Essays in Empirical Public Economics

Inaugural-Dissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)

an der
Ludwig-Maximilians-Universität München

2020

vorgelegt von

Carla Krolage

Essays in Empirical Public Economics

Inaugural-Dissertation

zur Erlangung des Grades
Doctor oeconomiae publicae (Dr. oec. publ.)

an der
Ludwig-Maximilians-Universität München

2020

vorgelegt von

Carla Krolage

Referent: Prof. Dr. Andreas Peichl

Korreferent: Prof. Dr. Clemens Fuest

Promotionsabschlussberatung: 22. Juli 2020

Acknowledgments

This doctoral thesis would not have been possible without the support of many colleagues, friends and family.

First and foremost, I would like to express my deepest gratitude to my supervisor Andreas Peichl for his valuable advice and encouragement during my time at ifo and ZEW. I would also like to thank my co-supervisor Clemens Fuest for his insightful comments and the many joint research and policy projects. I further thank Dominik Sachs for completing my dissertation committee.

Working at the ifo Center for Macroeconomics and Surveys as well as at ZEW's research group on International Distribution and Redistribution provided me with many opportunities that I am deeply grateful for. Not only did both institutes provide an inspiring research environment, they also enabled me to conduct many policy consulting projects, encouraged media appearances, and supported presentations of my work at international conferences.

While writing this dissertation, I benefited greatly from inspiring discussions with my co-authors Mathias Dolls, Florian Neumeier and Daniel Waldenström. I am also indebted to advisers, colleagues and friends at ifo, ZEW and the Universities of Mannheim and Munich, in particular Alina Bartscher, Maximilian Blömer, Florian Buhlmann, Marcell Göttert, Robert Lehmann, Sebastian Link, Tim Obermeier, Valentin Reich, Holger Stichnoth, Paul Schüle, Martin Ungerer, Johannes Voget, Daniel Weishaar, Sebastian Wichert, Christian Wittneben and Timo Wollmershäuser, as well as many others.

Last but not least, I thank my family and friends for their invaluable support throughout my life. Special thanks go to Julian Höhl and to my parents Gisela Krolage-Lung and Josef Krolage. Thank you for always caring, believing in me, and supporting me.

Contents

List of Figures

List of Tables

Introduction

This thesis consists of four essays that analyze the impact of public policy on issues at the heart of the current policy debate: Retirement insurance reform, the real estate market, and income inequality. The first three chapters empirically assess the impact of retirement and housing market reforms in Germany, using large-scale micro-datasets. These essays do not only thoroughly analyze behavioral responses and price effects and contribute novel insights to the literature, they also provide valuable insights for future policy-making in Germany and beyond. The fourth essay has a more methodological focus, addressing the measurement of top income shares. Here, calculations are applied to US data.

The first chapter *The Effects of Early Retirement Incentives on Retirement Decisions*, which is joint work with Mathias Dolls, analyzes behavioral responses to a major reform in the German public pension system. This 2014 reform allowed individuals with a long contribution history to retire without deductions before reaching the regular retirement age. Using administrative data from public pension insurance accounts, we first conduct an event study, which indicates that the probability of retiring immediately increases by approximately 10 percentage points upon becoming eligible. These large effects persist even if we limit the sample to individuals who do not become eligible at ages that are commonly framed as reference ages. Second, we employ a coarsened exact matching procedure to compare retirement entry decisions of eligible and non-eligible individuals. Our results show that individuals who are eligible for the early retirement scheme retire on average 6.4 months earlier than non-eligible individuals with identical characteristics. This translates into a decrease in the average retirement age by 0.74 months in response to a one percentage point reduction in deductions, a fairly large effect compared to the literature. With additional pension insurance expenditures of 10.4 billion euros and aggregate fiscal costs of 19.8 billion euros in the years 2014 to 2017, our subsequent fiscal cost projections are at the upper end of the range of previous back-of-the-envelope estimates.

The second chapter *Who Bears the Burden of Real Estate Transfer Taxes? Evidence from the German Housing Market* focuses on taxation of the real estate sector. This essay is joint work with Mathias Dolls, Clemens Fuest and Florian Neumeier. Using a rich micro dataset on German property prices covering the period from 2005 to 2018, we examine the effects of real estate transfer taxes (RETT) on house prices. We exploit a 2006 constitutional reform that allowed states to set their own RETT rates, leading to frequent increases in states' tax rates in subsequent years. Our monthly event study estimates indicate a price response that strongly exceeds the change in the tax burden for single transactions. I.e., twelve months after a reform, a one percentage point increase in the tax rate reduces property prices by on average 3%. Effects are stronger for apartments and apartment buildings than for single-family houses, and are robust to many specifications. We interpret these results in the context of a theoretical model that accounts for the effects of RETT on a property's resale value. If a property is expected to be traded more frequently in the future, the decline in its price can exceed the increase in the tax burden. Moreover, larger price effects can be explained by higher bargaining power of sellers, and by transaction taxes limiting which houses downpayment-constrained households are able to afford.

The third chapter *The Effect of Real Estate Purchase Subsidies on Property Prices* also focuses on government interventions in the housing market, assessing to which degree direct housing purchase subsidies are capitalized into property prices. Using a large-scale micro dataset on German property prices, I exploit that in 2018, the state of Bavaria introduced a much more generous subsidy scheme than other German states. My difference-in-difference estimations at the Bavarian interstate border indicate that following the reform, the prices of single-family homes increased by about 3.4% more in Bavarian border regions than in neighboring states. This is consistent with a full capitalization of the subsidy. No effect is found for apartments, whose purchasers seldom qualify for the subsidy. A heterogeneity analysis confirms that the price effect is larger in segments of the real estate market with a higher exposure to the subsidy scheme. I also provide suggestive evidence that the subsidy scheme slightly stimulated construction activity. Overall, my results indicate that instead of making house purchases more affordable for families, the subsidy scheme led to a rise in house prices and mainly benefited sellers of properties.

The fourth chapter *Richer or More Numerous or Both? The Role of Population and Economic Growth for Top Income Shares*, which is joint work with Andreas Peichl and Daniel Waldenström, makes a more methodological contribution to the inequality literature. The most commonly used measure for top income inequality is the income share of a fixed percentile of the population. While meeting many of the desired distributional criteria, this measure is sensitive to developments in the size of the underlying population or in the real economy. However, when measuring inequality over the long-run, accounting for population and productivity growth is important. This essay presents three alternative measures of top income shares that explicitly account for population and income growth. While the standard top share is a relative inequality measure, our three alternative measures take either an absolute or an intermediate point of view on inequality. We apply these measures to long-term income data from the United States and find that the U-shaped inequality trend over the past century holds up, but with important qualifications. For instance, we find more accentuated top income share growth since 1980 when allowing for variation in the population share of high-income individuals. Altogether, our findings suggest using several complementary top share measures when assessing long-term income inequality trends.

Appendices can be found at the end of the respective chapter. A consolidated bibliography is included at the end of this thesis.

1 The Effect of Early Retirement Incentives on Retirement Decisions

1.1 Introduction

Pension systems around the world face aging populations and demographic change, putting increased pressure on fiscal sustainability. Against this background, many countries have conducted pension reforms aimed at extending the working lives of the elderly population (OECD, 2017). These reforms encompass increases in the normal or the early retirement ages, tightening qualifying conditions and the introduction of actuarial deductions for early retirement. While Germany also increased the normal retirement age and closed specific pathways to retirement, a major reform in 2014, quite to the contrary, sharply increased early retirement incentives for individuals with a long contribution history.

This paper analyzes behavioral responses to this public pension reform that expanded the so-called *old-age pension for the especially long-term insured* by allowing individuals with at least 45 contributory years to retire without deductions as early as age 63. In the years before the reform, retiring without deductions was only possible at age 65, i.e., the reform implied a significant shift in the retirement age at which the long-term insured can retire without a financial penalty. Another element of the reform was to include additional pension-relevant periods such as periods of unemployment benefit receipt or of voluntary contributions to the public pension insurance as contribution years, thereby broadening the pool of eligible individuals. In the three years following the reform, the old-age pension for the especially long-term insured – commonly known as the '*retirement at 63*' scheme – has been the most important pathway into early retirement in Germany. Overall, it is the second most common pathway towards retirement, with on average 25% of new retirees exiting the labor market through this scheme, as compared to 33% of retirees retiring at the regular retirement age (Deutsche Rentenversicherung Bund, 2018).

We assess responses to this reform based on high-quality administrative data on pension claimants from public pension insurance accounts for the years 2013-2017. We first provide descriptive graphical evidence showing that the likelihood of retiring at the age of 63 in

---

This chapter is joint work with Mathias Dolls. An older version circulates as Dolls and Krolage (2019).

post-reform years rises for eligible, but not for non-eligible individuals. We likewise provide suggestive survey evidence that the reform has changed retirement expectations and has to some extent re-established age 63 as a retirement reference age for eligible individuals.

In our main analysis, we make use of two identification strategies to estimate the causal effect of the reform on retirement choices. First, we employ an event study design which exploits that individuals become eligible for claiming early retirement without deductions at different ages. The event study analysis shows that the probability of retiring immediately upon becoming eligible for early retirement without deductions increases by approximately 10 percentage points relative to the counterfactual probability of retiring at the same age with deductions. This is even the case for individuals who do not become eligible at ages commonly framed as reference ages for retirement. Looking at specific subgroups in our sample, we find stronger responses for men compared to women and for voluntarily insured and employed individuals compared to those in marginal employment.

Second, we use matching techniques and compare retirement choices of eligible and non-eligible individuals. Results from the coarsened exact matching analysis are in line with those from the event study analysis. In 2015 to 2017, the three years after the reform had been introduced, eligible individuals retire on average 6.4 months earlier than non-eligible individuals with identical characteristics. This indicates that reducing deductions by one percentage point lowers the average retirement age by 0.74 months. In a robustness check, propensity score matching estimates confirm that the reform leads to a preponement of retirement among eligible individuals of on average 6 months.

Finally, our study provides the first fiscal costs projection of the reform taking into account actual behavioral responses of eligible individuals. Our projections indicate additional pension insurance expenditures of 10.4 billion euros and aggregate fiscal costs of 19.8 billion euros between 2014 and 2017. These projections are at the upper end of the range of previous back-of-the-envelope estimates. They also exceed projected costs assumed in the draft government bill by more than 3 billion euros for the period under consideration.

The paper contributes to the literature studying how individuals respond to incentives in the retirement insurance system. A large empirical literature addresses the effects of pension reforms which increase the regular retirement age (Atalay and Barrett, 2015; Behaghel and Blau, 2012; Engels et al., 2017; Hanel and Riphahn, 2012; Mastrobuoni, 2009) or the early retirement age (Geyer and Welteke, 2020; Manoli and Weber, 2018; Staubli and Zweimüller, 2013). Mainly employing difference-in-difference or regression discontinuity designs, these studies find substantial labor market effects, albeit at varying magnitudes. In addition to increasing employment and an upward shift in retirement claiming ages, some of the papers also find evidence for program substitution towards unemployment insurance.

In contrast to these studies, our paper assesses a very salient reform that increased incentives to retire early, which is of particular interest as responses to changing incentives may be asymmetric. Evidence on such reforms is much more scarce. Previous research suggests that reforms granting the possibility to retire early lead to a reduction in the average retirement age (Börsch-Supan and Schnabel, 1998; Baker and Benjamin, 1999; Vestad, 2013). The reform we investigate has a substantially different approach though, as rather than enabling retirement at earlier ages than before, it drastically alters financial incentives for early retirement for a subgroup of the population. Evidence on responses to financial retirement incentives that do not entail changing retirement ages is mixed. Analyzing the introduction of deductions in a structural dynamic retirement model, Bönke et al. (2018) find sizable behavioral responses. In contrast, results by Brown (2013) and Manoli and Weber (2016) point to only a limited responsiveness of retirement choices to financial incentives.

Responsiveness to financial incentives also relates to the recent literature on reference dependence and social norms in retirement behavior[1]: Using bunching analysis, Seibold (2019) finds that retirement patterns cannot be explained by financial incentives alone. Rather, framing statutory ages as reference ages results in increased retirement probabilities at these thresholds. Likewise, Behaghel and Blau (2012) and Cribb et al. (2016) find behavioral responses to reforms of the statutory/early retirement ages that entail no or only limited financial incentives.[2] We also contribute to this literature on the role of reference ages. In our sample, all individuals are able to retire at the early retirement reference age 63, with the difference that financial incentives differ considerably between those who are eligible for early retirement without deductions and those who are not. While the early retirement age potentially serves as a more salient focal point for those eligible for the reform (and indeed gains in importance as revealed by our survey analysis), our results nevertheless show that large financial incentives affect retirement behavior. This is the case even if they do not coincide with ages commonly perceived as focal points.

The paper proceeds as follows. Section 1.2 describes the institutional background of the German pension system and the 2014 early retirement reform. In section 1.3, we provide information about the data sources used in subsequent analyses. In addition, we present a

[1] See e.g. Gustman and Steinmeier (1986), Rust and Phelan (1997) or Asch et al. (2005) for early contributions who do not find evidence for 'excess retirement' that cannot be explained by financial incentives.

[2] Rabaté (2019) shows that labor demand-induced job-exits as a consequence of mandatory retirement can explain part of the bunching at references ages.

survey analysis on retirement aspirations and first descriptive evidence of behavioral effects. Section 1.4 discusses our empirical strategy to identify the causal effects of the reform on retirement choices. Our main results are presented in section 1.5. Fiscal costs projections are reported in section 1.6. Section 1.7 concludes.

1.2 Institutional Background

1.2.1 The German Public Pension System

Covering almost all private and public sector employees[3], the German statutory pension system provides old-age pensions as well as invalidity and survivors' benefits. Financed as a pay-as-you-go scheme, the calculation of pension benefits is based on a person's contribution history. Entitlements are calculated according to a point system, where the number of pension points is determined by the ratio of individual annual earnings to average earnings across contributors in the same year. The system also features certain redistributive properties, such as pension points for child raising.

In recent years, the system has seen numerous reforms, affecting both the retirement age and the choice of pathways towards retirement. In light of demographic tensions, most of these reforms focused on increasing the retirement age or restricting pathways for accessing retirement. Most notably, recent years have seen a stepwise increase in the regular retirement age from 65 to 67. Retiring earlier is possible through several early retirement schemes, but usually requires deductions of 0.3% per month of retiring early. The accessibility of schemes depends on the insurance record, notably on the number of contributory years. Retiring as early as age 63 with deductions is possible for those with at least 35 contributory years. In addition to periods spent in employment, these also include periods spent raising children, voluntarily contributing or, under certain conditions, receiving unemployment benefits. Severely disabled individuals face both a lower regular and a lower early retirement age. For those born prior to 1952, two additional pathways were possible. Women with at least 15 contributory years, 10 of which have been spent actively contributing after age 40, may retire as early as age 60, but face deductions for each month of early retirement. Likewise, retiring earlier is also possible after unemployment or partial retirement.

[3] Civil servants are exempt from the statutory pension system. While self-employed individuals in certain vocations, such as craftspeople, are covered by compulsory insurance, other self-employed individuals have the option to opt into public pension insurance.

Table 1.1 shows the respective retirement ages by cohort. The reference age for deductions, which varies across cohorts, differs from the regular retirement age for some birth cohorts. Hence, this reference age is also depicted here. We also list the deductions a person faces when retiring at 63 through the 'regular' early retirement scheme (see fourth row of Table 1.1).

Table 1.1: Retirement ages by birth cohort

	1947	1948	1949	1950	1951	1952	1953	1954	1955	1956
Regular retirement	65+1	65+2	65+3	65+4	65+5	65+6	65+7	65+8	65+9	65+10
Early retirement (reference age)	65	65	65+1..3	65+4	65+5	65+6	65+7	65+8	65+9	65+10
Early retirement (with deductions)	63	63	63	63	63	63	63	63	63	63
Deductions at 63[a]	7.2%	7.2%	$\sim$ 8%	8.4%	8.7%	9.0%	9.3%	9.6%	9.9%	10.2%
Retirement at 63 (no deductions)	63	63	63	63	63	63	63+2	63+4	63+6	63+8
Retirement for disabled (ref.)	63	63	63	63	63	63+1..6	63+7	63+8	63+9	63+10
Retirement for disabled	60	60	60	60	60	60+1..6	60+7	60+8	60+9	60+10
Retirement for women	60	60	60	60	60	-	-	-	-	-
Retirement after unemployment	61+1..62	62+1..63	63	63	63	-	-	-	-	-

Notes: This table shows the earliest possible retirement ages for regular and early retirement schemes in years + months. For example, '65+2' refers to 65 years and two months. Early retirement schemes other than 'retirement at 63' require deductions of 0.3% per month, for which reference ages are shown. Severely disabled individuals face a lower reference age. Retirement ages continue to increase up to the 1964 birth cohort. To limit table size, we restrict this table to a selection of relevant birth cohorts.

a: Deductions when retiring at 63 with less than 45 contributory years, not considering foregone benefits due to the shorter contribution period.

1.2.2 The 2014 Early Retirement Reform

The 2014 reform of the *'old-age pension for the especially long-term insured'* contained two main elements. First, it abolished deductions at the early retirement age for individuals with long contribution histories. From July 2014 onwards, the so-called 'retirement at 63' scheme permitted individuals with at least 45 contributory years to draw a pension without deductions as early as at age 63. Between 2012 and 2014, the same was possible only at age 65. Prior to the reform, individuals who would now be eligible for the 'retirement at 63' scheme and who retired at 63 would have faced deductions for a period of 24 months, amounting to 7.2%. Second, the reform broadened eligibility criteria. Since 2014, further pension-relevant periods, in particular periods of unemployment benefit receipt and periods of voluntary contributions to public pension insurance, count towards the 45 contributory years. This enlarges the pool of potential recipients of the old-age pension for the especially long-term insured.

As a person retiring through the ‘retirement at 63’ scheme faces no deductions, the change in incentives upon becoming eligible for the scheme is very large. For example, a person born in 1950 with 44 years and 11 contributory months at age 63 could either retire at age 63 and face deductions of 8.4%, or keep on working for one more month and not face deductions at all. If the person earned an average income, delaying retirement by one month would thus increase retirement benefits by roughly 9.4% relative to pre-eligibility benefits.[4] This is much larger than the aforementioned deductions of 8.4% since deductions are generally framed relative to benefits without deductions. In turn, the 9.4% are set in relation to the comparatively lower pre-eligibility benefits with deductions. If retirement was delayed one further month, benefits would only increase by a further 0.19% due to accumulating more pension points during this month.[5]

Similar to the regular retirement age, the minimum retirement age for the ‘retirement at 63’ scheme increases across cohorts (see fifth row of Table 1.1). From the birth cohort 1953 onwards, it increases stepwise until reaching the age of 65 for the 1964 birth cohort.

The reform of the old-age pension for the especially long-term insured was part of a substantial and very salient retirement reform, which also increased pensions for mothers of children born before 1992 and increased invalidity benefits.[6] Being the first large project by the new grand coalition government that had formed at the end of 2013, it was widely discussed in the media. The reform was first announced in the new government’s coalition agreement in mid-December 2013, a first legislative draft was passed in January 2014 and the final law was passed in May. The Ministry of Labor dedicated a publicity campaign to the retirement reform in January 2014 (*‘not a gift, but well-deserved’*), claiming that hard-working individuals

[4] This calculation assumes the person has contributed for 45 years at an average income, and has thus earned 45 pension points. In this setting, working one further month at an average income increases pension benefits by 0.19%. Actual incentives differ across individuals due to heterogeneous earnings histories.

[5] Two further factors reduce incentives for delaying retirement once a person becomes eligible for the ‘retirement at 63’ scheme. First, while individuals who continue working beyond the regular retirement age receive an extra surcharge amounting to 6% in addition to the pension points earned after the regular retirement age, this extra surcharge is not granted to those who continue working after reaching eligibility for the ‘retirement at 63’ scheme. Second, the gradual introduction of deferred taxation of public pensions since 2005 implies that the taxable share of public pensions has been increasing by 2 percentage points per year recently. While the taxable share amounted to 68% in 2014, it has risen to 74% in 2017, the most recent year of our sample period.

[6] These other aspects of the reform should not confound our analysis. For once, invalidity benefits are commonly drawn at a much earlier age and with much fewer contributory years than required by the ‘retirement at 63’ scheme. While invalidity benefits play a very small role in our sample, we nevertheless exclude all individuals drawing invalidity benefits. Also, our analysis accounts for group demographics. Any potential income effect of mothers’ increased pensions should thus not exert a differential effect on control and treatment groups.

benefited from the reform as a reward for having contributed to society throughout their lives. Hence, the reform could hardly be anticipated in 2013, but was very salient from the beginning of 2014 onwards. The reform was also subject to much public discussion, with critics stating that the reform would increase benefits for those who already receive large retirement benefits at the expense of younger generations, whose contributions fund the reform. In this spirit, a pre-reform analysis indicated that eligible individuals are entitled to above-average retirement benefits and do not have worse health than non-eligible individuals (Börsch-Supan et al., 2015).

1.3 Data and Descriptive Evidence

1.3.1 Public Pension Insurance Accounts

For our main analysis identifying the causal effect of the early retirement reform on retirement choices, we employ high-quality administrative data on pension claimants from public pension insurance accounts (*Versichertenrentenzugang 2013-2017*). The scientific use file contains a 10% random sample of all individuals entering retirement between 2013 and 2017. As the dataset is process-produced, it mainly contains variables needed for calculating pension entitlements. Amongst others, we observe personal characteristics such as gender, marital status, education level and region of residence, as well as variables on the contribution history and retirement. These include the exact retirement age in months, the chosen retirement scheme, pension points, i.e., accumulated pension contributions, and pension-relevant periods, which enable us to determine eligibility for the early retirement scheme. The dataset provides further details on the three years preceding retirement, such as the respective socio-economic status and the annual salary.

We supplement our analysis with results from the *Survey of Health, Ageing and Retirement in Europe (SHARE-RV)*, a cross-national panel survey with a focus on the middle and old-age population, and focus on survey questions on early retirement aspirations, private retirement savings and physical health conditions. The German survey can be linked to administrative records from public pension insurance accounts (*Versichertenkontenstichprobe*). We exploit this feature in order to compare survey responses with actual retirement behavior observable in the public pension insurance accounts.

1.3.2 Sample for Estimating Behavioral Responses

Our analysis focuses on individuals with a long contribution history of 40 to 47 contribution years at age 63. While all individuals in this group are able to retire with deductions at age 63, their eligibility for early retirement without deductions differs. This allows us to identify a treatment (control) group of individuals who are affected (unaffected) by the reform. We exclude individuals who retired in prior years and are included in the dataset for administrative reasons, individuals with previous retirement spells, and individuals entering invalidity benefits (*Erwerbsminderungsrente*). Individuals who spent part of their career abroad, receive pensions according to the Foreign Pension Law or are subject to transitory regulations are likewise dropped from the sample. The overall sample contains roughly 110,000 individuals from birth cohorts 1947-1956 retiring between 2013 and 2017, of which we use different subsets for our two estimation approaches.

The calculation of relevant pension periods slightly differs from the calculation of contributory periods in other retirement schemes. In particular, the 'retirement at 63' scheme does not consider periods spent in education or periods of long-term unemployment. Periods of short-term unemployment with unemployment benefit receipt count towards the relevant pension period as long as they do not occur in the two years leading up to retirement. We thus subtract periods of education, unemployment extending 12 months, and unemployment occurring in the two years prior to retirement from the overall contributory period.[7]

Table 1.2 shows summary statistics across (i) the sample used in the event study and (ii) the weighted sample used in the coarsened exact matching procedure (cf. section 1.4). Both samples differ: Sample (i) corresponds to those who become eligible for the 'retirement at 63' scheme at some point prior to reaching the regular retirement age or who potentially could have become eligible had they not retired before. Pooling the data years 2015 to 2017, the event study encompasses the cohorts 1950 to 1954.[8]

[7] We do not observe unemployment benefit receipt in the data, but only the total number of months in unemployment. By counting only a maximum period of unemployment of 12 months towards the contributory period, we have chosen a conservative approach of identifying individuals eligible for the 'retirement at 63' scheme. Importantly, our sample includes individuals with long and stable employment biographies who typically only have short periods of unemployment, if at all.

[8] As the estimation strategy requires individuals in the sample to become eligible over time, the sample focuses on individuals born between 1950 and 1954. We do not include the data year 2014 as our data does neither contain the retirement month nor the birth month, but only the birth year and the retirement entry age in months. Due to this limitation, we do not observe for 2014 whether a person has become eligible before or after the introduction of the 'retirement at 63' scheme in July 2014.

Table 1.2: Public pension data: Summary statistics

	Event study	Matched sample
Female	46.9%	48.7%
Married	75.1%	75.7%
East German	31.7%	38.7%
Education		
None	6.7%	4.2%
Vocational degree	63.1%	59.6%
Advanced occupational degree	5.0%	3.8%
University degree	2.3%	2.0%
Unknown	22.9%	30.5%
Labor market status before turning 63		
Employed	75.7%	75.6%
Marginally employed	3.6%	3.3%
Voluntarily insured	3.3%	4.2%
Short-term unemployed	7.3%	4.9%
Receiving other social benefits	4.6%	2.9%
Credit period (sickness leave etc.)	1.6%	4.4%
Employer-sponsored early retirement^a	0.6%	0.2%
None/ unknown	3.3%	4.5%
Pension points at 63	43.8	41.9
Number of observations^b	61,318	62,375

Notes: This table shows summary statistics for the public pension insurance accounts sample, based on own calculations. Event study: 2015-2017. Matched sample: summary statistics after weighting, 2013-2017. a: Does not include partial retirement (Altersteilzeit). b: Unweighted sample size.

Sample (ii) encompasses individuals with 45-47 contribution years at age 63 and individuals who could at most attain 43-45 contribution years at the regular retirement age, i.e., individuals who are immediately eligible for the 'retirement at 63' scheme as well as a control group of individuals who are ineligible for the scheme. Individuals who become eligible between age 63 and the regular retirement age are included in sample (i) but not in sample (ii).

The matched sample encompasses the cohorts 1947 to 1956. A majority of those in the sample are employed prior to retirement and have earned a vocational degree. Compared to the universe of retirees, East German retirees, in particular East German women, are overrepresented due to their on average longer and more continuous labor market biographies.

1.3.3 Retirement Expectations and Descriptive Evidence

Prior to our estimations, we present descriptive evidence showcasing retirement expectations and retirement behavior. In a first step, we explore the effect of the reform on early retirement aspirations using survey responses from the *Survey of Health, Ageing and Retirement in Europe* (Börsch-Supan et al., 2013; Forschungsdatenzentrum der Rentenversicherung and Max-Planck-Institut für Sozialrecht und Sozialpolitik, 2019). This provides first suggestive evidence on changing reference ages in the wake of the reform. Exploiting that the German version of the panel survey (SHARE-RV) can be linked to administrative records from public pension insurance accounts, we identify eligible and non-eligible survey respondents and compare pre- and post-reform survey responses as well as post-reform retirement choices. We focus on survey waves conducted in the year prior to the introduction of the reform (wave 5, 2013) and one year after the reform became effective (wave 6, 2015). As in our matching analysis studying behavioral responses, we define a treatment and a control group. The treatment group consists of individuals in the survey who are immediately eligible for the 'retirement at 63' scheme when reaching the early retirement age or who become eligible at some point between the early and the regular retirement age. The control group includes survey participants who have accumulated 35-42.5 contribution years between the early and the regular retirement age and who hence will not become eligible before reaching the regular retirement age.[9]

We first assess whether retirement expectations changed for those affected by the reform, and whether expectations differ between treatment and control groups. The first row of Table 1.3 shows the age at which individuals expect to receive a public pension for the first time. The share of respondents in the treatment group who expect to retire at the age of 63 increases from 27% before the reform to 39% after the reform, which is a first indication that eligible individuals adjusted their retirement aspirations already shortly after the reform became effective. In contrast to the large increase in the treatment group, the share of respondents in the control group who expect to retire at the age of 63 does not change much (22% in 2013 vs. 24% in 2015). This indicates that the reform has to some extent re-established the age of 63 as a focal point for (early) retirement for those who are eligible, whereas little change can be seen for those who do not benefit from the reform.

[9] Note that the definition of treatment and control group differs from the matching analysis due to the much smaller sample size of the SHARE-RV survey. Information on the contribution history of survey participants is retrieved from linked public pension insurance accounts. Survey participants have to consent to the data linkage. In waves 5 (6), roughly 72% (79%) of respondents have agreed to link information from their public pension insurance accounts to their survey responses. This leaves us with N=181 in the treatment group and N = 121 in the control group.

Second, we compare survey responses from 2013, i.e., the year before the ‘retirement at 63’ reform was announced, with actual retirement choices after the reform became effective. Among those in the treatment group who retire through the ‘retirement at 63’ scheme, 41% had stated in 2013 not to look for early retirement. With roughly 17%, the corresponding share of early retirees with no aspirations for early retirement in the pre-reform year is much lower in the control group. A comparison of these shares suggests that the reform has induced a significant number of eligible individuals to change their retirement plans, speaking against the reform resulting in pure windfall gains.

Table 1.3: Survey evidence from SHARE-RV: Retirement expectations

Share of respondents	Treatment group	Control group
Expecting to retire at age 63		
...Wave 5, 2013	27.3%	21.8%
...Wave 6, 2015	39.3%	23.5%
Retiring before reaching the regular retirement age		
... and stating in wave 5 (2013) not to look for early retirement	41.0%	17.1%

Notes: Own calculations based on SHARE-RV, waves 5 and 6. Treatment group: Individuals eligible for the ‘retirement at 63’ scheme before reaching the regular retirement age (N=181). Control group: Individuals with 35-42.5 contribution years between early and regular retirement age (N=121). Treatment and control group only include individuals who retire after the ‘retirement at 63’ reform became effective (07/2014). Survey respondents in treatment and control group have consented to the linkage of their survey responses with administrative records of the German pension insurance.

We subsequently provide first descriptive evidence on retirement behavior in our main sample. Table 1.4 shows the fraction of retirees retiring through each retirement scheme in the full sample, regardless of contribution years.[10]

In the years following the reform, about 40% of all retirees in our sample retire at the regular retirement age, while about 30% exit the labor market through the ‘retirement at 63’ scheme. The remainder retires early through another retirement scheme, in most cases facing deductions. The share choosing to retire through one of the alternative retirement schemes declines over time as two of these schemes were only accessible to pre-1952 birth cohorts, and possibly due to substitution towards the ‘retirement at 63’ scheme.

[10] As our sample focuses on old-age retirement and excludes invalidity benefits, which may be drawn at any age by individuals who are physically unable to work, these fractions deviate from official statistics.

Table 1.4: Retirement choices by year in the full sample

	2013	2014	2015	2016	2017
Regular retirement	41.79%	44.78%	39.33%	39.44%	42.15%
Early retirement with deductions	17.82%	10.38%	16.03%	19.51%	18.60%
Retirement at 63^{a}	2.60%	18.40%	30.88%	28.89%	31.15%
Retirement for severely disabled	12.30%	9.71%	6.39%	7.15%	7.11%
Retirement for women	15.11%	9.59%	4.84%	3.72%	0.74%
Retirement after unemployment	10.37%	7.14%	2.52%	1.30%	0.24%

Notes: This table shows retirement choices in the public pension insurance account data. The sample includes retirees regardless of their contributory periods, excluding invalidity benefit recipients.
a: Prior to July 2014, a precursor of the 'retirement at 63' scheme allowed for early retirement without deductions for individuals with 45 contributory years at age 65. The data assigns individuals claiming an old-age pension for the especially long term insured after age 65 to the same retirement pathway as those entering early retirement without deductions at age 63 following the reform. For this reason, a small percentage of retirees enters retirement through this scheme prior to the reform.

To showcase the distribution of retirement ages, Figure 1.1 depicts the frequency of retirement decisions by retirement scheme and age in the pre-reform benchmark 2013 as well as in the post-reform years 2015 to 2017. While we observe the month of retirement in the most recent years, the retirement age is only available on a quarterly basis prior to 2014. For this reason, the 2013 retirement ages can only be depicted in 3-month age intervals, while more precise monthly retirement ages are available in the years that follow. Retirees retiring between age 60 and age 66 are included in the graphs regardless of their accumulated pension points.[11] The high frequency of early retirement via the 'retirement at 63' scheme in panel 1.1 (b) indicates that the shift towards early retirement in 2015 to 2017 is substantially driven by the retirement reform. A large fraction of those who retire early enter retirement at age 63, at 63 + 2, or at 63 + 4 months.

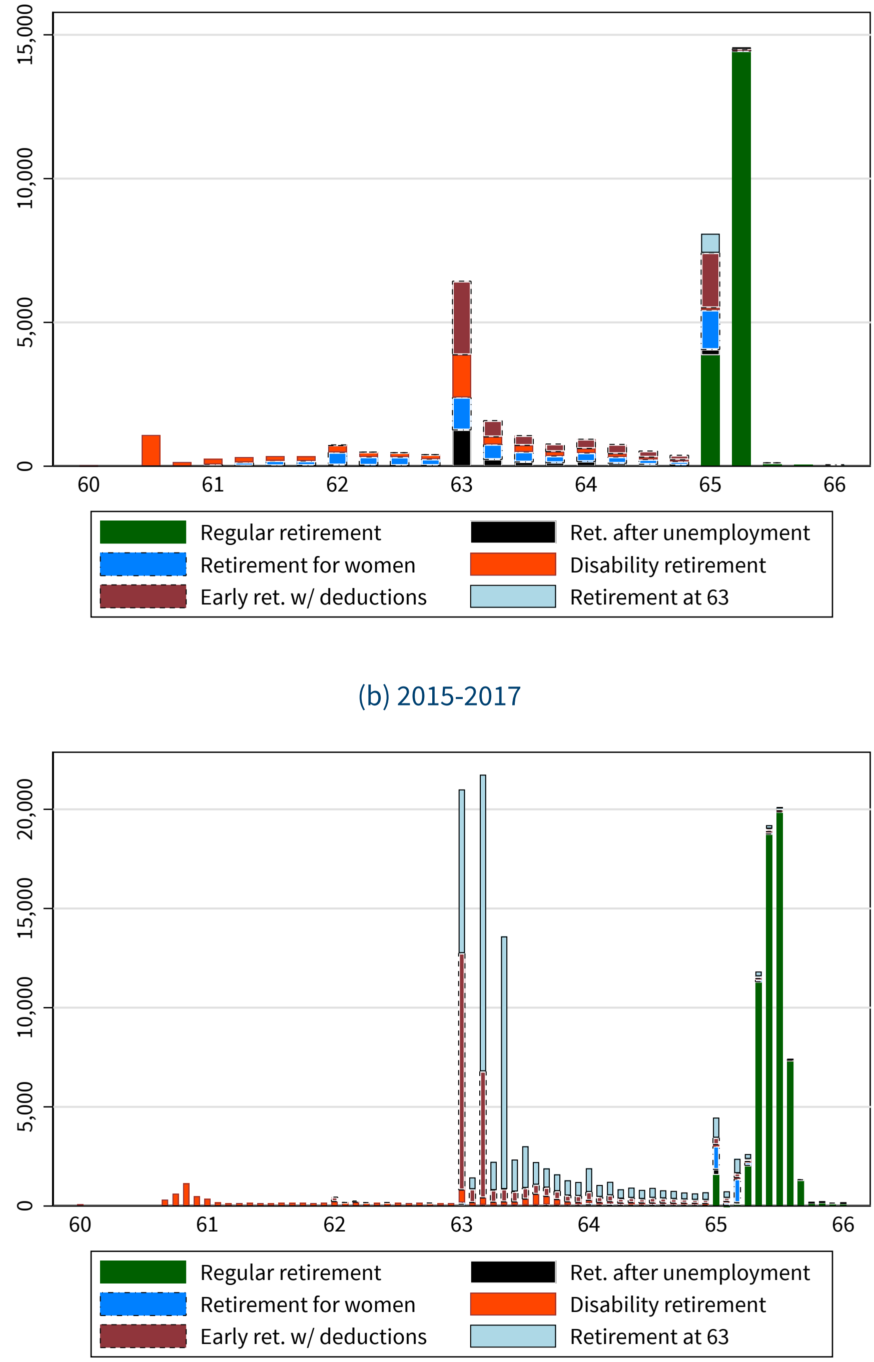
Figures 1.2 and 1.3 plot the Kaplan-Meier failure function, i.e., the empirical cumulative distribution function of retirement ages of those who are eligible for early retirement immediately once they reach the early retirement age (45-47 contributory years at age 63), those that will not reach 45 contributory years, and hence will not become eligible prior to age 65 (41-43 contributory years at age 63), and those in between that could potentially become eligible between ages 63 and 65 (between 43-45 contributory years at age 63).

[11] All other sample restrictions, such as the exclusion of individuals with previous retirement spells, do apply.

Figure 1.1: Frequency of retirement decisions by age and retirement type

(a) 2013

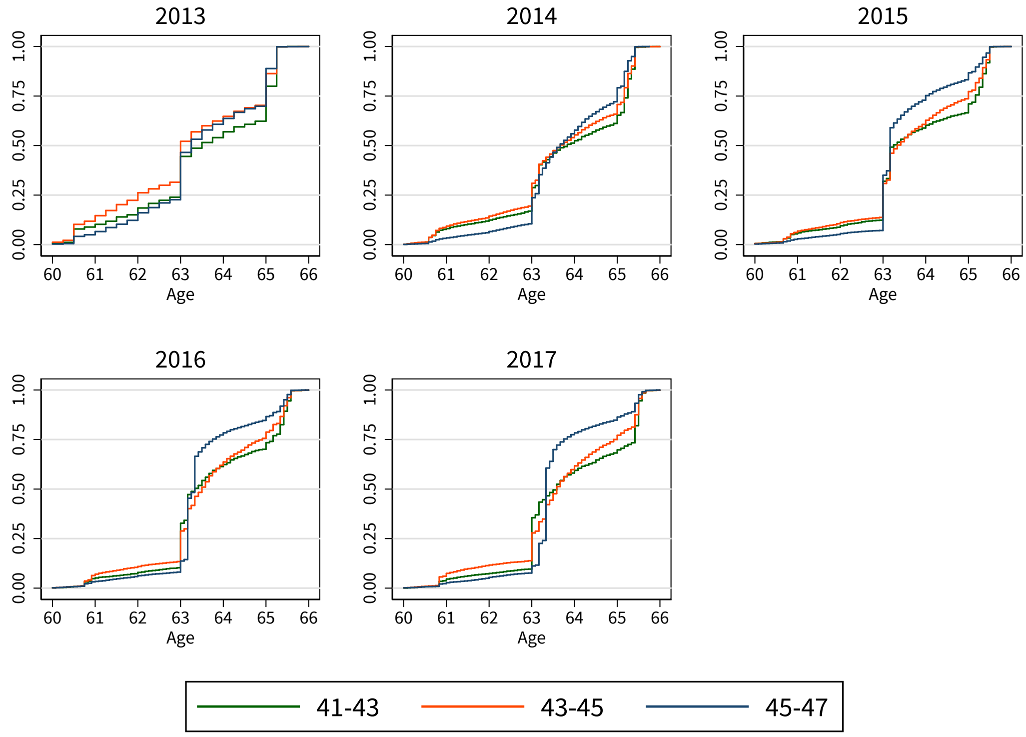
(b) 2015-2017



Notes: This figure shows the frequency of retirement decisions by retirement scheme and retirement age in the overall sample, irrespective of contribution years. For 2013, only a quarterly retirement age variable is available.

