Essays in empirical taxation and empirical public economics

Inauguraldissertation zur Erlangung des akademischen Grades eines Doktors der Wirtschaftswissenschaften der Universität Mannheim

Florian Buhlmann

March 15, 2020

University of Mannheim

Dean
Prof. Dr. Christian Becker

Referees
Prof. Dr. Johannes Voget
Prof. Dr. Andreas Peichl

Day of oral examination
May 11, 2020

Acknowledgments

First of all, I would like to thank my supervisors Johannes Voget and Andreas Peichl for their extraordinary support and guidance on the journey which led to this thesis and for giving me the freedom to pursue my own research ideas.

During my time at university Mannheim and ZEW Mannheim I learned a lot not only on how to conduct research and policy consulting but also personally. This thesis has benefited strongly from my coauthors Sebastian Blesse, Philipp Dörrenberg, Benjamin Elsner and Benjamin Loos by working with them I learned a lot on the things which are important for research but are not taught in the courses of any PhD program.

I am indebted to Michelle Hanlon who made it possible for me to visit the Massachusetts Institute of Technology in Cambridge MA. This was a great experience and offered me the opportunity to talk to many great economists.

In addition, I want to thank my coauthors from various policy consulting projects which are not part of this thesis but were part of my working days, Carla Krolage, Eric Sommer, Holger Stichnoth, Holger Bonin, Mathias Dolls, Maximilian Blömer, Maximilian Löffler and Sebastian Siegloch. I would like to highlight the support of Andreas Peichl, Holger Stichnoth and Sebastian Siegloch, who have been or are head of my department at ZEW. I am greatful to them for giving me the time to work on this thesis. Special thanks goes to my colleagues at ZEW, including the ones who moved to Munich, for many inspiring conversations.

Last but not least I would like to thank my beloved family for their ongoing support. I am especially greatful to my wife Vera for making my life much more enjoyable and for enduring the frustration I felt from time to time in writing this thesis and for building me up afterwards. I am indebted to my parents for lifelong support and for enabling me to study Economics in the first place.

Florian Buhlmann
Mannheim, March 2020

Contents

1 Preface

Modern welfare states rely heavily on taxes to finance their expenditures. The education system, social security and infrastructure are just a few amenities financed by welfare states through taxes. For example in Germany about 80 percent of state expenditures are financed through taxes and social security contributions (German Federal Bureau of Statistics 2019). It is in general undebated that governments need to raise taxes. However, it is heavily debated who should pay taxes and how much. In other words it is debated how the tax system should be designed. The question on the optimal design of the tax system cannot be answered by theory alone. Empirical work is needed to estimate model parameters (e.g. tax elasticities), to evaluate whether tax policies achieve the intended goals or to shed light on individuals' preferences. Therefore it is not only a normative but also a positive question.

Taxes are not only used to raise revenues but also to redistribute income and to guide individual behavior in certain directions. This thesis is empirical work and focuses especially on the latter dimension. The thesis is structured as follows. Chapter 2 quantifies unintended consequences of a negative income tax aiming to increase labor force participation. Chapter 3 studies individual preferences for tax simplification. Chapter 4 analyses the effects of a capital gains tax which was designed to deter individual investors from speculating on stock markets. While the data used in chapter 2 and 4 is data that shows the revealed preferences of individuals the question we address in chapter 3 cannot be answered with such type of data. Therefore we use survey and experimental data for a representative sample of the German population to study preferences for tax simplification.

Badly designed tax systems can have severe and unintended consequences. For example transfer programs often have high transfer withdrawal rates, this saves the money of tax payers and therefore make anti-poverty programs politically more feasible. At the same time this causes high marginal tax rates for transfer recipients. As a consequence, transfer recipients have little incentive to leave the transfer system and remain locked-in to the system. The earned income tax credit (EITC) in the US tries to reduce poverty of low wage earners and at the same time incentivize people to take up employment. The idea of the EITC is to motivate unemployed workers with low hourly market wages to take-up employment by paying a negative income tax depending on earned income. Still it has been shown that some of the design features cause tax payers to manipulate their income in order to maximize their tax refund (Saez et al. 2012). These responses are unintended in the sense that some tax payers reduce labor supply or evade taxes to get the maximal refund.

In chapter 2 which is coauthored with Benjamin Elsner and Andreas Peichl we use state level variation and a border discontinuity design to causally identify and quantify the change in income manipulation in response to a change in the EITC rates. We exploit that US states can set top-up rates, which means that, at a given point in time, workers with the same income receive different tax refunds in different states. Using an event study difference in difference approach as well as a border pair design, we document that a raise in the state-EITC leads to more bunching of self-employed tax filers at the first kink point of the tax schedule. While we document a strong relationship up until the Great Recession in 2007, we find no effect thereafter. These findings point to important behavioral responses to what is the largest welfare program in the US.

From a policy makers perspective these behavioral responses are in part desirable e.g. increases in labor supply at the extensive margin. On the other hand income manipulation can also be driven by adverse reactions like income misreporting, by reducing labor supply at the intensive or for married couples also on the extensive margin. The incentive to locate close to the kink point stems from the fact that at the first kink point the marginal tax rate (MTR) jumps upwards. Our results show further that if the jump in the MTR at the kink point becomes larger, it becomes more likely that taxpayers manipulate their income to locate close to the kink point. For policy makers it seems therefore advisable to decrease the jump in the MTR. This could be achieved by gradually reducing the rate at which the EITC is phased-in until the EITC does not increase anymore. By this means the first kink point could be abolished while maintaining positive labor supply incentives at the extensive and intensive margin. In addition, it would be advisable to individualize married couples with respect to the EITC, this would eliminate negative labor supply incentives for the second wage earner. If it is for some reason impossible to change the EITC schedule tax authorities could target tax audits to tax payers reporting income close to the kink point. This would make it more risky and more difficult for tax payers to misreport their income in order to maximize the refund.

In chapter 3 of this thesis, which is coauthored with Sebastian Blesse and Philipp Dörrenberg, we aim to improve the understanding of people's attitudes towards tax complexity. This is an important but largely unexplored question. It adds to the understanding on how the government should optimally design the tax system. We focus on complexity that is caused by deductions and allowances. Deductions and allowances are a main driver of complexity in a tax system. To see this imagine a tax system without them. In such a tax system gross income equals the taxable income and therefore there is no need to study the tax code, to document any expenditures or to discuss with tax authorities whether a certain expenditure is deductible. The only source for complexity remaining is then the statutory rate, which could be easily calculated by a computer program even if the calculation per se is very complex.

The general wisdom seems to suggest that most tax systems are overly complex and that tax simplification is generally desirable. Consistent with this general wisdom, we document in the first part of this chapter that more than 90% of respondents believe that the tax system should be simplified. However, there also are efficiency and equity arguments in support of a certain degree of tax complexity and it is puzzling why tax systems remain highly complex despite the conventional view in favor of more simplification. The second part of this chapter then investigates if the high support for tax simplification is driven by a lack of awareness about the trade-offs behind simple and complex tax systems. Our data show that the support for simplification decreases as we randomly provide economic arguments against simplification and as we ask respondents if the tax system should account for specific differences in living situations (such as costly care of elderly family members). Overall, our findings suggest that the high support for simpler taxes is to some extent driven by a lack of awareness about the implications of tax simplification.

From a policy perspective our results imply that in general voters find it quite difficult to file their taxes and want a simpler tax system. But our results also suggest that it is not in the interest of voters to implement a flat rate no deductions kind of tax system to get rid of any tax complexity. Instead there is widespread acceptance for at least certain deductions and therefore also a certain degree of tax complexity. Moreover, reactions to our information treatments imply that policy and public debate would benefit from a more balanced debate about the pros and cons of tax simplification. This could reduce information asymmetries and help to implement a more efficient tax policy.

Chapter 4 of this thesis which is coauthored with Philipp Dörrenberg, Benjamin Loos and Johannes Voget analyses how taxes affect the trading behavior of private stock-market investors. We exploit a large reform of capital-gains taxation in Germany combined with confidential portfolio-level daily panel data to study the causal effect of capital-gains taxes on individual stock-trading behavior and the the tendency of investors to sell gains with a higher probability than losses (disposition effect). In the course of the reform, Germany moved from a tax system that differentiated between short-term and long-term capital gains (with tax exempt long-term gains) to a system where all capital gains (independent of holding period) are subject to a flat tax of 25%. We find substantial spikes in selling probabilities of losses and gains around the intertemporal tax discontinuity for pre-reform years, and no such spikes in post-reform years. These findings provide evidence of a causal effect of taxes on holding periods. Using difference-in-bunching and hazard-rates techniques, we quantify the tax effect and identify interesting patterns of heterogeneity with respect to gender, age and trading experience. More experienced investors react stronger to the tax cutoff, i.e. the spike in selling probabilities around the discontinuity increases in experience. The experience effect is conditional on a set of controls including age. Therefore the experience effect is not confounded by age.

Results from this research project suggest that a capital gains tax rate decreasing in the holding duration is a useful tool to deter investors from speculating at stock markets. It works through providing an incentive for investors to delay the realization of capital gains and by this means reduces trading as long as the investor is borrowing constraint. Therefore, a well-designed capital gains tax can achieve similar goals as a transaction tax. At the same time, it has one additional benefit that comes into play if investors are reluctant to realize their losses. The latter investment mistake has been documented in the literature (Odean 1998). In a system where the capital gains tax rate decreases over time waiting reduces the tax shield provided by selling losses. Therefore, decreasing the rate over time gives an incentive to realize losses early. While giving a clear policy advice on how a optimal capital gains tax rate looks like is beyond the scope of this chapter. A well educated guess based on the results is to decrease the tax rate gradually over time, instead of having a decrease in the tax rate after a certain holding duration as currently in the US. The big advantage of gradually decreasing the tax rate is that selling decisions are affected from day one of the holding period. This seems to be unlikely in a tax system where the reduced rate applies after a holding duration of one year. In addition, letting the optimal capital gains tax rate converges to zero over time would eliminate firm level distortions coming from capital gains taxation. Since a long term oriented investor effectively has a zero rate.

2 Tax Refunds and Income Manipulation – Evidence from the EITC

2.1 Introduction

Assessing the responsiveness of individuals to policy changes holds key importance for the (optimal) design of tax-benefit systems and predicting the effects of policy reforms. Labor supply and taxable income responses have been studied extensively in the literature (see e.g. Blundell and MaCurdy 1999, Meghir and Phillips 2008, Keane 2011, Saez et al. 2012 and Bargain and Peichl 2016 for surveys). An important insight of this literature is that welfare programs aimed at reducing poverty can trigger adverse responses from recipients, who can maximize their welfare receipt by reducing labor supply or manipulating their taxable income. Because adverse responses are costly to the taxpayer, it is important for effective policy design to know the strength of these responses. One way to measure such behavioral responses is the degree of bunching at eligibility thresholds or kink points in the tax schedule (Saez 2010a; Chetty et al. 2013a).

In this paper, we document and quantify behavioral responses for the Earned Income Tax Credit (EITC), the largest welfare program in the United States. We exploit the discretion of each state in topping up the federal EITC, whereby recipients with the same taxable income receive higher tax refunds in some states than in others, leading to substantial variation in top-up rates across states and over time. Using event studies and a border pair design, we analyze the extent to which tax filers manipulate their taxable income in response to a change in the state top-up rate. To measure income manipulation, we use data by Chetty et al. (2013a) on the share of self-employed tax filers within a county who bunch around the first kink point of the EITC schedule.

In theory, one would expect that higher top-up rates lead to more bunching at the kink point because they give income manipulation a higher pay-off. Figure 1, which illustrates the main finding of our analysis, suggests that the theory is confirmed by the data. Here, we compare counties in states that raise their top-up rate to neighboring counties in a different state that do not experience a raise. After removing time trends, bunching in both groups follows a similar pattern before the raise but diverges thereafter. In states without a raise, it follows the same downward trend, while in states with a raise bunching significantly increases.

While this figure provides prima facie evidence of an adverse response, there are several endogeneity concerns that prevent us from interpreting this relationship as causal. One important concern is that states set top-up rates with adverse responses in mind. A state that expects a strong response may be reluctant to raise the top-up compared with a state that expects no or very little response. Alternatively, as shown by Neumark and Williams (2016), states may raise the top-up rate to encourage people to participate in the federal EITC, thereby increasing the inflow of federal EITC dollars into the state. Using a border pair design with multiple combinations of fixed effects, we address several important sources of endogeneity. In this research design, we compare the level of bunching in counties on opposite sides of a state border. In this setting, tax filers in treated counties receive a higher tax refund for the same income compared to those living in the control county across the state border.

Our estimates confirm the behavioral responses to a raise in the top-up rate observed in Figure 1. We consistently find a positive effect of the EITC top-up rate on the level of bunching at the kink point. In our preferred specification, an increase in the top-up rate by one within-county-pair standard deviation leads to an increase in bunching by about 8% of a standard deviation. To put this result in perspective, suppose that the average top-up rate would be raised from currently 3 percent by one standard deviation to 10 percent, which would be equivalent to raising the annual refund from USD 180 to 570. In this case, our estimates predict an increase in the degree of bunching by 0.9 percent. Across the US, in absolute numbers, this corresponds to an additional 930,000 self-employed EITC claimants, of which 20,000 would additionally bunch at the kink point.

We also document a change in the response to the EITC top-up rate during the Great Recession in 2008/09. While we observe a strong positive response up until 2007, we find small and statistically insignificant effects from 2008 onwards. This result appears to be driven by an overall higher number of self-employed workers claiming the EITC during the crisis. Because our outcome variable is the ratio of self-employed whose income is close to the kink point over all self-employed EITC claimants, the ratio remains unchanged when both the numerator and denominator are affected by the current economic situation.

Our results suggest that tax filers significantly respond to changes in the EITC schedule by manipulating their taxable income, through either changes in their labor supply or incorrect reporting of their income. Moreover, the response in the total number of EITC claimants points to knowledge effects as well as labor supply responses. When a state introduces a top-up rate, this decision is discussed in the media, which presumably raises the general awareness for the EITC. This may ultimately lead to more people claiming it, as well as more people claiming an amount close to the revenue-maximizing kink point. An alternative explanation for this effect is that the EITC induces people to shift income from employment to self-employment, in which case income manipulation is easier.

This paper adds to the growing literature on the economic and social impact of the EITC.[1] Several studies show that the EITC substantially improves the lives of low-

[1]For surveys, see Hotz and Scholz (2003), Eissa and Hoynes (2006), Meyer (2010) and Nichols and Rothstein (2016).

income families in the United States. For example, positive effects are found on infant health (Hoynes et al. 2015), children's education outcomes (Bastian and Michelmore 2018) as well as poverty reduction (Hoynes and Patel 2015). Other studies emphasize the distortive nature of the EITC by showing that the kink points in the tax schedule provide an important incentive to manipulate taxable income to maximize one's tax refund (Saez 2010a; Chetty et al. 2013a). This manifests itself through a visible degree of bunching of taxable incomes around this kink point, although it remains unclear whether this response is driven by income misreporting or an actual labor supply response.[2] While theses studies have documented and provided a rationale for bunching at the kink point, the contribution of our paper is to quantify the extent to which income manipulation responds to changes in the refund rates. Our results are important for assessing the effectiveness of the EITC and can inform policy-makers about the likely adverse responses of future increases in top-up rates.

More broadly, this paper contributes to the literature on behavioral responses to incentives provided by design features of public policies. A vast literature analyzes labor supply responses, especially to taxation, and numerous surveys and handbook articles have been written on this topic.[3] However, the variation in the magnitude of labor supply elasticities found in the literature is substantial (see Evers et al. 2008, Bargain et al. 2014), and there is little agreement among economists on the size of the elasticity that should be used in economic policy analyses (Fuchs et al. 1998). Heim (2007) and Blau and Kahn (2007) show that married women's wage elasticities have strongly declined over time in the US. A possible explanation for this finding is that a more stable attachment of women to the labor market is responsible for modest participation responses to financial incentives in the recent period. In addition to labor supply, more recent literature has investigated the elasticity of taxable income, following the seminal contributions by Feldstein (1995), Feldstein (1999) (see Meghir and Phillips 2008 and Saez et al. 2012, for surveys). There is also evidence that gross income is less responsive to tax changes than taxable income (Saez et al. 2012; Kleven and Schultz 2014a). Our paper shows that such incentives are also at play for the EITC, and tax filers significantly respond to them.

In the remainder of the paper, we first provide detailed information on the institutional background of the EITC (Section 2.2). In Section 2.3, we explain how we measure

[2] A key result of the existing literature on labor supply reactions to the EITC is that there are positive effects at the extensive margin (Eissa and Liebman 1996; Meyer and Rosenbaum 2001; Grogger 2003; Hotz and Scholz 2006; Gelber and Mitchell 2012). The latter result, which was found primarily for single mothers, does not hold true for secondary wage earners, for whom Eissa and Hoynes (2004) find a decrease in participation. In contrast to these findings, previous research suggests that there are none or only small effects at the intensive margin (Rothstein 2010; Chetty and Saez 2013).

[3] See e.g. Hausman (1985); Pencavel (1986), Killingsworth and Heckman (1986), Heckman (1993), Blundell and MaCurdy (1999), Meghir and Phillips (2008), Keane (2011), Keane and Rogerson (2012), McClelland and Mok (2012), Bargain et al. (2014).

income manipulation, describe the construction of the dataset and present descriptive evidence. In Section 2.4, we describe the empirical strategy. In Section 2.5, we present the main estimation results. Finally, Section 2.6 concludes.

2.2 Institutional Background

We begin by providing information about the federal EITC and the state-specific tax credits (state EITC). We show that EITCs considerably vary across states, such that workers with the same income receive higher tax refunds in some states than in others. We further describe bunching at the first EITC kink point, our outcome of interest, and provide a theoretical discussion concerning why one would expect bunching to increase after a raise in the state EITC.

2.2.1 The EITC

With 26.7 million workers receiving 63 billion dollars per year, the Earned Income Tax Credit (EITC) is arguably the largest and most important welfare program in the US (Nichols and Rothstein 2016). Its aim is to supplement a person's labor income and reduce the income tax burden of low-wage earners while providing incentives to work. The eligibility for the EITC and the amount of tax credit depends on the number of children as well as one's taxable income. To claim the EITC, eligible tax payers have to file a federal tax return. Their income tax liability is then reduced by the amount of the EITC. If the tax credit exceeds the tax liability, the taxpayer receives a tax refund. Taxes are in general paid in the state where the income is earned, although some states have reciprocity agreements that allow taxpayers to file their tax returns in their state of residence (Agrawal and Hoyt 2016).

The EITC tax schedule comprises three parts. In a phase-in region, starting at earnings of zero, the marginal refund increases with every additional dollar of labor income. At the plateau, for a range of annual earnings the tax credit remains constant, while it gets phased out above a certain threshold. For families with one child, for example, the tax credit is phased in at a rate of 34% starting from the first dollar of labor income, reaching the plateau at an annual income of $8,950 in 2009, the last year in our sample. Above the second kink point at $16,420, the tax credit is phased out at 16%. The maximum tax credit for a family with one child is $3,043, which they receive when their annual income lies between both kink points. If it lies above or below the kink points, the tax credit is reduced.[4] For workers without children, the maximum tax credit is very small ($457).

[4]See Figure 41 in Appendix A.1 for an illustration. For families with two children, the kink points 2009 are at $12,570 and $16,420. The maximum tax credit is $5,028, which results in steeper phase-in and phase-out regions compared to the schedule for families with one child.

2.2.2 State-specific tax credits

In our analysis, we exploit the variation in state-specific top-up rates over time. Besides the federal EITC, which is common to all eligible workers in the US, each state can decide to top up the federal tax credit by a certain percentage. The total tax credit is computed as

$$\text{total tax credit} \ = \ \text{federal EITC} \ \times (1 + \ \text{top-up rate}).$$

In some states, for example Minnesota and Wisconsin, the top-up rate depends on the number of children, whereby the top-up is only granted to families with children, or families with children receive higher top-up rates than singles or childless couples.[5] Moreover, some states refund the tax credit if the tax liability becomes negative, while others have a top-up of zero for negative tax liability. Over the years, the number of states with a top-up rate has steadily increased. While in 1996 six states granted a top-up, in 2009, the end of our sample period, it was 20 states. As shown in Figure 3, the top-up rates considerably vary across states, being zero in some states and as high as 40% in the District of Columbia (DC).[6]

EITC claimants in states with a low top-up rate are granted a significantly lower tax credit compared to claimants *with the same pre-tax income* in states with a high top-up rate. Figure 2 illustrates the difference in tax credit for EITC claimants with one child in a state with zero top-up and a state with a top-up rate of 40 percent. A claimant with an income at the first kink point would receive a tax credit of USD 3,043 in a state without a top-up, and USD 4,260 in DC, which has the highest top-up rate in the US. In both states, the kink points of the EITC schedule are the same, although the phase-in and phase-out region are steeper in the state with the high top-up rate. Therefore, in 2009, a family with one child receiving the maximum credit would receive an additional tax credit of USD 30 from a one percentage point increase in the top-up rate. The same family would gain USD 960 through moving from Cheshire county in New Hampshire to neighboring Windham county in Vermont. In 2009, New Hampshire and Vermont are the bordering US states with the largest difference in top-up rates (32 percentage points).

2.2.3 Bunching as a measure of income manipulation

With its two kink points, the EITC schedule provides incentives for recipients to manipulate their taxable income. For tax filers whose income is close to one of the kink points, it is optimal to manipulate their income to be exactly at the kink point. At the first kink

[5]Wisconsin has a top-up rate of zero for childless people, but top-up rates of 4%, 14%, and 43% for families with one, two, and three and more children, respectively

[6]We are aware that DC is technically not a state. However, it has its own EITC.

point, the marginal tax credit switches from a high positive value to zero, such that every additional dollar in earnings above the threshold does not result in higher tax credits. On the other hand, the tax liability increases with every dollar earned, regardless of the tax credit.

There are several margins along which EITC claimants can manipulate their taxable income, namely labor supply, income shifting and tax evasion. A legal margin is adjusting one's labor supply; for example, workers may decide to work fewer hours, thereby reducing their annual earnings while increasing their tax refund. Another way to adjust one's labor supply, especially for self-employed workers, is to smooth the stream of income over time. For self-employed workers whose income is close to the first kink point, it could pay off to postpone projects to the following year, thereby maximizing the tax credit in the present year. A further — yet illegal — margin of income manipulation is incorrectly declaring one's income in the annual tax return.

Such manipulations manifest themselves in a noticeable degree of bunching around the first kink point of the EITC schedule, as documented by Saez (2010a) and Chetty et al. (2013a). In the absence of income manipulation, one would expect the income distribution to be smooth. Instead, however, a large number of EITC claimants report an income that is very close to the first kink point, resulting in a spike in the earnings distribution.

Some groups of workers have a much greater scope for income manipulation than others. As shown by Saez (2010a), pure wage earners — i.e. regularly-employed workers — display no bunching at the kink point, because their taxable income gets directly reported to the Internal Revenue Service (IRS) by their employer, thus limiting the scope for mis-declaring one's income. In addition, work hours are usually fixed in a work contract, making it difficult to adjust one's labor supply. By contrast, self-employed workers have a much greater scope in terms of manipulating their taxable income, as they report the taxable income to the financial authorities themselves, and they are free to choose how much they work.

A raise in the top-up rate provides people with a higher payoff for income manipulation. Therefore, we would expect bunching to increase following a raise in the top-up rate, although we would only expect this effect for self-employed tax filers. Likewise, we would not expect any effect for tax filers without children, because their federal EITC is very small in the first place.

2.3 Data and Descriptive Evidence

In this section, we describe the construction of the dataset and provide descriptive statistics for the main variables. In addition, we produce event study graphs that provide descriptive evidence on an increase in bunching following a raise in the state EITC.

2.3.1 Data

We construct our dataset by linking county-level data on tax filing with state-level institutional data on the EITC, as well as county-level demographic data.

Main outcome of interest. Our main outcome of interest is the bunching of self-employed workers around the first kink point of the EITC schedule. For our analysis, we use data by Chetty et al. (2013a), which were compiled from the universe of individual tax records in the US. In this data, bunching is measured as the share of self-employed EITC claimants in an area whose income falls within a window of USD 500 around the first EITC kink point. The denominator of this share is the total number of self-employed EITC claimants in that area. In 2009, this represents about 600,000 people.[7] From Chetty et al. (2013a), this measure is available for all three-digit zip codes from 1996 to 2009.[8] The bandwidth of USD 500 is maintained over the entire sample period. While we do not have the underlying individual data, Chetty et al. (2013a) show that both the bunching measure itself as well as their regression results are unaffected by the choice of bandwidth.[9] If anything, a noisy measure of bunching at the kink point would increase the standard errors in our regression without leading to biased estimates.

In additional regressions, we consider three outcome variables representing the absolute number of EITC claimants, namely the number of self-employed claimants near the kink point (the numerator of the main outcome), the total number of self-employed EITC claimants (the denominator) as well as the total number of non-self-employed claimants.

Institutional data We combine the county-level data with institutional data on the state EITC from 1996 to 2009, as well as institutional features such as refunds not being granted to workers without children, or negative tax credits not being paid out. We take this data from the NBER TAXSIM database.[10]

[7] To put these numbers into perspective, in 2009, the total number of people with income from self-employment was 16.8 million, which represents 10.7% of the workforce (Source: Social Security Administration). According to Chetty et al. (2013a), the share of self-employed EITC claimants was 19.6%, whereas the share of EITC eligible filers among all tax filers was 18.9% (Source: Brookings Institution, Characteristics of EITC-eligible tax units 2015). Therefore, the share of filers that were both eligible for the EITC and had income from self-employment was around 3.7%.

[8] For this reason, our analysis spans these years, although in the future it would be desirable to have data past 2010, which would allow to study the effects of the EITC during and after the Great Recession. In Appendix A.4, we explain in greater detail how we convert zip-code-level information to the county level.

[9] As explained in footnote 14 in Chetty et al. (2013a), the results are robust to (i) defining the denominator of the bunching measure using only self-employed individuals rather than the full population, (ii) the choice of bandwidth around the kink point, and (iii), a measure whereby bunching is measured as the excess mass over a smoothly fitted polynomial within a certain bandwidth.

[10] See Feenberg and Coutts (1993) for a documentation.

County-level demographic data To run balancing tests as well as controlling for pre-treatment characteristics of counties, we use county-level data on population, employment as well as average wages. Data on employment and wages are taken from the Quarterly Census of Employment and Wages (QCEW), whereas population data are taken from the county-level population statistics provided by the Bureau of Labor Statistics.

2.3.2 Descriptive statistics

Table 1 reports descriptive statistics for the main variables of interest. Because in one of our research designs we only use counties that straddle a state border, we separately report statistics for border counties.

Overall, the outcome variables as well as the regressors of interest strongly increase over time. The first two panels show the evolution of the state EITC. We first consider a dummy that equals unity if a county is located in a state with a top-up rate, and zero otherwise. Over the sample period, the share of counties in states with top-up rates increased from 11.5% to 44%. Likewise, the average top-up rate across all counties increased over the same period. Due to the large share of zeros, it only amounted to 1.6% in 1996, whereas it increased to over 5% in 2009.

Panels 3)-5) display the mean and standard deviation of our outcome variables. The share of self-employed EITC claimants near the kink point corresponds to the bunching measure used in Chetty et al. (2013a). The variables displayed in Panels 4) and 5) represent the denominator and numerator, respectively, of the bunching measure. In addition, Panel 6) reports the total number of EITC claimants per county.

To compare border counties with all counties, we additionally report population and labor market statistics for 2004. According to these statistics, border counties do not differ in their demographic and economic structure from non-border counties. From 1,184 border counties, we construct a dataset of 1,308 border county pairs, whereby a county that straddles multiple counties in a neighboring state is part of multiple county pairs.

2.3.3 Descriptive evidence on top-up rates and income manipulation

The descriptive statistics in Table 1 show that both the top-up rate as well as the extent of bunching increase over the sample period. In a next step, we provide evidence on how both are related. We employ an event study design and use the sample of border pairs, whereby we pay particular attention to the timing of raises in the top-up rate. In order to be able to conduct a standard event study analysis in which the event dummy equals one if the EITC is raised and zero if it remains constant, we exclude from the sample the few county pairs in which the top-up rate decreased (55 pairs).[11] In addition, if a county pair experiences several changes over the sample period, we only include the first change.

As in Figure 1 in the introduction, we are interested in the time trends in bunching in counties that experience a raise in the EITC compared to those where the EITC remains constant. Within each pair, we consider as treated the county that is located in a state with a change in the top-up rate and as control the county located in a state without a change. If top-up rates were to have an effect on income manipulation, following a raise in the state EITC in the treatment group, we would expect to see an increase in bunching in the treatment but not in the control counties.

To provide more systematic evidence of a response in bunching, we estimate an event study equation of the form

$$y_{cpst} = \sum_{k=-4}^{3} \alpha_k \times \mathbb{1}_{[t=t^*+k]} + \sum_{k=-4}^{3} \beta_k \text{ treat}_s \times \mathbb{1}_{[t=t^*+k]} + \boldsymbol{X'_{st}\gamma} + \delta_t + \varepsilon_{cpst}, \tag{1}$$

whereby we consider the period beginning four years before the raise and running until three years after. The subscripts c, p, s and t refer to county, pair, state and time, respectively. We choose as base period the year before the raise, i.e. $t^* = -1$. Our coefficients of interest are β_k, which represent differential changes in bunching between the treated and untreated counties within a pair p relative to the base year. To control for time trends that are common to all counties, we include two distinct sets of fixed effects. The first set, $\mathbb{1}_{[t=t^*+k]}$, controls for average time trends before and after a raise in the top-up rate, regardless of the year in which the raise occurred. Because within our sample period of 14 years the raises occur in different calendar years, we additionally control for year fixed effects δ_t.[12] The year fixed effects ensure that the response to a raise in 1996 receives the same weight in the estimate of β_k as the response in, say, 2008. We also control for time-varying features of the tax code (X_{st}), namely whether the refund depends on the number of children, and whether a positive refund is given if a person's tax credit exceeds his/her tax liability. The error term ε_{cpst} captures all determinants of the outcome that are not explained by the regressors in the above estimating equation.

Figure 4 displays the estimates for β_k. Before the raise in the top-up rate, the estimates are close to zero and statistically insignificant. This is consistent with the parallel pre-trends shown in Figure 1. After the raise, we find significant positive effects on bunching in the treatment relative to the control counties. A raise in the top-up rate

[11] In our main analysis in Section 2.5, these county pairs will be included. We also performed the event study including these cases. The results remain unchanged. The tables are available from the authors upon request.

[12] This approach — controlling for leads and lags as well as year fixed effects — is similar to that used by Jäger (2016) and Fuest et al. (2018a).

increases the degree of bunching by half a percentage point, which amounts to 5% of the mean in 2009.

While these results provide strong evidence of tax filers responding to changes in top-up rates, there are endogeneity concerns that prevent us from interpreting these results as causal. The same economic factors that affect a state's decision to raise its top-up rate could also directly influence bunching. Despite the parallel pre-trends, we may not be able to appropriately control for these factors in the above regression. In the following sections, we address such endogeneity concerns by using a border pair design. In addition, we define here an event as a raise in the top-up rate, such that our estimates reflect the impact of an average raise. In the next section, we are able to quantify the marginal effect of raising the top-up rate by one percentage point.

2.4 Main Analysis - Empirical Strategy

While the event study shows an increase in income manipulation following a raise in the state top-up rate, there are several endogeneity concerns preventing us from interpreting these estimates as causal. In this section, we describe our identification strategy, which relies on a comparison of neighboring counties that are exposed to different EITC top-up rates.

2.4.1 Empirical model

To quantify the effect of the EITC top-up rates on income manipulation, we consider an empirical model of the form

$$y_{cpst} = \alpha + \beta \text{ top-up}_{st} + \boldsymbol{X'_{st}}\boldsymbol{\gamma} + FE(p, s, t) + \varepsilon_{cpst}. \tag{2}$$

The outcome y in county c, which is located in pair p and state s, at time t is regressed on the top-up rate in state s at time t. We control for time-varying state-level features of the EITC ($\boldsymbol{X_{st}}$), namely whether the refund depends on the number of children, and whether a positive refund is given if a person's tax credit exceeds his/her tax liability. In addition, we condition on fixed effects along several dimensions, namely pair, state, time, as well as combinations of these dimensions.

The error term ε_{cpst} captures all of the remaining determinants of the outcome. To account for serial correlation as well as cross-sectional correlation in the error term, we cluster the standard errors at the county and pair level. In addition, we asses our inference through permutation tests in Appendix A.6.

2.4.2 Identification

Given that the top-up rates are not randomly assigned to states but chosen by state governments, we cannot immediately interpret the estimate of β as causal. A causal interpretation requires that there is no correlation of the top-up rate with the error term conditional on controls and fixed effects,

$$cov(\text{top-up}_{st}, \varepsilon_{cpst} | \boldsymbol{X_{st}}, FE(p,s,t)) = 0. \tag{3}$$

There are at least three challenges to a causal interpretation. First, top-up rates may be set endogenously. A state government that expects a strong reaction of taxpayers to a raise in the top-up rate may choose a lower top-up rate than a state expecting a weak reaction. A second problem is economic shocks that affect EITC eligibility as well as the choice of top-up rate. A state that is hit by a negative economic shock may decide to raise the top-up rate to alleviate the consequences for low-income families. At the same time, the shock may lower incomes and thus increase the number of households eligible for the EITC. Therefore, an economic shock can result in a spurious relationship between tax refunds and income manipulation.

A third challenge is differential time trends in income manipulation and top-up rates. As shown by Chetty et al. (2013a), knowledge about the EITC schedule substantially varies across areas and over time. Initially, in some areas, tax filers seem to have no knowledge about the first kink point being income-maximizing, while in other areas there is a high concentration of tax filers with a taxable income around the kink point. Over time, as the knowledge of the EITC spreads, areas with initially zero bunching eventually catch up with those areas with a high degree of bunching from the outset. Unless appropriately controlled for, the estimated effect of top-up rates on income manipulation may reflect those differential time trends rather than a causal effect.

Border pair design. To circumvent these challenges, we apply a border pair design, whereby we compare neighboring counties that straddle a state border.[13] Taxpayers with the same income are eligible for different top-up rates on either side of the border. This setting has quasi-experimental character, as it allows us to compare the change in income manipulation in treated counties that experience a raise in top-up rates to changes in very similar control counties where the top-up rate remains unchanged. The border pair design differs from a conventional panel estimator in the definition of the control group. In the panel estimator, the control group is a weighted average of all other counties, whereas in the border pair design each treated county is assigned its neighbor as a control county. We implement the border pair design with two distinct sets of fixed effects.

[13] Similar approaches have been used by Dube et al. (2010) to evaluate changes in minimum wages in the US, and by Lichter et al. (2015) to estimate the impact of government surveillance in East Germany.

Pair and year fixed effects, $FE(p, s, t) = \delta_p + \delta_t$**.** In the first model, we condition on year and pair fixed effects, which restrict the identifying variation to within pairs over time. A positive estimate of β indicates that a widening of the gap in top-up rates within a county pair leads to a widening of the gap in the outcome. These fixed effects help us to overcome the first of the three challenges. The pair fixed effects control for the average top-up-rate differential in each pair and thus absorb any variation in states' differential setting of top-up rates.

Pair and year fixed effects and pair-specific time trends. While useful as a starting point, the two-way fixed effect model with pair and year fixed effect can yield biased estimates if county pairs diverge in their time trends, which have been shown to be present for bunching. To address this challenge, we additionally include pair-specific time trends in the regression. In this case, the coefficient β is identified off deviations from the time trend within a pair.

Pair-by-year fixed effects, $FE(p, s, t) = \delta_{pt}$**.** In a more demanding specification, we include pair-by-year fixed effects, which absorb all average differences in observable and unobservable characteristics between years within each county pair. Restricting the variation in way is useful to exclude that the estimation of β is confounded by local economic shocks or differential time trends between pairs. Take, for example, a pair that is hit by a negative shock, which in turn leads to a raise in the top-up rates as well as an increase in the level of bunching. Neither the pair nor the year fixed effects would account for that shock. However, the pair-by-year fixed effects absorb such shocks, which raises the plausibility that the identifying assumption (3) holds.

