to local tax rate differentials.

These results have important implications for our understanding of corporate business taxation under 'Formula Apportionment'. The political debate and numerous existing studies (see for example Devereux and Loretz, 2008) have been guided by the assumption that firms cannot adjust to local tax rate changes as easily as under the current system of 'Separate Accounting'. Our results demonstrate that this is not true, both in terms of the average response across all firms and with respect to some sectors in particular. Most importantly, while the response under the system of 'Separate Accounting' is mostly restricted to shifting paper profits, under 'Formula Apportionment' MJEs are locally adjusting and, at least to some extent, relocating real economic activity across jurisdictions, including jobs and wages.

Chapter 1

Individual Treatment Effects of Budget Balance Rules

This chapter is based on joint work with **Francesca Caselli** and **Philippe Wingender**, IMF Working Paper Nr. 2020/274.

1.1 Introduction

Fiscal rules have increasingly been used to strengthen fiscal discipline and constrain budget deficits: between 1985 and 2015, a total of 89 countries from all over the world introduced a budget balance rule.[1] Despite their rapid diffusion, the empirical evidence on the effectiveness of fiscal rules remains mixed because of challenges to establish a causal link with fiscal outcomes.[2] Empirical results suggest that countries that introduce a rule tend to have lower budget deficits (Poterba, 1997, Debrun et al., 2008, Badinger and Reuter, 2017), however, studies that correctly account for rules' endogeneity tend to produce insignificant results (Heinemann et al., 2018, Caselli and Reynaud, 2020). While most studies focus on the average effect of fiscal rules' adoption, some recent papers have investigated the heterogeneous impacts of fiscal rules on budget balances. Caselli and Wingender (2018), by constructing counter-factual distributions of government balances, uncover a 'magnet effect' of the three per cent fiscal ceiling in the European Union: the adoption of the rule reduces the occurrence of both large government deficits and surpluses. This implies not only that the rule had an effect on deficits when it was not complied with the rule, but also on prudent countries.[3]

However, as stressed in the program evaluation literature, the effect of a policy on the distribution of outcomes is not the same as the distribution of the effects (Heckman et al., 1997, Abbring and Heckman, 2007, Bedoya et al., 2017). While subtle, the difference between the impact on the distribution and the distribution of the impacts is important. Estimating the distribution of treatment effects is important to uncover further layers of heterogeneity by computing, for instance, the variance of treatment effects or what proportion of countries were 'helped' or 'hurt' by the rule. Also, it can shed light on whether countries changed their rank in the outcome distribution because of the introduction of the rule. However, estimating individual treatment effects is challenging because it requires estimating a

[1] Source: IMF Fiscal Rules Database.

[2] For example, countries with fiscal rules could have a stronger preference for fiscal prudence and discipline, even in the absence of the rule. Moreover, the adoption of such a rule could reflect the overall fiscal situation or can be part of a general fiscal overhaul. In both cases, countries with already smaller levels of public deficits would also be more likely to adopt a fiscal rule, thus rendering the causal interpretation of the impact of the fiscal rule on deficits invalid.

[3] Under a rank-invariance assumption, the authors also derive country-specific effects and find that all countries have seen their fiscal position improve on average because of the deficit rule.

counter-factual for every treated unit. In this paper we overcome this challenge by adopting a Synthetic Control Method (SCM) to estimate individual treatment effects of budget balance rules in a large sample of advanced and emerging economies, and low-income countries. We also address the lack of inference in most SCM studies exploiting the insight by Arkhangelsky et al. (2019), who propose inference methods for a version of the SCM with unit weights.

The use of SCM allows us to obtain country-specific counter-factuals and hence to go one step further and analyze the key determinants of rules' effectiveness. While previous studies have mostly relied on anecdotal evidence (for examples see Primo, 2007, and International Monetary Fund, 2018), we exploit the cross-country variation in the timing and design of budget balance rules and in the magnitude (and sign) of the treatment effects to quantify the impact of particular features of the rule in a systematic way.

Our findings point to a sizable heterogeneity in the size of the treatment effect. Most importantly, it is found that the effects of the fiscal rule at the top and the bottom of the balance distribution are of opposite signs: while the rule has reduced the deficit of high-deficit countries, it has increased it among low-deficit countries. Overall, the introduction of the rule has pulled the individual observations to the center of the distribution. From a methodological point of view, this shows that while usual approaches which are concerned with the average treatment effect correctly estimate the effect on the 'average country', it does not capture the full extent of the impact of the fiscal rule on all countries' deficits. Concretely, the average effect overstates the effect on high-balance countries and, more importantly, it understates the true effect for high-deficit countries.

Finally, when we link the individual treatment effects with rules' characteristics, our results suggest that the design of the fiscal rule is indeed critical for its effectiveness (see International Monetary Fund, 2018, and Caselli and Reynaud, 2020). Specifically, we find that maintaining a small overall number of fiscal rules is key. We also see that a monitoring process outside the government, especially at the supra-national level, improves the effect of the budget balance rule significantly while the existence of an escape clause has a diminishing effect.

The roadmap of this chapter is as follows: Sections 1.2 and 1.3 provide an introduction to the SCM, the specification of our baseline model, and the data used. Sections 1.4 and 1.5 present the baseline individual treatment effect results and a series of exercises to assess the goodness of the fit and its robustness to alternative specifications. In Section 1.6 we analyze the role of specific features of rules' design and their effectiveness. Section 1.7 concludes.

1.2 Synthetic Control Method and Model Specification

This section introduces the notation used throughout the paper and briefly reviews the SCM estimation. The SCM is a data-driven procedure to build counter-factual outcomes for observations that are subject to a treatment, e.g. the introduction of a fiscal rule. It is particularly well suited when units of observation are at an aggregate level such as countries for which it can be difficult to find a suitable counter-factual that resembles the treated country.[4]

[4]As argued in Abadie et al. (2010), with aggregate data and in an observational context it is more promising to use a weighted combination of several untreated units instead of a single untreated unit as a counter-factual. Doudchenko and

In a set of countries $i = 1, \ldots, N$ over T periods, country 1 is the only country that receives the treatment in period $s < T$. For a given value of the fiscal rule indicator $FR_i \in \{0, 1\}$ for whether the country is ever treated and values of the government budget balance $Y_{i,t}$, we define potential outcomes $Y_{i,t}\,(FR_i)$ as follows:

$$Y_{i,t}\,(FR_i) = \begin{cases} Y_{i,t}(0), & \text{if} \quad FR_i = 0 \\ Y_{i,t}(1), & \text{if} \quad FR_i = 1 \end{cases} \qquad \forall\ t \geq s$$

The size of the treatment effect on the outcome of interest would be straightforward to infer by comparing actual outcomes and their counter-factuals, i.e. $\tau_{1,t} = Y_{1,t}(1) - Y_{1,t}(0) \quad \forall\ t \geq s$, where $Y_{1,t}(0)$ denotes the counter-factual for country 1 had it not adopted the fiscal rule. However, we cannot observe a country in both states simultaneously: while $Y_{i,t}(1)$ is observable for those countries that have adopted a fiscal rule, $Y_{1,t}(0)$ is not directly observable in the data.

The SCM allows to build a counter-factual for each treated observation, i.e. the outcome of interest in the absence of the treatment. Concretely, it finds the weighted average of all potential comparison units which is 'closest' to the treated unit in terms of pre-treatment outcomes.

To do this, we use the following two-step nested minimization problem. We wish to match in the pre-treatment period both:

1. The weighted average of K predictors in the control group X_0 to the weighted average of those predictors for the treated unit X_1; and

2. The time path of the (weighted) average outcome among control units Y_0 to the outcome for the treated unit Y_1.

To match the treated country to the weighted average of the control units along the two dimensions, we need two sets of weights. A first set of country-level weights $W = (\omega_2, \ldots, \omega_N)$ is used to build weighted averages of the K predictors and the outcome variable among control units. A second set of predictor-level weights denoted by the diagonal matrix V is used to linearly combine the K predictors. Under the usual identifying assumptions, any difference between the treated unit and this synthetic control is due to the treatment itself, since both units have the same (weighted) average values for predictors and outcome in the pre-treatment period.

Formally, objective 1) is expressed as

$$\hat{W}(V) = \arg\min_{W}\ (X_1 - X_0 W)'\, V\, (X_1 - X_0 W) \tag{1.1}$$

with country-level weights $W = (\omega_2, \ldots, \omega_N)$, $\omega_i \geq 0$ and $\sum_{i=2}^{N} \omega_i = 1$ and taken as given the predictor-level weight matrix V, a diagonal positive semi-definite matrix of dimension $K \times K$ with trace equal to one.

For objective 2), we wish to find a linear combination of the K predictors that yields, through Equation 1.1, the set of country-level weights $\hat{W}(V)$ that in turn produces the smallest distance between the synthetic control and the treated unit. This can be expressed as

$$\hat{V} = \arg\min_{V}\ \left(Y_1 - Y_0 \hat{W}(V)\right)' \left(Y_1 - Y_0 \hat{W}(V)\right), \tag{1.2}$$

Imbens (2016) note that the SCM can be viewed as a generalization of the standard difference-in-differences approach, with the weights on the control units chosen to better match the pre-treatment trend of the unit that is exposed to the treatment.

noting that $\hat{W}(V)$, the solution to Equation 1.1, implicitly depends on the choice of matrix V.

The data-driven approach just described is highly flexible and allows for a very broad definition of X_0. To see why, consider for example that the set of predictors X_0 could be defined as the pre-treatment country-level average of k predictors. Alternatively, X_0 could also be defined as the set of annual observations of the same k predictors, so that the total number of predictors used in Equation 1.1 would instead be $K = s \times k$. Given some choice of X_0, Equation 1.2 would then select the matrix V that combines all predictors K to best match pre-treatment outcomes across the two groups. Note the matrix V does not have an economic interpretation but serves only to achieve the best pre-treatment match possible between treated unit and synthetic control unit.

Without economic theory to guide the choice of the weighting matrix V, there is a risk researchers will use specification searching to cherry-pick the set of K predictors that yields the desired outcome for the subsequent treatment effect estimation. This 'p-hacking' would affect the size of tests for statistical significance of treatment effects as shown by Ferman et al. (2020). Consequently, the authors recommend using all the pre-treatment outcome lags as predictors. This also simplifies the algorithm above to

$$\hat{W}(V) = \arg\min_{W} \, (Y_1 - Y_0 W)' \, (Y_1 - Y_0 W) \, ,$$

where we minimize the distance between outcomes for treated and synthetic control by searching directly for the country-level weights W.

After estimating country-level weights $\hat{W}$ that produce the synthetic control with the closest pre-treatment match, we can construct post-treatment counter-factual outcomes for the treated unit using a weighted average of the control units

$$\hat{Y}_{1,t}(0) = \sum_{i=2}^{N} \hat{\omega}_i Y_{i,t}, \quad \forall \, t \geq s.$$

We estimate the treatment effect at each point in time using

$$\hat{\tau}_{1,t} = Y_{1,t} - \sum_{i=2}^{N} \hat{\omega}_i Y_{i,t}, \quad \forall \, t \geq s.$$

If the number of pre-intervention periods in the data is sufficiently large, matching on pre-intervention outcomes can allow us to control for the heterogeneous responses to multiple unobserved factors. The intuition here is that only units that are alike along both observables and unobservables would follow a similar pre-treatment trajectory.

As a baseline, we choose the predictor-level weight matrix V such that pre-treatment outcomes are matched on the lags of the government balance with no additional covariates (see Doudchenko and Imbens, 2016, and Ferman et al., 2020).[5] In a set of robustness checks, we use instead other covariates in the two-step SCM algorithm outlined above. We allow the pre-treatment period for calibrating the donor weights to be as long as data availability allows but not longer than ten years.[6] We also allow for

[5] See also Kaul et al. (2015) for a discussion of covariates in SCM.

[6] In fact, given that the available data on fiscal balances start in 1980 and the first fiscal rules were introduced in the early 1990s, this is the longest pre-treatment horizon that the data would allow to use.

ten years after the introduction of the fiscal rule in our baseline scenario.[7] Thus, the sample window for every treated country can be at most 21 years in total. The pool of donor countries is restricted to those that have complete budget balance data over the same time horizon as the treated country and did not introduce a fiscal rule themselves within this time horizon. Moreover, as our baseline specification, we only consider donor countries that are in the same or in an adjacent income class group, as classified by the World Bank. For example, high-income countries are matched to other high-income countries or those from the upper middle-income group. Similarly, low-income countries are only paired with other low-income countries and those in the lower middle-income country group. To account for country-specific and time-invariant factors, we de-mean the government balance data by subtracting the pre-treatment average by country. If selection into treatment is only correlated with time-invariant common factors, e.g. a general taste for fiscal prudence, then the de-trended synthetic control estimator is unbiased. Finally, the use of SCM is particularly appealing because of its transparency in constructing counter-factual observations. Compared to standard regression methods, the chosen weights make explicit what each control unit contributes to the counter-factual.

To conduct inference, we make use of the recent result of Arkhangelsky et al. (2019) who provide a new perspective on the SCM treatment effect as a weighted least squares regression estimator with time fixed-effects of the following form:

$$\left(\hat{\mu}, \hat{\beta}_t, \hat{\tau}_{1,t}\right) = \arg \min_{\mu, \beta_t, \tau_{1,t}} \sum_{i=1}^{N} \sum_{j=-10}^{10} \left(Y_{i,t} - \mu - \beta_t - FR_i \tau_{1,t}\right)^2 \hat{\omega}_i, \tag{1.3}$$

where β_t are year fixed-effects and FR_i is an indicator variable that is equal to one in the case of the treated country. Importantly, the authors show that under standard assumptions, the estimation uncertainty of the SCM weights $\hat{\omega}_i$ can be ignored under plausible assumptions. This simplifies inference considerably as the weights obtained in the first step can be used in estimating Equation 1.3 without further adjustment.

1.3 Data

We investigate the country-specific impact of budget balance rules (BBRs) on the nominal government balance in a sample of 193 countries between 1980 and 2018. The sample combines data from the IMF World Economic Outlook database with historic fiscal balance series from Mauro et al. (2015) and the IMF Government Finance Statistics.[8] Among the 151 countries in our sample, 53 countries have adopted a budget balance rule with varying numerical targets.[9] Table A.1 lists the rule adopters, year of adoption, the numerical target of the rule, and its coverage (general or central government). Most

[7] This leaves enough time to let the effects of the fiscal rule materialize and become visible while being also short enough to minimize the risk that other major country-specific changes could potentially alter the true underlying country weights. It is not uncommon in the literature to use substantially longer post-treatment horizons. For example, Campos et al. (2019) analyze the country-specific growth effects of European integration and use the obtained weights for extrapolating counter-factual GDP series for as much as 30 years in some cases.

[8] We restrict the set of countries to those that have at least one million inhabitants in 2018.

[9] We exclude six countries (Cameroon, Chad, Estonia, Indonesia, Mali, and Singapore) due to limitations in data availability in the early 1990s. We also exclude India from our analysis as it had a budget balance rule in place only for one year.

countries have adopted fiscal rules restricting the deficit to levels not larger than three per cent or zero per cent of GDP.

Table 1.1 presents the summary statistics for the fiscal balance for countries that have adopted a fiscal rule ('Rulers') and those without fiscal rule ('Non-Rulers') between 1980 and 2018. Not surprisingly, deficits are larger on average for country-year observations without fiscal rules compared to the situation with fiscal rule in place. The table also shows that rulers and non-rulers are different with respect to other macroeconomic and institutional characteristics. This is particularly true for income per capita, inflation, and interest rates as well as the old-age dependency ratio. A simple comparison of the average deficits of rulers and non-rulers is therefore likely to give a biased estimate of the causal effect of fiscal rules on deficits.

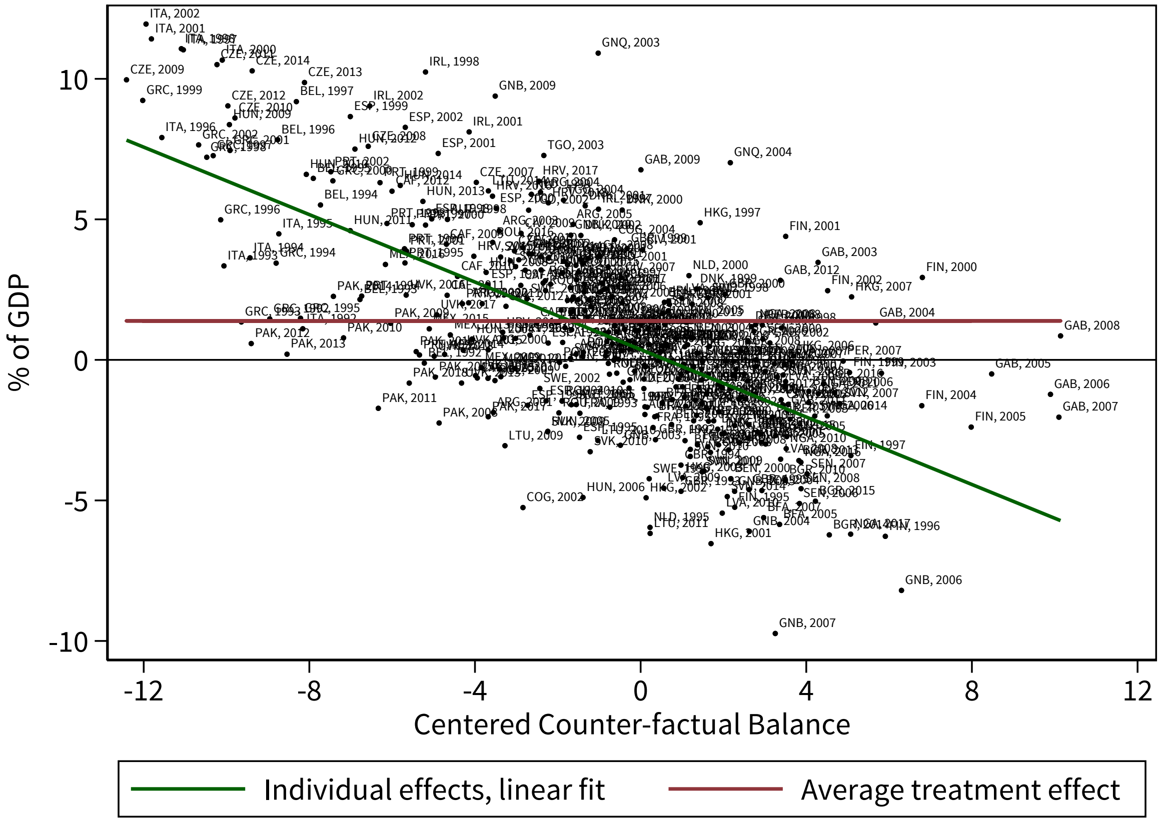
	Rulers				Non-Rulers			
	Obs.	Mean	Min.	Max.	Obs.	Mean	Min.	Max.
Government Balance, % of GDP	808	−2.17	−15.14	18.55	3817	−2.60	−19.80	19.39
Population Size, mio. inhabitants	815	24.72	0.67	200.96	4516	34.67	0.24	1395.38
GDP per capita, in 1,000 USD	815	24.67	0.57	70.03	4451	13.41	0.28	172.99
Government Debt, in % of GDP	811	61.01	0.05	233.72	4167	60.79	0.00	2092.92
Old Age Dependency Ratio, in %	809	18.73	4.07	35.59	4943	9.80	0.80	46.17
Inflation Rate, in %	815	2.62	−3.98	25.87	4422	260.78	−72.73	929 790.00
Trade openness, in % of GDP	812	91.79	20.72	442.62	4149	72.77	0.02	376.22
Short-Term Interest Rate, in %	691	3.68	−0.69	19.92	2641	34.20	−0.78	41 280.00
Federal State, 0/1-Dummy	815	0.19	0.00	1.00	5035	0.15	0.00	1.00
Participation in IMF Program	815	0.26	0.00	1.00	5035	0.34	0.00	1.00
Commodity Terms of Trade Ind.	802	100.06	48.56	111.81	4455	98.03	31.92	130.96

Source: IMF World Economic Outlook Database (April 2019), IMF International Financial Statistics (IFS), World Bank World Development Indicators Database, Monitoring of Fund Arrangements (MONA Database), and Gruss and Kebhaj (2019).

Table 1.1: *Summary Statistics on Main Economic Variables for Countries with and without Fiscal Rule*

1.4 Individual Treatment Effects

In this section we present the main set of results. Figure 1.1 reports the key finding of our analysis: it plots the individual effect of introducing a budget balance rule on countries' government balance. Countries that would have had large deficits in the absence of the rule exhibit positive and statistically significant treatment effects, thus reducing their budget deficits. On the other hand, countries with budget surpluses also respond to fiscal rules by reducing their efforts to maintain a budget surplus and move closer to the numerical target of the rule. This is in line with previous findings that the fiscal rule has exerted a 'magnet effect' on budget balances by pulling deficits and surpluses closer to the center of the distribution. Overall, the average treatment effect is significant, positive, and equal to around two to three per cent of GDP after ten years, as indicated by the horizontal line in the figure. From a policy perspective, it is also important to note that the average effect is not able to capture the heterogeneity in terms of the individual treatment effects. While it largely underestimates the effect for countries with large deficits, it vastly overstates (and predicts the wrong sign of) the treatment effect for countries with small deficits and surpluses as indicated by the difference between the average and linear fit.



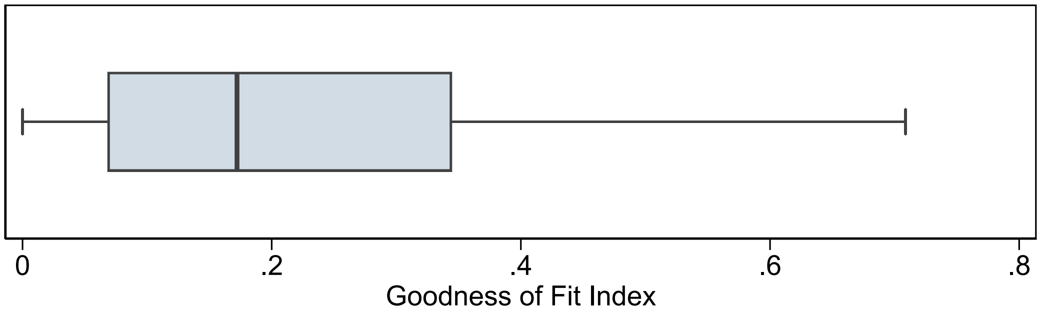
Note: The figure depicts the treatment effect for all countries with fiscal rules for the ten-year post-treatment period as obtained from the baseline SC specification as a function of their counter-factual budget balance, i.e. what the budget would have been if the rule would not have been introduced. The counter-factual balance on the horizontal axis is centered around the numerical value of the fiscal rule in each country.

Figure 1.1: *Individual Treatment Effects of Fiscal Rules by Level of Counter-factual Government Balance*

We compute the average treatment effect across countries together with the corresponding 95 per cent confidence interval obtained from the estimation of Equation 1.3. Figure A.2 shows that the size of the treatment effect in the post-treatment period increases gradually in the first six years after the introduction of the rule before reaching a long-run effect of around two to three per cent of GDP. This suggests that the introduction of the fiscal rule led to an economically and statistically significant reduction in the fiscal deficit for the average country. Figure A.3 plots the time paths of the 10^{th} and the 90^{th} percentiles of the actual and the counter-factual distribution before and after the introduction of the fiscal rule. The actual budget deficit of countries at the lower end of the budget balance distribution is considerably smaller in magnitude than in the case without fiscal rules in place. As mentioned before already, these countries have reduced their deficits substantially.

1.5 Quality of the SCM Estimates and Robustness Exercises

In this section we assess the quality of our baseline model and test alternative specifications. First, we want to assess the quality of the pre-treatment fit as it is the key aspect for the validity of the SC estimates. However, there is a lack of consensus on what constitutes a sufficiently good pre-treatment fit



Note: In order to assess the quality of the achieved pre-treatment fit for each treated country for which a counter-factual is constructed from the SC, Adhikari et al. (2018) propose to use the ratio of the Root Mean Squared Prediction Error (RMSPE) and the RMSPE obtained from the zero-fit model. The ratio of the two is denoted as 'Goodness of Fit Index'. A value of zero for the index indicates a perfect fit, while larger values of the ratio correspond to worse fits. According to the authors, a fit that yields an index larger than one should be discarded because the fit is performing not better than just predicting zeros. The above boxplot is based on all 46 indices, one from each country with fiscal rule in the baseline specification for which we constructed a synthetic counter-factual.