As opposed to section 1.5.2, where we use coarsened exact matching to ensure comparable characteristics of those who are eligible and ineligible for the reform, no weighting procedure is used here.[12] Hence, some of the differences may be due to differing characteristics across groups.

Figure 1.2: Kaplan-Meier failure estimates: ECDF of retirement ages by year



Notes: This figure shows the empirical cumulative distribution functions of retirement ages by contribution years at age 63 and by retirement year. Individuals with 45-47 contribution years are immediately eligible for the 'retirement at 63' scheme, while individuals with 43-45 contribution years attain eligibility over time. Individuals with 41-43 contribution years do not become eligible before reaching age 65.

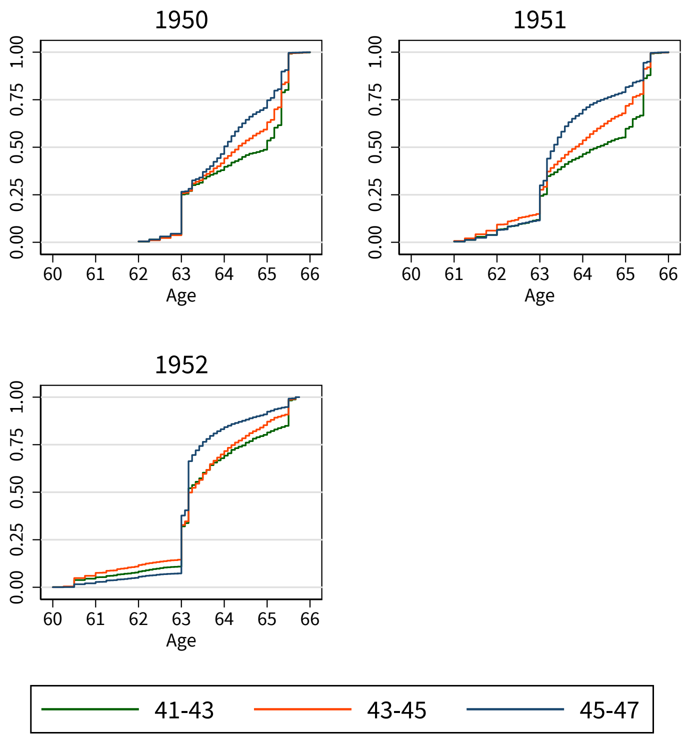
In all years, the Kaplan-Meier function displays a jump at age 63. While the size of the jump is roughly equal across different contributory year brackets in 2013, the jump is much larger for those with 45-47 than for those with fewer contributory years in 2015 to 2017. That is, retiring early occurs much more frequently for those eligible for the 'retirement at 63' scheme than for those who are not eligible. In 2016 and 2017, the jump of those with 45-47 contributory years takes place at a slightly later age than the jump in the cumulative distribution functions of those with fewer contributory years. This indicates that individuals in younger birth cohorts

[12] Note also that the definition of treatment and control groups is slightly simplified for expository purposes. We use a precise definition of treatment and control groups in our estimations.

postpone early retirement until reaching the age required to retire early without deductions (see Table 1.1). Note also that the Kaplan-Meier function of those with 43-45 contributory years is initially close to the distribution function of those who are ineligible at age 63, but then rises more steeply than the comparison groups' cumulative distribution function as more and more individuals with 43-45 contribution years at age 63 become eligible over time.

In a similar spirit, Figure 1.3 shows the Kaplan-Meier functions by birth cohort and contribution years at 63. For cohorts 1950, 1951 and 1952, a larger fraction of eligible than of ineligible individuals retires early. As opposed to Figure 1.2, the jump does not immediately occur at 63 for the 1950 cohort, and is smaller for the 1951 cohort, while the distribution function increases faster at later ages. This is due to the reform being passed in mid-2014, when those born in 1950 and many of those born in 1951 were already older than 63.

Figure 1.3: Kaplan-Meier failure estimates: ECDF of retirement ages by birth cohort



Notes: This figure shows the empirical cumulative distribution functions of retirement ages by contribution years at age 63 and by birth cohort. Individuals with 45-47 contribution years are immediately eligible for the 'retirement at 63' scheme, while individuals with 43-45 contribution years attain eligibility over time. Individuals with 41-43 contribution years do not become eligible before reaching age 65.

1.4 Empirical Strategy

We employ two different approaches to quantify responses to the pension reform. First, we conduct an event study to exploit that individuals become eligible at different ages. In this setting, the variation stems from individuals who reach 45 contributory years in differing months of their lives.

Second, we contrast the retirement behavior of individuals who are immediately eligible with the behavior of comparable individuals who are not eligible for this preferential retirement scheme. To ensure the similarity of the control vis-a-vis the treatment group, we employ a coarsened exact matching procedure.

1.4.1 Event Study

Using an event study design, we analyze the pathway towards retirement upon becoming eligible for the early retirement scheme. The sample includes individuals becoming eligible for the early retirement scheme prior to reaching the regular retirement age. In addition to those who reach at least 45 contributory years, we also consider those who would have reached 45 contributory years if they had not retired earlier through a different retirement scheme. This ensures that we do not only consider those who postpone their retirement in order to be eligible for the 'retirement at 63' scheme, but also include individuals with a comparable contribution history that choose to retire early regardless of the reform.

We estimate the following equation, where t corresponds to the age in months between age 62 and age 66:

$$Ret_{i,t} = \sum_{j=-12}^{24} \beta_j \delta Elig_{i,t-j} + \mu_i + \varsigma_{i,t} + \epsilon_{i,t} \tag{1.1}$$

Our dependent variable $Ret_{i,t}$ is a dummy whether person i is retiring at time t. $\delta Elig_{i,t-j}$ indicates a change in eligibility at time $t-j$, i.e., it is a dummy for the double condition of reaching 45 contributory years while also having reached age 63 (or age 63 + 2 (63 + 4) months for birth cohort 1953 (1954)). The coefficients β_j, which encompass the event window running from 12 months prior to 24 months after reaching eligibility, indicate differences in the probability of retirement entry at different points in the event window. We include individual fixed effects μ_i as well as monthly age x birth year fixed effects $\varsigma_{i,t}$. The year of birth interaction term is included in the time fixed effects to control for differing retirement conditions across birth cohorts (see Table 1.1). Notably, those born in 1953 (1954) can only claim early retirement benefits without deductions at age 63 + 2 (63 + 4) months, and should thus have different time fixed effects than individuals born in earlier cohorts. The error term $\epsilon_{i,t}$ is clustered at the individual level.

We estimate the equation across the event window and use β_{-12} as a benchmark set to zero in order to identify the other coefficients. This is a matter of scaling. For the ease of interpretation and as opposed to many other event studies, we do not use β_{-1}, i.e., the period prior to becoming eligible, as a benchmark as this period may reflect anticipation effects. In the months preceding eligibility, individuals might be less likely to retire as they are able to avoid deductions when they become eligible. Setting β_{-1} to zero would thus depict all coefficients in relation to a period in which soon-to-be eligible individuals have a lower propensity to retire than others.[13] Overall, the event study contrasts an individual's propensity to retire at a certain point in time prior to or after becoming eligible with a counterfactual probability of retiring with deductions at the same age. A positive coefficient β_j indicates that j periods after becoming eligible, a person is β_j percentage points more likely to retire than if they still had to wait 12 months until becoming eligible. As long as anticipation effects 12 months prior to becoming eligible are low, this approximates the increase in the propensity to retire j periods after becoming eligible due to the 'retirement at 63' scheme. Another way of interpreting the coefficient is that j periods after becoming eligible, individuals are more likely to retire than other, non-eligible individuals at the same age.

In order to set the magnitude of effects in relation to financial incentives, we subsequently calculate an *elasticity of the retirement entry probability* at $j = 0$. This measure sets the change in the retirement probability upon becoming eligible – the event study coefficient β_0 – in relation to the underlying change in financial incentives. The latter is defined as the percentage change between retirement benefits when retiring at $j = 0$, the month a person becomes eligible, and retirement benefits when retiring one month earlier.

$$\epsilon_0 = \frac{\beta_0}{ben_0/ben_{-1} - 1} \tag{1.2}$$

1.4.2 Matching

Subsequently, we contrast the behavior of those who are immediately eligible for the 'retirement at 63' scheme to those who are not eligible: The treatment group is composed of individuals with 45 to 47 contributory years at age 63, whereas individuals in the control group are not able to meet eligibility conditions for the 'retirement at 63' scheme by a short

[13] An event study in which β_{-1} is set to zero results in an upwards shift of all coefficients.

margin. The control group includes individuals who can at most achieve 43 to 45 contributory years at the regular retirement age if they continue working until then. Given that the regular retirement age is increasing across birth cohorts 1947-1956, the two-year window of contributory years at age 63 also differs across cohorts. It ranges from 40 years + 11 months to 42 years + 11 months for the 1947 cohort to 40 years + 2 months to 42 years + 2 months for the 1956 cohort.[14]

Other than that, the definition of the treatment and control group is analogous to the event study. As characteristics, preferences and expected retirement benefits might differ between groups, we employ a coarsened exact matching procedure (Iacus et al., 2012) to ensure the comparability of control and treatment groups. The following variables are used in the matching procedure: retirement year, gender, marital status, education level, a dummy for East Germany, socio-economic status at age 63[15], and the pension points an individual has earned at age 63 as a proxy for the earnings history and for expected retirement benefits. While all other variables require exact matching[16], the latter variable matches on coarsened intervals, where cutpoints are defined according to Sturges' rule (Scott, 2015).

In a second step, the retirement age in years is regressed on the eligibility dummy, with matching weights accounting for group differences. Any other variable affecting retirement preferences across groups should be controlled for by the matching procedure. The analysis is conducted separately for each year between 2013 and 2017, where the year 2013 constitutes a placebo test. The treatment and control groups only differ by their eligibility for the 'retirement at 63' scheme. As the reform was neither passed nor announced at this point, the behavior of individuals in the treatment and control group in 2013 should not be affected by the reform.

---

[14] For example, the regular retirement age for the 1950 birth cohort corresponds to 65 + 4 months. Individuals from this birth cohort who have accumulated at most 42 contributory years and 8 months at age 63 will not become eligible for the early retirement scheme before reaching the regular retirement age and hence enter the control group. In order to ensure comparability between the treatment and the control group, we also choose a 2-year window of contributory years at age 63 for the control group. For the 1950 birth cohort, this window ranges from 40 years + 8 months to 42 years + 8 months.

[15] This variable differentiates socially insured employees, marginally employed employees, voluntarily insured individuals including self-employed individuals who have opted into public retirement insurance, unemployment benefit recipients (*Arbeitslosengeld I*), recipients of other benefits, individuals with credit periods (*Anrechnungszeitversicherte*), e.g. due to sickness, and individuals who are neither working nor otherwise contributing towards retirement insurance, such as housewives.

[16] Treated individuals are dropped from the estimation if no individual in the control group shares their characteristics.

We subsequently calculate an *elasticity of the retirement age* with respect to the change in financial incentives. For this purpose, we set the treatment effect in relation to the percentage change in benefits due to deductions when retiring with deductions at the early retirement age.

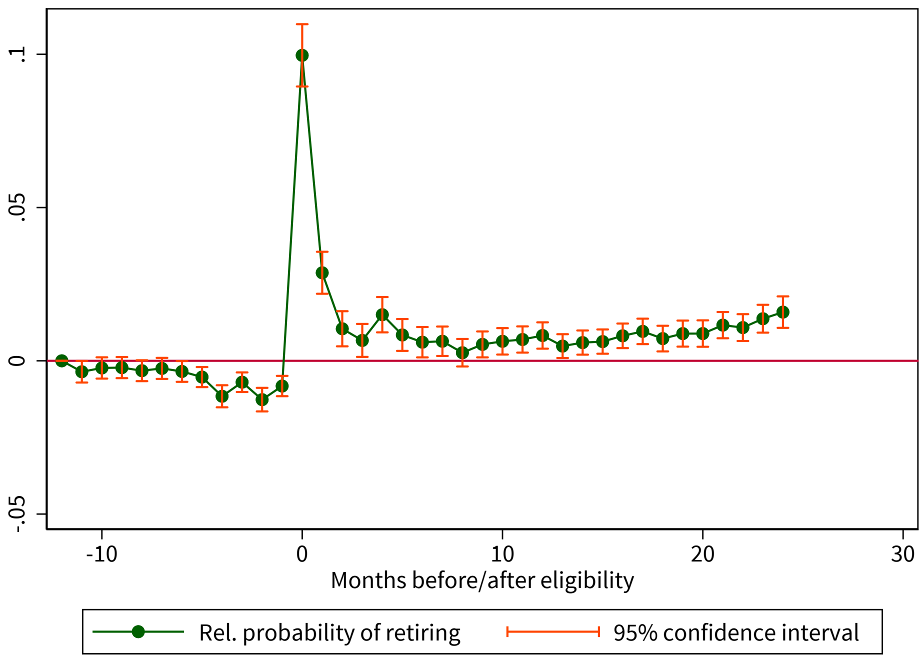
$$\eta = \frac{\Delta \, retirement \; age}{\% \; deductions} \tag{1.3}$$

1.5 Results

1.5.1 Event Study

Figure 1.4 displays the results of the baseline event study. The large spike at $j = 0$ shows that a substantial fraction of prospective retirees exits the labor force immediately once they are able to retire without deductions. The probability of retiring immediately upon becoming eligible exceeds the counterfactual probability of retiring at the same age with deductions by about 10 percentage points.

Figure 1.4: Event study: Baseline



Notes: This figure shows the event study coefficients β_j. The event window j ranges from -12 to +24 months and β_{-12} is set to zero. The specification includes age x birth year fixed effects. N=61,318.

With our sample averaging a 9.17% change in benefits upon becoming eligible, this amounts to an elasticity of the retirement entry probability of 1.09, i.e., a one percent increase in benefits increases the probability of retirement by 1.09 percentage points. We also find positive, albeit smaller coefficients in the months that directly follow eligibility, which is driven by employed individuals not immediately exiting the labor force, but remaining in their job for a few more months. Coefficients become slightly larger again towards the end of the event window, suggesting that eligible individuals have a slightly higher likelihood of retiring before reaching the regular retirement age than ineligible individuals.[17]

In the months preceding the eligibility date, a slightly negative effect arises, which indicates that some individuals postpone their retirement by a few months in order to benefit from the 'retirement at 63' scheme. There are small spikes in $j = -2$ and $j = -4$ which are mainly attributable to the 1953 and 1954 cohorts, for whom the early retirement age without deductions does not coincide with the early retirement age with deductions. While retiring with deductions is feasible at age 63, retiring through the 'retirement at 63' scheme is only possible two or four months after turning 63 for these cohorts (see Table 1.1). As individuals in these cohorts would face deductions of 9.3% (9.6%) when retiring at 63, but no deductions when retiring two (four) months later, there are strong incentives to retire at 63 + 2 (63 + 4) months. Hence, comparatively less of those eligible for the 'retirement at 63' scheme at age 63 + 2 or age 63 + 4 opt into early retirement with deductions at 63, leading to the larger negative effect at $j = -2$ and $j = -4$.

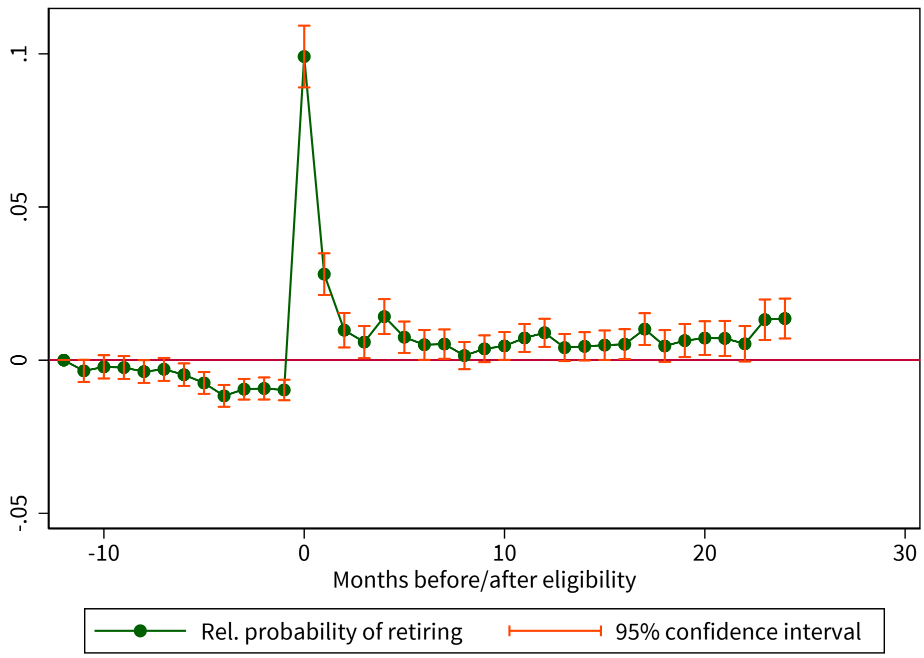
Analysis for Restricted Sample

An interesting question is to what extent our results are driven by the fact that the early retirement age for the 'retirement at 63' scheme is a focal point for early retirement (Seibold, 2019). In order to shed light on this question, we focus on the subset of individuals who reach 45 contributory years at some point between the early retirement age for the 'retirement at 63' scheme and the regular retirement age. In a first robustness check, we restrict the sample to individuals who reach eligibility only one month after turning 63 (or 63 + 2/ 63 + 4 months for the 1953/1954 birth cohort) or later. For this group, retiring early through the 'retirement at 63' scheme is not directly connected to the reference age of the 63rd birthday. In a second robustness check, we restrict the sample even further and focus on the subset of individuals who reach 45 contribution years at least one month after turning 64. For this group, eligibility neither coincides with age 63 nor with the 64th birthday, which may also serve as a focal point.

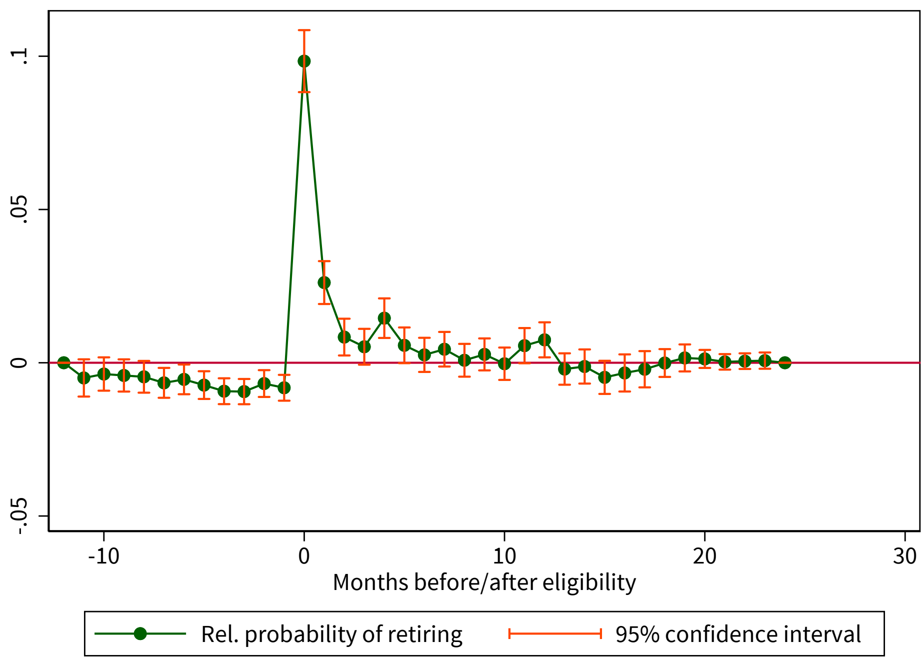
[17] Many of the individuals in the sample become eligible immediately at the early retirement age for the 'retirement at 63' scheme. For those individuals, the regular retirement age is beyond the event window. Note however that all coefficients are relative to $j = -12$, a period in which small anticipation effects might still be present.

Figure 1.5: Event study: Not immediately eligible individuals

(a) Individuals reaching eligibility later than age 63



(b) Individuals reaching eligibility later than age 64



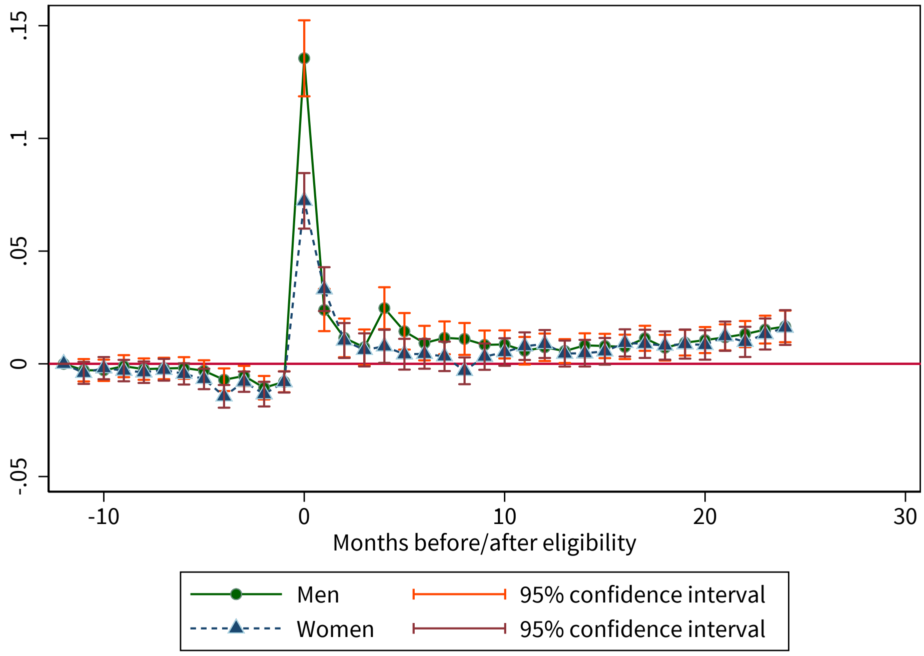
Notes: This figure shows the event study coefficients β_j for individuals who do not become eligible immediately. The event window j ranges from -12 to +24 months and β_{-12} is set to zero. Specifications include age x birth year fixed effects. (a) Only individuals becoming eligible after age 63 or, for birth cohort 1953 (1954), after age 63 + 2 (63 + 4) months. N=6,775. (b) Only individuals becoming eligible after age 64. N=4,393.

As shown in Figure 1.5, the overall trajectories of the event studies closely resemble those presented in Figure 1.4. In particular, the spikes at $j = 0$ are of a similar size as in our baseline event study, which implies that individuals exhibit strong responses to large and salient financial incentives, even if they do not coincide with reference ages.

Analysis Across Subgroups

We subsequently assess whether responses differ by gender. As shown by Figure 1.6, immediate effects are much stronger for men than for women. While the elasticity of the retirement entry probability amounts to 1.46 for men, it is only 0.80 for women.

Figure 1.6: Event study: Results by gender

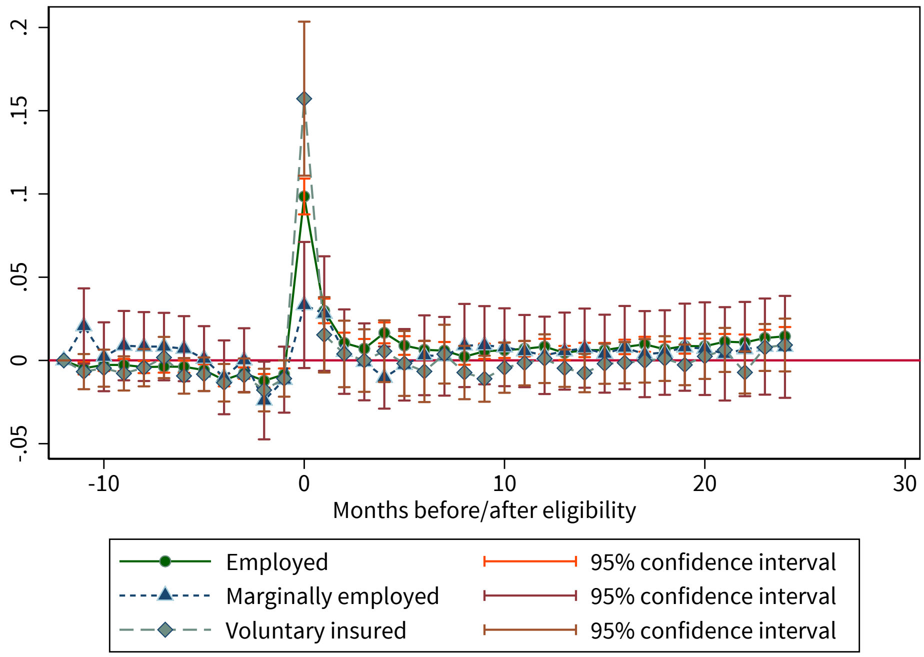


Notes: This figure shows the event study coefficients β_j by gender. The event window j ranges from -12 to +24 months and β_{-12} is set to zero. The specification includes age x birth year fixed effects. N=32,572 (Men) and N=28,746 (Women).

One possible explanation for these gender differences are cross-effects between spouses. Using Swiss data, Lalive and Parrotta (2017) show that women reduce their labor force participation once their partner reaches pension eligibility, whereas spousal retirement does not significantly affect male retirement behavior. A similar effect might play a role here. As on average, men tend to be older than their spouses, many 63 year old women will have spouses that become eligible for retirement schemes prior to them. Some of these women may choose to retire at the early retirement age regardless of incentives, resulting in a lower average responsiveness to becoming eligible for the 'retirement at 63' scheme.[18]

We also conduct the analysis by pre-retirement employment status. As the event study design requires variation in the age at which individuals become eligible, this subgroup analysis is only possible for employed, marginally employed[19] or voluntarily contributing individuals. Unemployment benefit receipt in the years prior to retirement, for example, does not increase contributory years. Hence, unemployed people are either eligible for the early retirement scheme right away, or will not become eligible at all. Figure 1.7 depicts the event study for socially insured employees, marginally employed employees and voluntary contributors.

Figure 1.7: Event study: Results by socio-economic status



Notes: This figure shows the event study coefficients β_j by employment status. The event window j ranges from -12 to +24 months and β_{-12} is set to zero. The specification includes age x birth year fixed effects. N=45,817 (Employed), N=2,268 (Marginally employed) and N=1,933 (Voluntarily insured).

[18] While the data contains marital status, we cannot match spouses and are hence unable to adequately assess the cross-effect of spousal retirement behavior.

[19] Marginally employed individuals only earn monthly wages of up to 450 euros and are exempt from social insurance contributions, but can opt into retirement insurance. This analysis is limited to marginally employed individuals who choose to contribute to retirement insurance and hence acquire further contributory periods.

The trajectory for socially insured employees looks rather similar to the baseline in Figure 1.4. There is a substantial spike of roughly 10 percentage points at $j = 0$, equivalent to an elasticity of the retirement entry probability of 1.08, followed by smaller positive effects in subsequent periods. In contrast, voluntarily insured individuals exhibit a much larger propensity to retire immediately, attaining an elasticity of the retirement entry probability of 1.83. Coefficients in follow-up periods are insignificant, which might be due to the fact that voluntary contributions increase retirement benefits only by very little once eligibility is reached.[20] Finally, we find only small and insignificant positive effects at $j = 0$ for those in marginal employment, suggesting that this group retires as early as possible, regardless of deductions.

1.5.2 Matching

Coarsened Exact Matching

Table 1.5 highlights the results of the baseline specification. The first specification (upper panel) shows OLS results without any matching. Specification 2 (lower panel) employs coarsened exact matching and matches on demographics and on characteristics related to the earnings history. The retirement age in years is used as dependent variable. After matching, eligibility for the scheme does not exert an effect on the retirement age in the placebo test (2013, first column, lower panel), adding credibility to the matching procedure.[21]

In 2015 to 2017, coefficients are large and significant, indicating a substantial behavioral response to the reform. Individuals in the treatment group retire on average 6.4 months (0.53*12) earlier than those in the control group. As average hypothetical deductions in our sample amount to 8.6%, our estimates translate into an elasticity of the retirement age with respect to deductions of 0.74. Otherwise put, reducing deductions by one percentage point decreases the average retirement age by 0.74 months.

In 2014, eligible individuals retire on average 2.3 months earlier, yielding an elasticity of 0.27. Amounting to roughly one third of the size of the 2015 to 2017 effects, the size of the coefficient for 2014 is in line with the other results as the scheme only became effective in the second half of the year. It is likely that two further factors lowered the 2014 coefficient. First, the reform's announcement at the beginning of 2014 might have led to anticipation effects as

[20] As opposed to socially insured employees who typically have a certain period of notice or who may take the requirements of their employer into account when deciding upon retirement, voluntarily insured individuals, many of them self-employed, may face fewer frictions.

[21] In the year 2013, retiring without deductions was possible from age 65 onwards (instead of e.g. at age 65 + 2 months for the birth cohort 1948) provided that a person had accumulated 45 contributory years. This might explain the small negative coefficient for 2013.

Table 1.5: Coarsened exact matching: Baseline specification

Retirement age					
	2013	2014	2015	2016	2017
Treat (no matching)	-0.074**	-0.081***	-0.298***	-0.293***	-0.279***
se	(0.035)	(0.030)	(0.025)	(0.022)	(0.023)
t	-2.09	-2.71	-12.14	-13.11	-11.90
N	8484	11240	14532	14012	14107
Treat (CEM, all observations)	-0.015	-0.192***	-0.541***	-0.520***	-0.536***
se	(0.064)	(0.063)	(0.051)	(0.043)	(0.044)
t	-0.23	-3.03	-10.53	-11.99	-12.24
N	7122	9418	12372	12001	12009

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Treatment group: 45-47 contribution years at age 63. Control group: Individuals who would reach 43 to 45 contribution years at the regular retirement age when contributing until the regular retirement age. Heteroskedasticity-robust standard errors. Specification 1: no matching. Specification 2: matching via demographics and total pension points at 63.

some individuals may have postponed their retirement to the second half of the year in order to benefit from the reform. Second, many of those becoming eligible once the scheme was passed were already older than 63, leaving less room for antedating retirement.

Table 1.6 depicts the results by gender, showing much larger coefficients for men than for women. While men retire on average 8.2 months earlier after the reform, women only retire 4.7 months earlier. This corresponds to a one percentage point reduction in deductions, lowering the average retirement age by 0.95 months for men and by 0.55 months for women. These results are in line with the lower responsiveness of women to the reform in the event study setting, shown in Figure 1.6.

Several reasons may play a role here: First, due to the 'retirement for women' scheme, women born before 1952 may already retire as early as age 60, albeit with deductions. Hence, some women in the treatment group may postpone retirement to age 63 in order to avoid deductions. This may compensate part of the counteracting effect of those antedating their retirement. Second, labor market affinity and selection effects due to gender roles could contribute to the smaller effects found for women. In the cohorts born in the late 1940s and early 1950s, female labor market participation was much lower than male participation. For this reason, those who accumulated 45 contribution years may constitute a selection of particularly labor market affine women, who on average derive a lower utility from retiring early. On the other hand, as discussed in section 1.5.1, some women may adjust their retirement behavior to their spouse's retirement choices, reducing the responsiveness to own financial incentives (Lalive and Parrotta, 2017).

Table 1.6: Coarsened exact matching: By gender

Retirement age					
	2013	2014	2015	2016	2017
Treat (CEM, all men)	-0.096	-0.353***	-0.673***	-0.686***	-0.700***
se	(0.083)	(0.086)	(0.076)	(0.061)	(0.064)
t	-1.16	-4.10	-8.89	-11.18	-10.92
N	3929	5608	6187	5478	5307
Treat (CEM, all women)	0.090	-0.060	-0.402***	-0.376***	-0.399***
se	(0.096)	(0.079)	(0.068)	(0.061)	(0.057)
t	0.35	-0.45	-5.94	-6.18	-6.95
N	3193	3810	6185	6523	6702

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Treatment group: 45-47 contribution years at 63. Control group: Individuals who would reach 43 to 45 contribution years at the regular retirement age when contributing until the regular retirement age. Heteroskedasticity-robust standard errors.

Next, we present matching results by socio-economic status. Table 1.7 focuses on those socio-economic statuses with a sufficiently large number of observations and comparability across years. Results are in line with the event study in Figure 1.7. Socially insured employees in control and treatment group do not exhibit a significantly different retirement behavior in 2013. In 2015 to 2017, those who are eligible retire on average 7.7 months earlier than those who are ineligible. Slightly lower effects are observed for voluntarily insured individuals, who retire on average 5.4 months earlier. These effects translate into the average retirement age decreasing by respectively 0.90 and 0.63 months if deductions are lowered by one percentage point. Fluctuations in coefficients across years may be attributable to the comparatively lower size of the subsample. In contrast, effects for marginally employed individuals are slightly positive, and, notably, significant in 2014, where those who become eligible retire 4.4 months later on average (elasticity of -0.51). This could be due to anticipation effects. As previously discussed, marginally employed individuals predominately retire immediately at age 63, even if this leads to deductions. In 2014, the reform only became effective in the second half of the year. This resulted in strong incentives to postpone early retirement to after July 2014. For the other socio-economic statuses, a similar effect is offset by comparatively earlier retirement in the second half of the year. In turn, individuals receiving unemployment benefits display a rather low level of responsiveness to the reform. Coefficients are negative and significant only in the year 2016, where unemployed individuals in the treatment group retire on average 3.3 months earlier (elasticity of 0.38).

Table 1.7: Coarsened exact matching: By socio-economic status at age 63

Retirement age					
	2013	2014	2015	2016	2017
Treat (CEM, employed)	-0.097	-0.360***	-0.684***	-0.634***	-0.615***
se	(0.093)	(0.075)	(0.063)	(0.053)	(0.051)
t	-1.05	-4.77	-10.83	-12.01	-12.13
N	4192	6359	8660	8538	9169
Treat (CEM, marginally employed)	0.038	0.365**	0.043	0.143*	-0.042
se	(0.165)	(0.181)	(0.084)	(0.082)	(0.090)
t	0.23	2.01	0.52	1.75	-0.47
N	327	381	577	587	460
Treat (CEM, unemployed (ALGI))	-0.100	-0.141	-0.180	-0.278**	-0.148
se	(0.131)	(0.166)	(0.135)	(0.121)	(0.119)
t	-0.76	-0.85	-1.33	-2.30	-1.25
N	715	411	572	543	349
Treat (CEM, voluntarily insured)	-0.127	-0.264	-0.563***	-0.382***	-0.407***
se	(0.148)	(0.185)	(0.131)	(0.131)	(0.128)
t	-0.86	-1.43	-4.29	-2.91	-3.19
N	344	447	526	472	437

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Treatment group: 45-47 contribution years at 63. Control group: Individuals who would reach 43 to 45 contribution years at the regular retirement age when contributing until the regular retirement age. Heteroskedasticity-robust standard errors.

Finally, we address possible cohort effects in previous estimations in order to check whether our results are affected by the changing availability of alternative retirement schemes across cohorts. Older cohorts in our sample had more retirement options. Women born prior to 1952 could enter the ‘retirement for women’ scheme, while men and women born prior to 1952 could choose ‘retirement after unemployment and partial retirement’. Therefore, a further analysis matches on the birth cohort instead of on the retirement year.[22] Table 1.8 shows the results for birth cohorts 1950, 1951 and 1952, respectively, based on pooled data years 2013 to 2017.[23] Results are again in line with previous findings: those in the 1950 cohort retire on

[22] We cannot match on both the retirement year and the birth cohort at the same time as this would restrict the possible difference in the retirement age between the control and the treatment group and bias the coefficient towards zero. In other words, the maximum treatment effect would be restricted to 11 months occurring if a person in the treatment group retiring in January was matched to a person in the control group retiring in December of the same year.

[23] Note that including the year 2013 in our estimations might result in a slight downward bias, given that individuals choosing to retire in 2013 do not yet adjust their behavior to the reform. However, only restricting the sample to the reform years 2014-2017 might, on the contrary, induce an upward bias. This would drop some of the least responsive individuals, in particular individuals born in 1950 who chose to retire early in the pre-reform

average 4.3 months earlier, while those born in 1951 and 1952 retire 6.5 months earlier. With elasticities of 0.51, 0.75 and 0.74, the responsiveness by cohort is very close to the baseline. The similar findings for birth cohorts 1951 and 1952 indicate that our results are not strongly affected by the availability of alternative pathways to early retirement for the 1951 birth cohort. The smaller coefficient for the birth cohort 1950 is due to the cohort's rather advanced age when the reform was passed, which implies that individuals in the treatment group could not antedate retirement as much as those in younger cohorts.

Table 1.8: Coarsened exact matching: By birth cohorts

Retirement age			
	1950	1951	1952
Treat (CEM, 2013-2017)	-0.355***	-0.545***	-0.554***
se	(0.048)	(0.062)	(0.052)
t	-7.38	-8.79	-10.65
N	7554	9188	11185

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Treatment group: 45-47 contribution years at 63. Control group: Individuals who would reach 43 to 45 contribution years at the regular retirement age when contributing until the regular retirement age. Heteroskedasticity-robust standard errors.

Robustness Check: Propensity Score Matching

As a further robustness check, we employ propensity score matching (Dehejia and Wahba, 2002). Instead of directly matching on observables, this methodology first predicts the probability of belonging to the treatment group using a logistic regression, and then matches observations in control and treatment group with a similar propensity score. We use Mahalanobis matching, matching on the same variables as with CEM.[24] As opposed to the CEM methodology, individuals with different characteristics, but a similar propensity score can be matched here. This results in an overall larger sample size than in the CEM procedure.

The results shown in Table 1.9 are in line with our previous findings using coarsened exact matching. While no significant effect can be found for the 2013 placebo test, coefficients for the years 2015 to 2017 are close to those presented in Table 1.5. According to the estimates

year 2013 regardless of incentives. Note also that the analysis by cohort is not conducted for birth cohorts 1953 and 1954 as our most recent data is from 2017, i.e., we do not observe individuals from these cohorts who retire at their regular retirement age.