To understand how β can be identified on top of pair-by-year fixed effects, it is instructive to use as a reference point a model with separate time and pair fixed effects. In that model, we exploit variation in top-up rates within pairs over time. A slightly more restrictive model would be one with pair-specific time trends, which exploits variation within pairs over time on top of the time trends. Our model with pair-by-year fixed effects goes yet another step further and allows for year-pair-specific economic shocks. It is possible to identify this model because the top-up rates as well as the outcomes vary *within* each pair. In the fixed-effect estimator for β, each pair-year combination receives equal weight. We no longer use variation within pairs over time, but rather use variation within and across pairs *after differencing out any pair-specific shocks.* In Appendix A.5, we show that a significant amount of variation remains even if we control for pair-by-year fixed effects.

Are changes in state EITCs exogenous? While the border pair design reduces — and in the best case eliminates — the influence of unobserved heterogeneity in explaining the results, there is a concern that both the state EITC and bunching rates are jointly determined by a third factor such as differences in minimum wages, tax rates, or the generosity of social benefits. To address this concern, in Appendix A.2, we investigate whether state-level variables predict changes in top-up rates. Consistent with Bastian and Michelmore (2018), we find no evidence that the generosity of the state EITC is driven by the business cycle, state tax revenues, welfare benefits, or minimum wage levels.[14] This result corroborates the identifying assumption that the level of the State EITC can be considered exogenous in our regressions.

2.5 Results

In the following, we present our estimates for the impact of the state EITC along several behavioral margins. We first present our main results for the border pair design, using different fixed effect specifications. In a further step, we analyze whether the response changed during the Great Recession in 2008/9. In both analyses, inference relies on parametric assumptions about the spatial and serial correlation of standard errors. To assess the robustness of our inference, we perform permutation tests in Appendix A.6, which confirm our main conclusions.

2.5.1 EITC refund rates and income manipulation

Table 2 presents OLS estimation results based on the regression model in Equation (2). We consider three fixed-effect specifications, four outcome variables and two treatment definitions. Each entry is the result of a separate regression of the outcomes listed in Panels A)-D) on the top-up dummy or rate. In Columns (1)-(3), the regressor of interest is a binary variable that equals unity if a state has a top-up rate, whereas in Columns (4)-(6), the regressor of interest is the top-up rate in percentage points (zero for counties located in states without a top-up rate).

Our main measure for income manipulation is the bunching of self-employed EITC claimants within a USD 500 interval around the first kink point of the EITC schedule. For each county, this measure is computed as the number of self-employed EITC claimants within this interval divided by the total number of self-employed EITC claimants. In Panels B and C, we separately estimate the impact of the top-up rate on both components that make up the bunching measure. This allows us to study whether the overall effect is driven by changes in the number of people around the kink point (numerator) or in the overall number of tax filers (denominator). Finally, in Panel D, we also consider as an outcome the number of non-self-employed claimants. If we found an effect of the top-

[14] This result is also consistent with the findings of Castanheira et al. (2012) for income tax reforms in Europe and Foremny and Riedel (2014) for local business taxes in Germany. Both studies show that tax setting is driven by political factors rather than the business cycle.

up rate on this variable, this would be indicative of knowledge effects and labor supply responses rather than manipulation of taxable income.

Effect of the state EITC on bunching. In Columns (1)-(3), we only consider changes in the top-up rate along the extensive margin. The coefficient $\widehat{\beta} = 0.365$ in Panel A, Column (1), means that when a state introduces a top-up rate, bunching increases in a treated county in that state by 0.365 percentage points relative to the neighboring county in a different state where the top-up dummy remains unchanged. This effect amounts to 4.4% of the mean level of bunching in 2004, as well as 19% of a within-pair standard deviation in bunching. The estimated coefficient is statistically significant at the 10% level. In Column (2), when we condition on pair-specific time trends, we find a similar point estimate, although the estimate is less precise and no longer statistically significant. In Column (3) — our most conservative specification — we condition on pair-by-year fixed effects, based on which we obtain an even larger point estimate of $\widehat{\beta} = 0.492$, significant at the 10% level. These results suggest that tax filers respond to the introduction of a state EITC with a higher share declaring an in-come closer to the revenue-maximizing kink point.

While these results provide a first indication of an effect, it should be noted that the effect is driven by changes in a limited number of states. Over the sample period, only 14 states introduced a top-up rate. Within a county pair, the identification comes from switches in the dummy from zero to one, which can only happen once per county over the sample period. By contrast, in Columns (4)-(6), we identify the effect off changes in the top-up rate along both the extensive and intensive margin.

In the model with separate pair and year fixed effects, shown in Column (4), we find no statistically significant effect of an increase in the top-up rate on bunching. However, once we condition on pair-specific time trends or pair-by-year fixed effects, the effect is large and statistically significant. For a within-pair standard deviation in the top-up rate ($sd = 5.43$), bunching increases by $5.43 \times 0.023 = 0.12$ percentage point, which is around 6.6 percent of a within-pair standard deviation in bunching.

Effect on the number of self-employed claimants. The results shown in Panel A represent the effect of an increase in the top-up rate on the *share* of EITC claimants whose income is close to the EITC kink point. This share comprises two components, namely in the numerator the number of self-employed tax filers close to the kink point and in the denominator the total number of self-employed tax filers. A positive effect in Panel A indicates that the numerator increases more than the denominator, leading to a higher share. To assess the relative contributions of both, we separately consider the effects of the EITC in Panels B and C. In Column (1), we find that the introduction of a top-up rate increases the number of tax filers near the kink point by 222, which is larger than the mean number across all sample years (123). At the same time, it leads to an increase in the total number of self-employed EITC claimants by 893, which is around 75% of the mean in 2004. In Column (4), we estimate that a one-percentage-point increase in the state EITC raises the number of self-employed claimants near the kink point by 8.6 (1.7% of a within-pair standard deviation) and increases the total number of self-employed claimants by 36.5 (1.6% of a standard deviation). With both regressors, the effect size increases when we condition on pair-by-year fixed effects. To sum up, the top-up rate increases both the numerator and the denominator, with the former increasing more than the latter.

Effect on non-self-employed EITC claimants. Finally, in Panel D, we estimate the impact of the EITC on the number of non-self-employed claimants. This group is interesting because they have little scope for manipulating their declared taxable income; rather, any effect here is indicative of a change in labor supply. The evidence on this channel is mixed. We find large and statistically significant results when we use the top-up dummy as a regressor, but small and statistically insignificant results when we use the continuous measure of the top-up rate. These results provide suggestive evidence for labor supply effects, although the marginal effect of an in-crease in the top-up rate on bunching appears to be driven by other channels. This is unsurprising given that in general it is (more) difficult to adjust labor supply at the intensive margin — i.e. the number of hours worked — due to frictions in the labor market. Nonetheless, it is possible that a higher state EITC increases labor supply at the extensive margin, which we cannot rule out but also not directly test with our data.

2.5.2 The impact of top-up rates before and during the Great Recession

While bunching had been steadily increasing up until 2007, there was a significant drop in 2008 and 2009, while at the same time the average top-up rate continued its upward trend. A possible reason for these developments is the Great Recession in 2008/09. As noted by Moffitt (2013), the role of the EITC during a recession is ambiguous. One the one hand, if families have lower work income, they may receive higher tax credits. On the other hand, unemployment leads to the loss of tax credits. The aggregate data, displayed in Appendix A.3, suggests that the number of claimants increased from 2008 to 2009 relative to the overall positive trend in the number of claimants. During the Great Recession, the US social safety net underwent a considerable expansion, in particular in the SNAP (Supplemental Nutrition Assistance Program) and unemployment insurance. In comparison, the expansion of the federal EITC was relatively small; eligible families with three or more children received higher tax credits. Figure 5 shows that, on average, top-up rates remained stable from 2009 onwards. If anything, states did not follow the previous trend of gradually raising the top-up rates.

To observe whether the impact of the top-up rate changes with the Great Recession, we estimate a regression with a full interaction of the top-up dummy or rate with dummies for the pre- and post-Great Recession period.

$$y_{cpst} = \beta_1 \text{ top-up}_{st} \times \mathbb{1}_{[t<2008]} + \beta_2 \text{ top-up}_{st} \times \mathbb{1}_{[t>=2008]} + \boldsymbol{X'_{st}\gamma} + \delta_{pt} + \varepsilon_{cpst}. \quad (4)$$

The first term is an interaction between the top-up rate and a dummy that equals one in the pre-crisis years, while the second term is an interaction with a dummy that equals one from 2008 onwards.[15] Our results point to a large and significant effect before 2008, although we find no consistent effects in 2008/9. In Column (1), the effect on bunching in 2008/9 is negative, which is the case because the denominator — the total number of self-employed claimants — reacts more than the number of claimants close to the kink point. These results are broadly consistent with those of Jones (2014) and Bitler et al. (2017), who show that — relative to other social security programs — the EITC played a minor role in alleviating poverty during the Great Recession. In addition, similar results have been found for Ireland by Hargaden (2015), who shows that the extent of bunching at notches in the Irish tax codes were three times larger before than during the Great Recession.

2.5.3 Discussion

Overall, our results support the hypothesis that higher tax refunds create a greater incentive for income manipulation and thus can trigger behavioral responses along several margins. While our data do not allow us to fully distinguish between false declaration of taxable income and labor supply responses at the extensive or intensive margin, our results suggest that both mechanisms are important. Our finding that a raise in the top-up rate increases the extent of bunching at the kink point suggests that there are adverse responses to the state EITC. If the effect was exclusively explained by labor supply responses — especially at the extensive margin — it would be unlikely that we find an effect on bunching. For labor supply responses along the extensive margin, we would rather expect that the numerator and denominator are similarly affected, i.e. the additional number of claimants near the kink point is proportional to the total additional number of claimants. By contrast, the positive effect on bunching suggests that the additional number of claimants at the kink point is much larger relative to the additional number of claimants. While not a proof, these overproportional changes at the kink point to false declarations of taxable income and potentially to labor supply adjustments at

[15] While these two dummies are multicollinear, it is possible to include these interactions in the regression because we do not include the dummies on their own.

the intensive margin.

Nonetheless, the effects on the total number of self-employed EITC claimants suggests that not all behavioral responses to the state EITC can be classified as adverse. One of the central aims of the EITC is to provide recipients with an incentive to work. The results in Panel C of Tables 2 and 3 and to some degree the results for non-self-employed workers in Panel D suggest that these incentives work. A higher top-up rate induces more people to work, and this additional labor supply appears to be spread out along the income distribution rather than concentrated at the kink point.

2.6 Conclusion

Virtually all public policies trigger behavioral responses by their recipients. In this paper, we document and quantify such behavioral responses for the Earned Income Tax Credit, the largest welfare program in the US. Using data on the extent of bunching at the first kink point of the EITC schedule, and exploiting variation in state-specific tax refunds over time, we find significant behavioral responses along several margins.

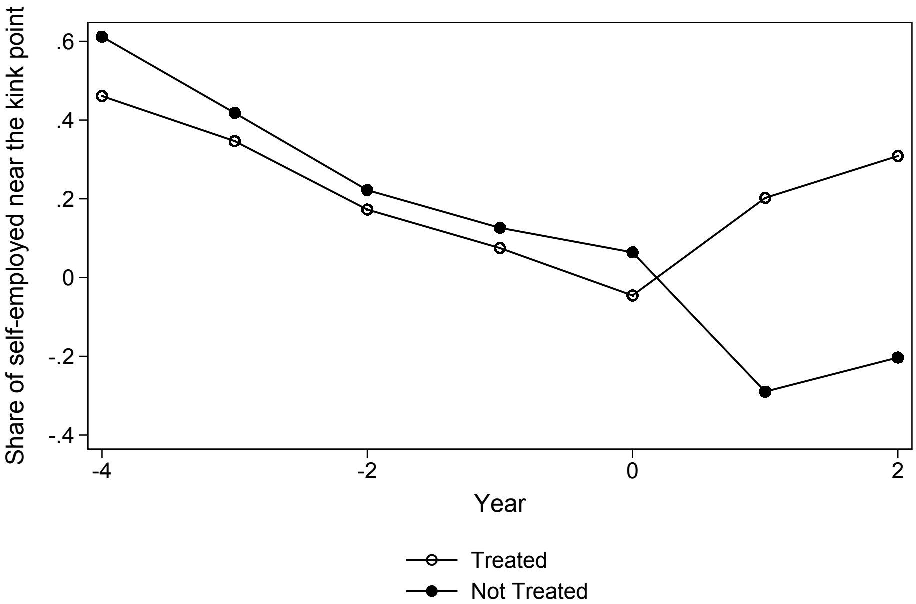
First, we document that higher EITC top-up rates increase the number of self-employed people who claim the EITC. This effect can either represent an increase in (self-employed) labor supply, or a change in tax filing behavior. LaLumia (2009), for example, shows that raises in the tax refund increase the likelihood that potential recipients declare their self-employed income.

Second, we show that a raise in the EITC top-up rate leads to an overproportional increase in the number of self-employed claimants who declare an income close to the income-maximizing first kink point of the EITC schedule. The increase in this number is considerably larger than that of the total number of self-employed EITC claimants, in turn leading to more bunching at the kink point. This result points to a significant adverse response, namely that tax filers choose to declare their taxable income or their labor supply or both in a way that maximizes their EITC receipt.

These results suggest that the EITC — like any other welfare program — triggers behavioral responses. To policy-makers, some of these — for example, labor supply at the extensive margin — are desirable, while adverse responses are not, such as false declaration of taxable income. While our results for the effect on bunching suggest that income manipulation is an important response, we would require more detailed data to fully disentangle labor supply effects from manipulation of taxable income through false declaration. For future work, we are hoping that such data become available.

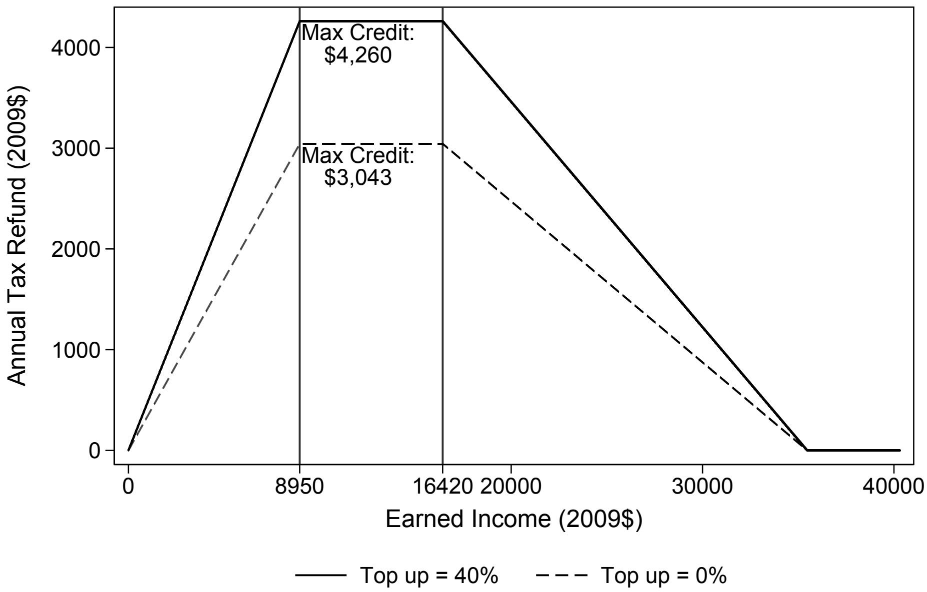
Figures and Tables

Figure 1: Bunching of self-employed near the kink point in counties with and without a raise in the top-up rate.



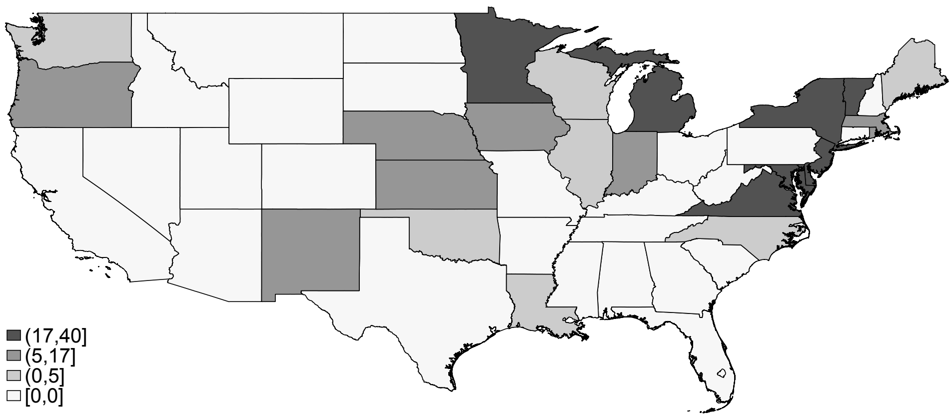
Notes: This figure compares the level of bunching before and after a raise in the top-up rate in the treatment counties — located in a state with a raise in $t = 0$ — with that in neighboring countries without a raise in the top-up rate. To make the counties comparable across years, year fixed effects have been controlled for.

Figure 2: Tax credit in states with high and low top-up rates



Notes: This figure displays the EITC schedule for claimants with one child in a state with zero top-up and a state with a top-up rate of 40 percent in 2009. The vertical lines mark the first and second kink points. Tax units with adjusted gross income above the earned income threshold are not eligible. Families with unearned income may be ineligible.

Figure 3: State-specific top-up rates in 2009



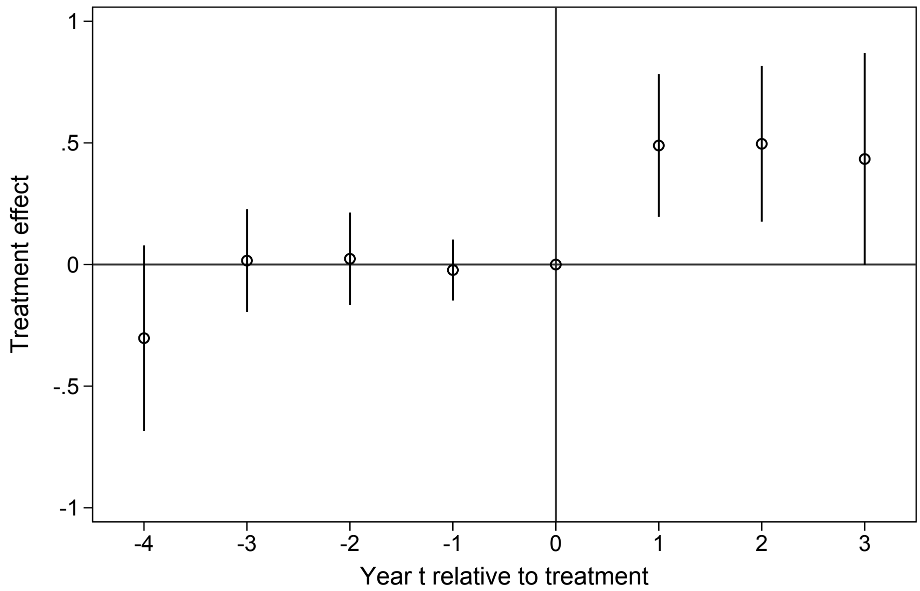
Notes: This Figure shows the variation in top up rates across states in 2009. Darker colors refer to higher top-up rates.

Table 1: Descriptive statistics

	All counties		Border Counties	
	Mean	SD	Mean	SD
1 Top-up dummy (1 if state has a top-up rate, in percent)				
1996	11.5	32.0	13.1	33.7
2000	22.8	42.0	25.7	43.7
2004	26.3	44.0	29.5	45.6
2009	43.8	49.6	46.6	49.9
2 Top-up rate (in percent)				
1996	1.60	5.94	2.17	7.58
2000	2.59	6.03	3.00	6.48
2004	3.14	6.99	3.71	7.61
2009	5.51	8.34	6.03	8.77
3 Share of self-employed EITC claimants near the kink point				
1996	5.04	1.55	5.00	1.61
2000	7.18	2.99	7.08	3.13
2004	8.50	3.98	8.29	3.96
2009	9.27	4.68	8.97	4.53
4 Self-employed EITC claimants				
1996	817	2,755	753	2,149
2000	866	3,235	826	2,957
2004	1,187	4,309	1,108	3,982
2009	1,434	5,004	1,326	4,782
5 Self-employed EITC claimants near the kink point				
1996	54	328	52	264
2000	91	572	99	702
2004	143	751	138	773
2009	194	902	178	904
6 Non-self-employed EITC claimants				
1996	4,714	13,244	4,458	12,659
2000	4,734	13,430	4,507	13,054
2004	5,006	13,135	4,736	12,768
2009	5,371	13,336	5,054	12,895
Population, 2004	93,320	302,015	93,581	260,604
Unemp rate, 2004	5.69	1.82	5.67	1.87
Empl rate, 2004	94.31	1.82	94.33	1.87
Average wage, 2004	28,805	6,141	28,909	6,219
Counties	3141		1184	
County pairs	NA		1308	
States	51		49	

Notes: This table reports descriptive statistics for the main variables of interest for selected years. The top-up dummy equals one if a county lies in a state with a top-up rate. The column on the left reports the statistics for all counties in the US, while the column on the right only reports the statistics for counties that straddle a state border.

Figure 4: Bunching before and after a raise in the top-up rate.



Notes: This graph displays the coefficient estimates of β_k in Equation (1). The specification includes year fixed effects and controls and is estimated on a sample restricted to counties straddling a same state border. The reference category is the year before treatment. The vertical line represents the period zero, i.e. the year before treatment.

Table 2: The Effects of Top-up rates on Bunching

	(1) Top-up Dummy	(2) Top-up Dummy	(3) Top-up Dummy	(4) Top-up Rate	(5) Top-up Rate	(6) Top-up Rate
A. Share of self-employed near the kink point						
Top-up	0.365*	0.310	0.492*	0.010	0.023**	0.029**
	(0.220)	(0.244)	(0.284)	(0.009)	(0.011)	(0.013)
B. Number of self-employed EITC claimants near the kink point						
Top-up	222.665**	237.897**	295.758**	8.587**	9.125**	11.439**
	(96.409)	(105.777)	(126.710)	(4.284)	(4.545)	(5.468)
C. Total number of self-employed EITC claimants						
Top-up	892.661**	951.592**	1168.111**	36.507**	35.294**	43.317**
	(376.278)	(410.242)	(492.554)	(17.155)	(17.257)	(20.765)
D. Total number of non-self-employed EITC claimants						
Top-up	1930.364*	2151.488*	2633.806*	58.175	61.947	77.751
	(1095.349)	(1210.460)	(1424.684)	(42.224)	(46.721)	(55.923)
Controls:						
Year FE	Yes	Yes	No	Yes	Yes	No
Pair FE	Yes	Yes	No	Yes	Yes	No
Pair-spec Time tr.	No	Yes	No	No	Yes	No
Pair × Year FE	No	No	Yes	No	No	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
N	36608	36608	36592	36608	36608	36592

This table displays the results of separate OLS regressions of our outcome variables on a dummy for having a top-up rate Column (1) – Column (2) and on top-up rates Column (3) – Column (4) , as well as the controls and fixed effects. Controls are: an indicator that equals unity if the refund depends on the number of children and an indicator that equals unity if positive refunds are given. The sample size differs between Columns (5) and (6) because in 16 county pairs, the information from one county was missing and, therefore, the pair is a singleton.[a] Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at county and pair level, are reported in parentheses.

[a]See Correia (2015) for an argument why singletons should be dropped from such a regression.

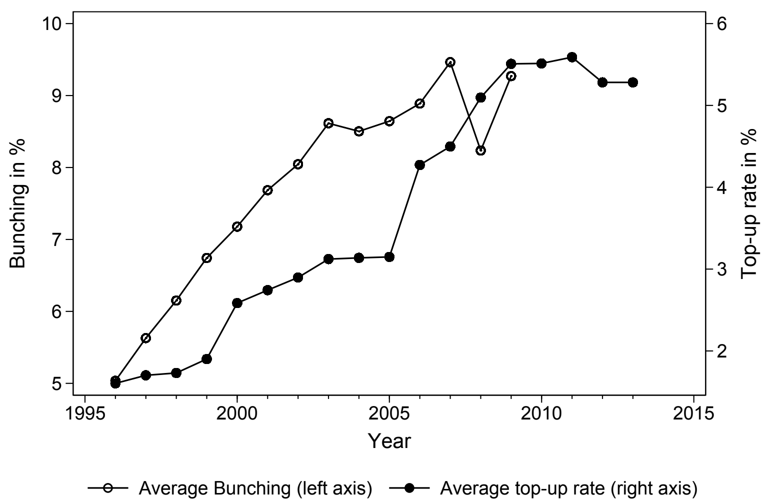


Figure 5: Top-up rates and bunching, 1996-2013

Notes: This figure shows the average level of bunching in percent (left axis), as well as the average top-up rate. Each dot represents the average across all counties within a given year. The data on bunching are only available up to 2009.

Table 3: Top-up rates and bunching before and during the Great Recession.

	(1) Top-up Rate	(2) Top-up Rate	(3) Top-up Rate
A. Share of self-employed near the kink point			
Top-up before 2008	0.019*	0.022*	0.035***
	(0.010)	(0.011)	(0.013)
Top-up 2008, 2009	-0.022***	0.026***	0.004
	(0.008)	(0.009)	(0.010)
B. Number of self-employed EITC claimants near the kink point			
Top-up before 2008	9.783**	10.341**	12.786**
	(4.621)	(4.898)	(5.753)
Top-up 2008, 2009	4.254	3.437	5.419
	(3.311)	(3.150)	(4.609)
C. Total number of self-employed EITC claimants			
Top-up before 2008	38.208**	38.659**	48.186**
	(17.277)	(17.859)	(21.389)
Top-up 2008, 2009	30.349*	19.552	21.548
	(17.339)	(15.167)	(19.960)
D. Total number of non-self-employed EITC claimants			
Top-up before 2008	65.071	68.770	90.632
	(44.377)	(48.234)	(58.560)
Top-up 2008, 2009	33.206	30.033	20.166
	(36.788)	(42.030)	(51.956)
Controls:			
Year FE	Yes	Yes	No
Pair FE	Yes	Yes	No
Pair-spec Time tr.	No	Yes	No
Pair $\times$ Year FE	No	No	Yes
Controls	Yes	Yes	Yes
N	36608	36608	36592

This table displays results of separate OLS regressions of our outcome variables on top-up rates, as well as the controls and fixed effects. Significance levels: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Standard errors, clustered at county and pair level, are reported in parentheses.

3 Do people really want a simple tax system? Evidence on preferences towards income tax simplification

3.1 Introduction

Should tax systems be simplified? The conventional wisdom seems to be: *yes*, tax systems should be simpler! As the literature shows, there are indeed many good reasons for supporting tax simplification. For example, recent studies show that the self-employed value tax simplicity and leave money on the table because of complex tax schedules (Aghion et al. 2017; Benzarti 2017), that taxpayers underreact to complex tax incentives (Abeler and Jaeger 2014), that the existence of complexity-adding tax expenditures facilitates tax evasion (Kleven et al. 2011; Paetzold and Winner 2016), and that tax complexity reduces the take up of tax refunds by firms (Zwick 2018). Complex tax systems also lend scope to lobby groups to achieve beneficial tax treatment for the groups they represent (Brusco et al. 2014), have negative effects on income inequality (Aghion et al. 2017), and possibly come with substantial resource costs (Pitt and Slemrod 1989). It is therefore maybe not surprising that many economists propose implementing tax reforms that make the system less complex, for example through lower rates and broader bases.[16] Not only many economists and academics support a simpler tax system; the conventional wisdom among policy makers and journalists also seems to hold that simplifying tax systems is generally desirable.[17]

However, despite many arguments in favor of tax simplicity, there are also economic arguments in support of a certain degree of tax complexity (see, e.g., OECD 2010b, Hines 2016, and Hines 2019). For example, a fairly complex tax system with a certain amount of tax expenditures i) makes it possible to tailor taxes to individual situations and to use 'tagging' components,[18] ii) allows to tax highly elastic goods at effectively lower rates, iii) avoids tax compounding (e.g., favorable tax treatment of pensions and retirement savings), and iv) enables to include Pigouivian elements into the tax system that correct for market failures or internalize negative externalities, e.g. research tax credits (Hines

[16]The simplification of the tax system is a key objective of many income-tax reform proposals by economists in various countries. For example, Gale (2001), Rohaly and Gale (2004) and Gravelle and Hungerford (2012) for the US, James et al. (1997) for Australia, New Zealand and the United Kingdom, Tran-Nam (2000) for Australia, and Fuest et al. (2008) and Wagner (2006) for Germany.

[17]See newspaper coverage for the US showing that many politicians and journalists make a case for a simplified tax system: e.g., Economist (2005), Economist (2013), NYT (2015), NPR (2015), Forbes (2017), as well as Vox (2017).

[18]Tagging is for example studied in Cremer et al. (2010). Gordon and Kopczuk (2014) study the selection of the income tax base and show that it is advantageous (in the sense of approximating a tax on ability as good as possible) to allow for particular tax expenditures (such as the dependents' deduction). Thus, there is an implicit rationale for not having the simplest possible tax system with a broad base and without any tax expenditures.

2016; Hines 2019). These elements of a complex tax system can contribute to making the tax system more efficient. Complex components of the tax system potentially also have redistributive purposes – think for example of deducting the costs of elderly care of family members or allowances for dependent children – and might therefore be viewed as equity enhancing and fair.[19]

The arguments in support of tax complexity do not feature prominently in the debate about tax complexity. The general-wisdom support of simple tax systems might therefore miss out some of the efficiency and redistribution aspects of tax simplification. In addition, the data show that most tax systems remain very complex and are characterized by the presence of a large amount of tax expenditures (e.g., OECD 2010b). Figure 6 shows for the US that there is an upward sloping trend in the growth of tax expenditures, suggesting that the tax system tends to get even more complex over time. It is thus a puzzle why most tax systems remain so complex although the general wisdom seems to hold that substantial tax simplification is desirable.

Considering important arguments both in favor and against simplification, and in light of lasting complexity of real-world tax systems, attitudes towards tax simplification among the public may be more nuanced than they seem on first glance. In this paper, we aim to shed light on the debate about tax complexity and collect new survey and experimental data to study preferences for tax simplicity among a representative sample of the German population.[20]

Our paper has two main objectives: First, we document preferences for tax simplicity and report which fraction of the population supports a simplified tax system. This first part of the paper particularly investigates if public support for tax simplicity is indeed consistent with the conventional wisdom and consensus that apparently exist in academic and public-press debates. Second, we investigate if preferences for tax simplification are driven by a lack of awareness about the implications and consequences of tax simplification. In other words: Is the general wisdom regarding tax simplicity driven by awareness and information deficits? Do individuals frequently express desires for tax simplification without appreciating the implications of tax simplification? To address these questions, we i) elicit if people are in favor of specific tax expenditures which add complexity to the tax system, and we ii) implement two randomized experiments to study if preferences for tax simplicity are elastic to information in favor and against tax simplification. That is, we study if preferences are shifted once people are made aware of the trade-offs behind

[19]The role of economic theory in this discussion is addressed by Hines (2016) who concludes that: "Economic theory does not say that an efficient and equitable income tax system has a broad base and a low rate, and in fact the theory has never said that." (Abstract).

[20]Germany has a considerably complex tax system with many tax expenditures. In addition, the simplification of the tax system is a frequently debated issue in the media and among politicians. The case of Germany might therefore be a well suited case to study preferences for tax complexity. See Section 3.2.2 for more details on tax complexity in Germany.

complexity and simplicity.

The working hypothesis throughout our paper is that preferences for tax simplification are possibly shifted once individuals are explicitly forced to think about their simplicity preferences in concrete applications or if they are made aware of information and aspects of tax complexity that are new to them. Presumably, such a shift in preferences is greater in response to information *against* tax simplicity than to information *in favor of* simplicity, because pro-simplicity arguments play a more prominent role in the public debate and misperceptions are thus likely to be less prevailing with regard to pro-simplicity arguments.

We included a set of questions in the context of tax simplification into the *German Internet Panel*, a representative survey of the German population ($N = 2464$). The survey questions are tailor-made and designed to speak to the two objectives of our paper. The concept of tax complexity is complex in itself, and for the purpose of the survey we decided to focus on one particular dimension of tax complexity: the number of tax expenditures.[21] While there are clearly more dimensions of tax complexity (e.g., documentation requirements, etc.), tax expenditures are a main source of tax complexity and a major issue in the debate about complexity; moving to a system without any tax expenditures would clearly make any existing tax system simpler, easier and more comprehensible. This definition of tax complexity is consistent with Slemrod and Kopczuk (2002) and Kopczuk (2005) who characterize an income tax system as complex when it features many deductions (also see the discussion about tax-complexity measurement in Abeler and Jaeger 2014). We further focus on the case of the personal income tax (PIT) which appears to be the natural choice for a survey on tax attitudes among the general public.

Our survey reveals the following main results. First, more than 90% of respondents have a preference for tax simplification. We survey this question on a scale from 1 to 6, where 6 means strong support for tax simplification, and find an average of 5.2. This result confirms that the prevailing view indeed holds that tax simplification is desirable. Apparently, supporting tax simplification seems to be the obvious choice and general-wisdom reply for the large majority of respondents. This finding is the starting point of our analysis, in which we aim to investigate if the high support for tax simplification persists as we make people aware of the trade-offs between tax simplicity and complexity. In other words, we move on from the finding that tax simplification is the obvious choice in the survey and study if the matter becomes less obvious as we highlight the trade-offs behind this choice.[22]

[21] A straight forward definition of tax expenditures is provided by the Tax Policy Center (2019): 'Tax expenditures are special provisions of the tax code such as exclusions, deductions, deferrals, credits, and tax rates that benefit specific activities or groups of taxpayers.'

[22] An analysis of the anatomy of simplification preferences shows that age, gender and the perceived

Second, we then dig deeper into preferences for tax simplicity and study if the large support for simplicity holds as we ask for the preferred tax system in particular contexts. For this purpose, we have a series of questions in which we present participants with the living situations of two fictitious taxpayers,[23] and then survey if these two taxpayers should pay the same amount of taxes or if any of the two should pay less in taxes than the other person. We designed the questions in a way where the two fictitious persons are similar in all tax-relevant means except for one particular aspect of their living situation. In particular, the two fictitious taxpayers were different with respect to i) the necessity to spend money on the elderly care of a family member, ii) the amount they donate for charity, and iii) the distance between their home and work place.

The results show that in all three scenarios – i) to iii) – a considerable fraction of respondents indicate that the two persons should *not* pay the same amount of taxes and that the taxpayer with the additional cost burden should pay less. In particular, more than 60% of respondents think that the tax system should account for elderly-care costs. In the case of the other two scenarios, roughly 40% and 30%, respectively, believe that the tax system should account for the respective difference between the taxpayers.[24] Interestingly, the share of people who indicate that the tax system should *not* account for certain differences in living situation (and thus be as simple as possible) is in all three scenarios considerably smaller than the 90%-share of respondents who generally support to have a simpler tax system.

These results are thus evidence that many people prefer a tax system which allows for a differential tax treatment of taxpayers in different living situations. Obviously, such a differential treatment of two otherwise identical taxpayers can only be achieved through tax expenditures, and thus through a substantial degree of tax complexity. The survey respondents might not even realize that their choices imply tax complexity. However, the exercise provides evidence that people implicitly have preferences for a tax system that is more complex than a system without any tax expenditures. As we force participants to think about specific complexity-adding expenditures, many indicate that they wish a tax system that differentiates between taxpayers in different situations.

We are able to show that the answers to these questions are not solely driven by self interest; the result picture remains as we condition on *not* benefiting personally from the

difficulty of filing a tax return are the strongest correlates of simplification preferences.

[23]This survey question technique is similar to some of the survey question types used by Weinzierl (2014), Saez and Stantcheva (2016) and Weinzierl (2017). These studies are not in the context of tax complexity though.

[24]The observation that the 'elderly care' scenario induces more participants to vote for differential taxation than the other two scenarios is interesting in light of the fact that the costs for elderly care are circumstantial (outside control of taxpayers) while the other two are choices. Our respondents thus exhibit preferences that are consistent with arguments in the literature strands of optimal taxation and equality opportunity according to which circumstantial differences should be accounted for in the tax system while choices should not.

respective tax expenditure. Our findings thus show that preferences for having certain tax expenditures in the tax system are not (only) driven by the desire to keep those tax expenditures from which someone benefits personally. In other words, preferences for a certain degree of complexity are beyond pure payoff-maximizing considerations.