Figure 1.2: *Assessment of the Quality of the Pre-Treatment Fit in the Baseline SC Specification*

of the synthetic counter-factual when compared to the actual outcome. We follow Adhikari et al. (2018) who propose to use the ratio of the Root Mean Squared Prediction Error (RMSPE) and a benchmark RMSE for each treated country. The RMSPE is defined as:

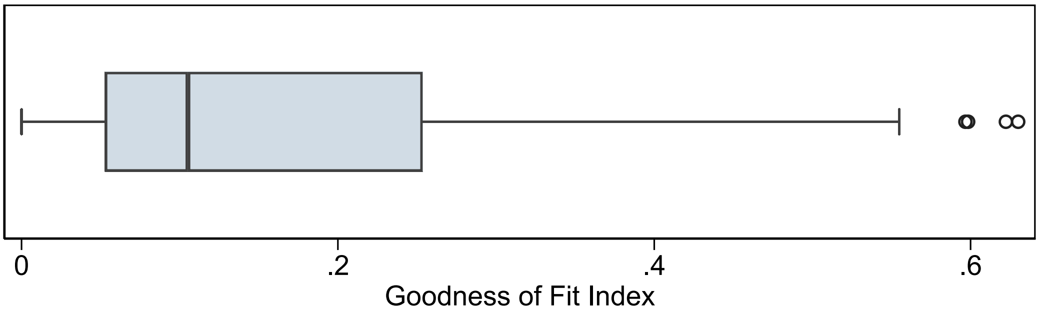
$$RMSPE = \sqrt{\frac{1}{10}\sum_{t=-10}^{-1}\left(Y_{i,t} - \sum_{i=2}^{N}\hat{\omega}_i Y_{i,t}\right)^2}$$

and the benchmark RMSPE as:

$$Benchmark\ RMSPE = \sqrt{\frac{1}{10}\sum_{t=-10}^{-1}\left(Y_{i,t}\right)^2}.$$

The index takes values greater or equal than zero, where smaller values indicate a better fit. Adhikari et al. (2018) suggest that if the index exceeds the value of one, the counter-factual fails to describe the actual path of the outcome variable of the treated country reasonably well and should be discarded. Figure 1.2 shows the distribution of the indices which are well below one, suggesting a sufficiently high quality of the fit.

Second, as already noted, another key feature of the SCM is to make the weights assigned to the treated unit explicit. Figure A.1 displays the weights of donor countries for every treated country in our sample. Treated countries are ranked according to their GDP per capita in 2018 in descending order on the horizontal axis, and the corresponding donor countries for each country on the vertical axis, also ranked according to GDP per capita. Each row contains the corresponding weights for every treated country, with larger weights displayed in darker red. The figure confirms that treated units tend to be matched with countries at a similar level of income. In our sample, positive weights are assigned to more than 40 countries from the pool of donor countries on average, but it can be as high as 60 in some cases. This comes with the advantage that potential country-specific shocks in the group of donor



Note: In this robustness test, the set of donor countries is not restricted to countries of the same or adjacent income groups. This could improve the quality of the fit if it was a binding constraint in the baseline SC. See notes below Figure 1.2 for an explanation of the 'Goodness of Fit Index' by Adhikari et al. (2018).

Figure 1.3: *Assessment of the Quality of the Pre-Treatment Fit in the Robustness SC Specification without Income Group Conditioning*

countries or cross-country spillover effects to geographically neighboring countries are negligible if single country-specific weights are small.

Third, we implement a series of robustness checks to assess the stability of our baseline results to a different set of donor countries and the use of covariates instead of lagged outcomes in the SCM algorithm. We also assess whether the results are influenced by restricting the pool of potential donors to countries only in the same or adjacent income level. Figure 1.3 shows the distribution of the Adhikari et al. (2018) index without donor pool restriction. This means that all countries, for which complete balance data is available, can be potentially used for the construction of the synthetic counter-factual independently of their level of economic development. Figure 1.3 shows that the shape of the distribution including those countries for which the pre-treatment fit is worst within the sample is only marginally changed as compared to the baseline scenario when conditioning on income levels (Figure 1.2). This suggests that conditioning on income classes is for most cases a non-binding constraint in terms of the quality of the pre-treatment fit.

Fourth, we assess the robustness of our main results by using relevant predictors of the fiscal balance in the nested optimization algorithm of Abadie et al. (2010). Specifically, we condition on the debt-to-GDP ratio, GDP per capita, inflation rates (all from the IMF World Economic Outlook database), a commodity terms of trade index (from Gruss and Kebhaj, 2019) and whether there has been an IMF program in place. Weights are estimated based on the pre-treatment averages of all regressors. Figures A.4 and A.5 confirm the magnet effect at the individual country-level and the positive average treatment effect of the rule. However, it is challenging to construct similar synthetic controls in terms of the pre-treatment trend with incomplete data on the economic characteristics used as regressors for some of the treated and untreated countries.

Finally, we test the robustness of the baseline specification to alternative pre- and post-treatment windows. We restrict the length of the pre- and post-treatment horizon to five years only. The result on the distribution of individual treatment effects is illustrated in Figure A.6 showing a very similar pattern to Figure 1.1.

1.6 Effectiveness of Fiscal Rules

The effective design of fiscal rules has been a longstanding question in the literature. A range of studies exploiting either cross-country or cross-state variation with respect to the 'strength' of fiscal rules document that stronger rules are generally better in restraining deficits (see e.g. Poterba, 1994, for the US and Badinger and Reuter, 2017, for the EU; Caselli and Reynaud, 2020). In many cases, these studies rely on a composite index of strength that aggregates several features, such as the coverage of the rule, the degree of independence of monitoring and enforcement bodies, existence of correction mechanisms and sanctions, etc. Based on a small set of case studies for countries that have adopted fiscal rules in recent years, International Monetary Fund (2018) compares the actual outcome of public debt and expenditures with their counter-factual paths in absence of the fiscal rule over a ten-year time horizon as derived from a SCM estimation. It then relates improvements with respect to those indicators (or the lack thereof) to some specific features of the rule, based on anecdotal evidence.

The granular information on country-specific treatment effects together with information on the specific design of fiscal rules in these countries allows us to test in a systematic way what makes fiscal rules more or less effective at constraining short-term fiscal policy. Special attention is devoted to some of the hypotheses that have been put forward in International Monetary Fund (2018): first, an increasing number and complexity of fiscal rules in place increases the potential of conflict between the rules and the probability of non-compliance. Second, the existence of monitoring and enforcement mechanisms can incentivize compliance with the rules, e.g. through raising public awareness.

The information on fiscal rule design is summarized in Table 1.2.[10] On average, countries in our sample have three rules in place, including rules at the national and the supra-national level. The most frequent combination is to have budget balance rules, many of them at the supra-national level, together with restrictions on public debt. We note that if a country has adopted a supra-national budget balance rule, it always includes a monitoring and an enforcement mechanism. For BBRs at the national level, this is only true in half of the cases. Differences exist with respect to the existence of escape clauses. It appears that this feature is included only in a minority of rules adopted.

Table 1.3 shows the estimation results for the number and types of rules in place. Specifically, we estimate the following regression equation:

$$\hat{\tau}_{i,t} = \alpha + \underbrace{\sum_{k=2}^{7} \beta_k \times \mathbb{1}_{k,i,t}}_{\text{Number of rules}} + \underbrace{\sum_{l \in \{\text{exp, rev, debt}\}} \gamma_l \times \mathbb{1}_{l,i,t}}_{\text{other rules in place}} + \underbrace{\sum_{m \in \{\text{sup, nat + sup}\}} \delta_m \times \mathbb{1}_{m,i,t}}_{\text{level of deficit rule}} + \epsilon_{i,t}.$$

The indicator variables $\mathbb{1}_{i,t}$ in all three parts of the regression capture different aspects of the design of fiscal rules. The first set of dummy variables refers to the total number of rules in place. In this case, the indicator variable is equal to one if country i has a total amount of k rules (of any sort) in place at time t. In general, there can be a total amount of eight different rules (rules on the deficit, debt, revenue, and expenditure – both at the national and the supra-national level). The largest number of

[10] We consider all the post-treatment years for the 46 countries with fiscal rules in place. For some countries, e.g. those that adopted the fiscal rule only after 2010 or that have removed the fiscal rule after a few years again, we have less than ten years of post-treatment observations.

Rule Feature	Obs.	Mean	St. Dev.
Number of national rules	420	0.89	1.03
Number of supra-national rules	420	2.09	1.14
Number of rules, national and supra-national	420	2.97	1.30
Dummy, expenditure rule	420	0.32	0.47
Dummy, revenue rule in place	420	0.21	0.41
Dummy, debt rule in place	420	0.87	0.33
Dummy, BBR at the national level only	420	0.18	0.38
Dummy, BBR at the supra-national level only	420	0.64	0.48
Dummy, BBR at both national and supra-national level	420	0.18	0.39
If BBR at supra-national level exists: Dummy, monitoring outside gov't	346	1.00	0.00
If BBR at national level exists: Dummy, monitoring outside gov't	166	0.42	0.50
Either BBR at national or supra-national: Dummy, monitoring outside gov't	420	0.90	0.30
If BBR at supra-national level exists: Dummy, enforcement mech.	345	1.00	0.00
If BBR at national level exists: Dummy, enforcement mech.	149	0.40	0.49
Either BBR at national or supra-national: Dummy, enforcement mech.	420	0.92	0.27
If BBR at supra-national level exists: Dummy, escape clause	345	0.45	0.50
If BBR at national level exists: Dummy, escape clause	148	0.30	0.46
Either BBR at national or supra-national: Dummy, escape clause	420	0.47	0.50

Source: IMF Fiscal Rules Database.

Table 1.2: *Summary Statistics on Fiscal Rule Designs*

rules that we observe in our sample is a total of seven. We are interested in the coefficient β_k related to each of the six potential outcomes (having only one rule in place is taken as the benchmark outcome). Column 1 shows that a smaller set of rules positively correlates with a positive effect of the rule on budget balances and the effect is statistically significant. For example, fiscal deficits in countries with two rules have improved by almost four per cent of GDP upon introduction of the rule. Column 2 adds to the discussion of how the existence of other, potentially conflicting rules, can have an impact on the effect of budget balance rules. We include an indicator variable in the regression that is equal to one if country i has, for instance, a debt rule in addition to a fiscal deficit rule in period t. The results do not suggest that the existence of other rules has an impact on the size of the treatment effect of budget balance rules, except for the case of debt rules which appears to have a positive effect. Similarly, also the fact that some countries have supra-national BBRs in addition to rules at the national level does seem to have a positive impact on the treatment effect size (column 3).

Next, we test the impact of rule design and institutional features, namely the existence of a monitoring mechanism, an enforcement procedure, and the possibility to trigger an escape clause. In the same fashion as before, we estimate the following regression equation:

$$\hat{\tau}_{i,t} = \alpha + \underbrace{\sum_{k \in \{\text{nat, sup, nat + sup}\}} \beta_k \times \mathbb{1}_{k,i,t}}_{\text{monitoring mechanism}} + \underbrace{\sum_{l \in \{\text{nat, sup, nat + sup}\}} \gamma_l \times \mathbb{1}_{l,i,t}}_{\text{enforcement mechanism}} + \underbrace{\delta_{\text{esc}} \times \mathbb{1}_{\text{esc},i,t}}_{\text{escape clause}} + \epsilon_{i,t}.$$

	Dep. variable: Size of the Treatment Effect		
	(1)	(2)	(3)
Dummy: 2 num. rules	4.17*** (0.81)		
Dummy: 3 num. rules	1.54** (0.65)		
Dummy: 4 num. rules	2.65** (0.75)		
Dummy: 5 num. rules	2.56** (1.19)		
Dummy: 6 num. rules	1.21 (1.12)		
Dummy: 7 num. rules	1.13 (1.83)		
Dummy: Exp. Rule		−0.16 (0.62)	
Dummy: Rev. Rule		−0.95 (0.97)	
Dummy: Debt Rule		2.33*** (0.81)	
Dummy: BBR at the supra-nat. level			2.32*** (0.80)
Dummy: BBR at the nat. and supra-nat. level			0.93 (0.61)
Constant	−1.10*** (0.19)	−0.30 (0.61)	−0.17 (0.41)
Observations	420	420	420
R^2	0.10	0.03	0.04

Note: Results are based on OLS estimation. Standard errors are clustered at the country-level and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.3: *The Impact of Rule Complexity on the Effect of Fiscal Rules*

Table 1.4, column 1 shows that the improvements of the fiscal balance are particularly strong for countries that have monitoring mechanisms at the supra-national level, as compared to countries that have either national monitoring mechanisms or none. The size of the treatment effect is large in magnitude and statistically highly significant. It would suggest that those countries with monitoring at the supra-national level would reduce their deficit by more than two per cent of GDP vis-à-vis countries with no monitoring at all. Similarly, countries with enforcement mechanisms at the supra-national level also improve their balances after introducing fiscal rules (column 2). The effect is smaller in magnitude, though. Lastly, the existence of an escape clause as part of the fiscal rule has a detrimental effect on the size of the treatment effect. The results indicate that allowing for the possibility to trigger an escape clause reduces the effect of the fiscal rule from around two per cent of GDP without escape clause by two percentage points to a zero-effect.

	Dep. variable: Size of the Treatment Effect		
	(1)	(2)	(3)
Dummy: Monitoring outside gov't at nat. level	1.03		
	(0.79)		
Dummy: Monitoring outside gov't at supra-nat. level	2.48***		
	(0.69)		
Dummy: Monitoring outside gov't at nat. and supra-nat. level	2.46***		
	(0.58)		
Dummy: Enforcement procedure at national level		−0.13	
		(0.75)	
Dummy: Enforcement procedure at supra-nat. level		1.93***	
		(0.69)	
Dummy: Enforcement procedure at nat. and supra-nat. level		1.95***	
		(0.47)	
Dummy: Existence of escape clause			−2.29**
			(0.86)
Constant	−0.64**	−0.09	2.56***
	(0.30)	(0.35)	(0.67)
Observations	420	420	420
R^2	0.03	0.03	0.07

Note: Results are based on OLS estimation. Standard errors are clustered at the country-level and are displayed in parentheses.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 1.4: *The Impact of Specific Rule Features on the Effect of Fiscal Rules*

1.7 Conclusions

This chapter investigates the heterogenous effects of budget balance rules on fiscal performances in a large sample of countries. In order to derive country-specific treatment effects of the fiscal rule under weak and reasonable assumptions, we apply the Synthetic Control Method (SCM) in combination with difference-in-differences estimation, drawing from recent results by Arkhangelsky et al. (2019). Our results indicate that countries with a fiscal rule improve their fiscal balance on average by two to three per cent of GDP in the long-run after its introduction, thus documenting an overall positive and statistically significant average effect of the rule. However, our results also illustrate the importance of going beyond the average treatment effect, as it masks significant heterogeneity in the country-specific impact of the rule. Consistent with previous findings, we show that the improvement of the fiscal balance is largest for countries with large deficits in the counter-factual case where no rule is adopted. Not unsurprisingly, also countries with budget surpluses are affected. Despite not being binding for them, these countries see their fiscal balance converge to the numerical target of the fiscal rule after its introduction. Taken together, these two results confirm the hypothesis that budget balance rules exert a ‘magnet effect’ on fiscal balances by pulling deficits and surpluses closer to the center of the distribution. The results are robust to various modifications of the SCM specification, including changes to the length of the pre- and post-treatment period as well as the selection of covariates based on which the SCM weights are fitted.

Our results also add to the discussion of what concrete features determine budget balance rules’ effectiveness in constraining fiscal deficits. Based on the country-specific treatment effects from the SCM analysis and significant cross-country variation with respect to the specific design of fiscal rules, our results indicate that its features and its interaction with the overall rule environment of the country under consideration are both critical determinants of its success. In particular, we demonstrate that a smaller overall number of rules and the existence of a credible and independent monitoring process for rule compliance are the most important contributors to the success of the budget balance rule. On the other hand, the existence of an escape clause appears to have a detrimental effect on the effectiveness of the rule. From a policy perspective, the results highlight the fact that fiscal rules, despite their overall potential to improve fiscal performances, are a complex matter and can trigger unexpected consequences, as in the case of diminishing fiscal surpluses. The treatment effects are also sensible to the specific rule design and have important interactions with the overall institutional setting.

Chapter 2

Tax Cuts Starve the Beast: Evidence from Germany

This chapter is based on joint work with **Clemens Fuest** and **Florian Neumeier**, CESifo Working Paper Nr. 8009.

2.1 Introduction

Over the past 25 years, many Western democracies have witnessed a rapid increase in public spending, and a corresponding increase in the tax burden. For instance, in the member states of the European Monetary Union (EMU), public spending in relation to GDP increased on average from 41.6 per cent to 46.7 per cent between 1990 and 2015. This implies that government spending has grown notably faster than private spending. Over the same period, public revenues have grown from 39.8 per cent to 44.6 per cent in relation to GDP.[1]

Whether this development should be considered good or bad in terms of economic welfare is disputed. Those who support higher government spending claim that public investment is highly productive, and that demand for publicly provided services and social insurance increase as incomes rise and the population ages. Critics object that the expansion of public spending and taxation reflects inefficient government growth. For both sides of this debate, it is important to understand the factors that enhance or inhibit the growth of public spending.

There is a widespread view among critics that the best way of containing public spending is to cut taxes. At an early stage, Milton Friedman has argued "[...] that the only effective way to restrain government spending is by limiting government's explicit tax revenue – just as a limited income is the only effective restraint on any individual's or family's spending." A similar argument was made by Alan Greenspan around the same time at a hearing of the Senate Finance Committee and later used by Ronald Reagan during his presidential election campaign in 1980.

The idea that lower tax revenues will restrain government spending later became known as the *starving the beast* hypothesis. Its proponents have argued that tax cuts compel legislators to enact spending cuts in order to ensure long-term fiscal sustainability. Critics, however, cast doubt on the notion that legislators will bow to the threat of a government's fiscal position worsening in the future. They argue that tax cuts will inevitably lead to accelerating public debt levels or to even larger tax increases later down the line, as opposed to spending cuts. For instance, Romer and Romer (2009)

[1] The figures represent the unweighted average public spending to GDP and public revenues to GDP ratios across the EMU11. Source: IMF Government Finance Statistics.

investigate the impact of tax cuts on US federal spending and find no evidence supporting the *starving the beast* hypothesis. They conclude that "[...] policymakers should be aware that the historical experience suggests that tax cuts tend to lead to tax increases rather than to spending cuts" (Romer and Romer, 2009: p. 198). However, the empirical approach adopted by Romer and Romer (2009) faces challenges. They try to identify effects of tax cuts on spending for the US federal government only, and they concede that "[...] because our estimates are not highly precise, the hypothesis that tax cuts exert some restraining influence on spending cannot be rejected" (Romer and Romer, 2009: p. 197).

In this chapter, we test the *starving the beast* hypothesis in a different institutional context, which enables the impact of tax cuts to be identified in a more convincing way. We use panel data from the German states (*Bundesländer*) spanning the years from 1992 to 2011 in order to study the fiscal consequences of changes in tax revenues driven by changes in tax legislation. The German states provide an ideal institutional setting for an empirical test of the *starving the beast* hypothesis. While the states have full autonomy regarding debt financing and the level and structure of public spending, they have only very limited influence on the level of their revenues. For most taxes, rates and bases are set collectively by the federal government and representatives of the state governments.[2] Each state then receives a certain share of total tax revenues, the size of which is determined by a formula-based multi-step fiscal equalisation scheme (see Section 2.3). Thus, changes in tax revenues at the state-level that are driven by changes in tax legislation, can be considered exogenous insofar as they are beyond the control of an individual state government. This is a prerequisite for the identification of the causal effect of tax cuts on public spending. It is worth noting, though, that this also implies German state governments only being able to choose between adjusting spending and debt financing in response to a tax cut, whereas they cannot raise taxes retrospectively. Methodologically, we apply the so-called narrative approach pioneered by Romer and Romer (2009) to ensure that the changes in tax legislation that we utilise in our empirical analysis, are unrelated to the current or expected future economic or fiscal situation. To this end, we collect information from official government sources to identify the motivation behind each tax bill.

Our main findings are as follows. First, unlike Romer and Romer (2009), we identify strong support for the *starving the beast* hypothesis. Our results suggest that an exogenous decrease in tax revenues triggers a reduction in public spending of roughly the same amount. The adjustment takes place with a considerable delay of two to three years, however. The decline in public spending following a tax cut, occurs more quickly than the increase in spending triggered by a tax revenue hike. Neither the size nor pace of the adjustment seem to be connected to the political and fiscal situation of the state. No matter whether single-party versus coalition governments, right-wing versus left-wing governments, states with high versus low levels of debt: they all react similarly. Regarding different types of public spending, we find that a revenue decline of one Euro reduces public spending on administration and, with a larger delay, social security, by 30 to 45 cents in each case. Spending on infrastructure declines by ten cents. We find no significant effects on spending on education, legal protection and public safety, or culture.

The remainder of this chapter is structured as follows. The next section briefly reviews the literature to which our paper refers, and highlights the contribution we make. Section 2.3 provides background

[2] Since 2006, states are allowed to set the rate of the real estate transfer tax. In 2005, the last year before the reform, revenue from this tax was only 2.7 per cent of overall state tax revenues.

information on the system of fiscal federalism in Germany, and the degree of tax and spending autonomy of the German states. The construction of our exogenous tax shock series is explained in Section 2.4. Section 2.5 discusses necessary assumptions for causal inference. In Section 2.6, we introduce our empirical model. Section 2.7 presents our results, followed by various extensions and robustness checks in Section 2.8. Section 2.9 reflects on the external validity of our findings. Section 2.10 concludes.

2.2 Related Literature and Contribution

By studying the relationship between public spending and revenues, our paper relates to a long-standing strand in the empirical public economics literature (see Payne, 2003, for a review of the literature). Early contributions to this literature focus on public spending and revenues at the federal level in the US and apply VAR models as well as Granger causality tests to analyze whether tax changes are followed by public expenditure adjustments and vice versa. With regard to the *starving the beast* hypothesis, the results are mixed. Anderson et al. (1986) as well as von Furstenberg et al. (1986), find that changes in public spending trigger subsequent changes in tax revenues, but not the other way around; thus, finding no support for the *starving the beast* hypothesis. Manage and Marlow (1986) as well as Blackley (1986), in contrast, provide evidence supporting the *starving the beast* hypothesis. Subsequent contributions to the literature that apply more advanced methods for the analysis of time-series data, especially vector error correction models and cointegration techniques, as well as studies focusing on countries other than the US, provide inconclusive results, too.[3]

Arguably, the lack of consensus in the empirical literature may be due to the difficulty in identifying changes in tax revenues unrelated to the underlying economic or fiscal situation. Since changes in the economic or fiscal situation potentially affect both tax revenues and public expenditure at the same time, identifying the causal effect of tax changes on public spending poses a challenge. Business cycle fluctuations, for instance, tend to give rise to adjustments in tax revenues and public spending in opposite directions; hence implying that estimates of the effect of tax changes on public spending would be biased downwards. Furthermore, tax changes may be enacted in anticipation of changes in the future fiscal situation (to finance spending hikes resulting from the enactment of government programmes or to reduce the debt burden). This indicates that decisions about taxation and expenditure are made concurrently, although they may become effective at different points in time.

The paper by Romer and Romer (2009) addresses those concerns by using a so-called narrative approach to identify tax changes unrelated to the current or expected future economic or fiscal situation. The idea is to use official sources in order to identify the legislator's motivation behind the introduction of a tax bill. Based on this information, we can separate tax bills that have been enacted for reasons unrelated to the economic or fiscal situation, from those that are not, thus allowing us to compile a list of exogenous tax shocks. Romer and Romer (2009) apply the narrative approach to study how exogenous tax changes affect the US federal government's fiscal policy stance. Their findings suggest that tax cuts lead to future tax hikes, but not to any adjustment in the level of public spending. Accordingly, the *starving the beast* hypothesis is not supported.

[3] Our analysis is also related to the literature on the impact of grants on subnational government spending and the flypaper effect (see Courant et al., 1979, and Hines and Thaler, 1995). We return to this issue in the discussion of our results.

In our empirical analysis, we adopt the narrative approach pioneered by Romer and Romer (2009) to study the effect of exogenous tax shocks on fiscal policy at the state-level in Germany. It is worth noting, however, that our study differs from Romer and Romer (2009) in at least three important aspects. The most important difference is the institutional framework. Unlike the US federal government, German state governments have almost no tax autonomy and would be unable to retrospectively react to tax cuts with the use of tax hikes, even if they wanted to. They can only respond by either decreasing public spending or increasing public debt. In this respect, our results should not be interpreted as a direct contradiction of Romer and Romer (2009). Rather, our results suggest that the *starving the beast* effect does work in the context of the German institutional framework. If we had found that states respond to tax revenue shocks by adjusting public debt, the *starving the beast* hypothesis would have to be rejected. Here, you could object that deficit financing has its limits because highly indebted states might damage investor trust. Yet, German states enjoy far-reaching bailout guarantees, implying that there are practically no risk premia on state level debt. In addition, our results also apply to states with low debt levels. The second major difference to Romer and Romer (2009) is that we do not only study the consequences of exogenous tax changes on aggregate public spending, but also on different types of spending. These include public administration, education, and infrastructure. This is important because one concern about pressure to cut spending is that it will harm public investment above all. A third difference is that we investigate whether the effects of tax changes on public spending vary across single-party governments and coalition governments, right and left-wing governments, as well as with the level of public debt.

2.3 Fiscal Federalism in Germany

The German federal system consists of three levels of government – the federal government, the state governments, and the municipal governments. Each level is endowed with its own legislative competencies and responsibilities, as specified by the German Constitution. Since the German reunification in 1990, there have been 16 German states (*Bundesländer*). Three of these 16 are so-called city-states (Berlin, Bremen, and Hamburg), which combine state and municipal-level competencies. The competencies assigned to the German states are extensive and mainly defined in Articles 71 to 74 of the constitution. They include policy areas such as public safety, education, infrastructure, social security, administration, and health.