[24] We use dummies for gender, marital status, being East German, education level, and labor market status before turning 63 to determine a person's propensity score. Pension points at age 63 enter the estimation as a third degree polynomial.

based on propensity score matching, individuals affected by the reform retire on average about 6 months earlier than those unaffected by the reform.

Table 1.9: Propensity score matching

Retirement age					
	2013	2014	2015	2016	2017
ATT (propensity score matching)	-0.095	-0.103	-0.490***	-0.526***	-0.489***
se	(0.075)	(0.078)	(0.051)	(0.047)	(0.048)
t	-1.27	-1.32	-9.64	-11.30	-10.24
N	8476	11230	14520	13999	14091

Notes: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. Treatment group: 45-47 contribution years at 63. Control group: Individuals who would reach 43 to 45 contribution years at the regular retirement age when contributing until the regular retirement age. Propensity score matching via demographics and a polynomial of total pension points at age 63.

Further Factors Possibly Affecting Retirement Behavior

In the public debate, advocates of the reform frequently stated that the reform enabled individuals with bad health outcomes due to their long working lives to retire early. As our main sample does not indicate the health status of retirees, we again draw on survey results from SHARE-RV to assess whether early retirement decisions are related to poor health conditions. For this purpose, we exploit survey questions on the self-assessed health status and on the presence of a long-term illness. As shown in Table 1.10, the share of respondents who perceive their health status to be ‘excellent’, ‘very good’ or ‘good’ is very similar in the treatment and the control group (66% vs. 67%). The same is true for the share of respondents with a long-term illness (51% vs. 48%). In both groups, respondents who feel to be in good health are slightly more likely to retire at the regular retirement age as compared to retiring before reaching the regular retirement age. However, retiring early with a self-assessed health status being ‘excellent’, ‘very good’ or ‘good’ occurs slightly more frequently in the treatment group compared to the control group (66% vs. 63%). While results should be interpreted with some caution due to the low sample size, they, in line with Börsch-Supan et al. (2015), do not lend support to the claim that individuals eligible for the ‘retirement at 63’ scheme have poorer health as a consequence of a long working life.

Second, we use the SHARE-RV data to shed some light on possible interactions between private retirement savings and early retirement. This could be of interest as financial constraints might affect retirement choices. For example, individuals with private retirement savings on top of their public pension might be less constrained by deductions that lower public pensions. It would also be conceivable that a lower discount factor induces people to work longer and to engage in private retirement savings. Both in the treatment and the control group, the share of respondents who declare to own a private retirement account is roughly 35% on average. In the treatment group, the share of individuals with a private retirement account is equal amongst those who choose to retire early and those who do not. In contrast, in the control group, the share of respondents with a private retirement account is higher among those retiring at the regular retirement age than among those retiring early with deductions. These results provide suggestive evidence that behavioral responses to the reform might be slightly stronger for individuals with private retirement savings.[25]

Table 1.10: Survey evidence from SHARE-RV: Further factors affecting retirement behavior

Share of respondents	Treatment group	Control group
Self-assessed health status excellent/very good/good (wave 5, 2013)	66.3%	67.2%
... among those retiring before reaching regular retirement age	65.8%	62.8%
... among those retiring at regular retirement age	70.7%	72.9%
With long-term illness (wave 5, 2013)	50.6%	48.3%
... among those retiring before reaching regular retirement age	50.4%	49.0%
... among those retiring at regular retirement age	53.7%	45.8%
With private retirement account (wave 5, 2013)	35.7%	34.6%
... among those retiring before reaching regular retirement age	35.7%	27.0%
... among those retiring at regular retirement age	35.7%	39.0%

Notes: Own calculations based on SHARE-RV, waves 5 and 6. Treatment group: Individuals eligible for the ‘retirement at 63’ scheme before reaching regular retirement age (N=181). Control group: Individuals with 35-42.5 contribution years between early and regular retirement age (N=121). Treatment and control group only include individuals who retire after the ‘retirement at 63’ reform became effective (07/2014). Survey respondents in treatment and control group have consented to the linkage of their survey responses with administrative records of the German pension insurance.

[25] SHARE-RV also contains a question on the monetary amount invested in the private retirement account. The mean value is 30,000 euros (25,600 euros) in the treatment (control) group, but the response rate is relatively low so that the number of observations becomes very small when we further differentiate between respondents retiring early and at the regular retirement age.

How Do Effects Compare?

On average, the abolition of deductions within the scope of the ‘retirement at 63’ scheme raises the probability of retiring immediately upon becoming eligible by 10 percentage points and lowers the average retirement age by 6.4 months. The magnitude of this effect is roughly equivalent to a one year increase in the early retirement age in Austria as found by Manoli and Weber (2018) and slightly exceeds the projected effect of increasing the regular retirement age in Germany from 65 to 66, which Seibold (2019) quantifies at 4 months. The effects in those papers are, however, not primarily driven by financial incentives, but rather by social norms or the role of references ages for retirement choices, while our event study effects hold when we focus on individuals reaching eligibility at non-reference ages.

As estimated by Bönke et al. (2018), the introduction of deductions in the German public pension system increased the average retirement age by 4.1 months, which would correspond to an elasticity of the retirement age of 0.57 with respect to deductions. With an estimated elasticity of 0.74, the magnitude of responses to the ‘retirement at 63’ scheme is moderately larger. Similarly and also in the German context, Engels et al. (2017) find that increasing deductions for early retirement for women by one percentage point reduces the average retirement rate by 1.9 percentage points. The latter effect is larger than our estimated elasticity of the retirement entry probability of 1.09, or 0.80 for women. However, Engels et al.’s estimated coefficients encompass both the effect of raising the regular retirement age and the increase in deductions, and are therefore not only driven by financial incentives.

One possible explanation for our comparatively large effect is that the scheme targets a set of particularly responsive individuals. As evidenced by our estimations, responses are especially strong amongst employed men – a group that forms a rather large share of those eligible for the reform. Moreover, the reform was very salient as a consequence of the government’s publicity campaign and a broad media coverage. Our finding that the largest response occurs in the month of becoming eligible suggests that frictions do not play an important role in the context of the ‘retirement at 63’ scheme. Another presumably important factor is the large size of the financial incentive. In our sample, the average change in pension benefits upon becoming eligible exceeds 9%. This contrasts with only minor increases in pension entitlements – 0.19% per month for an average income earner – if a person stays in the labor force after eligibility is reached. Finally, the rather large effects may possibly indicate a high responsiveness to a reform that facilitates retiring early, as opposed to reforms that tighten retirement conditions. Put differently, financial incentives may play a more (less) important role in settings opening (closing) pathways into early retirement.

1.6 Fiscal Costs

In the German public debate, the fiscal costs of the reform are controversially discussed. Current fiscal cost estimates vary substantially. At the lower end, the draft bill of the retirement reform estimated additional public retirement insurance expenditures amounting to 0.9 billion euros in 2014, 1.9 billion in 2015, 2.2 billion in 2016 and 2.0 billion in 2017, with costs slightly declining in subsequent years and then again increasing to 3.1 billion in 2030 (Deutscher Bundestag, 2014). Pimpertz (2017) provides a lower bound estimate of the reform's cost, assuming unchanged retirement behavior and focusing on foregone deductions only. This approach yields cost estimates ranging between 0.14 billion in 2014 and 1.2 billion euros in 2017. Using a simulation model, Werding (2014) projects somewhat higher costs of 0.5 billion euros in 2014, rising to 2.6 billion in 2015 and 3.2 billion in 2016 and 2017. Schnabel (2015) estimates that annual fiscal costs might rise to 6 billion euros if 125,000 individuals retire via the scheme per year – a figure which the number of actual claimants exceeds by far.[26] At the upper end, monthly costs of 1.3 to 2 billion euros were frequently circulated in the media[27] – an overestimation based on the total sum of pension benefits paid under the 'retirement at 63' scheme, neglecting that many of those claiming early retirement benefits would have otherwise retired early through another scheme.

Against this background of widely varying cost estimates, we strive to provide a more precise estimate of fiscal costs. Using our detailed microdata and the coarsened exact matching methodology described above, we are able to account for behavioral responses as well as for foregone deductions and contributions. To obtain counterfactual retirement choices, we match those retiring via the 'retirement at 63' scheme to the same control group as in section 1.5.2. Within each strata, we then compute the counterfactual retirement age, also accounting for specific retirement rules for women and individuals with disabilities. Overall, the reform's costs entail changing pension insurance expenditures as well as foregone social insurance contributions and tax revenues. For each of these dimensions, we calculate actual and counterfactual expenditures and revenues for each individual retiring via the 'retirement at 63' scheme and for each year under consideration. Individual fiscal costs, i.e., the difference between actual and counterfactual costs, are then aggregated and upweighted to match the total number of those retiring through the retirement scheme, taken from official statistics (Deutsche Rentenversicherung Bund, 2018). Our calculations encompass the following aspects:

---

[26] In 2016, 225,290 individuals retired via the 'retirement at 63' scheme (Deutsche Rentenversicherung Bund, 2018).

[27] See e.g. https://www.tagesschau.de/inland/rente-253.html.

Pension insurance expenditures. The retirement reform affects pension insurance expenditures along three dimensions: deductions, retirement timing, and accumulated pension points. Costs are assigned to the relevant fiscal year, assuming that retirement under the reform takes place in the middle of the year.

While retirement benefits under the reform are provided in the data, we impute counterfactual retirement benefits, accounting for changes in retirement timing, the ensuing change in pension points, as well as deductions. The abolition of deductions under the 'retirement at 63' scheme increases benefits of those retiring early, even absent any behavioral response to the reform. In addition, retiring earlier due to the reform results in less accumulated pension points, which reduces retirement benefits. At the same time, claiming benefits early raises fiscal costs in the year of retirement. Naturally, these effects reverse for individuals postponing their retirement in order to become eligible under the reform.

Income tax revenues. The timing of entering retirement and the amount of benefits affect taxable income and thus have direct implications for tax revenues. Furthermore, retirement benefits are only partially taxable. For example, for individuals retiring in 2017, only 74% of pension benefits are included in taxable income. Similar to the changes in pension insurance expenditures, we calculate each individual's actual and counterfactual tax base, which is composed of wages and retirement benefits. Individuals postponing their retirement in the counterfactual scenario are assumed to extend their pre-retirement employment status at their previous wage. Taxes are simulated by applying the German individual income tax schedule to taxable income. As we lack comprehensive data on partners' taxable income as well as on other income sources such as investment or rental income, we are nevertheless only able to provide a rough approximation of the reform's impact on tax revenues.

Social security contributions. Entering retirement likewise affects social security contributions. While employed individuals' employer and employee social security contributions amount to about 40%[28], contributions of retirees only correspond to roughly 18% of their retirement benefits. Considering that the public pension insurance covers about half of retirees' health insurance fees, retirees are only liable to contribute the remainder of about 10.5%. These low rates are due to retirees not contributing to public pension and unemployment insurance. At the same time, if retirement benefits increase due to foregone deductions, a small percentage of these costs is compensated by rising health insurance and nursing care insurance contributions. As before, calculations assess actual and counterfactual social security contributions in each year.

[28] Contribution rates slightly differ across years, between those with and those without children, for those with lower wages and by health insurance provider.

Our calculations do not account for second round effects, such as possible increases in social insurance contribution rates to compensate for foregone revenues following the reform. Table 1.11 presents the ensuing fiscal cost estimates. While our calculations suggest that aggregate pension insurance expenditures amounted to 10.4 billion euros between 2014 and 2017 - thus exceeding the fiscal costs projection of the government by more than 3 billion euros - total fiscal costs correspond to 19.8 billion. With estimated pension insurance expenditures of 3.74 billion euros and total fiscal costs of 7.26 billion euros in 2017, our fiscal cost projections are at the upper end of the range of cost estimates. This is also due to the high number of claimants, which had been underestimated when the reform was announced. Yet, costs per claimant lie between projections that assumed unchanged retirement behavior and projections assuming that all claimants would otherwise have retired at the regular retirement age.

Table 1.11: Fiscal costs of the early retirement reform

Costs in billion euros				
	2014	2015	2016	2017
Pension insurance expenditures	0.82	2.31	3.51	3.74
Total costs	1.51	4.42	6.64	7.26

Notes: This table shows annual fiscal cost estimates in billion euros. Estimations are based on counterfactual retirement choices based on coarsened exact matching of eligible and ineligible individuals. Total costs encompass pension insurance expenditures, income taxes and social security contributions.

1.7 Conclusion

This paper assesses the responses to a recent German pension reform that introduced incentives for retiring early. While the retirement age has been gradually increasing over time for most prospective retirees, the reform enabled individuals with 45 or more contribution years to retire early at age 63 instead of 65 without incurring a financial penalty. Exploiting high-quality administrative data from the German public pension insurance, we both employ an event study design and estimate the effect of becoming eligible for the early retirement scheme in a coarsened exact matching approach. In supplementary analyses, we make use of the SHARE-RV survey and study the effect of the reform on the reference age for early retirement as well as the role of potential determinants for claiming early retirement such as health conditions and private retirement savings.

Our results indicate that the probability of retiring early increases by roughly 10 percentage points in the month of becoming eligible relative to the counterfactual probability of retiring at the same age with deductions. This translates into a one percent increase in benefits increasing the probability of retirement entry by 1.09 percentage points. Responses are equally large if we restrict the sample to individuals who do not become eligible at age 63 or up to age 64. Even though in particular age 63 may serve as focal point for early retirement due to the scheme being commonly known as 'retirement at 63' – our survey analysis lends support to this hypothesis – becoming eligible at non-focal ages elicits similar responses.

Individuals eligible for the reform retire on average 6.4 months earlier than non-eligible individuals with identical characteristics. This means that reducing deductions by one percentage point decreases the average retirement age by 0.74 months. The effect is larger for men than for women and particularly large for individuals who have been working or voluntarily contributing to public pension insurance prior to becoming eligible. Our results based on the SHARE-RV survey do not suggest that responses are driven by poor health conditions or liquidity constraints.

We subsequently use our matching methodology to quantify fiscal costs. With additional pension insurance expenditures of 10.4 billion euros and aggregate fiscal costs of 19.8 billion euros between 2014 and 2017, our cost estimates exceed most previous back-of-the-envelope cost estimates including those presented in the draft government bill. The latter assumed additional pension insurance expenditure of just 7 billion euros over the period 2014–2017, which highlights the policy-relevance of precisely estimated behavioral parameters for ex-ante fiscal cost projections.

Our estimated elasticities are fairly large compared to the literature. In our view, there are several potential factors that may play a role. First, the reform analyzed in this paper was very salient. The large spike in the event study analysis in the month of becoming eligible is indicative that information or other frictions are negligible in the context of the 'retirement at 63' scheme. Second, the size of the financial incentive is very large as eligible individuals can retire early without any deductions. Third, financial incentives may be a more important driver of retirement choices when pathways into early retirement are opened rather than closed. While previous papers have shown that social norms and reference ages can be crucial for retirement choices in many settings, we find that responses to the reform are equally large even if they do not coincide with retirement ages framed as reference points. This suggests that, in our context, retirement choices may indeed be mainly influenced by financial incentives.

2 Who Bears the Burden of Real Estate Transfer Taxes? Evidence from the German Housing Market

2.1 Introduction

In many countries, taxes on real estate transfers are an important source of public sector revenue. However, they are often criticized for creating large distortions, reducing the number of transactions in the housing market. At the same time, transfer tax holidays are widely seen as an effective measure to stimulate the economy because there are strong behavioral reactions to these tax cuts, mostly regarding the timing of house purchases.[1]

This paper focuses on the effects of permanent changes in real estate transfer taxes, which are understudied in the literature. Most existing contributions focus on transitory tax changes or discontinuities in the tax schedule. Effects of permanent increases in transfer taxes are important because growing international mobility of both capital and people may increase pressure to raise more revenue from land and real estate. If that happens, a key question is who bears the tax burden. Is the tax capitalized into house prices so that those who own the house when the reform happens effectively pay the tax? Or do the buyers, who actually remit the tax, bear the burden? To provide answers to these questions, we combine theoretical modeling with empirical analyses. In the first part of the paper, we develop a simple and very stylized overlapping generations model where the price effects of transfer taxes depend (i) on the distribution of bargaining power between sellers and buyers as well as (ii) on the likelihood that the buyers will resell the house later. We use this model to derive hypotheses, which we then put to an empirical test. In the second part of the paper, we exploit a reform of the German federal fiscal system in 2006 that gave the German states the right to set the rate of the real estate transfer tax (RETT) to study the price effect of RETT rate changes. Before the reform, there was a nationwide uniform tax rate of 3.5%. After the reform, most states increased their tax rates, albeit at different points in time. Today, the highest tax rates are equal to 6.5%. Revenue from the RETT is significant. In 2005, just before the reform, revenue from this tax

---

This chapter is joint work with Mathias Dolls, Clemens Fuest and Florian Neumeier and circulates as Dolls et al. (2019).

[1] For instance, in the crisis of 2008-09, the UK government reduced the stamp duty tax on land to stimulate spending on house renovation. Best and Kleven (2018) find that the boost in spending caused by the temporary tax reduction was as large as the tax cut itself.

was just 2.7% of overall state tax revenue in Germany. In 2018, this number was equal to 4.5%. This increase partly reflects rising real estate prices in Germany since the financial crisis. In our empirical analysis, we utilize the variation in RETT rates across German states and over time to investigate the impact of RETT on house prices for different types of properties. To this end, we use a unique dataset covering roughly 18 million properties offered for sale over the period from 2005 to 2018. The data was collected by analyzing real estate advertisements from 140 different sources, including property portals such as ImmobilienScout24.de, as well as regional and trans-regional newspapers. Our dataset includes a large number of property characteristics, such as the asking price, the first and the last day the property was listed, floor size, the construction year, as well as several amenity features.

Our paper contributes to the literature studying the effect of property transfer taxes on house prices (Dachis et al., 2012; Besley et al., 2014; Kopczuk and Munroe, 2015; Slemrod et al., 2017; Best and Kleven, 2018). In contrast to the present paper, a large part of the existing literature focuses on temporary tax changes and tax holidays. The estimated price effects of property transfer taxes reported in this literature vary notably, suggesting that the price effects may depend on the institutional setting. Besley et al. (2014), for instance, exploit the UK stamp duty tax holiday 2008-2009 and find that roughly sixty percent of the tax relief accrued to buyers. Similarly, bunching results in Slemrod et al. (2017) for a transfer tax in Washington D.C. suggest that the burden of the transfer tax is equally split between buyers and sellers. In contrast, Dachis et al. (2012) estimate that the introduction of a (permanent) transfer tax in Toronto in 2008 led to a price reduction equal to the tax, suggesting that sellers bear the burden of the tax.[2] Kopczuk and Munroe (2015) and Best and Kleven (2018) study transfer taxes in New York/New Jersey and in the UK, respectively, and provide similar evidence. The authors find that price responses to tax thresholds exceed the size of the tax change. What is more, Kopczuk and Munroe (2015) show that a transfer tax notch even leads to an unraveling of the market in a certain price range.

Recent studies have also assessed the incidence of real estate transfer taxes in the German real estate market. However, these studies are mainly based on aggregate data. Using annual data at the state level, Petkova and Weichenrieder (2017) assess the effect of RETT increases on transaction prices and transaction volumes for single-family houses and apartments. For houses and vacant lots, the authors find that the number of transactions declines, while prices are not significantly affected. For apartments, however, the authors observe negative price

[2] This finding refers to the tax on one transaction. Since the tax increase will also apply to future transactions, the true burden implied by the tax increase is likely to be larger, as we will discuss further below, a point also made by Petkova and Weichenrieder (2017).

effects, but no effect on the number of transactions. In a similar vein, Fritzsche and Vandrei (2019) find negative effects of RETT increases on monthly state-level family home transaction volumes. Their results also indicate substantial anticipation effects of RETT reforms. Focusing on commercial property, Baudisch and Dresselhaus (2018) find both a decline in transactions and a decline in prices following a tax increase. In contrast to previous studies relying on state-level data, we exploit a micro-level data set on real estate prices, which allows for a more credible identification of price effects.

The literature also discusses the importance of different channels through which transfer taxes may affect house prices and transactions. Best and Kleven (2018) consider a model with downpayment constraints and leverage to rationalize large prices responses. Slemrod et al. (2017) study optimization frictions in a bargaining model. Their model allows both buyers and sellers to bear remittance responsibilities and accounts for the possibility that the seller may make quality adjustments to her house prior to the sale. Other bargaining models can be found in Besley et al. (2014) and Kopczuk and Munroe (2015). In the model proposed in the present paper, the price effects of transfer taxes depend on the bargaining power of the sellers vis-à-vis the buyers as well as the likelihood that the buyers will resell the house. In this framework, given that a tax increase is perceived as permanent, price effects are predicted to be larger for properties which are traded more often. In general, the model predicts that the price effects are likely to be larger than the tax change for a single transaction. Our empirical results are in line with the predictions of our model. The price effects we find are larger than those reported in earlier studies. On average, a one percentage point increase in the tax rate reduces property prices by 3% within a year after the tax reform. One interpretation of this result is that the growing tax burden on future transactions is capitalized into house prices. An alternative explanation for the large price effect is that buyers are crowded out of the market through downpayment constraints as emphasized by Best and Kleven (2018). As taxes are usually not mortgageable, an increase in the tax burden limits which houses downpayment-constrained households are able to afford.

The remainder of the paper is structured as follows: The next section describes the institutional background of the RETT in Germany. In section 2.3, we present a simple housing market model which motivates our empirical analysis and facilitates the interpretation of our results. Section 2.4 presents the data. In section 2.5, we describe our empirical approach and section 2.6 shows the results. Section 2.7 concludes.

2.2 Institutional Background

The RETT is an important source of revenue for the German states. With a revenue of 14.1 billion euros in 2018, which corresponds to 4.5% of state level tax revenues, its weight as a source of revenue is limited, but its importance is due to fact that it is the only significant tax where the states can set the tax rate.[3] The RETT is charged on the purchase price on all kinds of real estate, including residential and commercial properties as well as vacant lots.

Table 2.1: Real estate transfer tax rate changes

State	Initial Tax Rate	Date of Increase	New Tax Rate	First Legal Draft
Baden-Württemberg	3.5%	05.11.2011	5.0%	13.09.2011
Bavaria	3.5%	-	-	-
Berlin	3.5%	01.01.2007	4.5%	07.11.2006
		01.04.2012	5.0%	18.01.2012
		01.01.2014	6.0%	10.10.2013
Brandenburg	3.5%	01.01.2011	5.0%	13.09.2010
		01.07.2015	6.5%	04.03.2015
Bremen	3.5%	01.01.2011	4.5%	22.06.2010
		01.01.2014	5.0%	09.07.2013
Hamburg	3.5%	01.01.2009	4.5%	14.10.2008
Hesse	3.5%	01.01.2013	5.0%	25.09.2012
		01.08.2014	6.0%	13.05.2014
Lower Saxony	3.5%	01.01.2011	4.5%	31.08.2010
		01.01.2014	5.0%	17.09.2013
Mecklenburg-Western Pomerania	3.5%	01.07.2012	5.0%	14.02.2012
North Rhine-Westphalia	3.5%	01.10.2011	5.0%	10.05.2011
		01.01.2015	6.5%	28.10.2014
Rhineland-Palatinate	3.5%	01.03.2012	5.0%	23.11.2011
Saarland	3.5%	01.01.2011	4.0%	19.10.2010
		01.01.2012	4.5%	18.10.2011
		01.01.2013	5.5%	08.10.2012
		01.01.2015	6.5%	07.10.2014
Saxony	3.5%	-	-	-
Saxony-Anhalt	3.5%	02.03.2010	4.5%	30.09.2009
		01.03.2012	5.0%	28.09.2011
Schleswig-Holstein	3.5%	01.01.2012	5.0%	23.08.2010
		01.01.2014	6.5%	26.07.2013
Thuringia	3.5%	07.04.2011	5.0%	06.01.2011
		01.01.2017	6.5%	23.09.2015

[3] There is more tax autonomy at the local level. Local governments can set the property tax rate and the rate of the local business tax. The rates of the most important revenue sources, the income tax and the value added tax, are set at the federal level. Through the second chamber, the states participate in decisions regarding income and value added tax rates, and they receive a share of the revenue.

Before 2006, the tax rate was uniform across all states and equal to 2% prior to 1997 and to 3.5% until 2006. In 2006, a substantial constitutional reform permitted the states to set their own RETT rates. With the exception of Bavaria and Saxony, all states have increased their tax rates since, often multiple times (see Table 2.1). So far, no state has ever reduced its tax rate.

As shown by Büttner and Krause (2018), the German fiscal equalization scheme sets strong financial incentives for states to raise their RETT. Moreover, the German public debt ceiling ('debt brake') requires state governments to achieve structurally balanced budgets from 2020 onwards, which may explain why the need for budget consolidation is the most frequent official justification of RETT increases (Fritzsche and Vandrei, 2019).

2.3 A Simple Model of a Housing Market with Transfer Taxes

We consider a highly stylized model of an economy with overlapping generations. There are two types of agents, the young (Y) and the old (O). All agents live for two periods; they are young in the first period and old in the second. The number of households in each generation is normalized to unity. There is a stock of two units of housing in the economy. For simplicity we abstract from depreciation of housing capital and construction.

The utility for the young (old) of owning a house while young (old) is given by U^Y(U^O). Ownership of a house may or may not imply that a household actually occupies a house. There is a perfectly competitive rental market which makes sure that all households live somewhere. For the purposes of our analysis we do not need to model this market explicitly. We consider a housing market with frictions. At the beginning of each period, a fraction $0 < q < 1$ of the young enters the housing market.[4] Only old households consider selling a house. Each young household in the market is matched with an old household.

If no trade takes place, the old agent keeps the house while old and passes it on to the next generation, which generates a utility for the old household denoted by U^O.[5] The reservation

[4] A standard way of modeling frictions would be to assume that a share of $1 - q$ young households is liquidity constrained. A limitation of our model is that we do not endogenize q, which implies that changes in transfer taxes in our model do not influence the number of transactions. We make this assumption because our empirical analysis focuses on effects of tax changes on prices, not on the number of transactions. The main objective of our theoretical analysis is to highlight specific factors which are likely to affect the impact of tax changes on prices.

[5] This may or may not include the utility of the old of passing on a property to the next generation. Note that in equilibrium, at the beginning of each period before transactions take place, only the old households own houses, which is why only they can sell houses. After transactions have taken place, the old households still own $2 - q$ units of housing. At the end of period two, the old households die and the houses owned by the old are inherited by the next generation of old households.

utility of the young households is equal to zero. If a transaction takes place, the buyer pays a transfer tax equal to T percent of the house price.

It is straightforward to determine the equilibrium house price. When the young negotiate, they take into account that they will sell the house with probability q when they are old. With probability $1-q$ they will keep and use the house while old, so that the present value of the surplus from buying the house is given by

$$U^Y + \frac{q}{(1+\rho)} p_{t+1} + (1-q)\frac{U^O}{(1+\rho)} - p_t(1+T) \tag{2.1}$$

where t is the period index and ρ is the discount rate. The surplus of the old agent from selling is simply given by $p_t - U^O$.
The equilibrium house price in period t is thus given by maximizing the Nash maximand

$$\beta \ln\left(U^Y + \frac{q}{(1+\rho)} p_{t+1} + (1-q)\frac{U^O}{(1+\rho)} - p_t(1+T)\right) + (1-\beta)\ln(p_t - U^O) \tag{2.2}$$

over p_t, which yields

$$p_t^*(1+T) = \beta U^O(1+T) + (1-\beta)\left(U^Y + \frac{q}{(1+\rho)} p_{t+1} + (1-q)\frac{U^O}{(1+\rho)}\right). \tag{2.3}$$

Our analysis focuses on the house price effects of changes in the transfer tax T which are perceived as permanent. It is therefore sufficient to consider the tax effect on prices in the steady state, where prices are the same in each period in this stationary model. The steady state house price is given by

$$p^* = \left(1 + T - \frac{(1-\beta)q}{(1+\rho)}\right)^{-1}\left[\beta U^O(1+T) + (1-\beta)\left(U^Y + (1-q)\frac{U^O}{(1+\rho)}\right)\right]. \tag{2.4}$$

Denote the semi-elasticity of the house price with respect to the tax rate by $\varepsilon \equiv \frac{dp^*}{dT}\frac{1}{p^*}$. Consider first the two polar cases $\beta = 1$ (buyer has all the bargaining power) and $\beta = 0$ (seller has all the bargaining power). If the buyer has all the bargaining power it follows directly from (2.4) that $\varepsilon = 0$. Since the seller is always reduced to her reservation utility and the house price is the net of tax price, changes in T are always fully borne by the buyer and the house price does not change. In the opposite polar case, where the seller has all the bargaining power ($\beta = 0$), we get

$$\varepsilon = -\frac{1}{1 + T - \frac{q}{(1+\rho)}}. \tag{2.5}$$

Equation (2.5) yields various important insights. First, if q converges to zero, which implies that buyers do not expect further transactions during their lifetime, a one percentage point increase in the transfer tax ($dT = 0.01$) reduces the price by approximately one percent. But if q is positive, the decline in the price will be larger than one percent because the tax increase is also expected to be a burden on future transactions.

Unsurprisingly, the impact of future transactions is stronger, the lower the discount rate. Moreover, we should expect the price effects of a permanent increase in transfer taxes to be higher, the more likely it is that the property will be traded more than once over the relevant time horizon. In particular, we may then observe that a one percentage point increase in the real estate transfer tax reduces house prices by more than one percent. Consider finally the general case $0 < \beta < 1$, where:

$$\varepsilon = \left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-1} \left[\frac{\beta U^O}{p^*} - 1\right] < 0 \tag{2.6}$$

As we show in the appendix, equation (2.6) defines ε as a function of β and q, that is $\varepsilon = \varepsilon(\beta, q)$, with $\frac{\delta\varepsilon}{\delta\beta} > 0$ and $\frac{\delta\varepsilon}{\delta q} < 0$. For our empirical analysis, this implies that we would expect to see (i) a smaller price reduction in response to a tax increase in transactions where buyer bargaining power is higher and (ii) a larger price reduction in cases where the traded property is expected to be traded more frequently in the future.[6]

2.4 Data and Descriptive Statistics

In our empirical analysis we use a novel and large dataset on the German real estate market provided by F+B, a commercial real estate consultancy firm. The dataset covers roughly 18 million properties that were offered for sale in Germany during the period from January 2005 until December 2018. The dataset was created by analyzing real estate advertisements from 140 different sources, including online property portals, regional and trans-regional newspapers, as well as real estate agencies, using web-scraping techniques. The raw data was thoroughly cleaned to make sure that properties that were listed in more than one source at the same time only appear once in the final dataset. For all properties included in the final dataset, we know the first day the property was listed as well as the last day it was listed. Moreover, the final dataset includes the complete list of sources in which the property was advertised.

[6] Of course, the expected number of future transactions will itself be a function of the transfer tax. In the simple model considered here, the number of future transactions is exogenous because our empirical analysis focuses on price effects, not the quantity of housing transactions.

The dataset contains three price variables: the offering price of the property on the day it was first listed, the offering price on the last day the property was listed, and a proxy for the actual selling price of the property, which is equal to the offering price on the last day of the listing minus an estimated deduction. F+B estimates this deduction based on matching a subsample of the advert data to actual transaction data. In our analysis, we primarily focus on the final offering price. Note that we drop properties from our dataset in case the offering price (i) increased by more than factor two or (ii) decreased by more than 50% during the posting period since we are concerned that offering prices of those properties do not reflect market prices. Moreover, the dataset covers a wide range of property characteristics, such as floor space, the number of rooms, the construction year, as well as binary indicators for equipment and locational features, and the postal code of the property. The data is available for three different property types: apartments, single-family houses, and apartment buildings. In our empirical analysis, we study the price effects of a change in the real estate transfer tax separately for each property type.

Table 2.2 shows the sample means of important property characteristics separately for (i) the three different property types as well as (ii) three different time periods: 2005-2009, 2010-2014, and 2015-2018. A glance at the price variables suggest that property prices have increased notably over the past years. I.e., between 2015 and 2018, the average price per square meter for an apartment (single-family house) was roughly 650 euros (400 euros) higher than it was between 2010 and 2014. This corresponds to a price increase of about 34% (24%).

Table 2.2: Real estate data: Summary statistics by time period

	Apartments			Single-family Houses			Apartment Buildings		
	2005-09	2010-14	2015-18	2005-09	2010-14	2015-18	2005-09	2010-14	2015-18
First asking price	1,785	1,933	2,584	1,600	1,595	1,975	1,005	1,022	1,320
Last asking price	1,767	1,925	2,578	1,584	1,587	1,968	990	1,014	1,314
Floor size	92.6	94.3	102.4	150.0	152.0	153.3	385.5	312.3	325.5
Rooms	3.1	3.1	3.2	4.8	5.1	5.0	7.4	8.4	9.0
Construction year	1979	1979	1982	1981	1978	1980	1954	1952	1954
Kitchen	0.33	0.41	0.31	0.17	0.22	0.25	0.11	0.17	0.21
Parking spot	0.61	0.63	0.65	0.57	0.60	0.59	0.58	0.64	0.67
Garden	0.25	0.23	0.28	0.32	0.32	0.44	0.31	0.28	0.41
Balcony	0.47	0.37	0.43	0.31	0.24	0.39	0.33	0.23	0.40
Basement	0.28	0.46	0.46	0.35	0.40	0.42	0.31	0.39	0.45

Notes: The table shows the average realizations of different property characteristics for different property types and across different time periods. Floor space is measured in square meters. Asking prices refer to the price per square meter.

A closer inspection of our data reveals that by far the largest fraction of the properties included in our dataset (i.e., more than one-third) were advertised on the online property portal ImmobilienScout24.de, which is by far the largest online property portal in Germany. To check whether properties listed on ImmobilienScout24.de differ from those advertised in other outlets, we compare the characteristics of properties listed on ImmobilienScout24.de to the characteristics of properties that were solely listed in other outlets. The results are shown in Table 2.3.

Table 2.3: Real estate data: Summary statistics by data source

	Apartments		Single-family Houses		Apartment Buildings	
	IS24	Other source	IS24	Other source	IS24	Other source
First asking price	2,173	2,000	1,773	1,639	1,197	1,091
Last asking price	2,160	1,991	1,761	1,630	1,187	1,084
Floor size	94.1	97.5	148.8	153.9	312.3	336.3
Rooms	3.2	3.0	5.2	4.7	10.5	7.3
Construction year	1979	1980	1983	1977	1954	1953
Kitchen	0.39	0.33	0.14	0.26	0.10	0.22
Parking spot	0.68	0.59	0.62	0.56	0.69	0.61
Garden	0.28	0.22	0.31	0.38	0.30	0.35
Balcony	0.48	0.37	0.28	0.33	0.29	0.32
Basement	0.50	0.32	0.45	0.34	0.51	0.33

Notes: The table shows the average realizations of different property characteristics for different property types separately for properties listed on immobilienscout24.de vs. properties listed in other sources. Floor space is measured in square meters. Asking prices refer to the price per square meter.

The descriptive statistics in Table 2.3 indicate that properties listed on ImmobilienScout24.de do not appear to be representative of the German property market. On average, properties listed on ImmobilienScout24.de are more expensive than properties solely listed in other outlets. Also, the characteristics of properties advertised on ImmobilienScout24.de differ from the characteristics of properties listed in other sources. For instance, apartments advertised on ImmobilienScout24.de appear to be smaller, but are more likely equipped with a kitchen, a parking spot, a garden, a balcony, and a basement. This accentuates our dataset's higher degree of representativity compared to web-scraped Immobilienscout24.de data used by other studies on the German real estate market.

2.5 Empirical Strategy

We employ an event study design to assess the impact of changes in real estate transfer tax rates on residential property prices. For each property type, i.e., apartments, single-family houses, and apartment buildings, we estimate the following model:

$$ln(p)_{i,c,t} = \sum_{j=-12}^{23} \beta_j D_{c,t-j} + \nu X_i + \mu_c + \varsigma_{c,t} + \epsilon_{i,c,t} \tag{2.7}$$

Index i refers to the property, c to the postal code area the property is located in, and t to the month it was offered for sale. The dependent variable is the log of the property price per square meter. $D_{c,t-j}$ are monthly event study indicators for real estate transfer tax rate changes. The event window runs from 12 months prior to the tax change to 24 months after the tax change.[7] End points are adjusted in line with Schmidheiny and Siegloch (2019). Following Schmidheiny and Siegloch, we mainly focus on the size of the tax rate change $\Delta\tau$ as event study indicator. Additionally, a robustness check employs the change in the log net-of-tax rate as in Fuest et al. (2018). While the first set of indicators assumes a linear relationship between the change in the tax rate and the percentage change in the property price, the latter set of indicators captures the elasticity of property prices with respect to the net-of-tax rate $\eta = (\frac{\Delta p}{p})/(\frac{\Delta(1-\tau)}{1-\tau})$. We include postal code area fixed effects μ_c to account for time-invariant local characteristics that influence property prices. $\varsigma_{c,t}$ is a time-fixed effect for months and years which we interact with a set of four different dummy variables indicating the degree of urbanization. Indicators for the degree of urbanization (*Siedlungsstrukturelle Kreistypen*) are provided by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). That way, we account for the fact that property prices have experienced a stronger increase in urban areas over the last years (Baldenius et al., 2019). Standard errors are clustered at the postal code level.

Deviating from a standard event study setting, we choose $t-4$ as the reference period relative to which the change in property prices is measured. We do so for two reasons. First, the price of a property offered for sale shortly before a tax reform might already reflect the upcoming tax rate change. As it may take several months to complete a property transaction, setting an earlier reference period ensures that prices are compared to a time period in which the preceding tax rate still applies. Second, the median time between the first legal draft and the reforms' implementation amounts to 3.2 months. The 4-month window hence ensures that the pre-trend is not as much driven by announcement effects.

[7] The event study window runs until month 23 as the month of the tax change is coded as 0.

As a robustness check, we additionally control for property characteristics that may influence prices per square meter. The property characteristics that we consider are the same as in Tables 2.2 and 2.3, i.e., floor space, the number of rooms, dummy variables for construction year groups, as well as dummy variables indicating whether the property comes with a kitchen, a parking spot, a garden, a balcony, or a basement. In another robustness check, we also include county-level variables that might be related to regional property market developments. These variables include population figures, per-capita GDP, and the unemployment rate. Note that those variables are only available at an annual frequency. Moreover, per-capita GDP as well as unemployment rates are only available at the county level. The data are provided by the German Federal and the German States' Statistical Offices.