Third, we implement two randomized survey experiments to study the causal effect of information about the implications and consequences of tax simplification on preferences for tax simplification. These experiments are directly linked to our previous results, according to which a large majority indicate that they support tax simplification, but, at the same time, a large fraction of respondents prefer to account for the living situation of taxpayers through the tax system. The randomized experiments shed light on this puzzle and investigate if preferences for tax simplicity are elastic to information about the implications of tax simplicity. Eventually, the experiments have the objective of understanding if the high support for tax complexity is partly driven by a lack of awareness regarding what it possibly means to simplify taxes (i.e., the trade-off between arguments pro and con simplification).

To maximize power, we have one experimental intervention at the beginning of the survey and one experimental intervention towards the end. In each of the two experiments, we have a neutral control group, one group in which preferences are possibly shifted towards stronger support for tax simplification, and one group in which preferences are possibly shifted towards less support for tax simplification. This set up thus allows us to study if preferences for tax simplicity are elastic at all, and if yes, if they are elastic in both directions (the experimental groups are described in more detail in Section 3.3). The information and arguments that we use to shift preferences in the four treatment groups are inspired by some of the most debated issues in the context of tax complexity (see beginning of this Introduction): i) redistribution and social-policy aspects, ii) tax avoidance and evasion, iii) efficiency and iv) lobbyism and special-interest groups.

The first experiment includes three randomly assigned groups. A control group which is exposed to a neutral statement about the fact that there is an ongoing debate about whether the tax system is too complex because of various tax expenditures. The second group, labeled *Redistribution* group, is exposed to a statement highlighting that tax deductions can be used to reduce the tax burden for taxpayers which are disadvantaged by exogenous circumstances. A third experimental group, labeled *Avoidance* group, is exposed to an argument in favor of tax simplification, namely that a complex system with many deductions offers possibilities to avoid taxes and manipulate taxable income.[25]

[25]Respondents in all three experimental groups see the neutral statement. Respondents in the two treatment groups are thus exposed to the respective treatment in addition to the neutral statement, while control-group participants only see the neutral statement. The neutral statement in the control

The *Redistribution*-treatment significantly reduces the general support of tax simplification relative to the control group. This is evidence that preferences for tax simplicity are elastic to information about the potentially undesirable implications of tax simplicity. The *Avoidance*-treatment did not have any significant effects; coefficients are much smaller than the effects of treatment *Redistribution* and the standard errors are larger than the coefficients. This null-result is potentially explained by the more prominent role of arguments in favor of simplification in the public debate, and thus less misperceptions regarding arguments that support more simplification. However, while the *Avoidance* treatment did not affect preferences for tax simplicity, it did affect how people think about the distributional aspects of tax complexity: it increased agreement with the statement that tax expenditures benefit the rich more.

The second experiment also includes three groups. We again have a control group with a neutral statement. We then have a group in which we highlight that a system with deductions and exemptions provides better opportunities to tax individual capacities and ability; we label this group the *Efficiency* group. Participants in the third group, labeled *Special Interest Group*, are made aware that tax expenditures are potentially used by lobby and special interest groups to bargain tax exemptions for their clientele.[26]

The results of this second experiment show that the *Efficiency* argument significantly lowers the support for tax simplification, relative to the control group. This suggests that participants are not aware of the potential efficiency implications of tax complexity and confirms the finding of the first experiment that preferences for simpler taxes are elastic to information *against* simplification. The *Special Interest Group* treatment did not have an effect, relative to the control group. This is also consistent with the findings of the first experiment; arguments *in favor* of more simplicity do not have an effect on preferences for simplification. For both experiments, we observe that treatment effects are homogenous across different demographic groups.[27]

Fourth, we aimed at eliciting which type of simplifying tax reforms respondents prefer. We offered a choice of different reform approaches towards simplification and asked respondents to decide which ones they prefer (under the assumption that each of the offered approaches is revenue neutral). The most preferred reform (chosen by about 1/3 of the respondents) entails an increase in the degree of progressivity, but eliminates all deductions and tax expenditures. Overall, however, we observe that there is no consensus among respondents w.r.t. the type of tax-simplifying tax reform. These preferred policy

group serves the purpose of making the topic itself equally salient to respondents in all groups.

[26] As in the first experiment, respondents in all three experimental groups are exposed to the neutral statement.

[27] The experimental intervention also did not have an effect on the previously mentioned questions regarding the tax burdens of two similar taxpayers that differ w.r.t. one particular dimension.

choices are also not affected by the first experimental treatment.[28]

Contribution. We identify the following main contributions of our paper and its empirical findings (see section 3.2.1 for an overview of the literature to which we contribute). First, we implement the first nuanced survey in the context of tax complexity and integrate our questions into an established representative survey. The survey design itself therefore stands as a contribution. To this end, we add to a recent literature using tailor-made surveys to study specific research topics. Second, we document in a representative sample of the population that a large majority of individuals has strong preferences for tax simplification. These preferences in support of tax simplification are consistent with the large literature showing that complexity is costly. Third, we study if the large support for tax simplification depends on the extent of awareness about the consequences and implications of tax simplification. In particular, we show that the support for simplification weakens as respondents are confronted with scenarios and information which present potential arguments in support of a certain degree of complexity. This shows that preferences for simplification are elastic to information and context, and suggests that the debate about complexity could potentially benefit from a more nuanced discussion of the pros and cons of complexity. Fourth, we speak to the puzzle that tax complexity keeps increasing despite the largely prevailing view that tax systems should be simplified. In addition, we document that a large fraction of people would prefer a type of tax-simplifying reform that maintains the degree of redistribution/progressivity or even increases it. These two latter points relate to the literature on the political economy of taxes and tax complexity.

The remainder of the paper is organized as follows. Section 3.2 summarizes the related literature (3.2.1), elaborates on the German case of tax complexity (3.2.2), and derives the paper's working hypotheses (3.2.3). Section 3.3 provides an overview of the survey (incl. the randomized components) and its implementation. Section 3.4 presents the survey results. We conclude the paper in Section 3.5.

3.2 Related Literature, Tax Complexity in Germany, and Hypotheses

3.2.1 Related Literature

We relate to the following strands of literature. First, we speak to the literature on the consequences of tax complexity. Several papers show that tax complexity comes with resource costs and foregone money for firms and individuals (e.g., Pitt and Slemrod 1989,

[28] The question was asked before the second experiment, implying that we cannot test the effect of the second experiment on policy choices. The statistical inference of both experiments' results are robust to standard errors that are adjusted to multiple hypothesis testing and exact significance tests (see results section 3.4).

Aghion et al. 2017, Benzarti 2017, Zwick 2018). In addition, tax complexity has been shown to affect inequality (Aghion et al. 2017), and it facilitates lobbying for beneficial tax treatment by special-interest groups (Brusco et al. 2014) as well as tax evasion (Kleven et al. 2011; Paetzold and Winner 2016; Tsankova et al. 2019). Kopczuk (2012) shows that the introduction of a flat-tax reform with lower rates and less tax expenditures increases tax revenues.[29]

Second, we contribute to a related strand of literature showing that the complexity of taxes and other policy-measures distorts the responses to these government interventions and reduces their take-up. For example, Abeler and Jaeger (2014) study the causal effect of tax complexity on tax responses in a lab-experimental situation, and find that people underreact to complex tax incentives. Saez (2010b), along with the survey evidence of Fujii and Hawley (1988), suggests that individuals do not respond optimally to the incentives of the EITC. The complex structure of the EITC also seems to drive its low take-up (Kopczuk and Pop-Eleches 2007; Chetty and Saez 2013; Bhargava and Manoli 2015). Blaufus and Ortlieb (2009) show that complexity systematically distorts the decision to invest in retirement plans. A further set of papers shows that people systematically misperceive tax incentives – presumably due to tax complexity (de Bartolome 1995; Liebman and Zeckhauser 2004; Blumkin et al. 2012; Blaufus et al. 2013; Ito 2014; Feldman et al. 2016; Rees-Jones and Taubinsky 2016; Gideon 2017; Ballard et al. 2018). Furthermore, a simplified filing of the tax return affects filing and compliance behavior (Kotakorpi and Laamanen 2016; Fochmann et al. 2018). Tax complexity is also likely to be related to the established finding that tax salience is relevant for tax responses (Chetty et al. 2009; Finkelstein 2009; Goldin and Homonoff 2013; Feldman and Ruffle 2015). Tax responses also depend on the existence of complexity-adding deductions (Neisser 2017; Doerrenberg et al. 2017; Paetzold 2019).[30]

Third, we relate to (rarely made) arguments that highlight some potential upsides of tax complexity and express a rationale for the implementation of tax expenditures. James Hines discusses why it is not necessarily desirable to have the simplest tax system with a broad base and low rates (Hines 2016; Hines 2019). He particularly focuses on tax expenditures and provides efficiency and equity arguments for why it is costly to reduce or eliminate tax expenditures (see points i) to iv) above on potential arguments in favor of some degree of complexity). He also discusses several potential justifications to have a comprehensive and simple system with a broad base and low rates, and elaborates that many of these justifications for simplicity do not withstand economic reasoning. OECD (2010a) also discuss the rationale for implementing tax expenditures. They particularly point out arguments of i) tax administration costs (costs of broadening the base might

[29]See the first paragraph of the Introduction for more context for some of these papers.

[30]Somewhat related also is the finding by Brown et al. (2017) that complexity complicates the ability of consumers to value life annuities (such as social security benefits).

exceed the corresponding efficiency gains), ii) equity and social-policy considerations (tax provisions might have the same purposes as social benefits), iii) correcting of market failures (internalize positive external effects), and iv) a political-economy argument, that they borrow from Hettich and Winer (1999), according to which the elimination of tax expenditures possibly reduces tax revenues (abolishing tax expenditures implies that government will be less able to discriminate among heterogeneous taxpayers and voters, which will lead to an increased overall opposition to taxation). Given that the existence of tax expenditures adds complexity to the tax system, Hines (2016), Hines (2019) and OECD (2010a) thus provide arguments for keeping a certain degree of complexity and not move to the simplest possible system. As discussed in footnote 18, Gordon and Kopczuk (2014) also provide a rationale for not having the simplest possible tax system.

Fourth, we touch upon a literature on the political economy of taxes and tax reforms (e.g., Brennan and Buchanan 1980; Meltzer and Richard 1981; Bierbrauer et al. 2018; Bierbrauer and Boyer 2018). We point to an apparent puzzle that tax complexity remains high in the real-world despite the seemingly wide support for tax simplification. To this end, we for example relate to Hettich and Winer (1988) who model the existing tax system with several expenditures as the result of a political process and a government that maximizes political support. A few papers explicitly study tax complexity in a political-economy set-up and investigate how tax complexity arises in the interaction between voters and politicians (Warskett et al. 1998; Galli and Profeta 2009). Our paper speaks to these papers as it suggests that arguments against tax simplicity could play a more prominent role in the voting process if voters were more aware of the trade-offs behind tax complexity and simplicity. To the extent that our paper shows that individuals have misleading information regarding taxes, we also relate to literature showing that such information frictions induces the government to implement inefficient tax policy (Boccanfuso and Ferey 2019).

Fifth, we join a set of papers that set up tailor-made surveys with randomized components to study a particular research topic. Topics that were investigated in such tailor-made surveys include preferences for redistribution (Cruces et al. 2013; Kuziemko et al. 2015; Alesina et al. 2018; Roth and Wohlfart 2018), beliefs about behavioral responses to taxes (Cappelen et al. 2018), immigration (Alesina et al. 2018), social preferences (Kerschbamer and Müller 2017), inheritance taxation (Bastani and Waldenstroem 2019), reforms in the Euro area (Dolls and Wehrhoefer 2018), tax-compliance attitudes (Doerrenberg and Peichl 2018), education (Lergetporer et al. 2018; Lergetporer et al. 2018), road mileage user fees (Duncan et al. 2014), misperceptions in the context of different economic policies (Stantcheva 2020), and macro-economic expectations (Roth and Wohlfart 2019). As we do, the randomized components in these surveys show that information can have an effect on attitudes and preferences. Several of these studies rely on commercial providers who conduct the surveys online and establish representativity through a reweighting of the initially non-representative sample. We implement our questions within the GIP, an established survey with a representative sample of the German population that was explicitly build up for research purposes.

3.2.2 Tax Complexity in Germany

Our survey is conducted in a country with a fairly complex income-tax system. For example, Germany's tax schedule presumably includes more than 500 deduction possibilities, according to Kirchhof (2011). A study by Blaufus et al. (2014) finds that the large number of tax expenditures along with other particularities of the German tax system translate into considerable income-tax compliance costs of filing taxes. Using survey data, the study estimates aggregate compliance costs for Germany of 6-9 billion EUR, corresponding to 3.1-4.7% of total 2007 tax revenues. Blaufus et al. (2017) show that expenses for a professional tax preparer are smaller than the savings in tax liability that are due to the tax preparer. In other words, it pays off to have a professional tax preparer. This is a further indication that the system is complex; in a world with the easiest possible tax system it would most likely not pay off to have the tax return filed by a professional. Overall, studying the topic of tax complexity in the context of Germany thus appears a sensible choice given its complex income tax system. In light of the large number of tax expenditures, studying complexity through its dimension of the number of tax expenditures is also reasonable.

A further reason for why Germany is an interesting case to study complexity is that there are frequently returning debates about tax simplification in the public, media and among politicians. One prominent example of this debate is the proposal by prominent politicians (particularly in the conservative center-right party) to simplify the tax system in a way that makes it possible to file the income-tax return on a sheet of paper that is not larger than a usual German beer coaster (such proposals were originally made in 2003 and kept coming back ever since; see FAZ 2004 or Goettinger Tageblatt 2018). Another salient example is the proposal of a prominent academic tax lawyer (Paul Kirchhof) during election campaigns to introduce an income-tax system with a flat rate of 25% and considerably less tax expenditures (see e.g., FAZ 2005).

3.2.3 Main Objectives and Hypotheses

As sketched in the Introduction, our paper has two main objectives. Based on the related literature and the public discussion about tax complexity (as described above), we derive the following hypotheses regarding these two objectives.

Our first objective is to document preferences for tax simplification among a representative sample of the population and to understand related aspects of tax simplification.

Our expectation with respect to this objective is that the support for tax simplification is high. This expectation builds on the observation that both the public debate and the professional discussion (in academic literature and press) are centered around critiques about overly complex tax systems and proposals to simplify taxes, while economic arguments in support of a certain degree of tax complexity play a considerably less prominent role. In light of the debates, asking people about their desire to simplify taxes then presumably triggers a natural and obvious reply, namely that the tax system needs simplification.[31] We provide first survey evidence documenting if the support for tax simplification is as high as we would expect in light of the debates about this topic.

Our second objective immediately builds on the first objective. We aim to understand if the (presumably large) overall support for tax simplicity is to some extent driven by a lack of awareness about the trade-offs behind complex and simple tax systems. Our first strategy towards this objective is forcing people to think about tax-relevant situations that are potentially familiar to them and then let them decide how the tax system, in their view, should deal with these situations. Because the debate about complexity is leaning towards simplification, we deliberately confront individuals with scenarios which they do not immediately relate to the debate about complexity and deliberately do not mention to them that their decisions could have implications for tax complexity. This allows us to investigate individuals' preferences towards complexity-adding components in the tax system in the absence of the public-discussion-induced prejudices towards more tax simplicity. Our expectation for this part of the paper is that the share of respondents who indicate that the tax system should *not* account for certain differences in living situation (and thus be as simple as possible) is considerably smaller than the share of respondents who generally support to have a simpler tax system. It is of course possible to believe that the system should account for the described differences in living situations and at the same time think that the overall tax system should be somewhat simpler. However, if the general support for simplification is considerably higher than the share of respondents who think that the system should not account for differences across people, then this could point in the direction that some people are not aware of certain aspects of tax complexity.

The three scenarios that we present to respondents differ with respect to their degree of being circumstantial (exogenous) to taxpayers or the result of a choice. Building on the literature on optimal taxation and equality of opportunity (e.g., Alesina and Angeletos 2005; Durante et al. 2014; Ooghe and Peichl 2015), we further hypothesize that the share of people who believe that circumstantial living situations should be accounted for in the tax system, is higher than the share of people who believe that chosen living situations

[31] To some extent (and certainly exaggerated), asking for general simplification preferences could turn out to be similar to asking people if they wish to have a higher disposable income.

should be accounted for.

The second strategy in the context of our second objective builds on randomized survey components that expose participants to information/arguments against and in support of tax simplification. The rationale for this approach is simple: if the provided information shift individuals' preferences, then individuals are indeed not aware of certain aspects of tax simplification. We hypothesize that misperceptions and a lack of awareness are less prevalent when it comes to information *against* tax simplification. As a result, shifts in preferences are presumably larger in response to information against simplification than to responses in support of simplification. This hypothesis (again) builds on the observation that both the public debate and the professional discussion are centered around tax simplification.

Overall, our paper and the hypotheses relate to papers that emphasize the importance of the public opinion and the role of taxpayers as voters for the design of the tax system (e.g., Hettich and Winer 1988). We expect that the public opinion is dependent on context and information, and that attitudes towards tax simplicity of taxpayers/voters are more nuanced than it apparently seems on first glance.

3.3 The Survey

3.3.1 German Internet Panel

Our questions are embedded in the *German Internet Panel* (henceforth: GIP).[32] The GIP is a longitudinal survey that is operated and administered at the University of Mannheim in Germany.[33] The main purpose of the panel survey is to collect 'data on individual attitudes and preferences relevant in political and economic decision making processes'. GIP data are collected online on a bi-monthly basis. The survey is representative for the German population aged 16 to 75 (see Blom et al. 2015 for more details on its representativity). Recruitment was conducted offline with face-to-face interviews, during which respondents were invited to the online panel. To ensure the representativeness of the sample, the GIP includes respondents without prior computer or Internet access by providing them with the necessary equipment and training (Blom et al. 2017).

The survey includes repeated questions (included in every wave) as well as questions only included in single waves. We included a block of questions in wave 36, which went

[32]The background information about the GIP in this subsection are partly based on the respective subsection in Doerrenberg and Peichl (2018).

[33]To be more precise, the survey is based at the "Collaborative Research Center 884 on Political Economy of Reforms", which is funded by the German Science Foundation (*Deutsche Forschungsgemeinschaft*, SFB 884). See `http://reforms.uni-mannheim.de/` for background information on the research center. Also see the general survey description in Blom et al. (2015) and at `http://reforms.uni-mannheim.de/internet_panel/home/`. Examples of GIP-based papers include Kerschbamer and Müller (2017), Müller and Renes (2017), Dolls and Wehrhoefer (2018), Engelmann et al. (2018), Doerrenberg and Peichl (2018) and Blesse and Heinemann (2019).

to the field in July 2018 and included 2464 participants (summary statistics below).

3.3.2 Survey Structure and Questions

We designed a block of 10 survey questions and integrated these questions into the GIP. Our question block focuses on matters of tax simplicity and for our analysis we can complement the results from our questions with the results from other questions in the same wave and other waves of the GIP (for example, background characteristics of the participants are available although they are not surveyed in our question block). Our questions were embedded in a regular wave of the GIP (wave 36, July 2018) and were surrounded by other GIP questions.

Our survey questions and the respective reply categories are shown in full in the Appendix.[34] The GIP has a professional and experienced team of survey experts who supported us in developing and formulating our survey questions. Our questions therefore meet up-to-date standards of survey methodology. The survey and its structure (in chronological order) are summarized in the following.

- **Introduction:** Opener stating that the next set of questions will be about the tax system in Germany and in particular about whether the German income-tax system is complicated or easy to understand. The opener also includes a general statement that the degree of complexity particularly depends on the number of possible tax expenditures. This latter statement thus explains to participants which dimension of complexity we are particularly interested in. The opener also makes all respondents, independent of treatment status, aware of the topic and ensures that the topic is made equally salient to all respondents.

- **Q1: Difficulty of filing a tax return:** We ask participants how difficult they find it to file their tax return. We use this question to derive a proxy for the perceived difficulty of the tax system and to investigate whether other questions and treatment responses depend on the degree of perceived tax complexity.

- **Randomized Experiment 1:** Participants are randomly assigned to three groups that are exposed to different information and arguments in the context of tax simplification. See below for more info.

- **Q2: Preferences for tax simplification:** We ask participants if they think whether the income-tax system in Germany generally needs to be simplified. This question elicits preferences for tax simplification and also is a potential outcome variable in the analysis of the effects of the randomized information.

[34] These are the translated survey questions. The original German questions are available upon request from the authors and will also be available on the GIP website (`https://reforms.uni-mannheim.de/internet_panel/Questionnaires/`).

- **Q3: General need for tax reform:** The question surveys if participants in general think that the German tax system is in need of reform.

- **Q4: Distributional implications of tax expenditures:** The question measures participants' beliefs about the distributional effects of tax expenditures. We particularly survey whether someone beliefs that tax expenditures contribute to a fairer distribution of income or if they tend to benefit high-income taxpayers.

- **Q5: Which type of tax-simplifying reform:** We offer different types of tax reforms that potentially contribute to simplification of the tax system. Participants are asked to indicate which type of reform they prefer under the assumption that all the listed reforms are revenue neutral.

- **Q6-Q8: Should the tax system account for differences in living situations?:** In each of these three questions, we present to the survey respondents the living situations of two fictitious taxpayers A and B (see Weinzierl 2014, Saez and Stantcheva 2016 and Weinzierl 2017 for similar survey question techniques). Respondents are told that A and B earn the same gross income and are very similar in all other (tax relevant) means, but only differ in one particular dimension. We have three different scenarios of varying living situations, and for each scenario we ask participants if A and B should pay the same amount of taxes or if any of the two should pay more. In the three presented scenarios, A and B differ with respect to i) the amount that has to be paid for elderly care of a poor mother, ii) the amount that is given to charity, and iii) the distance they commute to work. The three scenarios are presented in random order to avoid any order effects.

  The choice of these three type of tax expenditures is motivated by their economic importance and real-world prevalence. Tax expenditures for commuting, charitable giving and elderly care are substantial and large in size, with commuting expenditures being the largest income related expense recognized by tax authorities in Germany; in 2017, tax expenditures related to commuting alone amounted to about 5 billion Euro (WiWo 2017). Moreover, these items are used by many taxpayers; about 35 percent of the taxpayers use the commuting expenditure, about 24 percent of the taxpayers deduct charitable contributions from their tax base, and approximately 8 percent of the taxpayers use the elderly care deductions (own calculations based on German administrative tax records (FAST); see FAST 2010).[35] Overall,

[35]The Factually Anonymous Income Tax Statistic (FAST) is a 10% stratified random sample of the German Income Tax Statistics, comprising information about taxable income, family situation, income sources, granted deductions and exemptions, revenues and sources of revenues, income tax burden, etc. The data are available as cross-section scientific use files. For the tax-expenditure calculations here we use the most recent available year of 2010. See Boenke and Schroeder (2017) for more information.

the tax expenditures that we chose to rely on in these questions are very important, salient and likely to be familiar to most of our survey respondents.

In addition, these three tax expenditures represent three different rationales for tax expenditures, namely i) circumstances that are outside of the control of the respective taxpayer (elderly care), ii) positive externalities associated with the (self chosen) expenditure (charitable giving) and iii) items representing job-related choices of taxpayers (commuting).

- **Randomized Experiment 2:** Participants are again randomly assigned to three groups that see different information and arguments in the context of tax simplification (renewed randomization). See below for more info.
- **Q9: Preferences for tax simplification:** We again elicit preferences for tax simplification (as in Q2). We explain to participants that we ask the same question again because the topic was subject of some of the previous questions and participants might have developed a different view on it in the course of the survey questions. The question primarily serves as an outcome variable for the second set of experiments.
- **Q10: Own use of tax expenditures:** We survey which tax expenditures participants usually make use of in their annual income tax declaration.

3.3.3 Randomized Survey Experiments

We include two randomized components into our survey block on tax simplification (see the survey structure above). The two experiments are preceded by separate randomization processes. In both experiments, respondents are randomly assigned to either a control group or one of two treatment groups (i.e. between-subjects design with three groups). An alternative to having two separate experiments would have been one single experiment with more treatment groups. However, in light of the number of participants and the rather subtle experimental interventions, we decided to choose to include two randomized components in order to maximize statistical power.

Both experiments are structured in the same way: i) We first have a short opener that serves as a connecting passage to the subsequently provided information. The opener again explains that tax expenditures potentially contribute to the complexity of the tax system. All participants (control group and treatment groups) see this opener. The opener therefore ensures that the issues of tax expenditures and complexity are made equally aware to control-group participants and treatment-group participants. Any treatment effects are therefor not driven by differences in the extent of topic awareness across the groups. ii) After the opener, respondents in the two treatment groups are provided short information/arguments in the context of tax simplification. iii) Respondents in all three groups move on to the next survey question.

The information that we provide in the treatments pick up some of the most frequently debated issues and empirical findings in the context of tax simplification; the treatments are all reflected in the cited literature and presented arguments in section 3.1 and 3.2.1. In each of the two experiments, one treatment aims to shift preferences towards tax simplification and the other one aims to shift preferences away from tax simplification. Our treatments thus reflect that there are arguments both against and in support of tax simplification. In addition, we are able to investigate if preferences for tax simplification are more elastic with respect to arguments in support or against tax simplification.[36] We describe and motivate the two experiments in the following.

Experiment 1. The first experiment includes two treatments which we label *Redistribution* treatment and *Avoidance* treatment. The two treatments are preceded by an opener that is shown to all respondents (i.e., both treatment groups and control group). The opener is everything that control-group respondents see in the context of the first experiment before they move to the next survey question. It ensures that respondents in all treatment groups are made equally aware of the topic of tax complexity. The opener reads as follows:

> *In Germany there is an ongoing debate on whether the income tax system is too complicated because of many possible deductions and allowances.*

The *Redistribution* treatment highlights that tax expenditures, which add to tax complexity, potentially have redistributional effects and can be used to reduce the tax liability of taxpayers who are disadvantaged by circumstances. The treatment addresses the point that tax expenditures can serve as a social-policy measure and presents a potential argument in support of a certain degree of complexity. Those respondents who are initially not aware of the link between tax expenditures and social-policy aspects might reconsider their tax-simplicity preferences in response to the treatment and become less supportive of tax simplification. The treatment text follows directly after the opener and reads as follows:

> *However, it is sometimes also argued that a tax system with many possible deductions and allowances has an important social-policy role, particularly in relation to income redistribution. For example, tax deductions can be used to reduce the tax burden of taxpayers who are disadvantaged by circumstances.*

[36]The treatment structure is *not* augmented, meaning that respondents in the second treatment group do not see both the information in the first treatment group and the information from the second treatment group, but only see the information from the second treatment group.

The *Avoidance* treatment highlights the frequently debated point that the existence of many complexity-adding tax expenditures potentially facilitates tax avoidance and tax evasion. Assuming that most people disapprove tax avoidance and evasion, respondents who were initially not fully aware of the potential link between tax complexity and avoidance/evasion might shift their preferences towards more simplification in response to being exposed to this treatment. The treatment text follows directly after the opener and reads as follows:

> *In this context, one argument is that a tax system with many possible deductions and allowances offers greater opportunity for tax avoidance and tax adjustment. For example, when individuals have a better knowledge of the tax system or make unjustified declarations, they can reduce their tax burden by taking advantage of certain allowances or deductions.*

Experiment 2. The second experiment includes two treatments, labeled *Efficiency* treatment and *Special interest* treatment. The two treatments are preceded by an opener that is shown to all respondents (i.e., both treatment groups and control group). As before in the first experiment, this opener is everything that control-group respondents see in the context of this second experiment. It ensures that respondents in all treatment groups are equally aware of the general topic, tax complexity. The opener reads as follows:

> *We would like to once again address the ongoing debate concerning whether the income tax system is too complicated due to the many possible deductions and allowances.*

The *Efficiency* treatment highlights the argument that efficiency is potentially higher in a complex system with many tax expenditures because such a system provides the opportunity to tailor taxes to individual situations and, thus, to tax individual capacity and ability. The treatment therefore increases awareness for a potential argument against tax simplification, and potentially shifts preferences away from tax simplification – at least among those respondents who did not consider this argument initially. The treatment is presented immediately after the opener and reads as follows:

> *One argument that is often used against tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances provides better opportunities to tax individuals in accordance with their ability to pay and is therefore economically more efficient.*

The *Special Interest* treatment highlights that a complex system with many tax expenditures is more vulnerable to the lobbying activities of special interest groups. The argument is that special interest groups try to bargain favorable tax treatment and the existence of many complexity-adding exemptions facilitates the groups' efforts; a system with a narrow tax base and without tax expenditures would make it more to difficult to implement special interests in the tax system. Provided that most people agree that special interests should not be accounted for in the tax system, this second treatment provides an argument in support of tax simplification. The treatment text, that follows right after the opener, is formulated as follows:

> *One argument that is often used in favor of tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances offers special interest groups greater opportunity for obtaining exemptions for their clientele.*

Discussion of Experimenter Demand Effects and Power Analyses. One frequently raised concern with survey experiments (and surveys in general) is that experimenter demand effects drive the survey responses and results. A recent study by Mummolo and Peterson (2019) presents a large and carefully conducted test of experimenter demand effects in the context of survey experiments. They run online survey experiments with more than 12,000 participants and randomly assign information about experimenter intent. They find that providing these information does not affect treatment effects; even financial incentives to respond in line with experimenters' intent did not trigger any demand effects. These findings provide evidence that survey experiments are on average robust to experimenter demand effects. The findings are consistent with the results of de Quidt et al. (2018). They use a similar approach in online experiments and find that experimenter demand effects are 'small'.

The main survey question in our paper, preferences for tax simplicity, asks respondents for their view on a specific aspect of policy. There is neither a correct or false answer to this question nor is it in anyhow ethically critical. Participants are therefore not under the impression that they must provide a particular answer and social-desirability bias should thus not matter here. In addition, the question is very similar to the questions that GIP-participants are used to. The information that treatment participants receive prior to replying to the tax-simplicity preferences are provided in a very neutral and objective way, and thus do not induce subjects to provide a certain answer. Overall, the intention behind our survey question and experimental interventions was certainly considerably more subtle than in the above described two studies in the literature. In light of the findings in the literature and the nature of our questions, we argue that experimenter demand is not a critical concern in our survey experiment.

Note that performing ex-ante power analyses during the design stage of our survey experiment was very difficult. The main survey question, preferences for tax simplicity, has neither been included in the GIP before nor are we aware of any other survey that includes a similar question. It was therefore not possible to rely on any reliable predictions regarding the standard deviation (and mean) for our main survey question at the point of time when we designed the survey experiment. In light of a lack of comparable studies, we could neither form any good expectations regarding the effect sizes that would occur from our treatment interventions. However, these parameters are of course crucial for a meaningful power analysis. In addition, we faced a given number of participants in the GIP and it would have been difficult to adjust the sample size in response to the results of an ex-ante power analysis. We therefore do not present the results of any ex-ante power analyses.

3.3.4 Sample Characteristics

Table 4 provides descriptive statistics with respect to the demographics of our survey participants.[37] Most demographics in Table 4 were not surveyed in the context of our specific survey block on tax simplification, but in other parts of the same wave or in other waves (some variables that do not change over time are linked to the current survey wave through the panel character of the GIP). The descriptive results for the questions of our survey block are not in this table, but are instead presented further below in the results section 3.4.

The table shows that we have a survey sample with balanced gender composition (48% female) and that we cover all age categories (with 36% of the participants being older than 58, and 23% retired). 61% of the respondents are married. 17% of the sample participants live in single households, 46% in 2-person households and 18% in households with three people. The distribution of education levels is also very reasonable. We split participants in different income categories and see that 11% are quite poor (net monthly income less than 1500 EUR) and 15% are relatively rich (net monthly income greater than 4500 EUR). The share of people in the three income classes in between poor and rich are quite balanced.

Corresponding with low current unemployment rates in Germany, only about 2% of the survey participants are unemployed. In terms of political affiliations, we see that about 38% of the sample are in the rather conservative political spectrum and 47% are rather left-wing. 8% indicate that they do not have any partisan preferences (left-right preferences are elicited on a 11-point scale from right to left, where we classify 'conservative' as <= 5 on this scale).

[37]Note that the GIP is designed to be representative of the German population, which is why it is not necessary to compare the summary statistics of our sample with statistics from other representative data, such as census data (see Blom et al. 2015 for more details on the GIP's representativeness).

3.3.5 Balancedness across Experimental Groups

Table 12 presents the results of balancing checks for our first experiment. Following the strategy in Alesina et al. (2018), we test balance across groups as follows: For each covariate, we run three OLS regressions of the form $y_i = \beta Covariate_i + \epsilon_i$, where $Covariate$ is the respective covariate that we test. The three dependent variables for which we run the regressions are dummies indicating the treatment groups – redistribution, avoidance, and control group. As a result of this procedure, we have the results of 30 OLS regressions (one regression for each combination of 10 covariates and 3 outcome dummies). Reassuringly, we find strong evidence that randomization worked well and our covariates do not predict treatment status. Out of 60 estimated coefficients, only 5 are significant at the 10% level and only one is statistically significant at the 5% and 1% level, respectively. This is well in line with these coefficients being significant by chance within their margin of error.[38]

The equivalent strategy was used for testing balancedness of the second experiment; see Table 13 which is structured just as the corresponding table for the first experiment (Table 12). We here restrict the sample to respondents who were in the control group of the first experiment (because we are interested in the effect of the second experiment for this 'unencumbered' group; see 3.4.4 for a more detailed explanation). The results are again quite reassuring. Out of 60 coefficients, 7 are significant at conventional levels of significane (10% or lower). Overall, randomization apparently worked out well, which is not surprising given that the GIP computer system assigned respondents randomly to treatment groups and selection into groups was not possible. Further below in our regressions, we present specifications that condition on all observable covariates to mitigate all remaining concerns regarding balancedness.

3.4 Results

This section presents the results of our survey on tax complexity. We proceed as follows. First, we document in subsection 3.4.1 the preferences for tax simplification (Q2) and then investigate the 'anatomy' of these preferences (i.e., which observable characteristics are correlated with the preferences). Second, in subsection 3.4.2, we report the results for further survey questions on the topic of tax complexity. In particular, we document the results for our survey questions w.r.t. difficulty of filing a return (Q1), general need for tax reform (Q3), distributional aspects of simplification (Q4), and which tax expenditures the participants use themselves (Q10). We then go and explore in steps 3 and 4 of our analysis to which extent preferences for tax simplification are affected by awareness w.r.t. the pros and cons of tax simplification; in particular: Third, we document in

[38] With 60 estimated coefficients, one would expect six coefficients with a significance level of 10% even in the absence of any systematic difefrences between groups.

subsection 3.4.3 if participants believe that the tax system should account for differences in the living situations of taxpayers (Q6-Q8). Fourth, we present the results of the two randomized survey experiments in subsection 3.4.4. Therein, we are particularly interested in the experimental effects on preferences for tax simplification. Fifth, subsection 3.4.7 documents which type of tax-simplifying reform the survey respondents favor (Q5).

3.4.1 Preferences for Tax Simplification

Preferences Across all Respondents. We elicit preferences for tax simplicity using a question which surveys whether people believe that the income-tax system in Germany needs to be simplified (Q2 in the survey structure above). The reply categories were on a 6-point scale from 1 'Absolutely not' to 6 'Absolutely'.

Figure 7 presents the share of respondents in each reply category across all survey participants. A large majority believes that the tax system needs to be simplified: Among all question respondents (i.e., those who gave a non-missing and non-*I-don't-know* reply), 53% (= 46.9/(100 − 11.1)) checked reply category six, meaning that the system 'absolutely' needs to be simplified. Another 23% of the respondents chose the second highest reply category 5. This then implies that about 76% of the respondents have strong (either category 5 or 6) preferences for tax simplification. 16% are in category 4, which also implies a preference for tax simplification. Overall, more than 90% of the respondents thus express a preference for a simplified tax system. Only about 8% of the respondents chose categories 1, 2 or 3, which indicate rather weak preferences for tax simplification. The mean response across all respondents is 5.16.

The replies of respondents who are in one our information treatments might be affected by the treatment information. However, the support for tax simplification is also very high among respondents in the control group who did not receive any information; the mean reply in the control group was 5.22.