Although equipped with far-reaching legislative competencies, fiscal autonomy at the state-level is restricted. While enjoying full discretion about the level and priorities of public spending (at least in those policy areas for which they are responsible), the German states have scarcely any tax setting authority. With few exceptions, taxes are levied and tax revenues are collected by the federal government. In order to ensure that each state has sufficient means to perform its functions and to harmonise living conditions across all 16 German states, tax revenues are allocated across states applying a multi-step mechanism (the so-called *Länderfinanzausgleich* or inter-state fiscal equalization scheme).

In a first step, the vertical allocation mechanism, tax revenues from the income tax, corporate tax, capital gains tax, and value added tax are allocated across different levels of government according to fixed ratios. For example, the federal government and the state governments each receive 42.5 per cent of the income tax, 50 per cent of the corporate tax, and 44 per cent of the capital gains tax. With regard to the value added tax, the federal government receives 51.4 per cent and the state governments 46.4 per cent of the revenues. The remainder is for the municipal governments. Revenues from some other tax measures are fully allocated either to the federal government (for instance energy and tobacco taxes) or the state governments (such as inheritance tax).

In the second step, the horizontal allocation mechanism, each single state's share in the total amount of tax revenues allocated to the state-level is determined. Income tax revenues are distributed in order to roughly match the amount of taxes paid by each state's inhabitants. The allocation of corporate tax revenues follows a similar principle. Each state's share in total corporate tax revenues depends on the amount of taxes paid by the firms who have their headquarters or affiliated production units in that state. The allocation of VAT revenues largely serves the purpose of harmonising tax revenues across states. Up to 25 per cent of the total VAT revenues are assigned to those states that have received below-average per capita tax revenues from other tax sources. The rest is distributed according to the number of inhabitants in each state.

The inter-state fiscal equalization scheme comprises a third and fourth step, which both aim at further mitigating the differences in per capita tax revenues across states. In the third step, states with higher than average per capita tax revenues pay transfers to states with below-average per capita tax revenues. This redistributive scheme is justified by the fact that each state is believed to require financial resources of a comparable level in order to properly fulfill its functions. As a final step, the federal government pays vertical grants to those states that still have below-average financial resources.

In 2017, the sum of tax revenues allocated to the German states was roughly 300 billion Euro, which is only slightly below the level of tax revenues of the federal government (approx. 310 billion Euro). The tax revenues of German municipalities amounted to some 105 billion Euro. Thus, the state governments possess more than 40 per cent of total tax revenues collected in Germany.

The only channel available to state governments to influence tax legislation is through the Second Federal Chamber, that is, the *Bundesrat*. The *Bundesrat* represents the interests of the state governments vis-à-vis the federal government. Its members are not elected, and are instead delegated by the state governments. In general, the extent to which the *Bundesrat* participates in the legislation process depends on the nature of the proposed legislation. Legislation affecting states' interests requires the approval of the *Bundesrat* (so-called 'Consent Bills'). This includes legislation on all taxes where the states (or local governments) participate in revenues (Art. 105 III, German Constitution). In principle, the *Bundesrat* may also propose changes or amendments to (tax) bills introduced by the federal government. However, it is ultimately the German Federal Parliament that decides whether changes proposed by the *Bundesrat* will be adopted.

2.4 Constructing an Exogenous Tax Shock Series

We adopt the narrative approach pioneered by Romer and Romer (2009) in order to identify the causal effect of tax shocks on fiscal policy outcomes. Over the past years, the narrative approach has become a popular tool to investigate the impact of legislated tax changes on macroeconomic aggregates, especially GDP (see Romer and Romer, 2010, for the US, Cloyne, 2013, for the UK and Hayo and Uhl, 2014, for Germany). As a starting point, we collect information on all discretionary changes in the federal tax legislation in Germany over the period from 1988 to 2011, together with the expected impact of each legislated tax change on tax revenues. Our source of information is the *Finanzbericht* (fiscal report), which is the Federal Ministry of Finance's annual publication. The *Finanzbericht* contains detailed information on every piece of tax legislation, including (i) the date the tax bill was passed, (ii) the motivation and objective behind the tax bill, and (iii) forecasts of the annual revenue impact over the coming years for each level of government; for instance federal, state, and municipal level. Note that the forecasts for the state and municipal level represent aggregates, in other words, the *Finanzbericht* reports the prospected cumulative revenue impact for all states and municipalities combined.

We proceed in several steps in order to construct a measure of exogenous tax shocks. First, we assign each piece of tax legislation to one out of seven different categories, depending on the motivation behind the tax bill, thereby, closely following the schemes applied in Romer and Romer (2010), Cloyne (2013), and Hayo and Uhl (2014). By rules of the parliamentary procedure, the motivation for any tax change needs to be explained in detail in the draft bill.

The first category comprises tax changes primarily intended to stabilise supply or demand-driven fluctuations in aggregate output. We label these tax measures 'counter-cyclical'. There are several examples of tax changes falling into this category, most prominently the fiscal stimuli packages that were implemented during the economic and financial crisis in 2010 (*Konjunkturpaket*; economic stimulus package).[4]

The second category are 'spending-driven' tax changes. This label refers to tax measures that are implemented in anticipation of higher future public expenditure over the short and mid-term. Examples include the increase in taxes on cigarettes in 2003, explicitly adopted with the aim of financing the fight against international terrorism. Related to this are tax measures that were implemented in order to consolidate public finances. These 'deficit-driven' tax changes merely constitute tax hikes. A prime example is the value-added tax increase (the VAT rate rose from 16 per cent to 19 per cent) of 2006 with a prospected rise in public revenues of roughly 24 billion Euro each year.

Fourth, some policy measures were taken in response to 'macroeconomic-shocks', such as German reunification and the introduction of the Euro as a common currency in Europe. In 2000, for example, the German government issued a bill converting and rounding amounts denoted in the German tax law in *Deutsche Mark* (such as allowances or income thresholds important for determining tax rates) into Euro amounts.

The fifth category includes tax changes adopted in an attempt to steer taxpayers' behaviour into the desired direction ('Pigou taxes'). The intention behind this type of tax is generally to force taxpayers to internalise some sort of externality. Most of the tax bills falling into this category are environmental taxes.

Our sixth category is for tax measures that were implemented for 'structural reasons'. This label applies to tax instruments intended to improve the long-term economic conditions, while being unrelated to the contemporary economic situation at the same time. Examples include tax measures that aim to reduce the bureaucratic burden of taxation for firms or to improve conditions for private investments.

Seventh, some tax changes reflect the transposition of EU law into national legislation, or they are passed due to a ruling of the European Court of Justice. For instance, in 2010, a bill was introduced ex-

[4] Cloyne (2013) explicitly distinguishes between demand and supply-driven tax policy reactions within this broader category of 'counter-cyclical' tax policy measures. This distinction is not important for our purpose, as we are only concerned with the identification of exogenous tax measures.

tending the possibility to deduct donations to Germany-based charitable organisations from the taxable income to donations directed to charitable organisations based in other EU countries. Introducing these tax bills into national legislation is compulsory and the German government generally has only little – if any – leeway with regard to the details of the law's content. Those tax bills are typically intended to harmonise tax legislation across EU member states.

In a second step, we classify each tax category as either endogenous or exogenous. In the process, we closely follow the examples of Romer and Romer (2009), Cloyne (2013), and Hayo and Uhl (2014). Discretionary tax changes that can be unambiguously labelled as exogenous, should not be correlated with factors that may concurrently affect government spending and revenues. One example of a tax category that clearly fails to meet this criterion are counter-cyclical tax measures. Tax cuts (tax hikes) implemented during downturns (upswings), tend to be accompanied by spending hikes (cuts) so as to amplify the impact of the fiscal stimulus (contraction). Moreover, automatic stabilisers mechanically trigger expenditure adjustments in the presence of business cycle fluctuations, and therefore coincide with counter-cyclical tax measures. However, they are not caused by tax changes.

In addition, exogenous tax changes must not automatically entail decisions about public expenditure or be triggered by them. Tax hikes adopted for the purpose of financing a specific government programme for instance, clearly violate this condition, and thus, must be considered endogenous. Tax hikes that aim at consolidating public finances, on the other hand, may be inversely related to spending changes. Hence, including these tax changes would most certainly induce a downward bias in our coefficient estimates. All remaining tax changes are labelled as exogenous for two reasons. Firstly, they are neither the result of changes in public expenditure, nor are they related to economic or fiscal variables that affect tax revenues and public expenditure at the same time. Secondly, since they reflect decisions taken at the federal government-level and are, thus, 'externally' imposed, they are unrelated to the political situation in the state where the revenue 'shock' occurs.

Finally, for each year of our sample period, we compute the aggregated revenue impact of past and present exogenous tax changes. For each tax measure, the *Finanzbericht*, a yearly report about public finances in Germany, includes the forecasted annual revenue impact for k consecutive years, with k varying across publications and tax changes. We simply add up the tax revenue changes projected for year t of all tax measures adopted between t and $t-k$. Therefore, our exogenous (indicated with x) tax shock measure is equal to

$$\Delta\tau_t^x = \sum_{j=0}^{k} \Delta\tau_{t|t-j}^x,$$

where $\Delta\tau_{t|t-j}$ is the sum of tax revenue changes projected for year t across all tax measures introduced in year $t-j$.

Note that, in contrast to Hayo and Uhl (2014), who omit tax measures with a prospected revenue impact of less than 0.1 per cent of GDP from their analysis, our analysis includes all pieces of tax legislation that are adopted during our sample period. The reason is that several tax changes are introduced each year and we are interested in their cumulative impact. Even if each single tax measure implemented in a particular year has a rather modest impact on tax revenues, the cumulative effect of all tax changes introduced that year may be large enough to exert a significant influence on the public budget. In total, our dataset covers 129 pieces of legislated tax changes, of which we consider 93 to be exogenous and 36 to be endogenous.

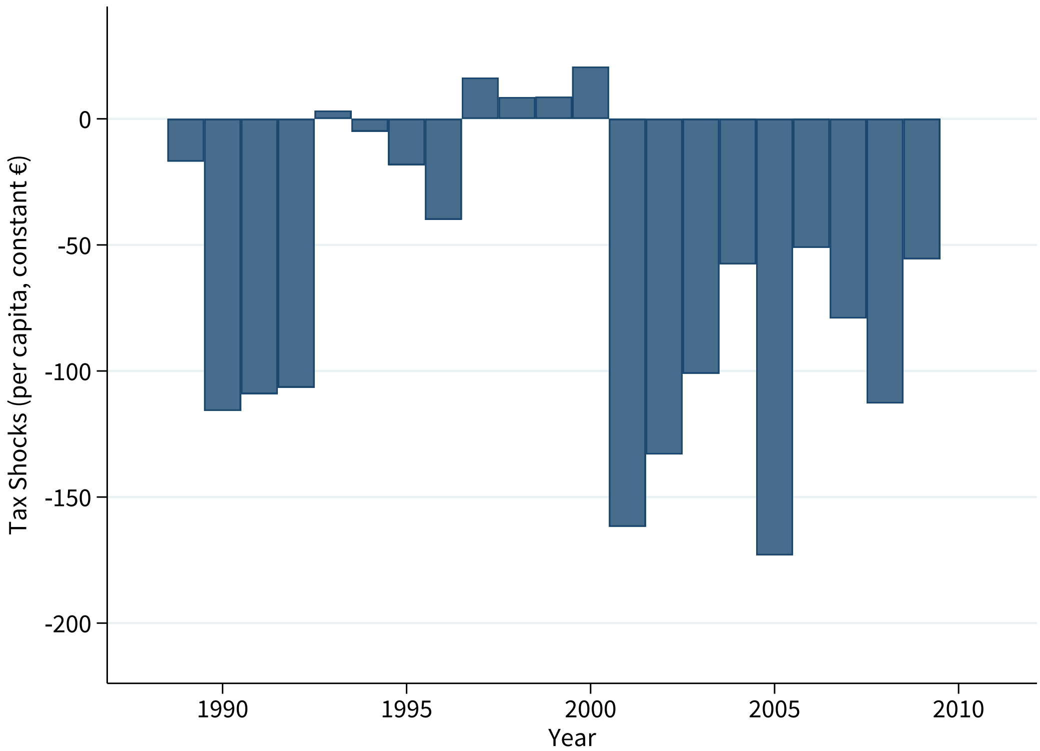


Figure 2.1: *The Series of Exogenous Tax Revenue Shocks at the State-Level*

The resulting exogenous tax shock series is depicted in Figure 2.1 with the projected per capita tax revenue changes at the state-level in constant prices on the vertical axis. We can see that the projected aggregate revenue change is positive in only five out of 21 years. Moreover, in those five years, the projected increase was of rather modest size. A very similar picture emerges when focusing on the single tax bills. We find that around two thirds of the exogenous tax changes are estimated to have a negative revenue impact. What is more, on average the projected revenue impact of negative tax shocks is larger than that of positive tax shocks (in absolute terms). Some years stand out in this context: substantial drops in tax revenues in our exogenous tax shock series in the early and mid-2000s are mostly driven by structural changes in labour, as well as corporate profit taxation predominantly enacted in the year 2000 when the economy experienced strong growth. The corresponding tax bills were intended to spur domestic demand and private investment over the longer-term. Similarly, the negative tax revenue shocks of 2007/08 that we label exogenous, were not a reaction to the upcoming financial and economic crisis but instead reflect an accumulation of smaller tax changes aimed at changing the long-run conditions. For example, the largest tax change in terms of revenue impact at that time emanates from the introduction of tax deduction possibilities for private pension plans. This instrument was implemented in light of the projected demographic change over the coming decades. Clearly, these changes are not driven by contemporary economic conditions but focus on solving longer-term challenges.

2.5 Conditions for Causal Inference

The aim of our analysis is to identify the causal effect of changes in tax revenues on the fiscal policy of German state governments in order to empirically test the *starving the beast* hypothesis. The following conditions must apply in order for our identification strategy to be valid.

2.5.1 Accuracy of Revenue Projections

Tax projection experts at the Federal Ministry of Finance provide the revenue projections reported in the *Finanzbericht*. Unfortunately, there is little information on how and on which basis they are obtained. Given that we use these projections to construct our key explanatory variable, our identification strategy depends on the accuracy of these projections. In order to assess their accuracy, we regress the absolute change in state tax revenues per capita on our exogenous tax shock series. Table 2.1 shows the regression results for different specifications of the regression model. Since the point estimate of the contemporary effect of the tax shock series is close to one across all three specifications (and never significantly different from one, as the bottom row of Table 2.1 indicates), we have no reason to believe that the ministry's projections are systematically biased. We are therefore confident that they are sufficiently precise for our purpose.

	Dep. variable: Total revenues per capita		
	(1)	(2)	(3)
$\Delta\tau_t^x$	0.889^{***}	0.882^{***}	1.146^{***}
	(0.257)	(0.252)	(0.326)
linear trend	✘	✓	✓
quadr. trend	✘	✘	✓
Observations	224	224	224
R^2	0.049	0.051	0.059
H_0: $\beta_{\Delta\tau_t^x} = 1$	$p = 0.67$	$p = 0.64$	$p = 0.65$

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 2.1: *Accuracy of Revenue Projections from the Annual Fiscal Report of the Ministry of Finance*

2.5.2 Exogeneity of Tax Shock Series

To establish a causal link between changes in tax revenues and public expenditures, it is important that our tax shock series is unrelated to the underlying economic and fiscal conditions. Applying the narrative approach ensures that the tax changes we consider are not motivated by current economic conditions. What about the timing of those tax changes, however? Even if a tax change may not be motivated by the current economic situation, policymakers might still take the business cycle into account when making decision on the timing of a tax change. For instance, policymakers may be reluctant to implement tax hikes during downturns or tax cuts during upswings, which would render our identification strategy invalid. However, when comparing the realisations of our tax shock series depicted in Figure 2.1 to contemporary income growth rates, it becomes evident that the tax changes we consider to be exogenous are not related to business cycle fluctuations. Some of the largest negative tax shocks have been adopted in 'normal' economic times characterised by income growth rates that are close to the sample average. For example, in the early 1990s, income growth was about 1.5 per cent p.a., and in the early 2000s it was roughly 1.9 per cent. Other peaks in the tax shock series seem to be evenly distributed across economic downturns and upswings. In 2005, when the German economy reached the height of an upswing phase with an annual income growth rate of 3.8 per cent, a large negative tax shock occurred. The next sizable negative tax change was implemented in 2009, at a time Germany was still suffering from the repercussions of the financial and economic crisis, and income growth was negative.

2.5.3 No Anticipation Effects

Another concern that may be raised is that state governments might anticipate future tax changes and adjust public spending in advance, not least because they could potentially influence the legislation process through the Second Federal Chamber, for example the *Bundesrat* (see section 2.3). State governments' ability to influence tax legislation may raise concerns about reverse causality. The question arises whether states adjust their expenditure in anticipation of future tax changes, which they could potentially influence. To test the validity of this concern, we regress our tax shock series on contemporary and past changes in state public expenditure per capita.[5] The results are presented in Table 2.2. Our findings do not indicate that state governments anticipate future tax changes. The coefficient estimates for present and past changes in public spending are not only statistically insignificant, but also economically negligible. Thus, we are confident that reverse causality is not an issue.

2.6 Empirical Approach

Our aim is to evaluate the influence of legislated tax changes on fiscal policy outcomes. To this end, we utilise panel data from the German states spanning the years from 1992[6] to 2011, the latest year for which state-level fiscal data are currently available.[7] We estimate the following empirical model:

$$\Delta y_{it} = \alpha + \sum_{j=0}^{4} \beta_j \Delta \tau^x_{t-j} + \sum_{j=1}^{4} \gamma_j y_{it-j} + X_{it}\delta + \epsilon_1 t + \epsilon_2 t^2 \tag{2.1}$$

The index t refers to the year, and i to the state. In our baseline specification, the dependent variable is the absolute change in total public spending per capita. As part of an extension, we also focus on

[5]Note that according to Reutter (2007), the time span between the initiation of a bill and its adoption is, on average, 250 days. In fact, all the tax bills covered in our analysis became effective the year after they were introduced to parliament. Therefore, we believe that it is sufficient to control for the first lag of public expenditure in order to test for reverse causality.

[6]Note that state-level forecasts of the revenue impact of tax changes are only reported from 1988 onwards. In the *Finanzbericht* of 1988, revenue forecasts are provided for the years from 1988 to 1992. Thus, if we were to utilise data from before 1992 in our analysis, we would not take into account the estimated state-level revenue impact of tax measures introduced before 1988.

[7]State-level fiscal data are published with a considerable time-lag as the variables need to be made comparable across states by the Federal Statistical Office.

	Dep. variable: Tax Shock Series ($\Delta\tau_t^x$)		
	(1)	(2)	(3)
Δy_t	−0.015	0.014	−0.006
	(0.023)	(0.022)	(0.021)
Δy_{t-1}	0.009	0.001	−0.020
	(0.025)	(0.021)	(0.020)
linear trend	✘	✓	✓
quadr. trend	✘	✘	✓
Observations	224	224	224
R^2	0.002	0.203	0.303

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.2: *Reverse Causality Check: The Impact of Spending Adjustments on the Tax Revenue Shock Series*

public spending in different policy areas (see Section 2.8.2). Our explanatory variable of main interest is our measure for exogenous tax shocks $\Delta\tau^x$. As described in Section 2.4, our tax shock indicator is equal to the sum of the changes in tax revenues projected for year t across all tax changes adopted between t and $t-k$. Since for most volumes of the *Finanzbericht*, k is equal to four (and never smaller), our empirical model includes four lags of the tax shock variable. X_{it} is a vector of state-specific covariates. This vector includes the level of net income per capita, the unemployment rate, the level of public debt per capita, the dependency ratio (for example, the share of people below 25 or above 65 years of age), as well as dummy variables for election years, coalition governments, and left-wing governments (such as governments led by the Social Democratic Party). Note that net income, the unemployment rate, and the level of public debt are lagged by one year to mitigate concerns about reverse causality. In order to account for gradual budget adjustments, we also include the four lags of the dependent variable in our empirical model. To test whether the coefficient estimates of our tax shock series are sensitive to the inclusion of additional explanatory variables, we consecutively add the lags of the dependent variable and the vector of covariates to our empirical specification. Finally, we add a linear and a quadratic time trend to our model to account for homogeneous trends. Data on all fiscal and economic variables are provided by the Federal Statistical Office (*Statistisches Bundesamt*), while the political variables are taken from the State Election Commissioners (*Landeswahlleiter*). The economic and fiscal variables are adjusted for inflation using the national CPI.[8] We estimate Equation 2.1 using ordinary least squares (OLS).

Given that the *Finanzbericht* only reports forecasts of the aggregate tax revenue effect for all German states combined, we need to make an assumption about how the aggregate tax revenue effect influences the budgets of individual states. Since, generally speaking, the allocation of tax revenues across states is primarily determined by the number of citizens residing in a state, we divide our tax shock variable by national population figures and our dependent variables by state population figures. That way, we

[8]The base year is 2000; the data are also taken from the Federal Statistical Office.

implicitly assume that the change in tax revenues in each state resulting from a legislated tax change, is proportional to its number of inhabitants (and that the per capita revenue impact of a legislated tax change does not vary across states).

Three of the 16 German states are so-called city states (*Stadtstaaten*), namely Berlin, Bremen, and Hamburg. These states adopt functions at both the state and local government-level. As a result, they receive tax revenues designated for the state and local level, which is why the tax shock series for the city states combines prospected changes in tax revenues at both state and the municipal level.[9]

Since we use the first difference of public spending as the dependent variable, the coefficients β_j can be interpreted as the tax shock-induced deviations of state-level public spending from its long-term trend. In the full model that includes net income and the unemployment rate as covariates, the dependent variable can be interpreted as the cyclically-adjusted long-term trend in public spending. Note that the inclusion of a linear and a quadratic time trend in our empirical model implies that we allow the trend to vary over time.

The lack of precision with which the state-level revenue impact of a tax shock is measured, represents an important limitation for our empirical analysis. Our tax shock series is constructed based on forecasts of the aggregate revenue impact of legislated tax changes across all 16 states. Nevertheless, as witnessed in Section 2.5, those forecasts appear to be accurate on average, yet we know nothing about how well they predict the revenue changes in each individual state. Thus, our independent variable of main interest is likely measured with noise. Consequently, under the classical errors-in-variables assumption, our estimates of β_j will suffer from an attenuation bias, meaning that they are biased toward zero. Our estimates can thus be interpreted as a lower bound of the true parameter β_j.

2.7 Results

The estimation results for Equation 2.1 are presented in Table 2.3. The first column of Table 2.3 shows the estimates when omitting the AR(4) term and the covariates, the second column when omitting only the covariates, and the third column for the full model. Our results indicate a strong and statistically significant impact of legislated tax changes on aggregate spending at the state-level. However, spending adjustments only occur with a significant delay. According to our estimates, an increase (decrease) in tax revenues by one Euro is associated with an increase (decrease) in public spending by 0.40 to 0.70 Euro two years after the occurrence of the tax shock, and by another 0.40 to 0.70 Euro after three years. Note that for each specification, the sum of significant coefficients never substantially differs from one, implying that any tax hike (tax cut) is followed by an increase (a reduction) in public spending of the same amount over the long-run. The delayed reaction of public spending after a legislated tax change, suggests a delay in political decision-making. State governments take time to adapt to revenue change and agree on an adjustment of public spending.

[9] Our baseline results in Table 2.3 do not change if we also include municipality-level revenue projections for all other states.

	Dep. variable: Aggr. Spending per capita		
	(1)	(2)	(3)
$\Delta\tau_t^x$	0.068	−0.079	−0.162
	(0.212)	(0.235)	(0.210)
$\Delta\tau_{t-1}^x$	0.288	0.107	−0.182
	(0.205)	(0.187)	(0.164)
$\Delta\tau_{t-2}^x$	0.585**	0.734***	0.417**
	(0.242)	(0.238)	(0.205)
$\Delta\tau_{t-3}^x$	0.491***	0.569***	0.424**
	(0.183)	(0.190)	(0.189)
$\Delta\tau_{t-4}^x$	−0.282	−0.103	−0.162
	(0.218)	(0.194)	(0.180)
AR(4)	✘	✓	✓
Covariates	✘	✘	✓
linear trend	✓	✓	✓
quadr. trend	✓	✓	✓
Observations	224	218	218
R^2	0.125	0.193	0.341

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 2.3: *Baseline Regressions: The Impact of Revenue Shocks on Aggregate Spending*

2.8 Extensions and Robustness Checks

2.8.1 Heterogeneous Spending Adjustments

Tax Hikes vs. Tax Cuts

We do not differentiate between tax hikes and tax cuts in our baseline specification. Instead, we study their effects combined. The *starving the beast* hypothesis, however, suggests exclusively focusing on tax cuts. In addition, Figure 2.1 of Section 2.4 illustrates that our sample mainly comprises tax cuts. In fact, tax hikes have been a rare event during our sample period, and if they did occur, they have only been of modest size. In light of that, it seems natural to merely focus on the effect of tax cuts. To this end, we re-estimate Equation 2.1 after omitting all tax hikes from our sample. The results are presented in the second column of Table 2.4. To facilitate the comparison with our baseline estimates, the first column of Table 2.4 replicates the results from our baseline specification.