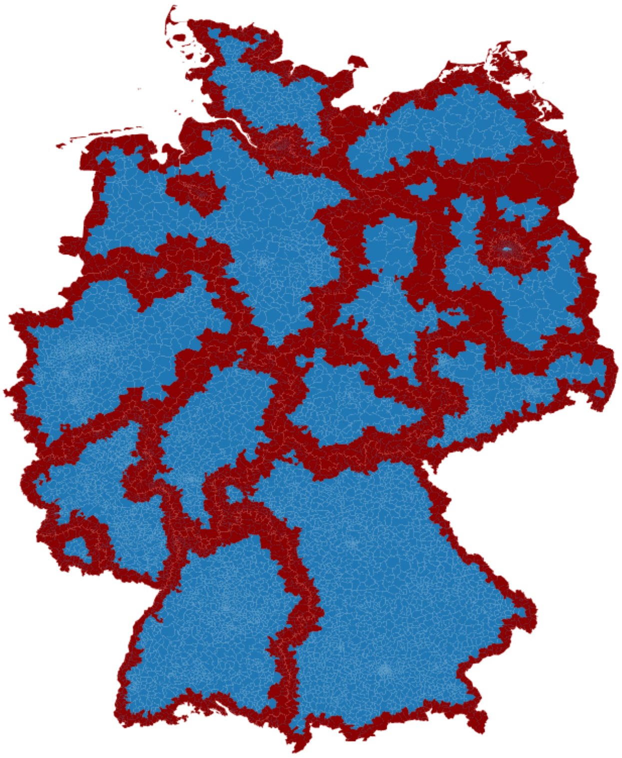
In a final robustness check, we address the concern that effects in border regions may be partially driven by spillover effects. I.e., an increase in a state's tax rate may shift demand to border regions in neighboring states, which might result in higher prices in the control group. This could lead to an overestimation of price effects. We therefore estimate a specification without observations in the vicinity of a border. More precisely, we exclude postal code areas that either directly adjoin a state border, or for which the postal code's centroid is located at a distance of up to 10 kilometers to the border. Figure 2.1 indicates which postal codes areas are excluded in this specification.

Our theoretical model highlights the importance of transaction frequencies as well as the distribution of bargaining power between the seller and the buyer of a property for the price effect of the RETT. Unfortunately, we neither directly observe the intended holding period of a specific property nor the distribution of bargaining power between buyers and sellers. However, we approximate both factors using two different municipality-level indicators. As a proxy for transaction frequencies, we use the fraction of property advertisements relative to the number of residential properties in a municipality. Data on the number of residential properties is taken from the 2011 census. We then compute transaction frequency quartiles within each state and estimate the event study separately for each quartile.

We approximate the bargaining power of sellers using population growth between 2005 and 2017. Arguably, the more people move to a municipality, the higher the demand for houses, implying more bargaining power for the seller. We again compute quartiles within states and estimate our model separately for each quartile.

Note that when assessing the importance of transaction frequencies and bargaining power for the price effects of RETT, we restrict our analysis to single-family houses. The reason is that apartments and apartment buildings are often considered an investment and, consequently, bought by institutional investors, which is why population growth may only represent a

Figure 2.1: Postal codes in the vicinity of state borders



Notes: This figure shows all German postal code areas, distinguished by their distance to state borders. Red areas indicate postal codes that either directly adjoin a state border or whose centroid is located at a distance of up to 10 kilometers to a border.

poor proxy for bargaining power. Likewise, if a property is bought as an investment, the link between the transaction frequency and the fraction of properties offered for sale relative to the property stock should be less strong. In contrast, the vast majority of single-family houses is owner-occupied, which is why our proxies for transaction frequencies and bargaining power should be better suited for this analysis. Also note that when constructing the proxy for transaction frequencies, we omit all municipalities that have less than 1,000 buildings.[8]

2.6 Results

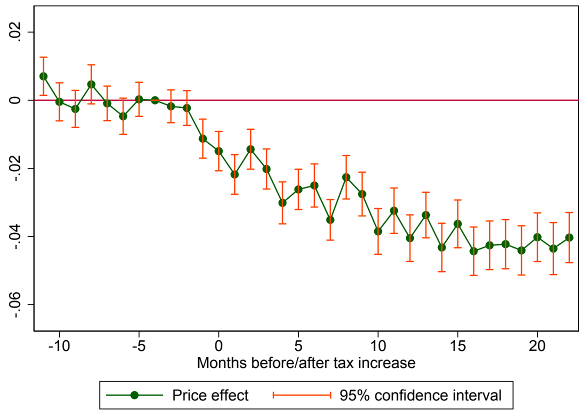
2.6.1 The Effect of RETT on Property Prices

Figure 2.2 displays the results for the baseline specification in which we employ the change in the RETT rate $\Delta\tau$ as event study indicator.

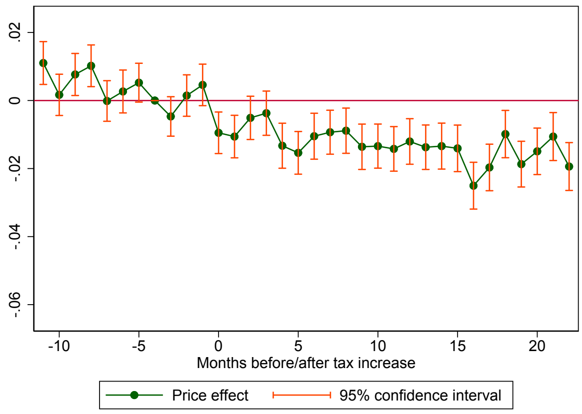
[8] Without a minimum municipality size, very small municipalities might be assigned to high transaction frequency categories if just few buildings are on offer.

Figure 2.2: Baseline: Effects of changes in the RETT rate $\Delta\tau$

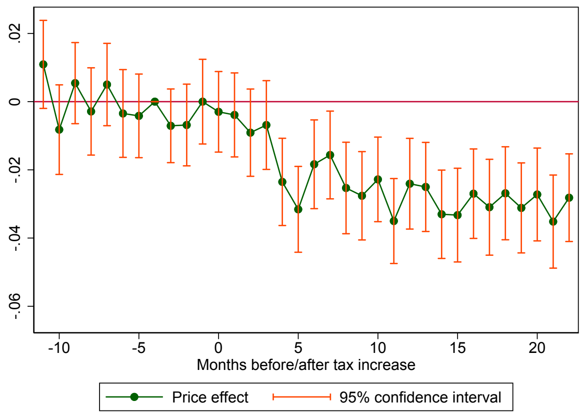
(a) Apartments



(b) Single-family houses



(c) Apartment buildings



Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands for the baseline event study specification. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level.

For all three property types considered in our analysis, prices start to drop immediately after the tax hike becomes effective, whereas the magnitude of the response increases (in absolute terms) over time.[9] Apartment prices gradually decline until the price response reaches a minimum at around -0.04 after one year. This indicates that an increase in the tax rate by one percentage point reduces the price of apartments by up to 4%. For single-family houses and apartment buildings, the price response is somewhat smaller – -0.015 and -0.03, respectively –, and reaches its minimum more quickly. Nevertheless, for all three property types, the price decrease exceeds the increase in the tax burden. Note that this finding is well in line with our theoretical model. The larger price effect for apartments compared to single-family houses and apartment buildings may be due to a shorter average holding period: While houses are mainly bought by families who plan to live in the property for many years (and may even have a bequest motive), apartments are more frequently bought by investors who may intend to resell the property at some point in time (Petkova and Weichenrieder, 2017; Deutsche Bundesbank, 2018).

Note that we observe a small pre-trend for single-family houses, although the post-reform effect rapidly stabilizes at a rather constant level. Also, price effects fluctuate prior to the reform. This could partly be driven by anticipation effects: Depending on the state and the reform, a tax increase may have been announced just two months or an entire year prior to the reform. This could have induced anticipatory responses at different points in time preceding the respective reforms.

Over the last decades, we have seen a very heterogeneous development of property prices in Germany. Some large German cities, such as Munich or Berlin, as well as some more rural areas in their vicinity, have experienced a rapid increase in property prices over the last decades, considerably driven by a substantial growth in population size. At the same time, there are some predominately rural areas in Germany that suffer from a population drain, leading to declining property prices. Although our baseline specification accounts for heterogeneous time trends across regions characterized by different degrees of urbanization, we are still concerned that our results might be affected by some outliers that have experienced extreme migration patterns during our sample period. In order to address this concern, we winsorize our sample according to municipal population growth between 2005 and 2017.[10] I.e., we drop all municipalities with a population growth rate that is smaller than the population-weighted

[9] Results are virtually identical when using F+B's proxy variable for actual transaction prices as a dependent variable (not depicted here).

[10] Administrative data on municipal population size was only available until 2017 at the time of writing the paper.

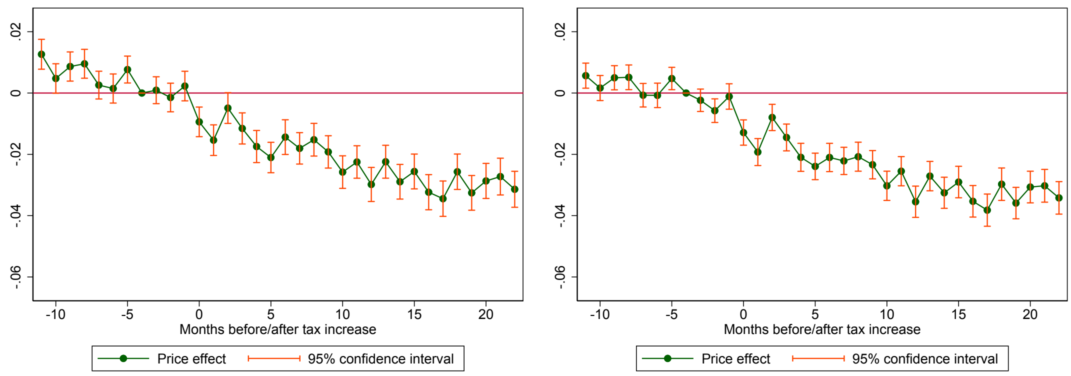
5% quantile or larger than the population-weighted 95% quantile of the population growth rate.[11]

Results are depicted in Figure 2.3. While the coefficients are very similar to those illustrated in Figure 2.2, no significant pre-trend remains once municipalities are dropped that exhibit extreme population growth rates. The decline in prices starts in the three months prior to the reform. This is in line with our expectations and reflects the importance of anticipation effects. If a property is offered for sale shortly before the RETT rate change becomes effective, it is unlikely that the transaction will be completed before the implementation date, implying that the higher RETT rate will apply. Therefore, we already observe a decrease in property prices before the implementation of the reform.

Figure 2.4 estimates the event study on a combined sample of apartments, houses, and apartment buildings. As before, the pre-trend disappears once municipalities with particularly strong growth or decline in population size are excluded. With a coefficient of roughly 3% after one year, the aggregate effect lies, as expected, in between the effects of the individual specifications.

Figure 2.4: Joint estimation for apartments, houses and apartment buildings

(a) Full sample (b) Winsorized

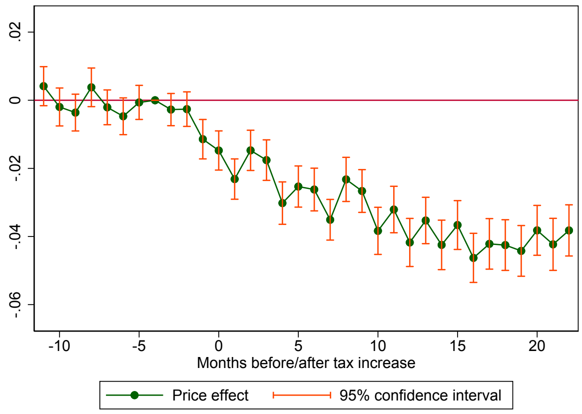


Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands for a joint sample including apartments, houses and apartment buildings. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. Panel (a) considers the full sample. Panel (b) excludes the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017.

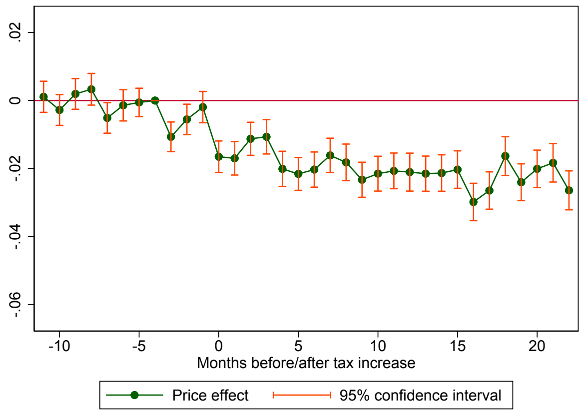
[11] Dropping properties offered for sale in areas that experienced particularly large increases and declines, respectively, in population growth, rather than directly winsorizing based on property prices, ensures that the selection of our sample is not endogenous, that is, related to price changes induced by a decrease in RETT rates.

Figure 2.3: Winsorizing municipal population growth rates

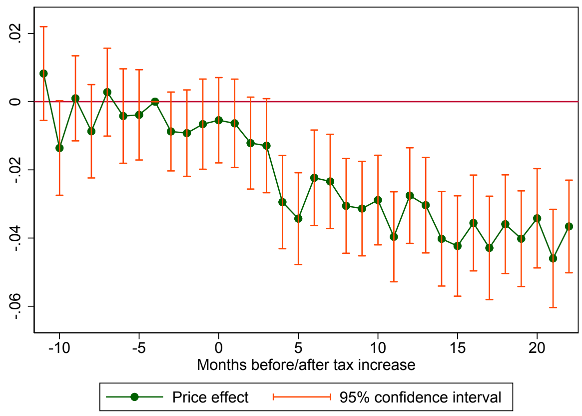
(a) Apartments



(b) Single-family houses



(c) Apartment buildings



Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. This specification excludes the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017.

To check the robustness of our results, we modify our empirical specification in several additional ways. Note that all further robustness checks are based on the winsorized sample. As a first robustness check, we replace the change in the RETT rate by the log net-of-tax rate. A glance at Figure 2.5 reveals that our results remain qualitatively unchanged. For apartments and apartment buildings, the estimated price elasticity with respect to the log net-of-tax rate reaches a maximum of 4, while the estimated elasticity for houses is somewhat smaller.

Second, we add property-specific control variables to equation 2.7. Specifically, we control for the floor space in square meters, the number of rooms, construction year categories, as well as various amenities. Controlling for property characteristics ensures that our findings are indeed due to changes in offering prices and not driven by composition effects. This concern would be relevant if a change in RETT has an effect on the pool of properties that are offered for sale. The results are displayed in Figure 2.6.

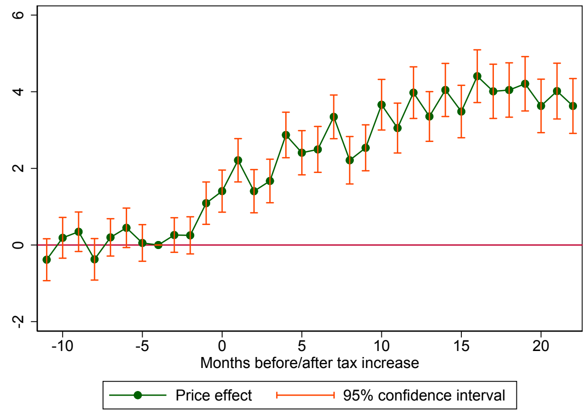
While the figure closely resembles Figure 2.3, the coefficients' magnitude slightly decreases. However, with the confidence bands becoming more narrow, the significance of the coefficients remains unaffected.

In a third robustness check, we add several variables to our empirical model that cover regional housing market conditions. We control for county-level GDP, population size, and the unemployment rate. The results are illustrated in Figure 2.7. Again, we find that our results remain qualitatively unchanged.

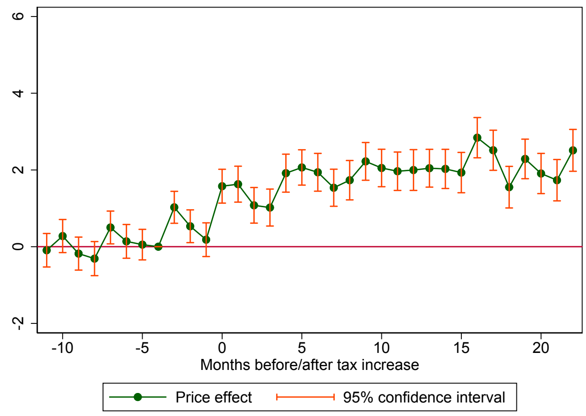
In a final robustness check, we exclude properties located in postal codes in the vicinity of a state border. The reason is that there may be spillover effects of RETT changes into regions that are located close to the border of a state that has implemented the RETT change. I.e., suppose there is a region located in state A and bordering state B. If state B increases the RETT, but state A does not, we may observe an increase in the demand for properties located in that region because of its proximity to state B. Figure 2.8 shows the results. While the estimated price effects for single-family houses and apartment buildings are virtually identical to the ones obtained in our baseline estimation, the magnitude of coefficient estimates slightly increases for apartments. This alleviates the concern that the rather large effect measured in the baseline specification is attributable to spillover effects into border regions of tax-increasing states. Instead, coefficients are slightly more negative than in the baseline. This could also be due to the specific municipalities that are excluded here: amongst others, not considering border regions almost fully excludes the three city states from the sample.

Figure 2.5: Robustness check: Effects of changes in the log net-of-tax rate

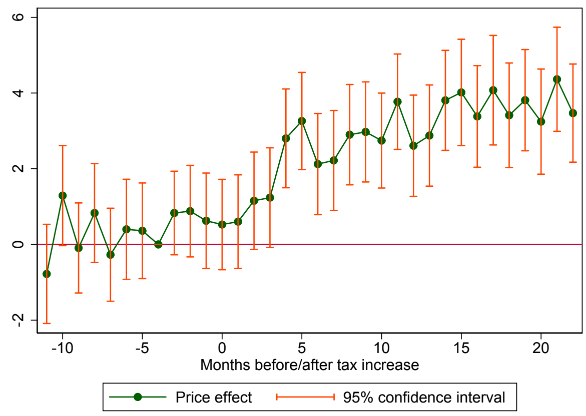
(a) Apartments



(b) Single-family houses



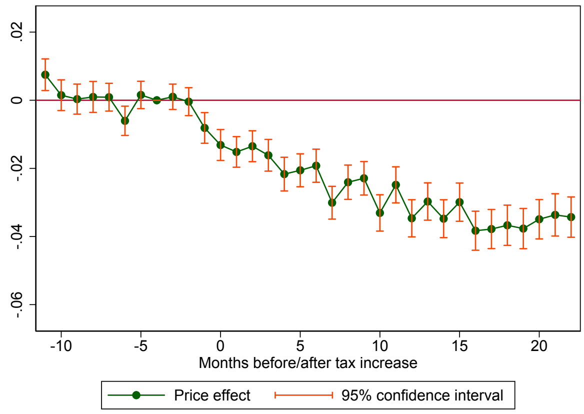
(c) Apartment buildings



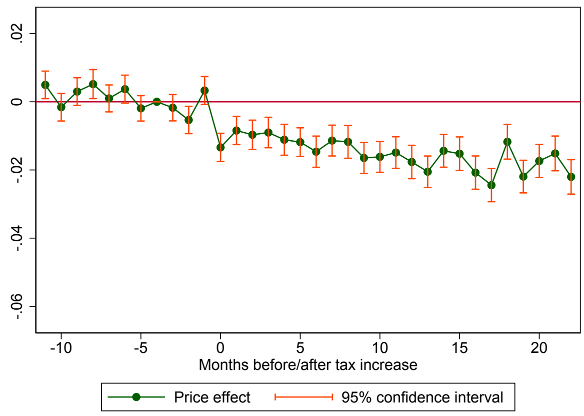
Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the log net-of-tax rate. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level.

Figure 2.6: Robustness check: Property-specific control variables

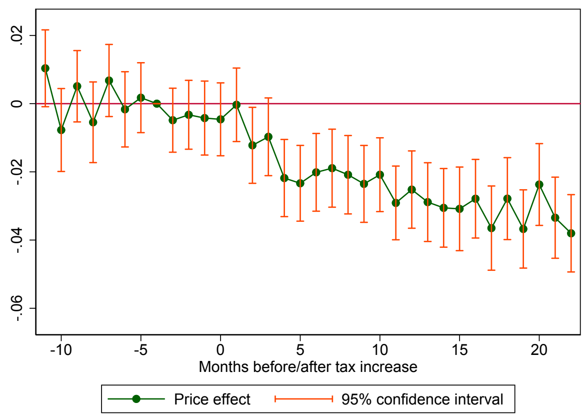
(a) Apartments



(b) Single-family houses



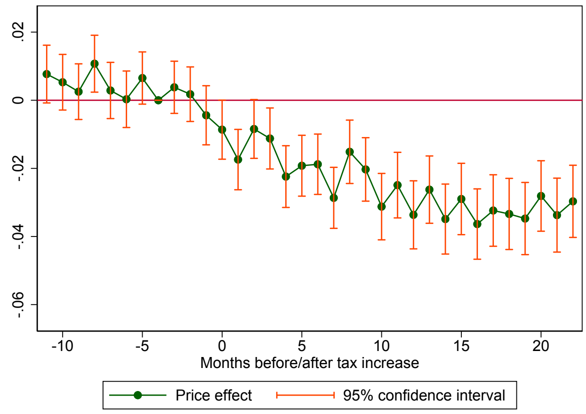
(c) Apartment buildings



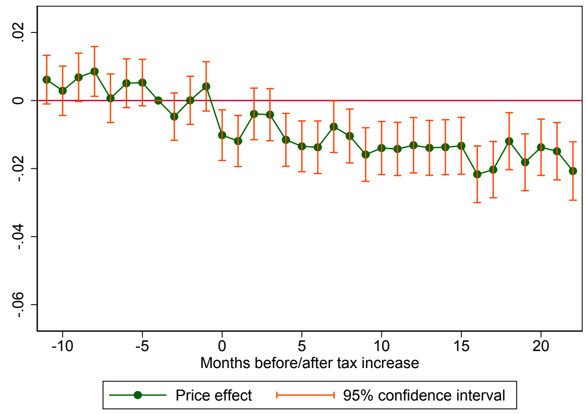
Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. Specifications control for area in square meter, the number of rooms, the construction year, as well as whether the property has a basement, a parking spot, a garden, and a kitchen, and exclude the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017.

Figure 2.7: Robustness check: Regional control variables

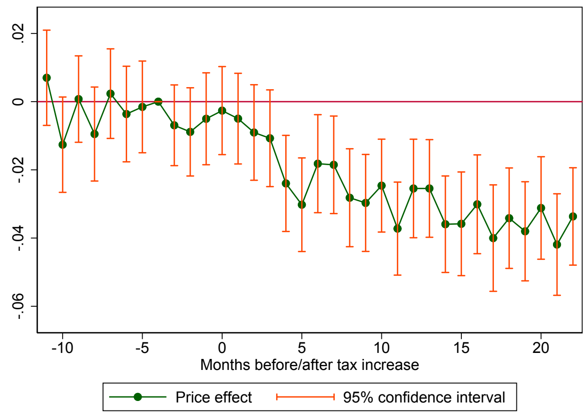
(a) Apartments



(b) Single-family houses



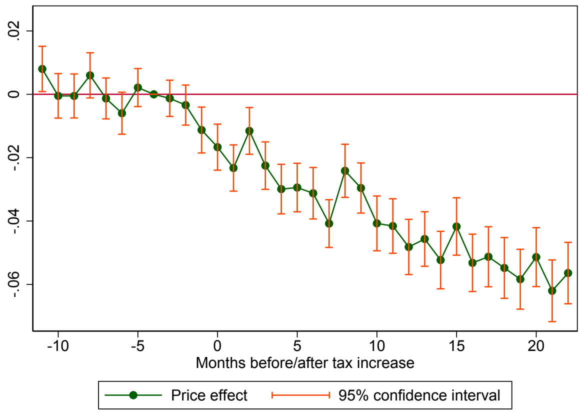
(c) Apartment buildings



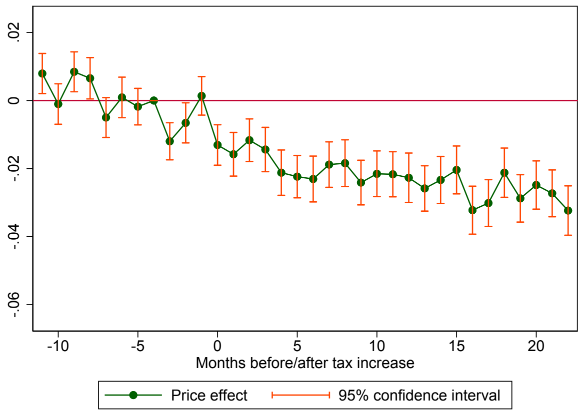
Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. Specifications control for annual county-level population, GDP, and the unemployment rate, and exclude the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017.

Figure 2.8: Robustness check: Without postal codes within 10 km of the border

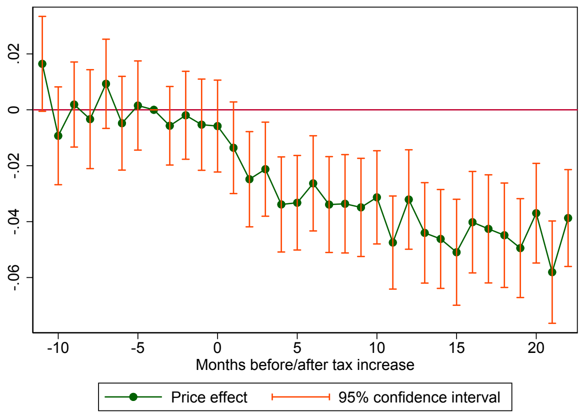
(a) Apartments



(b) Single-family houses



(c) Apartment buildings



Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. Specifications exclude the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017, as well as postal codes that either directly adjoin a border or whose centroid is located at a distance of up to 10 kilometers to the border.

To sum up, our findings indicate that property prices decline by more than the magnitude of the tax increase. A one percentage point change in the RETT reduces the prices of apartments and apartment buildings by roughly 3-4% and house prices by 1.5-2%. While prices start responding immediately to the reform, the response increases in magnitude over the course of the first year after a reform. Responses seem to be more immediate for house prices than for apartment prices.

While in contrast to previous findings for Germany, the observed overshifting is consistent with Best and Kleven (2018), Kopczuk and Munroe (2015) and Davidoff and Leigh (2013), who likewise find a reduction in real estate prices which by far exceeds the increase in the tax rate. Several factors may contribute to this overshifting. First, the capitalization of RETT in a property's future resale value might lower property prices, as described in section 2.3. This might also help explain why effects are substantially larger for apartments than for single-family houses. While single-family houses are frequently bought by individuals who intend to live in their house for many decades, and hence do not pay as much attention to the future resale value, apartments more often serve as an investment property.

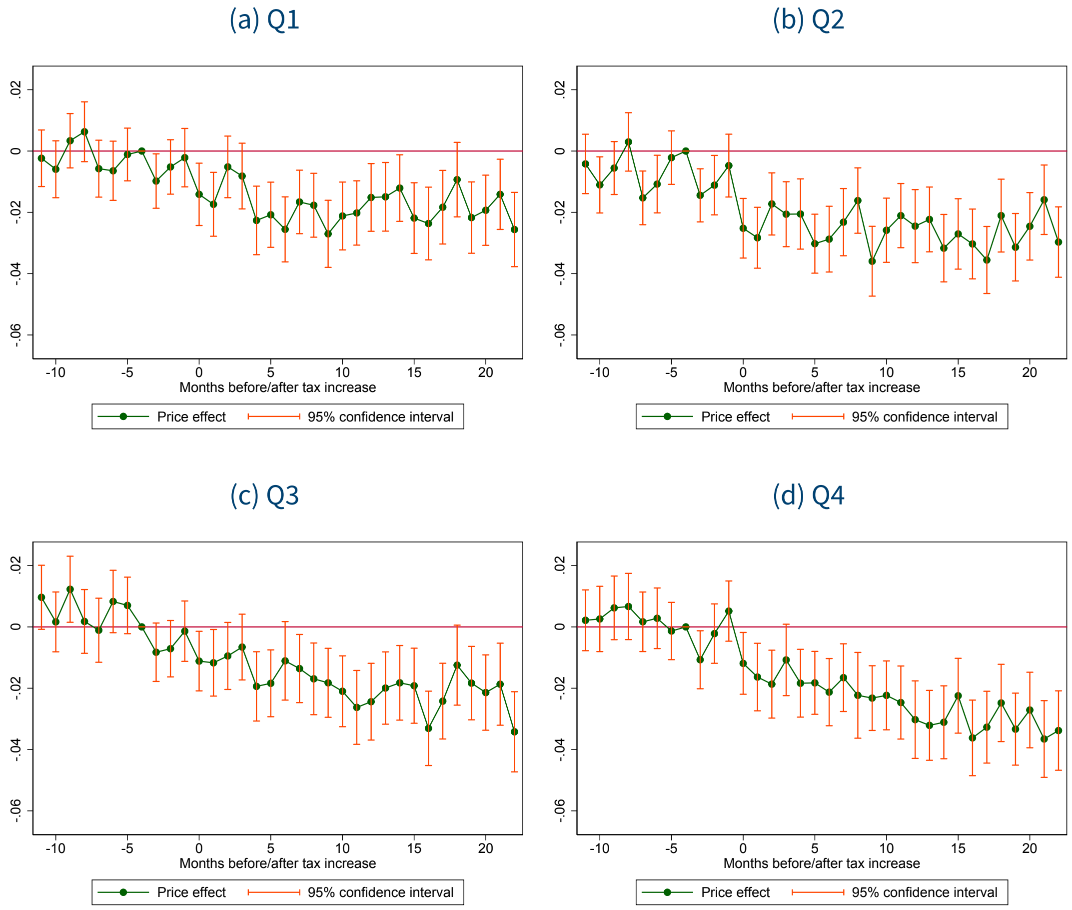
Second, downpayment constraints might play an important role here (Best and Kleven, 2018). While it is possible to debt-finance a large share of the property price, taxes need to be paid upfront and are usually not mortgageable. For credit-constrained buyers, an increase in the tax burden thus has a much larger impact on the affordability of a property than a property price change of the same magnitude. For downpayment-constrained households, one can thus expect a price response that exceeds the change in the tax burden.

2.6.2 Price Effects by Transaction Frequency and by Bargaining Power

Our theoretical model suggests that the price effect of a RETT change is larger when (i) the bargaining power of buyers is lower and (ii) when a property is expected to be traded more frequently in the future (see section 2.3). We provide evidence for the accuracy of both predictions by approximating properties' transaction frequencies and buyers' and sellers' bargaining power using two different indicators that vary at the municipal level.

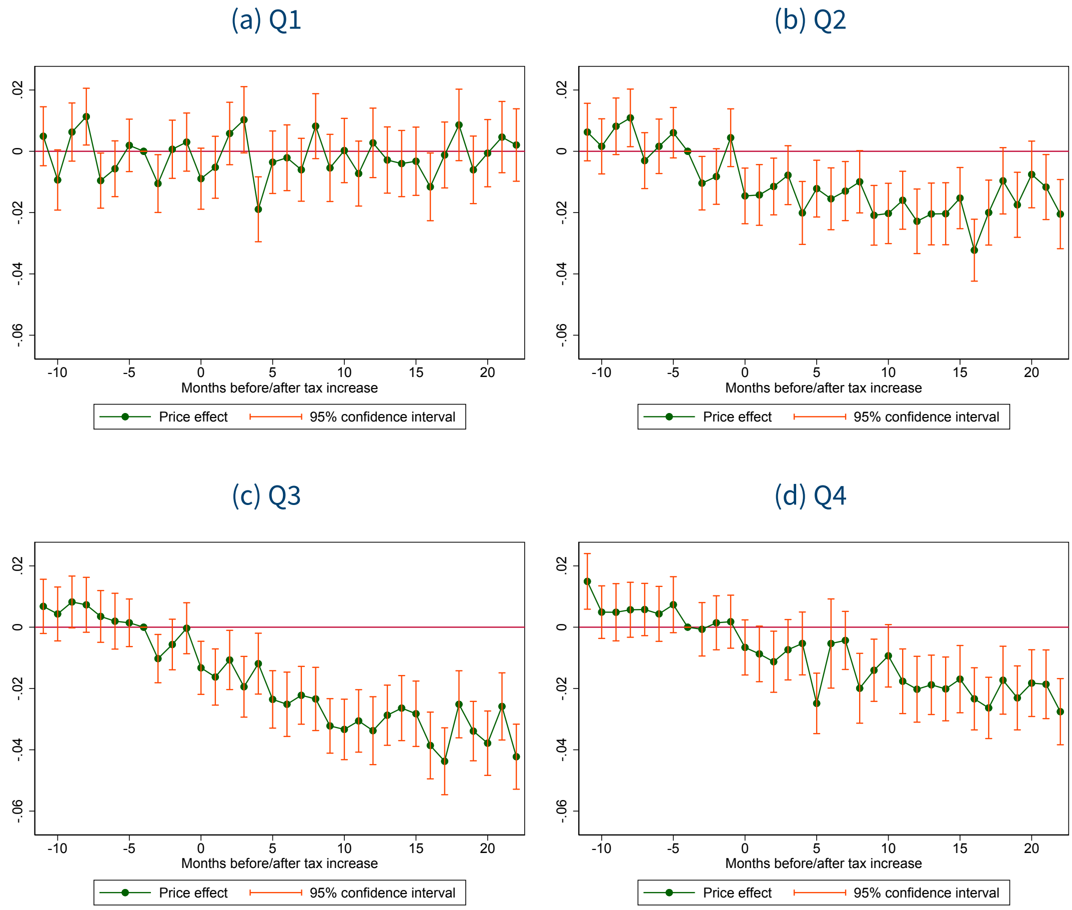
To analyze whether property prices respond more strongly to a RETT increase when transaction frequencies are higher, we compute a proxy for transaction frequencies at the municipal level. In a first step, we divide the aggregate number of houses offered for sale in a municipality by the total number of houses in the municipality. Then, we calculate the state-specific quartiles of this variable and assign each municipality to one of the four quartiles. Finally, we re-estimate equation 2.7 separately for each quartile. Figure 2.9 shows the results, with transaction frequencies increasing from Q1 to Q4. In line with the theoretical prediction of our

Figure 2.9: Transaction frequencies: Effects for houses by transaction frequency quartiles



Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands for each of four transaction frequency quartiles. Quartiles are formed within states and based on overall postings across all years relative to the housing stock. Transaction frequencies increase from Q1 to Q4. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level. This specification excludes the population-weighted top and bottom 5% of municipalities according to municipal population growth between 2005 and 2017.

Figure 2.10: Bargaining power: Effects for houses by population growth quartiles



Notes: The figure plots monthly event study estimates and corresponding 95% confidence bands for each of the four population growth quartiles. Quartiles are formed within states and based on municipal population growth between 2005 and 2017, also including municipalities with very high and very low population growth. Population growth rates increase from Q1 top Q4. The dependent variable is a property's log price per square meter, and event study indicators correspond to the change in the tax rate $\Delta\tau$. Specifications include postal code and month-year x urbanization level fixed effects. Standard errors are clustered at the postal code level.

model, price responses seem to be slightly stronger in case housing transaction occur more frequently, even though the differences between the quartiles are of a rather modest size.

In a similar vein, we proxy bargaining power of sellers and buyers using municipal population growth rates. The idea is that the larger a municipality's population growth, the higher is the demand for properties and, consequently, the better a seller's bargaining position. As before, we compute state-specific population growth quartiles, determine the quartile each municipality belongs to, and re-estimate equation 2.7 separately for each quartile. Figure 2.10 displays results. The price effects in the lowest quartile, that is, Q1, are not significantly different from zero. In those municipalities, sellers have little bargaining power. In line with the predictions derived from our model, the house price is close to the net-of-tax price, and no significant price response occurs. In turn, price responses are pronounced when population growth is (relatively) large and, hence, sellers (buyers) have high (low) bargaining power. Yet again, the evidence that we find supports the conclusions that can be drawn based on our theoretical model.

2.7 Conclusion

This paper exploits a constitutional reform that was passed in Germany in 2006 to study the effect of changes in the real estate transfer tax (RETT) on property prices. The reform granted the German states the right to set the rate of the RETT. Over the following years, 14 out of the 16 states executed this right and increased the RETT rate, often multiple times. Up to date, there have been 27 tax hikes. Before the reform, there was a uniform RETT rate of 3.5% that applied to all German states. Today, the highest RETT rate amounts to 6.5%.

We combine the information on RETT rate changes at the state level with a large micro dataset covering roughly 18 million properties that were offered for sale during the period from January 2005 until December 2018. The dataset was created by collecting information from property advertisements using web-scraping techniques. This information was collected from 140 different sources, including online property portals, regional and trans-regional newspapers, as well as property brokers. The list of variables includes the offering price as well as a large set of property characteristics. Importantly, our dataset contains the exact day the property was listed as well as the postal code of the property. Based on this dataset, we analyze the effect of an increase in the RETT rate on property prices using an event study design. We conduct our analysis separately for apartments, single-family houses, and apartment buildings.

Before turning to the empirical analysis, we set up a stylized theoretical model of the housing market to derive empirically testable predictions. One of the main insights is that in the Nash equilibrium, an increase in the RETT rate may result in a decline in property prices that exceeds the tax increase. Our model predicts that the semi-elasticity of the house price with respect to the RETT may be larger than one if the bargaining power of the seller is high and if a property is expected to be traded more frequently in the future.

Our empirical findings lend support to our theoretical model. We find that a one percentage point increase in the RETT reduces prices of apartments and apartment buildings by roughly 3-4% and single-family house prices by 1.5-2% in the twelve months after the reform. These results are robust to several modifications to our empirical specifications. What is more, we indeed find evidence that the magnitude of the price effect in response to a RETT rate change is positively related to the bargaining power of sellers in the housing market as well as to properties' transaction frequencies.

Our findings bear great importance for economic policy-makers. In light of the increasing international mobility of capital and labor, taxes on property are often regarded as a particularly attractive source of public revenue. However, only little is known so far about the distribution of the tax burden for a permanent increase in real estate transfer taxes. Our results suggest that the biggest share of the tax burden is borne by the seller, not the buyer of a property.

Appendix

In this appendix we show that $\frac{\delta\varepsilon}{\delta\beta} > 0$ and $\frac{\delta\varepsilon}{\delta q} < 0$, as claimed in the main text.