Overall, the results provide clear evidence that preferences for tax simplicity in the German population are very strong. We are able to confirm that the prevailing view indeed is in strong favor of tax simplification. As a matter of fact, the strong preferences for simplifications suggest that the support of simplification is the obvious choice for participants as they fill out this survey question. We investigate further below if this choice becomes less obvious as we increase awareness w.r.t. the trade-offs behind tax complexity and tax simplification.

Anatomy of Preferences for Tax Simplification. In a next step, we study the 'anatomy' of simplification preferences and investigate which groups (in terms of observable characteristics) are more likely to have strong preferences for tax simplification. For this purpose, we simply regress (using OLS) our measure of simplification preferences on a wide set of observable characteristics. These characteristics comprise demographic factors, including gender, age, marital status, household size, employment status, retirement status and education, as well as a measure of perceived tax difficulty, household income and political preferences. We report results with robust standard errors. The coefficients in this regression are (conditional) correlations and should not be given a causal interpretation. However, they nevertheless shed light on the heterogeneity of preferences and allow us to gain a more nuanced picture.

The results for this anatomy analysis are presented in Table 5. Important demographic correlates of simplification preferences are age and gender. Older people tend to have stronger preferences for simplification, and women have weaker preferences. Age and gender differences are further investigated in Figures 19 and 20 which illustrate unconditional differences between age groups and between men and women, respectively. Figure 19 shows that the average support for simplification steadily increases in age; the support is 16% higher among respondents older than 58, relative to respondents younger than 29. As illustrated in Figure 20, the support for simplification among men is roughly 3% greater than among women.

Another important correlate of simplification preferences is the perceived difficulty of filing a tax return. Respondents who find it easy to file a tax return have lower simplification preferences than respondents who find it difficult. The unconditional relationship between simplification preferences and perceived filing difficulty are displayed in Figure 21. The Figure confirms the intuitive result that the perceived difficulty to declare income taxes is positively associated with support for tax simplification.

Interestingly, respondents who do not file their tax return themselves or employ a tax preparer are more supportive of simplification than those who file themselves and find it easy. These non-filers, however, have lower support for simplification than self-filers who find it difficult to prepare the tax return.

3.4.2 Further Survey Questions

We surveyed further aspects in the context of tax complexity in order to learn more how respondents think about the topic. We present the results question by question in the following.

Difficulty of Filing a Return. We survey the perceived difficulty of filing a tax return on a 5-point scale from 1 'Very Easy' to 5 'Very Difficult' (Q1). This question particularly allows us to investigate if the substantially high preferences for tax simplification correspond with the perceived difficulty to file a return.[39] The results for this survey

[39]Recall that the question of perceived difficulty of filing a tax return was asked before treatment information were presented.

question are summarized in Figure 8, which, again, presents the share of respondents in each reply category.

The right part of the Figure shows that more than 1/3 (34.7%) of the respondents did not indicate their perceived difficulty, either because they do not file a return or because they have their return filed by someone else (e.g., a tax preparer, spouse). Among all respondents who file a return (i.e., the other 61.9%), about 18% (11.1/61.9) find it 'very difficult' to file their return and 32% (19.6/61.9) checked category 4, which also indicates a fair degree of difficulty. The medium category 3 was checked by about 28% (17.5/61.9). About 22% ((10.8 + 2.9)/61.9) of the respondents find it rather easy to file the return. The mean reply for this question is 3.41 (on a scale of 1-5) among all respondents who file a tax return.

Overall, there is a tendency that tax returns are perceived to be fairly difficult, but the picture is not as strong as in the case of preferences for tax simplification. This corresponds with the 'anatomy' result above: the positive correlation between simplification preferences and perceived difficulty to file a return is not perfectly linear. This suggests that the strong preference for tax simplification is not entirely motivated by own experiences with too-difficult tax returns.

General Need for Tax Reforms. We also surveyed if participants believe that the German tax system generally needs to be reformed (Q3). The question was asked on a scale from 1 'Absolutely not' to 6 'Absolutely'. Figure 9 shows that a large fraction of 46.7% (= 39.4/(100 − 15.6)) of question respondents (i.e., those who gave a non-missing and non-*I-don't-know* reply) think that the tax system 'absolutely' (reply category 6) needs to be reformed. Another 27.7% also have a strong preference for reforming the system (reply category 5). A negligible share of people do not see a need to implement reforms; only about 8% of the respondents checked reply categories 1, 2 or 3. The mean reply for this question is 5.06. The mean response in the control group is similar (5.09) to the overall mean. Overall, this part of the survey provides clear evidence that Germans believe that the tax system in their country is in strong need of reform.

Distributional Implications of Tax Expenditures. One frequently raised concern in the context of tax complexity is that the rich are able to exploit tax expenditures better than low-income taxpayers; for example, because they afford professional tax advisors or because they have income sources with more possibilities for tax planning. However, given that many tax expenditures also have a redistributive purpose, it is interesting to survey the public opinion in this context. We therefore survey beliefs about the distributional implications of tax expenditures. In particular, we ask if deductions and allowances contribute to equality or if high-income taxpayers tend to benefit from them (Q4). The reply categories are 1 'Equality' to 6 'High Incomes Benefit' and the question results are summarized in Figure 10.

The results are unambiguous: the majority of question respondents (i.e., those who gave a non-missing and non-*I-don't-know* reply) believe that allowances and deductions benefit high-income taxpayers, rather than contributing to equality. 33.1% (= 27.3/(100−17.5)) were in the corner category 6 and another 20.3% are in the second-largest reply category 5. That is, more than 50 percent of the respondents were in those two categories which indicate the highest beliefs that high-income taxpayers benefit from deductions and allowances. Only about 29% of all respondents checked reply categories 1, 2 or 3. The mean reply for this question is 4.33. This mean response for the overall sample is very similar to the mean response in the control group (4.28). Overall, the large majority beliefs that tax expenditures mostly benefit richer taxpayers.

Which Tax Expenditures are Used? We also survey which type of tax expenditures respondents use regularly (Q10). This survey question mainly serves the purpose of evaluating if survey answers about particular type of expenditures (see below section 3.4.3) are driven by self-interest. The question, however, is also interesting in itself and we therefore briefly summarize the results in Figure 11. The Figure presents the share of people who use particular tax expenditures (note that multiple answers were possible so the shares do not add up to 100). The list of itemized deductions is, of course, not exhaustive. The most frequently used tax expenditures in our sample are the commuting-to-work allowance, the deduction of other type of work expenses (e.g., work-related costs for books, clothes, etc.), charitable donations, and deductible expenses for pension and retirement savings. Child allowances and so-called 'standard deductions'[40] are also quite frequently used.

3.4.3 Different Tax Burden for Taxpayers in Different Living Situations?

In the next set of questions (Q6-Q8), we face respondents with two fictitious taxpayers who differ in one aspect of their living situation, and then ask if these two taxpayers should pay the same amount of taxes. While a differential tax treatment of the two fictitious taxpayers would add complexity to the tax system, we do not mention this complexity aspect of the presented scenario explicitly to respondents. These questions allow us to evaluate if respondents prefer to account for different living situations through the tax system at the cost of adding complexity to the tax system. In other words, if people indicated that specific differences in living situations should matter for the tax burden, this would imply that they do not desire the simplest possible tax system and are willing to accept a certain degree of tax complexity.

[40]This represents the lump sum deduction amount for taxpayers who do not exceed the thresholds in other deduction categories.

All three questions are structured in the same way. We ask respondents to imagine two fictitious taxpayers, A and B, who are comparable in all tax relevant aspects, and only differ along one of the following dimensions:

i) Person A has to spend a considerable amount for the elderly care of her mother, while Person B does not have to bear such costs.

ii) Person A spends a considerable amount of income on charitable giving, while Person B does not donate.

iii) Person A has to travel a considerable distance to work, while Person B lives close to work.

We then ask who of the two persons, A or B, should pay more taxes (where the order of reply categories and the order of presented scenarios was randomized). The results are presented in Figures 12 to 14.[41]

Figure 12 shows that a majority of about 60% of the survey respondents believe that the costs for elderly care should reduce the tax burden. In other words, a majority of about approximately 2/3 of respondents think that the tax system should account for this difference in living situation, and that Person B should pay more taxes. Almost 40% indicate that taxpayers with and without costs for elderly care should pay the same amount of taxes, and almost nobody thinks that A should pay more in taxes.

Figure 15 shows that these effects are not driven by self-interest. We split the sample into those who make use of deductions for care costs themselves and those who do not. The survey responses among these two groups look very similar. Even among those who do not use care deductions themselves, a majority of almost 60% believes that Person B, who does not have care costs, should pay more taxes. Among those who use the deduction themselves, a little bit more than 60% think that Person B should pay more.[42] Overall, these survey responses provide clear evidence that people favor a system in which the costs for elderly care are deductible from the tax base.

For the survey questions regarding charitable donations (Figure 13) and expenses for commuting (Figure 14) we see that a majority of about 66% and 59% of the respondents think that both persons, A and B, should pay the same amounts of taxes, respectively. That is, about 2/3 think that differences in charitable donations and commuting expenses should not imply differential tax payments.

However, a fraction of 32% and 39% of all respondents yet think that higher donations and commuting costs should imply lower tax burdens. That is, roughly 1/3 of the

[41] Note that the responses here were not affected by the randomized interventions.

[42] We acknowledge that it might be possible that a few taxpayers, who do not currently use this tax expenditure, expect to use it in the future. A support of this tax expenditure might then be driven by self-interest, even if they do not currently make use of the tax expenditure.

respondents believe that differential expenses in these areas should result in a reduced tax burden. This is a considerably smaller share than in the case of elderly care, but 1/3 of respondents still is a substantial fraction that is in favor of accounting for these living situations in the tax system.

The difference between the 'elderly care' situation and the other two situations is interesting: Costs for elderly care are circumstantial and outside the control of the respective taxpayer, while donations and commuting distance are choices of the taxpayer. Consistent with the literature on optimal taxation and equality of opportunity (e.g., Alesina and Angeletos 2005; Durante et al. 2014; Ooghe and Peichl 2015), our survey respondents have the intuition that circumstantial differences should be accounted for to a larger extent by the tax system than deliberate and self-chosen differences.

In line with this, the responses for donations and commuting expenses are more affected by self-interest, as compared to the responses for elderly care. Figures 16 and 17 present the results separately for those who use the respective tax expenditure themselves and those who do not. In the case of donations, a quarter of those respondents who do not use the donation expenditure themselves think that donations should reduce the tax burden, while the share is 45% among those who do use the donation expenditure. The pattern is similar for the case of commuting expenditures: among those who do not use the commuting expenditure, 34% believe that it should reduce in a lower tax burden. Among those who do use the commuting expenditure, the share of people who believe commuting should reduce tax payments stands at 47%.

Overall, the differences between those who use the respective tax expenditure and those who do not are thus larger in the case of donations and commuting than for the case of elderly care. However, even for donations and commuting we still see that a large share of those who do not use the expenditure support the notion that the tax system should account for the respective living situation. This suggests that the result for none of the three different tax expenditures is entirely driven by self-interest.

3.4.4 Randomized Survey Experiments

Empirical Strategy. We now present the results of the two randomized survey experiments. In case of the first experiment, we use OLS regressions (with robust standard errors) in which we regress the respective outcome variable on dummy variables indicating the two information treatments. The resulting coefficients then present the effect of the respective treatment relative to the omitted control group. In our preferred specification, we include control variables to improve precision of the treatment effects.

In case of the second experiment, we expect that the treatments of the first experiment impact the treatment effects of the second experiment. For example, consider a respondent who was assigned to the con-simplification treatment in the first experiment and to the pro-simplification treatment in the second experiment. A positive effect of the pro-argument in the second experiment might then cancel out with the negative effect of the con-argument of the first experiment and, as a result, we see no effect in the second experiment, although there actually is a positive effect. We circumvent this concern as follows: We first fully interact dummies indicating treatment status of the second experiment with dummies indicating treatment status of the first experiment (the control group always being the reference category), and then use OLS (with robust standard errors) to regress the outcome variable of interest on the full set of interactions. We only report the coefficients of the treatment dummies of the second experiment (and not the interactions). These reported coefficients then present the effects of the second experiment for those respondents who were in the control group of the first experiment. These respondents have not received any prior treatment in the context of simplification and therefore are 'unencumbered' when they enter the second experiment.[43] As with the first experiment, our preferred specifications include control variables which improve precision of the treatment coefficient of interest.

The main outcome variables are the responses to the question of whether the tax system should be simplified; i.e., Q2 in the case of the first experiment and Q9 in case of the second experiment. These are the variables that follow immediately after the respective randomized intervention. The variable that we use in the regressions is coded just as the original survey question, on a six-point scale, in order to not throw away any information. In the context of the first experiment, we further study the treatment effects on the survey question regarding the perceived distributional effects of expenditures (Q4). This variable is also coded as the original survey variable (on a 6-point scale). We also looked at the effects of the experimental intervention on the question about the general need to reform the tax system (Q3). However, we did not detect any effects for this question and therefore do not report the results.

We use OLS for reasons of eased interpretation. Ordered probit models, which account for the discrete and ordered nature of the outcome variables, are presented in robustness checks.

3.4.5 Experiment 1

Main Effects. Table 6 presents the main results for the first experiment, in which we provide information about the social-policy role of tax expenditures (*Redistribution* group) and about expenditure-induced tax avoidance opportunities (*Avoidance* group) in

[43]The coefficients that we report for the second experiment are identical to coefficients that are estimated in regressions in which the sample is restricted to respondents who were in the control group in the first experiment. We use the full interaction model, and not the sample-split variant, because this approach improves precision and the resulting coefficients are based on the same sample that is used for the regressions for the first experiment.

complex tax systems. Preferences for tax simplicity (Q2) is the outcome variable in all specifications of the table.

Column (1) of the table shows the effects of the treatment dummies in a regression specification without conditioning on any additional covariates. The other columns gradually add further variables in order to increase efficiency and test the sensitivity with respect to covariates. Column (2) adds several demographic control variables, and Columns (3), (4) and (5) additionally condition on the perceived difficulty to file a tax return, household income and political preferences, respectively.

The regression results in all specifications show negative effects of the *Redistribution* treatment on preferences for tax simplification (all estimates statistically significant at the 5% level). The coefficients are remarkably stable across the five different specifications. In Column (5), our preferred specification where we include all covariates, the support for tax simplification is reduced by about 2.6% ($-0.133/5.22$ = coefficient/control-group average), relative to the control-group average. The regressions thus provide evidence that preferences for tax simplicity are elastic to information *against* tax simplification.

The effect size is not very large, but it has to be considered in light of the fact that the overall support for tax simplification is substantial and, given the debate in the public and press outlets, presumably is strongly anchored among respondents. Our treatment thus affects preferences for tax simplification *although* the conventional wisdom on the topic is very clear and strong. For these reasons, we argue, the effect size should be interpreted as non-negligible.

The *Avoidance*-treatment does not have a significant effect on simplification preferences. The coefficients are small and not statistically significant throughout the five specifications. The standard errors in all five specifications are considerably greater than the respective coefficient. Statistical precision is thus much weaker than in the case of the *Redistribution*-treatment. The coefficients are also considerably different: across all specifications, the coefficients of the *Redistribution*-treatment are at least 2.7 times larger than the coefficients of the *Avoidance*-treatment, and the difference between the two is statistically significant throughout specifications (3) to (5) (with p-values in the range of 0.064 to 0.057).[44] We also tried different specifications of the outcome variable (e.g., a dummy variable indicating very high support for simplification) but never find a significant effect of the *Avoidance*-treatment.

The results thus show that preferences for simplification are not elastic to the information *in support of* tax simplification. This null-result might be explained with the very prominent role of arguments in favor of simplification in the public debate. As a result of these salient arguments, participants presumably have less misperceptions regarding in-

[44]Negative coefficients of the *Avoidance*-treatment are consistent with the treatment having no effect. If two independent samples are drawn from the same population, it is very likely that one sample is smaller than the other one.

formation that support simplification. The high initial support for tax simplicity among participants (which does not leave much room for even more support) might also play a role.[45]

We also investigated the effect of the first experimental intervention on respondents' views about the distributional implications of tax complexity (Q4). The results of this exercise are presented in Table 7, which is structured like the previously discussed Table 6. The *Redistribution*-treatment does not affect these beliefs (relative to the control group). The coefficients are close to zero and not significantly different from zero. However, the information about possible complexity-induced avoidance possibilities in the *Avoidance*-group somewhat affect the distributional beliefs. The treatment coefficient is statistically significant in specifications (3) to (5), and indicates that the treatment increases beliefs that tax expenditures add to income inequality (the coefficients in specifications (1) and (2) are imprecisely measured). Considering the specification in column (5), which includes all covariates, the treatment increased the distributional-beliefs variable by about 4% (0.169/4.285), relative to the control-group average. Comparing the coefficients of the *Redistribution*-treatment and the *Avoidance*-treatment, we find statistical significant differences for specifications (3) to (5) with p-values ranging from 0.09 to 0.064.

Results for the experimental effects on both tax simplification attitudes and distributional views are robust to using Ordered Probit regressions that account for the discrete nature of the outcome variables; the respective results are shown in Tables 14 and 15.

Randomization Tests and Multiple Hypothesis Testing. In a next step, we investigate if the (robust) OLS standard errors that we reported above are robust to other ways of computing standard errors. In particular, we adjust standard errors using i) randomization tests in the spirit of Fisher (1935) and ii) tests for multiple comparisons that follow the procedure proposed by Westfall and Young (1993). Note that the coefficients are not affected by the alternative types of statistical inference that we present in the following.

First, we perform randomization tests following Young (2018). The Young (2018)-procedure performs exact tests which test the sharp null hypothesis that the effect of the information treatment is zero for all individuals receiving our treatment. That is, it does not test whether the average treatment effect is zero (which is what we tested in our main analysis), but whether the treatment effects are zero across all repondents. The randomization-test procedure, which is in the spirit of Fisher (1935), is more conservative in computing standard errors: Young (2018) reports that, using his approach, the number

[45]Note that the constant decreases as we subsequently add control variables across the regression specifications. This suggests that our control variables can explain a considerable part of the high baseline support for simplification.

of significant results of randomized experiments is considerably reduced relative to conventional tests of individual treatment effects. Compared to classical asymptotic-based testing procedures, these randomization tests have the advantage that they are robust against concentrated leverage and do not rely on sample size or the characteristics of the error (Young 2018).[46]

Using the Young (2018)-procedure with 5000 draws (to approximate the p-value of the Fisher distribution), the effect of the *Redistribution*-treatment on preferences for simplification yields a p-value of 0.015 (in our preferred specification with the full set of controls). The coefficients are thus statistically significant and the levels of significance of the classical testing method – as reported above – are confirmed. This stricter procedure for computing p-values also confirms the insignificant effect of our *Avoidance*-treatment on preferences for tax simplicity; the p-value for the *Avoidance*-dummy is computed to be 0.6 and thus far off conventional levels of statistical significance.

Overall, all p-values based on the randomization tests are very similar to the ones obtained by ordinary OLS with robust standard errors. This is reassuring and lends credibility to the inference used in our main analyses above (which used classical hypothesis testing). The similarity between p-values might be interpreted as an indication that the treatment effects in our setting are constant among individuals; as noted by Ding (2017), the sharp-null hypothesis and the null hypothesis of zero average causal effect are equivalent in the case of constant causal effects.

Second, we use the method proposed by Westfall and Young (1993), and for example recently applied by Blattman et al. (2017), to adjust standard errors for multiple comparisons. As Blattman et al. (2017), we take a rather conservative approach that adjusts for comparisons across treatments and outcomes: in our first experiment, the combination of three outcome variables and two treatments implies that six hypothesis are tested (i.e., for each outcome variable, two treatment effects are tested relative to the control group). We tested the effect of our information treatments on the following three outcome variables: preferences for tax simplification (Q2), general need for tax reform (Q3), distributional implications of tax expenditures (Q4). Note that we only reported in detail the results for outcomes Q2 and Q4 because we did not find any effects of the treatment on Q3. However, since we initially intended to study the effect on all three outcome variables, the correct procedure here requires that we adjust standard errors to the case with three outcomes and two treatments.

Using the Westfall and Young (1993)-procedure to adjust standard errors for multiple comparisons, we find a standard error of 0.083 for the effect of the *Redistribution*-treatment on preferences for tax simplification (based on our preferred specification with

[46]We implement the randomization tests using the ado file provided by Alwyn Young on his website; the exact testing procedure is described in (Young 2018). We report randomization-t tests since the author finds in practice "randomization-t to be superior to the -c".

the full set of control variables). The effect of the *Avoidance*-treatment on preferences for tax simplification is insignificant with a p-value of 0.84. We thus confirm the classical p-values regarding the treatment effects on our main outcome variable, preferences for tax simplification. The p-value of *Avoidance* on the perceived distributional implications, which is significant in the classical inference approach, stands at 0.24 with this method and therefore turns insignificant. All other hypotheses are insignificant with p-values greater than 0.7.

Heterogenous Treatment Effects. In a next step of the analysis, we investigate if particular groups of respondents respond differently to the treatments of the first experiment than other groups of respondents. For this purpose, we interact the treatment-group dummies with the observable characteristics of the sample population; in particular we test if there are heterogenous effects with respect to the following variables: age, gender, marital status, household size, income, education, political preferences, difficulty of filing a tax return, trust in government, the perceived quality of tax use for public spending, taste for redistribution (from wave 34 of the GIP), beliefs in luck or effort, and social mobility perceptions (from wave 33 of the GIP). Overall, the effects of the treatments seem to be very homogenous. We mostly do not see any significant interactions. For reasons of brevity and given these results, we do not report the regression results. We acknowledge that it is possible that the interaction models for detecting heterogeneity lack statistical power, rather than providing evidence of homogenous treatment effects. The finding that treatment effects seem to be rather constant across observable characteristics is consistent with the above finding that the classical standard errors and the adjusted standard errors using the Young (2018)-procedure are very similar.

3.4.6 Experiment 2

Main Effects. The main results for our second randomized intervention are presented in Table 8, which is organized as the corresponding table for the first experiment. This second experiment includes a control group, a group that is presented an *Efficiency* argument against tax simplification, and a group that is presented a *Special interest group* argument in favor of tax simplicity. The dependent variable is the question surveying tax-simplicity preferences (note that Q9 is the dependent variable here, not Q2 which we use for the first experiment). Consistent with the results from the first experiment, we observe that preferences for tax simplicity are elastic towards information against tax simplification, and not elastic to information in favor of simplification.

The estimated coefficient for the *Efficiency*-treatment is negative and statistically significant throughout all five specifications of the regression table. Column 5, our preferred specification with all covariates, shows that the efficiency argument reduced support for tax simplicity by about 5% (= 0.240/5.084), relative to the control-group average. The effect size should again be considered in light of the fact that the general wisdom clearly holds that tax simplification is desirable; we therefore consider a 5% effect size to be non-negligible.

The estimates for the effect of the *Special interest group* argument in favor of tax simplicity are very close to zero and non-significant in all of the regression specifications. Notably, the coefficient of the *Efficiency*-treatment is at least three times larger than the coefficient of the *Special interest group*-treatment across the five specifications. However, these differences between the two treatment estimates are not statistically significant, presumably due to power reasons since we only compare reactions for participants who have been in the control group in the first experiment. The null result of the *Special interest group*-treatment is, again, likely to be driven by the more prominent role of arguments in favor of simplification in the public debate, which reduce misperceptions regarding pro-simplification arguments.

As shown in Table 16, these results are robust to using ordered probit models.

Randomization Tests and Multiple Hypothesis Testing. As with the first experiment, we again adjust standard errors using Young (2018)-type randomization tests and Westfall and Young (1993)-type tests for multiple comparisons. Note that we only have one outcome variable (preferences for tax simplicity) and two treatment groups here, implying that we test only two hypotheses in the context of this second experiment.

First, the randomization tests come with a p-value for the effect of the *Efficiency*-treatment on preferences for tax simplification of 0.019 (in our preferred specification with full set of controls). That is, the previously reported significance for the *Efficiency*-treatment is confirmed. The effect of the *Special interest group*-treatment remains insignificant with a p-value above 0.5. As with the first experiment, the p-values are remarkably similar to the p-values from classical testing methods. This is reassuring and again indicates that our treatment effects are constant across participants.

Second, the Westfall and Young (1993) method finds adjusted p-values of 0.043 for the *Efficiency*-treatment and 0.48 for the *Special interest group*-treatment (both in specifications with the full set of control variables). The procedure thus confirms the classical inference procedure that treatment *Efficiency* has a significant effect, while treatment *Special interest group* does not.

Heterogenous Treatment Effects. As in the case of the first experiment, we investigate if particular groups of respondents respond differently to the treatments of the first experiment than other groups. We run the same interaction models as in the case of the first experiment (with the same interacted observable variables) and again find that effects of the treatments are very homogenous across different demographic groups; we mostly do not detect any significant interactions. For reasons of brevity and given these results, we again do not report the regression results. We acknowledge, again, that it is possible that the interaction models for detecting heterogeneity lack statistical power, rather than providing evidence of homogenous effects. However, the lack of heterogeneity is consistent with the finding that exact p-values following Young (2018) are very similar to the classical p-values.

3.4.7 Which Simplifying Tax Reform?

In light of the conventional wisdom that tax simplification is desirable, we included a question to survey *how* policy should reform the tax system in order to make it simpler. For the purpose of this question, we provided respondents with a list of potential tax-simplifying reforms and they could chose which of the offered alternatives they prefer. This list is of course not exhaustive, it yet features some of the most debated type of reforms. Respondents are explicitly informed that they should consider each of the reforms under the assumption of tax-revenue neutrality (i.e., no effects on tax revenues through the respective reforms) in order to abstract from revenue considerations.[47]

The results for this question are summarized in Figure 18. The most frequently chosen type of reform (33%) increases the progressivity of the tax system and abolishes all types of tax expenditures. About 20% of the respondents would prefer a flat-rate system which features the same amount of tax expenditures as in the status quo. 15% of respondents also want a flat-rate system, but without any possibilities for deductions or allowances.

About 18% of the respondents prefer a different type of tax simplification. Instead of reforming rates or the amount of tax expenditures, they prefer to change the tax-filing process through pre-filled tax returns that require less effort to file a return. 6% of our respondents have a preference for keeping the status-quo and do not implement any tax reform.

These results are evidence for heterogeneity in the preferred approach for moving towards a simplified tax system. While the results from the Taxpayer-A-vs-B part of our survey and the randomized experiments show that there is a lack of awareness about the trade-offs behind complexity and simplicity, the results here suggest that, in addition, there is no consensus w.r.t. the tax simplifying reform to be implemented. Both of these empirical observations add to an explanation for the puzzle that real-world tax systems are so complex although the conventional wisdom holds that simplicity is strongly desirable.

We also investigated if the policy-reform preferences are affected by our experimental intervention, i.e. the first experiment as the second experiment was implemented after

[47]See Q5 in Appendix for the detailed question design.

the reform survey question. We do not find any evidence that this is the case. This is somewhat in line with other recent survey experiments finding that policy preferences are often relatively inelastic to information treatments (e.g., Kuziemko et al. 2015; Alesina et al. 2018).

3.5 Conclusion

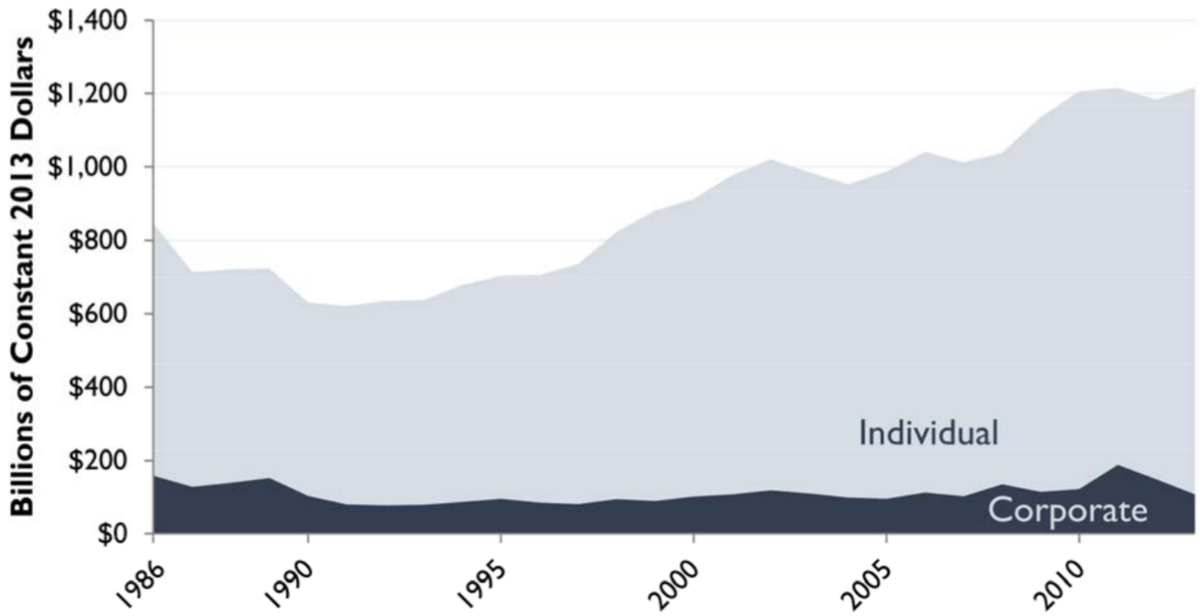
The prevailing view in the academic literature and public debate seems to be that most tax systems are too complex and should be simplified. However, there also are economic arguments in support of a certain degree of tax complexity and it is puzzling why tax systems remain highly complex despite the conventional view in favor of more simplification. Using new experimental and survey data for a representative sample of the German population, we shed light on preferences for tax simplification. We find that most people are indeed in favor of a simpler tax system. However, once we make people aware of the trade-offs between simplicity and complexity, preferences for simplicity are reduced. For example, a large share of respondents believes that the tax system should account for different circumstances in living situations (such as costly care of elderly family members). This suggests that respondents implicitly favor to add complexity to the tax system by allowing to deduct the associated costs from the tax base.

Our survey also includes two randomized experiments in which we make respondents aware of the possible consequences of tax simplification. Both randomized experiments consistently provide evidence that the support for tax simplification is elastic to information *against* tax simplification, while arguments *in support of* simplification do not impact preferences for simplification. The results thus suggest that misperceptions are more relevant in the context of arguments against simplification, which, in turn, is consistent with the observation that arguments in favor of simplicity are more prominent in the debates and that the general wisdom holds that simplicity is desirable. Overall, we show that the high support for simpler taxes is to some extent driven by a lack of awareness about the implications of tax simplification. Individuals apparently frequently express desires for tax simplification without appreciating the implications of tax simplification.

Overall, our findings suggest that the (policy, academic, and public) debate about tax simplification potentially benefits from a more nuanced discussion of the pros and cons of tax simplification. As a result of more nuanced discussion, the matter would not be dominated by a general wisdom view anymore and instead potentially gain objectivity. Recent work shows that information deficits among individuals in the context of taxation can induce governments to implement inefficient tax policy (Boccanfuso and Ferey 2019). More nuanced discussions and better information could also mitigate this source of inefficiency.

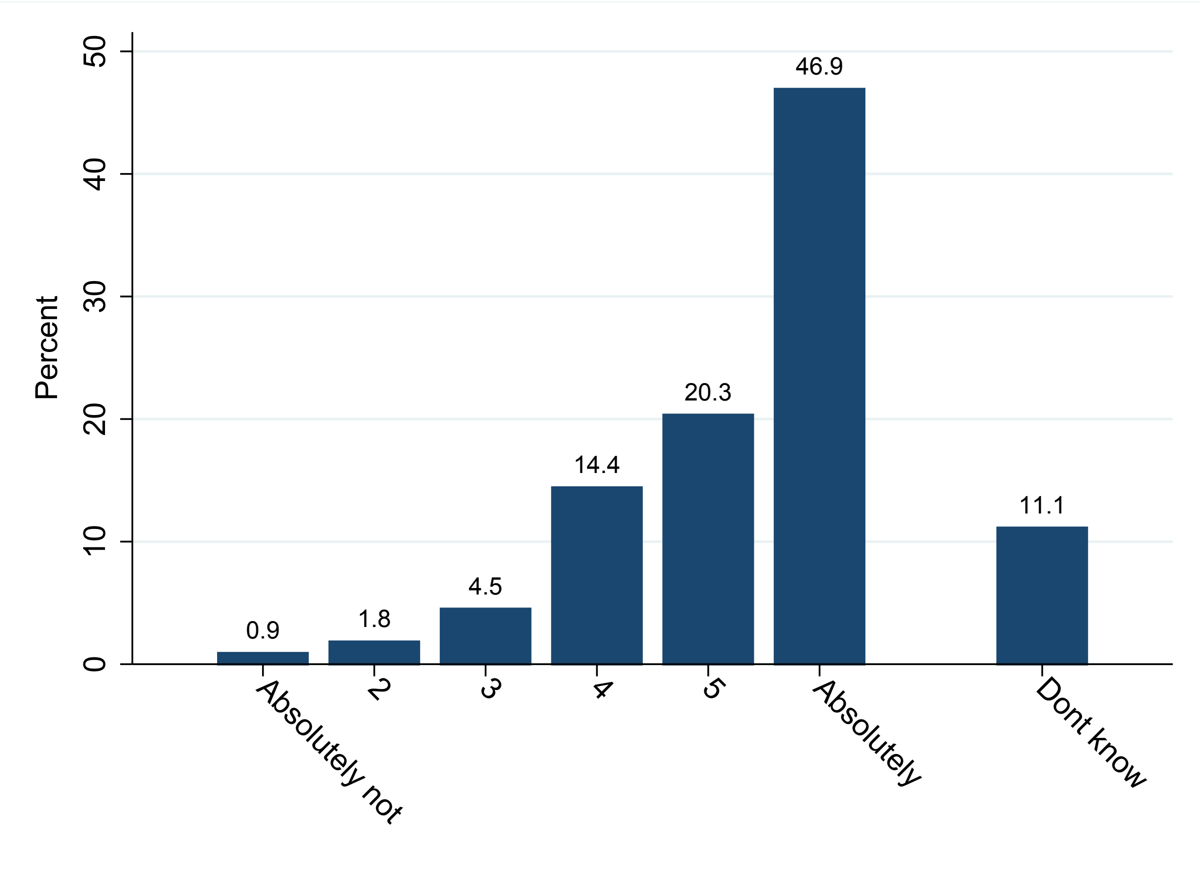
Figures and Tables

Figure 6: Growth of Tax Expenditures over Time in the US



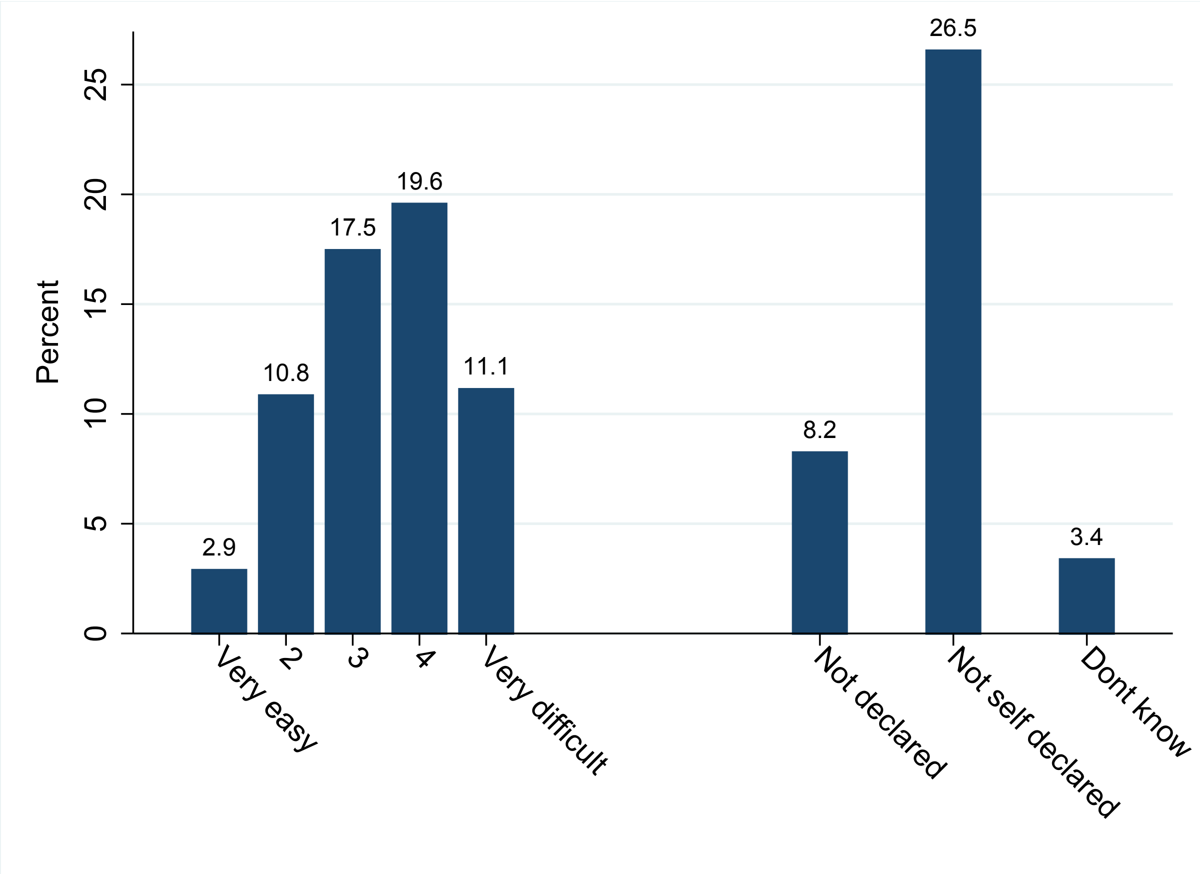
Notes: US Treasury estimates of tax expenditures, 1986-2013, adjusted for inflation to 2013 dollars. Source: Tax Foundation, Fiscal Fact, A Brief History of Tax Expenditures. Available online: `https://files.taxfoundation.org/legacy/docs/ff391.pdf`.