The findings indicate that state governments appear to react faster to a tax cut than a tax hike. When omitting tax hikes from our sample, we see a statistically and economically significant adjustment in public spending one year after a tax shock occurs. Moreover, the coefficient estimates for the first three lags of our tax shock series become notably larger when only focusing on tax cuts. The sum of significant coefficient estimates is roughly equal to 2.6, thus indicating that the decrease in public spending far exceeds the magnitude of the tax cut.[10] A possible explanation is that state governments seek to repay the debt they incurred due to the delayed adjustment of public spending. On the whole, our findings lend strong support to the *starving the beast* hypothesis.

	Dep. variable: Aggr. Spending per capita	
	All tax changes	Only tax cuts
$\Delta\tau^x_t$	−0.079	0.396
	(0.235)	(0.356)
$\Delta\tau^x_{t-1}$	0.107	0.649**
	(0.187)	(0.275)
$\Delta\tau^x_{t-2}$	0.734***	1.127***
	(0.238)	(0.323)
$\Delta\tau^x_{t-3}$	0.569***	0.831***
	(0.190)	(0.227)
$\Delta\tau^x_{t-4}$	−0.103	−0.490
	(0.194)	(0.342)
AR(4)	✓	✓
linear trend	✓	✓
quadr. trend	✓	✓
Observations	218	218
R^2	0.193	0.239

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 2.4: *The Impact of Exogenous Tax Revenue Cuts on Aggregate Spending*

Highly vs. Modestly Indebted States

The degree of flexibility governments have when it comes to decisions about public spending depends on the level of public debt. It is typically assumed that the marginal costs of public debt increase with the level of debt. This implies that a high debt level restricts the government's room for manoeuvre and forces the government to react faster to a negative tax shock. To test this conjecture, we construct a dummy variable that takes the value one if a state's level of public debt in a given year is above the sample median and zero otherwise. Subsequently, we interact this dummy variable with our tax shock series in order to estimate separate coefficients for highly and modestly indebted states. The results are presented in Table 2.5. The coefficient estimates for highly and modestly indebted states are virtually

[10]The lower and upper bound of the 95 per cent confidence interval is [1.6; 3.7].

identical and scarcely differ from one another. Thus, our conjecture is not supported by the data. Note that this conclusion does not change when exclusively focusing on tax cuts.[11]

	Dep. variable: Aggr. Spending per capita	
	High debt level	Low debt level
$\Delta\tau_t^x$	-0.155	-0.189
	(0.233)	(0.263)
$\Delta\tau_{t-1}^x$	-0.239	-0.075
	(0.231)	(0.213)
$\Delta\tau_{t-2}^x$	0.534^{*}	0.478^{*}
	(0.297)	(0.258)
$\Delta\tau_{t-3}^x$	0.529^{*}	0.448^{*}
	(0.269)	(0.230)
$\Delta\tau_{t-4}^x$	-0.298	-0.091
	(0.256)	(0.204)
AR(4)	✓	✓
linear trend	✓	✓
quadr. trend	✓	✓
Observations	218	218
R^2	0.278	0.278

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}\ p < 0.10$, $^{**}\ p < 0.05$, $^{***}\ p < 0.01$

Table 2.5: *The Impact of Revenue Shocks on Aggregate Spending for States with High and Low Debt Levels*

Single-Party vs. Coalition Governments

Another potentially important factor that may affect a government's reaction to a tax shock, is a phenomenon referred to as the 'war of attrition' (e.g., Poterba and von Hagen, 1999, and Persson and Tabellini, 1999). Here, the argument is that divided governments or coalition governments have an incentive to delay an adjustment in response to a fiscal shock because they have different preferences regarding the measures of adjustment. For example, in the event of a tax shock, the parties forming a coalition may opt for spending hikes or cuts in different policy areas. The first party that concedes will alienate its constituents and thus bear a larger share of the political costs associated with spending adjustment. This is why each party tries to outlive the other(s) (Padovano and Venturi, 2001). For that reason, we would expect that single-party governments adjust public spending faster than coalition governments when a tax shock occurs. To test this hypothesis, we combine our tax shock series with a dummy variable that adopts the value one for coalition governments and zero for single-party

[11] Results available on request.

governments. We thus obtain separate coefficient estimates for both types of government. Our results do not support the 'war of attrition' hypothesis since the coefficients are virtually identical (see Table 2.6).

	Dep. variable: Aggr. Spending per capita	
	No Coalition	Coalition
$\Delta\tau_t^x$	0.101 (0.250)	−0.175 (0.296)
$\Delta\tau_{t-1}^x$	−0.271 (0.229)	0.280 (0.245)
$\Delta\tau_{t-2}^x$	0.675** (0.272)	0.758** (0.307)
$\Delta\tau_{t-3}^x$	0.651*** (0.245)	0.563** (0.257)
$\Delta\tau_{t-4}^x$	−0.031 (0.251)	−0.148 (0.231)
AR(4)	✓	✓
linear trend	✓	✓
quadr. trend	✓	✓
Observations	218	218
R^2	0.209	0.209

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.6: *The Impact of Revenue Shocks on Aggregate Spending for States with and without Coalition Governments*

Differences between Left-Wing and Right-Wing Governments

Finally, we investigate whether left (for example, governments led by the Social Democratic Party; SPD) and right-wing (such as governments led by the Christian Democratic Union; CDU) governments react differently to tax shocks. We do so by interacting the tax shock variable with a dummy for left-wing governments. A first glance at the coefficients presented in Table 2.7 suggests that left-wing and right-wing governments differ with regard to the timing of spending adjustments. Two years after a tax cut, left-wing governments appear to have fully adjusted public spending (in other words the decrease in tax revenues is mitigated by a decrease in public spending of the same amount). Right-wing governments, however, seem to require three years to reduce public spending by the amount necessary to balance the budget. It is worth noting, though, that despite the differences in their magnitudes, the coefficients do not significantly differ from one another at conventional levels of significance.[12]

	Dep. variable: Aggr. Spending per capita	
	Left-wing Party	Right-wing Party
$\Delta\tau_t^x$	0.006	−0.169
	(0.334)	(0.225)
$\Delta\tau_{t-1}^x$	0.299	−0.126
	(0.292)	(0.222)
$\Delta\tau_{t-2}^x$	0.994***	0.481*
	(0.323)	(0.285)
$\Delta\tau_{t-3}^x$	0.515*	0.618***
	(0.307)	(0.208)
$\Delta\tau_{t-4}^x$	−0.108	−0.028
	(0.305)	(0.233)
AR(4)	✓	✓
linear trend	✓	✓
quadr. trend	✓	✓
Observations	218	218
R^2	0.220	0.220

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}\,p < 0.10$, $^{**}\,p < 0.05$, $^{***}\,p < 0.01$

Table 2.7: *The Impact of Revenue Shocks on Aggregate Spending for States for Left- and Right-Wing Governments*

2.8.2 Spending on Sub-Categories

Until now, we have focused on the effects of legislated tax changes on aggregate public spending. Our findings provide strong support for the *starving the beast* hypothesis. They suggest that tax cuts are associated with reductions in public spending of a similar amount. However, the pending question is in which policy areas public spending is reduced in response to a tax cut. To answer this question, we re-estimate Equation 2.1 using per capita spending in different policy areas as the dependent variable. We focus on eight different policy areas: public administration, education, public safety and legal protection, health, infrastructure, social security, science, and culture. These are the largest spending items in German state government budgets.

[12]The p-value for a test of the null that the coefficient of the second lag of the tax shock variable is the same for left and right-wing governments is 0.193.

The results are shown in Table 2.8. To economise on space, we only report the results for those policy areas for which we find robust significant effects.[13] Our findings indicate that both the timing as well as magnitude of the adjustment in public spending varies considerably across the different policy areas. We see particularly large adjustments to tax shocks in public spending for administration and social security. Depending on the specification, the estimates suggest that an increase (a decrease) in tax revenues by one Euro triggers an increase (a reduction) in spending on administration by 0.30 Euro to 0.45 Euro in total. The adjustment in social security spending is of a similar amount. However, while spending on public administration is already reduced the year after a tax change is realised, the change in social security spending only occurs two to three years after the tax shock. Arguably, the difference with regard to the timing of the adjustment indicates that spending on social security is more prevalent than spending on public administration.

	Dep. variable: Spending per capita on ...							
	Administration		Health		Infrastructure		Social Security	
	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
$\Delta\tau^x_t$	0.135**	0.088	0.020	0.024	0.091**	0.097	0.061	−0.022
	(0.059)	(0.058)	(0.025)	(0.026)	(0.046)	(0.062)	(0.102)	(0.098)
$\Delta\tau^x_{t-1}$	0.315***	0.297***	0.018	0.022	0.090**	0.094	−0.003	−0.111
	(0.077)	(0.065)	(0.031)	(0.033)	(0.051)	(0.070)	(0.112)	(0.120)
$\Delta\tau^x_{t-2}$	0.038	−0.003	0.104***	0.103***	0.083	0.086	0.153*	0.043
	(0.078)	(0.073)	(0.031)	(0.032)	(0.054)	(0.238)	(0.081)	(0.088)
$\Delta\tau^x_{t-3}$	−0.005	−0.037	0.083***	0.082***	0.024	0.027	0.348***	0.251***
	(0.055)	(0.051)	(0.029)	(0.029)	(0.055)	(0.065)	(0.098)	(0.095)
$\Delta\tau^x_{t-4}$	−0.007	−0.036	0.019	0.014	0.104**	0.101**	0.110	0.076
	(0.059)	(0.057)	(0.018)	(0.018)	(0.041)	(0.043)	(0.068)	(0.064)
AR(4)	✓	✓	✓	✓	✓	✓	✓	✓
Covariates	✘	✓	✘	✓	✘	✓	✘	✓
linear trend	✓	✓	✓	✓	✓	✓	✓	✓
quadr. trend	✓	✓	✓	✓	✓	✓	✓	✓
Observations	160	160	160	160	160	160	160	160
R^2	0.480	0.551	0.227	0.254	0.178	0.206	0.255	0.292

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: *The Impact of Revenue Shocks on Spending on Specific Sub-Categories*

Aggregate spending on infrastructure features smaller spending increases in the beginning, and slightly more generous spending over the longer term. This is not surprising since many construction projects require considerable time for planning. Spending increases in the longer run, for example after four years by around 0.10 Euro per capita.

[13] All additional results are available on request.

We conducted the same exercise for additional expenditure categories. These include spending on education, science, legal protection, and public safety as well as culture. Point estimates for the impact of exogenous revenue changes for spending on these categories, were not statistically significantly different from zero at the five per cent level.

2.9 A Note on External Validity

Does a decline in tax revenue lead to a decline in public spending, as implied by the *starving the beast* hypothesis? Or does it cause higher deficits and subsequently higher taxes but no change in spending, as suggested by Romer and Romer (2009)? Our analysis of the fiscal consequences of tax shocks at the state-level in Germany supports the *starving the beast* hypothesis. It is important, though, to interpret our findings in light of the specific institutional framework of German fiscal federalism. German states have little tax autonomy. They cannot react to today's tax cuts by borrowing more and raising taxes tomorrow. Yet, they have other margins of adjustment. They could borrow more and lobby for higher taxes at the federal level. Or they could rely on future bailouts by the federal government. We can see that they prefer to cut spending. There are some federal states whose tax autonomy is also limited and tax sharing arrangements dominate. Examples include Austria and Belgium. In other federations like the US, Canada, and Switzerland, the states have more tax autonomy, and federal bailouts are less likely. In this type of institutional environment, the revenue shocks may result in different spending reactions. To shed some light on this issue, it is interesting to look at the literature on the so-called 'flypaper effect' (Courant et al., 1979 and Hines and Thaler, 1995), which observes that unconditional federal grants received by US states usually lead to higher spending but not to tax cuts. Standard theory would predict a combination of both. This observation could be interpreted as an indirect confirmation of the *starving the beast* hypothesis – provided that the reaction to changes in grants is symmetric.

2.10 Conclusions

The present chapter empirically tests the validity of the *starving the beast* hypothesis. According to this hypothesis, government spending can be restrained by limiting tax revenues. This conjecture relies on the argument that tax cuts commit the government to enact spending cuts in order to ensure that fiscal policy remains sustainable over the long-run.

In our empirical analysis, we use data from the German states covering the period from 1992 to 2011. We take advantage of the fact that the institutional setting of German fiscal federalism is ideally suited to study the consequences of tax changes for public spending. While the German states have full discretion about the level and priorities of public spending as well as the level of debt financing, they have an almost complete lack of tax setting autonomy. Both tax rates and tax bases are determined at the federal level. We use official governmental publications to compile a list of almost 130 legislated tax changes in Germany – each of them carefully classified by motivation in order to identify those that are unambiguously exogenous with respect to other economic or political fluctuations – together with the corresponding projected tax revenue changes for the states. We then use these projected tax revenue changes to study the causal influence of tax cuts on aggregate public spending, as well as spending in eight different policy areas.

We find that legislated tax changes have a strong and positive impact on state governments' aggregate expenditures. Our results suggest that a one Euro decrease in tax revenues leads to a decrease in public spending of about the same amount. However, it takes up to two or three years until the spending cut occurs. Our findings thus support the *starving the beast* hypothesis. Moreover, our results stand up against the inclusion of a range of important state-level covariates. Furthermore, with regard to the timing or size of the fiscal adjustment, we do not identify notable differences between unified and divided governments, left-wing and right-wing governments as well as across states with high versus low levels of debt.

When focusing on the structure of public spending, we find substantial differences across spending categories. While spending on public administration decreases immediately and substantially in response to a tax cut, health, social security, and infrastructure-related expenditure appears to decline only in the longer term, and in the case of spending on health and infrastructure to a minor extent. Education spending is largely unaffected. That undermines the widespread view that pressures to cut spending are counter-productive because governments respond by reducing public investment.

These results should be interpreted in light of the specific institutional setting of German fiscal federalism. German states have very limited tax autonomy and cannot react to revenue shocks by changing tax rates individually. They can only try to push for collective tax changes at the federal level or increase debt financing. In other federations with more tax autonomy at the state-level, state governments may behave differently.

Chapter 3

The Quality of Local Public Good Provision and Electoral Support

3.1 Introduction

Do voters take into consideration the quality of public services when going to the polls on election day? This question is not easy to answer, particularly because it is challenging to find an adequate metric that captures the quality of local public good provision at a granular geographic level. Specifically, consider issues such as whether sidewalks are clean, whether park benches need repair or whether abandoned cars block streets. Thus, it is not surprising that little is known about the effect of the quality of public service provision at the local level on electoral support of incumbent local politicians and parties. While some studies examine this relationship at the national and sub-national level (see for example Johnston and Pattie, 2001, Bartle, 2003 and Boyne et al., 2009), studies at the local level face data availability difficulties which limit both their internal and external validity. Much of the existing literature at the local level is restricted to city-specific case studies or cross-sectional variation therein, casting doubts on the generalizability of the results. For example, Burnett and Kogan (2017) exploit within-city cross-neighborhood variation with respect to the number of pothole complaints and electoral behavior in San Diego to identify the effect of pothole complaints on the vote share of the incumbent. Obviously, these neighborhoods might differ in other characteristics that affect electoral outcomes, or causality might be reversed. A second reason for concern relates to the choice of the outcome measure, which often appears to be guided by data availability rather than sound economic theory. Many of these outcomes used in retrospective voting, especially in the context of public services, are not necessarily salient to voters and may be poor measures of the actual quality of public good provision. For example, Arnold and Carnes (2012) use the number of city employees as a proxy for municipal service quality and find no effect on approval ratings of New York City mayors. It is unclear whether this is due to the fact that the number of local city employees is an inadequate proxy for the actual quality of municipal services, whether voters are unaware of the number of local administrative employees, or whether it is indeed irrelevant to their judgement. The same holds true for other approaches based on the size of the local council budget or its specific budget distribution across various spending categories, as used, for example, by Balaguer-Coll et al. (2015) and Litschig and Morrison (2012). Instead, it would be more convincing to focus on specific outcomes that are more visible to voters since they are more likely to retrospectively evaluate the work of the local council based on their daily experiences (Popkin, 1994).

Against this backdrop, this paper breaks new ground by deriving a new proxy for public goods and service provision at the local level that allows for a more rigorous testing of the effect in a more general econometric framework that is not limited to specific cities. I use complaints posted on the complaint-platform FixMyStreet.com. This platform is particularly popular in the United Kingdom with more than 2 million complaints posted to date since its launch in early 2007. Meanwhile, the platform is available for all regions in the UK and forms an integral part of the local complaint management infrastructure. Via their posts, users can alert their local authorities to problems such as potholes, broken streetlights, graffiti, abandoned cars, flypaper etc. Complaints are automatically geo-located by the application on the cell-phone and can thus be directly assigned to specific neighborhoods. Between May 2007 and May 2015, the dates of major local elections in the United Kingdom, around 600,000 posts were made by users in England. I use these observations for all 7678 wards of the country, the most disaggregated geographic area in the United Kingdom, to compute granular performance measures over four-year pre-election horizons based on how quickly these complaints are resolved by the local authority. The share of complaints not resolved after 12 months, which I use as an indicator for the quality of public services, is, as expected, negatively correlated with various municipal expenditure items, e.g., local waste management, and the number of employees in the public sector.

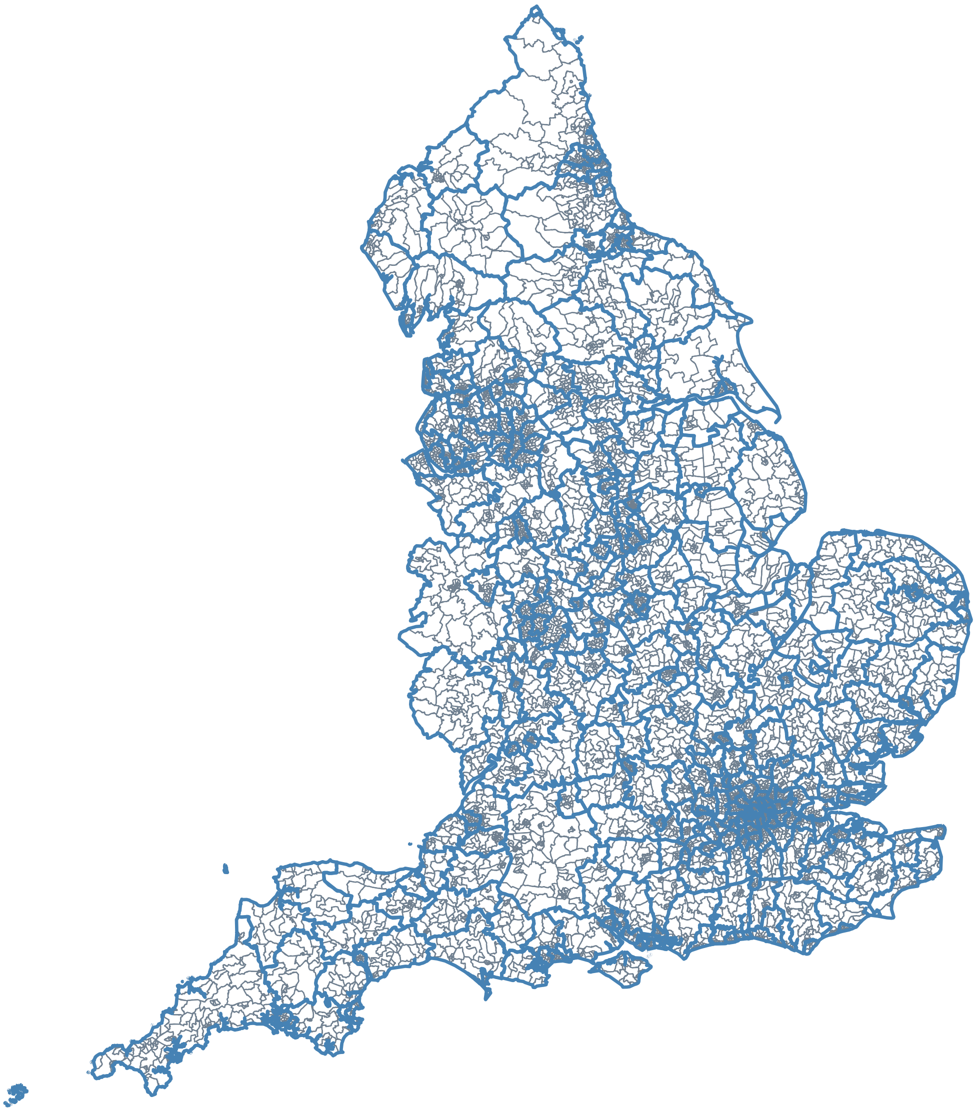
By leveraging within-neighborhood variation of the quality indicator over time, I show that these local performance measures are indeed highly relevant to the re-election of the local incumbent party in subsequent local elections. The estimation results provide compelling evidence for a strong punishment effect by voters: Parties in neighborhoods with better performance measures, i.e. with a relatively smaller share of complaints not resolved, also have higher re-election chances. The effect seems to be particularly strong for the quick removal of dog excrement and fly-tipping and is robust to various fixed effects specifications that account for both time-constant neighborhood characteristics and time-varying trends at the regional level.

This study draws on the online activity of citizens holding service providers accountable for the quality of service they provide (see Andrabi et al., 2017, and Banerjee et al., 2007 for additional examples in different contexts). Greater citizen participation has traditionally been hampered by existing practical constraints, in particular a lack of awareness of existing programs (Banerjee et al., 2010), and low participation rates due to classic collective action problems (Barr et al., 2014), as well as financial, time, and even social burdens on citizens (see Grossman et al., 2017, and Blair et al.). The growing availability of text-messaging platforms has significantly facilitated the process of monitoring service providers and providing feedback on their performances. Not only have they lowered the cost of participation through (anonymous) reporting, but they have also made the interaction with authorities more effective: While citizens are in a better position to judge the quality of provided services from their daily experiences, elected officials can enforce a higher quality of service delivery (Grossman et al., 2017).

The remainder of the paper is structured as follows. Section 3.2 provides a brief description of the local governmental system in the United Kingdom and England in particular. Section 3.3 describes the data on which the analysis is based and the computation of the quality indicators. Section 3.4 presents the estimation strategy and the results, followed by a series of robustness checks in Section 3.5. Section 3.6 concludes.

3.2 Local Electoral System in England

The structure of local government in the United Kingdom varies from area to area. For example, in most parts of England, there are two tiers – namely regions and local authority districts, or first and second-tier – between which responsibilities for local council services are divided. London as well as other metropolitan areas and some parts of shire England operate under a single tier structure that combines local council responsibilities for all services in their area. The level below the district level consists of roughly 8,500 so-called wards in the United Kingdom. Every ward elects councilors to be members of the local city or district council. Figure 3.1 depicts the electoral structure in England with districts highlighted in blue and wards highlighted in gray.



Note: Local Authorities are colored in blue and local wards are colored in gray.

Figure 3.1: *Electoral Structure in England: Local Authorities and Wards*

On average, each ward has an electorate of approximately 5,500 people, but ward population can vary to some extent. In cases of larger ward population, the inhabitants of the ward can nominate two, and in rare cases even three, local councilors so that the ratio of inhabitants per representative is kept approximately the same in all wards. If there is more than one vacant council seat, each party can nominate two or three candidates to stand for election, except for independent candidates without party affiliation. Councilors are ordinary members of the public elected by the local residents only. In this system, known as the 'multi-member plurality system', the candidate who received the most votes wins. There is no proportional representation as known from local elections in other countries. Councilors are usually elected for a period of four years. However, councils may be elected in full every four years, or alternatively 'by thirds', with one third of the councilors elected each year and one year without elections in between. Recently also the 'by halves' system has been allowed, in which half of the council is elected every two years. Elections usually take place on the first Thursday in May.[1]

The main task of a councilor is to represent his or her constituency in the next highest parliament. Councilors play an important role in planning, implementing, monitoring and developing local council activities. The competencies of local district, borough and city councils are far-reaching and include, for example, waste collection, recycling practices, housing development, as well as other planning applications such as repair work or local infrastructure maintenance and development. It is these services to which most of the requests on the online complaint platform relate. It is important to note that no individual councilor has sole decision-making authority. Instead, day-to-day decisions are made at the district-level by specific committees and the full Cabinet, which are made up of a number of councilors. Therefore, the councilors do not directly deal with individual complaints, e.g. those from FixMyStreet. Nevertheless, as observed by Dipoppa and Grossman (2020), many of the local civil servants who are responsible for complaint management are appointed by councilors whose strategic decisions they execute.[2] Moreover, assuming that addressing reported complaints does not only increase re-election chances of incumbent councilors but also the prospects of re-appointment of council staff, the incentives of both should align closely in this respect.[3]

Local governments in England have limited revenue-raising options. Funding comes from a combination of business rates, different types of central government grants and local council taxes. Smaller revenue sources include rents, fees and other charges at the local level. Earmarked central government grants make the largest source of revenue for local authority districts. These 'specific grants' are intended for specific purposes, such as the operation of local schools. 'General grants' make up a smaller fraction of central government funding and can be spent by district councils on any service. In addition to government grants, local authorities receive substantial funding from the 'Business Rates Retention Scheme'. These are taxes set by the central government and paid by occupants of non-domestic properties. Before 2013, revenues from business rates were collected by local councils and forwarded to the Treasury before being redistributed to local authorities via central government grants. Since 2013, councils have been allowed to retain up to 50% of these business rates revenues, with central grants being adjusted accordingly. If the 50% share of the business rates retained at the local level is considered too much or too little revenue in order to meet necessary spending requirements at the local level, the

[1] Electoral results at the ward level between 2007 and 2015 are retrieved from the *Local Elections Archive Project* (LEAP). The results can be accessed online at www.andrewteale.me.uk.