From equation (2.6) we can derive

$$\begin{aligned}\frac{\delta\varepsilon}{\delta\beta} = &- \left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-2} \frac{q}{(1+\rho)} \left[\frac{\beta U^O}{p^*} - 1\right] \\ &+ \left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-1} \frac{U^O}{p^{*2}} \left[p^* - \beta\frac{\delta p^*}{\delta\beta}\right]\end{aligned} \tag{2.A.1}$$

Note that the first term on the right hand side of equation (2.A.1) is positive because $\frac{\beta U^O}{p^*} - 1 < 0$. The second term on the right hand side of equation (2.A.1) is also positive because the price declines with increasing bargaining power of the buyers, i.e.

$$\frac{\delta p^*}{\delta\beta} = -\left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-1} \left[\frac{p^* q}{(1+\rho)} - U^O(1+T) + U^Y + (1-q)\frac{U^O}{(1+\rho)}\right] < 0 \tag{2.A.2}$$

From equation (2.6) we can also derive

$$\begin{aligned}\frac{\delta\varepsilon}{\delta q} = &\left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-2} \frac{(1-\beta)}{(1+\rho)} \left[\frac{\beta U^O}{p^*} - 1\right] \\ &- \left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-1} \frac{\beta U^O}{p^{*2}} \frac{\delta p^*}{\delta q}\end{aligned} \tag{2.A.3}$$

The first term on the right hand side of equation (2.A.3) is negative because $\frac{\beta U^O}{p^*} - 1 < 0$. The second term is also negative because

$$\frac{\delta p^*}{\delta q} = \left[1 + T - \frac{(1-\beta)q}{(1+\rho)}\right]^{-1} \frac{(1-\beta)}{(1+\rho)}(p^* - U^O) > 0. \tag{2.A.4}$$

q.e.d.

3 The Effect of Real Estate Purchase Subsidies on Property Prices

3.1 Introduction

Rising rents and property prices have fueled a debate on the affordability of housing in Germany, as well as in other countries around the world. This has led to calls for housing subsidies, and to the introduction of numerous measures aiming to reduce housing costs. Amongst others, recent years have seen the introduction of rent control, of a temporary accelerated depreciation schedule for the construction of residential units, and of subsidies for the acquisition of property by owner-occupiers. While many previous initiatives to make housing more affordable targeted renters and poorer households, increasing attention has lately been put on the costs of acquiring real estate. Both the German federal and the Bavarian state government implemented housing purchase subsidies in 2018, aiming to reduce purchase costs for owner-occupiers.

Although intended to foster homeownership and to make the acquisition of property more affordable, in particular for families, housing subsidies may well exert adverse effects by driving up real estate prices. This would especially be the case if housing demand is driven up by the subsidy scheme while housing supply is rather inelastic. According to the federal government, the federal subsidy is unlikely to lead to large windfall gains, and it claims a lack of evidence on price effects of housing purchase subsidies (Deutscher Bundestag, 2019). However, several features of the subsidy design speak in favor of potentially large price effects. First, due to generous income thresholds, roughly three quarters of German families with minor children – and in the case of Bavaria three quarters of households regardless of family structure – would be eligible for subsidies when buying a property. Second, federal subsidy provisions are set to expire in 2020. This could in turn further stimulate housing demand between 2018 and 2020. With the German construction sector operating at its capacity limits, housing supply is however rather inelastic (Gornig et al., 2019). As the application window for the federal scheme is confined to three years, incentives for the construction sector to expand and develop additional capacity are limited. Contrary to claims by the government, one could thus expect a considerable pass-through into prices.

Against this background, this paper investigates to which degree direct housing subsidies are capitalized into home prices. My study is the first to assess the price effects of direct housing purchase subsidies that are not intended as a stimulus measure. For this purpose, I exploit that Bavaria, Germany's second largest federal state by population, introduced a much more extensive subsidy scheme than the federal scheme available in all states. I use this policy discontinuity at the Bavarian interstate border to assess the effect of subsidies on home prices, using a rich micro-dataset on German house prices. My findings indicate that in the second half of 2018, single-family home prices increased by roughly 3.4% more in Bavarian border regions than in neighboring regions of other states. These results are consistent with a full shifting of subsidies into the prices of single-family homes. In contrast, no effect can be observed for apartments. This is likely due to apartments seldom being bought by owner-occupiers who could claim the subsidy. Splitting the sample into houses with a comparatively high or low subsidization probability also points to heterogeneous effects: price effects tend to be larger in sectors of the real estate market with a larger exposure to the subsidy scheme. I also provide suggestive evidence that the subsidy scheme slightly stimulated construction activity of single-family houses, while possibly leading to a partial crowding-out of the construction of apartment buildings. Providing a clean identification of subsidy effects, my findings provide an important contribution to both the literature and the current policy debate at a time at which the affordability of housing is considered a key policy issue in many countries.

Evidence on housing purchase subsidies in other countries also suggests a significant capitalization into real estate prices. While the German and Bavarian schemes grant flat-rate direct subsidies, other countries tend to subsidize the purchase of real estate through the tax code by granting mortgage interest deductions. Generally, most empirical evidence indicates that such tax subsidies do not increase the homeownership rate and are passed-through into property prices (see Bourassa et al., 2013 for a survey). In a general equilibrium model of the US housing market, Sommer and Sullivan (2018) show that eliminating the mortgage interest deduction would result in declining property prices, increasing homeownership and improved welfare. Hilber and Turner (2014) point out that a subsidy's effects on homeownership decisions and house prices depend on the elasticity of the housing supply: Homeownership only rises in areas with lax land-use regulations, whereas subsidies are capitalized into home prices in tightly regulated, rather inelastic housing markets. This house price effect might even result in an adverse effect on homeownership. Davis (2018) exploits the variation of US state-level tax legislation to assess capitalization effects of mortgage interest deductions on houses on both sides of the state border. His results indicate strong capitalization effects, with a one percentage point increase in the tax rate applied to mortgage interests leading to a 0.8 percent increase in house prices. Similarly, Berger et al. (2000) show a full capitalization of after-tax interest rate subsidies in Sweden. Using a Danish tax reform with a differential effect on mortgage interest deductions across tax brackets, Gruber et al. (2020) estimate long-term effects of housing tax subsidies. Their findings indicate zero effect on homeownership, but a sizable effect at the intensive margin as well as suggestive evidence that tax subsidies are capitalized into house prices.

However, the institutional setup of a mortgage interest subsidy considerably differs from the German subsidy schemes. While the latter grant flat-rate direct subsidies to households below an income threshold, the size of a mortgage interest subsidy depends on both the price of a property and individual marginal tax rates. Due to the interaction between tax progressivity and the mortgage interest subsidy, high-income households with high marginal tax rates benefit the most from these subsidies.

Evidence on direct subsidies is much more scarce. Also, in contrast to the German setting, governments tend to resort to direct subsidy programs as a stimulus when the economy is weak. In the wake of the financial crisis, the United States introduced a homebuyer tax credit to counter dropping demand in the housing market (Dynan et al., 2013). While first designed with a repayment requirement, the tax credit was granted as a subsidy in 2009 and 2010. In 2009, first-time homebuyers up to a certain income threshold were eligible for a refundable tax credit of 10 percent of the purchase price, capped at 8,000 USD. For most claimants, this is equivalent to a flat-rate subsidy, as in the Bavarian case. In a general equilibrium model, Floetotto et al. (2016) show that such homebuyer tax credits temporarily increase home prices and transaction volumes, but lead to negative welfare effects. Dynan et al. (2013) exploit regional variation in housing markets, finding only a small and temporary effect on sales. However, as credits were available throughout the country and the housing market underwent rapid changes, identifying a control group for an empirical analysis on prices is difficult. Similarly, the UK subsidizes the acquisition of new built homes below a certain property value with an equity loan for up to 20% (40% for London) of the property value. Exploiting spatial discontinuities in the scope of the scheme, Carozzi et al. (2019) find strong capitalization effects in the supply-constrained London area, the size of which suggests an overcapitalization, but no effect on construction. In a region with rather elastic supply, the subsidy is instead shown to stimulate construction.

This paper proceeds as follows. Section 3.2 provides an overview of the subsidy schemes implemented in 2018. Section 3.3 describes the data sources used in my analysis. In section 3.4, I subsequently present my methodological approach. This encompasses a description of the border difference-in-difference design and of the analysis of geodata. Results are presented in section 3.5. Section 3.6 concludes.

3.2 Institutional Background

While real estate prices were stagnating in Germany between 1995 and 2010, nominal prices have risen by roughly 50% in the last decade (Baldenius et al., 2019; Mense et al., 2019). Following the debate on increasing home prices, both the German federal government and the state of Bavaria introduced housing purchase subsidies in 2018. As the Bavarian subsidy program is supplementary to the nation-wide subsidy program, overall housing purchase subsides are much more extensive in Bavaria.

The Bavarian housing purchase subsidy (*Bayerische Eigenheimzulage*) constitutes an immediate subsidy of 10,000 euros and is paid to eligible households who purchase or build a house or apartment for personal residence after June 30, 2018. The aim of this subsidy is to encourage the acquisition of property, increase home ownership rates and create additional housing (Bayerische Eigenheimzulagen-Richtlinien, 2018). The subsidy is only granted to households who have resided in or been employed in Bavaria for at least one year. Income thresholds are rather generous. While singles with taxable incomes below 50,000 euros are eligible for the subsidy, the threshold increases to 75,000 euros for married couples and to 90,000 euros for households with one child. Each additional child increases this threshold by a further 15,000 euros. I.e., a family with two children would be eligible if their household income is below 105,000 euros. Overall, about three quarters of households meet these income requirements, and would potentially be eligible for the subsidy when purchasing or building real estate (see section 3.3.2).

In the same year, the German federal government implemented a housing subsidy program for families. In all states, families with at least one child can claim the federal child benefit for building (*Baukindergeld*) of 1,200 euros per child and year for a period of ten years. This subsidy is available nation-wide, independent of the Bavarian housing purchase subsidy. Income thresholds coincide with the Bavarian scheme. After the subsidy was enacted in May 2018, applications have been possible from September 18, 2018 onwards. While this time frame roughly corresponds to the Bavarian subsidy scheme, housing purchases and construction permits are retroactively eligible from January 1 onwards. However, this subsidy is only available for a limited time: The application window ends on December 31, 2023, while the building permit or purchase contract needs to be issued by December 2020.

In addition, Bavaria introduced a top-up of the federal child benefit of 300 euros per child and year (*Bayerisches Baukindergeld Plus*). This top-up has the same residency and employment requirements as the Bavarian housing purchase subsidy.

Table 3.1 indicates the maximum housing subsidy per household type in Bavaria and in other German states. Overall, eligibility conditions are broader and the average subsidy is much larger in Bavaria. Note also that the Bavarian housing purchase subsidy is paid up-front upon approval, whereas child benefits are paid over a period of ten years. This may have different implications for downpayment-constrained households as imminent payments may be more readily considered by mortgage brokers[1]: Subsidy payments that banks consider equivalent to equity may lead to more favorable interest rate conditions.

Table 3.1: Scope of housing subsidies

	Bavaria			
	No children	One child	Two children	Three children
Bavarian purchase subsidy	10,000	10,000	10,000	10,000
Federal child benefit	0	12,000	24,000	36,000
Bavarian child benefit	0	3,000	6,000	9,000
Total subsidy	10,000	25,000	40,000	55,000
	Other states			
	No children	One child	Two children	Three children
Federal child benefit	0	12,000	24,000	36,000

Notes: The table indicates the maximum amount of housing subsidies in euros in Bavaria and in other German states.

A similar nation-wide scheme was abolished in 2006 due to its limited cost-effectiveness and its resulting windfall gains (Deutscher Bundestag, 2005). With a volume of 11.4 billion euros in 2004, the subsidy scheme had been one of the largest subsidy schemes at the time.[2] While the policy was widely criticized on the grounds of being costly and inequitable, leading to windfall gains and potentially driving up prices (see e.g. Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2003); Bundesamt für Bauwesen und Raumordnung (2002); Färber (2003)), studies on this scheme are only descriptive in nature.

[1] According to one of Germany's largest real estate platforms, the child benefit for building is not considered equivalent to equity by banks, also due to the long payment window: https://ratgeber.immowelt.de/a/baukindergeld-2018-wer-es-bekommt-wie-viel-es-gibt-und-was-die-voraussetzungen-sind.html

[2] As under current legislation, households with incomes below a certain threshold were eligible for the subsidy for the purchase or construction of an owner-occupied property. The subsidy was paid as a direct subsidy for a period of eight years, and consisted of a base subsidy tied to a property's acquisition costs and an additional child allowance. Until 2003, the construction of new properties was subsidized at twice the rate of the subsidy for purchases of existing homes. In 2004 to 2005, lower and uniform base subsidy levels were granted, while child supplements increased.

As opposed to other countries such as the United States, mortgage interest on owner-occupied housing cannot be deducted from income taxes. Therefore, interaction effects between housing purchase subsidies and mortgage interest taxation do not need to be accounted for. However, the federal government has introduced a temporary accelerated depreciation schedule for the construction of new residential units. This reform enables an additional 5 percent depreciation rate, subject to an upper bound, on residential units for rent constructed between September 2018 and December 2021. While this measure does not directly affect owner-occupiers, it adds to the strain on the construction sector and might drive property prices.

These reforms are implemented at a time of historically high capacity utilization in the construction sector (Gornig et al., 2019). As the application window of the child benefit for building and the accelerated depreciation schedule is confined to a period of three years, the incentive for construction companies to expand capacities are limited. Against this background, one could expect a substantial effect on property prices.

3.3 Data and Descriptive Statistics

3.3.1 Microdata on Real Estate Prices

My empirical analysis is based on a large and detailed micro dataset on the German real estate market provided by the real estate consultancy firm F+ B (c.f. chapter 2). The dataset encompasses property adverts from 140 different sources, ranging from online property portals to newspaper adverts and real estate agents. Data collection was conducted via web-scraping. The raw dataset was subject to data cleansing and consistency checks to ensure that properties listed concurrently in multiple sources are only included once.

The final dataset contains 307,517 houses and 273,786 apartments that were offered for sale within 50 km of the Bavarian interstate border in 2016 to 2018. While F+B provides data from 2005 onwards, I restrict the data to the years around the reform to ensure that the estimation of pre-reform postal code fixed effects are unbiased by further state-level policies, such as long-term infrastructure investments or increases in real estate transfer tax rates.

Table 3.2 shows sample means of property characteristics for houses and apartments in the border regions of Bavaria and of neighboring states, both for the full data set (within 50 km of the border) and the data used in my main specifications (within 25 km of the border). The main variable of interest is a property's final asking price per square meter. While F+B provides both the first and the final asking price, I focus on the latter as it is likely closer to the actual transaction price. As shown in Table 3.2, asking prices of houses in Bavaria amount to 299,742 euros on average, or 1,952 euros per square meter (281,645 and 1,825 euros, respectively, for the narrower sample). These are slightly lower than average asking prices in neighboring states. These price differences are at least partly driven by the slightly higher frequencies at which houses in other states are equipped with amenities, such as a garden or a balcony. My estimations employ postal code fixed effects to account for initial price level differentials, and some specifications account for amenities.

Table 3.2: Real estate data: Summary statistics

	Houses				Apartments			
	<50km		<25km		<50km		<25km	
	Bavaria	Other	Bavaria	Other	Bavaria	Other	Bavaria	Other
Asking price	299,742	348,419	281,645	324,619	240,083	288,001	234,306	283,982
Price per sqm	1,952	2,215	1,825	2,084	2,434	2,736	2,292	2,679
Area in sqm	157.3	158.7	157.6	158.3	104.2	105.4	108.0	105.9
Number of rooms	5.3	5.3	5.3	5.3	3.3	3.4	3.4	3.4
Balcony	36.6%	39.9%	36.3%	39.1%	43.3%	43.6%	43.2%	41.8%
Garden	39.1%	43.7%	38.3%	43.1%	27.1%	28.1%	27.8%	28.9%
Basement	49.3%	49.7%	49.5%	49.4%	49.2%	51.7%	48.3%	51.6%
Parking spot	55.4%	57.9%	56.0%	56.9%	72.4%	72.6%	72.7%	71.6%
Number of observations	109,485	198,032	65,653	85,458	84,356	189,430	46,706	80,115

Notes: Overall sample: Houses and apartments within 50 km and 25 km of the Bavarian border, 2016-2018. Other states encompass Baden-Württemberg, Hesse, Thuringia, and Saxony. Source: F+B and own calculations.

3.3.2 Income and Consumption Survey Data

I supplement my analysis with data from the German Income and Expenditure Survey (*EVS, Einkommens- und Verbrauchsstichprobe*) 2018. Conducted by the Federal Statistical Office every five years, the EVS constitutes a representative survey of German households. In the 2018 wave, the dataset encompasses 58,278 households. Amongst others, the survey contains data on incomes, homeownership and living conditions. This enables me to assess the household and property characteristics of households that meet eligibility requirements for the subsidy scheme.

Table 3.3 presents summary statistics by property type in the EVS data. The vast majority of households living in houses are owner-occupiers, whereas only about one fifth of households in apartments own their own property. Also, houses are more frequently inhabited by families with minor children. On average, houses in the EVS sample are a bit smaller than in the advert data, but more frequently equipped with a parking spot.[3]

Table 3.3: EVS data: Summary statistics for households by property type

	Houses	Apartments
Owner-occupiers	83.4%	21.2%
Minor children	27.8%	14.7%
Area in sqm	129.8	73.8
Number of rooms	4.6	2.7
Parking spot	86.5%	49.8%
Number of observations	24,029	34,249

Notes: Overall sample: Households in the EVS data. Source: EVS 2018 and own calculations.

Table 3.4 indicates the fraction of Bavarian households with incomes below the eligibility threshold. While eligibility is based on gross taxable income, EVS data provides binned net household incomes. Therefore, I first apply a tax-benefit calculator on household-type specific gross income eligibility thresholds. Households with incomes below the resulting net income threshold are then classified as eligible. I use linear extrapolation to determine the fraction of eligible households whose income lies in the same income bin as the eligibility threshold.[4]

Table 3.4: EVS data: Share of eligible households in Bavaria

	All households	Owner-occupiers
All Bavarian households	74.8%	66.3%
Singles	82.6%	74.0%
Childless couples	72.6%	69.1%
Households with one child	76.7%	67.3%
Households with two children	77.2%	73.6%
Households with three or more children	83.4%	81.6%
Number of observations	8,402	4,702
All German households	80.4%	69.3%
Number of observations	58,278	28,808

Notes: Overall sample: Bavarian households in the EVS data, all households in the EVS data. Calculations for German households according to the Bavarian eligibility criteria. Source: EVS 2018 and own calculations.

[3] This may be due to different resale frequencies of property types, as well as to differing geographic scopes of both data sets. While Table 3.3 provides summary statistics on German households, Bavarian border regions are less urban than the German average. As homes in urban areas tend to be smaller, this might contribute to the difference between both data sets.

[4] Take an eligibility threshold of 4,600 euros per month, for example, which lies in the net income bin of 4,500 to 5,000 euros. In this case, calculations for Table 3.4 assume that 20% of households in this income bin are eligible. Results barely change, though, when either classifying all or no households in this income bin as eligible.

As shown in Table 3.4, about three quarters of households would be eligible for the subsidy when purchasing or building real estate. Amongst owner-occupiers, roughly two thirds of households meet the subsidy schemes' income criteria. This group might be more indicative of households who purchase a house.

3.3.3 Construction Permit Statistics

In addition to estimating the subsidy schemes' effect on property prices, I assess whether the availability of subsidies exerts a differential effect on construction activity. For this endeavor, I employ municipality-level administrative data on authorized residential construction projects (*Statistik der Baugenehmigungen*). This data set is based on a full census of residential construction projects for which either a construction permit was granted, or which required a notification of municipal authorities in lieu of an application for a construction permit.[5] The data set thus covers the universe of planned residential construction activity in the year formal approval was acquired. For ease of reference, I will refer to all cases as construction permits.

As larger cities issue much more construction permits than smaller municipalities, the number of residential construction permits varies between zero and several hundred permits per municipality and construction year. To account for differing municipality sizes, I scale construction activity in relation to the building stock. The latter is based on administrative data on the number of residential buildings in each municipality in 2017. Table 3.5 shows summary statistics on the number of construction permits for residential buildings, both in absolute terms and in relation to the overall municipal building stock.

Table 3.5: Construction permit data: Summary statistics

	Residential construction	Single-family houses	Multi-family houses
Total construction permits	9.8	7.3	2.5
Per 1000 buildings	6.7	5.4	1.3

Source: Statistical Offices of the Federal States and own calculations. Overall sample: Annual municipal residential construction permits in the vicinity of 25 km of the Bavarian interstate border, 2016-2018.

[5] Whether the construction of a property requires a construction permit depends on state laws as well as local building regulations and development plans.

3.4 Methodology

3.4.1 Estimation Strategy

I employ a border difference-in-difference approach to estimate the price effect of the real estate purchase subsidy. This approach assesses whether property price trends diverge after the introduction of the subsidy, while controlling for different local price levels and property characteristics. Allowing for differential regional time trends, the estimation strategy also accounts for changing local conditions that may impact real estate prices. I hence estimate the following equation:

$$ln(p)_{i,c,t} = \beta \, Subsidy_{c,t} + X_i'\theta + \delta_c + \gamma_{a(c),t} + \epsilon_{i,c,t} \tag{3.1}$$

Subscript i indicates the respective property, t the month it was offered for sale, and c the postal code area the property is located in. As explained more thoroughly in section 3.4.2, postal codes are allocated to cross-border regions $a(c)$ to capture regional trends. A property's log square meter price $ln(p)_{i,c,t}$ is used as dependent variable. The main variable of interest, $Subsidy_{c,t}$, is a dummy for properties posted in Bavaria after July 2018. A positive coefficient indicates that prices on the Bavarian side of the border have risen more than prices in neighboring regions after the implementation of the subsidy scheme. The specification accounts for postal code fixed effects δ_c, which capture persistent differences in local property prices due to possibly unobserved factors, such as natural amenities, traffic accessibility, or school quality. Region-month fixed effects $\gamma_{a(c),t}$ permit differential time trends across regions. Several specifications also control for property characteristics X_i, which encompass the number of rooms, a property's area in square meters, and the presence of amenities that may affect property prices. The latter include dummy variables for whether a property comes with a parking spot, a balcony, a garden or a basement.[6] Standard errors $\epsilon_{i,c,t}$ are clustered at the postal code level to account for a possible spatial correlation in local property price shocks.

My main estimations focus on house prices as houses are predominately acquired by owner-occupiers, whereas apartments tend to be more frequently bought by investors (Petkova and

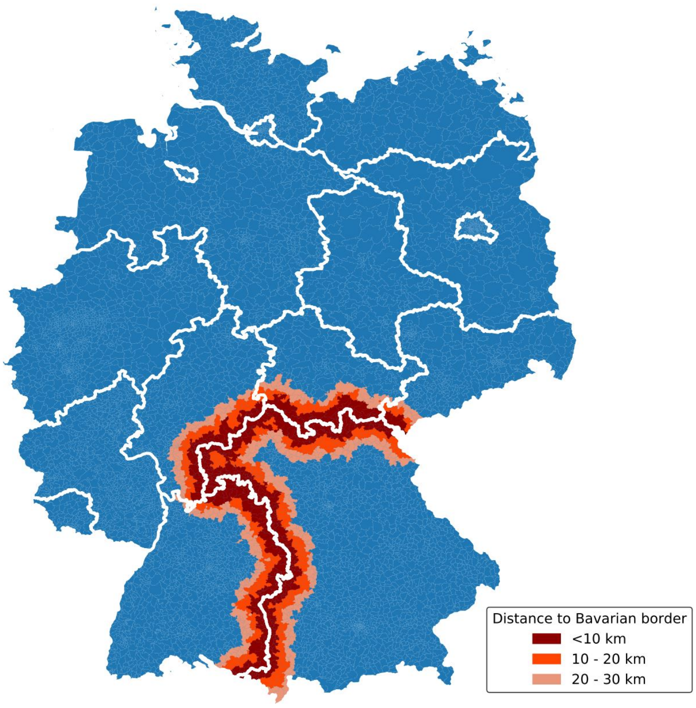
[6] I do not account for more subjective property characteristics, such as whether a property is described as modern, well-equipped or luxurious. These assessments might be partially driven by the market environment, such as sellers' market power, and might hence not be orthogonal to the reform. Likewise, I do not account for the construction year. This is the case as the construction year is missing for 19.8% of houses in the sample. Whether a seller discloses the construction year is however not random, and might be correlated with other conditions in the real estate market. Hence, either controlling for construction years, or excluding observations with missing construction years, might lead to a bias in the estimations.

Weichenrieder, 2017; Deutsche Bundesbank, 2018). This is also in line with EVS data, which show that a vast majority of residents of houses are owner-occupiers, while most households living in apartments are renters. As the subsidies are only granted to owner-occupiers, I expect much stronger price effects for houses. A further specification investigates whether this prediction holds and provides results on apartment prices.

3.4.2 Geographic Location Data

Each postal code is allocated to a distance band around the Bavarian interstate border according to the minimum distance between the postal code's centroid and the border. While postal codes in the immediate vicinity of the border are arguably subject to rather comparable time trends, trends may diverge more strongly the larger the distance to the border. This implies that there is a trade-off between the number of observations and, thus, estimation efficiency on the one hand, and unbiasedness on the other hand. For this reason, I estimate equation 3.1 for different distance bands around the interstate border. Figure 3.1 showcases the assignment of postal codes to distance bands.

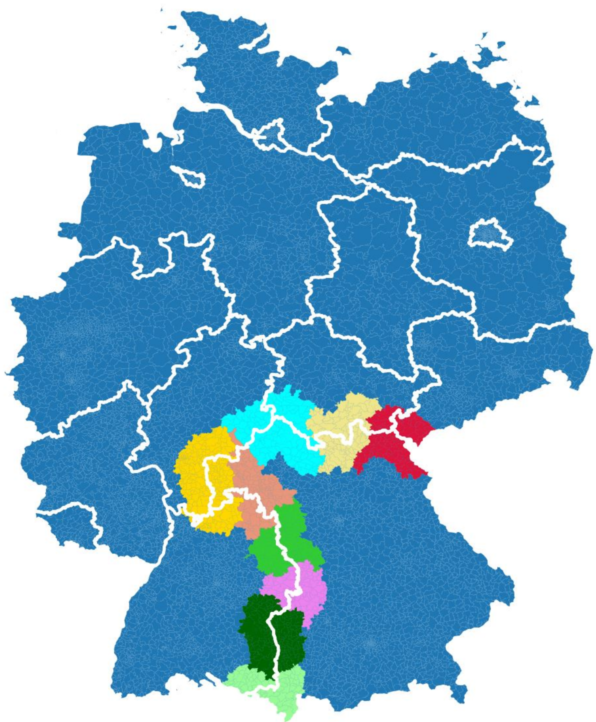
Figure 3.1: Postal codes in proximity of the Bavarian border



Notes: This figure shows postal codes in the proximity to the Bavarian interstate border and their allocation to distance bands around the border.

As economic conditions may vary along the border over time, I subsequently segment border regions based on spatial planning regions (*Raumordnungsregionen*). A spatial planning region combines several NUTS-3 regions within a state according to regional structure and commuting patterns. These regions are commonly used for spatial observation and monitoring by German institutions, such as the German Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR), but are not endowed with administrative autonomy. As spatial planning regions are defined within states, I generate cross-border regions by matching postal codes in bordering states to the closest Bavarian region. As a first step, I assign Bavarian postal codes to their respective spatial planning region along the border. Subsequently, postal codes in neighboring states are matched to the closest Bavarian spatial planning region. This matching is based on the minimum geographic distance between the postal code's centroid and the border of the spatial planning region. Using rather wide distance bands includes some Bavarian postal codes in the sample that are located in a non-border spatial planning region. These postal codes are assigned to the closest spatial planning region that adjoins the border. Figure 3.2 shows which region postal codes are assigned to.

Figure 3.2: Matched regions in proximity of the Bavarian border



Notes: This figure shows the allocation of postal codes to cross-border regions, based on the proximity to spatial planning regions in Bavaria.

3.4.3 Accounting for Tax Reforms

Other concurrent reforms may possibly exert a differential impact on real estate prices. Most notably, the neighboring state of Thuringia increased its real estate transfer tax (RETT) rate from 5.0 to 6.5% at the beginning of 2017 (see Table 2.1 in chapter 2). This presumably had an impact on real estate prices in Thuringia. As shown in chapter 2, a one percentage point increase in the real estate transfer tax rate reduces house prices by 1.5-2%, and lowers apartment prices by 3-4%. While this reform predates the introduction of housing purchase subsidies by more than a year, it likely resulted in a downward shift in prices in the pre-period, which would not be adequately captured by postal code fixed effects and cross-border regional time trends. In consequence, the estimated price effect of the Bavarian real estate purchase subsidy might be biased. Two different strategies are used to address possible confounding effects of Thuringia's RETT increase. One set of specifications drops all properties in regions intersected by the Thuringian border. I.e., estimations exclude the three north-eastern regions of Figure 3.2. A second set of specifications retains all observations, but introduces dummies intended to capture differential price trends in Thuringia. As indicated by the event studies in chapter 2, house prices begin to decline in the quarter prior to RETT reforms, with most of the pass-through taking place within half a year of a tax increase. In line with these findings, I account for RETT effects with dummies in the state of Thuringia for the quarters during which one could expect a gradual pass-through into house prices – Q4, 2016, Q1 2017, and Q2 2017 – as well as a dummy variable for the time period in which house prices would be expected to have adjusted to the new price level, i.e., Q3 2017 to Q4 2018. However, the latter specification would not account for spillover effects of the Thuringian tax increase into border regions of Bavaria, Hesse and Saxony. In this setting, spillover effects are more of a concern than in case of the real estate purchase subsidy: While the subsidy requires prior residence or prior employment in the state of Bavaria, the RETT increase applies to all households regardless of their prior residence.

3.5 Results

3.5.1 Real Estate Prices

Table 3.6 shows results for houses in postal codes within 25 km of the Bavarian interstate border. Specification (1) does not allow for regionally differentiated trends and neither controls for the real estate transfer tax reform in Thuringia, nor for property characteristics. Regional time trends are added in specification (2). Coefficients are positive and significant in both specifications, albeit at a lower level than in subsequent specifications which account for a bias due to Thuringia's RETT reform: Estimated effects are larger when excluding border regions of Thuringia (specification (3)) or using dummy variables to control for the RETT reform (specification (4)). Controlling for property characteristics results in coefficients of respectively 0.0345 and 0.0264 in specifications (5) and (6). This indicates that in the second half of the year 2018, Bavarian house prices increased by roughly 2.6 to 3.4% more than house prices in neighboring states. Specifications that use dummy variables to capture differential price trends in Thuringia yield lower effects than specifications that exclude Thuringian border regions. This could either be due to a lower responsiveness of prices in the predominately rural north-eastern border region[7], spillover effects between Thuringia and neighboring states, or the dummy variables not adequately capturing the timing of the pass-through of RETT reforms[8]. Hence, further robustness checks primarily focus on specification (5). With pre-subsidy house prices averaging 318,700 euros in the Bavarian border region (276,400 euros when including border regions with Thuringia), findings would be consistent with a full shifting of the Bavarian real estate purchase subsidy into house prices: 10,000 euros correspond to 3.3% of 300,000 euros.

Table 3.6: Subsidy effects on asking prices of single-family houses

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	0.0211*	0.0287**	0.0410***	0.0334***	0.0345***	0.0264**
	(0.0098)	(0.0129)	(0.0126)	(0.0126)	(0.0120)	(0.0120)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month	Month x region	Month x region	Month x region	Month x region	Month x region
Controls for Thuringia	✗	✗	Exclusion	Dummies	Exclusion	Dummies
Property controls	✗	✗	✗	✗	✓	✓
Max km to border	25	25	25	25	25	25
N	151,111	151,111	113,917	151,111	113,917	151,111

Notes: This table shows the differential effect of housing subsidies in Bavaria on house prices. The treatment dummy indicates properties listed in Bavaria between July and December 2018. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

To verify that trends within cross-border regions are indeed comparable, I conduct a placebo test on a sample limited to the pre-reform years 2016-2017. In analogy to the baseline, this specification estimates whether price trends of houses available for sale in Bavaria in the second half of the year 2017 differ from bordering states.

[7] A robustness check in Table 3.12 however finds no differences in the pass-through for urban and rural regions.

[8] As opposed to other states' RETT reforms, Thuringia announced its tax increase more than a year in advance in mid-2015. This might conceivably lead to anticipation effects and diverging pass-through patterns.

As indicated by Table 3.7, the placebo test yields no significant difference in the evolution of property prices, underlining the validity of my identification strategy.

Table 3.7: Placebo test for asking prices of single-family houses

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	-0.0027	-0.0064	0.0129	-0.0032	0.0050	-0.0106
	(0.0116)	(0.0119)	(0.0117)	(0.0119)	(0.0111)	(0.0113)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month	Month	Month	Month	Month	Month
		x region	x region	x region	x region	x region
Controls for Thuringia	✗	✗	Exclusion	Dummies	Exclusion	Dummies
Property controls	✗	✗	✗	✗	✓	✓
Max km to border	25	25	25	25	25	25
N	96,237	96,237	74,126	96,237	74,126	96,237

Notes: This table shows the results of a placebo test for differential trends in house prices in Bavaria. The treatment dummy indicates properties listed in Bavaria between July and December 2017. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

As a robustness check, I conduct the estimation for different distance bands around the interstate border. Table 3.8 shows results that correspond to specification (5) in Table 3.6, i.e., estimations that exclude border regions with Thuringia and control for property characteristics.

Table 3.8: Subsidy effects on asking prices of single-family houses for different distance bands to the interstate border

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	0.0240**	0.0300***	0.0355***	0.0358***	0.0372***	0.0442***
	(0.0096)	(0.0102)	(0.0114)	(0.0123)	(0.0135)	(0.0164)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month	Month	Month	Month	Month	Month
	x region	x region	x region	x region	x region	x region
Controls for Thuringia	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion
Property controls	✓	✓	✓	✓	✓	✓
Max km to border	50	40	30	20	15	10
N	227,475	183,380	134,451	92,593	77,410	55,401

Notes: This table shows the differential effect of housing subsidies in Bavaria on house prices. The treatment dummy indicates properties listed in Bavaria between July and December 2018. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

For a range between 15 and up to 40 km around the border, results are in line with each other and the coefficient on the subsidy dummy amounts to on average 0.035. This is again consistent with a full shifting of subsidies into property prices. However, coefficients gradually increase when the band around the border becomes more narrow. In particular, results are larger for a very narrow distance band of 10 km, although the coefficient of 0.0442 does not significantly differ from the coefficients for larger distances. Two factors might play a role here: First, even though restricted by prior residency and employment requirements, spillover effects across the border might exert effects on real estate prices on both sides of the border. This would be the case if households who used to live in neighboring states purchased houses in Bavaria in response to the reform, or if Bavarian households who would have otherwise considered moving to a neighboring state decided to remain in Bavaria. This effect attenuates with an increasing bandwidth around the border. Second, the common trend assumption might not hold up as well for the very narrow sample.

Results of the 2017 placebo test for different distance bands point in this direction (Table 3.9): while coefficients are insignificant for all distance bands, they are larger for the 10 kilometer band around the border.

Table 3.9: Placebo test for asking prices of single-family houses for different distance bands to the interstate border

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	0.0051	0.0059	0.0077	0.0006	0.0056	0.0231
	(0.0089)	(0.0096)	(0.0107)	(0.0120)	(0.0129)	(0.0146)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month x region	Month x region	Month x region	Month x region	Month x region	Month x region
Controls for Thuringia	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion
Property controls	✓	✓	✓	✓	✓	✓
Max km to border	50	40	30	20	15	10
N	148,462	119,950	87,957	60,019	50,207	35,935

Notes: This table shows the results of a placebo test for differential trends in house prices in Bavaria. The treatment dummy indicates properties listed in Bavaria between July and December 2017. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

In contrast to houses, effects for apartments are insignificant and close to zero (see Table 3.10). This is also the case for various distance bands around the border, as shown in Table 3.11. The absence of any notable effect is consistent with expectations, given that owner-occupiers only constitute a small share of apartment residents, and investment decisions on rental properties remain unaffected by the reform. The subsidy scheme might also exert a counterbalancing effect on apartment prices: Some tenants of apartments may decide to purchase a house and vacate their rental apartment in response to the subsidy. With rental revenues decreasing, this could conceivably lead to a small downward shift in the demand for apartments.

Table 3.10: Subsidy effects on asking prices of apartments

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	0.0048	0.0067	-0.0025	0.0053	-0.0063	0.0018
	(0.0129)	(0.0140)	(0.0140)	(0.0140)	(0.0131)	(0.0132)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month	Month x region	Month x region	Month x region	Month x region	Month x region
Controls for Thuringia	✗	✗	Exclusion	Dummies	Exclusion	Dummies
Property controls	✗	✗	✗	✗	✓	✓
Max km to border	25	25	25	25	25	25
N	126,821	126,821	106,970	126,821	106,970	126,821

Notes: This table shows the differential effect of housing subsidies in Bavaria on apartment prices. The treatment dummy indicates apartments listed in Bavaria between July and December 2018. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 3.11: Subsidy effects on asking prices of apartments for different distance bands to the interstate border

Dependent variable: log price per sqm						
	(1)	(2)	(3)	(4)	(5)	(6)
Subsidy	-0.0008	-0.0082	-0.0100	-0.0053	-0.0093	-0.0139
	(0.0107)	(0.0112)	(0.0124)	(0.0141)	(0.0156)	(0.0185)
PLZ FE	✓	✓	✓	✓	✓	✓
Time FE	Month x region	Month x region	Month x region	Month x region	Month x region	Month x region
Controls for Thuringia	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion	Exclusion
Property controls	✓	✓	✓	✓	✓	✓
Max km to border	50	40	30	20	15	10
N	228,503	177,544	127,145	83,664	68,160	46,871

Notes: This table shows the differential effect of housing subsidies in Bavaria on apartment prices. The treatment dummy indicates apartments listed in Bavaria between July and December 2018. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Overall, these findings confirm the validity of the house price estimations: If results for house prices were driven by a spurious correlation with other policy changes, this would likely show up in all property prices.

3.5.2 Heterogeneity Analysis and Quantification of Effects

As shown by my previous analysis, the subsidy scheme's aggregate effect on house prices is consistent with a full capitalization into house prices. At an average pre-reform house price of 318,700 euros in Bavarian border municipalities (276,400 when Thuringian border regions are included in the sample), my preferred specification's coefficient of 0.0345 translates into a price increase of roughly 11,000 euros (see the upper panel of Table 3.12). Yet, one could conceivably expect differential effects across segments of the property market.

For once, findings by Hilber and Turner (2014) and Carozzi et al. (2019) suggest differential effects by the degree of urbanization. As building plots might be more readily available for development in rural areas, housing supply might be more elastic. This could result in a comparatively lower capitalization in house prices. I assess whether price responses differ by the level of urbanization by estimating separate treatment coefficients for rural and urban counties. Counties are classified in line with a categorization (*Siedlungsstrukturelle Kreistypen*) by the Federal Institute for Research on Building, Urban Affairs and Spatial Development (BBSR). However, treatment coefficients hardly differ between urban and rural counties (second panel of Table 3.12). This might be due to capacity constraints in the construction sector: In the short-run, housing supply might be fairly inelastic, even if developable land were readily available. As average house prices are higher in urban regions, a 3.5% price increase nevertheless translates into a higher price growth in absolute terms in urban compared to rural counties.