Figure 7: Preferences for Tax Simplification



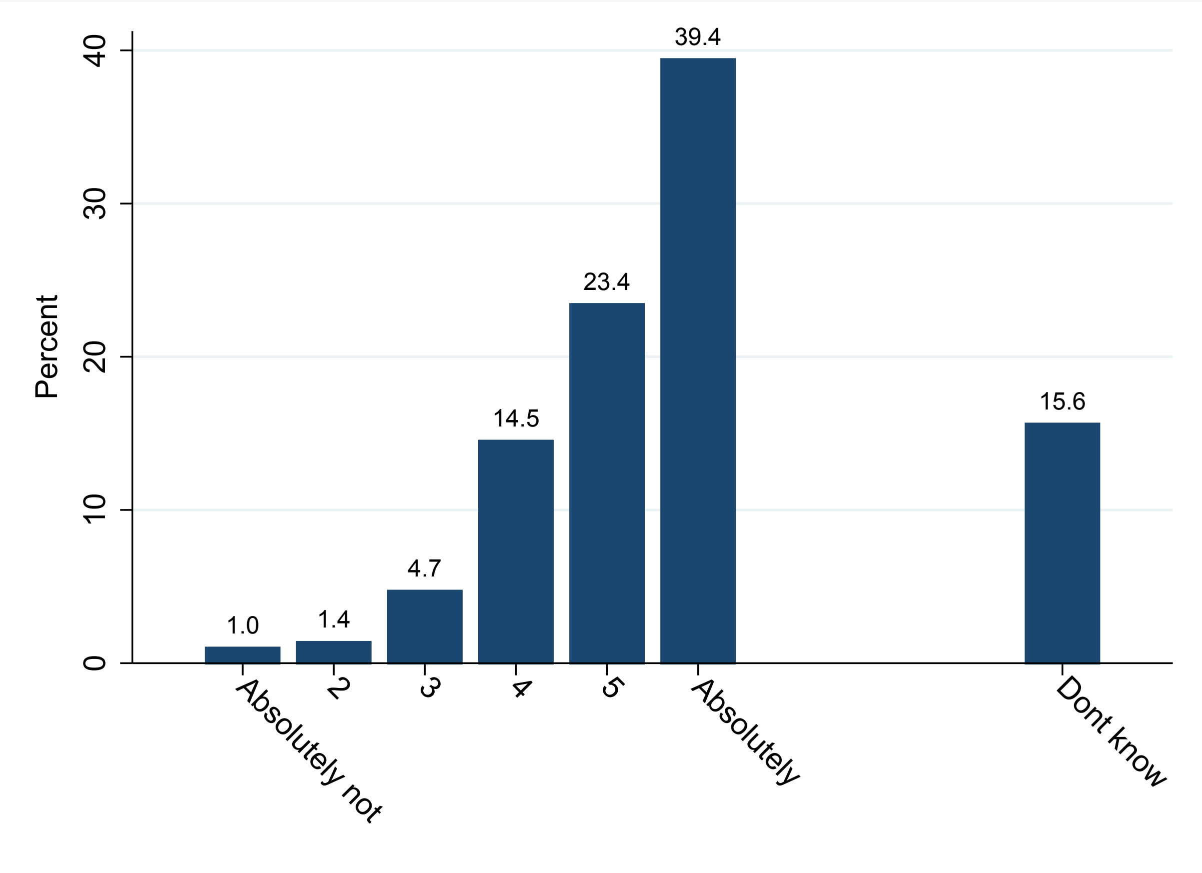
Notes: This figure depicts the percentage share of respondents in the respective categories of the question "Do you generally think that the income tax system in Germany needs to be simplified?" Respondents could pick one of the following categories: 1 Absolutely not; ... ; 6 Absolutely; I do not know. The figure is based on 2,423 non missing observations. The mean answer is 5.16. Source: Own calculations based on German Internet Panel.

Figure 8: Perceived Difficulty of Filing a Tax Return



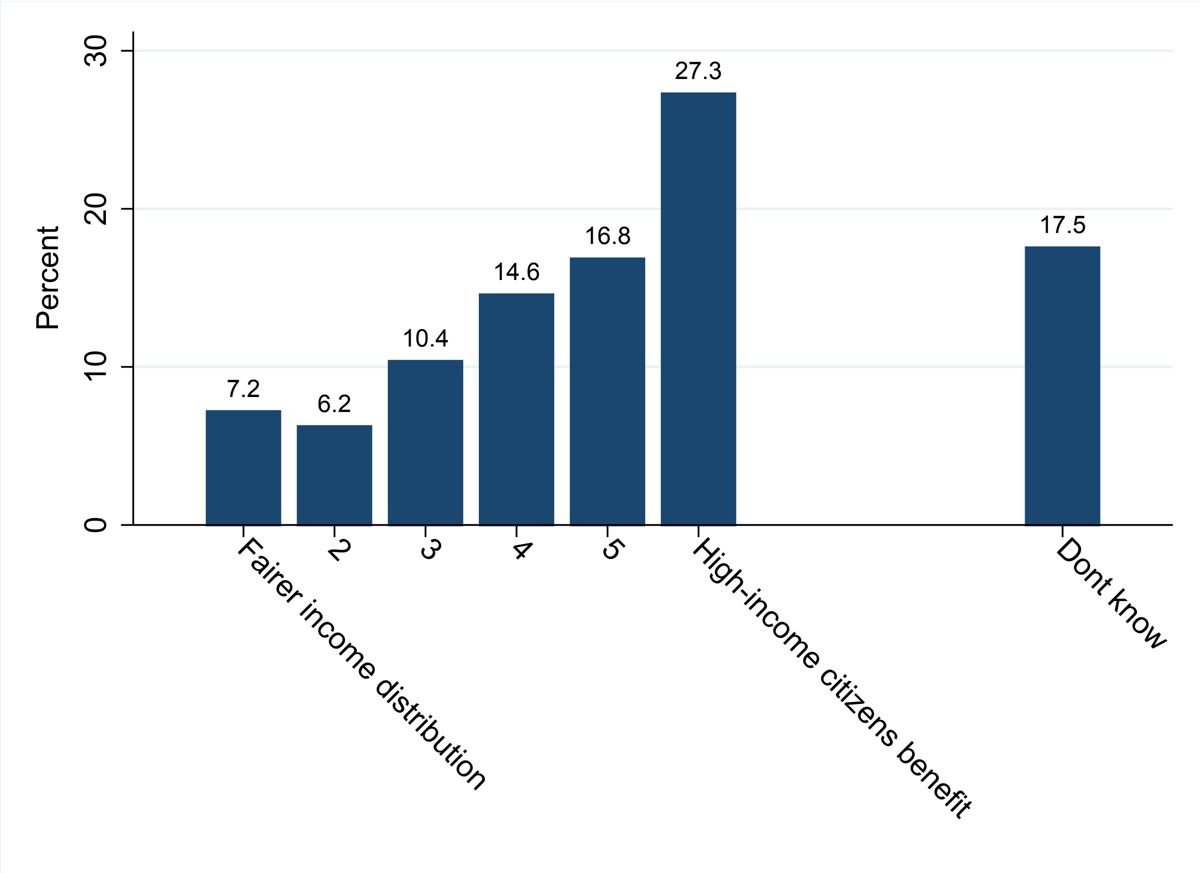
Notes: This figure depicts the percentage share of respondents in the respective categories of the question "How difficult is it for you to fill out your tax return?" Respondents could pick one of the following categories: 1 Very easy; ... ; 5 Very difficult; I do not know because no taxes are declared in my name; I do not know because I do not declare taxes myself (rather, my partner or a tax consultant, etc. does this); I do not know. The figure is based on 2,424 non missing observations. The mean answer for categories 1 to 5 is 3.41. Source: Own calculations based on German Internet Panel.

Figure 9: Need for Tax Reform



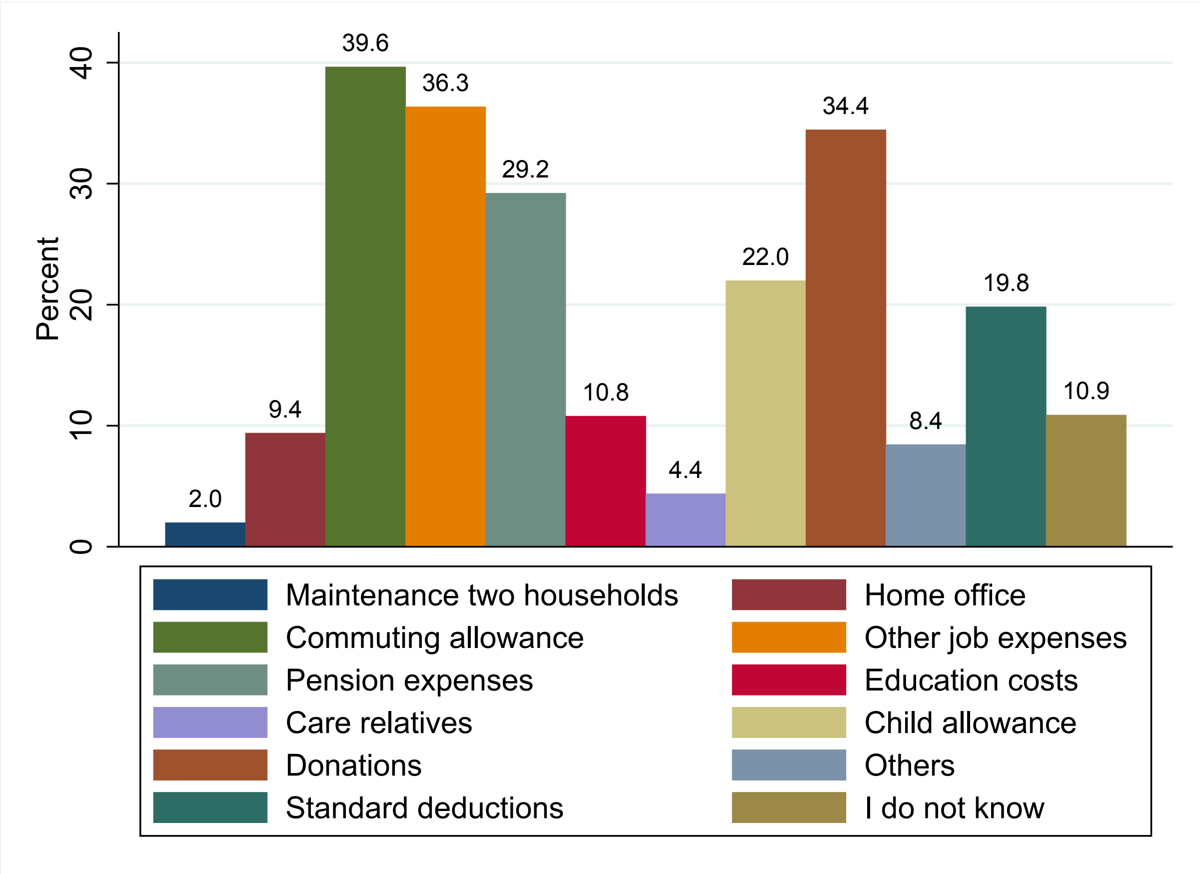
Notes: This figure depicts the percentage share of respondents in the respective categories of the question "Do you generally believe that the income tax system in Germany is in need of reform?" Respondents could pick one of the following categories: 1 Absolutely not; ... ; 6 Absolutely; I do not know. The figure is based on 2,423 non missing observations. The mean answer is 5.06. Source: Own calculations based on German Internet Panel.

Figure 10: Perceived Distributional Implications of Complexity



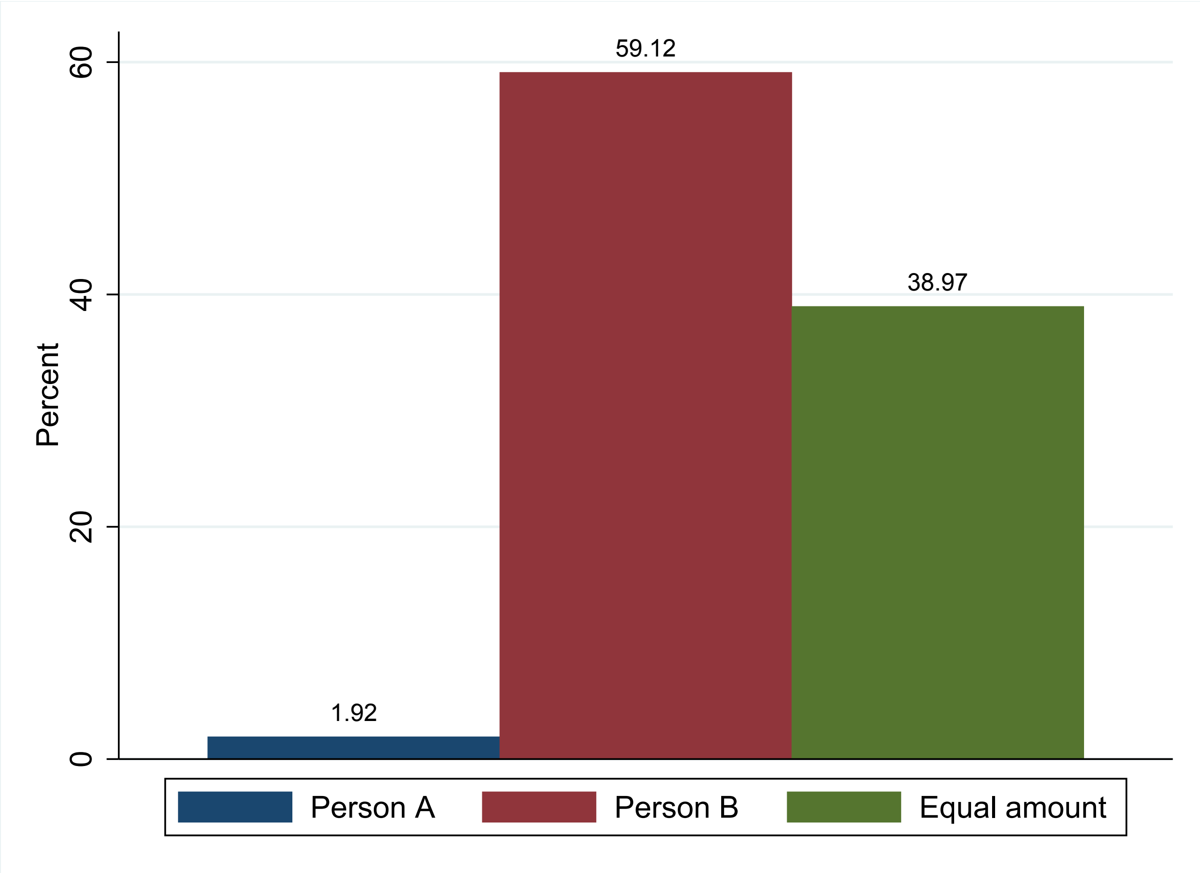
Notes: This figure depicts the percentage share of respondents in the respective categories of the question "Do you think that numerous deductions and allowances contribute to a fairer distribution of income, or do you believe that high-income citizens benefit more from these deductions and allowances?" Respondents could pick one of the following categories: 1 They contribute to fairer income distribution; ... ; 6 High-income citizens benefit; I do not know. The figure is based on 2,423 non missing observations. The mean answer is 4.33. Source: Own calculations based on German Internet Panel.

Figure 11: Which Deductions and Allowances are used?



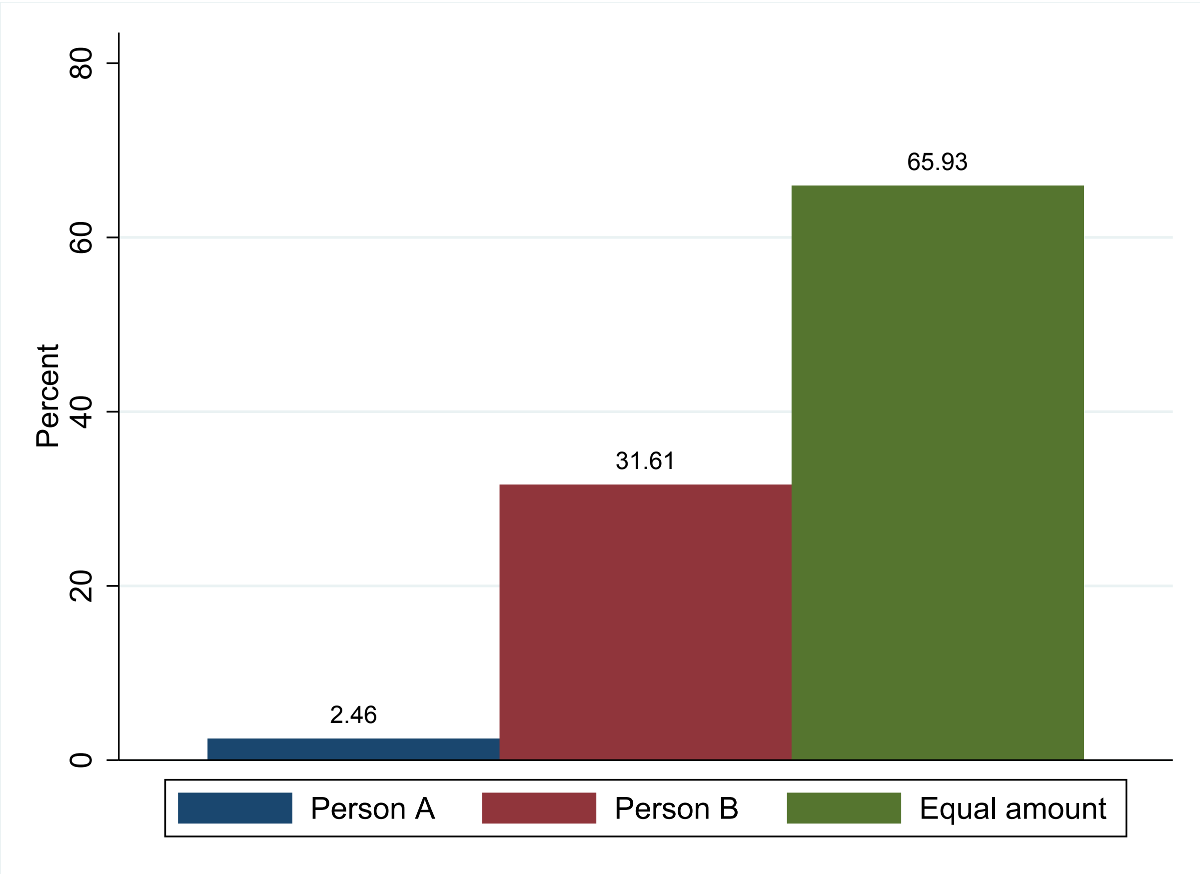
Notes: This figure depicts the percentage share of respondents in the respective categories of the question: "Which of the following deductions and/or allowances do you usually use when filing your income tax?" Respondents could pick one of the following categories: Maintenance of two households; Home office; Commuting allowance; Other job related expenditures; Pension expenses; Education costs; Care relatives; Child allowance, childcare; Donations; Others [insert text]; No deductions; I do not know. The figure is based on 2,215 non missing observations. Note shares do not add up to one because respondents could check multiple items. Source: Own calculations based on German Internet Panel.

Figure 12: Who should pay more taxes? Elderly Care



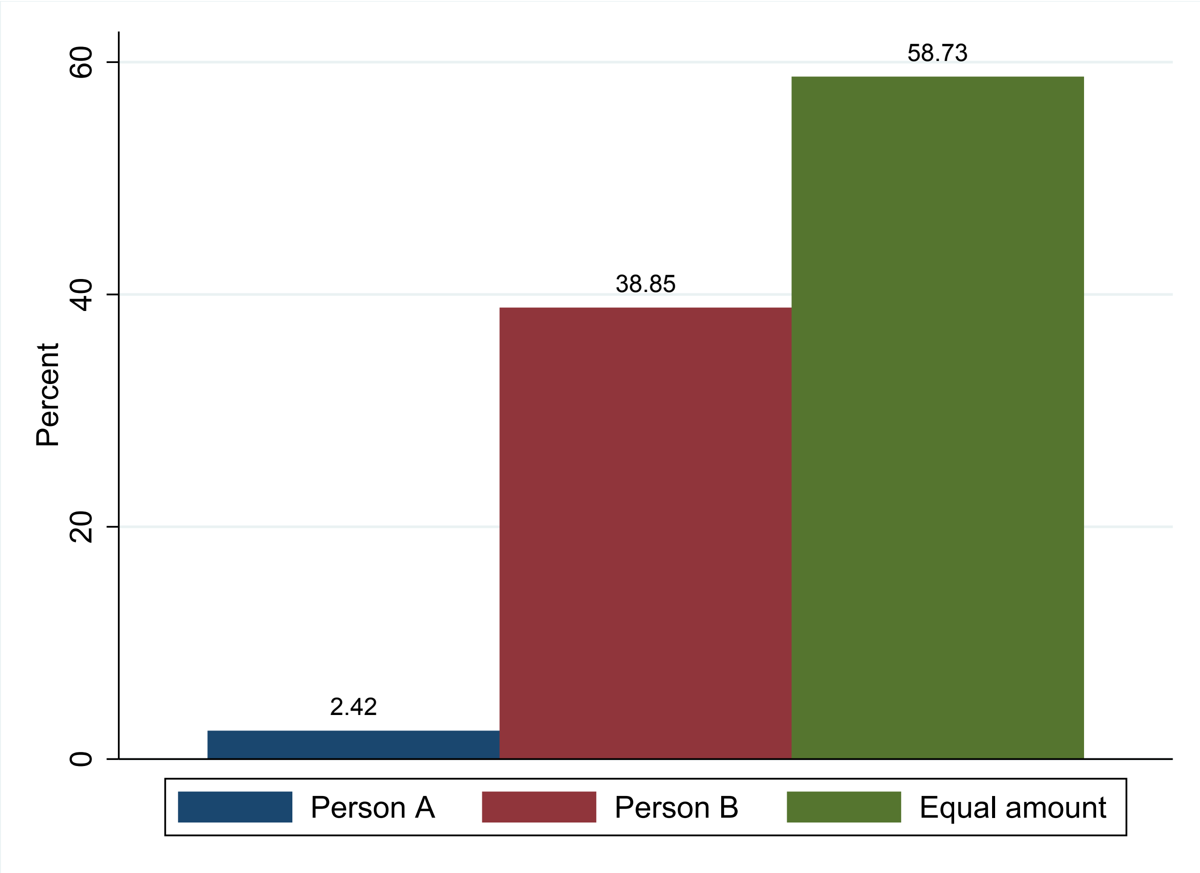
Notes: This figure depicts the percentage share of respondents in the respective categories of the question: "In contrast to Person B, Person A has a poor mother in need of elderly care and has to spend a considerable amount of her income for the care of her mother. Person A and B have the same gross income and are very similar in all other respects." Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The figure is based on 2,397 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 13: Who should pay more taxes? Donations



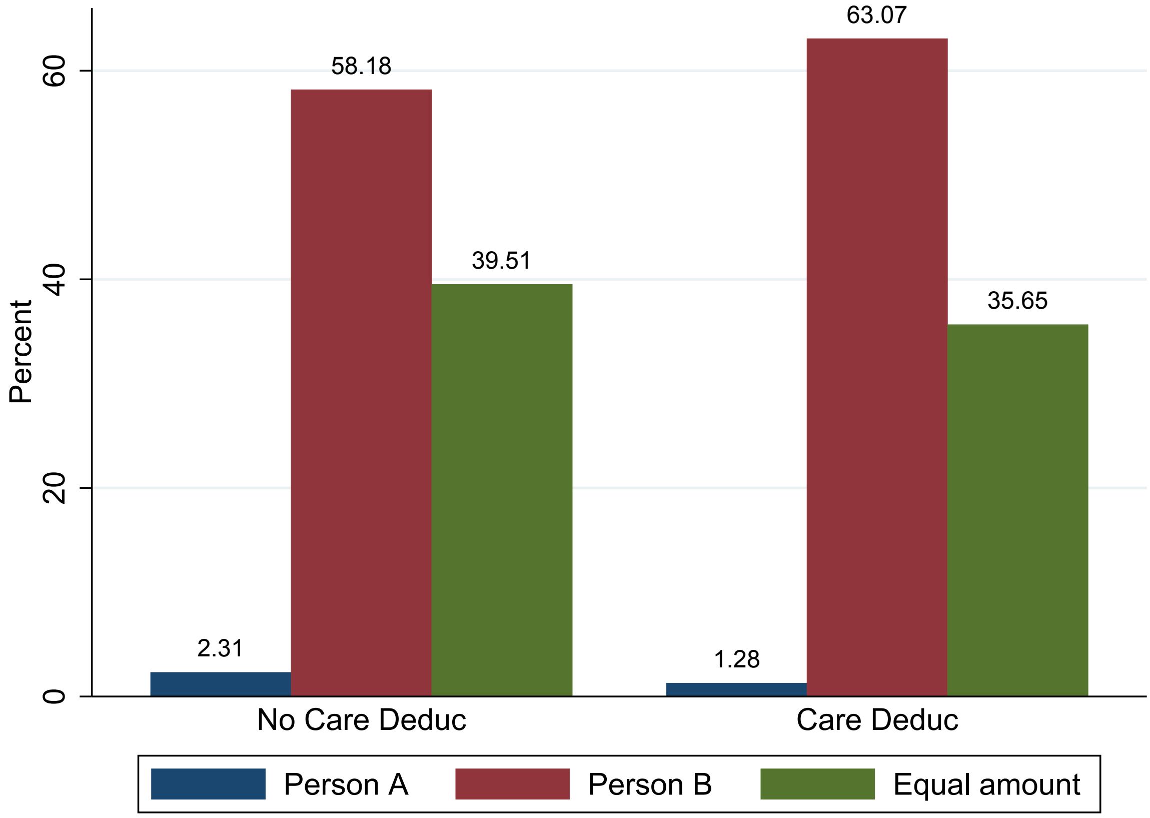
Notes: This figure depicts the percentage share of respondents in the respective categories of the question: "Person A spends a considerable amount of her income on charitable giving. Person B does no such thing. Both Person A and B have the same gross income and are very similar in all other respects." Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The figure is based on 2,398 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 14: Who should pay more taxes? Commuting To Work



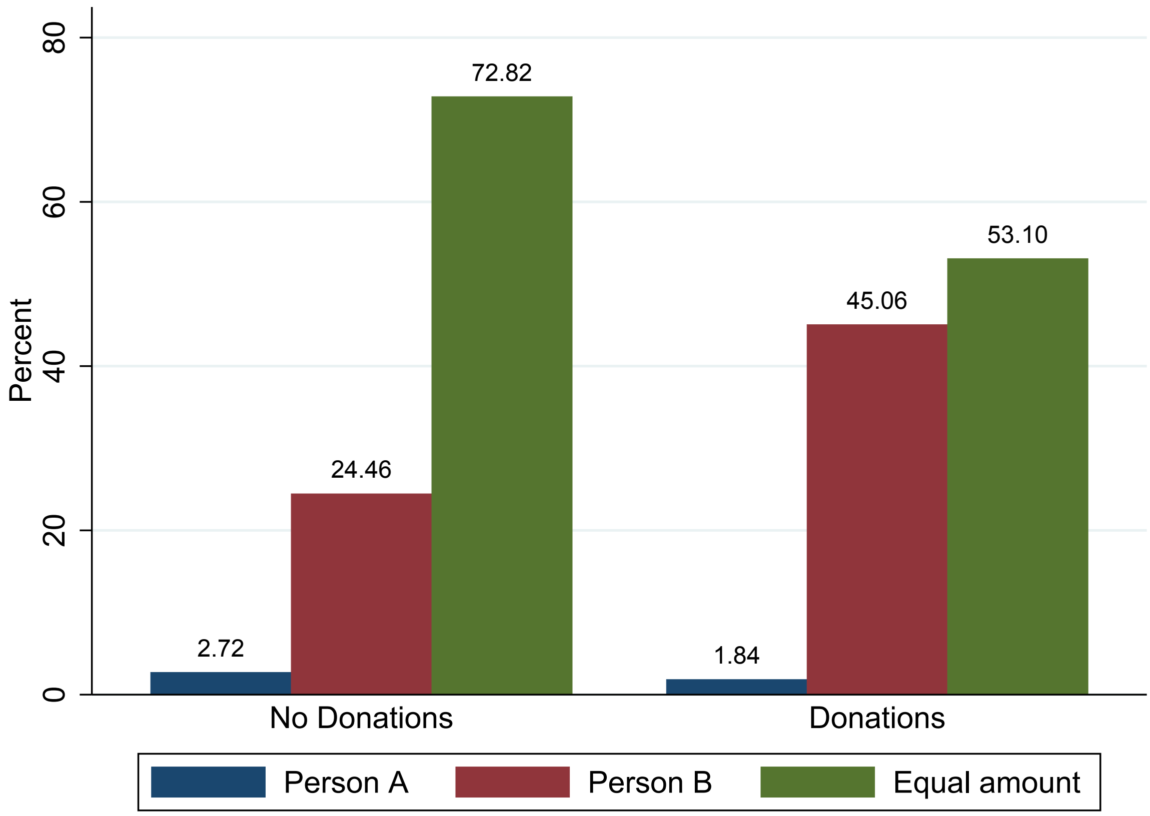
Notes: This figure depicts the percentage share of respondents in the respective categories of the question: "Person A has to travel a considerable distance to work. Person B lives very close to work. Both Person A and B have the same gross income and are very similar in all other respects." Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The figure is based on 2,394 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 15: Driven by self interest? Elderly Care



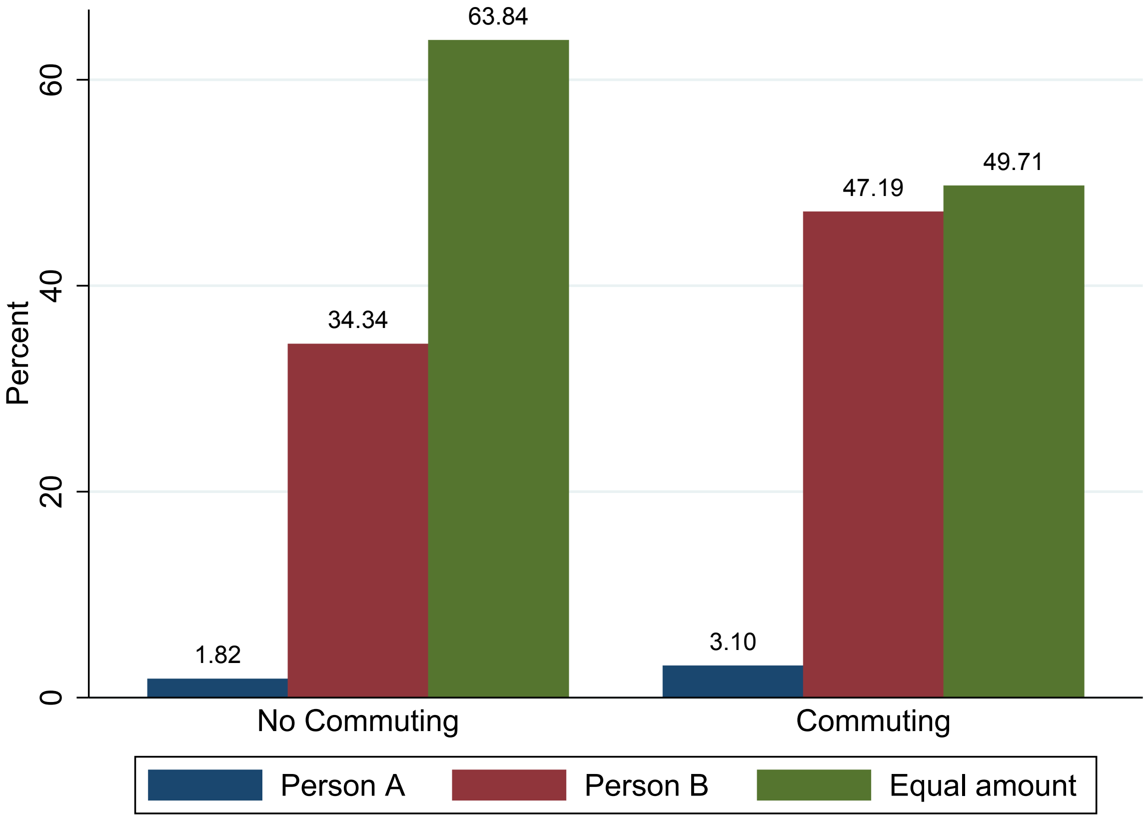
Notes: This figure depicts the percentage share of respondents in the respective categories by respondents claiming care deductions. Since only few people claim deductions for elderly care we consider the broader category of care deductions. Care deductions include deductions for elderly and child care. Question: “In contrast to Person B, Person A has a poor mother in need of elderly care and has to spend a considerable amount of her income for the care of her mother. Person A and B have the same gross income and are very similar in all other respects.” Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The left part shows replies for respondents who do not use the deduction for elderly care. The right part shows replies of respondents who do use the deduction for elderly care. The figure is based on 2,397 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 16: Driven by self interest? Donations



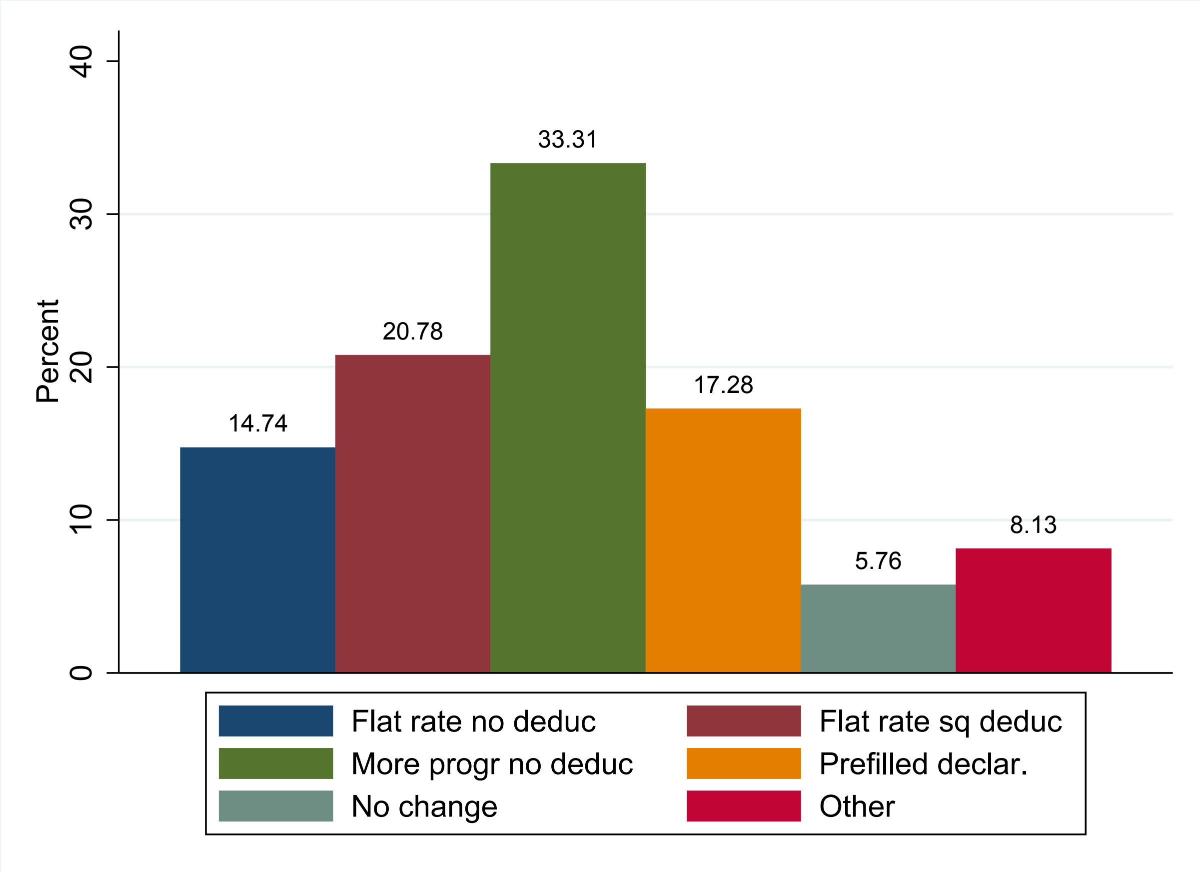
Notes: This figure depicts the percentage share of respondents in the respective categories by respondents claiming the deduction. Question: "Person A spends a considerable amount of her income on charitable giving. Person B does no such thing. Both Person A and B have the same gross income and are very similar in all other respects." Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The left part shows replies for respondents who do not have deductible donations. The right part shows replies of respondents who do have deductible donations. The figure is based on 2,398 non missing observations.Source: Own calculations based on German Internet Panel.