[2] According to the Quarterly Public Sector Employment Survey by the Office of National Statistics more than 12% of the total local government staff in the UK was employed on a non-permanent contract in the first quarter of 2015, with some local councils employing more than 40% or even 50% of their employees on a temporary basis.

[3] Illustrative examples of councilors who documented their (supposedly positive) record of solving FixMyStreet complaints on their professional websites can be found in Dipoppa and Grossman (2020).

central government can impose a 'tariff' or grant a 'top-up' in order to offset local spending capacity to some extent. Additional revenues come from local council taxes which are intended to balance local spending requirements and other revenues, in particular from the central government. Unlike the central government, local councils cannot borrow money in order to cover their day-to-day expenses, i.e. they must either maintain a balanced budget or draw on savings from previous years.

On the spending side, councils typically follow a 4-year planning cycle for their budgets. Major budget decisions are debated and approved by the Cabinet, the main decision-making body at the district level, in order to provide its local services and to maintain and develop local infrastructure. Central government can exert pressure on local authorities if councils fail to provide local services according to national standards, or if spending and council taxes increase substantially.[4] Education, adult and police services account for the largest spending items in local authorities' annual budgets, followed by expenditures on highways services. Much of the spending by local councils takes the form of earmarked central government grants, as noted earlier. However, 'general grants' and other revenue sources give councils leeway to prioritize specific budget categories over others by topping-up available resources.

3.3 Data and Quality Indicators of Local Public Good Provision

The analysis is based on all complaints that were posted on the app-based platform FixMy-Street.com between May 2007 and May 2015, the dates of general local elections in the UK. Using the application on their cell-phones, users can post complaints of various categories on the platform.[5] To do so, users first enter a UK postcode or street address, or they can be located automatically. On the map, users can then pinpoint the exact location of the incident and add a subject line, a category, a detailed description and pictures. Once completed, the complaint report is forwarded, usually within minutes, to an email address of the responsible local authority, including the specific geo-coordinates of the complaint.[6] Users do not need to register on the website, nor are they required to use their real name. Only a valid email address must be used. Figure 3.2 shows an example of a complaint from the 'Fly-Tipping' category. The description on the left side of the screen-shot shows when the complaint was posted (and by whom) and if (and when) the complain twas solved by the local authority.

As Solymosi et al. (2017) note, users of FixMyStreet are also consumers of the platform in the sense that they have an intrinsic interest in seeing their complaints resolved. Thus, it appears reasonable to assume that the reporting mechanism creates sufficient incentives for users to provide accurate information, in particular with respect to the exact location of the incident, so that complaints can be solved accurately and quickly.

In principle, complaints can be labeled as "solved" by all users, i.e. local council staff, the person who posted the complaint or any other user of the app. Councils typically use official council accounts with names such as " Oxfordshire County Council". Of the resolved complaints in this sample, 18% were labeled as "solved" by a user with an alias containing the string "Council" (or "council"). In 38% of the

[4]For example, under the 2011 amendments to the Localism Act, the local council must hold a referendum on the proposal if spending increases such that local council taxes must to be raised by more than 2% in order to maintain a balanced budget.

[5]In July 2017, MySociety, the charity organization that provides the open source software for the FixMyStreet platform, reported that the application was downloaded 40,000 times in the UK: see their blog post at mysociety.org/2017/07/21/the-big-one-million-celebrating-fixmystreet.

[6]The email address needs to be added by the local authority when activating the service of the platform for its area.

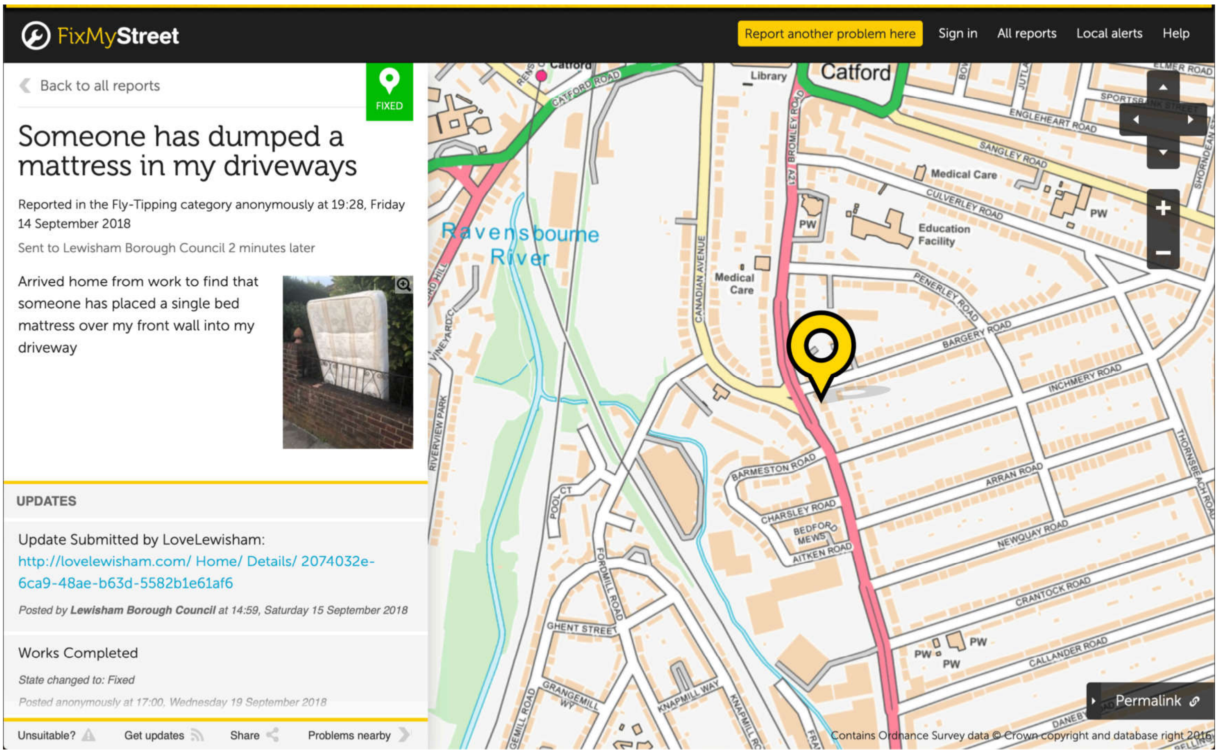
cases it was a person without an explicit user name ("Posted anonymously", see also the example from Figure 3.2) and the remaining 44% provided either their real names or variants of it. Thus, judging by the users' online names, councils appear to play only a minor role in reporting of resolved complaints on the platform. Nevertheless, while less likely for users reporting problems, council staff responsible for the communication on the platform but without specific knowledge of the status of repair work could potentially lead to measurement errors in the online data used for this analysis. This could apply to both the status of open complaints (fixed or not?) and the date the repair work was completed (when fixed?). In this case, the resulting measurement error in the neighborhood-specific quality indicator causes the parameter estimate in the linear regression model to be biased towards zero (Griliches and Hausman, 1986). Econometric specifications with unit fixed-effects are particularly susceptible to this 'attenuation bias' when the true regressor, in this case the actual quality of local public good provision, is (highly) correlated over time, while the measurement error is merely serially uncorrelated noise. Despite the inherent advantage of avoiding bias from omitted time-consistent characteristics, the focus on deviations from unit-specific means in the fixed-effects specification removes much of the variation in the variable of interest in this case. Of course, the fixed-effects specification is particularly useful if the measurement error is a fixed effect in itself. For example, some ward's or council's complaint reporting procedure might consistently over- or understate the true quality indicator by a constant share, say ten percentage points. In this case, the inclusion of ward-specific fixed effects completely eliminates the measurement error in the regressor.

In addition to measurement errors, it is possible that problems are reported by multiple users unaware that these complaints have already been posted by others.[7] To assess the extent of this potential issue, I apply an algorithm that identifies clusters of geographically close complaint items from the same category (potholes, abandoned cars etc.) and from the same quarter of the year in order to group complaints by time and category. The DBSCAN (short for *Density-based spatial clustering with applications and noise*) technique, first introduced by Ester et al., 1996, is particularly suitable for this exercise. The algorithm treats the number of clusters as endogenous, i.e. it does not need to be specified *a priori*, which is an inherent advantage over other algorithms in this context. Instead, the number of clusters (of arbitrary shape) is derived from the data based on the parametrization of two critical factors: the maximum distance between two points to be considered as belonging to the same geographic cluster, and the minimum number of observations necessary to form a cluster. In this case, the minimum number of complaints is set to 1, so that every post could potentially form a cluster on its own. What remains to be assessed is its sensitivity to the maximum distance between geographic locations of the complaint items. Table 3.1 summarizes the results the algorithm yields for various reasonable values of the maximum distance.[8]

When the distance between individual complaints is gradually increased in steps of ten meters (first column), the algorithm shows that some complaints within groups of the same category and quarter of the year are indeed duplicates of other existing posts. For example, when grouping complaints which are within ten meters of each other, the algorithm indicates that around 16.5 thousand out of around 580,000 complaints are redundant. In relative terms, this amount of excluded complaints equals less then

[7] In principle, when locating the incident on the map during the reporting process, users can see other open posts in the same neighborhood and have the possibility to check if a complaint has already been reported.

[8] A Python procedure for this routine is described by Boeing (2018).



Note: The example is taken from fixmystreet.com/report/1413342.

Figure 3.2: *Example Complaint Posted Online on FixMyStreet.com*

3% of the initial number of posts.[9] When increasing the radius around complaint items, the number of suspected duplicates increases, albeit at a decreasing rate in absolute amounts. However, even with a radius of 60 meters around complaints the average cluster size barely exceeds 1.1 complaints per cluster. Hence, the problem of double-posting does not appear to pose a threat to the validity of the data.

In what follows, I apply a geographic clustering algorithm that uses a maximum of 20 meters as the threshold for grouping complaints. The threshold value is large enough to eliminate a small number of obvious complaint duplicates and at the same time small enough not to combine an unreasonable number of complaints. Within the same cluster of complaints, posts are sorted based on whether they were resolved within 12 months or less and the first post is kept in order to provide accurate information about when the problem was first reported.

Table 3.2 gives an overview of categories, the number of complaints in each category, as well as the share of complaints that were solved and how long this took on average. In descending order, the categories with the largest number of posts are complaints related to 'Potholes', 'Vegetation, Weed and Hedges' and 'Roads and Road Cleaning'. Smaller categories include, for example, 'Pavement and Footpaths' and 'Fly-tipping'. On average, over the eight-year time horizon considered here, around 40% of all complaints are resolved and this process takes less than two months on average. However, considerable

[9]Complaints without a category, to which the algorithm cannot be applied, are not dropped from the sample in any of the cases discussed in the table.

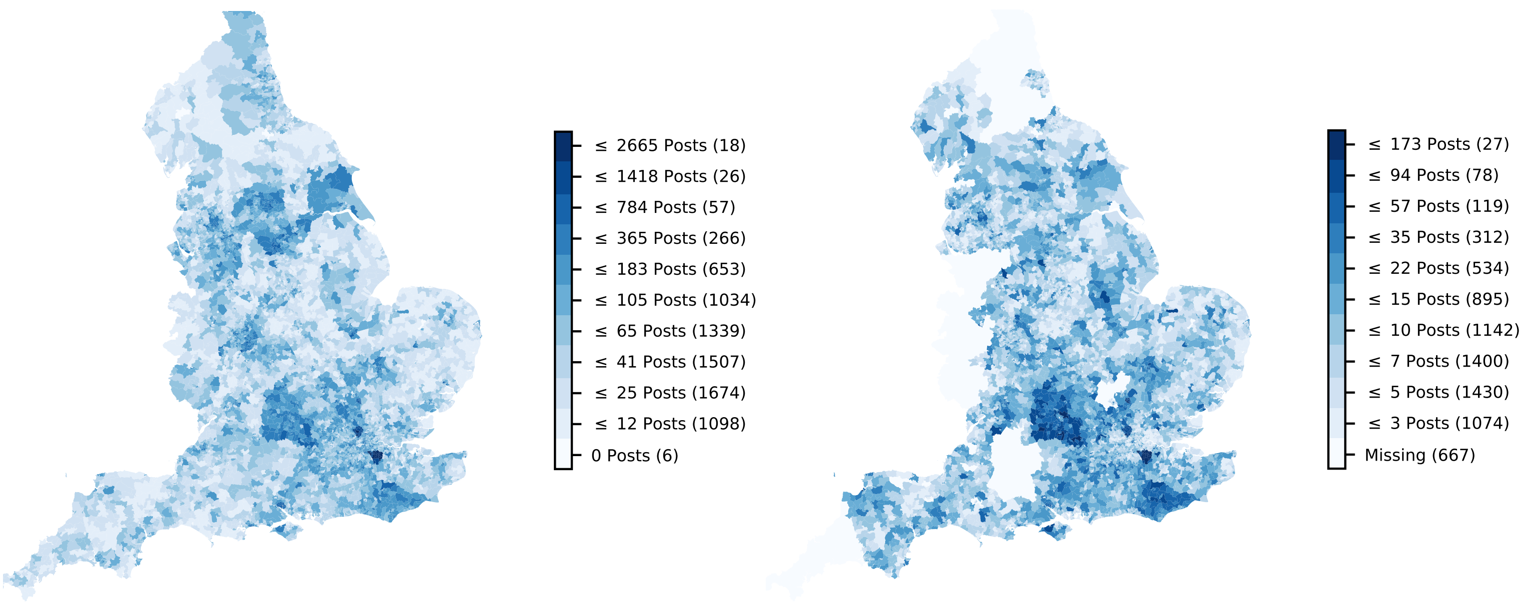
Maximum Distance between Complaints (in m)	Spatial-Time-Category-Duplicates Removed	Rel. Share of Rem. Duplicates (in %)	Average Cluster Size	Maximum Cluster Size
0	0	0.00	1.000	1
10	16 587	2.87	1.031	42
20	26 438	4.58	1.051	43
30	34 008	5.89	1.067	50
40	40 161	6.96	1.080	50
50	45 844	7.94	1.093	52
60	51 096	8.85	1.104	71

Note: Complaints from the same category and the same quarter of the year are grouped together before applying the spatial clustering algorithm. Complaints with no category assigned are not dropped.

Table 3.1: *Data Cleaning for Removal of Duplicate Complaints*

differences between the categories can be observed: for example, relatively complex defects that would be expected to require more time and materials to fix, such as complaints regarding 'Dangerous Construction' take considerably more time than comparatively 'simple' matters such as removing trash and dead animals or emptying accumulated waste. Similarly, the proportion of resolved complaints from these categories is considerably higher.

The geographic distribution of complaints is shown in Figure 3.3, both in terms of absolute counts (Panel (a)) as well as relative counts per 1000 residents (Panel (b)).[10] For virtually all wards in England, at least one complaint has been posted on the platform between May 2007 and May 2015, except for six wards which I exclude from the analysis.



(a) *Absolute Number of Complaints*

(b) *Relative Number of Complaints (per 1,000 residents)*

Figure 3.3: *Geographic Distribution of Complaints between May 2007 and May 2015*

[10] Wards for which population data at the ward-level are not available are labeled as "Missing" in Panel (b) of Fig. 3.3.

Category	Complaints (total #)	Share fixed (in %)	Average Duration until fixed (in days)
Abandoned Vehicles	12 589	44.99	51.81
Accumulated Litters	4494	49.93	27.26
Bus Stops	1551	34.69	71.50
Car Parking	15 371	18.55	82.52
Dangerous Construction	1009	29.93	102.17
Dead Animals	1070	64.11	13.86
Dog Fouling	13 524	24.03	52.27
Drainage	6651	46.04	64.00
Fly-tipping	38 366	51.50	41.18
Graffiti	14 436	48.69	52.20
Parks	6982	36.98	64.28
Pavement and Footpaths	50 937	35.79	67.41
Potholes	131 321	42.49	57.46
Roads and Road Cleaning	97 273	33.97	67.14
Rubbish	12 354	59.82	30.44
Traffic Lights	7010	44.94	53.57
Vegetation, Weed and Hedges	98 708	48.86	51.26
No Category	36 929	29.31	71.82
Total	550 575	40.80	56.37

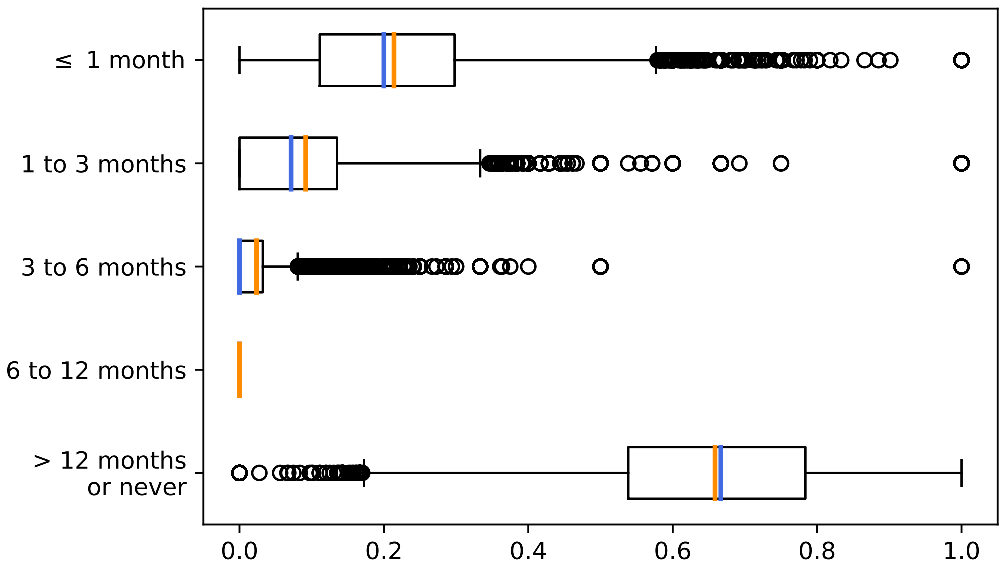
Note: Complaints with a duration between the date when reported and the date when labeled as fixed of more than two years are not included in the average duration of the main categories.

Table 3.2: *Summary of Descriptive Statistics for Major Complaint Categories*

Over eight years, the average number of complaints within a ward is approximately 31 (or 4.8 per 1000 residents), which I consider sufficiently large to compute robust quality measures. Some regions, especially larger or metropolitan regions and cities, have higher absolute and relative complaint counts per capita than more rural areas, for instance, in the south-west of England.

Following the approach of Dipoppa and Grossman (2020), neighborhood-specific measures of the quality of local public good provision are based on the council's responsiveness to online complaints. Based on detailed information about when complaints were reported on the platform, the time it takes for the local council to solve the issue is determined for each recorded complaint. I then compute for every ward in England the share of complaints that are solved within certain time horizons, ranging from shorter ones, such as 30 days, to longer ones, such as six months. Obviously, a larger share of complaints that are solved within a month or even less speaks to a higher responsiveness of local councils, while a larger proportion of complaints only addressed after a year (or never) indicates poor management of citizen complaints. Unlike other measures used in the literature, such as self-reported performance indicators (Carey et al., 2009) or citizens' assessments of the quality of elected representatives (Clinton et al., 2004), the proposed indicator has the advantage of being a more objective measure that captures spatial and temporal variation in the responsivenss of councils responsiveness at a very granular geographical level for all of England. What is more, resulting measures are highly visible to local voters and have the potential to affect their everyday experiences, which is an important determinant of voters' retrospective evaluation of the council's work (Popkin, 1994).

The distribution of each quality indicator, computed for all English wards for two four-year windows between election dates (2007-2011 and 2011-2015), is illustrated in Figure 3.4. Focusing on five different time horizons, the figure displays boxplots to illustrate the range and the distribution of quality indicators across English neighborhoods, with the median of each distribution highlighted in blue and the mean in orange in the interquartile range box. The share of complaints that are solved relatively quickly within one month (first row) and the share of complaints that are solved only after twelve months or not at all (last row) show the largest variation among all time horizons considered, as can be readily seen from the wider interquartile range boxes. In most wards, complaints fall in either category while only little activity can be documented for the other three time horizons. In most cases, complaints that are resolved after twelve months or never account for the largest proportion. As indicated by the fact that the value of the first quarter of the distribution of this measure is approximately 50%, three quarters of the neighborhoods in England do not manage to solve half of the reported complaints after one year. Nevertheless, the distribution is skewed to the left, suggesting that some neighborhoods perform better than others in this category, with some even solving all requests within twelve months. The results are closely mirrored by the distribution of the share of complaints resolved within one month. The distribution has a median of roughly 20% and is skewed to the right. With little repair activity recorded between one and twelve months, the sum of the fraction of complaints solved either within 30 days, only after one year, or not at all is fairly close to one for most neighborhoods, thus describing two sides of the same coin. Given a slightly higher variation in the share of complaints solved only with considerable delay or never, I stick to this indicator of the quality of local public good provision for the remaining analysis.



Note: Horizontal lines in blue and orange denote median and mean of each distribution, respectively.

Figure 3.4: *Quality Indicators at the Ward-Level: Fraction of Complaints solved in Different Time Horizons*

To provide empirical support for the proposed proxy, Table 3.3 explores how the proposed quality indicator $Q_{i,t-4}$ and its sub-components correlate with other observable indicators, particularly in terms of local-level spending and the number of employees in the public administration. For the pair-wise correlations I use population-weighted averages of of the quality indicator at the level of the local authority for which other indicators are available as well. Over the same four-year time periods as for the quality indicators, I compute average spending and divide it by population counts in 2011 and 2015, respectively. The number of public sector employees is expressed per thousand inhabitants of the respective ward.

As shown in Table 3.3, the suggested quality indicator $Q_{i,t-4}$ as well as the performance measure in specific sub-categories are negatively correlated with overall per capita spending of the local administration and explicit local council spending on service provision. Although they are smaller in absolute terms, pair-wise correlations with the number of public sector employees are also negative. As expected, the correlation table confirms that higher expenditures per inhabitant correlate with smaller, i.e. better, quality indicators. The correlation is stronger for complaints that fall into the categories 'roads and road cleaning', 'pavements, footpaths and potholes' and 'traffic lights'. The fact that the correlation coefficients are not larger in absolute magnitude, say above 0.7, is not surprising given that idiosyncratic shocks at the local level in combination with other district-specific features, such as the efficiency of the organization of the administration, are likely to co-determine the quality of public services and, thus, the quality indicators. In fact, simply regressing the share of complaints related to potholes, the largest category in terms of complaint counts, on the district's spending per capita on services suggests that more than 15 per cent of the total variation in the quality indicator across neighborhoods and years can be explained by differences in expenditures, providing further evidence that the quality indicator, when aggregated to higher geographic levels, closely aligns with alternative indicators which are not available at more granular geographic levels.

3.4 Estimation Strategy and Baseline Results

In order to identify the effect of the quality of local public good provision on the re-election chances of incumbent parties, I rely on variation in the quality of public good provision within spatial units over time. More specifically, I estimate variants of the following two-way fixed-effects model

$$P(Change_{i,t}) = \beta_1 Q_{i,t-4} + \beta_2 log(compl.p.c._{i,t-4}) + \eta_i + \eta_t + \gamma X^e_{i,t} + \eta_{la,t} + \epsilon_{i,t}$$

where $Q_{i,t-4}$ corresponds to the ward-specific quality performance indicators over the four-year time period prior to the election date in t. To account for time- and local ward-specific characteristics, I include time and neighborhood fixed effects, denoted η_t and η_i, respectively. The relevant outcome variable is the probability that the current party of the councilor in ward i is voted out of office and replaced by another party. For larger wards with multiple councilors potentially from different parties, at least one party must be replaced by another for this to count as a change.[11] In the baseline specifications, I focus on party-level outcomes, as opposed to the re-election chances of individual politicians.

[11] I include only those wards in which there were no elections between 2007 and 2011 and between 2011 and 2015, i.e., I discard wards that held elections 'by thirds' or 'by halves'.

	Total spending p.c.	Spending on all services p.c.	# public sector employees
$Q_{i,t-4}$	−0.1000 (678, 0.0092)	0.0462 (221, 0.4942)	−0.0465 (514, 0.2929)
$Q_{i,t-4}^{\text{roads, road cleaning}}$	−0.1536 (678, 0.0001)	−0.3724 (221, 0.0000)	−0.0909 (514, 0.0394)
$Q_{i,t-4}^{\text{dog fouling, fly-tipping}}$	−0.1410 (678, 0.0002)	−0.3082 (221, 0.0000)	−0.0749 (514, 0.0896)
$Q_{i,t-4}^{\text{graffitis}}$	−0.0760 (678, 0.0480)	−0.0791 (221, 0.2418)	−0.0311 (514, 0.4813)
$Q_{i,t-4}^{\text{pavements, footpaths, potholes}}$	−0.1635 (678, 0.0000)	−0.3890 (221, 0.0000)	−0.0978 (514, 0.0266)
$Q_{i,t-4}^{\text{abandoned vehicles}}$	−0.1161 (678, 0.0025)	−0.2542 (221, 0.0001)	−0.0871 (514, 0.0485)
$Q_{i,t-4}^{\text{vegetation, weed, hedges}}$	−0.1233 (678, 0.0013)	−0.2729 (221, 0.0000)	−0.1032 (514, 0.0192)
$Q_{i,t-4}^{\text{traffic lights}}$	−0.1484 (678, 0.0001)	−0.3454 (221, 0.0000)	−0.0834 (514, 0.0588)

Note: The data on council-level spending and public sector employment numbers are retrieved from the *Local Government Association*. The number of observations based on which pair-wise correlations are computed and the p-value is shown in brackets below each correlation coefficient.