Table 3.12: Subsidy effects on asking prices of single-family houses: Heterogeneous effects

Dependent variable: log price per sqm					
	(1) Subsidy	(2) SE	(3) Average area	(4) Average price	(5) Effect in euros
All	0.0345***	(0.0120)	157	318,702	10,995
Rural counties	0.0335**	(0.0155)	158	295,509	9,900
Urban counties	0.0354**	(0.0147)	160	347,101	12,287
Small houses	0.0362**	(0.0156)	112	250,939	9,084
Medium-sized houses	0.0425***	(0.0154)	146	298,222	12,674
Large houses	0.0244	(0.0162)	220	414,548	10,115

Notes: This table shows the differential effect of housing subsidies in Bavaria on house prices. The treatment dummy indicates properties listed in Bavaria between July and December 2018. Specifications are equivalent to column (5) of Table 3.6, i.e., encompass postal codes within 25 km of the interstate border, exclude border regions to Thuringia, and account for postal code fixed effects, control for property characteristics and include month x region fixed effects. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Second, I assess whether effects differ by house size. All estimations use a property's price per square meter as a dependent variable. All else equal, capitalization of flat-rate subsidies into prices per square meter should be larger for smaller houses. I split the sample into small, medium-sized and large houses, based on tertiles of the house size distribution. Treatment coefficients for all tertiles are jointly estimated. Results are depicted in the bottom panel of Table 3.12. Effects are positive for all house types and significant for small and medium-sized houses. Medium-sized houses exhibit the largest price growth, both in percentage and in absolute terms. However, coefficients might also capture a different effect: houses of different sizes may have a different propensity to be acquired by recipients of the subsidy. Average subsidies might also differ between house types as families are granted a higher subsidy due to the child supplement. For example, small houses with few rooms may not be attractive for families with children. In consequence, the subsidy scheme may have a comparatively lower impact on the demand curve for small houses.

3.5.3 Extension and Discussion

Analysis by Likelihood of Subsidization

While real estate adverts data is well-suited for an analysis of aggregate price effects of subsidy schemes, it does not provide any information on a property's buyer. Therefore, I cannot directly infer whether a property's purchaser is eligible for the Bavarian housing purchase subsidy or for additional child benefits for building. This complicates assessing how a differential scope of subsidies is capitalized into prices. In order to assess whether effects differ across subsidy levels, I instead impute subsidization probabilities based on EVS data. This allows for a differential analysis of houses whose characteristics make them more or less likely to be acquired by beneficiaries of the subsidy scheme.

As a first step, I estimate a probit model for all houses in the EVS data. This estimates the probability that a house is inhabited by owner-occupiers whose incomes comply with eligibility requirements, taking account of house characteristics contained in both data sets.[9] The estimated coefficients are then used to predict subsidization probabilities in the real estate advert data. These predicted probabilities are indicative of how likely a house is to be subsidized, but should not be taken at face value. Not only is explanatory power limited at the first stage, the categorization of houses might also be prone to omitted variable bias: Both the size of houses and the share of households above income thresholds may be correlated with the regional price level. I.e., in areas with a higher initial price level, households with a given income may

[9] Variables include a polynomial of a house's area in square meters and dummy variables for the number of rooms, a parking spot, and broad construction year categories as defined in the EVS data.

on average acquire smaller houses, and houses with given characteristics may on average be acquired by households with higher incomes. Lacking detailed geographic information in the EVS data, I cannot account for this correlation. Furthermore, housing choices might be endogenous to the subsidy scheme, with subsidies inducing the acquisition of larger homes (Gruber et al., 2020). Finally, while EVS constitutes a representative household sample, its results are not necessarily representative for the cross-section of advertised properties. As average housing tenure may be related to property characteristics, some property types might comprise a larger share of housing transactions than of the housing stock. The probability that a specific house is inhabited by an eligible household might thus differ from the probability that the house is acquired by the very same household.

Therefore, I only conduct a broad-level analysis with heterogeneous effects for houses that are more or less likely to be subsidized. For this purpose, I define the upper half of the probability distribution as houses with a high subsidization probability. While individual probability predictions might be biased, houses in the upper half of the distribution should on average have a higher likelihood of being subsidized. To assess differential effects for the subset of houses with a high subsidization probability, I extend equation 3.1 with an interaction term between the treatment variable and an indicator for houses with a high subsidization probability (HP_i):

$$ln(p)_{i,c,t} = \beta_1\, Subsidy_{c,t} + \beta_2\, (Subsidy_{c,t} * HP_i) + \nu\, HP_i + X_i'\theta + \delta_c + \gamma_{a(c),t} + \epsilon_{i,c,t} \quad (3.2)$$

These estimations are then conducted for households that are eligible for different subsidy levels. I.e., I estimate several probit models with different dependent variables. I first assess overall eligibility for the Bavarian purchase subsidy scheme, and subsequently estimate the probability that a specific house is inhabited by a family that is also eligible for child supplements for at least one, two or three children. As families receive higher subsidies due to the Bavarian top-up of the federal child subsidy, this helps assessing capitalization across subsidy levels.

Table 3.13 presents results for the heterogeneity analysis. The coefficient on the interaction term shows to what extent the price effect for houses with a comparatively high exposure to the subsidy scheme differs from the remainder of houses in the sample. As before, the analysis includes regional time trends, excludes border regions of Thuringia, and controls for property characteristics. While the coefficient is positive for all subsidy schemes, it is only significant for the sample of houses that is most likely to be inhabited by eligible families with at least one, or by eligible families with two or more children. These findings confirm heterogeneous effects across property types, contingent on the exposure of properties to the subsidy scheme.

Table 3.13: Subsidy effects on asking prices of single-family houses: High and low subsidization probability

Dependent variable: log price per sqm				
	(1)	(2)	(3)	(4)
	overall	1+ child	2+ children	3+ children
Subsidy	0.0255*	0.0217*	0.0254**	0.0294**
	(0.0135)	(0.0130)	(0.0129)	(0.0129)
Subsidy * high subsidy probability	0.0177	0.0319**	0.0235*	0.0142
	(0.0138)	(0.0138)	(0.0138)	(0.0132)
PLZ FE	✓	✓	✓	✓
Time FE	Month x region	Month x region	Month x region	Month x region
Controls for Thuringia	Exclusion	Exclusion	Exclusion	Exclusion
Property controls	✓	✓	✓	✓
Max km to border	25	25	25	25
N	113,917	113,917	113,917	113,917
R-squared, first stage	0.0335	0.0826	0.1029	0.1281
Average price, baseline	333,869	306,118	304,117	287,117
Effect in euros, baseline	8,514	6,643	7,725	8,441
Average price, high probability	290,900	335,945	339,146	368,919
Effect in euros, high probability	12,567	18,007	16,584	16,085

Notes: This table shows the differential effect of housing subsidies in Bavaria on house prices. The treatment dummy indicates properties listed in Bavaria between July and December 2018. Average prices refer to prices in Bavaria prior to July 2018. Standard errors are clustered at the postal code level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

I subsequently quantify price effects based on the average pre-subsidy prices of Bavarian houses in both subsamples. In all low-probability samples, prices increase by less than 10,000 euros on average. The subsidy is only partially capitalized in segments of the real estate market that are in comparatively lower demand by subsidy recipients. In contrast, subsidies are fully capitalized for homes that are frequently demanded by eligible families. For an average house in the high probability subsample, the price effect closely resembles the difference between subsidies in Bavaria and in neighboring states. For example, a family with two children would receive up to 40,000 euros in subsidies in Bavaria, and up to 24,000 euros in other states. While subsidy levels differ by 16,000 euros, house prices increase by a just slightly larger amount in the corresponding high probability sample. This indicates that on average, families do not benefit from the subsidy scheme as it is fully capitalized into prices. Rather, the main beneficiaries are developers and existing homeowners that benefit from the appreciation in house prices. Subsidy recipients may however benefit from the reform if they choose to acquire properties that are less frequently bought by eligible households and, in particular, by eligible families.

Effects on Construction Activity

Subsequently, I follow the same methodological approach as in my baseline estimation to assess the subsidy scheme's effects on construction activity: I regress the number of annual construction permits per 1000 existing buildings on a treatment dummy for Bavarian municipalities in 2018, while accounting for municipality and time fixed effects. Standard errors are clustered at the municipality level. As before, I allow for differential regional time trends and estimate specifications without border regions to Thuringia.

Several aspects distinguish these specifications from prior estimations. First, local administrative data on construction permits is only available on an annual basis. Therefore, I am only able to estimate a treatment effect for the year 2018, pooling construction permits granted under the subsidy scheme with construction permits granted in prior months of 2018. This attenuates explanatory power vis-a-vis a setting which distinguishes construction permits granted early in the year and once the subsidy scheme became effective. Note however that in the absence of anticipatory effects in the first half of the year, the estimated effect should capture the change in the number of construction permits following the introduction of the scheme. Second, while price effects estimations control for a property's postal code, data on construction permits is only available at the municipal level, which often, but not always coincides with postal code areas. Larger municipalities and cities encompass several postal codes. To ensure a high degree of similarity between price and construction permit data, I weigh each municipality with its number of postal codes that are located within the distance band around the Bavarian interstate border. Results are shown in Table 3.14. Analogous to Table 3.6, estimations are based on municipalities within 25 kilometers of the Bavarian interstate border.

Specifications (1)-(3) assess the effect of the subsidy scheme on overall residential construction activity. Akin to Table 3.6, specification (1) neither allows for regionally differentiated trends, nor controls for the real estate transfer tax reform in Thuringia. Regional time trends are added in specification (2), while specification (3) additionally excludes border regions of Thuringia. Treatment effects are then decomposed into single family homes (specifications (4)-(6)) and houses with two or more apartments (specifications (7)-(9)). No significant effects can be observed for any specification. Note however that while the coefficients on overall construction activity and on single-family homes are positive, larger buildings with several units display a negative coefficient. While insignificant, these findings would be in line with the subsidy scheme slightly stimulating the construction of single-family homes, possibly accompanied by a partial crowding-out of multi-unit construction. As the construction sector has been operating at its capacity limits over the course of 2018, the latter could conceivably be related to price effects of the subsidy scheme on the construction sector.

Table 3.14: Subsidy effects on construction activity

Dependent variable: number of residential construction permits per 1000 buildings									
	All			Single-family houses			Multi-family houses		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Subsidy	0.266	0.184	0.381	0.381	0.326	0.524	-0.116	-0.142	-0.143
	(0.346)	(0.365)	(0.467)	(0.312)	(0.330)	(0.397)	(0.106)	(0.112)	(0.162)
Municipality FE	✓	✓	✓	✓	✓	✓	✓	✓	✓
Time FE	Month	Month x region	Month x region	Month	Month x region	Month x region	Month	Month x region	Month x region
Controls for Thuringia	✗	✗	Exclusion	✗	✗	Exclusion	✗	✗	Exclusion
Max km to border	25	25	25	25	25	25	25	25	25
N	3,264	3,261	2,139	3,264	3,261	2,139	3,264	3,261	2,139

Notes: This table shows the differential effect of housing subsidies in Bavaria on construction activity. The treatment dummy indicates Bavarian municipalities in 2018. Standard errors are clustered at the municipality level. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

3.6 Conclusion

This paper assesses the effects of direct housing subsidies on property prices. Intending to reduce purchase costs for owner-occupiers, both the German federal and the Bavarian state government introduced flat-rate direct housing purchase subsidies in 2018. Exploiting that Bavaria implemented a much more extensive subsidy scheme, I quantify capitalization effects in a difference-in-difference setting across the Bavarian interstate border. Based on a rich micro dataset on properties offered for sale, my results indicate that house prices increased by roughly 3.4% more in Bavarian border regions than in neighboring states. This is consistent with a full capitalization of the subsidy into the prices of single-family homes. In contrast, no significant effect arises for apartment prices, which can be attributed to apartments being rarely inhabited by owner-occupiers.

These results indicate that subsidy recipients do not necessarily benefit from the subsidy scheme. Instead, the subsidy scheme leads to an upsurge in housing demand, which is capitalized into prices. While subsidy recipients in market segments with lower price appreciation might still gain individually, prices of properties that are most likely to be inhabited by eligible households rise by the full subsidy amount. Thereby, the subsidy scheme also affects households who do not receive the subsidy, but nevertheless face higher prices. Homeowners who acquired their properties in prior years gain the most from the reform due to the appreciation of house values. On aggregate, the subsidy scheme thus redistributes from prospective towards preexisting home owners.

My results are consistent with the literature on real estate subsidies: While the German direct subsidy design substantially differs from other countries' subsidization through the tax code, substantial capitalization effects are well in line with the literature.

These findings are of high importance for the policy debate. My results show that due to the significant capitalization of subsidies into property prices, the recently introduced subsidy schemes fail to deliver on its promise to make housing more affordable.

While my results capture short-term effects, future research might address long-term effects on house prices and construction activity. As housing supply might be more elastic in the medium and long-run, long-term capitalization effects may plausibly differ from my findings.

4 Richer or More Numerous or Both? The Role of Population and Economic Growth for Top Income Shares

4.1 Introduction

Income inequality is a multi-faceted concept, and no single measure can capture all its relevant aspects. When studying top incomes and their importance in the income distribution, the most commonly used measure is the income share of a fixed percentile of the population. This top income share measure meets many of the desired distributional criteria that a proper inequality measure should meet. However, it is sensitive to developments in the size of the underlying population or in the real economy. A growing population implies that the size of the top group grows regardless of whether the incomes of its top earners change in real or even relative terms. Productivity growth that lifts a majority of the population to welfare levels previously only enjoyed by the rich would not result in higher top income shares if this growth affects everybody equally since the standard measure captures the income share of a fixed fractile of the population. Whether these properties are desirable for a top income inequality measure in times of population and economic growth or not depends on the research question. Nonetheless, it is important to notice that the standard top income share measure is only one of several ways of representing the relative status of top earners. The present study is the first to extend the scope of analyzing top incomes by proposing alternative measures of top shares that explicitly account for population and income growth. We apply these new measures to data from the United States for the period between 1917 and 2014 to analyze if and how long-run inequality trends differ when using different measures.

The seminal article by Piketty and Saez (2003) was the first to compute top income shares for the US for the entire twentieth century.[1] In its aftermath, many studies have analyzed different aspects of top incomes in the US and largely corroborated the main findings of Piketty and Saez.[2] However, this literature has paid little attention to how secular trends in

---

This chapter is joint work with Andreas Peichl and Daniel Waldenström and circulates as Krolage et al. (2018).

[1] Previous studies of US top income shares by Kuznets (1955) and Feenberg and Poterba (1993, 2000) focused on shorter time periods.

[2] See, e.g., Wolff (2002), Kopczuk et al. (2010), Atkinson et al. (2011).

real incomes, productivity and the population size have influenced top income shares. As over the past century, the US population has tripled and real per capita GDP has increased more than fivefold (see Figures 4.A.1 and 4.A.2 in the Appendix), it may be of first-order relevance to analyze how these factors influence inequality trends. To be able to conduct such analyses, alternative inequality measures are necessary as complements to the standard approach.

We propose three alternative ways to compute top income shares, all aimed at reflecting different aspects of income inequality. The new measures capture the roles of aggregate income and population growth in a different way than the baseline top income share measure does.[3] The first two measures fix an income threshold, which corresponds to a certain top fractile – say, the top 1 percent – in a given year, and then deflate that threshold over time using either consumer prices (our first alternative measure) or GDP (our second measure). The third measure instead defines a fixed top group in terms of number of units, for example the top 1 million earners, and then tracks their income shares over time.

The idea behind using these alternative top income shares is that they capture different, and yet relevant, aspects of top income inequality than the standard top share measure does. By fixing a real top income threshold above which all individuals are counted as top earners, we get a measure allowing both a higher average income and a larger group size to contribute to an increasing top income share. If the economy grows such that more people can consume the same amount of welfare (in constant consumer prices or constant overall output) as previous top earners, the economy produces more top income earners. Our new top share measures that use either CPI- or GDP-deflated top income thresholds thus allow us to answer another top inequality question: "Have the income rich become more numerous?". This question cannot be answered by the standard top share measure since it fixes the top group size relative to the size of the whole population.[4] By fixing the number of top earners, as in our final alternative measure, we remove the impact of population growth on the top income share. In the standard measure, top shares may increase as the population grows simply because the size of the top group has increased as it is a fixed share of the total. Of course, the population growth effect on top income shares can be either positive or negative. A positive effect arises if population growth mainly implies adding low-income earners, perhaps due to immigration of low-skilled people. This would make the top group expand mechanically as it is a fixed share of the total, and lead to an inclusion of more and more relatively well-paid individuals. The standard top

[3] We also compare our results with the new series using distributional national accounts (DINA) using the data from Piketty et al. (2018) that are available since 1960 (c.f. section 4.4.3).

[4] These measures are thereby related to the headcount measures commonly used in poverty analysis. See Peichl et al. (2010) for a comprehensive discussion and analysis of such measures in analyzing affluence.

share measure would then increase even though the real incomes of the top earners have not changed at all.

This analysis is closely related to the literature on absolute versus relative inequality measurement. The baseline top income share measure is a purely relative measure, while our three alternative measures are either hybrids (those fixing a relative income threshold and then deflating to capture absolute income changes) or a purely absolute measure (fixed top group). Relative inequality measures have been by far the most widely used, and for good reason. Relative income measures are scale-invariant, allowing for comparisons over time despite changes in overall population size or incomes earned. However, the inability of relative income measures to reflect changes in the structure of incomes or income groups has spurred discussions among inequality researchers about whether one should always favor relative over absolute notions of inequality (see, e.g., Atkinson and Brandolini, 2010; Bosmans et al., 2014, in the context of global inequality measurement). Related to this, survey evidence suggests that many people perceive absolute (rather than relative) differences in incomes as being an important aspect of inequality (see, e.g., Amiel and Cowell, 1992, 1999). Therefore, in addition to relative measures, other inequality measures that reflect either an absolute or intermediate notion of inequality have been proposed in the literature. We contribute to this literature by providing the first alternative top income shares that allow for other notions of inequality compared to the standard approach which takes a relative point of view.[5]

Our empirical application to the US income distribution shows that the broad U-shaped top income share trends documented by Piketty and Saez are also apparent in our new measures, but with some interesting qualifications. Compared to the standard measure, the decrease in inequality up to the 1980s is larger when using GDP-deflated top incomes and a fixed top group but smaller when using CPI-deflated top incomes. This variation reflects that the economic hardship during the wars and the economic crises caused the top groups to shrink in size. In the post-1980 era, the standard top shares have increased by an order of magnitude, e.g., the top percentile share has almost doubled between 1980 and 2010. However, both the CPI- and GDP-deflated top income shares have increased at an even faster pace. In addition to experiencing a higher income growth, these measures reflect that the top groups have also grown in size; more people today qualify as top income earners than they did 30 years ago when required to earn the same absolute income as in that era. The steeper top income share trend when using GDP-deflated incomes suggests that productivity growth in society has been

[5] This exercise is also related to the recent debate about absolute versus relative social mobility in the United States (see, e.g., Chetty et al., 2014a, 2014b, 2017).

disproportionately reaped by top earners.[6] Looking at the measure using a fixed number of top earners, the share is actually lower than the standard measure, reflecting a relative shrinkage of this group.

We also find interesting patterns for the top groups just below the highest percentile, suggesting that influences from population and economic growth are not homogeneous within the top tail of the income distribution. In contrast to the expanding group size observed in the top percentile, we observe rather stable population shares in lower top brackets using the two fixed-income threshold measures (the fixed number of earner-measure is, of course, unchanged). Their average incomes are also relatively similar to the standard top measures, which jointly results in limited growth, or even slight decrease, in the income share of these groups. In other words, the proportion of individuals with above-average productivity gains are found almost exclusively in the highest-earning percentile, not in top groups lower down the distribution.

We subsequently decompose the different top percentile measures' post-1980 growth into their underlying factors. This allows us to more precisely identify to which extent the development of top income share measures is driven by changes in average top incomes, changes in the number of taxpayers and changes in the lower part of the earnings distribution. Notably, the number of taxpayers who increase their real income, as well as the number of taxpayers who benefit more than average from economic growth, increase at faster rates than the overall population.

Finally, when running the analysis across income sources, we investigate the extent to which the increasing importance of wage income is attributable to increasing numbers of taxpayers belonging to the top percentile. While the baseline results show an increasing importance of wage incomes at the top, our results point to a more nuanced development. We find that for the very top, the relative importance of wages vis-à-vis capital and entrepreneurial income has been slightly declining over the past three decades.

The contributions of our study are directed primarily at the income inequality literature. While we are not the first to examine how different measurement approaches influence top income share trends, previous studies have focused on the concept of income, for example the effect of subtracting taxes and adding unrealized capital gains (Armour et al., 2013), on changing the computation of business income (Alstadsaeter et al., 2016) or on using national accounts-equivalent income measures (Piketty et al., 2018). The recent literature on distributional national accounts (DINA) is of particular relevance for our analysis because it explicitly exam-

6 See Dew-Becker and Gordon (2005) for an early discussion of the distribution of US productivity growth.

ines the distributional effect of including the entire national income instead of only the fiscal income concepts used so far in the top income literature. For this reason, we conduct a further analysis of how our measures perform when using DINA-incomes rather than fiscal incomes.[7] In addition, Auten and Splinter (2019) re-estimate top shares accounting for tax base changes, income sources absent from tax records and changing marriage rates. With their methodology, income shares are shown to increase at a much lower rate. By contrast, our analysis keeps the income concept unchanged throughout and instead focuses on different statistical measures of top shares and their composition.

Our findings also add to the research literature assessing long-run trends in top income shares in the Western world, specifically in the US. Most of these studies use the baseline definition of top income shares (Piketty and Saez (2003), Atkinson and Piketty (2007, 2010), Leigh (2009), Atkinson et al. (2011) and Roine and Waldenström (2015)), but in some cases, they use other data sources to compute the top share (Burkhauser et al. (2012)). Our analysis complements these studies by asking how the picture would change if one considers further aspects, such as the variation in the size of top groups as in the fixed threshold measures. Using different top share measures and a richer compositional analysis that comes with it could also provide insights for cross-country comparisons of the historical evolution of inequality.

4.2 Empirical Approach

4.2.1 Methodology

Following Piketty and Saez (2003), the literature typically uses top income shares, i.e. the share of total income going to the top x percent of the population, as inequality measures for the top of the income distribution. This standard top income share measures the amount of income of a fixed fraction of the population but without accounting for changes in the composition of the population or in the distribution of income among the rich.[8]

In order to investigate the impact of population and real income growth on top income inequality, different measures are necessary to complement the standard approach. Therefore, we propose three alternative ways to compute top income shares. Our main analysis then

[7] There is a current debate about how to compute DINA incomes, for example, in the US case (Auten and Splinter, 2019). While our analysis is about the relative performance of different top share measures rather than which underlying income one wishes to use, we nevertheless apply our methodology to both standard and DINA top income shares.

[8] See Peichl et al. (2010) for a comprehensive discussion of alternative measures to analyze top income inequality.

consists of comparing the trends in top shares for the different measures. Comparing the differences between the four measures allows us to single out the contributions of population and economic growth to the observed inequality trends. The purpose is not only to obtain a picture of the sensitivity of these trends to these variations but also to explicitly account for the growth of the population and the overall economy. To do so formally, we also propose a decomposition analysis that separates how much of inequality growth is due to rising incomes at the top versus declining incomes of the remainder of the distribution in addition to changes in population size.

As discussed in the introduction, relative inequality measures have been by far the most widely used in empirical analyses. However, other inequality measures have been proposed in the literature which do not satisfy the relative scale invariance axiom but rather give an absolute or intermediate notion of inequality. We contribute to this literature by providing the first alternative top income shares that allow for absolute or intermediate notions of inequality. The aim of this exercise is not to show that the standard approach (and hence the scale invariance axiom) is incorrect. Rather, as in the literature on global inequality (see, e.g., Atkinson and Brandolini, 2010; Bosmans et al., 2014), the aim is to complement the standard approach by combining relative, absolute and intermediate measures in the same analysis to provide a fuller picture of inequality developments.

We compute top income shares for four different top groups that differ in whether their population share and group size are variable or fixed:

1. *Standard measure*: Top income share used in Piketty and Saez (2003): fixed population share, variable group size (in number of earners).

2. *CPI-deflated threshold*: Top income or population share of those earning above CPI-deflated income threshold: variable population share, variable group size.

3. *GDP-deflated threshold*: Top income or population share of those earning above GDP-deflated income threshold: variable population share, variable group size.

4. *Fixed group size*: Constant number of top earners: variable population share, fixed group size.

Measure A is the baseline definition that has been used in the top income literature. It is defined as the share of total incomes earned by a fixed fractile (e.g., top decile or top percentile) of the population. This standard top income share is a relative inequality measure that, among others, satisfies the scale invariance axiom, i.e. inequality remains unchanged if all incomes are multiplied by the same factor.

Measure B refers to income earners with an income above a level that is linked to an income threshold of a certain top group (e.g., the 99th percentile threshold for the top percentile group) in a specific year. We deflate this threshold using the CPI, where the base year is either the year 1980, which serves as a focal point in the inequality literature due to the substantial tax reforms implemented in the early 1980s, or the first or the last year in our sample, 1917 and 2014, respectively.[9] Since this measure varies in both income and population shares, we compute both top income shares and top population shares (headcount ratios).

Measure C top shares are computed in the same way as those for measure B, except that we deflate the income thresholds using per capita GDP instead of CPI to capture the overall productivity growth in the economy. This step helps investigate the extent to which top incomes have grown more quickly than the overall economy.[10] If increases in top real incomes were attributable solely to economic growth and not to changes in the income distribution, income and population shares above the GDP-deflated thresholds should remain roughly constant over time. In a robustness check, where we use the DINA data, we apply the same national income price index as Piketty et al. (2018).

Finally, measure D is the top income share of a constant number of top earners, such as the top one million earners in the distribution. We include this measure since rising income shares of measures B and C may reflect two effects: a rising number of taxpayers above the fixed income thresholds and rising incomes of these earners. By fixing a number of high-earning taxpayers, we isolate the latter effect. If population growth was distributionally neutral, the income share of a given number of top earners should decline proportionally with population growth. The extent to which the observed pattern deviates from this pattern is informative about inequality trends within the top group.

To sum up, measure A is a relative top income share measure, measure D is an absolute one, while measures B and C are hybrid measures providing an intermediate view. Which one to use in an empirical analysis depends on the context and research question at hand. We suggest to use all of them to investigate the full picture of inequality trends.

[9] To be precise, we use the same inflation measure as Piketty and Saez (2003) which is the so-called CPI-U-RS series. In a robustness check, where we use the DINA data, we apply the same national income price index as Piketty et al. (2018).

[10] Per capita growth is chosen instead of overall growth to net out the growth effect of the changing population size.

Decomposition analysis. We conduct two different decomposition analyses of the top share measures in order to gain further insight into the forces driving them. The first is to decompose changes in the top percent income share into the contributions of population size, group size (for measures B and C), overall income and average income growth in the top and bottom groups. More precisely, the income share S_i of top fractile i can be decomposed as

$$S_i = \frac{\bar{Y}_i N_i}{\bar{Y} N} \Leftrightarrow \Delta ln S_i = \Delta ln \bar{Y}_i + \Delta ln N_i - \Delta ln \bar{Y} - \Delta ln N, \quad (4.1)$$

where $\bar{Y}$ ($\bar{Y}_i$) indicates average income (in fractile i) and N (N_i) indicates the number of tax units (in fractile i).

This decomposition allows us to more precisely identify the extent to which changes in top income share measures are driven by changes in average top incomes, changes in the number of taxpayers, and changes in the denominator, i.e., in the incomes and size of the remaining population.

In a second decomposition, we analyze the effect of different sources of income on overall income trends: wages, capital income (excluding realized capital gains) and business income. As before, we strive to neutralize the effect of population growth by fixing a certain number of top taxpayers and tracking the development of their incomes over time. We take the same fixed numbers of taxpayers, calculate their average revenue derived from each income category and derive the share of each group's respective income categories in aggregate US income.

4.2.2 Data

Our estimates are based on the standard source of international top income data: The World Inequality Database, and its predecessor, the World Wealth and Income Database. These estimates encompass income shares and percentile thresholds for the US over the years 1917–2014, using incomes from all sources before tax and deductions and most transfers. These data come from tax returns statistics compiled by the IRS and have been adjusted to consider changes in the tax law (Piketty and Saez, 2003). Thresholds and annual incomes are computed in real terms to ensure their comparability across years. Realized capital gains are not included in the baseline calculations, but we show in the appendix that our findings are not sensitive to the treatment of capital gains.

As is common in this literature, the units of analysis are the income tax units (single or married households) and they are related to the total number of potential tax units in the population calculated from census data. Income thresholds and shares are calculated assuming a Pareto distribution to approximate income shares of top fractiles.

Due to the large income growth over the past century and the fact that the World Wealth and Income Database provides data on only the top income decile, we sometimes need additional data for those cases when more than 10 percent of tax units surpass a given threshold. If needed, we supplement the data with IRS SOI[11] tax statistics on larger income brackets than given in the World Wealth and Income Database. The corresponding data sets are available from 1986 onwards and capture the income shares and thresholds of rather broad income brackets.[12]

In addition, we contrast our findings with recent DINA measures. These are based on different concepts for calculating top income shares and strive to capture the full scope of national accounts (Piketty et al., 2018).

4.3 Top Income Share Trends Across Measures

We start our empirical analysis with a broad comparison of the four different top share measures by analyzing both income and population shares of the top group and their long-run developments. In the next subsections, we analyze our new measures in more detail by looking at subgroups of the top 5 and the top 1 percent. In section 4.4, we dig deeper by decomposing the various measures into their components as well as investigating differential trends by income source.

4.3.1 The Top 1 Percent Share

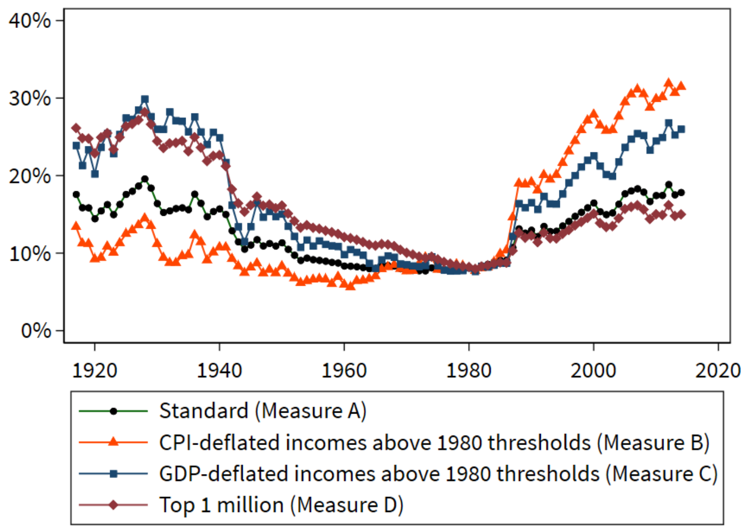
Figure 4.1 presents the evolution of the top 1 percent pre-tax income share in the US between 1917 and 2014 according to the four top income share measures presented above.[13] Since it commonly serves as a focal point in the literature, we use the year 1980 as a reference year (see Figures 4.A.6 and 4.A.7 in the Appendix for reference years 1917 and 2014). Measure A, the standard measure as used by Piketty and Saez, exhibits the marked U-shaped pattern that has been described numerous times in the past literature, showing a share beginning high at 15-20 percent, decreasing to under ten percent and finally rising to prewar levels in

[11] Source: https://www.irs.gov/uac/SOI-Tax-Stats-Historical-Table-3

[12] As Piketty and Saez's computation procedure results in a slight divergence between World Wealth and Income data and IRS data, adjustments were made to ensure that the 10-percent threshold of the IRS data matched the 10-percent level of our base dataset.

[13] Figure 4.A.4 in the Appendix shows graphs for different top income groups.

Figure 4.1: Different measures for the top 1 percent



Notes: This figure plots income shares for the four different top inequality measures (as described in section 4.2.1): the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980 (see Appendix Figure 4.A.5 for 1917 and 2014 as baseline year). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

recent years. Measure B, which shows the income share of earners with an income above the CPI-deflated 1980 income threshold (that is, the 99th percentile) looks different. It is much lower in the prewar era, hovering around ten percent, where it stays until the mid-1980s, when it rises rapidly, reaching over 30 percent in the 2000s. Measure C, the GDP-deflated 1980 income threshold is also different. This share drops drastically from over 30 percent in the interwar era to ten percent in the 1960s, a fall that is twice as large as in the baseline series. Moreover, while the share remains rather similar to the baseline until the mid-1980s, it displays a substantially higher growth rate afterwards and again reaches the level it had amounted to a century ago. Finally, measure D shows the income share of the top 1 million tax units – the approximate number of tax units in the baseline around 1980. This series is the most similar to the baseline, with deviations primarily reflecting the effect of differences in group size (being relatively larger before 1980 and relatively smaller afterward).[14]

[14] Qualitatively, the results do not change by much when we analyze incomes including capital gains (c.f. Figure 4.A.8 in the Appendix).

The main message in Figure 4.1 is that developments within shorter periods (though still several decades) are apparently sensitive to how one defines the top income shares. Our analysis thus shows that having different types of top income share series provides new insights on the evolution of income inequality. The difference between the baseline and the fixed threshold series reflects that not only the relative incomes of top earners but also the size of their group matter. Deflating the threshold using GDP provides information about the distribution of the economy's overall productivity gains: from the 1930s to the 1950s, they went mostly to the bottom 99 percent, while they subsequently were roughly equally distributed between the top 1980 percentile and the rest till the late 1980s. From then onwards, they predominately accrued to the upper part of the earnings distribution. However, this development is partly attributable to a rising fraction of individuals sharing the productivity gains above the GDP-deflated 1980 threshold.

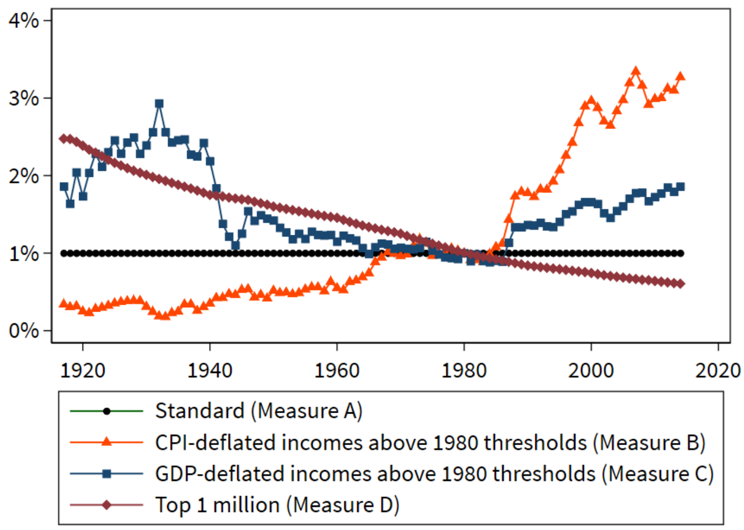
Next, we examine the evolution of top population shares using the four different measures. This analysis differs from the analysis of top income shares and it addresses related but yet slightly distinct questions. For example, top population shares are informative regarding the degree of concentration among the rich (see also Atkinson (2008)) and the absolute number of high-earning individuals in an economy. Figure 4.2 shows the population shares of the top percentile income level for each of the four different measures.

Measure A, the baseline, is by construction fixed at the one percent level and entirely uninformative about inequality trends. Similarly, measure D falls steadily along with population growth since the share of a fixed group size (the top 1 million) falls as the population grows.

By contrast, measures B and C are more informative. Their developments correspond relatively well to what was shown in Figure 4.1. The share of tax units above the CPI-deflated 1980 threshold (measure B) increases throughout the past century. This effect is due to both productivity growth – increasing real incomes over time – and changes in the income distribution. Between the 1980s and the 2000s, the top population share increased threefold, which reflects that top incomes increased more than consumer prices.

The share of taxpayers above the GDP-deflated 1980 top percentile threshold (measure C) – which should follow a flat trajectory if tax units along the entire income distribution benefit similarly from productivity growth – has been subject to different developments during the past century. The share decreases quite sharply in the first half of the twentieth century, and increases again from the 1990s onwards, albeit at a much lower level than the CPI-deflated shares. In other words, when requiring that a top percentile income should match the 1980 level deflated using GDP, less and less top earners have been able to make it to the top in the previous century. This group has again expanded in recent years.

Figure 4.2: Population share of top 1 percent earners across measures



Notes: This figure plots population shares for the four different top inequality measures (as described in section 4.2.1): the standard measure (A; equal to 1 percent by construction), the population shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the population share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980 (see Appendix Figure 4.A.5 for 1917 and 2014 as baseline year, respectively). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

Table 4.1 presents an overview of average annual growth rates of the top percentile share for the different measures during subperiods over the past century. The table distinguishes between the growth in income shares and the growth in population shares, where the latter corresponds to zero for the baseline shares but is more meaningful for the other measures. Looking first at the income share growth rates, there is a fairly large correspondence in the signs of growth rates across measures within subperiods. In one case, the sign differs: the CPI-deflated top share during the 1950-1980 period shows a positive rate, while rates are negative for the other measures. However, the magnitudes of growth differ quite notably between the baseline and the other measures in several cases, especially during the postwar era. The largest growth rates are observed in the 1980-2000 period. Incomes above the GDP-deflated threshold have exhibited a much higher growth rate than the underlying population share, while this gap is much smaller for the CPI-deflated threshold. This underlines that over this period, top earners have particularly benefited from economic growth.