Figure 17: Driven by self interest? Commuting



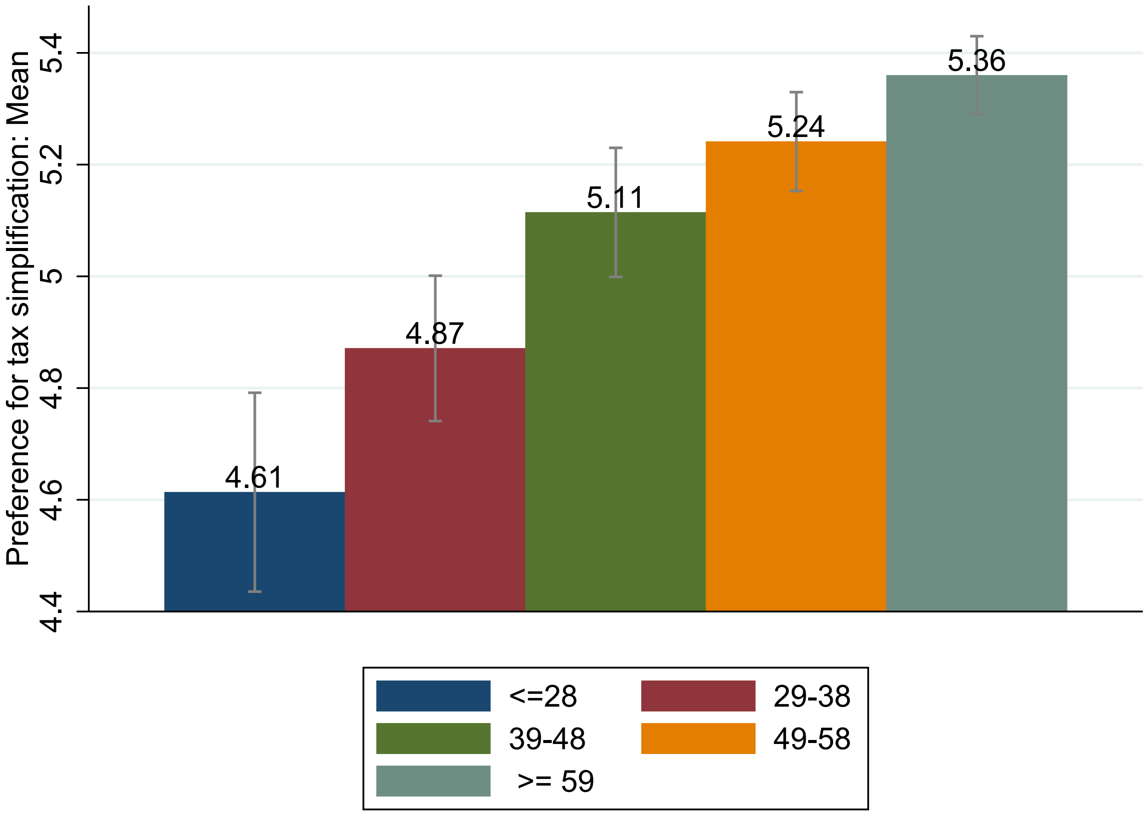
Notes: This figure depicts the percentage share of respondents in the respective categories by respondents claiming the deduction. Question: "Person A has to travel a considerable distance to work. Person B lives very close to work. Both Person A and B have the same gross income and are very similar in all other respects." Respondents could pick one of the following categories (order of answer categories was randomized): Person A should pay higher taxes; Person B should pay higher taxes; Person A and B should pay equal tax amounts. The left part shows replies for respondents who do not use the deduction for commuting to work. The right part shows replies of respondents who do use the deduction for commuting to work. The figure is based on 2,394 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 18: Which Revenue-Neutral Reform?



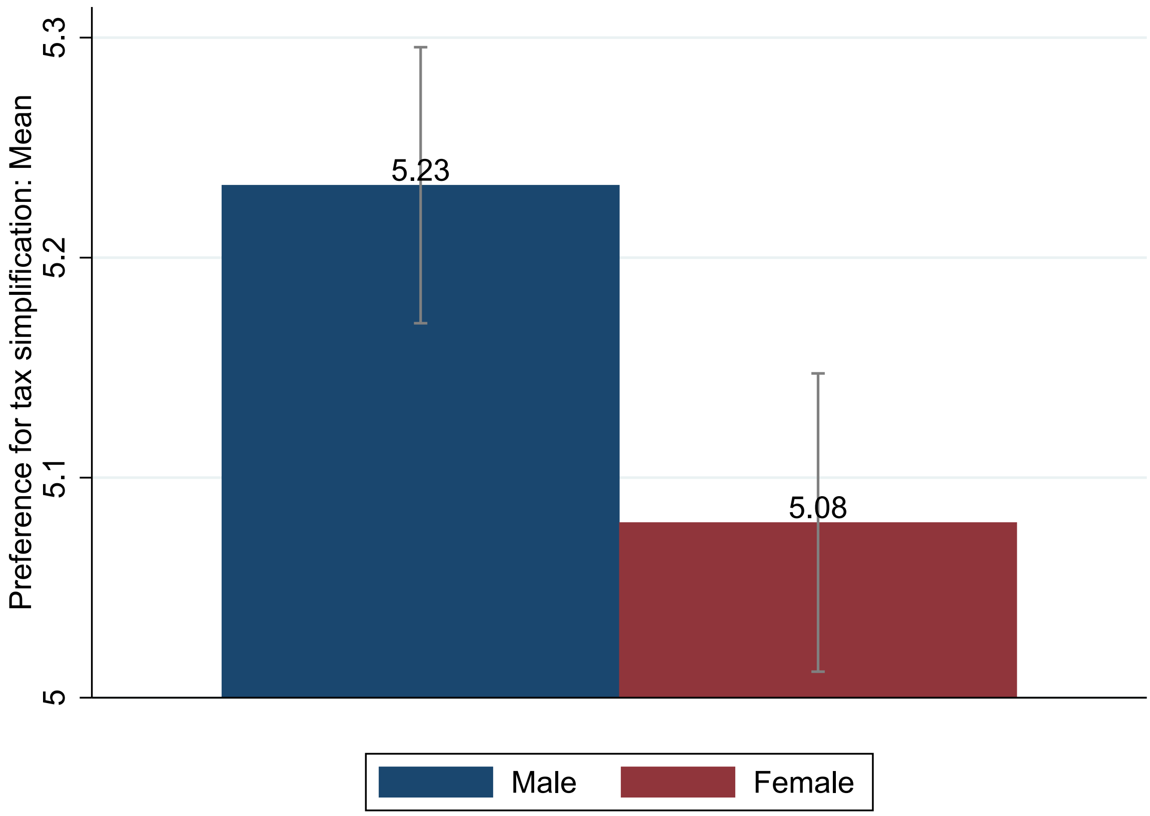
Notes: This figure depicts the percentage share of respondents in the respective categories of the question: "Which of the following measures to simplify the income tax system would you like the most? Assume the proposed measures will lead to unchanged tax revenues in each case." Respondents could pick one of the following categories: Same rate for all but no deductions and allowances; Same rate for all and same deductions and allowances as under current system; More progressive tax rates and no deductions and allowances; Automatic determination of amounts in income tax declaration; No change; Other measure [insert text]; I do not know. The figure is based on 1,771 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 19: Preferences for Tax Simplification by Age Categories



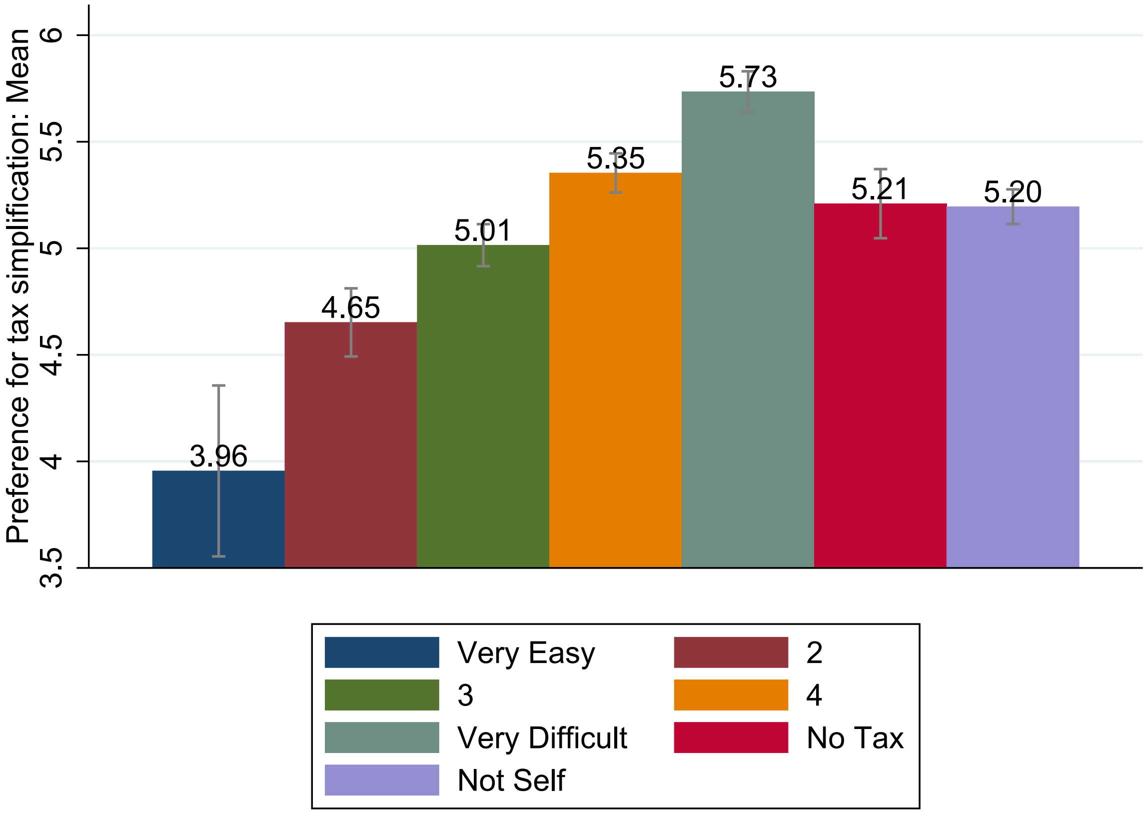
Notes: Average preference for tax simplification by age categories. The outcome variable is the survey-based preference for tax simplification as described in Section 3.3. The figure is based on 2,189 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 20: Preferences for Tax Simplification by Sex



Notes: Average preference for tax simplification by sex. The outcome variable is the survey-based preference for tax simplification as described in Section 3.3. The figure is based on 2,190 non missing observations. Source: Own calculations based on German Internet Panel.

Figure 21: Preferences for Tax Simplification by Perceived Difficulty to File a Return



Notes: Average preference for tax simplification by perceived difficulty. The outcome variable is the survey-based preference for tax simplification as described in Section 3.3. The figure is based on 2,164 non missing observations. Source: Own calculations based on German Internet Panel

Table 4: Descriptive Statistics

	N	Mean	Std.Dev.	Min	Max
Experiment 1					
Redistribution	2424	0.33	0.47	0	1
Avoidance	2424	0.33	0.47	0	1
Control	2424	0.33	0.47	0	1
Experiment 2					
Efficiency	2419	0.33	0.47	0	1
Special interest	2419	0.33	0.47	0	1
Control	2419	0.33	0.47	0	1
Demographics					
Single households	2463	0.17	0.38	0	1
2	2463	0.46	0.50	0	1
3	2463	0.18	0.38	0	1
4	2463	0.14	0.35	0	1
5+	2463	0.05	0.22	0	1
Age <=28	2461	0.09	0.29	0	1
Age 29-38	2461	0.15	0.36	0	1
Age 39-48	2461	0.15	0.36	0	1
Age 49-58	2461	0.24	0.43	0	1
Age >=59	2461	0.36	0.48	0	1
Married	2464	0.61	0.49	0	1
Female	2463	0.48	0.50	0	1
Unemployed	2463	0.02	0.13	0	1
Retired	2463	0.23	0.42	0	1
Low education	2401	0.03	0.17	0	1
Low-med education	2401	0.45	0.50	0	1
High-med education	2401	0.23	0.42	0	1
High education	2401	0.30	0.46	0	1
Difficulty in declaring taxes					
No difficulty	2381	0.03	0.17	0	1
2	2381	0.11	0.32	0	1

3	2381	0.18	0.38	0	1
4	2381	0.20	0.40	0	1
Very difficult	2381	0.12	0.32	0	1
No taxes declared	2381	0.09	0.28	0	1
Not self declared	2381	0.27	0.45	0	1
Household net income					
Poor	2464	0.11	0.32	0	1
2	2464	0.19	0.39	0	1
3	2464	0.20	0.40	0	1
4	2464	0.16	0.37	0	1
Rich	2464	0.15	0.36	0	1
No income stated	2464	0.11	0.32	0	1
Not merged	2464	0.07	0.26	0	1
Political orientation					
Conservatives	2464	0.38	0.48	0	1
Left-wing	2464	0.47	0.50	0	1
Non partisans	2464	0.08	0.27	0	1
Not merged	2464	0.07	0.26	0	1

Notes: The table depicts the summary statistics for all treatment group dummies and all variables in our tailored survey block on tax complexity. Variables are defined as follows: experiment 1 and experiment 2 realizations represent the respective group allocations of respondents in either experiment; household size comprises single households and household with 2, 3, 4 and 5+ members; age categories are ≤28, 29-38, 39-48, 49-55 and ≥59; Married equals 1 if respondent is married, 0 otherwise; Female equals 1 if respondent is female, 0 otherwise; Unemployed equals 1 if respondent is unemployed, 0 otherwise; Retired equals 1 if respondent is retired, 0 otherwise; education categories comprise low (secondary schooling, no job training), low to medium education (upper secondary schooling or finished job training), high to medium education (upper secondary schooling and finished job training) and high eduction (tertiary education); household income variables define net monthly household incomes on a 5-point scale from poor, i.e. 1 (≤1500 Euro), 2 ($1500 \geq x < 2500$ Euro), 3 ($2500 \geq x < 3500$ Euro), 4 ($3500 \geq x < 4500$ Euro) to 5 being rich (≥4500 Euro) as well as a dummy for no answers (No income stated) and a dummy for those observations which had not been in the GIP wave where the income question was asked; conservatives equals ≤5 on a 11-scale left-right placement variable, for > 5 left-wing equals 1. Non partisans did not report a score for the left-rich placement variable and a dummy for those observations which had not been in the GIP Wave where the political preference question was asked. Data comes from German Internet Panel (GIP) wave 36, except for the following items: political preferences derived from variable left-right placement (wave 31) as well as household incomes (wave 31).

Table 5: Anatomy of tax simplification preferences

	(1)	(2)	(3)	(4)	(5)
Household size	0.030	0.021	0.021	0.021	0.021
	(0.026)	(0.024)	(0.024)	(0.024)	(0.024)
Age	0.180***	0.168***	0.168***	0.167***	0.169***
	(0.025)	(0.024)	(0.024)	(0.024)	(0.024)
Married	-0.054	-0.087	-0.092	-0.093	-0.093
	(0.058)	(0.056)	(0.057)	(0.057)	(0.057)
Female	-0.128***	-0.135***	-0.134***	-0.140***	-0.138***
	(0.047)	(0.045)	(0.046)	(0.046)	(0.046)
Unemployed	0.137	-0.046	-0.038	-0.042	-0.034
	(0.199)	(0.201)	(0.204)	(0.204)	(0.203)
Retired	-0.003	0.023	0.025	0.025	0.030
	(0.063)	(0.063)	(0.063)	(0.063)	(0.063)
Education	-0.017	0.006	0.004	0.003	0.005
	(0.027)	(0.026)	(0.027)	(0.027)	(0.027)
Difficulty 2		0.690***	0.686***	0.682***	0.675***
		(0.212)	(0.213)	(0.213)	(0.214)
Difficulty 3		1.033***	1.033***	1.034***	1.028***
		(0.204)	(0.204)	(0.204)	(0.205)
Difficulty 4		1.375***	1.373***	1.370***	1.364***
		(0.203)	(0.203)	(0.203)	(0.204)
Very difficult		1.725***	1.722***	1.721***	1.720***
		(0.204)	(0.204)	(0.204)	(0.206)
No taxes declared		1.251***	1.253***	1.245***	1.242***
		(0.214)	(0.215)	(0.215)	(0.216)
Taxes not self declared		1.209***	1.206***	1.205***	1.203***
		(0.202)	(0.202)	(0.202)	(0.203)
Income gr 2			-0.031	-0.030	-0.034
			(0.100)	(0.100)	(0.100)
Income gr 3			-0.018	-0.016	-0.019
			(0.101)	(0.101)	(0.100)
Income gr 4			0.059	0.061	0.056
			(0.103)	(0.103)	(0.102)
Rich			-0.016	-0.011	-0.014
			(0.110)	(0.111)	(0.110)
No income stated			0.036	0.035	0.032
			(0.109)	(0.109)	(0.109)
Not merged			-0.050	-0.912***	-0.938***
			(0.125)	(0.124)	(0.125)
Conservative				-0.039	- 0.039
				(0.048)	(0.048)
Non partisans				0.028	0.025
				(0.103)	(0.103)
Not merged				0.850***	0.863***
				(0.119)	(0.120)
Redistribution					-0.133**

					(0.055)
Avoidance					-0.029
					(0.053)
Constant	4.566***	3.407***	3.419***	3.406***	3.456***
	(0.145)	(0.239)	(0.246)	(0.246)	(0.249)
N	2132.000	2109.000	2109.000	2109.000	2109.000
R2	0.046	0.146	0.147	0.148	0.150

Notes: The table presents the determinants of Preferences for Tax Simplicity using OLS regressions of preferences for tax simplicity on various covariates. Each column (1)-(5) presents the results of one regression with different sets of covariates. Robust standard errors are in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Variables are defined as follows: experiment 1 and experiment 2 realizations represent the respective group allocations of respondents in either experiment; household size comprises single households and household with 2, 3, 4 and 5+ members; age categories are ≤28, 29-38, 39-48, 49-55 and ≥59; Married equals 1 if respondent is married, 0 otherwise; Female equals 1 if respondent is female, 0 otherwise; Unemployed equals 1 if respondent is unemployed, 0 otherwise; Retired equals 1 if respondent is retired, 0 otherwise; education categories comprise low (secondary schooling, no job training), low to medium education (upper secondary schooling or finished job training), high to medium education (upper secondary schooling and finished job training) and high eduction (tertiary education); household income variables define net monthly household incomes on a 5-point scale from poor, i.e. 1 (≤1500 Euro), 2 (1500≥x<2500 Euro), 3 (2500≥x<3500 Euro), 4 (3500≥x<4500 Euro) to 5 being rich (≥4500 Euro) as well as a dummy for no answers (No income stated) and a dummy for those observations which had not been in the GIP Wave where the income question was asked; conservatives equals 1 if ≤5 on a 11-scale left-right placement variable, for > 5 left-wing equals 1. Non partisans did not report a score for the left-right placement variable and a dummy for those observations which had not been in the GIP Wave where the political preference question was asked. Data comes from German Internet Panel (GIP) wave 36, except for the following items: political preferences derived from variable left-right placement (wave 31) as well as household incomes (wave 31).

Table 6: Exp 1: Effect on Preferences for Tax Simplification

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Redistribution	-0.115**	-0.123**	-0.133**	-0.134**	-0.133**
	(0.058)	(0.058)	(0.055)	(0.055)	(0.055)
Avoidance	-0.042	-0.039	-0.032	-0.032	-0.029
	(0.057)	(0.056)	(0.053)	(0.053)	(0.053)
Constant	5.215***	4.610***	3.453***	3.469***	3.456***
	(0.040)	(0.149)	(0.242)	(0.248)	(0.249)
N	2190	2132	2109	2109	2109
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions on preferences for tax simplification. This is estimated by OLS regressions of preferences for tax simplification on treatment dummies. Tax simplification is measured on a 6 point scale based on the question: "Do you generally think that the income tax system in Germany needs to be simplified?" The experimental groups are: Control group, Redistribution group and Avoidance group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "In Germany there is an ongoing debate on whether the income tax system is too complicated because of many deduction possibilities and allowances." Participants in the Redistribution group receive the following information: "However, it is sometimes also argued that a tax system with many deduction possibilities and allowances has a social-policy and redistributive compensation role. For example, tax deductions can be used to reduce the tax burden of taxpayers who are disadvantaged by circumstances." Participants in the Avoidance group receive the following information: "In this context, one argument is that a tax system with many deduction possibilities and allowances offers more scope for tax avoidance and tax adjustment. For example, tax deductions can be used to reduce one's own tax burden through better knowledge of the tax system or through unjustified specifications in the tax return." Columns (1)-(5) differ in the included sets of covariates. (1): no covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. Robust The scale of the outcome variable is 1 (absolutely not) to 6 (absolutely). Robust standard errors are in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 7: Exp 1: Effect on Perceived Distributional Effects of Complexity

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Redistribution	0.009	0.005	0.001	0.009	0.007
	(0.090)	(0.090)	(0.090)	(0.090)	(0.089)
Avoidance	0.118	0.137	0.157*	0.158*	0.169*
	(0.089)	(0.090)	(0.091)	(0.091)	(0.090)
Constant	4.285***	3.506***	3.136***	3.127***	3.126***
	(0.065)	(0.215)	(0.295)	(0.315)	(0.314)
N	1998	1946	1931	1931	1931
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions on believes about whether people think that deductions work in favor of the rich. This is estimated by OLS regressions of beliefs on treatment dummies. The outcome is measured on a 6 point scale based on the question: 'Do you think that numerous deductions and allowances contribute to a fairer distribution of income, or do you believe that high-income citizens benefit more from these deductions and allowances?" The experimental groups are: Control group, Redistribution group and Avoidance group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "In Germany there is an ongoing debate on whether the income tax system is too complicated because of many possible deductions and allowances." Participants in the Redistribution group receive the following information: "However, it is sometimes also argued that a tax system with many possible deductions and allowances has an important social-policy role, particularly in relation to income redistribution. For example, tax deductions can be used to reduce the tax burden of taxpayers who are disadvantaged by circumstances" Participants in the Avoidance group receive the following information: "In this context, one argument is that a tax system with many possible deductions and allowances offers greater opportunity for tax avoidance . For example, when individuals have a better knowledge of the tax system or make unjustified declarations, they can reduce their tax burden by taking advantage of certain allowances or deductions." Columns (1)-(5) differ in the included sets of covariates. (1): no covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. The scale of the outcome variable is 1 (add to a fair income distribution) to 6 (higher incomes benefit). Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 8: Exp 2: Effect on Preferences for Tax Simplification

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Economic Efficiency	-0.197*	-0.216**	-0.229**	-0.237**	-0.240**
	(0.109)	(0.109)	(0.105)	(0.105)	(0.105)
Special Interest	-0.064	-0.062	-0.055	-0.068	-0.067
	(0.097)	(0.097)	(0.094)	(0.095)	(0.095)
Constant	5.084***	4.588***	3.928***	3.993***	3.960***
	(0.066)	(0.160)	(0.222)	(0.232)	(0.232)
N	2187	2134	2114	2114	2114
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions of the second experiment on preferences for tax simplification. This is estimated by OLS regressions of preferences for tax simplification on treatment dummies and a full set of interactions of the treatment groups of the first and second experiment. Tax simplification is measured on a 6 point scale based on the question: " Now that we have dealt extensively with various aspects of the German tax system in this survey, we would like to ask again: do you generally believe that the income tax system should be simplified in Germany?" The experimental groups are: Control group, Economic efficiency group and Special interest group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "We would like to once again address the ongoing debate concerning whether the income tax system is too complicated due to the many possible deductions and allowances." Participants in the Economic efficiency group receive the following information: "One argument that is often used against tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances provides better opportunities to tax individuals in accordance with their ability to pay and is therefore economically more efficient." Participants in the Special interest group receive the following information: "One argument that is often used against tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances offers special interest groups greater opportunity for obtaining exemptions " Columns (1)-(5) all include a full set of interactions of the treatment groups of the first and second experiment, they differ in the additionally included sets of covariates. (1): no additional covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. The scale of the outcome variable is 1 (absolutely not) to 6 (absolutely). Robust standard errors are in parentheses *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

4 How do taxes affect the trading behavior of individual investors? Evidence from individual portfolio data

4.1 Introduction

Many aspects of the trading behavior of individual investors are well documented in the literature (see Barber and Odean 2013 for an extensive overview of the behavior of individual investors). One aspect of individual trading behavior which is less understood concerns the causal effect of taxes on trading behavior. Realized capital gains are subject to taxes in most countries around the world and it is therefore important to have a proper understanding of their effects on investment behavior. In theory, realization-based taxes on capital gains induce investors to defer the realization of gains (lock-in effect) and to realize losses as they accrue (because losses can be used to offset taxable gains).[48] However, it has been suggested that such effects of taxes on individual trading behavior are often swamped or offset by non-tax considerations (Hanlon and Heitzman 2010). In particular, the well documented disposition effect, according to which investors are more likely to realize gains than to realize losses (Shefrin and Statman 1985; Odean 1998), runs in opposite direction than the effect of capital-gains taxes on individual investment behavior.

These considerations motivate the research questions in this paper: First, we study the causal effect of capital-gains taxes on individual-level holding periods of private stock-market investments. Second, we study the causal effect of taxes on the disposition effect. The literature has touched upon these two research questions (for example in Barber and Odean 2004 and Ivković et al. 2005; see below and section 4.2 for more literature), but the evidence is surprisingly limited and our paper aims to move beyond existing studies in understanding the role of investor-level taxes in trading markets. To address the research questions, appropriate micro-level data need to be combined with an institutional set-up that offers plausible exogenous variation in taxes. However, there do not exist many micro-level data for individual investors and exogenous variation in capital-gains tax rates is very seldom (Poterba 2001, for example, discusses the difficulties of identifying tax effects in investment behavior). As a result, the combination of appropriate micro data with a convincing quasi-experimental institutional set-up rarely exists in the existing literature.[49]

One strand of literature uses data from individual tax returns to study the link

[48]The theoretical effects of taxes on trading behavior are for example discussed in Constantinides (1984) and Ivković et al. (2005). We elaborate on the theoretical predictions in the context of our set-up further below in the Introduction.

[49]In the following, we provide a brief overview of different literature strands to illustrate the contribution of our study. An extensive review of the literature is presented in section 4.2.

between overall capital gains and taxes (e.g., Feldstein et al. 1980 and Jacob 2018. An early survey is Poterba 2002). However, tax-return data usually do not include information that are potentially important for a comprehensive understanding of tax effects on trading behavior; in particular, they typically only have aggregated annual information on the total amount of capital gains and as such lack information on single sales (such as the holding period of realized sales).[50] Tax-return data further do not include information on unrealized sales (because these are not tax relevant), and they do not have information about trading activities in non-taxable accounts. Studies from a different strand of literature use firm and stock level data to shed light on the effect of investor-level taxes (a review is in Hanlon and Heitzman 2010). However, because these data are on the firm or stock level, they do not allow studying the individual tax responses of investors.

Another set of papers overcomes these data challenges and uses individual-level investor data obtained from brokerage houses to study tax effects on investor behavior (e.g., Odean 1998, Grinblatt and Keloharju 2001, Barber and Odean 2004, and Ivković et al. 2005). In light of the limited availability of quasi-experimental institutional set ups, the identification of tax effects in this literature is often based on the comparison of trading behavior in taxable accounts and tax-deferred accounts, rather than exogenous variation in tax rates. However, there potentially exist differences in trading behavior between taxable and non-taxable accounts for non-tax reasons and it is therefore difficult to isolate the tax effect in such a setting. Another approach in this literature is to compare trading patterns in December to trading patterns during the rest of the year, and attribute the December differences to taxes because of end-of-year tax planning. However, such an approach offers no direct evidence of tax effects and might for example be confounded by the momentum effect, window dressing or an overall tendency of investors to 'clean-up' their portfolios towards the end of the year (see section 4.2 for a more nuanced discussion of these identification issues).

In our paper, we add to the existing literature by combining individual-level investor data with a large tax reform that is exploited for causal identification. In particular, we use confidential portfolio-level data provided by a large commercial bank in Germany.[51] These data contain daily information about the entire trading behavior (including purchases and sales of stocks and other assets) in a panel of approximately 100,000 individual investors for the period 1999 to 2016. Benchmarking with official statistics and the comparable US data set used in e.g. Odean (1998), we show that our sample of investors

[50] Dowd and McClelland (2019) is an exception that uses more frequent tax-return data on sales. See section 4.2 for more details.

[51] The type of data are comparable to the frequently used US data set which is propriety data from a discount brokerage house (e.g. Odean 1998; Barber and Odean 2000; Barber and Odean 2001). Our data have for example been used by Leuz et al. (2017).

is representative for the overall population of German investors and similar to U.S. investors. We focus on the trades of stocks in our analyses and explore the effect of taxes on the holding duration of stocks.

To identify causal tax effects, we exploit the institutional setting of capital-gains taxation in Germany before and after a large reform in 2009. This reform consisted of two major changes: i) Before the reform, the gains arising from sales of assets with a holding period of less than one year were subject to a so-called 'speculation' tax. The rate for this speculation tax was equal to the marginal income tax rate of the selling investor. The gains of sales with a holding period longer than one year were tax exempt. Short-term losses with a holding period of less than 365 days could be used to offset tax-relevant gains. As a result, the before-reform tax system created a holding-period based intertemporal discontinuity in the taxation of capital gains. The pre-reform system is thus similar to the tax set-up in the US that also differentiates between long-term and short-term gains, though the German pre-reform system has a larger tax differential between long-term and short-term gains (with tax free long-term gains). This speculation tax was abolished in the context of the reform. ii) After the reform, all capital gains are subject to a flat tax of 25%. That is, capital-gains taxes became independent of the individual marginal tax rate and independent of the holding duration of the sold asset. Capital gains from share sales can be offset by losses from share sales.

We start our empirical analysis with an investigation of the number of realized sales around the holding-period dependent tax discontinuity. For this purpose, we non-parametrically plot the number of sales (in bins of seven days) by holding duration before and after the reform and separately for losses and gains. Theoretically (following e.g., Constantinides 1984), we expect that tax-sensitive investors realize losses as long as they can be deducted from the tax base (i.e., before the intertemporal tax discontinuity is crossed). This implies that we should see an increased number of realized losses *before* the tax discontinuity in pre-reform years. On the other hand, tax-sensitive investors should delay the sale of gains until they qualify for the preferential tax treatment. We thus expect an increased number of realized gains *after* the tax discontinuity in pre-reform years.[52] As the holding period is not tax relevant in post-reform years, we do not expect to find any irregularities in the number of sold gains or losses around the 365-days holding period.

The empirical findings are consistent with the predictions. We see in pre-reform years that the number of sold losses spikes sharply just before the 365-days cutoff. The

[52]The model in Constantinides (1984) further predicts that gains should be realized immediately once they qualify for the lower long-term tax rate. This implies that we should see a spike in the number of sales to the right of the tax discontinuity (i.e, during the first week after 365 holding-period days). Losses, on the other hand, should be realized as they accrue, according to the model, and their realizations do not necessarily spike anywhere in the short-term-tax period.

number of sales in the seven days before the cutoff is roughly 3.2 times as large as the number of sales during the seven-day bin just after the cutoff. We further see that investors defer sales of gains until they have reached the 365-days holding period. We further see a sharp spike in the number of sold gains in the weeks just after the 365-days cutoff.[53] We further show that our findings are not driven by a few very tax-sensitive investors; the number of distinct investors who sell share packages also spikes around the time discontinuity.

In the post-reform years where the 365-days cutoff is not tax relevant, we see no spikes or other irregularities around the holding period of 365 days. The absence of any spikes whatsoever in post-reform years clearly suggests that the pre-reform spikes are not driven by any non-tax factors and can indeed be attributed to a causal effect of the tax.

We use a difference-in-bunching method to estimate the elasticity of the length of the holding period with respect to the tax rate. Conventional bunching methods estimate the counterfactual distribution by extrapolating the (seemingly) unaffected region away from the discontinuity to the region in the neighborhood of the discontinuity. As we elaborate below in section 4.2, in our context the regions away from the discontinuity are plausibly affected by the discontinuity as well and therefore do not qualify to predict a counterfactual. To overcome this issue, we use data from a time period without discontinuity (the post reform years) to construct the counterfactual distribution (e.g., Brown 2013, Kleven 2016). For gains, we find a tax elasticity of the holding period between 0.185 and 0.56 (depending on the applicable tax rate of investors which we do not see in our data). This translates to a tax-induced increase in the holding period of 16 days for gains. The results for losses are similar: the elasticity estimates range between 0.195 and 0.59 and the change in the holding period is roughly 17 days.

The difference-in-bunching approach also sheds light on the question of where the excess mass in the number of sales around the time discontinuity (in pre reform years) comes from. Are the spikes that we see to the right (for gains) and to the left (for losses) of the discontinuity 'fed' by sales that investors would have realized before or after the

[53]It is consistent with the model prediction that we observe a larger number of realized losses to the left of the discontinuity and a larger number of realized gains to the right of the discontinuity (also see below where we discuss in the bunching setting where the 'excess mass' comes from). The spike in realized gains that we see in the first week after the discontinuity is also consistent with predictions. For losses, the model does not predict that realizations should spike just before the discontinuity (see footnote 52). However, the spike may be consistent with loss averse investors who need a self-control mechanism to realize their losses – as for example described in Shefrin and Statman (1985, section I.D.). Loss averse investors are reluctant to realize losses and only realize their losses when there is an external self-control mechanism (commitment device) that induces them to sell losses. The intertemporal tax discontinuity, which is salient and known to investors, potentially serves as such an external self-control mechanism (commitment device) because the losses lose their valuable tax-shield function once the discontinuity is crossed. As a result, losses are not realized immediately as they accrue (because investors do not like to realize losses) and instead are realized shortly before the time discontinuity (because of its role as a commitment device).

cutoff in the absence of the tax? For gains, we see that the mass mostly comes from the left of the discontinuity; this suggests that investors delay the sales of gains until they qualify for tax exemption. For losses, we see that the mass of investors mostly comes from the right side of the discontinuity; this suggests that investors move forward the realization of sales in order to count them against their tax-relevant gains.