Table 3.3: *Pair-wise Correlations of Quality Indicators with Alternative Measures*

The reason is that local elections, unless they take place in very closed settings such as schools (Berry and Howell, 2007), are usually "low-profile, low-cost, and low-information affairs" (Bonneau and Cann, 2015). With little knowledge about the personal characteristics of the incumbent and the opposing candidates and their specific programs, voters would likely be guided by their party preferences. I conduct the analysis again in one of the robustness checks at the individual politician-level. Furthermore, following Krebs (1998) and Grossman and Michelitch (2018), I also control for election-specific characteristics related to competitive pressure, such as the ratio of candidates to available councilor seats and the vote margin in the last election.[12] District-specific time-varying factors which could potentially affect local voting decisions in neighborhood i are captured by an additional fixed-effect $\eta_{la,t}$. These include, for example, regional unemployment and inflation rates, local price levels, crime rates, prevailing council party majorities, government formations, party campaign spending, newspaper endorsements and other district-level trends. What is more, these fixed-effects also capture the size of the district council budget and other budget characteristics, including the allocation to specific expenditure items and the level of the council tax.

The corresponding regression results are displayed in Table 3.4. Throughout all specifications of the model, we see that an increase in the share of complaints solved only after 12 months or not at all goes along with an increase of more than eight percentage points in the probability that the incumbent party is voted out of office. The results provide compelling evidence that voters reward local politicians for

[12]It is true that personal characteristics of a politician, e.g. education, experience, age, sex and race, can also play a role. The *Local Government Association* regularly censuses among local councilors, but unfortunately the data is not publicly available.

	Dep. variable: P(At least one party loses council seat)				
	(1)	(2)	(3)	(4)	(5)
Share of Complaints fixed after 12 months or never	0.099**	0.098**	0.097**	0.086*	0.084**
	(0.013)	(0.014)	(0.016)	(0.063)	(0.041)
Logarithm of Complaints per capita		−0.005	−0.010	−0.010	−0.001
		(0.525)	(0.459)	(0.459)	(0.952)
ward FE	✓	✓	✓	✓	✓
year FE	✘	✘	✓	✘	✘
election covariates	✘	✘	✘	✓	✘
local authority-year FE	✘	✘	✘	✘	✓
Observations	5630	5630	5630	5190	5630

Note: Results are based on OLS estimation. I use clustered standard errors at the level of the local ward. Corresponding p-values are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 3.4: *Baseline Results: The Impact of the Quality of Public Services on Re-election Chances*

improving the timely provision of local public services. Including election co-variates or district-year fixed-effects in the regression specification leads to slight, but not statistically significant, changes in the coefficient estimate. Interestingly, the number of complaints *per se* does not seem to matter. The point estimate is both statistically and economically insignificant, suggesting that it is not the occurrence of a complaint that is important but how the complaint is handled.[13]

We can also 'unpack' the quality indicator and focus on its specific components in order to assess what matters most to the electorate in local elections. Figure 3.5 shows the coefficient estimates from a two-way fixed-effects model, equivalent to model (3) in Table 3.4, when using only the quality indicators computed for complaints from specific categories. Among the sub-categories for which sufficient observations are available, the largest coefficients and the only statistically significant effects can be found for 'dog fouling and fly-tipping' as well as 'vegetation and hedges'. The effects of the other categories are not statistically distinguishable from zero.

3.5 Robustness Tests

A series of robustness checks support the baseline results. In the first column, I focus explicitly on 'contested wards', i.e. those neighborhoods in which the councilor had a vote margin of five percentage points or less in the 2007 election. The fact that the coefficient estimate is three times larger than that in the baseline results seems to suggest that voters in these wards are much more thorough in examining the quality of public goods than in uncontested wards. The number of complaints per capita remains irrelevant to the probability of re-election.

One might also assume that voters are to some degree affected by the quality of public good provision in adjacent neighborhoods. For example, poor provision of public goods in nearby neighborhoods could

[13]When altering the specific form of the regressor, such as the logarithm of the number of complaints or complaints per capita, the results do not change.

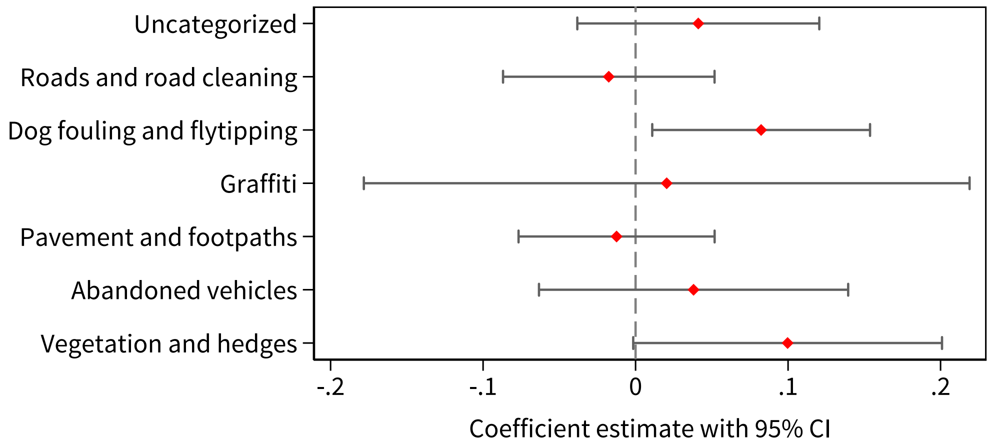


Figure 3.5: *The Impact of the Quality of Services of Sub-Categories on Re-election Chances*

potentially contribute to a general feeling of dissatisfaction among constituents about the efficiency of the work of local representatives, regardless of the actual quality of public good provision in their own neighborhood. With this in mind, it makes sense to extend the baseline specification to include spatially lagged regressors that control for the quality indicators in the direct neighborhood of ward i. Following Dubin (1998), I extend the specification as follows,

$$P(Change_{i,t}) = \beta Q_{i,t-4} + \rho \sum_j w_{i,j} Q_{j,t-4} + \beta_2 log(compl.p.c._{i,t-4}) + \eta_i + \eta_t + \epsilon_{i,t},$$

with $w_{i,j}$ being the ij-th cell of a spatial weights matrix W and $Q_{j,t-4}$ being the quality performance of neighborhood $j \neq i$. The matrix W assigns non-zero values if neighborhood j shares a border (of unspecified length) with neighborhood i. Row values are standardized by the number of geographic neighbors of i, so that the additional regressor $\sum_j w_{i,j} Q_{j,t-4}$ corresponds to the average quality indicator in direct neighborhoods of ward i.[14] Including the average quality indicator of direct neighbors of i does not alter the previous results, suggesting that spatial spill-over effects from surrounding neighborhoods are not relevant for the effect of the quality of local services in neighborhood i.

Similarly, as an alternative functional specification, the logit regression with unit and year fixed effects confirms the sign and significance of both the quality indicator and the complaints per capita.

In the last column, I move from the party-level to the individual politician-level and model the probability of an incumbent politician being voted out of office as a function of the quality indicator in his or her constituency since the last election. Including fixed-effects for each politician, the analysis is restricted to those who served as councilor for a specific neighborhood and were up for re-election in the subsequent election four years later. In total, this applies to 9020 politicians. The results of the regressions are remarkably similar to those at the party-level. While the number of complaints is

[14] Some English neighborhoods in the East share borders with Wales and others in the North share borders with Scotland. When computing the average performances in the neighborhood of i, the quality of public good provision in these neighborhoods outside of England is not considered.

virtually irrelevant, the share of complaints resolved within 12 months or thereafter is an important determinant of politicians' re-election chances.[15]

	Dep. variable: P(At least one party loses council seat)			
	Contested Wards	Spatial auto-corr.	Logit Model	Politician-Level
Share of Complaints fixed after 12 mon's/never	0.331*** (0.002)	0.090** (0.025)	0.711** (0.018)	0.040** (0.037)
Logarithm of Complaints per capita	−0.030 (0.346)	−0.006 (0.491)	−0.073 (0.439)	0.001 (0.919)
unit & year FE	✓	✓	✓	✓
Observations	1576	5576	1160	9020

Note: Results are based on OLS estimation. I use clustered standard errors at the level of the local authority and the level of the individual politician, respectively. Corresponding p-values are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 3.5: *Robustness Checks: The Impact of the Quality of Public Services on Re-election Chances*

3.6 Conclusion

In this paper, I investigate the link between the quality of local public services and the chances of incumbent parties to be re-elected in local elections. Finding a suitable metric for the quality of public good provision at a granular level that captures differences in the quality of public good provision across spatial units and time has been a shortcoming of previous studies, limiting both the internal and external validity of the results. I use 550,000 complaints from the online platform FixMyStreet.com that were submitted between May 2007 and May 2015, the dates of major local elections in the UK, to derive indicators for all 7500 wards in England, the most granular geographic level in the UK. I focus on the share of complaints that were resolved only after 12 months or never, with larger shares indicating poorer provision of public services. At the district-level, the next level up in the geographic structure of the United Kingdom, the indicator correlates negatively with various expenditure items of local councils and the number of employees in the public sector.

The results provide compelling evidence that in wards where public services are of poorer quality, i.e. wards with a large fraction of unresolved complaints, the chances of the incumbent parties to be re-elected in 2011 and 2015 are about nine percentage points smaller. The results are robust to various fixed-effects specifications and the inclusion of election-specific co-variates at the ward-level. The timely removal of dog waste and fly-tipping appear to be particularly important to voters. Interestingly, the number of complaints per capita in each ward does not influence re-election probabilities, suggesting that what matters in evaluating the work of the councilor is not the incident of a complaint *per se* but how it is handled.

[15] Including fixed effects for the party that the councilor is affiliated with (or whether he/she is without party affiliation) does not change the results.

Chapter 4

The Tax-Elasticity of Corporate Profits under Formula Apportionment

This chapter is based on joint work with
Clemens Fuest and **Florian Neumeier**.

4.1 Introduction

The international system of corporate taxation is under tremendous pressure. Calls for a substantial reform are coming from national governments as well as international organizations, and they are getting louder. One of the main points of criticism is that the current system fails to ensure that corporate profits are taxed where value is created. A series of recent leaks indicates that multi-jurisdictional enterprises (MJEs) exploit inconsistencies between national tax rules and loopholes in existing regulations in order to shift profits to jurisdictions where they are taxed at low rates or, in some cases, not at all. The process of digitization, allowing firms to serve markets without physical presence, as well as MJEs' increasing reliance on intangible assets have aggravated this issue. The strategic location of intangibles, such as patents, trademarks, and copyrights, is believed to be one of the main channels through which MJEs shift profits (see Heckemeyer and Overesch, 2017, and OECD, 2018, on this topic).

The vulnerability of the international tax system to MJEs' profit shifting and tax avoidance strategies is related to its design. The current system is based on separate accounting (SA), meaning that taxable profits are calculated separately for each jurisdiction in which an MJE operates. Since the definition of 'taxable profits' varies across jurisdictions, MJEs have an incentive to organize their international activities as to minimize their overall tax burden. Moreover, under SA, the valuation of cross-border transactions between affiliates of the same MJE group bears great importance for the distribution of profits. The arm's length principle (ALP), which is supposed to ensure that prices for intra-firm transactions reflect market values, has severe limitations, since many goods and services that are exchanged between affiliates belonging to the same MJE group are hard, if not impossible, to value. This makes the valuation of intra-firm transactions prone to manipulation and therefore offers considerable scope for profit shifting.

Many policy-makers and economists see the solution to these problems in a change from SA to a system of formula apportionment (FA). Under FA, the accounts of an MJE's affiliates are consolidated and then divided among the jurisdictions in which the MJE operates based on one (or a list) of apportionment factors that measure economic activity in the jurisdictions, such as, for instance, payroll costs, asset values, and sales. The jurisdictions can then apply their own tax rates to determine the corresponding tax liability. The consolidation of profits and the formulaic apportionment of the tax base would – at least in theory – resolve any inconsistencies between national tax regulations and make the valuation of intra-firm transactions superfluous. Thus, it would be ensured that corporate profits are taxed where they are generated.

In light of its potential benefits, the European Commission proposed adopting an EU-wide FA regime in 2001 (European Commission, 2001). The initial proposal was adapted and refined in subsequent years (European Commission, 2011, 2015). According to this proposal, the profits of companies with cross-border activities within the EU would be consolidated at the EU-level, creating a so-called common consolidated corporate tax base (CCCTB). The CCCTB would then be distributed among EU member states on the basis of three apportionment factors: assets, labor (payroll and number of employees), and sales. Thus far, though, FA regimes are only found at the sub-national level. Examples include Canada, Japan, Germany, and the US.

However, the adoption of FA may have downsides as well. McLure Jr (1981) as well as Gordon and Wilson (1986) highlight that under FA, the corporate tax effectively becomes a tax levied on the apportionment factors, and that tax rate differences may distort the allocation of these factors across jurisdictions. This, in turn, also affects the distribution of profits. Simply speaking, under FA, MJEs have an incentive to relocate apportionment factors to low-tax jurisdictions in order to minimize their tax burden, while under SA, they may only shift paper profits by exploiting gaps in national tax legislation. Despite this insight, existing studies estimating the corporate tax revenue effects of an introduction of FA on a global scale (e.g., Cobham and Loretz, 2014), in the EU (e.g., Devereux and Loretz, 2008; Fuest et al., 2007), or unilaterally by the US (e.g., Clausing and Lahav, 2011; Shackelford and Slemrod, 1998) ignore potential behavioral responses of MJEs to corporate tax rate differences.

So far, empirical evidence on the consequences of tax rate differences for the distribution of MJEs' profits and apportionment factors under FA is scarce. This study aims at filling this gap by using firm-level data from Germany. In Germany, profits are taxed both at the federal level and at the local level. The local business tax is levied by the municipalities, which can set their tax rates independently. Local business taxation takes place in the form of an FA regime. The profits of firms with entities in multiple municipalities are consolidated at the national level before being divided between municipalities that host entities of an MJE based on the relative payroll share. Consequently, in our analysis, we focus on the tax sensitivity of MJEs' profits and payroll costs. The data that we employ is provided by German tax authorities through the Federal Statistical Office (local business tax statistics; *Gewerbesteuerstatistik*). It covers the universe of German firms and allows us to identify which firms operate in multiple municipalities as well as how profits and payroll costs are distributed between host municipalities. Our sample covers the years from 2010 to 2014 and includes roughly 170,000 MJEs per year. A unique tax identification number for each firm allows us to link the data to a panel. While the definition of the tax base is uniform across all municipalities, local business tax rates vary notably both across time and space, which makes Germany particularly well-suited to study the behavioral responses of MJEs to tax rate differences under FA.

Our results provide compelling evidence for a strong response of MJEs to local tax rate changes. In particular, it is found that the tax semi-elasticity of MJEs' profits is approximately −1.2 per cent, that is an increase in the LBT by ten percentage points would lead, *ceteris paribus*, to a decline of local taxable profits of twelve per cent. A high sensitivity of firms to cross-jurisdictional tax rate differentials raise strong concerns about the effectiveness of a FA regime to curb the extent of profit shifting by MJEs.

The rest of the chapter is organized as follows. Section 4.2 provides a brief review of the existing literature on the extent of profit shifting activities under FA, followed by a description of the local corporate tax system in Germany in Section 4.3. Sections 4.4 and 4.5 introduce the data we use for this analysis and the empirical approach. Main results follow in Section 4.6 together with a breakdown of the response by industries in Section 4.7. Finally, Section 4.8 concludes.

4.2 Related Literature

While profit shifting under SA has been studied extensively (see Beer et al., 2020; Dharmapala, 2014; Heckemeyer and Overesch, 2017), empirical evidence on the tax-sensitivity of profits and apportionment factors in FA regimes is scarce. Most of the existing empirical studies focus on the US, where FA is used to divide MJEs' domestic profits between states. An interesting feature of the US case is that the states cannot only choose corporate income tax (CIT) rates autonomously, but also the weights they apply to the specific factors in the apportionment formula, namely payroll, assets, and sales. The implicit tax levied on the apportionment factors through the formulaic approach can thus be manipulated by US states by either changing the CIT rate or the formula weights. Results of different studies are conflicting, though. Weiner (1994) finds that manufacturing sales in a state are inversely related to the implicit tax on sales. Goolsbee and Maydew (2000) provide similar evidence regarding employment. In contrast, Klassen and Laplante (2012) report that employment and assets are insensitive to changes in the respective formula weights and/or CIT rates. Clausing (2016) analyzes the sensitivity of investment, employment, and sales with respect to changes in formula weights and/or CIT rates, but does not find a (robust) significant association. A disadvantage of these studies is that they use data aggregated at the state-level. Due to that, the dependent variables employed in the analyses capture the activities of both MJEs and single-jurisdictional enterprises (SJEs). The results therefore do not allow any conclusions to be drawn with regard to the tax sensitivity of MJEs' profits or the distribution of apportionment factors.

In contrast, Mintz and Smart (2004) draw on Canadian data allowing them to differentiate between two types of firms: MJEs operating in multiple provinces through a single corporate entity and MJEs operating through separate subsidiaries. While the former firm type is subject to FA, the profits of the latter type are taxed according to SA rules. The authors find that the profits of MJEs operating in an FA regime are considerably less sensitive to CIT rates than the profits of MJEs preparing separate accounts. However, the data Mintz and Smart (2004) use are aggregated across all firms of the same type at the province-level. One of the advantages of using firm-level data, as we do in the present paper, is that it allows taking the level of taxation in other jurisdictions hosting entities of an MJE into account. This is important since theoretical models suggest that distortions in the allocation of apportionment factors are caused by tax rate differences between host jurisdictions. While firm-level data allow assessing the sensitivity of profits and apportionment factors to such tax rate differences, studies based on aggregate data only take local tax rate levels into account.

To the best of our knowledge, the only study with a focus on Germany is Riedel (2010).[1] Despite using the same data source, the are several important differences between her paper and ours: first, Riedel (2010) assesses the tax-sensitivity of the payroll-to-capital ratio, whereas our focus is on profits and payroll.[2] Second, her data only cover two years with a time gap in between: 1998 and 2001. Consequently, her estimates of the tax sensitivity of the payroll-to-capital ratio is based on tax rate changes that potentially occurred at different points in time in between the two sample years, making it difficult to distinguish between short and medium-run effects. Moreover, a nationwide unique identifier for MJEs, their subsidiaries, and permanent establishments only became available in the local business tax statistics in 2001. Before 2001, the data only included an identifier provided by the host municipalities. As a result, entities that relocated from one municipality to another were assigned a new identifier. What is more, some municipalities even assigned an entity a new identifier in case it relocated within the municipality. Riedel's decision not to use information on entities that cannot be linked across the two sample years results in a loss of 80 per cent of observations. Thus, her sample is considerably smaller than ours (and selection into the sample may be endogenous). Third, we test whether the tax-sensitivity of profits and payroll varies across industries, which is not done by Riedel (2010).

4.3 Business Taxation in Germany

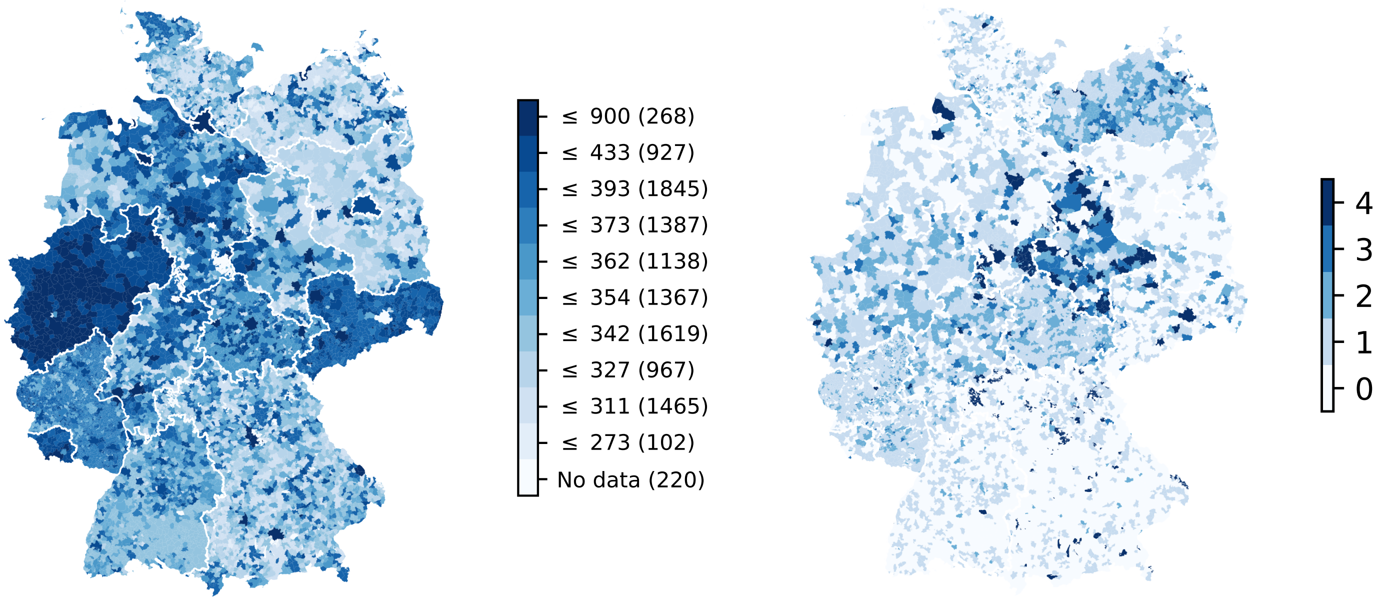
In Germany, business profits are taxed at two levels. At the national level, depending on a firm's legal status, profits are either subject to the CIT (applies to corporate firms) or the personal income tax (PIT; applies to non-corporate firms). While the CIT is characterized by a uniform rate of 15 per cent, the PIT is progressive, with an initial rate of 14 per cent and a maximum rate of 45 per cent. At the municipality-level, the profits of both corporate and non-corporate firms are subject to the local business tax (LBT).

A firm's LBT liability is the product of three factors: the tax base, the federal basic factor (*Steuermesszahl*), and the local scaling factor (*Hebesatz*). The LBT rate is thus equal to the national basic factor times the local scaling factor. Both the tax base and the basic factor are determined at the federal level. They are uniform across municipalities and cannot be changed by them. Since 2008, the basic factor is equal to 3.5 per cent (before: 5.0 per cent). In contrast, municipalities can set the local scaling factor autonomously.[3] Figure 4.1 illustrates that the scaling factor exhibits a considerable degree of variation, both across municipalities (left panel) and within municipalities over time (right panel). In 2014, the last year of our sample, it ranged from 200 per cent to 900 per cent, with an average of 354 per cent. This implies an average LBT rate of 12.39 per cent (= 3.5% × 354%) and a range from 7.0 per cent to 31.5 per cent. 56 per cent of the municipalities changed their scaling factor at least once during the sample period, and three per cent changed it in every sample year. Of all scaling factor changes covered in our sample, 97 per cent were hikes and only three per cent were cuts.

[1] Buettner et al. (2011) use data from Germany to analyze whether tax rate differences across municipalities affect MJEs' choice to integrate legally independent entities in case the costs associated with remaining disintegrated are reduced.

[2] The choice of the dependent variable in Riedel (2010) is motivated by a theoretical model indicating that FA distorts both labor and capital demand. Unfortunately, though, information on capital investment is not available in the local business tax statistics anymore.

[3] The multiplier must be at least 200 per cent.



(a) *Local scaling factors in 2014 (in %)* (b) *Scaling factor changes between 2010 and 2014*

Figure 4.1: *Local Scaling Factors in Germany*

Firms with entities in more than one municipality are subject to formula apportionment: the profits of all entities are consolidated at the national level in order to create a common tax base. Consolidated profits are then apportioned across municipalities in which the company operates based on the entities' payroll shares.[4] An entity's payroll costs include the salaries, wages, and other compensations paid to persons formally employed at the entity, but exclude compensations for apprentices and interns (§31 Local Business Tax Law/*Gewerbesteuergesetz*).

4.4 Data

The analysis is based on administrative data provided by the German tax authorities through the Statistical Offices of the States. The data comprise information from the tax returns of all German firms that are subject to the local business tax (*Gewerbesteuerstatistik*). The units of observations are municipality-firm-years, i.e., the information is aggregated across all of a firm's entities (including the headquarter, subsidiaries, and permanent establishments) located in a municipality in case of multiple affiliates in the same municipality. Our data set covers the years from 2010 to 2014 and includes approximately four million observations per year. A rich set of accounting information is available at the firm-municipality level for every year, including the tax base, i.e., the profits of the company which are apportioned to a specific municipality, the payroll costs, the main industry of the firm, its specific legal status as well as the LBT rate. Furthermore, the data set contains a unique tax identification number that is uniform across all municipalities in which the company operates and over time (*Steueridentifikationsnummer*).