Table 4.1: Average annual growth rates of income and population shares for different top 1 percent measures

Annual growth in income and population shares								
Period	Standard (Measure A)		CPI-deflated top thresholds (B)		GDP-deflated top thresholds (C)		Top 1 million (D)	
	Income	Pop.	Income	Pop.	Income	Pop.	Income	Pop.
1917-1929	2.4	0	3.4	3.1	2.0	0.1	1.4	-1.6
1929-1950	-2.1	0	-1.8	3.0	-2.1	-1.4	-2.3	-1.1
1950-1980	-0.7	0	1.0	2.9	-1.1	-0.8	-1.8	-1.5
1980-2000	4.4	0	7.5	6.2	6.7	2.7	4.0	-1.5
2000-2014	0.4	0	0.6	0.5	0.7	0.0	0.0	-1.5

Notes: This table shows average annual growth rates of the top percentile share in percent for the different measures during subperiods over the period of analysis. The four different top inequality measures (as described in section 4.2.1) are: the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980. The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

4.3.2 Top Groups Above Fixed Thresholds (Measures B and C)

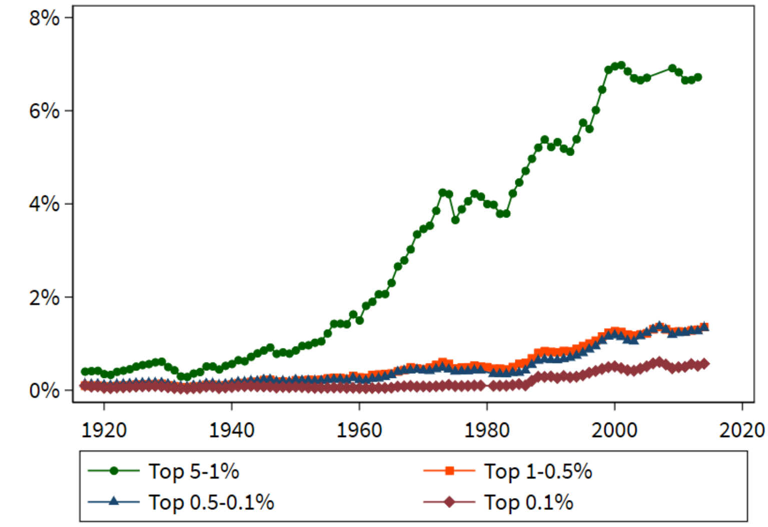
We now zoom in on the different alternative top share measures by taking a closer look at the long-run developments of different top fractiles and their income shares in the different top share measures.

Measure B is based on CPI-deflated thresholds. Figure 4.3 shows the income shares and the corresponding shares of taxpayers for income brackets above the CPI-deflated 1980 thresholds. Comparing real incomes across years instead of comparing relative positions on the income distribution in a year constitutes one approach to accounting for population growth. We consider the top 5, 1, 0.5 and 0.1 percent thresholds. In order to highlight differences along the top, shares are displayed for the brackets in between the thresholds. This analysis hence decomposes the top 1 percent shares reported in Figures 4.1 and 4.2 into three groups: top 1–0.5, 0.5–0.1 and top 0.1 percent and additionally shows the group just below: top 5–1 percent.

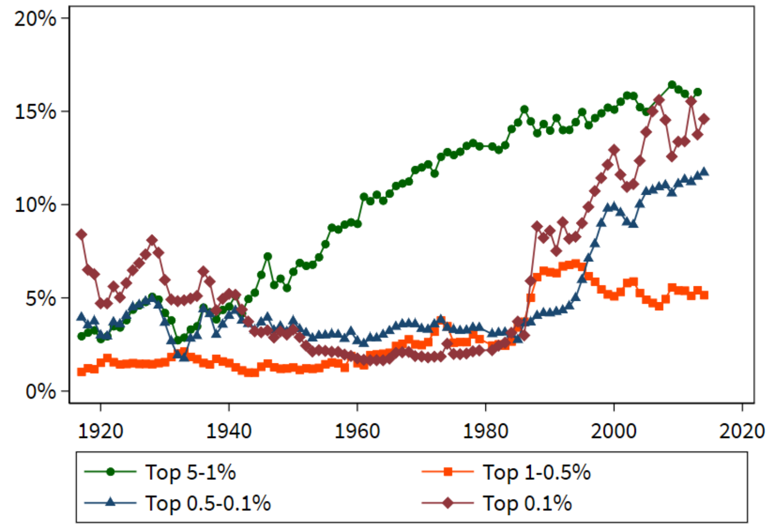
Accounting for population growth reveals different developments than the well-known top fractile results (c.f. Figure 4.A.3 in the appendix). The respective trajectories follow different patterns, deviating from the familiar U-shaped one. The developments of income shares do not closely follow the growth of population shares above thresholds. Instead, the income

Figure 4.3: Top shares of those above CPI-deflated 1980 top thresholds (Measure B)

(a) Percentage of taxpayers above CPI-deflated 1980 top thresholds



(b) Income shares of taxpayers above CPI-deflated 1980 top thresholds



Notes: This figure plots population and income shares of those earning above the CPI-deflated 1980 income threshold (Measure B, as described in section 4.2.1). The measure is constructed such that it equals the standard top 1 percent share in 1980 (see Appendix Figure 4.A.6 for 1917 and 2014 as baseline years). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

share between the top 1-0.5 percent thresholds remains rather constant, with some growth in the 1990s. In contrast, the income share above the uppermost threshold has experienced a large increase since the 1980s, while declining and then remaining at a roughly constant level in previous decades. The income share of taxpayers between the 1 percent and 5 percent thresholds has almost continuously grown since the early 20th century. These developments cannot be fully explained by economic growth or by a simple fanning out of the income distribution. A proportional growth of all incomes or an increased dispersion of incomes would shift the tail of the distribution outward, leading to higher income shares for all – not just some – upper thresholds. Instead, Figure 4.3 points to a more nuanced development than one might infer from the evolution of standard top income shares. Possible drivers are discussed in detail in section 4.4.2.

Measure C is based on GDP-deflated thresholds. The above results are driven partly by economic growth. We use per capita GDP-deflated thresholds (see Figure 4.A.2 in the appendix for GDP growth) to assess whether top income earners have more than proportionally benefited from economic growth. If economic growth is equally distributed across the income distribution, the population and income shares above the GDP-deflated thresholds should remain roughly constant over time.

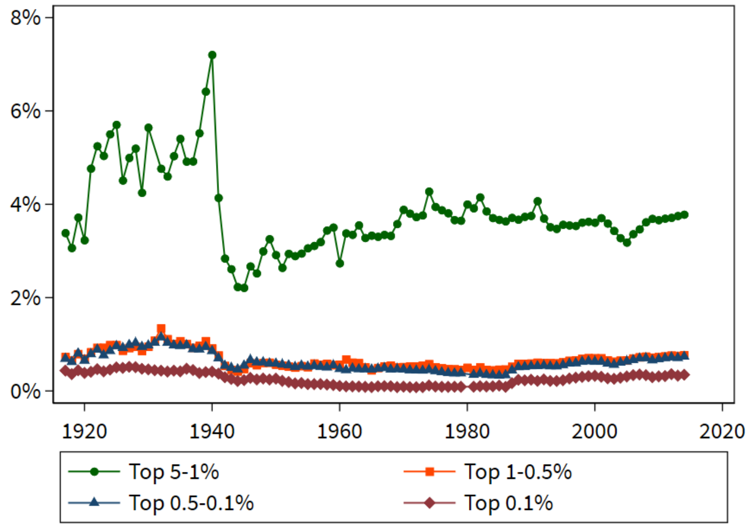
As shown in Figure 4.4, such constant population and income shares cannot consistently be found in the data for GDP-deflated thresholds. As before, the figure depicts developments for the GDP-deflated top 5-1, 1-0.5, 0.5-0.1 and 0.1 percent 1980 thresholds.[15] Differing findings emerge over time and for rather narrow and larger GDP-deflated thresholds. As shown in subfigure (a), the percentage of tax units above all GDP-deflated top thresholds declined around World War II. The 1950s and 1960s witnessed diverging developments: more and more tax units exceeded the GDP-deflated 5 percent threshold during the mid-twentieth century, whereas the fraction of taxpayers above the GDP-deflated top 0.1 percent threshold shrank. Hence, more and more people with high – but not extremely high – incomes have benefited more than proportionally from economic growth. As, at the same time, the percentage of tax units above very high income thresholds continued to decline throughout the mid-twentieth century, incomes at the larger top of the distribution became more equal. In the late 1970s, the result reverses. From then onwards, the share of tax units above the 5 percent threshold ceased to grow, whereas the number of tax units at the very top grew significantly.

For income shares, the picture looks a little different. In line with other findings in the literature, the income share of those above the GDP-deflated 1980 0.1 percent threshold – the highest in our computations – experienced a stark decline throughout the first half of the twentieth

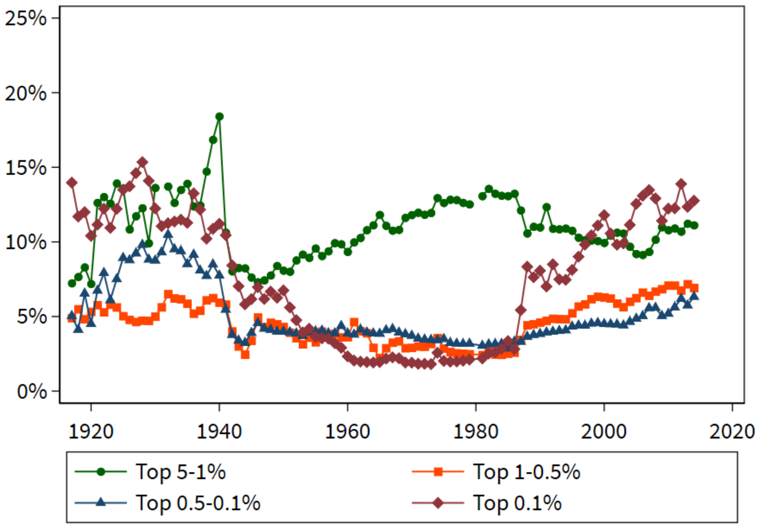
[15] Similar findings emerge for 1917 and 2014 thresholds. See Figure 4.A.7 in the appendix.

Figure 4.4: Top shares of those above GDP-deflated 1980 top thresholds (Measure C)

(a) Percentage of taxpayers above GDP-deflated 1980 top thresholds



(b) Income shares of taxpayers above GDP-deflated 1980 top thresholds



Notes: This figure plots population and income shares of those earning above the per capita GDP-deflated 1980 income threshold (Measure C, as described in section 4.2.1). The measure is constructed such that it equals the standard top 1 percent share in 1980 (see Appendix Figure 4.A.7 for 1917 and 2014 as baseline years). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

century. Remaining rather low until roughly 1980, this group's income share has seen rapid increases over the past three decades. Since the beginning of the twenty-first century, the income share of the top bracket has been subject to even more pronounced fluctuations than the fraction of tax units above those thresholds. However, whilst decline and growth periods coincide with the baseline, the share develops at a much larger magnitude.

The income shares between the GDP-deflated top 1-0.5 and 0.5-0.1 percent brackets shrank in the first half of the twentieth century and began growing again in the 1980s. The time trend is substantially less pronounced than for the tax units above the top GDP-deflated threshold. In contrast, the decline of the 5-1 percent share halted after World War II, with the share then rising until the early 1980s. From the late 1980s onwards, it experienced a slight decline, only to remain roughly constant in more recent years. Hence, the dispersion of incomes does not evenly affect all high incomes.

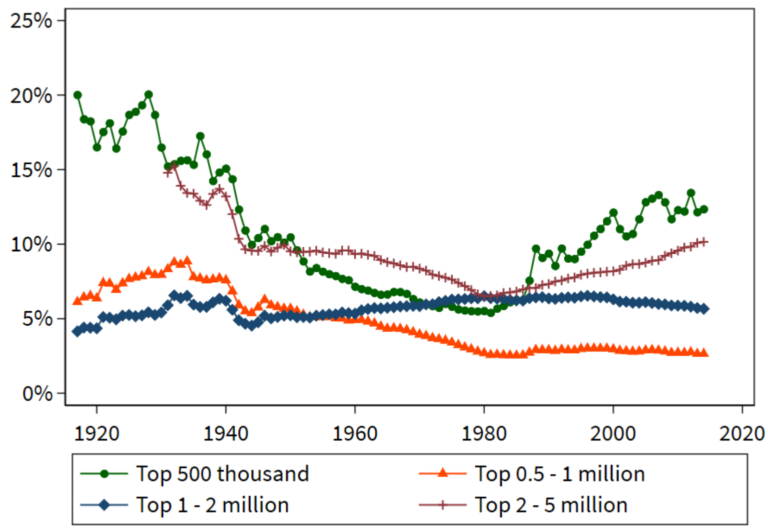
4.3.3 Constant Number of Top Tax Units (Measure D)

While the comparison of income shares above thresholds indicates a pronounced growth-exceeding increase only for incomes at the very top, it is driven by two effects: first by changes in the number of taxpayers above those income thresholds and, second, by changes in the income allocated to them. Hence, from Figure 4.4 alone, one cannot seamlessly infer which part is attributable to the presence of more or fewer taxpayers exceeding the threshold, and which is attributable to individual taxpayers above the threshold increasing their respective incomes.

To net out the effect of population growth and separate these two effects, we analyze how the income share of a fixed number of taxpayers evolves over time (see Figure 4.5). For this, we assess the income shares of the top 500 thousand, 500 thousand - 1 million, 1-2 million, and 2-5 million taxpayers.[16] Recall that distributionally neutral population growth should lead to a proportional decline in this measure. However, this is not the case and this measure likewise shows an increase in inequality in recent decades. Splitting top income shares into sub-brackets yields more thorough insights on the effects at work. Most importantly, the different income brackets at the top do not seem to follow the same general trend. Instead, while incomes at the very top experience the widely discussed increase since the 1980s, this increase does not necessarily apply to the slightly lower tier of top income recipients.

[16] For the early 1900s, the income share of the top 5 million tax units cannot be calculated because they constituted more than 10 percent of all taxpayers in these years.

Figure 4.5: Top shares of fixed numbers of earners (Measure D)



Notes: This figure plots income shares of fixed numbers of top tax units (Measure D, as described in section 4.2.1). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

In line with previous research, the income share of the uppermost income group – here, the top 500,000 – sharply declined until 1980, when it began to substantially increase. A possible explanation is that the richest individuals derive a substantial share of their income from capital and entrepreneurial activities and from performance-tied compensation, e.g., via bonus payments and stock options. These sources of earnings are more volatile and more tied to the business cycle than wages in lower income brackets. In line with this, the spikes and troughs in recent years may be explained by the Dot-com bubble and the Great Recession. The picture completely changes for the top 0.5 to 2 million, who did not improve their income share over time. This starkly contrasts with the top 2 to 5 million, who, similar to the very top, have seen increasing income shares since the 1980s. General trends in the developments are fairly robust to excluding or including capital gains.

The above findings accentuate that the top 1 percentile (which would have encompassed several of the above groups in 2014) is far from being a homogeneous group. Instead, there seem to be differential forces at work that distinguish the top segments from one another. An explanation might well be found by differentiating why certain taxpayers belong to the top. In other words, to what extent are rentiers, who derive most of their income from capital, entrepreneurs, and top managers, who all receive rather high incomes, subject to different time trends? The decomposition in section 4.4.2 sheds light on this question.

4.4 Decomposition Analysis and Robustness

4.4.1 The Role of Income and Population Changes

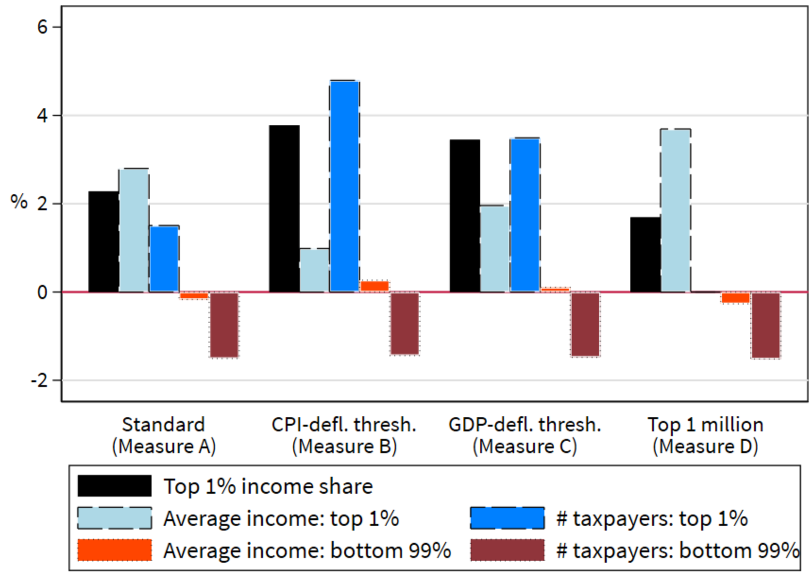
We now implement the decomposition analysis presented in section 2 (see equation 4.1), which decomposes changes in the log income share of the top percentile into changes in its average income, in the average income of the lower 99 percent, and of overall population growth. As a starting point, Figure 4.6 depicts the average annual changes in log income shares over the 1980-2014 period for the previously discussed core measures: the Piketty and Saez baseline top 1 percent income share, the income shares above CPI- and GDP-deflated 1980 top 1 percent thresholds, and the top 1 million taxpayers. These changes are then decomposed into their contributions. That is, we display average changes in average incomes for the top 1 percent and for the bottom 99 percent measures as well as the changes in the number of taxpayers in the respective top 1 percent and bottom 99 percent. As indicated by equation 4.1, positive changes in average top incomes increase income shares, while increases in the bottom 99 percent's average incomes lower income shares. At the same time, increases in the number of tax units above (below) the threshold have a positive (negative) effect on the top income share. Therefore, in order to be consistent with equation 4.1, increases in the bottom population and income share appear as negative numbers in the figure.

The income share above the CPI-deflated threshold grows more on average than the unadjusted top income share. While average incomes above the threshold experience positive average growth, the effect is driven substantially by an increasing number of taxpayers exceeding the threshold. More precisely, the number of taxpayers above the threshold experiences much faster growth than the overall number of taxpayers. That is, in both absolute and relative terms, an increasing number of taxpayers earned at least as much real income as the top 1 percent in 1980.

To a lesser degree, this also holds true for the income share above the GDP-deflated 1980 threshold. The growing income share is substantially driven by the increasing number of taxpayers at the top, again rising at a much faster pace than the overall number of taxpayers. Growing average income at the top also plays an important role. This development is however partly attributable to the threshold increasing with per-capita GDP, which ceteris paribus results in growing average income above the threshold.

By construction, the number of taxpayers in the top 1 million stays constant. The positive growth in their income share is thus driven entirely by the top 1 million experiencing substantially higher income growth than the rest of the population. Nevertheless, the remainder of the population was also able to increase their real incomes, albeit at a much smaller scope.

Figure 4.6: Decomposing the top percentile growth, 1980-2014



Notes: This figure plots the results from the decomposition analysis presented in section 2 (see equation 4.1), which decomposes changes in the log income share of the top percentile into changes in its average income, in the average income of the lower 99 percent, and of overall population growth (number of taxpayers in top vs. bottom). Results of this decomposition over the 1980-2014 period are presented for standard measure (A), the income shares above CPI- (B) and GDP-deflated (C) 1980 top 1 percent thresholds, and the top 1 million taxpayers (D). In order to be consistent with equation 4.1, increases in the bottom (top) population and income share appear as negative (positive) numbers in the figure. The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

That is, not only do the various measures yield differing developments, but the factors driving these measures also vary. The number of taxpayers increasing their real income, and the number of those that benefit more than average from economic growth, grow faster than the overall population as reflected in the baseline.

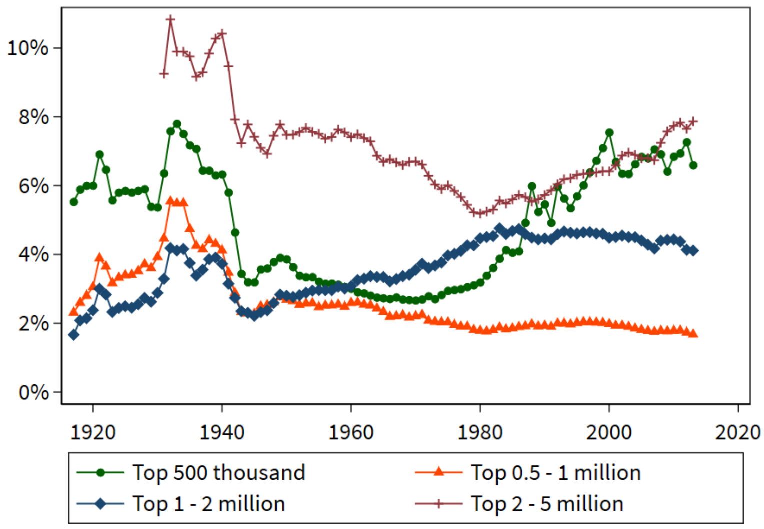
4.4.2 Differential Effects by Income Source

To further assess which factors drive the differential development of top income shares, we approximate the contribution to income shares of wages, capital (dividends, interest and rents) and entrepreneurial income. According to past research (cf. section 4.1), wage income became increasingly important for top percentiles in the second half of the twentieth century. At the same time, the relative importance of capital income declined. While these results have been widely discussed, our previous question also applies here: To what extent are these findings driven by changes in the denominator? As above, with the population increasing threefold over the course of the past century, taxpayers with comparably lower incomes – and hence a larger share of wage income on average – moved to higher fractiles of the income distribution. Hence, to what extent is the increasing importance of wage income attributable to more and more taxpayers belonging to the top 1 or top 0.1 percentiles?

For this analysis, we track the same top 500 thousand, 1 million, 2 million, and 5 million taxpayers as above and decompose their incomes according to their sources. When the number of taxpayers in each income bracket is held constant, a different picture emerges than that for top percentiles, which do not account for population growth. The following describes the developments by income source.

Wage income. This category encompasses all income derived from dependent labor, i.e., wages, salaries and pensions. Figure 4.7 displays wage income's share in the overall population's aggregate income. The shares of top income groups' wages in aggregate income have grown since the 1980s for both the top 500,000, as well as for the top 2 to 5 million.

Figure 4.7: Wage shares of top income brackets (Measure D)



Notes: This figure plots wage income shares of fixed numbers of top tax units in aggregate income (Measure D, as described in section 4.2.1). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

In particular, the overall income shares of the uppermost category's wages have roughly tripled during the past three decades. In contrast, the top 0.5 to 2 million's wages have not increased relative to overall income.

These findings diverge from the results computed with total income. While wages became increasingly important for earners at the very top and in the middle-upper class, a similar effect fails to manifest for income earners between these groups. Comparing wage income to entrepreneurial and capital income shares, though, the relative importance of wage income has even been slightly declining at the very top of the income distribution. Note, though, that while we track a constant number of top income earners, we cannot track individuals. Salary increases for top-earning executives might have lifted them into a higher bracket, in turn shifting taxpayers who are less dependent on wage income – and therefore did not benefit from a similar increase – to a lower income bracket. Hence, the lack of an increase in the top 0.5 to 1.5 million's wage shares might be partly attributable to changing compositions of taxpayers at the top.

However, the recent growth at the very top is consistent with the 'superstar' hypothesis of Rosen (1981). Globalization and technological change, particularly in the realm of information and communication technologies, have led to expansions of scale during the past three decades. Hence, those with the very highest abilities have managed to obtain larger and larger rents. The reach of those with 'second-best' abilities, however, is limited by these 'superstars', explaining why the importance of wages has not risen for the top 0.5 to 2 million.

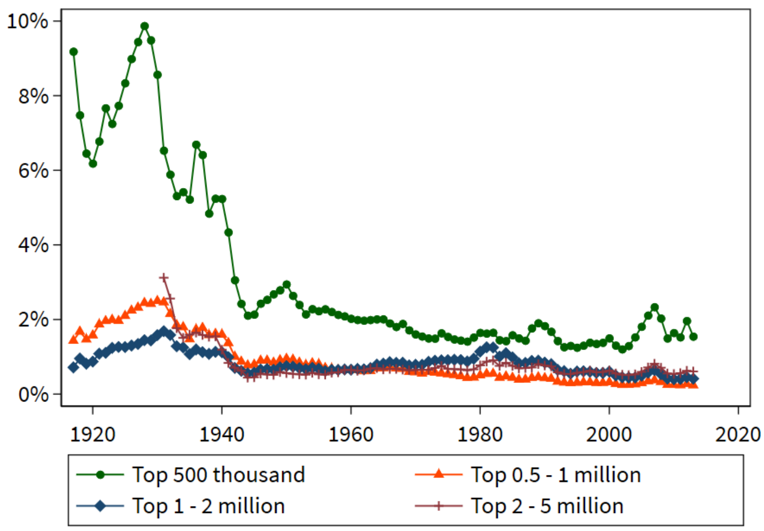
Another potential contributing factor to the rising importance of wages is the increased assortative mating that has occurred since the 1970s (Schwartz, 2010). Because income is measured at the tax unit level, the increased propensity to marry a spouse with a similarly high income should increase both the importance of wages and overall income shares over time.

Another popular explanation for this development are tax reforms, particularly the 1981 and 1986 Tax Reform Acts (Bargain et al., 2015; Feldstein, 1995; Auerbach and Slemrod, 1997; Hausman and Poterba, 1987). In addition to a broad range of measures, the 1986 TRA reduced top marginal tax rates from 50 to 28 percent but broadened tax bases. The 1980s constituted a tipping point in the development of top wage incomes. After the tax reforms, the shares of top wages in overall national incomes increased, but the previously increasing importance of wages relative to other income sources of top earners came to a halt. That is, while top earners became relatively richer and their wages grew over time, their earnings increases were driven by both wages – possibly driven by the developments described above – and other income sources. Other reforms of income tax rates did not have such pronounced effects. Notably, the

Omnibus Budget Reconciliation Act of 1993 increased top marginal tax rates from 31 to 39.6 percent, but this increase can be associated at best with small fluctuations in wage shares.

Capital income. Capital income is composed of rents, dividends and interest. The share of capital income followed a rather flat trajectory from the 1940s to the 1970s for the top 500,000 to 5 million. This holds for both the share in aggregate income and its importance relative to top earners' other income sources.

Figure 4.8: Capital shares of top income brackets (Measure D)



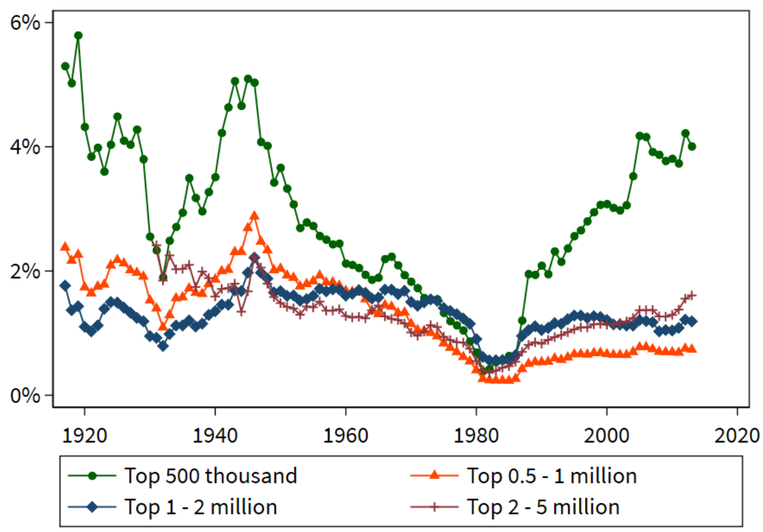
Notes: This figure plots capital income shares of fixed numbers of top tax units in aggregate income (Measure D, as described in section 4.2.1). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

For the uppermost 500,000, however, capital income has increasingly contributed to their income in recent years. Nevertheless, the recent increase pales in comparison to the income shares that top earners obtained from dividends, interest and rents prior to World War II.

The spike in the early 1980s was accompanied by a sharp drop in business income (see Figure 4.9). Part of this development may therefore be related to reclassification of incomes in response to changing tax incentives. Some of the development may also be attributable to changing tax avoidance and evasion opportunities over time.

Business income. Revenues from entrepreneurial activities have been subject to large changes. While for all top income brackets, the importance of entrepreneurial income declined from World War II throughout the 1970s, it has seen a sharp rise since the 1980s. Not only has the proportion of income generated by entrepreneurs multiplied, but overall entrepreneurial revenues have also contributed to overall US incomes at increasing rates for the top 500,000 and, to a much lower extent, the top 2–5 million. Strikingly, while entrepreneurial income has played less of a role for the top 500,000 than for taxpayers with incomes just below theirs, this relation reversed after 1980.

Figure 4.9: Entrepreneurial income shares of top income brackets (Measure D)



Notes: This figure plots entrepreneurial income shares of fixed numbers of top tax units in aggregate income (Measure D, as described in section 4.2.1). The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

Much of the large jump in Figure 4.9 may be attributable to reclassifications and incentives related to the 1986 Tax Reform Act (Feldstein, 1995; Auerbach and Slemrod, 1997; Hausman and Poterba, 1987). First, the abolition of the general utilities rule made C-corporations less attractive. Prior to the reform, such corporations had allowed for lower tax rates than the personal income tax. As a result of the reform, many C-corporations were converted into S-corporations. Thereby, previously excluded corporate income was included on personal tax returns, counting towards entrepreneurial income (Feldstein, 1995). Top earners' higher capacity for tax avoidance might also explain why the effect was larger for the top 500,000 than for the subsequent high-income earners.

4.4.3 Using Data from the Distributional National Accounts

Recently, Piketty et al. (2018) analyzed income distribution trends with broader measures that strive to capture the full scope of national accounts. Based on tax, survey and national accounts data, DINA measures are more comprehensive than standard top share measures. In particular, the pre-tax national income measure captures all income flows to capital and labor, accounting for pensions, unemployment and disability insurance. I.e., it corresponds to income before government intervention.

Our measures can likewise be applied to this income concept. As opposed to the standard top income share data, sufficient DINA data is only available from the 1960s onwards.[17] In addition to being more comprehensive, the DINA calculations differ in further aspects, which are attributable to differences in the data sets.[18]

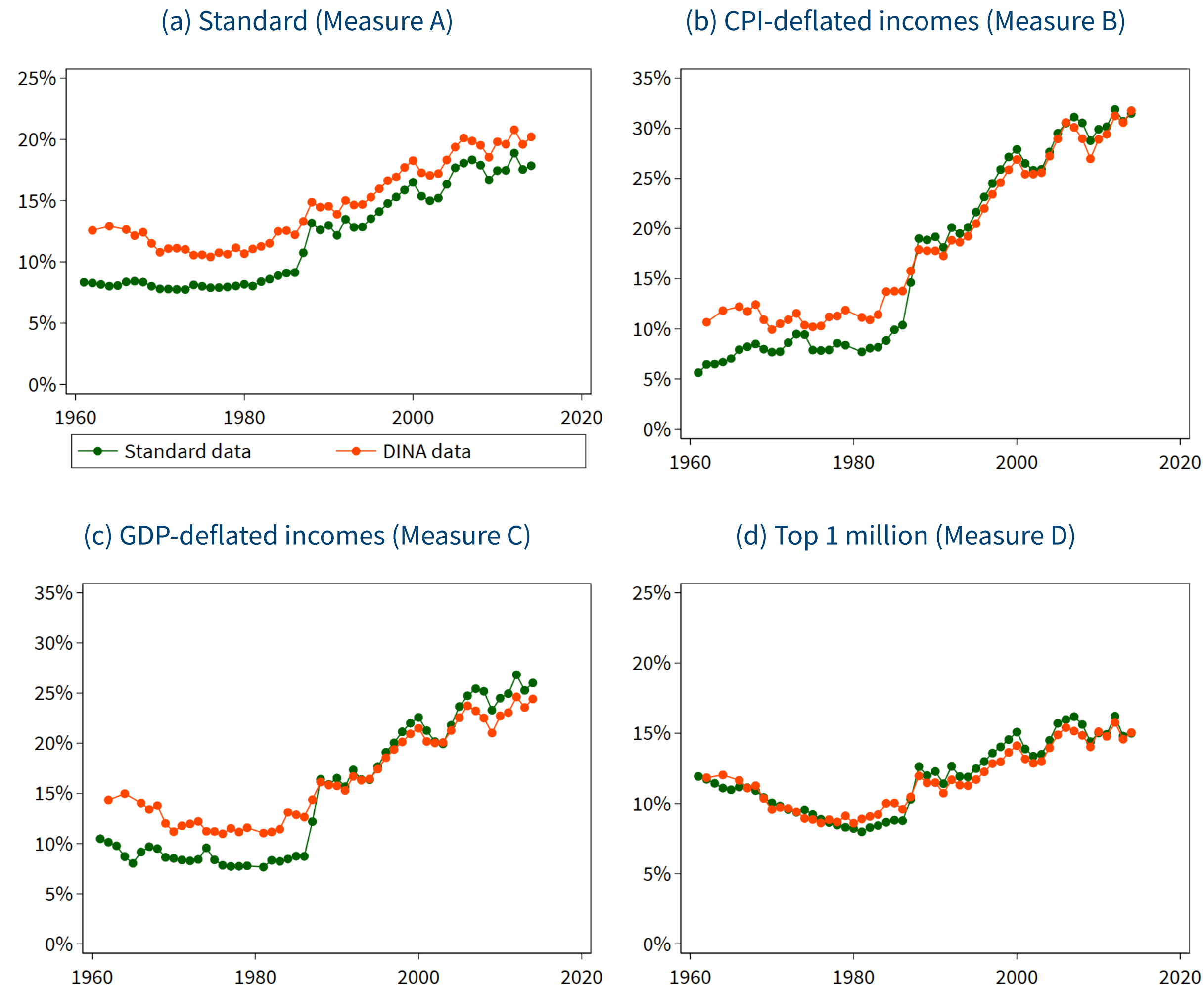
Despite the methodological differences, the DINA-based measures for the top percentile in Figure 4.10 follow a very similar trajectory compared to the different top share measures using the standard data. Interestingly, the biggest difference between the two data sources can be found for the standard top income share measure. Both population and income shares above the CPI-deflated top 1 percent thresholds have risen rapidly in the last three decades. Income and population shares above the GDP-deflated top 1 percent threshold follow a similar, but less pronounced trend.

The very high degree of similarity between DINA and standard measures grants further credibility to our measures and indicates that our measures can be applied to a wide array of income concepts.

[17] Prior years lack information on top thresholds and only contain limited information on top shares. Calculating our alternative top share measures with a sufficient degree of precision is thus not possible.

[18] Instead of tax units, DINA considers individuals age 20 or older, between whom incomes are equally split within households. We correspondingly deflate thresholds by per adult GDP. Instead of the CPI, the data set also uses the national income price index rather than the CPI for deflating thresholds. Moreover, from the 1960s onwards, top thresholds and shares are available at a much finer grid than the standard top share data, with separate measures for every percentile and an even finer grid within the top percentile. We exploit this different data structure and use piecewise linear interpolation to obtain measures above thresholds.

Figure 4.10: Standard vs. DINA data



Notes: This figure compares top percentile income shares for the four different top income inequality measures using two different data concepts: standard data and DINA data. The four different top inequality measures (as described in section 4.2.1) are: the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980. The figure is based on authors' calculations using data from the World Wealth and Income Database and the World Inequality Database (see section 4.2.2 for details). The income concept is, respectively, pre-tax fiscal income, excluding capital gains, and pre-tax national income.

4.5 Concluding Discussion

The measurement of income inequality trends has been discussed extensively in the academic inequality literature. The recent top income literature has proposed using top income shares to capture inequality dynamics, and it has received tremendous attention and spurred much new research. However, little attention has been paid to the fact that the standard top income share measure is only one of several variants of top income share measures. By focusing on incomes earned by a fixed share of the population, this measure does not fully reflect changes in the economy that could make the top group either shrink or grow relative to the rest of the population.

Our study offers three alternative top share measures that try to address this aspect, focusing on the influence of real income growth and population growth. The first two measures define an income threshold in a certain year above which income earners are said to belong to the top (different thresholds refer to different top fractiles). This threshold is then extended to later or earlier years, deflating either by CPI or GDP to reflect the relative influence of increased purchasing power or relative shares in overall productivity growth. The third measure fixes the top group in terms of number of earners, making it insensitive to population growth.

Our empirical application of these measures to the long-run trends in US top income shares shows that our measures offer a broadly consistent picture with the one previously provided by Piketty and Saez. We find a decline in top shares in the first half of the twentieth century and a rise in the period after 1980. However, our findings point to several additional patterns, in particular concerning top share measures that are anchored at the absolute level of income in one particular year and then deflated over time using either consumer prices or GDP. Most notable are the differences between threshold levels in the CPI- and GDP-deflated measures. While groups at the very top are roughly similar to the baseline measures, trajectories of groups in the upper middle class vary a great deal.

It should be noted that these alternative measures are complementary to the standard approach. In fact, we believe that using them in parallel could offer a valuable strategy for the study of top incomes, and we think that this is underlined by our analysis of US historical experiences.

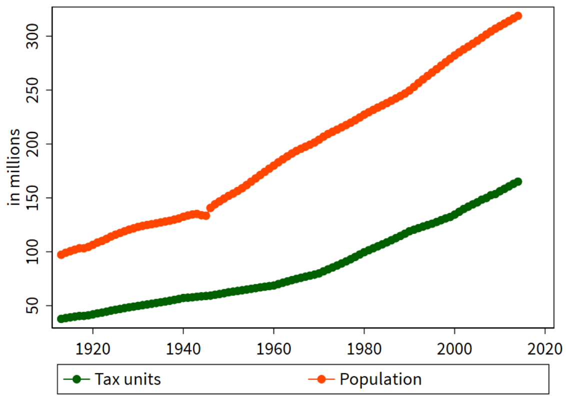
The analysis of the role of GDP-deflated real incomes relates to the recent studies on inequality outcomes using a broader income concept that includes all of national income instead of just tax return-based fiscal incomes (Piketty et al., 2018; Bozio et al., 2016; Auten and Splinter, 2019). We apply our alternative measures to these distributional national accounts incomes and find broadly consistent results. Again, our contribution is to offer alternative top income share measures, and these can be used on any underlying concept of income.

We hope that future research will continue working on refining the way income inequality is measured. Our alternative top income measures offer one way of approaching similar re-assessments of top shares for other countries, and even of conducting cross-country analyses of the development of top income shares over time. This analysis could shed additional light on the multitude of factors driving the differing developments across and within countries.

Appendix

Population growth. As depicted in Figure 4.A.1, the US population has more than tripled over the course of the past century. The number of tax units has grown by a factor of four across the same time period. Data are taken from the World Wealth and Income Database.

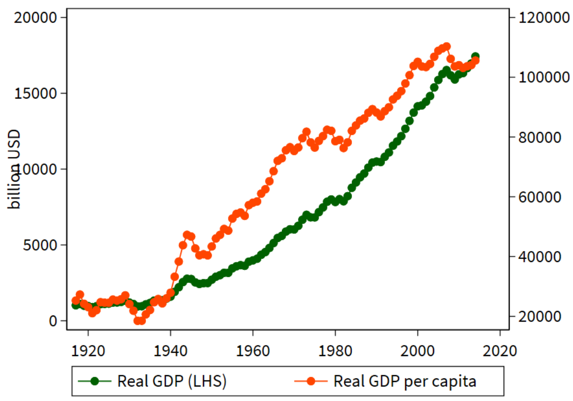
Figure 4.A.1: US population growth



Notes: This figure plots the US population size and the number of tax units over time. The figure is based on data from the World Wealth and Income Database.