The next steps of our analysis are based on hazard-rate regressions which estimate for each day of the holding period the probability that a given asset is sold on this day of the holding period. In contrast to the previous approaches (plotting of the number of sales by holding period and bunching), hazard-rate regressions allow i) for the use of daily data, ii) to exploit stocks that are not sold, iii) to include control variables and interactions which make it possible to study heterogeneous effects, and iv) to study the disposition effect (literature on the disposition effect usually uses hazard-rates as well). We present all results of the hazard-rate regressions in graphs that plot for each day of holding period the coefficient estimating the probability of sale.

The hazard-rate regressions confirm our previous results. In pre-reform years, we estimate strongly increased selling probabilities just before holding periods of 365 days for losses and just after 365 days for gains. We see no increased selling probabilities around the 365-days cutoff in the post-reform years, neither for losses nor gains, which is further support of a causal tax effect. We also estimate hazard-rates regressions separately for each year in our sample period. We see spikes around the 365-days discontinuity in all pre-reform years for both gains and losses, but we never see any spikes or irregularities around the 365-days holding period in any of the post-reform years. Our main results are thus not driven by a few exceptional years in our sample. We interpret these findings as additional evidence that the tax discontinuity clearly affects trading behavior.

Average effects of taxes potentially mask heterogeneity across different types of investors. Our rich data allow us to study several sources of heterogeneity and understand which types of investors exhibit the largest tax responses. We focus on three sources of heterogenity which have received considerable attention in the trading literature (e.g., Barber and Odean 2001; Seru et al. 2009; Korniotis and Kumar 2011): age, experience and gender. We find strong evidence that tax responsiveness is increasing in trading experience. This finding is based on the observation that spikes in selling probabilities around the discontinuity increase in experience. Notably, this effect of experience is conditional on a set of covariates including age, implying that the experience effect is not driven by age. We further see that the tax response increases in age (conditional on experience), in particular in the context of losses. Regarding gender, we find that men are less likely to sell their losses during the days before the discontinuity. We further explore heterogeneity w.r.t the magnitudes of gains and losses, which has been shown to be potentially relevant in Ivković et al. (2005). We find that the tax responsiveness increases in the size of the gains or losses and that this effect is about double as large for losses relative to gains.

A robust finding in the literature on trading behavior is that investors have a larger propensity to realize gains than to realize losses, the so-called disposition effect. Considering that the disposition effect and tax effects potentially run in opposite directions (see intuition above), we study how the disposition effect interacts with tax effects. In post-reform years (without intertemporal tax discontinuity), we observe the disposition effect on each single day of the holding period; that is, gains are always sold with a higher probability than losses. This confirms findings in the large literature that documents the disposition effect. In pre-reform years, however, we detect the disposition effect only for holding-period days which are sufficiently distant to the tax relevant 365-days discontinuity. In the neighborhood to the left of the intertemporal discontinuity, we observe that gains are sold with a much smaller probability than losses. To the right of the discontinuity, gains are sold with a greater probability than losses, but this increased probability is much larger than the 'usual' disposition effect that we see in post-reform years and further away from the cutoff. We also find suggestive evidence that the tax discontinuity affects the disposition effect even on holding-period days distant from the cutoff. Compared to the post-reform benchmark (without tax relevant cutoff), the disposition effect in pre-reform years tends to be lower during the first year of the holding period and higher after 365 days holding period have passed on days distant to the 365-days cutoff. This suggests that the time discontinuity affects the disposition effect even on holding-period days distant to the discontinuity.[54]

These findings confirm in a credibly identified set-up the findings of, for example, Odean (1998) and Ivković et al. (2005) that an intertemporal tax discontinuity affects the disposition effect. An additional implication of our finding for papers in the literature on the disposition effect arises because we find that, in a system with an intertemporal tax discontinuity (such as the U.S.), taxes have an effect on trading behavior and the disposition effect throughout the entire year and not only in December. Many studies in the literature on trading behavior and the disposition effect control for tax effects by allowing for separate December effects; our findings imply that this December approach is not sufficient to adjust for tax effects. More generally, our findings provide novel evidence on the causal determinants of the disposition effect. As recently suggested by Frydman and Wang (2020), the causes of the disposition effect are still subject to debate, and we are able to add to this debate in that we show that capital-gains taxes have an

[54]Previous literature finds for the U.S. that older and more experienced investors are less prone to the disposition effect. The findings from our heterogeneity analysis (see above) indicate that age and gender effects on the disposition effect are driven by tax effects and that heterogeneity in the disposition effect along the age and experience dimensions would be mitigated in the absence of intertemporal tax discontinuities – see the Conclusion (section 2.6) for more discussion on this.

impact on the disposition effect. To this end, we for example relate to a recent stream of papers showing that purchase prices (or their salience) causally affect the disposition effect (Frydman and Rangel 2014; Frydman and Wang 2020; Loos et al. 2020).

The paper proceeds as follows. Section 4.2 provides an overview of the related literature and discusses how our paper contributes to existing findings. Section 4.2.1 describes the institutional background of capital-gains taxation in Germany during our sample period. Section 4.3 provides information on the data and the calculation of holding periods in this data set. We describe the empirical strategy and causal identification in Section 4.4. The results are presented in Section 4.5. Section 4.6 concludes the paper.

4.2 Contribution to the Literature

We relate to (empirical) studies in different fields and literature strands. We therefore believe that a systematic and extensive overview of the literature studying the tax effects on trading behavior may be valuable to readers. In the following, we describe the approaches and findings in the related empirical literature and elaborate on our contribution relative to the existing studies. We organize the literature review along the different strands of literature that we identified to be relevant for our paper.

Literature using tax-return data. First, we speak to studies that use data from individual tax returns to study the link between capital gains and taxes (an early survey is Poterba 2002). This literature usually finds a negative relation between realizing capital gains and taxes (e.g. Feldstein et al. 1980; Bogart and Gentry 1995; Daunfeldt et al. 2010; Jacob 2013; Dowd et al. 2015; Jacob 2018).[55] Our findings on the behavioral effects of capital gains taxes relate to this literature and we confirm that capital gains taxes induce investors to defer the realization of gains. However, as mentioned above, studies using tax-return data typically only have aggregate annual information on the total amount of capital gains and as such lack information on single realized sales; i.e., whether a single realized sale is a gain or loss or how long the respective asset had been held by the investor. Tax return data also do not include information on unrealized sales because these are not tax relevant, and they do not have information about trading activities in non-taxable accounts. Our paper uses portfolio-level data that allow us to overcome most of the data restrictions in this literature. For example, one main finding in our paper relies on the differentiation between gains and losses of single sales, and our hazard-rate regressions account for unrealized assets and exploit the daily frequency of our data set.

[55]Saez (2017) studies the behavioral responses of reported incomes to the 2013 tax reform in the US. Using annual IRS income statistics, the paper finds considerable responses of reported income to the reform, with much of the effect being driven by realized gains.

Two recent studies use US tax-return data that include information on sales at a less aggregated frequency. Hoopes et al. (2016) have daily data on sales, but their study is not about tax effects. Dowd and McClelland (2019) use American IRS data on capital realizations for directly held assets on the level of the single transaction. These data are based on Form 8949 which requires taxpayers to report the purchase and sale price as well as the date of acquisition and disposition. Using these data for the tax year 2012, the authors are able to calculate the holding period (in weekly bins) for single assets and study whether the holding period is affected by the intertemporal tax discontinuity in the US that differentiates between short-term and long-term gains. Consistent with our findings, Dowd and McClelland (2019) find that the number of realized gains spikes in the first week in which the lower long-term rate is available; i.e., they see a spike in realized gains to the right of the discontinuity. Losses spike on both sides of the tax discontinuity. As acknowledged by the authors, this is somewhat surprising since a rational investor should sell losses as short-term in order to offset short-term gains. The paper then focuses on gains and uses bunching methods to estimate a counterfactual distribution that is constructed by predicting sales in the neighborhood of the time discontinuity using sales further away from the discontinuity. Based on this approach, the authors estimate that a 10 percent reduction in tax rates for high-income taxpayers leads to a 7.3 percent increase in realized gains shifted to just beyond the one-year holding period.

Our paper moves beyond Dowd and McClelland (2019) along a number of dimensions. i) We have an institutional setup where the intertemporal time discontinuity was first in place and is then abolished in the course of a reform. This setting yields a proper counterfactual and allows us to compare the effects of the time discontinuity to years where the discontinuity did not exist. Dowd and McClelland (2019), who only have one year of data, use the conventional bunching approach to construct a counterfactual. This approach relies on the assumption that the number of realized sales away from the cutoff (which are used to estimate the counterfactual distribution of sales just around the discontinuity) are not affected by the tax discontinuity. Theory as well as our results show that this assumption is unlikely to hold. For example, some of the gains that are realized just after 365 days might have been realized considerably before the 365 days cutoff in absence of the tax discontinuity. Our hazard rates confirm this as they show that the entire distribution of sales probabilities to the left of the cutoff is shifted downwards in pre-reform years relative to post-reform years (without discontinuity). ii) In addition to plotting the number of sales by holding period around the intertemporal time discontinuity, we estimate hazard-rate regressions which, as discussed in the Introduction, have several advantages. In particular, they allow to exploit the daily frequency of the data and account for sales that are not realized (whereas Dowd and McClelland 2019 use weekly data and do not have information on unrealized sales). iii) In contrast to

Dowd and McClelland (2019), we use the institutional setting of the intertemporal tax discontinuity, *and* its abolishment, to shed new light on the interaction of taxes and the disposition effect. iv) We exploit a panel over several years whereas Dowd and McClelland (2019) only have one year of data. v) We study how capital-gains taxes affect the disposition effect, vi) Looking at the case of Germany, we exploit a set-up where the tax differential between short-term and long-term gains is larger than in the US case.

Literature using firm and stock level data. Second, we relate to a strand of literature that uses firm and stock level data to investigate the effects of investor-level capital-gains taxes (see the overview in Hanlon and Heitzman 2010). This strand of literature for example studies the effects of investor-level taxes on asset prices, end-of-year market irregularities, acquisition premiums, turnover patterns, and the role of taxes in reactions to news disclosures (e.g., Reese Jr 1998; Lang and Shackelford 2000; Seida and Wempe 2000; Poterba and Weisbenner 2001; Shackelford and Verrecchia 2002; Blouin et al. 2003; Ayers et al. 2003; Ayers et al. 2008; Dai et al. 2008; Blouin et al. 2009). These papers typically find that capital-gains taxes matter and that they affect asset prices. These findings are indirect evidence that capital-gains taxes affect individual selling behavior, but they do not allow for the identification of tax effects on individual behavior directly. Studying tax effects on individual investors is generally difficult with firm level data.[56]

Literature using investor-level trading data. Third, we relate to a set of papers that overcomes data issues and studies the link between taxes and individual trading behavior using portfolio-level micro data. One of the papers in this literature is Ivković et al. (2005) who use data from a discount brokerage. These data allow the authors to track the single investments of individual investors – their US data are very comparable in spirit to the German data that we use. To shed light on taxation effects, Ivković et al. (2005) compare trading behavior in taxable accounts and trading behavior in tax-deferred accounts (IRAs or Keogh plans). The paper finds a negative relation between accrued gains and the selling probability in taxable accounts for stocks with a holding period of more than one year, while it does not observe such a relation in tax deferred accounts. In light of the presumption that taxes should induce investors to defer the realization of gains (see explanation above) and because this negative relation is only observed in taxable accounts, the authors suggest that their finding is an indication of taxation effects (lock-in effect) on trading behavior. The results of the paper also speak to the interaction between taxes and the disposition effect. The relation between accrued gains and selling probabilities in taxable accounts is (as described above) negative once a stock has been held for more than 12 months, and it is positive in the first few months of the holding

[56] A related stream of papers studies the tax sensitivity of institutional investors (Blouin et al. 2017; Sikes 2018).

period. This suggests that the disposition effect outweighs tax effects only in the first few months after the stock was purchased and that tax effects matter more for longer holding periods.

Another paper in this literature is Barber and Odean (2004) who also use investor-level data from brokerages and compare trading in taxable and tax-deferred accounts. They find that the realization probabilities of gains and losses are very similar across these two types of accounts, except in December when loss realization is more pronounced in taxable accounts than in tax-deferred accounts. The authors attribute this 'December' difference between the two types of accounts to tax-loss selling. This somewhat contrasts the results of Ivković et al. (2005) who find that investors are more likely to realize losses in taxable accounts than in tax-deferred accounts throughout the entire year, not just in December. It is therefore an open question if realization probabilities are always different between these two type of accounts or just in December.

Overall, rather than exploiting exogenous variation in taxes, identification of tax effects in Barber and Odean (2004) and Ivković et al. (2005) thus comes from the comparison of taxable and tax-deferred accounts. However, differences in trading behavior between these accounts are not necessarily fully attributable to taxation effects. Trading behavior might be different between these two types of accounts for non-tax related reasons (even conditional on investor fixed effects and exploiting that many investors have both taxable and tax-deferred accounts). For example, investors usually use tax-deferred accounts to save for retirement, and they might therefore be inclined to invest in different types of assets in these accounts than in taxable accounts. In their tax-deferred retirement accounts, investors might seek to invest in less risky assets or purchase assets for these accounts with a much longer investment horizon and hence with the explicit goal of trading these assets less frequently. This assertion that investments in taxable and tax-deferred accounts are different for non-tax reasons is supported by the literature: e.g., theory contributions on the optimal allocation of assets come to the result that certain assets, such as taxable bonds and actively-managed mutual funds, should be held in tax-deferred accounts, whereas other asset types, such as tax-exempt bonds, passively-managed mutual funds and stocks, should be located in taxable accounts (Huang 2001; Dammon et al. 2001; Shoven and Sialm 2004). Consistent with the assertion that non-tax considerations make a difference for investment behavior across these two accounts, Barber and Odean (2004) find that turnover is higher in taxable accounts than in tax deferred accounts. This finding of Barber and Odean (2004) induces Ivković et al. (2005, page 1617) to acknowledge that investors may view taxable and tax-deferred accounts differently and that their estimates for tax-motivated trading might therefore be biased.

We move beyond Barber and Odean (2004) and Ivković et al. (2005) in that we use a similarly rich data set of individual investors, but combine it with quasi-experimental variation in tax rates which comes from the intertemporal tax discontinuity and the abolishment of this discontinuity. Another difference to Barber and Odean (2004) and Ivković et al. (2005) is that we use data with daily frequency, rather than monthly frequency. The daily data allow us to zoom in the trading behavior along each day of the holding period, which is especially useful in analyzing trading behavior around the holding-period based intertemporal tax discontinuity.

A further set of papers documents in individual-level data sets that trading behavior in December is different than trading behavior in other months of the year. For example, Odean (1998) and Grinblatt and Keloharju (2001) find evidence of the disposition effect in all months of the year, except in December. These papers interpret these findings as evidence that i) capital-gains taxes matter for investment behavior and that tax-loss selling is prevalent and ii) that the disposition effect is affected by taxes. However, differences in trading behavior in December vs. other months of the year are only indirect evidence of tax effects. These papers do not rely on exogenous variation in tax rates and it is therefore not clear to which extent the 'December' finding is driven by taxes or other seasonality patterns. For example, it is not clear why tax-loss selling should not occur throughout the year. In addition, as noted by Grinblatt and Keloharju (2004, pages 52-53), the December effect could also be explained with the momentum effect or window dressing. It has been shown that the momentum effect for losses is much larger in December than in other months of the year and also much larger than the December momentum effect for gains (Grinblatt and Moskowitz 2004). This then implies that it could be rational for investors to sell losses in December even in the absence of tax considerations. Window dressing may also play a role: December often is the time to recap one's portfolio and investors may be embarrassed to carry on losers to the next year. Considering these concerns, Grinblatt and Keloharju (2004) study if investors sell losers and then immediately repurchase the same stocks (so called wash sales) and indeed find evidence for this behavior. This then is a better indication that tax considerations matter, but is yet no direct evidence of tax effects.

Using exogenous variation in rates and the abolishment of a large intertemporal tax discontinuity, our paper provides clear and direct evidence for tax effects on trading behavior and on the interaction of taxes and the disposition effect that does not rely on trading patterns in December. We therefore complement the existing literature in that we study a set-up where the concerns about the roots of differential trading behavior in December do not play a role. Our finding on the interaction between taxes and the disposition effect more generally relates to papers that demonstrate that tax considerations of individual investors are sometimes swamped by non-tax considerations or behavioral aspects (Hanlon and Heitzman 2010).

Literature on the disposition effect. Fourth, our paper contributes to the general literature on the disposition effect (in non-tax contexts). As summarized in the handbook chapter by Barber and Odean (2013), the evidence is very robust that individual investors sell gains with a higher propensity than losses. An important question in this literature is whether and how the disposition effect is causally affected by other factors. This is potentially relevant because an understanding of the causal determinants of the disposition effect can help to improve investment behavior. However, as stated by Frydman and Wang (2020, page 233), the cause of the disposition effect is still debated. A few recent papers provide causal evidence on the determinants of the disposition effect. Frydman and Rangel (2014), Frydman and Wang (2020) and Loos et al. (2020) show that changes in purchase prices or changes in the salience of purchase prices affect the disposition effect. We relate to these papers on the causal drivers of the disposition effect and provide novel evidence that taxes affect the disposition effect and can even reduce it temporarily.

A further contribution of our paper to the literature on the disposition effect is to show that it is not sufficient to have separate December effects to control for tax effects. Consistent with our finding that taxes affect the disposition effect, the literature has acknowledged that tax effects should be controlled for in disposition-effect settings. We provide evidence that taxes affect trading behavior and the disposition effect throughout the entire year, which then implies that December adjustments will not fully control for tax effects.

Literature on the behavioral responses to taxes. Fifth, we also contribute to the large literature on behavioral responses to taxes using individual-level data. This literature studies the causal effects of taxes along many dimensions, often relying on bunching approaches and/or large tax reforms for identification.[57] We contribute to this literature in that we add micro-level evidence on the behavioral effect of taxes along a dimension that has rarely been investigated before, namely individual-level trading behavior.

4.2.1 Institutional Background

Our analysis is based on the system of capital-gains taxation in Germany between 1999 and 2016 (i.e., the time period of our data set). We focus on the trade of shares and describe in this section how capital gains occurring from shares are treated in Germany.

[57] This large literature for example studies causal effects of taxes on: taxable income (Chetty et al. 2011; Saez et al. 2012; Kleven and Schultz 2014b), investment behavior (Yagan 2015), dividend payments (Chetty and Saez 2005), education (Abramitzky and Lavy 2014), wealth accumulation (Jakobsen et al. 2018), housing prices (Best and Kleven 2018), wages (Suarez Serrato and Zidar 2016; Fuest et al. 2018b), consumption (Chetty et al. 2009), migration (Kleven et al. 2014; Agrawal and Foremny 2019) and labor supply (Eissa and Liebman 1996; Martinez et al. 2018).

A major reform of capital-gains taxation was implemented in 2009 and therefore falls into our sample period. Both before and after the reform, capital gains from shares are generally only taxed upon realization (i.e., taxes are due when the share is sold).

Taxation of capital gains before 2009. Before the reform, the tax treatment of capital-gains was dependent on the holding period of the underlying asset. The gains and losses of assets sold within a holding period of 365 days or less were subject to taxation. This tax was commonly refereed to as a 'speculation tax'. The tax rate was identical to the personal income-tax rate (PIT) of the investor. The PIT depends on the sum of all income types (wage income, self-employment, etc). The top income tax rate (PIT rate) was, for example, 42% in 2008 and applied for overall annual taxable income greater than 52,152 EUR. The entry tax rate in 2008 was 15%. Losses from sales with a holding period of $\leq$ 365 days could be used to offset gains from capital gains. Between 2001 and 2008 a so-called half-income method applied: one half of the gains/losses from capital gains with holding periods $\leq$ 365 days were subject to the tax.

For illustration, consider a fictitious investor who is subject to the top-income tax rate of 42%. She realizes gains worth 2000 EUR from shares that she had held less than 365 days, and she sells other shares within the 365-days holding period which come with losses of 200 EUR. The resulting capital-gains tax liability for this investor then was $1/2 \times (2000 - 200) \times 0.42 = 378$ EUR.

The gains resulting from assets with a holding period of more than 365 days were not subject to any taxes; the resulting tax liability on gains was zero if the underlying asset was held for more than 365 days. Accordingly, losses resulting from assets with a holding period of more than 365 days could not offset positive capital gains.

This system of capital-gains taxes applied to assets such as stocks (as long as the investor is not a substantial shareholder), funds, certificates (except guarantee certificate) and capital gains from bonds (except zero bonds). Overall, the system creates large incentives to realize gains after the relevant holding period of 365 days, while losses should be realized within the 365-days holding period to reduce the tax base.

Taxation of capital gains since 2009. The tax treatment of capital gains was substantially reformed as of January 2009. In stark contrast to the old system, the holding period of assets is not tax relevant anymore. That is, the holding-period based 'speculation tax' was abolished in the context of this reform. It was replaced by a system where all capital gains and capital losses (independent of holding duration) are subject to a flat tax of 25% or, if the PIT rate is smaller than 25%, the PIT rate. That is, the tax on capital gains/losses is min(25%,PIT rate). Losses can be used to offset gains. The half-income method was abolished.

Consider again an fictitious investor who is subject to the top PIT rate (which is $> 25\%$) and who has capital gains of 2000 EUR and capital losses of 200 EUR. Her tax liability after the reform is independent of the holding periods of the underlying assets and sums up to: $(2000 - 200) \times 0.25 = 450$ EUR. Importantly, any tax incentives to hold assets for a certain time period were abolished. The old pre-2009 tax rules applied to all assets bought before January 2009 (resulting in grandfathered assets).

4.3 Data

4.3.1 Data Description and Summary Statistics

We use individual investor and portfolio data from a large German online bank. The full-service bank has more than half a million customers and operates across the entire country. We obtain a sample of about 110,000 investors which is randomly drawn out of all of the bank's clients. Variants of this data set were for example used by Leuz et al. (2017) and Loos et al. (2020). For each investor, we have the complete trading history for the period January 1999 to May 2016. These data allow us to construct an individual-level panel of daily trading activities over almost 18 years. Trading information in the data include type of traded asset, transaction volumes, prices, order types (with or without limit) and dates for purchases and sales. We further have investor information on age, gender, zip code of residence, marital status, employment type, and for how long the investor has had the trading account. In addition, the data include self-reported information about education, income (in categorical ranges), total wealth and risk tolerance.

For the purpose of our paper (in which we focus on the trading of stocks), we restrict the sample to all investors who have purchased at least one stock during the sample period. This leaves us with about 93,000 investors. These investors bought about 8.4 million share packages with an overall purchase value of 49 billion EUR during our sample period (the unit of analysis in most of our analyses will be a share package; see section 4.3.2 below for a definition and more information). Table 9 provides summary statistics for all investors in our analysis sample. The average portfolio value (incl. all assets in the portfolio) is 51,725 EUR and the investors in our sample make on average roughly 78 trades in total over the observation period. The average monthly portfolio turnover[58] was 10.86 percent, which implies that most investors have quite active accounts. Most investors in our sample are male (83%) and their average age (by the end of 2015) was 52 years. 6% work in the financial sector and 16% of our sample is self-employed. The average investor in our sample has held the account at this bank for more than 13 years (as of the end of 2015); we use this measure as a proxy for trading experience. The share

[58]Monthly portfolio turnover is calculated as in Barber and Odean (2001) as one-half of the monthly sales turnover plus one-half the monthly purchase turnover. Sales (purchase) turnover is defined as value of shares sold (purchased) divided by the portfolio value in the beginning of the month.

of investors in our sample with a PhD-level degree is 6%, whereas the share over the entire German population is only about 1.5%. This is in line with prior evidence showing that individuals with investment portfolios are more educated than the population average (van Rooij et al. 2011; Cole et al. 2014; Leuz et al. 2017).

We investigate if our sample is representative for the population of investors and does not only include special groups of investors or play-money accounts. For this purpose, we provide several comparisons of our data sample with i) the German population of investors and ii) with other comparable data sets that have been used in the literature (these comparisons build on Leuz et al. 2017 who use a very similar data set). The portfolio value in our sample (51,725 EUR) is very comparable to the number that the German central bank (*Bundesbank*) reports to be the average portfolio value of German equity investors: 48,000 EUR in 2013 (Deutsche Bundesbank 2013). We further construct a variable that measures the ratio of portfolio value over annual income for our data and benchmark this ratio with official statistics reported by the German Federal Office. As income in our data set is reported in several categorical ranges, we use either the midpoint or the lower end of each range as a proxy for investor income. Using the midpoint, the mean ratio of the average portfolio value (over the entire sample period) to annual income is 1.3. Using the lower ends of each income range as a proxy for annual income, this ratio is calculated to be 1.2. These numbers are very close to the ratio of total financial assets to gross household income in the German population: 1.1 (German Federal Bureau of Statistics 2008b; German Federal Bureau of Statistics 2008a).[59] In addition, the ratio of the median portfolio value to median gross income for the German investors surveyed by Dorn and Huberman (2005) is 0.6 and it turns out to be 0.6 for our sample as well.[60] Overall, these comparisons allow us to conclude that our investor sample is representative of the population of German investors and should not be significantly biased by play money accounts.

Demographic and portfolio characteristics of the investors in our sample are also well comparable to the well-established investor data set used by, for example, Odean (1998) and Barber and Odean (2001). Their data are obtained from an U.S. online brokerage house and are similar in spirit to the data that we use. For example, average age (50.4 vs 52.26) and the share of males (79% vs 86%) is fairly similar across these data sets. Furthermore, the average portfolio value of about 51,000 EUR is in the same order of magnitude (considering the different time periods) as the average portfolio value of 47,000 USD that is reported in Barber and Odean (2001).[61]

[59]We manually calculate this value from total financial assets and the monthly gross income reported in the above sources.

[60]We manually calculate this from the values given in Tables 1 and 2 of Dorn and Huberman (2005, pages 443 and 447).

[61]The EUR-USD exchange rate throughout our sample period was at 1.16 in Jan. 1999 and 1.11 in

4.3.2 Unit of Analysis

We are interested in the number of stock realizations around the intertemporal tax discontinuity. That is, our primary interest is not regarding the number of investors trading around the discontinuity (although we analyze this too in one series of analyses) and we thus do not employ the investor as the unit of analysis. To study the number of stock realizations, we use 'share packages' as the unit of analysis throughout most of the paper. One 'share package' is independent of the number of shares that are included in this share package. For example, if an investor buys 100 shares and sells this 'package' of 100 shares ten days letter, we generate one observation with a holding period of 10 days (see below for more on the measurement of holding period). If another investor buys 10,000 shares and sells her 'package' of 10,000 shares 10 days letter, we also generate one observation with a holding period of 10 days. We selected this unit of analysis in order to avoid that our results are driven by the behavior of a few large-scale investors or penny stocks. Our approach reflects that we are eventually interested in the individual behavior of investors and we want to avoid that the individual behavior is weighted with the number of shares that an individual investor moves. In the previous example (one investor selling 10 and one selling 10,000 shares), both of these investors are given the same weight in our analysis because we are interested in the tax-induced trading behavior of both these investors. If single shares were the unit of analysis, the behavior of the smaller 10-shares investor would be almost negligible relative to the behavior of the bigger 10,000-shares investor.

4.3.3 Measuring the Holding Period

We measure the holding period as the difference in days between purchase date and sales date of a share package. For example, if a fictive investor buys five shares of some firm on the second of October and sells all of them on the 15^{th} of October, this would result in one observation with a holding period of 13 days. If the first purchase of a share package occured outside our sample period (i.e., prior to January 2009), we cannot calculate the holding period and have to drop the share package from our analysis.

If there are multiple buys before the first sale occurs we apply the first-in-first out principle (which is in line with the German tax law). For example, if an investor buys two shares on Oct 5, ten shares of the same firm on Oct 10, and then sells all 12 shares on Oct 20, we generate two observations with holding periods of 15 days and 10 days, respectively.

If one purchase is sold on several days we create one observation for each sale. For example, consider an fictive investor who buys five shares on October 2, then sells three of these shares on October 4 and two shares on October 15. We then create two

May 2016.

observations: one with a holding period of two days and the other with a holding period of 13 days.

Sometimes shares change their isin or shares are splitted or reverse splitted. We account for this by using hand collected data for isin changes and data on splits and reverse splits from datastream.[62] In cases were shares have been splitted or reverse splitted, we adjust prices such that purchase and sale price have the same basis.[63] Since we cannot guarantee 100% that we are able to detect all events with isin changes or changes in the volume of a trade, we compare the trade volume from the transaction data with the final stock at each end of the month, which is reported in a separate data file. Per deposit-isin combination we keep trades until we detect any difference between stock and accumulated trading volume. We drop all trades for the affected deposit isin combination for points in time after the first difference.

4.3.4 Final Sample

Our analysis is based on several million share packages. For the years before the reform, we include 2.74 million observations of appreciated share packages(gains) and 2.47 million depreciated share packages (losses). In the after-reform years, we have 1.34 million appreciated share packages and 0.85 million depreciated share packages. Restricting the sample to half a year before and after the intertemporal tax discontinuity, we rely on 313,000 appreciated share packages and 380,000 depreciated packages during pre-reform period, and 212,000 gains and 136,000 for the after-reform period.

4.4 Empirical Strategy

Our empirical strategy aims at identifying the causal effect of capital taxes on trading behavior, in particular on holding periods and the probability to sell an asset. We focus on the trading of share packages throughout our analyses. In addition, we shed light on the interaction of taxes and the disposition effect.

4.4.1 Number of trades in weeks around the cutoff

The starting point for our analyses are figures in which we plot the number of sold share packages around the holding-period tax cutoff of 365 days. We group the number of sold share packages in weekly bins of seven days and primarily look at the year window around the 365-days cutoff. That is, we plot the number of sales in each week during the 26 weeks

[62]We also try to identify from the data (reverse) splits which were not reported in datastream and isin changes. For those cases, we manually check whether there was indeed an isin change or (reverse) split.

[63]Since the total value of a position is unaffected by the split or reverse split, the price basis before and after the split is not the same anymore. For example, consider 100 shares with a value of 200 Euro that are splitted by 2. Without adjustments, the price before the split is 2 Euro while it is just 1 Euro after.

prior and 26 weeks after the discontinuity. We use bins of seven days although we have access to daily sales data because of a mechanical pattern in the daily data which causes noise and is smoothed out in weekly data. Since it is not possible to trade on weekends, some day-measured holding periods occur more often than others. For example, a seven day holding period is possible for sales made on all five weekdays, whereas a four days holding period is only possible for sales made on Mondays, Thursdays or Fridays.

We plot the number of trades around the 365-days cutoff separately for years before and after the 2009 reform and separately for gains and losses. Since the holding period became tax irrelevant in the course of the reform, we expect a smooth distribution of trades around the 365-days cutoff for the years after the reform. A causal effect of taxes on trading behavior would imply that we see, in pre reform years, an increased number of trades of appreciated assets (gains) in the weeks after the 365-days cutoff, and an increased number of trades of depreciated assets (losses) in the weeks before the cutoff.

To investigate if potential tax effects are driven by a few large tax sensitive investors who sell many share packages around the time discontinuity, we also plot the number of *distinct* investors who sell share packages in a given week of the holding period. For this purpose, and analogous to the above strategy, we group the number of distinct investors who sell a share package in bins of seven days and plot the number of investors in each bin during the 26 weeks before and after the time discontinuity.

4.4.2 Difference-in-Bunching

We go on and use bunching methods to quantify the tax effects around the 365-days cutoff. Bunching approaches go back to Saez (2010b) and are now commonly used (see the recent overview by Kleven 2016).[64] We use a difference-in-bunching approach where we use the sales distribution in the post-reform periods as a counterfactual for the pre-reform distribution (as in e.g., Brown (2013); also see Kleven (2016)).

To make the pre and post reform distributions comparable and to obtain a good counterfactual, we account for level differences in the number of sales before and after the reform. We divide all weekly counts by the respective total number of share packages that have a holding period of at least 26 weeks. We include all share packages in the denominator, including those which have not been sold, when we perform these divisions. This is necessary because we observe share packages in the pre reform period for a longer amount of time than in the post period. As a result, the probability that we observe the sale for a share package bought in the pre reform period is higher than for a share package bought in the post reform period. We apply this procedure separately for gains and losses. We therefore need to determine whether an unsold share package is treated

[64] Bunching applications for example include: Chetty et al. (2011), Chetty et al. (2013b), Bastani and Selin (2014), Best et al. (2015), Best and Kleven (2018), and Almunia and Lopez-Rodriguez (2018).

as a gain or a loss. We count unsold share packages as a gain if the last observed price of that share packages was higher than the purchase price. Similarly, we count unsold share packages as a loss if the last observed price of that share packages was lower than the purchase price.

In many bunching applications, the counterfactual distribution is estimated through predicting the distribution in the region close to the cutoff using the distribution in the region further away from the cutoff. Our approach is advanategous to this conventional approach because it does not rely on the assumption that the distribution further away from the cutoff is unaffected by the discontinuity at the cutoff. This is particularly relevant because it is very likely that the large and salient discontinuity in our setting also affects sales which, in the absence of the discontinuity, would have been sold on days further away from the 365-days discontinuity. Our hazard-rate Figures suggest that this is really the case; pre-reform hazard rates away from the cutoff are different than post-reform hazard rates away from the cutoff.

The identifying assumption in our set-up then is that the post-reform distribution (without discontinuity) is a plausible counterfactual for the pre-reform distribution (which has the discontinuity). Looking at our plotted Figures, this assumption appears plausible, and we find it less strong than the assumption of the conventional bunching method. Note that the exact counterfactual distribution is particularly relevant for calculating the exact elasticity of the tax effect. To make the plain case that the spikes that we see in our data are tax effects, it is sufficient to show that there are no irregularities at all around the 365-days holding period (neither in the non-parametric plots nor in the hazard-rate regressions).

The size of the causal tax effect will be proportional to the excess mass in bunching relative to the counterfactual distribution. To quantify the tax effect, we estimate parameter b which describes the excess mass around the cutoff relative to the counterfactual distribution. This parameter is then used to calculate an implied elasticity which describes the percentage change in holding period in response to a one-percent change in the tax rate. Based on graphical evidence, we define our bunching window for gains to be the first four weeks of the holding period after the 365-days discontinuity. The spike for losses is somewhat more concentrated around the last week before the 365-days cutoff. We therefore choose the bunching window for losses to be the three last weeks before the cutoff. Note that an increased bunching window generally simply increases the excess mass and therefore the tax effect.

We calculate a standard error for the excess mass b using a bootstrap procedure. To do so, we randomly draw share packages from our sample with replacement on the investor level to generate a new set of counts and reestimate the excess mass b^j. Repeating this for a thousand times gives us an estimate for the distribution of b^j. We use the standard deviation of the b^j as our estimate for the standard error of the excess mass.

Following for example Chetty et al. (2011) and Glogowsky (2016), we calculate the elasticity parameter based on the excess mass b using the following equation:

$$e = \frac{\frac{\Delta b}{b^*}}{ln(\frac{1-t}{1-t-\Delta t})}. \tag{5}$$

Recall that, in pre reform years with tax discontinuity, the applicable tax rate for realized stock trades with a holding period of less than one year was the personal income tax (PIT) rate of the investor. At the tax discontinuity, the tax rate thus falls from the PIT rate of the individual investor to zero. Since we do not have individual tax rates or taxable income in our data, we calculate two sets of elasticities that differ w.r.t. to the PIT rate that we use: i) using the top income tax rate in 2008 (42%), ii) and the minimum income tax rate (15%).

Note that there are no strictly dominated regions in our set-up. There are at least four reasons why it might be rational to sell an appreciated share even on the day before it can be sold tax free. First, loss carryforwards: if the investors has sufficient loss carryforwards, she can sell an appreciated share tax free even if still the long term rate applies. Second, time discounting: for example the investor needs liquidity in the time before the cutoff. Third, expected prices: if the investor assumes that the price will drop strongly on the day after the cutoff, selling on the days before the cutoff might be advantageous for her. Fourth, risk aversion: even if the investor assumed that prices remain constant in expectation, it might be optimal for her to sell before the cutoff in cases where prices might fall with a small positive probability.