[4]The law provides for exceptions to be made for producers of electricity and heat (§29 Local Business Tax Law/*Gewerbesteuergesetz*).

The tax return data have several important advantages over other, mostly commercial data sets, such as the *Bureau van Dijk's* Orbis and Amadeus data. First, Orbis and Amadeus report ownership information only for the most up to date version of the available data. This creates the possibility of misclassification of ownership structures in case of changes to the ownership, for example in case of company take-overs or mergers. Second, other available data sets typically report financial statement information rather than tax return information. This distinction is important, even though it is reasonable to assume that most countries achieve a satisfying degree of book-tax conformity. Third, and most important for our research question, the Orbis and Amadeus data sets often have a limited coverage of affiliates in countries with particularly generous tax-favoring conditions ('tax havens'), thus providing only an incomplete picture (see Tørsløv et al., 2018, on this issue). The tax return data used here provide information on *all* municipalities in Germany in which the company operates through affiliates.

Descriptive statistics of the variables included in our analysis are shown in Table 4.1. Between 2010 and 2014, the average local scaling factor across all municipalities in which entities of MJEs were located was approximately 383 per cent, which translates into an average (statutory) LBT rate of 13.4 per cent. However, there is considerable variation with respect to the local scaling factor and the resulting LBT rate as can be seen from the fact that the difference between the scaling factor's 5$^{\text{th}}$ and 95$^{\text{th}}$ percentiles is almost 200 percentage points (300 per cent vs. 475 per cent). The tax rate difference between the entities of MJE in one municipality and the average tax rate in other municipalities in which the MJE has affiliates is centered around 0 and varies between −3 and 3 percentage points. The median of the apportioned tax base and payroll costs per year and municipality are roughly 860 Euro and 0.6 million Euro, respectively. Large standard deviations in both cases as well as substantially larger sample means indicate that there are outliers at the top of the distributions, for which we account in a robustness test by trimming the bottom and the top percentile of both variables.

	5$^{\text{th}}$ Perc.	Mean	Median	95$^{\text{th}}$ Perc.	St. Dev.
Local Scaling Factor (in %)	300.00	383.33	380.00	475.00	53.19
LBT Rate (in %)	10.50	13.42	13.30	16.63	1.86
LBT Diff. to other Group Entities (in pp.)	−2.91	0.03	0.00	3.09	1.89
App. Tax Base (only pos., in Euro)	42.00	6222.87	861.00	13 933.00	132903
Payroll Costs (only pos., in Mio. Euro)	0.01	6.98×10^{8}	0.59	945.00	8.31×10^{11}

Source: Local business tax statistics.

Table 4.1: *Descriptive Statistics for Multi-Jurisdictional Entities*

4.5 Empirical Approach

To estimate the sensitivity of MJEs' profits and payroll costs with regard to LBT rates, we follow the existing profit shifting literature (see Beer et al., 2020, and Heckemeyer and Overesch, 2017) and adopt a modified version of empirical model proposed by Hines and Rice (1994):

$$y_{i,m,t} = \beta_0 + \beta_1(\tau_{m,t} - \overline{\tau}_{i,-m,t}) + \phi_{i,m} + \phi_t + \varepsilon_{i,m,t} \tag{4.1}$$

where i refers to the MJE, m to the municipality, and t to the year. The dependent variable $y_{i,m,t}$ is either the logarithm of the tax base (*Steuermessbetrag*) or the payroll costs of MJE i in municipality m and year t. The main explanatory variable of interest is $\tau_{m,t} - \overline{\tau}_{i,-m,t}$, which denotes the difference between the LBT rate of municipality m and the average LBT rate of all other municipalities in which MJE i has entities. An increase in municipality m's scaling factor would lead, *ceteris paribus*, to an increase in the LBT rate differential and, thus, increases MJE i's incentive to shift profits out of municipality m. We thus expect the coefficient estimate β_1 to be negative in case MJEs are responsive to changes in the LBT rate.

In the profit shifting literature, it is common to control for different firm characteristics as well as variables that depict the economic, demographic, and political situation in a jurisdiction. However, since the choice of covariates tends to be arbitrary, we instead add a comprehensive set of fixed-effects to our empirical model. To check the sensitivity of our results with respect to modifications to our empirical model, we estimate three specifications that differ regarding the set of fixed-effects that we include. All three specifications include MJE-municipality fixed-effects, which account for any (time-invariant) characteristics of the MJE i's entities located in municipality m. In the first specification, we add year fixed-effects, which account for changes in economic conditions, such as business cycle fluctuations, at the national level. In the second specification, we replace the year fixed-effects with year-district fixed-effects, thus allowing economic fluctuations to vary across regions. Since there are 401 districts in Germany, this specification allows us to control for changes in economic conditions at a granular regional level. In the third specification, the year fixed-effects are replaced by year-district-industry fixed-effects, thus allowing economic fluctuations to vary not only at the regional level, but also across industries within regions.

4.6 Results

The estimation results for Equation 4.1 are shown in Table 4.2. The left panel shows the estimates when using the tax base as the dependent variable, the right panel when using payroll costs.

Our results indicate a statistically and economically significant relationship between the tax rate differential and the tax base. An increase in the LBT rate differential by one percentage point is associated with a decrease in the tax base by roughly 1.2 per cent. The estimated semi-elasticity is remarkably robust across the three specifications of our empirical model. Interestingly, our estimates are somewhat larger (in absolute terms) than the average semi-elasticities reported by Beer et al. (2020) as well as Heckemeyer and Overesch (2017). Their estimates are based on meta-analyses of the literature on cross-country profit shifting under SA. Beer et al. (2020) find an average tax semi-elasticity of corporate profits of -1.0, Heckemeyer and Overesch (2017) of -0.8. Against this background, our finding may be interpreted as evidence that FA does not reduce profit shifting activities.

The estimated tax semi-elasticity of payroll costs is even larger than the semi-elasticity of the tax base. According to our estimates, a one percentage point increase in the tax rate differential is associated with a decrease in payroll costs of 3.1 per cent. The fact that the magnitude of the estimated tax semi-elasticity is around three times larger for the payroll costs than for the tax base suggests that jobs are not relocated to entities located in another municipality, in which case the two tax semi-elasticities would be of similar size. Instead, it appears that local payroll costs are reduced without an equivalent increase of payroll costs elsewhere.

The descriptive statistics reported in Table 4.1 indicate that the sample distribution of the tax base and payroll costs is heavily skewed in the sense that there are few firms with tax bases and payroll costs that are far larger than the sample means. To check the robustness of our results with regard to these outliers, we repeat the analysis based on trimmed data where we discard the smallest and largest one per cent of observations in terms of tax bases and payroll costs. However, our results remain virtually unchanged (Table B.1).

	Dep. variable: Log(Tax Base$_{i,m,t}$)			Dep. variable: Log(Payroll$_{i,m,t}$)		
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_{m,t} - \overline{\tau}_{i,-m,t}$	-1.236^{***}	-1.175^{***}	-1.108^{***}	-8.779^{***}	-3.118^{***}	-3.112^{***}
	(0.258)	(0.273)	(0.274)	(1.900)	(0.419)	(0.420)
Affiliate FEs	✓	✓	✓	✓	✓	✓
Year FEs	✓	✘	✘	✓	✘	✘
District-Year FEs	✘	✓	✘	✘	✓	✘
District-Year-Ind. FEs	✘	✘	✓	✘	✘	✓
Observations	955,277	955,275	954,055	1,356,317	1,356,315	1,355,622
$\overline{R}^2$	0.89	0.89	0.89	0.79	0.84	0.85

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table 4.2: *The Effect of Tax Rate Differentials on Tax Base and Payroll Costs of Affiliates of MJEs*

4.7 Profit Shifting Responses by Industries

Our results indicate that the local reduction of payroll costs in response to a LBT rate hike is not compensated by an increase in payroll elsewhere. Instead, the reduction in local payroll costs apparently leads to a reduction in the MJE's overall payroll costs. In principle, there are two ways to reduce payroll costs. A firm can either cut the wages of its employees or reduce the level of employment, implying that either employees are laid off or working hours are reduced.

However, it seems unlikely that our results are driven by wage cuts, since nominal wages tend to be rigid, at least in the short-run. Moreover, nominal wages are specified in the working contract and changes to the contract require the agreement of the employee. Supporting this notion, Fuest et al. (2018) study the incidence of the German LBT on real wages and find that LBT rate hikes lead to a significant real wage decline only after two years. They argue that this effect most likely comes from a slower nominal wage growth.

What about the level of employment then? In principle, employees can be laid off at short notice. For instance, the legal period of notice is one month for employees who have been employed at the company for a maximum of five years, four months for employees who have been with the company for a maximum of ten years, and seven months for employees who have been with the company for a maximum of 20 years. However, in practice, laying off employees turns out to be more difficult, as the works council can object to a dismissal.

	Percentage of Respondents ...		
Industry	Paid by the Hour	with Mini-Jobs	with Paid Overtime
Art/entertainment	10.47	18.42	5.45
Catering/lodging	32.94	29.94	10.26
Construction	46.45	5.36	15.36
Education	4.56	10.57	4.70
Finance	0.65	3.91	9.54
Health care	9.42	11.52	10.02
IT	3.35	5.40	7.89
Manufacturing	43.39	4.69	16.60
Other business services	46.52	28.73	12.39
Other services	15.94	15.89	6.52
Property	9.76	12.40	4.28
Scient./techn. services	2.28	8.74	7.16
Trade	22.00	20.76	12.58
Transport	39.99	13.15	15.11
Water and waste	38.42	4.95	14.69

Source: Own calculations based on data from *Socio-Economic Panel* (v.34).

Table 4.3: *Differences in the Prevalence of Payment by Hours, Mini-Jobs, and Paid Overtime by Industry*

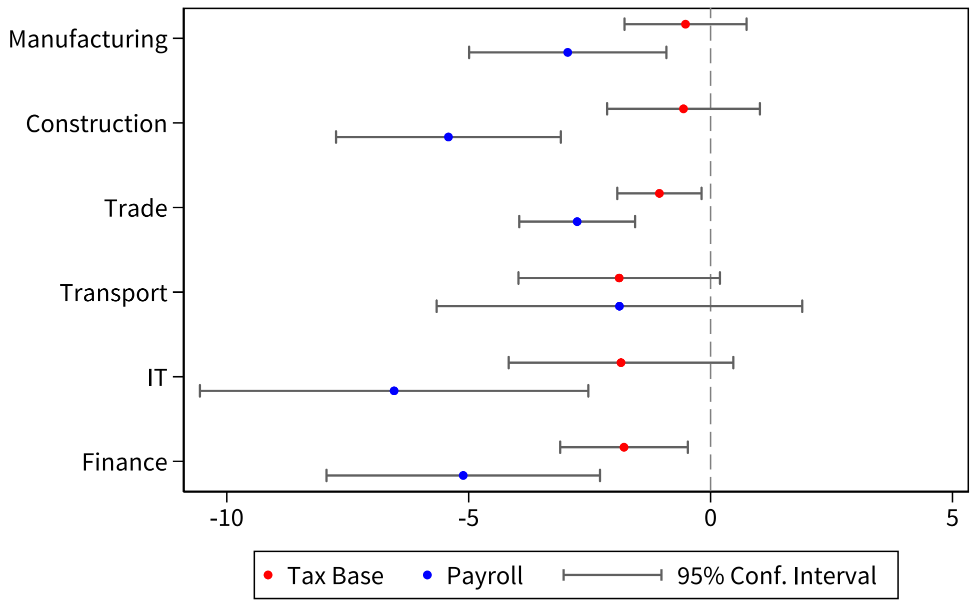
An easier way for a company to reduce the level of employment is to reduce working hours. In principle, there are two types of employees: those who receive a fixed remuneration each month (*Angestellte/Gehaltsbezieher*) and those who are paid by the hour (*Arbeiter/Lohnbezieher*). The latter group typically comprises low-skilled workers as well as so-called 'mini-jobbers', that is, persons who work part-time and earn less than 450 Euro per month (400 Euro before 2013). If the working hours of paid-by-the-hour workers and 'mini-jobbers' are reduced, payroll costs decrease as well. Moreover, the payroll costs relevant for the apportionment of the tax base under FA also include compensations for working overtime. Reducing overtime work thus represents another possible way of reducing payroll costs.

Unfortunately, our data do not include any information about the relevant components of payroll costs. However, as Table 4.3 shows, there is notable variation in the prevalence of workers paid by the hour and paid overtime work across industries. On these grounds, we estimate separate tax semi-elasticities by industries. The coefficient estimates are illustrated graphically in Figure 4.2.[5]

The results suggest that there are substantial differences between industries regarding the tax-sensitivity of the tax base and payroll costs. Turning to the tax base, we obtain particularly large estimates in the 'Transport', 'Finance', and 'IT' sector (although the latter effect is insignificant). In line with our conjecture, the 'Transport' sector is characterized by a relatively high share of paid-by-the-hour and paid overtime workers. This is not true, though, for the 'Finance' and 'IT' sector. In contrast the estimated tax semi-elasticities for 'Manufacturing', 'Scientific and Technical Services', and 'Other Business Services' are smaller than the baseline estimate and insignificant. However, at least in 'Manufacturing'

[5]Note that the Figure only includes the results for industries with at least 35,000 observations. The results for all industries can be found in Figure B.1 in the Appendix.

and ‘Other Business Services’, the share of paid-by-the-hour workers is particularly large. All in all, our findings are thus inconclusive regarding the potential channels through which local payroll costs are reduced. Further research is needed in this context.



Note: The point estimates for each industry are obtained from the same specification as Table 4.2, columns (2) and (5).

Figure 4.2: *The Extent of Tax Base and Payroll Cost Responses to Tax Rate Differentials for the Largest Industries*

4.8 Conclusion

The international system of corporate taxation is often considered to be a relic of the ‘brick and mortar’ economy and not to be fit for our modern, highly globalized, and digitalized world, as it does not ensure that profits are taxed where value is created. In light of this criticism, many policy-makers and economic scholars advocate for a change from the current system of separate accounting (SA) to a formula apportionment (FA) regime. The general hope is that under FA, the international distribution of profits may better reflect the distribution of economic activity instead of being a result of multi-jurisdictional enterprises (MJEs) aiming at minimizing their tax burden.

However, one aspect that is often ignored is that under FA, the existence of tax differentials across jurisdictions may lead to a distortion in the distribution of apportionment factors. That is, under FA, MJEs have an incentive to relocate economic activities to low-tax jurisdictions, whereas under SA, profits are merely shifted on paper.

Our results indicate that the introduction of FA is not an effective measure to reduce profit shifting by MJEs. Utilizing data from Germany, where FA is applied at the local level and apportionment is based on relative payroll shares, we obtain an estimate for the tax semi-elasticity of MJEs’ profits of −1.2 per cent, indicating that MJEs are very sensitive to tax rate differentials under FA. Moreover, we find that the distortions in the distribution of apportionment factors under FA can be large: the corresponding tax semi-elasticity of payroll costs is -3.1 per cent.

Appendices

Appendix A

Details: Chapter 1

A.1 Details about National Fiscal Rules of Treated Countries

Country	Year of Adoption	Num. Target (% of GDP)	Coverage of FR	Included in baseline sample
Antigua and Barbuda	1998	−3	GG	no, < 1 million population in 2018
Argentina	2000	0	GG	Yes
Austria	1995	−3	GG	Yes
Belgium	1992	−3	GG	Yes
Benin	2000	−3	CG	Yes
Bulgaria	2006	−2	GG	Yes
Burkina Faso	2000	−3	CG	Yes
Cabo Verde	1998	−3	CG	no, < 1 million population in 2018
Cameroon	2002	0	CG	no, missing balance data
Canada	1998	0	CG	Yes
Central African Republic	2002	0	CG	Yes
Chad	2002	0	CG	no, missing balance data
Congo, Republic of	2002	0	CG	Yes
Croatia	2013	−3	GG	Yes
Cyprus	2004	−3	GG	no, < 1 million population in 2018
Czech Republic	2004	−3	GG	Yes
Côte d'Ivoire	2000	−3	CG	Yes
Denmark	1992	−3	GG	Yes
Dominica	1998	−3	GG	no, < 1 million population in 2018
Equatorial Guinea	2002	0	CG	Yes
Estonia	1993	−3	GG	no, missing balance data
Finland	1995	−3	GG	Yes
France	1992	−3	GG	Yes
Gabon	2002	0	CG	Yes
Georgia	2014	−3	CG	Yes
Germany	1992	−3	GG	Yes
Greece	1992	−3	GG	Yes
Grenada	1998	−3	GG	no, < 1 million population in 2018
Guinea-Bissau	2000	−3	CG	Yes
Hong Kong SAR	1997	0	GG	Yes
Hungary	2004	−3	GG	Yes
India	2008	−3	GG	no, FR was only for 1 year in place
Indonesia	1980	−3	GG	no, FR already in place since 1967
Ireland	1992	−3	GG	Yes
Israel	2007	varying	CG	Yes
Italy	1992	−3	GG	Yes
Kosovo	2013	2	GG	Yes
Latvia	2004	−3	GG	Yes
Lithuania	2004	−3	GG	Yes
Luxembourg	1992	−3	GG	no, < 1 million population in 2018
Mali	2000	−3	CG	no, lack of fiscal balance data
Malta	2004	−3	GG	no, < 1 million population in 2018
Mexico	2006	0	CG	Yes
Montenegro, Rep. of	2014	−3	CG	no, < 1 million population in 2018
Netherlands	1992	−3	GG	Yes
Niger	2000	−3	CG	Yes
Nigeria	2007	−3	CG	Yes
Pakistan	2008	0	CG	Yes
Paraguay	2015	−1.5	CG	Yes
Peru	2000	-2/-1.5	CG	Yes
Poland	2004	−3	GG	Yes
Portugal	1992	−3	GG	Yes
Romania	2007	−3	GG	Yes
Senegal	2000	−3	CG	Yes
Singapore	1980	0	CG	no, FR already in place since 1962
Slovak Republic	2004	−3	GG	Yes
Slovenia	2004	−3	GG	Yes
Spain	1992	−3	GG	Yes
St. Kitts and Nevis	1998	−3	GG	no, < 1 million population in 2018
St. Lucia	1998	−3	GG	no, < 1 million population in 2018
St. Vincent and the Grenadines	1998	−3	GG	no, < 1 million population in 2018
Sweden	1995	1	GG	Yes
Togo	2000	−3	CG	Yes
United Kingdom	1992	−3	GG	Yes

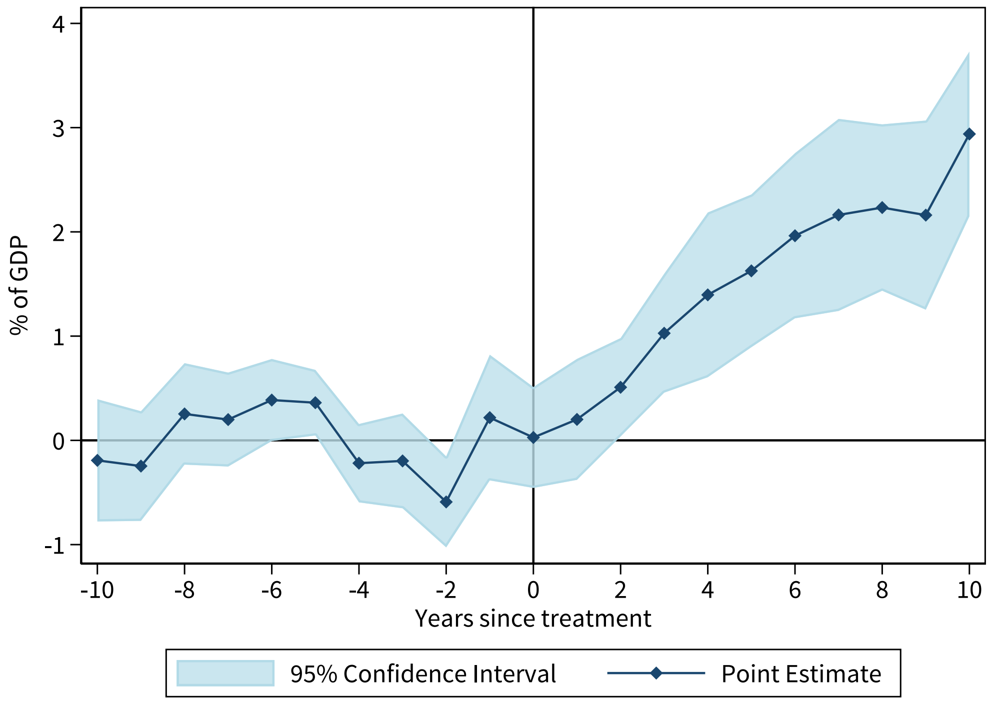
Source: IMF Fiscal Rules Database.

Table A.1: *Countries included in the Sample and their National Fiscal Rules*

A.2 Matrix of Weights from Synthetic Control Approach and Additional Results

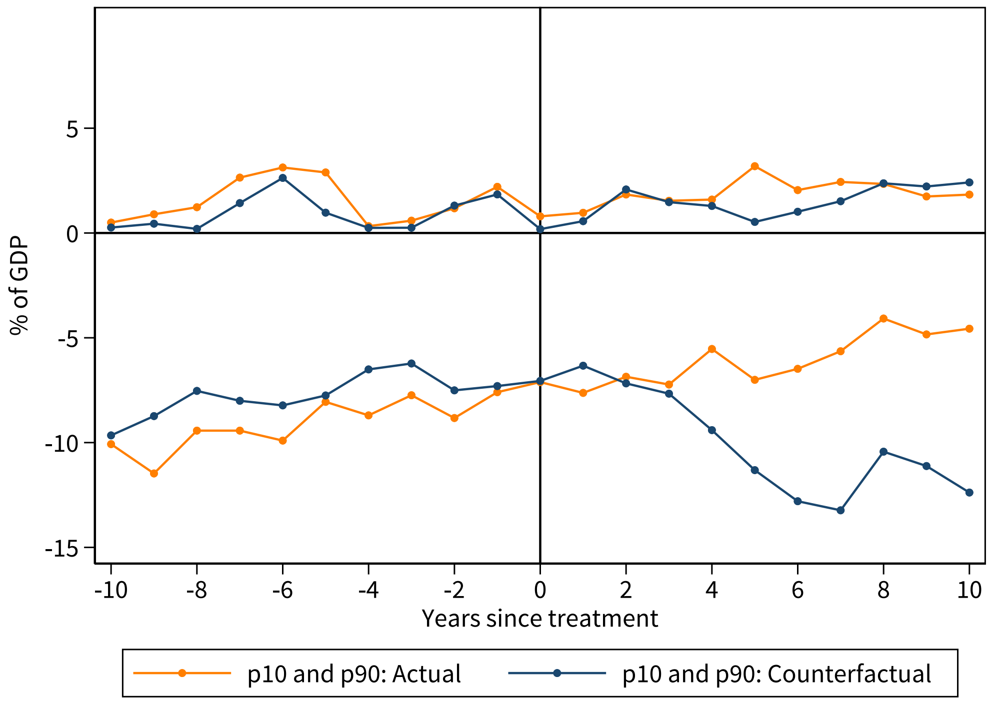
Note: The matrix displays countries with fiscal rules on the horizontal axis and donor countries on the vertical axis. Assigned weights obtained from the baseline SC specification are shown in the corresponding cells of the matrix with larger weights in darker red. Cells in white indicate zero weights.

Figure A.1: *Matrix of Donor Weights for Baseline SC Specification*



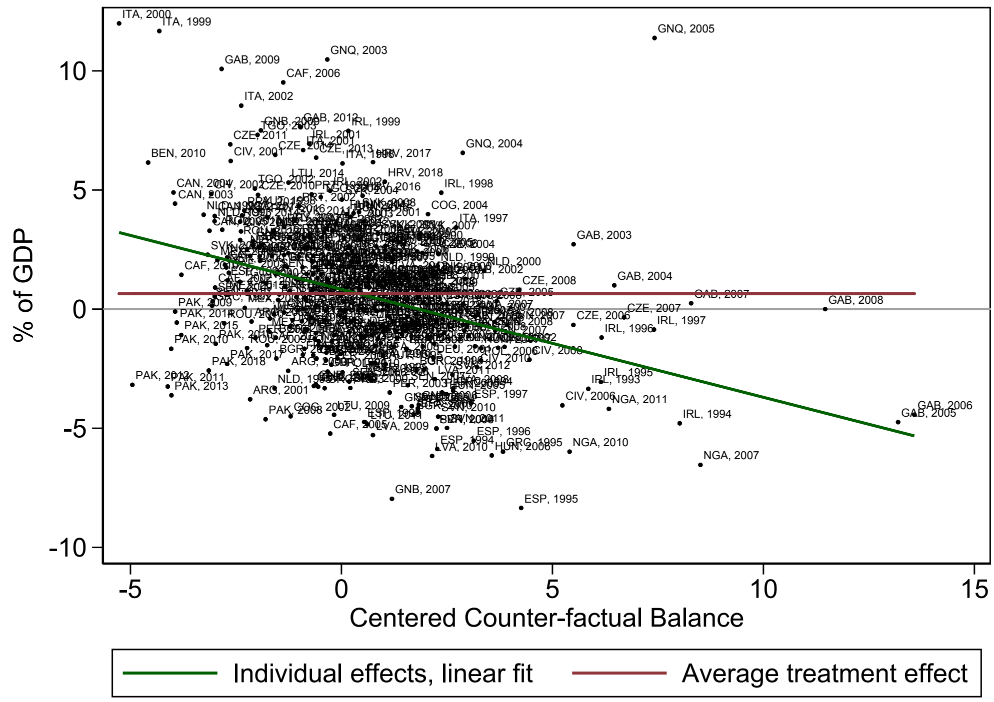
Note: The plot displays the average treatment effect as obtained from a linear combination of country-specific treatment effects from the baseline SC specification and Equ. 1.3. The fiscal rule is introduced in year 0.