Economic growth. Figure 4.A.2 displays the development of overall and per tax unit real GDP in 2014 terms. Data are taken from the World Wealth and Income Database and from Johnston and Williamson (2018).

Figure 4.A.2: Real GDP growth

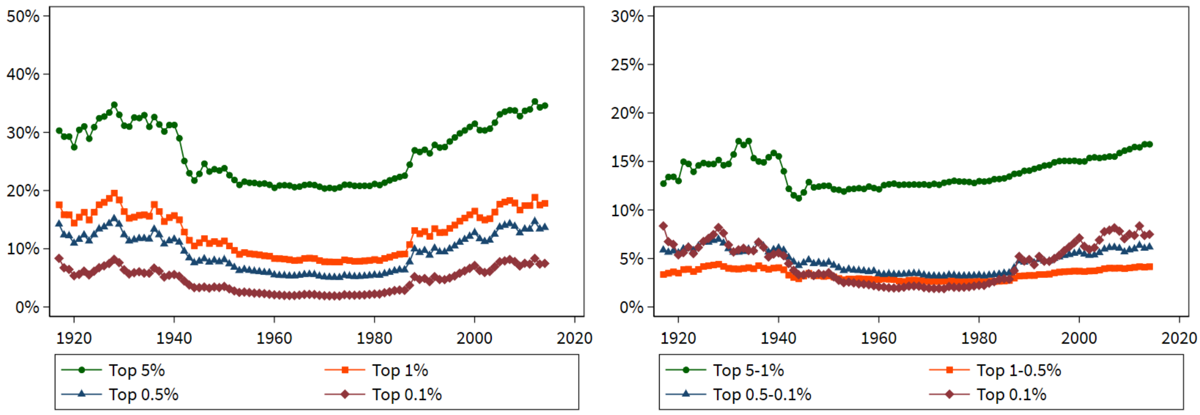


Notes: This figure plots real GDP and real per tax unit GDP in 2014 USD over time. The figure is based on data from the World Wealth and Income Database and from http://www.measuringworth.org/usgdp/ (Johnston and Williamson, 2018).

The key results of Piketty and Saez. As a baseline, Figure 4.A.3 displays the evolution of top income shares excluding capital gains over time. The upper figure depicts the unadjusted top income shares as in Piketty and Saez (2003) and Piketty and Saez (2006). As shown in the graph, top income shares stay roughly constant in the mid-twentieth century, but experience substantial increases since the 1980s, especially at the very top. The bottom figure shows the development of income shares in top income share brackets, such as the top 5-1 percent. This provides a benchmark against which the alternative measures can be compared.

Figure 4.A.3: Development of unadjusted top income shares (Measure A)

(a) Top income shares (b) Income shares of top income brackets

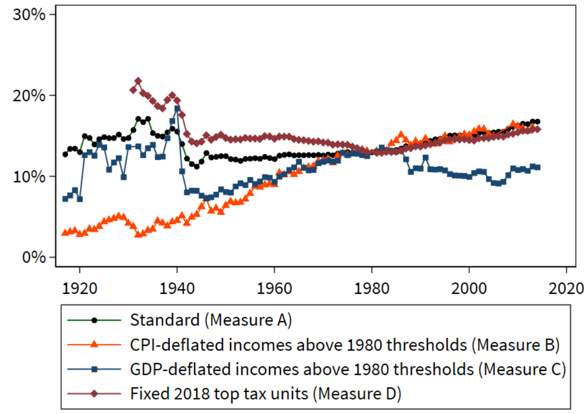


Notes: This figure plots standard top income shares (Measure A, as described in section 4.2.1). The figure is based on data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

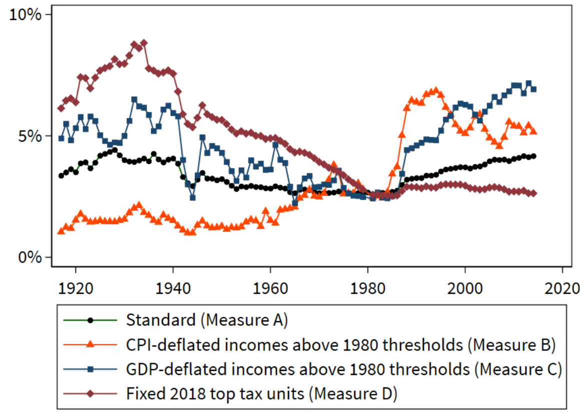
Different measures for varying top percentiles. Similar to Figure 4.1, Figure 4.A.4 compares top income shares for different parts of the 1980 top earnings distribution.

Figure 4.A.4: Different measures for varying top percentiles

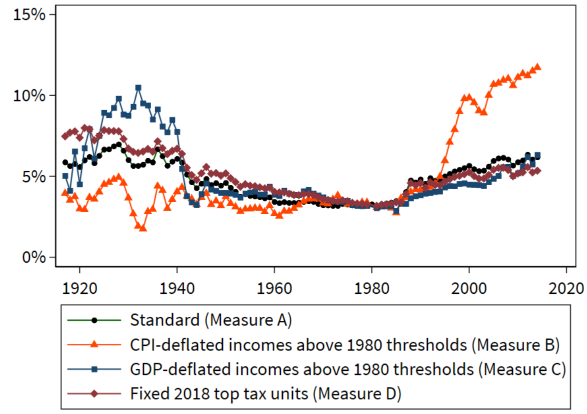
(a) Top 5-1 percent



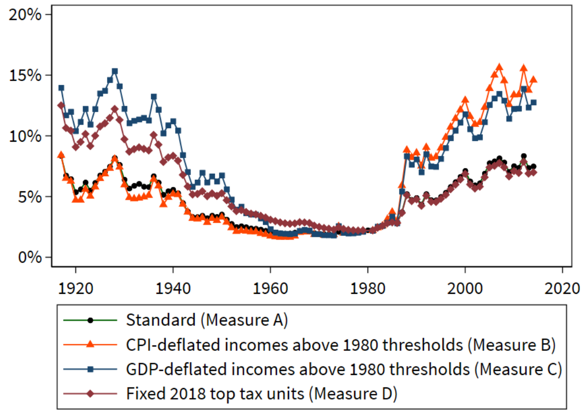
(b) Top 1-0.5 percent



(c) Top 0.5-0.1 percent



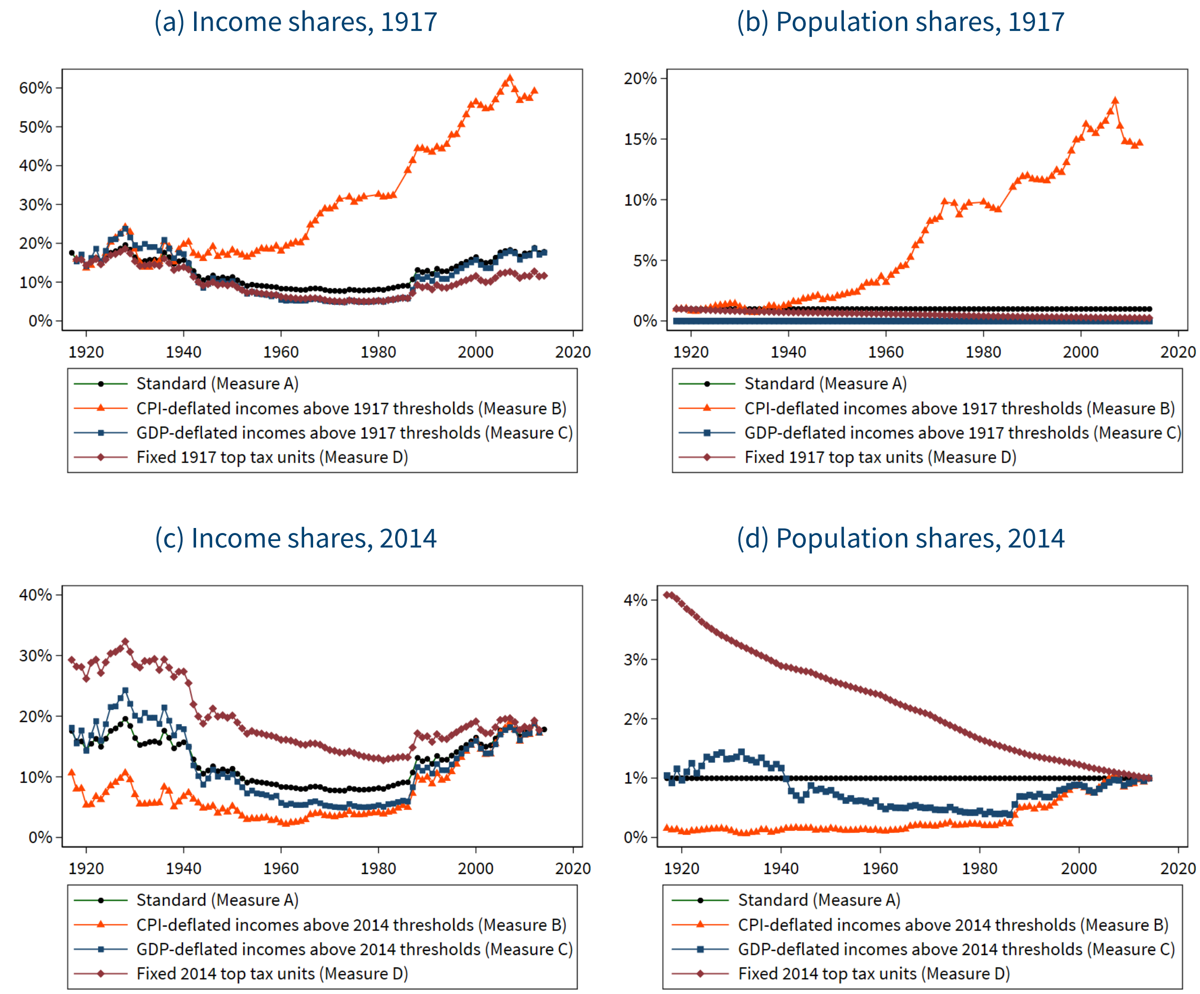
(d) Top 0.1 percent



Notes: This figure plots income shares for the four different top inequality measures (as described in section 4.2.1): the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of a fixed number of taxpayers (D). The measures are constructed such that they equal each other in 1980. The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

1917 and 2014 thresholds. While our main results focus on 1980 thresholds, similar findings emerge for CPI- or GDP-deflated 1917 and 2014 thresholds (4.A.5). Results for CPI-deflated thresholds (fig. 4.A.6) and GDP-deflated thresholds (fig. 4.A.7) fit well with the 1980 results. While the richest group of taxpayers has become significantly richer since the 1980s and has obtained more than an equal share in the benefits of growth, the effect is not that clear-cut for the middle upper class. Again, the groups that form the lower top (e.g. the 2014 top 5-1 percent) increase in size and gain larger income shares than those above them, but below the very top.

Figure 4.A.5: Different measures for the top 1 percent: 1917 and 2014

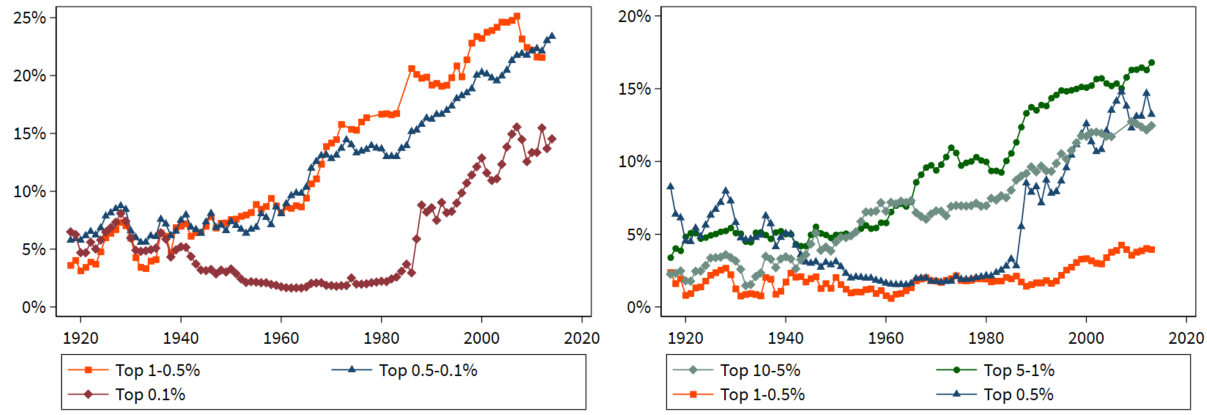


Notes: This figure plots population and income shares for the four different top inequality measures (as described in section 4.2.1): the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1917 and 2014 income thresholds, as well as the income share of the top 1 million taxpayers (D). Measures are constructed such that they equal each other in 1917 or 2014. The figure is based on authors' calculations using data from the World Wealth and Income Database as well as IRS SOI data (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

Figure 4.A.6: Income shares above CPI-deflated 1917 and 2014 thresholds (Measure B)

(a) Income shares above CPI-deflated 1917 thresholds

(b) Income shares above CPI-deflated 2014 thresholds

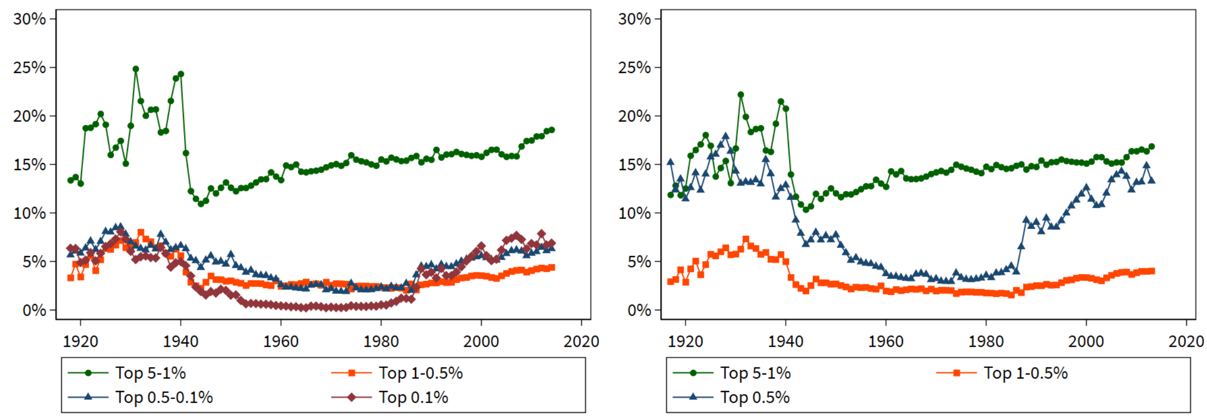


Notes: This figure plots population and income shares of those earning above the CPI-deflated 1917 and 2014 income thresholds (Measure B, as described in section 4.2.1). The measure is constructed such that it equals the standard top 1 percent share in 1917 or 2014. The figure is based on authors' calculations using data from the World Wealth and Income Database as well as IRS SOI data (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

Figure 4.A.7: Income shares above GDP-deflated 1917 and 2014 thresholds (Measure C)

(a) Income shares above GDP-deflated 1917 thresholds

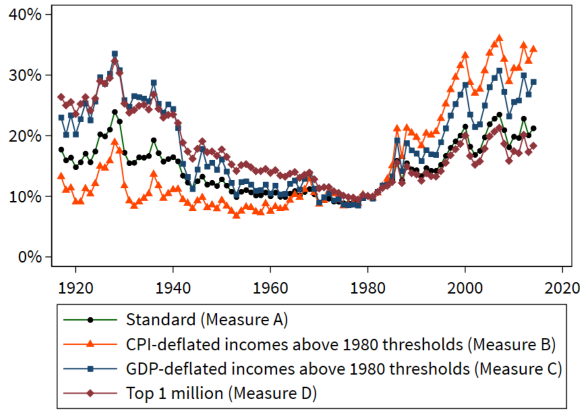
(b) Income shares above GDP-deflated 2014 thresholds



Notes: This figure plots population and income shares of those earning above the per capita GDP-deflated 1917 and 2014 income thresholds (Measure C, as described in section 4.2.1). The measure is constructed such that it equals the standard top 1 percent share in 1917 or 2014. The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, excluding capital gains.

Inclusion of capital gains. As a robustness check, Figure 4.A.8 displays 1980 top 1 percent income shares including capital gains. As can be expected, income shares including capital gains fluctuate more than those net of capital gains in Figure 4.1. As for total incomes excluding capital gains, top income shares start rising rapidly in the 1980s.

Figure 4.A.8: Top 1980 1 percent shares including capital gains

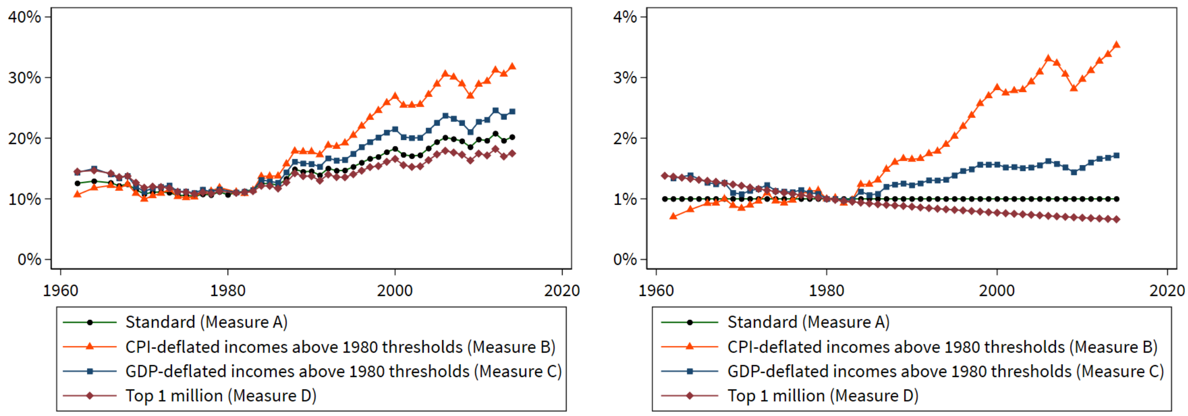


Notes: This figure plots income shares for the four different top inequality measures (as described in section 4.2.1): the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980. The figure is based on authors' calculations using data from the World Wealth and Income Database (see section 4.2.2 for details). The income concept is pre-tax fiscal income, including capital gains.

DINA 1980 top shares. As a further robustness check, Figure 4.A.9 shows the DINA equivalent to the top shares presented in Figures 4.1 and 4.2.

Figure 4.A.9: Top 1980 1 percent shares based on DINA

(a) Income shares (b) Population shares



Notes: This figure plots income shares for the four different top inequality measures (as described in section 4.2.1) using DINA data: the standard measure (A), the income shares of those earning above the CPI-deflated (B) or GDP-deflated (C) 1980 income threshold, as well as the income share of the top 1 million taxpayers (D). The measures are constructed such that they equal each other in 1980. The figure is based on authors' calculations using data from the World Inequality Database (see sections 4.2.2 and 4.4.3 for details). The income concept is pre-tax national income.

Bibliography

Alstadsaeter, A., M. Jacob, W. Kopczuk, and K. Telle (2016). Accounting for Business Income in Measuring Top Income Shares: Integrated Accrual Approach Using Individual and Firm Data from Norway. *NBER Working Paper 22888*.

Amiel, Y. and F. Cowell (1992). Measurement of Income Inequality: Experimental Test by Questionnaire. *Journal of Public Economics 47*(1), 3–26.

Amiel, Y. and F. Cowell (1999). *Thinking About Inequality: Personal Judgment and Income Distributions*. Cambridge University Press.

Armour, P., R. V. Burkhauser, and J. Larrimore (2013). Deconstructing Income and Income Inequality Measures: A Crosswalk from Market Income to Comprehensive Income. *The American Economic Review 103*(3), 173–177.

Asch, B., S. J. Haider, and J. Zissimopoulos (2005). Financial Incentives and Retirement: Evidence from Federal Civil Service Workers. *Journal of Public Economics 89*(2-3), 427–440.

Atalay, K. and G. F. Barrett (2015). The Impact of Age Pension Eligibility Age on Retirement and Program Dependence: Evidence from an Australian Experiment. *Review of Economics and Statistics 97*(1), 71–87.

Atkinson, A. B. (2008). More on the Measurement of Inequality. *Journal of Economic Inequality 6*(3), 277–283.

Atkinson, A. B. and A. Brandolini (2010). On Analyzing the World Distribution of Income. *The World Bank Economic Review 24*(1), 1–37.

Atkinson, A. B. and T. Piketty (Eds.) (2007). *Top Incomes Over the Twentieth Century: A Contrast Between Continental European and English-Speaking Countries*. Oxford University Press.

Atkinson, A. B. and T. Piketty (Eds.) (2010). *Top Incomes: A Global Perspective*. Oxford University Press.

Atkinson, A. B., T. Piketty, and E. Saez (2011). Top Incomes in the Long Run of History. *Journal of Economic Literature 49*(1), 3–71.

Auerbach, A. J. and J. Slemrod (1997). The Economic Effects of the Tax Reform Act of 1986. *Journal of Economic Literature 35*(2), 589–632.

Auten, G. and D. Splinter (2019). Income Inequality in the United States: Using Tax Data to Measure Long-term Trends. *Mimeo*.

Baker, M. and D. Benjamin (1999). Early Retirement Provisions and the Labor Force Behavior of Older Men: Evidence from Canada. *Journal of Labor Economics 17*(4), 724–756.

Baldenius, T., S. Kohl, and M. Schularick (2019). Die neue Wohnungsfrage: Gewinner und Verlierer des deutschen Immobilienbooms. *Mimeo*.

Bargain, O., M. Dolls, H. Immervoll, D. Neumann, A. Peichl, and N. Pestel (2015). Tax Policy and Income Inequality in the U.S., 1979–2007. *Economic Inquiry 2*(53), 1061–1085.

Baudisch, C. F. and C. Dresselhaus (2018). Impact of the German Real Estate Transfer Tax on the Commercial Real Estate Market. *Finanzwissenschaftliche Arbeitspapiere Nr. 100 – 2018, Justus-Liebig-Universität Gießen*.

Bayerische Eigenheimzulagen-Richtlinien (2018). Richtlinien für die Gewährung eines Zuschusses zum Bau oder Erwerb von Wohnraum zu eigenen Wohnzwecken – EHZR . *Bekanntmachung des Bayerischen Staatsministeriums für Wohnen, Bau und Verkehr*.

Behaghel, L. and D. M. Blau (2012). Framing Social Security Reform: Behavioral Responses to Changes in the Full Retirement Age. *American Economic Journal: Economic Policy 4*(4), 41–67.

Berger, T., P. Englund, P. H. Hendershott, and B. Turner (2000). The Capitalization of Interest Subsidies: Evidence from Sweden. *Journal of Money, Credit, and Banking 32*(2), 199–217.

Besley, T., N. Meads, and P. Surico (2014). The Incidence of Transaction Taxes: Evidence from a Stamp Duty Holiday. *Journal of Public Economics 119*, 61–70.

Best, M. C. and H. J. Kleven (2018). Housing Market Responses to Transaction Taxes: Evidence from Notches and Stimulus in the UK. *The Review of Economic Studies 85*(1), 157–193.

Bönke, T., D. Kemptner, and H. Lüthen (2018). Effectiveness of Early Retirement Disincentives: Individual Welfare, Distributional and Fiscal Implications. *Labour Economics 51*, 25–37.

Börsch-Supan, A., M. Brandt, C. Hunkler, T. Kneip, J. Korbmacher, F. Malter, B. Schaan, S. Stuck, and S. Zuber (2013). Data Resource Profile: The Survey of Health, Ageing and Retirement in Europe (SHARE). *International Journal of Epidemiology 42*(4), 992–1001.

Börsch-Supan, A., M. Coppola, and J. Rausch (2015). Die “Rente mit 63”: Wer sind die Begünstigten? *Perspektiven der Wirtschaftspolitik 16*(3), 264–288.

Börsch-Supan, A. and R. Schnabel (1998). Social Security and Declining Labor-Force Participation in Germany. *The American Economic Review 88*(2), Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association, 173–178.

Bosmans, K., K. Decancq, and A. Decoster (2014). The Relativity of Decreasing Inequality Between Countries. *Economica 81*(322), 276–292.

Bourassa, S., D. Haurin, P. Hendershott, and M. Hoesli (2013). Mortgage Interest Deductions and Homeownership: An International Survey. *Journal of Real Estate Literature 21*(2), 181–203.

Bozio, A., T. Breda, and M. Guillot (2016). Taxes and Technological Determinants of Wage Inequalities: France 1976-2010. *PSE Working Paper No. 2016-05*.

Brown, K. M. (2013). The Link Between Pensions and Retirement Timing: Lessons from California Teachers. *Journal of Public Economics 98*, 1–14.

Bundesamt für Bauwesen und Raumordnung (2002). Bericht zur Inanspruchnahme der Eigenheimzulage in den Jahren 1996-2000. *Bonn: Bundesamt für Bauwesen und Raumordnung*.

Burkhauser, R. V., S. Feng, S. P. Jenkins, and J. Larrimore (2012). Recent Trends in Top Income Shares in the United States: Reconciling Estimates from March CPS and IRS Tax Return Data. *Review of Economics and Statistics 94*(2), 371–388.

Büttner, T. and M. Krause (2018). Föderalismus im Wunderland: Zur Steuerautonomie bei der Grunderwerbsteuer. *Perspektiven der Wirtschaftspolitik 19*(1), 32–41.

Carozzi, F., C. Hilber, and X. Yu (2019). The Economic Impacts of Help to Buy. *Mimeo*.

Chetty, R., D. Grusky, M. Hell, N. Hendren, R. Manduca, and J. Narang (2017). The Fading American Dream: Trends in Absolute Income Mobility Since 1940. *Science 356*(6336), 398–406.

Chetty, R., N. Hendren, P. Kline, and E. Saez (2014). Where is the Land of Opportunity? The Geography of Intergenerational Mobility in the United States. *Quarterly Journal of Economics 129*(4), 1553–1623.

Chetty, R., N. Hendren, P. Kline, E. Saez, and N. Turner (2014). Is the United States Still a Land of Opportunity? Recent Trends in Intergenerational Mobility. *American Economic Review 104*(5), 141–147.

Cribb, J., C. Emmerson, and G. Tetlow (2016). Signals Matter? Large Retirement Responses to Limited Financial Incentives. *Labour Economics 42*, 203–212.

Dachis, B., G. Duranton, and M. A. Turner (2012). The Effect of Land Transfer Taxes on Real Estate Markets: Evidence from a Natural Experiment in Toronto. *Journal of Economic Geography 12*(2), 327–354.

Davidoff, I. and A. Leigh (2013). How Do Stamp Duties Affect the Housing Market? *Economic Record 89*(286), 396–410.

Davis, M. (2018). The Distributional Impact of Mortgage Interest Subsidies: Evidence from Variation in State Tax Policies. *Mimeo*.

Dehejia, R. H. and S. Wahba (2002). Propensity Score-Matching Methods for Nonexperimental Causal Studies. *Review of Economics and Statistics 84*(1), 151–161.

Deutsche Bundesbank (2018). Methodenbericht zu den Wohnimmobilienpreisindizes. *Frankfurt am Main, 29. August 2018*.

Deutsche Rentenversicherung Bund (2018). Rentenversicherung in Zeitreihen. *DRV Schriften Band 22*.

Deutscher Bundestag (2005). Entwurf eines Gesetzes zur Abschaffung der Eigenheimzulage. *Drucksache 16/108*.

Deutscher Bundestag (2014). Entwurf eines Gesetzes über Leistungsverbesserungen in der gesetzlichen Rentenversicherung. *Drucksache 18/909*.

Deutscher Bundestag (2019). Antwort der Bundesregierung auf die Kleine Anfrage der Abgeordneten Christian Kühn (Tübingen), Daniela Wagner, Lisa Paus, weiterer Abgeordneter und der Fraktion BÜNDNIS 90/DIE GRÜNEN - Drucksache 19/2105 - Wirkungen des so genannten Baukindergeldes. *Drucksache 19/2684*.

Dew-Becker, I. and R. J. Gordon (2005). Where Did the Productivity Growth Go? Inflation Dynamics and the Distribution of Income. *Brookings Papers on Economic Activity 2005*(2), 67–128.

Dolls, M., C. Fuest, C. Krolage, and F. Neumeier (2019). Who Bears the Burden of Real Estate Transfer Taxes? Evidence from the German Housing Market. *ifo Working Paper No.308*.

Dolls, M. and C. Krolage (2019). The Effects of Early Retirement Incentives on Retirement Decisions. *ifo Working Paper No. 291*.

Dynan, K., T. Gayer, and N. Plotkin (2013). An Evaluation of Federal and State Homebuyer Tax Incentives. *Washington, DC: The Brookings Institution*.

Engels, B., J. Geyer, and P. Haan (2017). Pension Incentives and Early Retirement. *Labour Economics 47*, 216–231.

F+B Forschung und Beratung für Wohnen, Immobilien und Umwelt GmbH. *Property Price Data Prepared for Ifo Institute*. Data set.

Feenberg, D. R. and J. M. Poterba (1993). Income Inequality and the Incomes of Very High-Income Taxpayers: Evidence from Tax Returns. *Tax Policy and the Economy 7*, 145–177.

Feenberg, D. R. and J. M. Poterba (2000). The Income and Tax Share of Very High-Income Households, 1960-1995. *American Economic Review 90*(2), 264–270.

Feldstein, M. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act. *Journal of Political Economy 103*(3), 551–572.

Floetotto, M., M. Kirker, and J. Stroebel (2016). Government Intervention in the Housing Market: Who Wins, Who Loses? *Journal of Monetary Economics 80*, 106–123.

Forschungsdatenzentren der Statistischen Ämter des Bundes und der Länder. *Einkommens- und Verbrauchsstichprobe 2018 Grundfile 1 und 2*. Data set.

Forschungsdatenzentrum der Rentenversicherung. *Versichertenrentenzugang 2013-2017*. Data set.

Forschungsdatenzentrum der Rentenversicherung and Max-Planck-Institut für Sozialrecht und Sozialpolitik (2019). *SHARE-RV*. Release version: 7.0.0. SHARE-ERIC. Data set.

Färber, G. (2003). Wirkungen der Eigenheimzulage. *Wohnungswirtschaft und Mietrecht 4*, 196–200.

Fritzsche, C. and L. Vandrei (2019). The German Real Estate Transfer Tax: Evidence for Single-Family Home Transactions. *Regional Science and Urban Economics 74*, 131–143.

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

Geyer, J. and C. Welteke (2020). Closing Routes to Retirement for Women: How Do They Respond? *Journal of Human Resources*, forthcoming.

Gornig, M., C. Michelsen, and M. Bruns (2019). Construction Industry Momentum Continues – State Stimulus Impacts Prices. *DIW Weekly Report 9*(1/2), 3–14.

Gruber, J., A. Jensen, and H. Kleven (2020). Do People Respond to the Mortage Interest Deduction? Quasi-Experimental Evidence from Denmark. *American Economic Journal: Economic Policy*, forthcoming.

Gustman, A. L. and T. L. Steinmeier (1986). A Structural Retirement Model. *Econometrica 54*(3), 555–584.

Hanel, B. and R. T. Riphahn (2012). The Timing of Retirement – New Evidence from Swiss Female Workers. *Labour Economics 19*(5), 718–728.

Hausman, J. A. and J. M. Poterba (1987). Household Behavior and the Tax Reform Act of 1986. *Journal of Economic Perspectives 1*(1), 101–119.

Hilber, C. and T. Turner (2014). The Mortgage Interest Deduction and its Impact on Homeownership Decisions. *Review of Economics and Statistics 96*(4), 618–637.

Iacus, S. M., G. King, and G. Porro (2012). Causal Inference Without Balance Checking: Coarsened Exact Matching. *Political Analysis 20*(1), 1–24.

Internal Revenue Service. SOI Tax Statistics: Historical Table 3: Number of Individual Income Tax Returns, Income, Exemptions and Deductions, Tax, and Average Tax, by Size of Adjusted Gross Income.

Johnston, L. and S. H. Williamson (2018). What Was the US GDP Then? *MeasuringWorth*.

Kopczuk, W. and D. Munroe (2015). Mansion Tax: The Effect of Transfer Taxes on the Residential Real Estate Market. *American Economic Journal: Economic Policy 7*(2), 214–257.

Kopczuk, W., E. Saez, and J. Song (2010). Earnings Inequality and Mobility in the United States: Evidence from Social Security Data Since 1937. *The Quarterly Journal of Economics 125*(1), 91–128.

Krolage, C., A. Peichl, and D. Waldenström (2018). Richer or More Numerous or Both? The Role of Population and Economic Growth for Top Income Shares. *CESifo Working Paper No. 7385*.

Kuznets, S. (1955). Economic Growth and Income Inequality. *The American Economic Review 45*(1), 1–28.

Lalive, R. and P. Parrotta (2017). How Does Pension Eligibility Affect Labor Supply in Couples? *Labour Economics 46*, 177–188.

Leigh, A. (2009). Top Incomes. *The Oxford Handbook of Economic Inequality*, 150–176.

Manoli, D. and A. Weber (2016). Nonparametric Evidence on the Effects of Financial Incentives on Retirement Decisions. *American Economic Journal: Economic Policy 8*(4), 160–182.

Manoli, D. and A. Weber (2018). The Effects of the Early Retirement Age on Retirement Decisions. *Mimeo*.

Mastrobuoni, G. (2009). Labor Supply Effects of the Recent Social Security Benefit Cuts: Empirical Estimates Using Cohort Discontinuities. *Journal of Public Economics 93*(11-12), 1224–1233.

Mense, A., C. Michelsen, and K. Kholodilin (2019). The Effects of Second-Generation Rent Control on Land Values. *AEA Papers and Proceedings 109*, 385–388.

OECD (2017). *Pensions at a Glance 2017: OECD and G20 Indicators*. OECD Publishing, Paris.

Peichl, A., T. Schaefer, and C. Scheicher (2010). Measuring Richness and Poverty: A Micro Data Application to Europe and Germany. *Review of Income and Wealth 56*(3), 597–619.

Petkova, K. and A. J. Weichenrieder (2017). Price and Quantity Effects of the German Real Estate Transfer Tax. *CESifo Working Paper No. 6538*.

Piketty, T. and E. Saez (2003). Income Inequality in the United States, 1913–1998. *The Quarterly Journal of Economics 118*(1), 1–41.

Piketty, T. and E. Saez (2006). The Evolution of Top Incomes: A Historical and International Perspective. *The American Economic Review 96*(2), 200–205.

Piketty, T., E. Saez, and G. Zucman (2018). Distributional National Accounts: Methods and Estimates for the United States. *Quarterly Journal of Economics 133*(2), 553–609.

Pimpertz, J. (2017). Kosten der schwarz-roten Rentenpolitik - eine Heuristik. Was kosten die zusätzliche Mütterrente und die abschlagfreie Rente mit 63? Reicht die Rente künftig noch über das Grundsicherungsniveau? *IW Policy Paper 3/2017*.

Rabaté, S. (2019). Can I Stay or Should I Go? Mandatory Retirement and the Labor-Force Participation of Older Workers. *Journal of Public Economics 180*(104078).

Rosen, S. (1981). The Economics of Superstars. *The American Economic Review 71*(5), 845–858.

Rust, J. and C. Phelan (1997). How Social Security and Medicaire Affect Retirement Behavior in a World of Incomplete Markets. *Econometrica 65*(4), 781–831.

Sachverständigenrat zur Begutachtung der gesamtwirtschaftlichen Entwicklung (2003). *Staatsfinanzen konsolidieren – Steuersystem reformieren*. Jahresgutachten 2003/2004.

Schmidheiny, K. and S. Siegloch (2019). On Event Study Designs and Distributed-Lag Models: Equivalence, Generalization and Practical Implications. *CEPR Discussion Paper No. DP13477*.

Schnabel, R. (2015). Das Rentenpaket 2014 - Eine ökonomische Beurteilung. *Wirtschaftsdienst 95*(1), 22–27.

Schwartz, C. R. (2010). Earnings Inequality and the Changing Association between Spouses' Earnings. *American Journal of Sociology 115*(5), 1524–1557.

Scott, D. W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization, Second Edition*. John Wiley & Sons.

Seibold, A. (2019). Reference Dependence in Retirement Behavior: Evidence from German Pension Discontinuities. *CESifo Working Paper No. 7799*.

Slemrod, J., C. Weber, and H. Shan (2017). The Behavioral Response to Housing Transfer Taxes: Evidence from a Notched Change in D.C. Policy. *Journal of Urban Economics 100*, 137–153.

Sommer, K. and P. Sullivan (2018). Implications of US Tax Policy for House Prices, Rents, and Homeownership. *American Economic Review 108*(2), 241–274.

Statistische Ämter des Bundes und der Länder. *Bevölkerungsstand (EVAS 12411), Arbeitsmarktstatistik der Bundesagentur für Arbeit (EVAS 13211), Bruttoinlandsprodukt je Einwohner (EVAS AI017-1)*. Data set.

Statistische Ämter des Bundes und der Länder. *Regionaldatenbank Deutschland, Fortschreibung des Wohngebäude- und Wohnungsbestandes 2017, EVAS 31231*. Data set.

Statistische Ämter des Bundes und der Länder. *Regionaldatenbank Deutschland, Gebäude- und Wohnungszählung 2011 (Zensus), EVAS 31211*. Data set.

Statistische Ämter des Bundes und der Länder. *Regionaldatenbank Deutschland, Statistik der Baugenehmigungen 2016-2018, EVAS 31111*. Data set.

Staubli, S. and J. Zweimüller (2013). Does Raising the Early Retirement Age Increase Employment of Older Workers? *Journal of Public Economics 108*, 17–32.

Vestad, O. L. (2013). Labour Supply Effects of Early Retirement Provision. *Labour Economics 25*, 98–109.

Waldenström, D. and J. Roine (2015). *Handbook in Income Distribution, vol 2A*, Chapter Long-Run Trends in the Distribution of Income and Wealth. Amsterdam: North-Holland.

Werding, M. (2014). Demographischer Wandel und öffentliche Finanzen: Langfrist-Projektionen 2014-2060 unter besonderer Berücksichtigung des Rentenreform-Pakets der Bundesregierung. *Arbeitspapier, Sachverständigenrat zur Begutachtung der Gesamtwirtschaftlichen Entwicklung*.

Wolff, E. N. (2002). *Top Heavy: The Increasing Inequality of Wealth in America and What Can Be Done About It*. New Press.

World Inequality Database, formerly: World Wealth and Income Database. *https://wid.world*. Data set.