Figure A.2: *Average Treatment Effect of the Fiscal Rule on Governments' Balances*



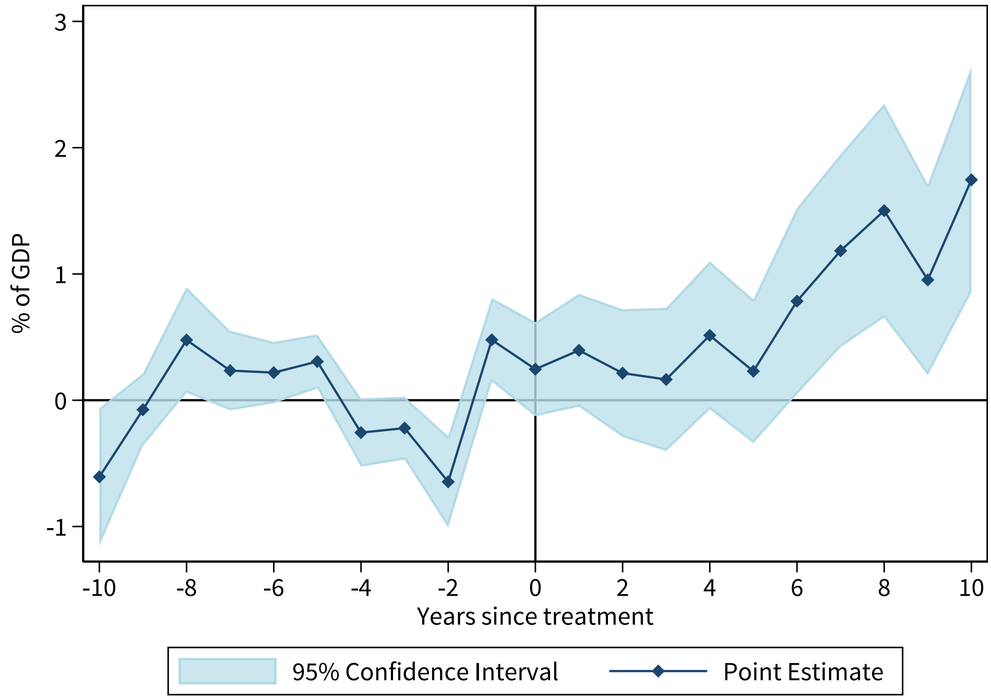
Note: The time series depict the 10^{th} and the 90^{th} percentiles of the actual budget balance series and the counter-factual balance series as obtained from the baseline SC specification. The fiscal rule is introduced in year 0.

Figure A.3: *Time Paths of the 10^{th} and the 90^{th} Percentiles of the Treated and Counter-factual Balances*



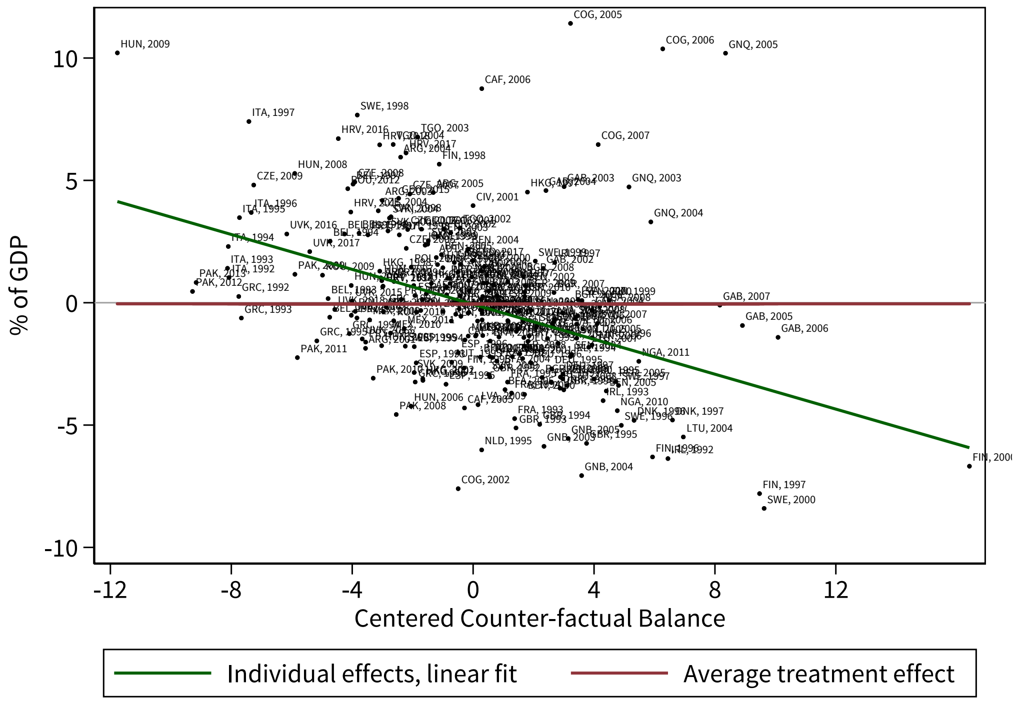
Note: The scatter plot depicts the individual treatment effects by level of counter-factual budget balance as obtained from a SC estimation based on economic covariates instead of lagged outcome variables of the budget balance. The counter-factual balance on the horizontal axis is centered around the numerical value of the fiscal rule in each country.

Figure A.4: *Individual Treatment Effects by Level of Counter-factual Government Balance from Estimation based on Covariates*



Note: The figure displays the average treatment effect as obtained from a linear combination of country-specific treatment effects from the SC specification based on covariates and Equ. 1.3. The fiscal rule is introduced in year 0.

Figure A.5: *Average Treatment Effects of Fiscal Rule Introduction from Estimation based on Covariates*



Note: The figure depicts the individual treatment effects by level of the counter-factual budget balance as obtained from a SC specification with only five years of pre- and post-treatment windows. The income filter according to which only countries in the same or adjacent income groups are matched with each other is kept.

Figure A.6: *Individual Treatment Effects by Level of Counter-factual Government Balance with Shortened Pre- & Post-Treatment Horizon*

Appendix B

Details: Chapter 4

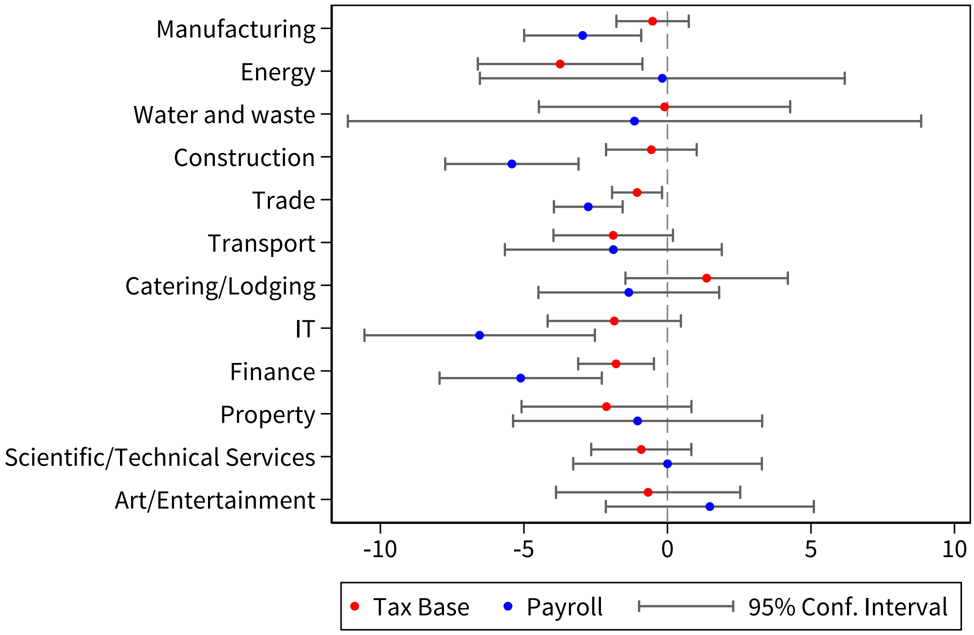
B.1 Robustness Check for Trimmed Data

	Dep. variable: Log(Tax Base$_{i,m,t}$)			Dep. variable: Log(Payroll$_{i,m,t}$)		
	(1)	(2)	(3)	(4)	(5)	(6)
$\tau_{m,t} - \overline{\tau}_{i,-m,t}$	-1.132^{***} (0.266)	-1.058^{***} (0.273)	-0.998^{***} (0.274)	-8.659^{***} (1.832)	-3.150^{***} (0.426)	-3.160^{***} (0.425)
Affiliate FEs	✓	✓	✓	✓	✓	✓
Year FEs	✓	✘	✘	✓	✘	✘
District-Year FEs	✘	✓	✘	✘	✓	✘
District-Year-Ind. FEs	✘	✘	✓	✘	✘	✓
Observations	894,212	894,212	892,960	1,335,731	1,335,731	1,335,022
$\overline{R}^2$	0.84	0.84	0.85	0.79	0.84	0.84

Note: Results are based on OLS estimation. We use heteroskedasticity robust standard errors which are displayed in brackets below. The coefficient of the constant is omitted to ease readability.
$^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$

Table B.1: *The Effect of Tax Rate Differentials on Tax Base and Payroll Costs of Affiliates of MJEs Based on Trimmed Data*

B.2 Industry-Specific Results



Note: The point estimates for each industry are obtained from the same specification as Table 4.2, columns (2) and (5).

Figure B.1: *The Extent of Tax Base and Payroll Cost Responses to Tax Rate Differentials for all Industries*

Bibliography

Abadie, A., Diamond, A., and Hainmueller, J. Synthetic Control Methods for Comparative Case Studies: Estimating the Effect of California's Tobacco Control Program. *Journal of the American Statistical Association*, 105(490):493–505, 2010.

Abbring, J. H. and Heckman, J. J. Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation. *Handbook of Econometrics*, 6:5145–5303, 2007.

Adhikari, B., Duval, R., Hu, B., and Loungani, P. Can Reform Waves Turn the Tide? Some Case Studies Using the Synthetic Control Method. *Open Economies Review*, 29(4):879–910, 2018.

Anderson, W., Wallace, M. S., and Warner, J. T. Government Spending and Taxation: What Causes What? *Southern Economic Journal*, 52(3):630–639, 1986.

Andrabi, T., Das, J., and Khwaja, A. I. Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets. *American Economic Review*, 107(6):1535–1563, 2017.

Arkhangelsky, D., Athey, S., Hirshberg, D. A., Imbens, G. W., and Wager, S. Synthetic Difference in Differences. Technical report, National Bureau of Economic Research, 2019.

Arnold, R. D. and Carnes, N. Holding Mayors Accountable: New York's Executives from Koch to Bloomberg. *American Journal of Political Science*, 56(4):949–963, 2012.

Badinger, H. and Reuter, W. H. The Case for Fiscal Rules. *Economic Modelling*, 60:334–343, 2017.

Balaguer-Coll, M. T., Brun-Martos, M. I., Forte, A., and Tortosa-Ausina, E. Local Governments' Re-election and its Determinants: New Evidence Based on a Bayesian Approach. *European Journal of Political Economy*, 39(C):94–108, 2015.

Banerjee, A. V., Cole, S., Duflo, E., and Linden, L. Remedying Education: Evidence from Two Randomized Experiments in India. *The Quarterly Journal of Economics*, 122(3):1235–1264, 2007.

Banerjee, A. V., Banerji, R., Duflo, E., Glennerster, R., and Khemani, S. Pitfalls of Participatory Programs: Evidence from a Randomized Evaluation in Education in India. *American Economic Journal: Economic Policy*, 2(1):1–30, 2010.

Barr, A., Packard, T., and Serra, D. Participatory Accountability and Collective Action: Experimental Evidence from Albania. *European Economic Review*, 68:250–269, 2014.

Bartle, J. Partisanship, Performance and Personality: Competing and Complementary Characterizations of the 2001 British General Election. *Party Politics*, 9(3):317–345, 2003.

Bedoya, G., Bittarello, L., Davis, J., and Mittag, N. *Distributional Impact Analysis: Toolkit and Illustrations of Impacts Beyond the Average Treatment Effect*. The World Bank, 2017.

Beer, S., De Mooij, R., and Liu, L. International Corporate Tax Avoidance: A Review of the Channels, Magnitudes, and Blind Spots. *Journal of Economic Surveys*, 34(3):660–688, 2020.

Berry, C. R. and Howell, W. G. Accountability and Local Elections: Rethinking Retrospective Voting. *The Journal of Politics*, 69(3):844–858, 2007.

Blackley, P. R. Causality Between Revenues and Expenditures and the Size of the Federal Budget. *Public Finance Quarterly*, 14(2):139–156, 1986.

Blair, G., Littman, R., and Paluck, E. L. Inciting Action Against Corruption in Nigeria. Unpublished Manuscript.

Boeing, G. Clustering to Reduce Spatial Data Set Size. 2018.

Bonneau, C. W. and Cann, D. M. Party Identification and Vote Choice in Partisan and Nonpartisan Elections. *Political Behavior*, 37(1):43–66, 2015.

Boyne, G. A., James, O., John, P., and Petrovsky, N. Democracy and Government Performance: Holding Incumbents Accountable in English Local Governments. *Journal of Politics*, 71(4):1273–1284, 2009.

Buettner, T., Riedel, N., and Runkel, M. Strategic Consolidation under Formula Apportionment. *National Tax Journal*, 64(2):225–257, 2011.

Burnett, C. M. and Kogan, V. The Politics of Potholes: Service Quality and Retrospective Voting in Local Elections. *The Journal of Politics*, 79(1):302–314, 2017.

Campos, N. F., Coricelli, F., and Moretti, L. Institutional Integration and Economic Growth in Europe. *Journal of Monetary Economics*, 103:88–104, 2019.

Carey, J. M., Niemi, R. G., and Powell, L. W. *Term Limits in State Legislatures*. University of Michigan Press, 2009.

Caselli, F. and Reynaud, J. Do Fiscal Rules Cause Better Fiscal Balances? A New Instrumental Variable Strategy. *European Journal of Political Economy*, 2020.

Caselli, F. G. and Wingender, P. *Bunching at 3 Percent: The Maastricht Fiscal Criterion and Government Deficits*. International Monetary Fund, 2018.

Clausing, K. The Effect of Profit Shifting on the Corporate Tax Base in the United States and Beyond. *National Tax Journal*, 69(4):905–934, 2016.

Clausing, K. A. and Lahav, Y. Corporate Tax Payments under Formulary Apportionment: Evidence from the Financial Reports of 50 Major US multinational Firms. *Journal of International Accounting, Auditing and Taxation*, 20(2):97–105, 2011.

Clinton, J., Jackman, S., and Rivers, D. The Statistical Analysis of Roll Call Data. *American Political Science Review*, 98(2):355–370, 2004.

Cloyne, J. Discretionary Tax Changes and the Macroeconomy: New Narrative Evidence from the United Kingdom. *American Economic Review*, 103(4):1507–1528, 2013.

Cobham, A. and Loretz, S. International Distribution of the Corporate Tax Base: Implications of Different Apportionment Factors under Unitary Taxation. 2014.

Courant, P., Gramlich, E., and Rubinfeld, D. The Stimulative Effects of Intergovernmental Grants or Why Money Sticks Where It Hits. In Mieszkowski, P. and Oakland, W., editors, *Fiscal Federalism and Grants-in-aid*. 1979.

Debrun, X., Moulin, L., Turrini, A., Ayuso-i Casals, J., and Kumar, M. S. Tied to the Mast? National Fiscal Rules in the European Union. *Economic Policy*, 23(54):298–362, 2008.

Devereux, M. P. and Loretz, S. The Effects of EU Formula Apportionment on Corporate Tax Revenues. *Fiscal Studies*, 29(1):1–33, 2008.

Dharmapala, D. What Do We Know about Base Erosion and Profit Shifting? A Review of the Empirical Literature. *Fiscal Studies*, 35(4):421–448, 2014.

Dipoppa, G. and Grossman, G. The Effect of Election Proximity on Government Responsiveness and Citizens' Participation: Evidence From English Local Elections. *Comparative Political Studies*, 53(14): 2183–2212, 2020.

Doudchenko, N. and Imbens, G. W. Balancing, Regression, Difference-in-Differences and Synthetic Control Methods: A Synthesis. Technical report, National Bureau of Economic Research, 2016.

Dubin, R. A. Spatial Autocorrelation: A Primer. *Journal of Housing Economics*, 7(4):304–327, 1998.

Ester, M., Kriegel, H.-P., Sander, J., and Xu, X. A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise. In *Proceedings of 2nd International Conference on Knowledge Discovery and Data Mining*, 1996.

European Commission. *Towards an Internal Market Without Tax Obstacles: A Strategy for Providing Companies with a Consolidated Corporate Tax Base for Their EU-wide Activities*. 2001.

European Commission. *Proposal for a Council Directive on a Common Consolidated Corporate Tax Base (CCCTB)*. 2011.

European Commission. *Action Plan for a Fair and Efficient Corporate Tax System in the EU*. 2015.

Ferman, B., Pinto, C., and Possebom, V. Cherry Picking With Synthetic Controls. *Journal of Policy Analysis and Management*, 39(2):510–532, 2020.

Fuest, C., Hemmelgarn, T., and Ramb, F. How Would the Introduction of an EU-wide Formula Apportionment Affect the Distribution and Size of the Corporate Tax Base? An Analysis Based on German Multinationals. *International Tax and Public Finance*, 14(5):605–626, 2007.

Fuest, C., Peichl, A., and Siegloch, S. Do Higher Corporate Taxes Reduce Wages? Micro Evidence from Germany. *American Economic Review*, 108(2):393–418, 2018.

Goolsbee, A. and Maydew, E. L. Coveting thy Neighbor's Manufacturing: The Dilemma of State Income Apportionment. *Journal of Public Economics*, 75(1):125–143, 2000.

Gordon, R. and Wilson, J. D. An Examination of Multijurisdictional Corporate Income Taxation under Formula Apportionment. *Econometrica*, 54(6):1357–1373, 1986.

Griliches, Z. and Hausman, J. A. Errors in Variables in Panel Data. *Journal of Econometrics*, 31(1):93–118, 1986.

Grossman, G. and Michelitch, K. Information Dissemination, Competitive Pressure, and Politician Performance Between Elections: A Field Experiment in Uganda. *American Political Science Review*, 112 (2):280–301, 2018.

Grossman, G., Michelitch, K., and Santamaria, M. Texting Complaints to Politicians: Name Personalization and Politicians' Encouragement in Citizen Mobilization. *Comparative Political Studies*, 50(10): 1325–1357, 2017.

Gruss, B. and Kebhaj, S. *Commodity Terms of Trade: A New Database.* International Monetary Fund, 2019.

Hayo, B. and Uhl, M. The Macroeconomic Effects of Legislated Tax Changes in Germany. *Oxford Economic Papers*, 66(2):397–418, 2014.

Heckemeyer, J. H. and Overesch, M. Multinationals' Profit Response to Tax Differentials: Effect Size and Shifting Channels. *Canadian Journal of Economics*, 50(4):965–994, 2017.

Heckman, J. J., Smith, J., and Clements, N. Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts. *The Review of Economic Studies*, 64(4):487–535, 1997.

Heinemann, F., Moessinger, M.-D., and Yeter, M. Do Fiscal Rules Constrain Fiscal Policy? A Meta-Regression Analysis. *European Journal of Political Economy*, 51:69–92, 2018.

Hines, J. and Rice, E. M. Fiscal Paradise: Foreign Tax Havens and American Business. *The Quarterly Journal of Economics*, 109(1):149–182, 1994.

Hines, J. R. and Thaler, R. H. The Flypaper Effect. *Journal of Economic Perspectives*, 9(4):217–226, 1995.

International Monetary Fund. *Second-Generation Fiscal Rules: Balancing Simplicity, Flexibility, and Enforceability.* 2018.

Johnston, R. and Pattie, C. Dimensions of Retrospective Voting: Economic Performance, Public Service Standards and Conservative Party Support at the 1997 British General Election. *Party Politics*, 7(4): 469–490, 2001.

Kaul, A., Klößner, S., Pfeifer, G., and Schieler, M. Synthetic Control Methods: Never Use All Pre-Intervention Outcomes Together With Covariates. 2015.

Klassen, K. J. and Laplante, S. K. Are U.S. Multinational Corporations Becoming More Aggressive Income Shifters? *Journal of Accounting Research*, 50(5):1245–1285, 2012.

Kneller, R., Bleaney, M. F., and Gemmell, N. Fiscal Policy and Growth: Evidence from OECD Countries. *Journal of Public Economics*, 74(2):171–190, 1999.

Krebs, T. B. The Determinants of Candidates' Vote Share and the Advantages of Incumbency in City Council Elections. *American Journal of Political Science*, 42(3):921–935, 1998.

Litschig, S. and Morrison, K. M. Government Spending and Re-election: Quasi-Experimental Evidence from Brazilian Municipalities. Working Papers 515, Barcelona Graduate School of Economics, 2012.

Manage, N. and Marlow, M. L. The Causal Relation between Federal Expenditures and Receipts. *Southern Economic Journal*, 52(3):617–629, 1986.

Mauro, P., Romeu, R., Binder, A., and Zaman, A. A Modern History of Fiscal Prudence and Profligacy. *Journal of Monetary Economics*, 76:55–70, 2015.

McLure Jr, C. E. The Elusive Incidence of the Corporate Income Tax: The State Case. *Public Finance Quarterly*, 9(4):395–413, 1981.

Mintz, J. and Smart, M. Income Shifting, Investment, and Tax Competition: Theory and Evidence from Provincial Taxation in Canada. *Journal of Public Economics*, 88(6):1149–1168, 2004.

OECD. *Tax Challenges Arising from Digitalisation – Interim Report 2018*. 2018.

Padovano, F. and Venturi, L. Wars of Attrition in Italian Government Coalitions and Fiscal Performance: 1948-1994. *Public Choice*, 109(1-2):15–54, 2001.

Payne, J. E. A Survey of the International Empirical Evidence on the Tax-Spend Debate. *Public Finance Review*, 31(3):302–324, 2003.

Persson, T. and Tabellini, G. Political Economics and Macroeconomic Policy. In Taylor, J. B. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1 of *Handbook of Macroeconomics*, chapter 22, pages 1397–1482. Elsevier, 1999.

Popkin, S. L. *The Reasoning Voter: Communication and Persuasion in Presidential Campaigns*. University of Chicago Press, 1994.

Poterba, J. M. State Responses to Fiscal Crises: The Effects of Budgetary Institutions and Politics. *Journal of Political Economy*, 102(4):799–821, 1994.

Poterba, J. M. Do Budget Rules Work? *Fiscal Policy: Lessons from Economic Research*, 1997.

Poterba, J. M. and von Hagen, J. Introduction to 'Fiscal Institutions and Fiscal Performance'. In *Fiscal Institutions and Fiscal Performance*. National Bureau of Economic Research, 1999.

Primo, D. M. *Rules and Restraint: Government Spending and the Design of Institutions*. University of Chicago Press, 2007.

Reutter, W. Struktur und Dauer der Gesetzgebungsverfahren des Bundes. *Zeitschrift für Parlamentsfragen*, 38(2):299–315, 2007.

Riedel, N. The Downside of Formula Apportionment: Evidence on Factor Demand Distortions. *International Tax and Public Finance*, 17(3):236–258, 2010.

Romer, C. D. and Romer, D. H. Do Tax Cuts Starve the Beast? The Effect of Tax Changes on Government Spending. *Brookings Papers on Economic Activity*, 40(1):139–214, 2009.

Romer, C. D. and Romer, D. H. The Macroeconomic Effects of Tax Changes: Estimates Based on a New Measure of Fiscal Shocks. *American Economic Review*, 100(3):763–801, 2010.

Shackelford, D. and Slemrod, J. The Revenue Consequences of using Formula Apportionment to Calculate US and Foreign-Source Income: A Firm-level Analysis. *International Tax and Public Finance*, 5(1): 41–59, 1998.

Solymosi, R., Bowers, K. J., and Fujiyama, T. Crowdsourcing Subjective Perceptions of Neighbourhood Disorder: Interpreting Bias in Open Data. *British Journal of Criminology*, 58(4):944–967, 2017.

Stone, M. Literature Review on Complaints Management. *Journal of Database Marketing & Customer Strategy Management*, 18(2):108–122, 2011.

Tørsløv, T. R., Wier, L. S., and Zucman, G. The Missing Profits of Nations. Working Paper 24701, National Bureau of Economic Research, 2018.

von Furstenberg, G. M., Green, R. J., and Jeong, J.-H. Tax and Spend, or Spend and Tax? *Review of Economics and Statistics*, 68(2):179–188, 1986.

Weiner, J. M. Company Taxation for the European Community. *PhD Dissertation, Harvard University, Cambridge*, 1994.