4.4.3 Hazard-Rate Regressions

We complement the bunching analyses with hazard-rate regressions (as used for example in Chang et al. 2016). These hazard-rate regressions estimate for each day of the holding period the probability that a given share package is sold on this holding-period day. For this purpose, we set up our data set such that it contains one observation per share package, individual investor and day of the holding period.[65] For example, this would give us 11 observations for a share package that an individual investor has held for 10 days (0, 1, 2, ..., 10). We then create a dummy variable – *Sell* – that indicates for each day of the holding period if the asset was sold on this respective day. We merge the resulting dataset with daily price information for all assets, extracted from *Datastream*. For each day of the holding period, we estimate separate regressions in which we regress the *Sell*-dummy on a constant. The resulting coefficient for the constant then describes

[65]To avoid selection in assets because of missing prices in datastream, we assign the last observed price to shares where the price is missing. This is the case for about 10% of all assets in our sample.

the probability that a share package is sold on this particular day of the holding period. We again focus on the year around the tax discontinuity. Formally, we estimate the following regression separately for each day of the holding period t:

$$Sell_{ijd} = \beta_0 + \varepsilon_{ijd}, \tag{6}$$

where indices indicate a share package i of individual investor j on calender-day date d. Note that we would not yet need indices j and d for this regression model here, in which we simply regress the sale dummy for a share package j on a constant. However, further below we will introduce investor-level (j) variables, which partly vary by calender-day date (d), and we therefore already introduce indices j and d at this point.

To summarize, our hazard-rate regressions estimate separately for each day of the holding period the probability that a share package is sold on this day of the holding period. Here, and across all hazard-rate specifications throughout the paper, the unit of analysis thus is the share package. Since we focus on the one year holding period window around the time discontinuity, our approach implies that we estimate one regression for each of holding period days 185-545. We estimate this set of regressions separately for pre-reform and post-reform years to see if selling probabilities around the 365-days holding-period cutoff are different before and after the reform. In light of the differential tax incentives for gains and losses, we further run separate regressions for gains and losses. As a result, we thus have estimates four all four combinations of pre and post years as well as gains and losses.[66]

For illustrative purposes, we plot the estimated β_0 coefficients for each day of the holding period (separate plots for gains and losses, and post and pre reform). The β_0 coefficients measure the probability of sale on a given day of the holding period. This procedure provides graphical evidence whether the selling probabilities are affected by taxation. One particular advantage of the hazard-rate approach, relative to the bunching approach, is the possibility to make use of our daily data. To complement our main hazard-rate regressions (which bundle all pre-reform or all post-reform years), we also provide hazard-rate regressions separately for each year in our data sample.

Heterogeneity w.r.t. investor characteristics. Our dataset includes several demographic variables which allow us to study heterogeneity across different type of investors. We use the hazard-rate regression setup for this purpose and add investor-level characteristics to the share-package level data. We then run the following type of regression for each day of the holding period to study heterogeneity:

[66]In our baseline specifications, we exclude grandfathered assets from the regressions. We provide robustness checks in which we relax this sample choice.

$$Sell_{ijd} = \beta_0 + \beta_1 Demogr_{ijd} + X_{ijd}\beta^{'} + \varepsilon_{ijd}, \tag{7}$$

where i, j and d again indicate share packages, investors and calender-day dates, respectively. Variable *Demogr* is the respective variable along which heterogeneous effects may occur. We focus on three different sources of heterogeneity which have received attention in the trading literature (e.g., (Barber and Odean 2001; Seru et al. 2009; Korniotis and Kumar 2011)): age of the investor, investor experience and gender. In the 'investor experience' regressions, β_1 estimates the increase in selling probability as experience increases by one year. In cases where the focus is on age, β_1 indicates the effect of one additional year of age on the selling probability. Gender is coded such that β_1 in the 'gender regressions' measure the difference in selling probability on a given day for male investors relative to female investors.

In all these regressions, we conditional on a set of observable control variables which are all included in vector **X**. These control variables include: age, investor experience, gender, birth year, income category, wealth category, dummies indicating employment in the financial sector, having a doctoral degree, and being self-employed. The respective heterogeneity variable of interest, *Demogr*, is of course omitted from vector **X** in the respective regression (for example, in cases in which we are interested in gender heterogeneity, the gender variable is included in *Demogr* and not included in **X**). In regressions in which we are interested in heterogeneity w.r.t. age (i.e., variable *Demogr* is age), we exclude birth year from the vector of control variables because age and birth year are strongly correlated and we do not want to 'control away' cohort effects when investigating age heterogeneity. The corresponding summary statistics for all variables used here are reported in Table 9. Including these control variables for example implies that the effect of investor experience is going to be conditional on age. In our results graphs, we plot the β_1 coefficients of this regression. These show if the selling probabilities are different across the groups, and we are of course particularly interested in the differential selling probabilities around the intertemporal time discontinuity.

Heterogeneity w.r.t. magnitude of gains and losses. The hazard-rate approach also allows us to estimate if potential tax effects depend on the magnitude of the gain or loss of an investor. This is potentially relevant because an investor with a large loss faces larger incentives to sell the share package before the 365-days cutoff because deducting a large loss reduces the tax base by more than a small loss. In addition, if the loss is only small the investor might want to wait and see if share package prices rise. Equivalently, a large gain would trigger a larger tax liability, which increases the incentive to sell a gain after the cutoff. Studying heterogeneity w.r.t. the magnitude of gains and losses also relates to Ivković et al. (2005) who provide a similar analysis. For this purpose, we include an additional variable into the above regressions which measures the percentage change in the value of the share package. In this context, we estimate the following regressions for each day of the holding period t:

$$Sell_{ijd} = \beta_0 + \beta_1 Change_{ijd} + \varepsilon_{ijd}, \tag{8}$$

where the *Change* variable describes the change between the share price at holding-day t and the purchasing ($\frac{p_{ijtd}-p_{ij0d}}{p_{ij0d}}$). To avoid that the regression results are driven by extreme outliers which could be caused by mistakes in the price databank, we exclude observations for which the price change is not included within the first and 99th percentile. The constant in these regressions describes the selling probability for share packages without a price change, $(\beta_0 + \beta_1)$ estimates the selling probability for changes of size 1, and β_1 measures the difference between the selling probabilities of share packages without any change and a large change of 1. We again estimate these models separately for losses and gains and pre and post reform periods. In our graphs, we plot the β_1 coefficients.

4.4.4 Taxes and the Disposition Effect

We aim to test if taxes affect the extent of the disposition effect. The starting point of the analysis is to measure the existence and magnitude of a potential disposition effect in our data. Following the literature (e.g., Chang et al. 2016) and using our previous hazard-rate regression framework, we regress a *Sale*-dummy (see above) on a dummy indicating whether a share package comes with a gain on this day of the holding period. Formally, we estimate the following regression for each day of the holding period and using all sample years and shares with both gains and losses:

$$Sell_{ijd} = \beta_0 + \beta_1 Gain_{ijd} + \varepsilon_{ijd}. \tag{9}$$

If β_1 is greater than zero in this regression, this is evidence of a disposition effect; i.e., gains are sold with larger probabilities than gains. The coefficient for β_1 measures the magnitude of the disposition effect. We plot these β_1 coefficients in our result graphs.

We estimate the above regression separately for pre-reform and post-reform years. Any difference between pre and post reform years, especially around the 365-days cutoff, sheds light on the tax effects of the disposition effect. The difference in the disposition effect between post-reform and pre-reform years can also be estimated in a DiD-type regression of the following form:

$$Sell_{ijd} = \beta_0 + \beta_1 Pre + \beta_2 \mathbb{1}(Gain_{ijd}) + \beta_3 Pre \times \mathbb{1}(Gain_{ijd}) + \varepsilon_{ijd}, \tag{10}$$

where Pre indicates years before the reform (when the holding period mattered for the tax liability). The interaction of pre-years and the gain dummy, β_3, measures the difference in disposition effect before the reform relative to after the reform. We again estimate this regression separately for each day of the holding period, which allows to check if the difference between post and pre years is particularly pronounced around the 365-days cutoff. We plot β_3 when we present the graphical results for this approach.

4.5 Results

This section presents the empirical results. All of our empirical findings are presented in graphs which aim to visualize the effects and make them approachable. The chapter is organized along the same order as the description of the empirical strategy in section 4.4.

4.5.1 Number of trades in weeks around the cutoff

Figure 22 depicts the number of traded *gains* (i.e., appreciated share packages) in weekly bins around the intertemporal (365-days) time discontinuity separately for pre-reform and post-reform years. The red vertical line at zero marks the 365-days holding period.

In pre-reform years, in which the 365-days cutoff was tax relevant, the number of gains that are sold spikes sharply in the first week after the 365-days cutoff. The number of sold gains in this first week after the discontinuity is more than 2.5 times as high as in the week before the 365-days cutoff. In week 2 after the discontinuity the number of sales is roughly 1.8 times as high as in the week before the reform. This trend then continues in subsequent weeks: the number of sold gains remains higher than before the cutoff, but the difference becomes smaller as we move further to the right from the cutoff.

Is the spike in the number of sold gains driven by the capital-gains tax discontinuity? In post-reform years, in which the 365-days cut-off is not tax relevant, we see a smooth development of the number of sales around 365 days. Specifically, the number of sold gains does not exhibit a spike just to the right of the discontinuity. This is clearly indicative that the large spike in pre-reform years is driven by the capital-gains tax system. We will compare the number of sales in pre and post years further below when quantify the tax effect using the bunching approach.

Figure 23 presents the equivalent plot for the number of sold *losses* (i.e., depreciated share packages). In pre-reform years (with tax-relevant 365-days cutoff), we see a sharp spike in the number of sold losses in the week just before the 365-days cutoff. The number of sold losers is more than 3 times as large in the week before the cutoff than in the week just after the cutoff. In week -2, the spike is still clearly visible but considerably smaller than in week -1; the number of sold losers is about 1.7 times larger in week -2 compared to the week after the cutoff. Importantly, we see a smooth development in the number of realized losses around the 365-days cutoff in post-reform years where crossing the 365-days holding period does not have any tax implications. The spike in pre-reform years, along with the absence of any spike in post-reform years, provides clear evidence that the tax discontinuity affects trading behavior.

Overall, our findings for both losses and gains are consistent with the notion that investors try to realize losses within the holding period that allows using them as a tax shield, whereas investors defer the realization of gains until they are tax free.

Number of distinct selling investors in weeks around the cutoff. The previous results showed the number of appreciated and depreciated share packages around the intertemporal tax discontinuity. Are the spikes in the number of sales around the discontinuity driven by a few investors who are tax aware and sell many of their share packages around the cutoff? We shed light on this question by plotting the number of *distinct* investors who trade in a given week. As before, we plot the weekly numbers separately for investors who trade gains and losses, as well as for pre and post reform years.

Figures 24 and 25 present the plots for gains and losses, respectively. In pre-reform years, the number of investors selling gains spikes sharply in weeks to the right of the discontinuity and the number of distinct investors trading losses spikes sharply in weeks to the left of the discontinuity. We see no spikes in post-reform years in which the 365-days cutoff is not tax relevant. The spikes in pre-reform years, along with the absence of spikes in post-reform years, again indicates a causal tax effect.

Overall, this exercise suggests that the sharp spikes in the number of share packages above is not driven by a few tax-sensitive investors selling many share packages around the discontinuity. Apparently, many different investors respond to the tax incentives in a way that is consistent with our expectations. We study different sources of potential heterogeneity in tax responses among different investors further below.

4.5.2 Difference-in-Bunching

The Difference-in-Bunching results are presented in Figures 26 (for gains) and 27 (for losses). As described in section 4.4, we use the post-reform years (without tax discontinuity) as the counterfactual distribution for the tax-affected pre-reform years. Recall that we divide the number of sales by the respective total number of share packages (including the ones which have not been sold) in order to account for level differences between pre and post reform years. in order to account for differences in levels across pre-reform and post-reform years. In the Figures, the vertical red line depicts the 365-days holding period and the blue and red line present the weekly bins for the pre- and post-reform periods, respectively. The patterns in both Figures are (not surprisingly) similar to the patterns that we saw above in the Figures that simply plot the number of sales. In particular, the density of realized gains spikes sharply in the week after the 365-days cutoff in pre-reform years and no such spike is observed in post-reform years. The density of realized losses has a large spike in the week before the cutoff in pre-reform years and, again, there is no spike in post-reform years. The main purpose of our Difference-in-Bunching approach is to quantify the magnitude of the tax effect and to estimate an elasticity of the holding duration with respect to the tax rate. In other words, we aim to calculate the percentage change in holding-period days in response to a one-percent change in the tax rate.

We estimate an excess mass of 2.32 (standard error: 0.07) for gains (see Figure 26) and an excess mass of 2.43 (standard error: 0.07) for losses (see Figure 26). To derive an elasticity, these excess-mass estimates can be related to the change in tax rates once the holding-period of 365 days is crossed. As we described in Section 4.2.1, the applicable tax rate for assets sold within the first 365 days after purchase is the individual personal marginal income-tax rate. Our portfolio data do not include personal marginal tax rates for investors and we therefore calculate the elasticity based on two scenarios: i) the top marginal income-tax rate of 42% applies, ii) the minimum tax rate (lowest bracket) of 15% applies.[67] Using the top marginal income tax rate of 42%, we estimate an elasticity of 0.185 for gains and an elasticity 0.195 for losses. Using the minimum income-tax rate of 15%, we estimate elasticities of 0.56 for gains and 0.59 for losses. Our estimates translate to a tax-induced change in the holding period of 16 days for gains and 17 days for losses.

Where does the excess mass come from? Are the spikes 'fed' by regions to the left or to the right of the cutoff? For gains, we see that the mass mostly comes from the left of the discontinuity; this suggests that investors delay the sales of gains until they qualify for tax exemption. For losses, we see that the mass of investors mostly comes from the right side of the discontinuity; this suggests that investors move forward the realization of sales in order to count them against their tax-relevant gains.

4.5.3 Hazard-Rate Regressions

We present the results of our main hazard-rate regressions in Figures 28 (for gains) and 29 (for losses). The red vertical line again indicates a holding period of 365 days. The blue line plots the daily-estimated coefficients for the selling probability of either gains or losses in pre-reform years. The red line plots the equivalent line of coefficients for post-reform years. That is, we plot the β_0 coefficients (i.e., the coefficients for the constant) of regression equation 6 in these Figures. The faded area around the line of coefficients indicates 95% confidence intervals.

Although we now use regression methods based on daily data, the results are very much consistent with the patterns that we saw in the preceding analyses. In particular, we see in pre-reform years that the probability to sell an appreciated share package spikes sharply during the holding-period days just after the 365-days cutoff, whereas the probability to sell depreciated share packages spikes sharply during the days just before

[67]The top marginal income tax rate and minimum tax rate were 42% and 15% during most of the years in our data sample. Note that most of the investors in our data are likely to be high earners and closer to the top rate than to the minimum rate.

the 365-days holding period. We do not see any spikes in selling probabilities around the time discontinuity in post-reform years.

The magnitudes of the spikes are considerably large. As Figure 28 shows, the probability to sell a gain on a given day of the holding period jumps from around 0.02 during the days before a 365-days holding period to approximately 0.07 on the day after the cutoff. No such jump is observed in the post-reform years, again indicating that the tax incentives have a clear effect on trading behavior. Comparing pre- and post-reform hazard rates further away from the cutoff, the Figure suggests that investors indeed defer the realization of gains until they qualify for preferential tax treatment; the pre-reform selling probabilities tend to be below the post-reform probabilities to the left of the discontinuity, and then sit above the post-reform probabilities on days after the 365-days holding period. While plausible, this observation rests on a comparison of different time periods and should therefore be viewed with caution.

For losses, as shown in Figure 29, the jump is even more considerable than for gains; the selling probability is below 0.02 during the days after the 365-days holding period and stands at almost 0.08 on day 364. Along with an absent jump in the post-reform period, this is further evidence that the tax discontinuity induces investors to realize their losses as long as they can be used to offset gains. Comparing pre- and post-reform hazard rates further away from the cutoff, the Figure is suggestive that investors reduce the holding period of losses for tax reasons. The hazard-rates to the right of the cutoff tend to be higher in post-reform years than in pre-reform years. This difference in probabilities could 'feed' the spike to the left of the discontinuity. Again, comparing pre- and post-reform years away from the cutoff rests on a comparison of different time periods and should be treated with caution.

Hazard-Rate Regressions by Year. To shed light on the yearly dynamics and to examine if a few exceptional years drive our main results above, we estimate the daily hazard-rate regressions separately for each year in our data sample. The resulting Figures, which are to be interpreted as our main hazard-rate Figures above, are presented in Appendix Figures 49 to 58. Each of these Figures presents the hazard-rate results for three consecutive years. We again estimate the hazard rates separately for gains and losses. To make all yearly Figures comparable, the scale of the y-axis is held constant across all Figures.

The results for gains in pre-reform years (i.e., where 365 days holding preiod was tax relevant) are presented in Figures 49 to 51. In 1999, the first year of our sample, the spike to the right of the 365-days discontinuity is clearly visible and very pronounced. This spike becomes smaller in the years 2000-2002, though it remains clearly visible across this time period. The smaller magnitude of the spike in these years is reasonable given that gains were less prevalent during the burst of the *dot.com* bubble and many investors presumably had losses that they could use to offset gains and which made it less necessary to sell gains in the tax-free domain. The spike to the right of the discontinuity is then again large and very prevalent in the years 2003-2007.[68] Overall, we see a spike in selling probabilities to the right of the tax-relevant time discontinuity in all pre-reform years of our sample period. The results for gains in post-reform years (i.e., where the 365-days cutoff is not tax relevant anymore) are presented in Figures 52 and 53. We do not see any spikes or irregularities in selling probabilities around the holding period of 365 days in any of the six post reform years.

The results for losses in pre-reform years are shown in Figures 54 to 56. We observe clear and substantial spikes in selling probabilities just before the 365-days holding period in each pre-reform year (1999-2007). The smallest spike is observed in 2003, but it is clearly visible even in this year. The results for losses in post-reform years are shown in Figures 57 and 58. As with gains, we do not observe spikes or irregularities around the 365-days holding period in any of the six post-reform years

Overall, selling probabilities of both gains and losses spike around the 365-days holding period in all pre-reform years, but we do not see spikes in any of the post-reform years. We interpret this finding as clear evidence that the tax discontinuity affects trading behavior.

Heterogeneity w.r.t. investor characteristics. We study heterogeneity with respect to three different investor characteristics: age, investor experience (both measured in years) and gender (dummy indicating males). The underlying regression models condition on a set of other investor-level characteristics (see section 4.4.3).

Figures 30 and 31 depict the effect of an additional age year on selling probabilities on each day of the holding period. We particularly see age heterogeneity in the context of loss-selling behavior (see Figure 31). The likelihood of selling a loss shortly before the discontinuity sharply increases in age in pre-reform years. That is, older workers are more likely to sell gains for tax reasons. We see no such effect in the post-reform years in which the discontinuity is abolished. Age heterogeneity is not very pronounced in the context of gains and we cannot conclude from the data that older and younger investors respond differently to the tax discontinuity when it comes to selling gains. In addition, we see no difference in selling probabilities between older and younger investors for holding-period-days further away from the discontinuity (this goes for both losses and gains). Importantly, all our age effects are conditional on our measure of experience; that is, they are not confounded by trading experience.

[68]Note that we do not present results for the year 2008 because share packages bought before 2009 are excluded on calendardates later than 12/31/2008 (see section 4.4.3 for a detailed discussion). Hence, even a share package bought on 01/01/2008 is not reaching the 365 days holding period threshold.

Figures 32 (for gains) and 33 (for losses) illustrate the coefficients for investor experience. The result is unambiguous for both losses and gains: experienced investors react stronger to the tax. This is reflected in the finding that selling probabilities around the discontinuity sharply increase with each year of experience in pre-reform periods. In other words, the probability to sell a stock for tax purposes around the time discontinuity increases in trading experience. Further distant to the discontinuity, we do not see any significant effects of experience on the probability of selling gains, neither in pre nor in post reform years. This is different for losses: experienced traders are more likely to sell losses throughout the entire set of holding-period days before the discontinuity in pre-reform years. This difference disappears for days to the right of the discontinuity. Note, again, that these effects of experience are conditional on age of the investor.

Heterogeneity with respect to gender is plotted in Figures 34 (for gains) and 35 (for losses). We do not see any conclusive evidence for gender heterogeneity in the context of gains. For losses, we see a large negative spike just before the discontinuity in pre-reform years. This finding indicates that men are less likely to sell their losses on the day before the cutoff, implying that men are less tax responsive in the context of loss realizations.

Heterogeneity w.r.t. magnitude of gains and losses. Figures 36 and 37 plot the β_1 coefficients of regression equation for each day of the holding period around the 365-days discontinuity. These Figures shed light on the question of whether responses to the tax depend on the magnitude of the loss or gain. The pronounced spike in the blue line in Figure 36 just after the one year threshold implies that investors become much more likely to dispose those stocks which had the largest gains. This effect then levels off over the subsequent weeks. The pattern disappears completely once the flat tax regime is introduced (red line). The relationship is similar but even stronger for the size of losses: The strong decrease of the blue line in Figure 37 in the three weeks prior to the one year threshold implies that investors become much more likely to dispose of those stocks which have performed the worst.[69] Apparently, the last opportunity to at least preserve some additional value in the form of a tax shield gives an extra impetus to dispose of the more extreme loss makers. This feature may be particularly valuable from an optimal investment perspective because investors are in general more hesitant to dispose of the largest losses as implied by the coefficient plots in the positive range in Figure 37 after the reform (red line) and before the reform (blue line) – except, as discussed for the blue line, for the last few weeks before the one year threshold.

[69]Losses are measured as negative values. Hence, a negative coefficient corresponds with an increased likelihood to dispose of larger losses.

4.5.4 Taxes and the Disposition Effect

Figure 38 plots the disposition effect on each day of the holding period separately for pre- and post-reform years. That is, we plot the β_1 coefficients of regression equation 9. In the absence of a tax discontinuity in post-reform years, we observe the disposition effect on each day of the holding period. That is, the probability to sell gains is higher one each day of the holding period than the probability to sell losers. This result is consistent with the literature where the disposition effect has been shown to be very robust. How does the magnitude of our disposition effect compare to estimates in the literature? According to the overview handbook chapter by Barber and Odean (2013), the selling probability of gains is about 20-70% higher than that of losses. To make our estimates comparable with these numbers, we divide the sales probability of gains by the sale probability of losses. Technically, this means we use coefficients from regression equation 9 and divide the coefficient of the gain dummy, by the constant which indicates the probability to sell a loss, for each day of the holding period. The results of this exercise are plotted in Figure 39 (that is, Figure 39 plots the ratio β_1/β_0). For the purpose of comparing our disposition effect to the estimates in the literature, we mostly consider the post reform period (without tax discontinuity) because, as we see below, the disposition effect in the pre reform period is heavily affected by the intertemporal tax discontinuity. On average over the entire holding period of days 185-545 in the post period, we observe that the probability to sell a gain is 67% higher than the probability to sell a loss. This finding is well in line with the findings in the literature.

Looking at pre-reform years with the tax relevant time discontinuity in Figure 38, it is clearly visible that the disposition effect is affected by the capital-gains taxes. To the left of the 365-days cutoff the disposition effect is first reduced and then steadily drops. The disposition effect then turns negative during the days before the cutoff and exhibits a sharp negative spike on the last day before the 365-days holding period is reached. This reveals that the desire to sell losers before the cutoff for tax reasons is larger than the disposition to sell gains with larger probability than losses. The pattern is reversed as we explore the days just after the 365-days cutoff. The disposition effect is heavily accelerated as compared to its usual magnitude; we see a substantial spike in selling probabilities of gains during the days after the cutoff. On subsequent days, the disposition effect remains higher than usually and it takes about 35 holding-period days to go back to the usual level. The findings are consistent with investors selling gains once they are tax free.

Figure 38 provides clear evidence that the disposition effect is affected by the tax around the days of the discontinuity. Does the tax discontinuity also impact the magnitude of the disposition effect on holding-period days more distant to the cutoff? To shed light on this question, we require a benchmark against which the disposition effect away from the cutoff can be compared. We use the post-reform periods (without discontinuity) as the benchmark. This exercise obviously relies on the assumption that the post-reform disposition effect is a good counterfactual for the pre-reform years. Effectively, comparing the post- and pre-reform years is a before-after comparison that should naturally be viewed with caution. Having this caution in mind, the Figure indicates that, away from the cutoff, the disposition effect tends to be lower during the first year of the holding period and higher after 365 days holding period have passed. This suggests that the time discontinuity affects the disposition effect even on holding-period days distant to the discontinuity.

All above results are also visible in Figure 40 which plots the coefficients of the DiD set-up (β_3 in equation 10). These coefficients compare the disposition effect between pre-reform and post-reform years. The Figure particularly confirms that the days around the discontinuity are substantially different between post and pre years, and additionally adds to the suggestive evidence that the disposition effect is affected by the discontinuity even on holding-period days away from the discontinuity.

4.6 Concluding Remarks

In this paper, we contribute to a better understanding of the role of capital-gains taxes for the stock-market trading behavior of private investors. We provide causal evidence on two interrelated questions: i) How do capital-gains taxes affect the holding period of private stock market investments? ii) How do taxes interact with disposition effect? The literature has addressed certain aspects of these two questions, but the evidence is surprisingly limited. The lack of evidence is presumably attributable to the challenge of finding appropriate micro level data on trading behavior in combination with an institutional set up that allows for identification of causal tax effects. Our paper overcomes this challenge in that it combines high-frequency portfolio-level data (which we confidentially obtained from a large German bank) with an intertemporal tax discontinuity, and its abolishment, in the German capital-gains tax system.

Our findings provide clear evidence that capital-gains taxes affect the trading behavior of individual investors. Selling probabilities, which we estimate on a daily basis, are heavily affected by the tax discontinuity and disappear in years after the abolishment of the discontinuity. Interesting patterns of heterogeneity reveal that more experienced and older investors respond stronger to the tax incentives.

We also find that the disposition effect – the tendency to sell gains with a larger propensity than losses – is strongly affected by capital-gains taxes. Depending on the type of sale – gain or loss – the disposition effect is accelerated or mitigated due to the tax. Previous studies have found that more experienced and older investors exhibit smaller disposition effects (e.g., Feng and Seasholes 2005, Dhar and Zhu 2006 and Seru et al. 2010). However, as our heterogeneity analyses suggest, this is not an intrinsic direct effect of age or experience. We find that it is salient intertemporal tax discontinuities which induce the more experienced investors to dispose of their loss-making positions. When the salient tax discontinuities are removed, there is no difference in the probability to dispose of losses anymore between more or less experienced investors or older and younger investors. This implies that, in the absence of the discontinuity, the disposition effect is not different between older and younger and between more and less experienced investors. Hence, if the U.S. were to smoothen the tax schedule for capital gains, the seemingly stronger resistance of more experienced (or older) investors to behavioral biases may disappear as well because it is the time discontinuity in the tax schedule which helps these types of investors to focus their minds / make up their minds on loss-making positions.[70]

How do our results relate to the predictions from theoretical models such as Constantinides (1984)? First, our results are consistent with theory in that we see that the discontinuity induces investors to delay the sale of gains until they qualify for preferential tax treatment and to realize losses earlier, both relative to a counterfactual without intertemporal tax discontinuity. Second, the sharp spike in selling probabilities of losses shortly before the discontinuity is not necessarily consistent with standard theoretical predictions. However, this result is consistent with the notion that the discontinuity serves as a self-control device that commits loss averse investors to take care of their losses. The idea of a self-control mechanism to realize losses was first developed by Shefrin and Statman (1985). According to this idea, investors are reluctant to realizing losses and they need some nudge to overcome this reluctancy and sell their losses. Our results show that investors do not realize losses as they accrue and instead wait until the quickly approaching discontinuity nudges them to realize the loss. To this end, our paper provides some indication that taxes can serve as a commitment device for investors with behavioral biases such as loss aversion.

[70]A complete smoothening of the tax schedule in the U.S. would imply not only the same tax rate on short and long run capital gains but also a loss carry-back option for the deductibility of capital losses against ordinary income or an abolishment of the deductibility against ordinary income.

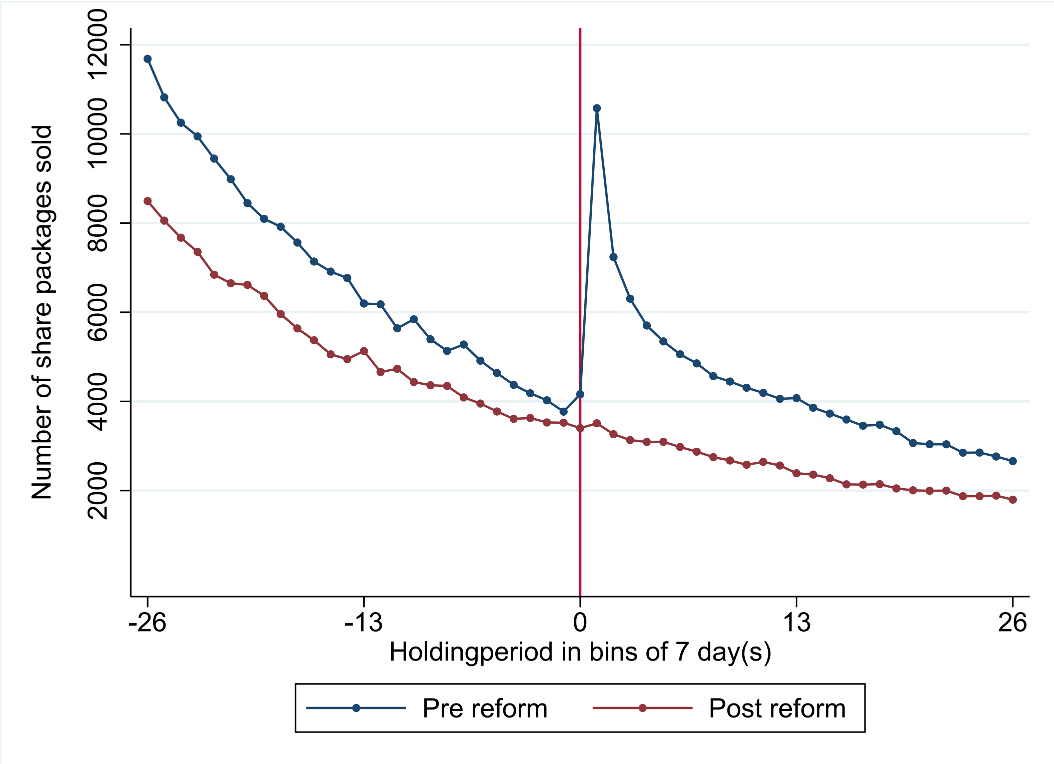
Figures and Tables

Table 9: Descriptive statistics for all investors in the sample

	N	Mean	Std.Dev.	Min	Max
Unit of observation: Investor					
Birthyear	93186	1962.74	13.23	1905	2010
Age end of 2015	93186	52.26	13.23	5	110
Trading experience in years end of 2015	93186	13.52	4.28	-0	22
Male	93186	0.86	0.35	0	1
Works in financial sector	93186	0.06	0.24	0	1
Self-employed	93186	0.16	0.36	0	1
Wealth $\leq 30,000$	93186	0.20	0.40	0	1
Wealth $> 30,000 < 100,000$	93186	0.19	0.40	0	1
Wealth $\geq 100,000$	93186	0.07	0.25	0	1
Wealth information missing	93186	0.54	0.50	0	1
Income $\leq 40,000$	93186	0.15	0.36	0	1
Income $> 40,000 < 100,000$	93186	0.30	0.46	0	1
Income $\geq 100,000$	93186	0.04	0.19	0	1
Income information missing	93186	0.51	0.50	0	1
Holding a PhD	93186	0.06	0.24	0	1
Number of trades	93186	77.79	218.29	0	19877
Number of trades 0.5-1.5 years	93186	11.27	24.87	0	876
Average monthly turnover	93109	10.86	15.39	0.00	99.66
Average monthly turnover < 2009	82618	11.80	16.13	0.00	99.41
Average monthly turnover ≥ 2009	87319	9.05	16.12	0.00	100.00
Average portfolio value	93109	51726	239157	0.03	57774533
Average percentage gain per trade	81688	32.63	27.61	0.00	263.64
Average percentage loss per trade	78926	-31.49	18.99	-96.83	-0.01
Average gain (EUR) per trade	86486	9.23	658.07	-5429.97	5345.57
Unit of observation: Share package					
Sale in December	7248978	0.08	0.27	0	1
Sale in December: Gain	3925440	0.07	0.26	0	1
Sale in December: Loss	3323538	0.08	0.27	0	1

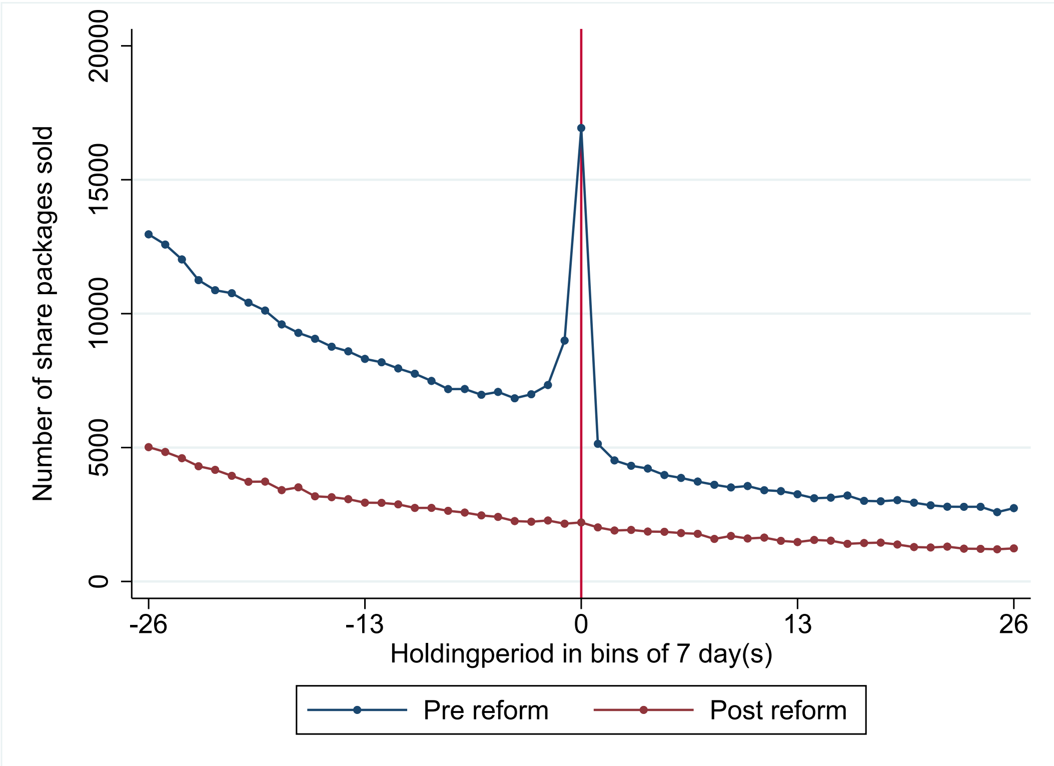
Notes: The table depicts the summary statistics for all variables used in our analysis. Variables are defined as follows: Birthyear is the birth year of the investor; Age and trading experience end of 2015 are the age and the trading experience measured by the number of years the investor has a depot at that bank at 12/31/2015; Male, works in the financial sector, holding a PhD and self-employed are dummy variables information comes from the MiFID documentation; Wealth $\leq 30,000$, Wealth $> 30,000 < 100,000$; Wealth $\geq 100,000$ and Wealth missing are 4 mutually exclusive wealth dummies indicating whether the investor belongs to one of the respective wealth groups. Income $\leq 40,000$, Income $> 40,000 < 100,000$, Income $\geq 100,000$ and Income information missing are 4 mutually exclusive income dummies indicating whether the investor belongs to one of the respective income groups. The information for wealth and income stems from the MiFID documentation and is self-reported. Number of trades is the investor average of the total number of share packages (see section 4.3.2 for a definition) sold; Number of trades 0.5-1.5 years is the investor average of the total number of share packages sold with holding periods in between 185 and 545 days. Average monthly turnover is the investor average of the average monthly portfolio turnover. Monthly portfolio turnover is calculated as in Barber and Odean (2001) as one-half of the monthly sales turnover plus one-half the monthly purchase turnover. Sales (purchase) turnover is defined as value of shares sold (purchased) divided by the portfolio value in the beginning of the month. Average monthly turnover < 2009 and average monthly turnover ≥ 2009 show the average monthly turnover for monthes prior and after January 2009 respectively. Average portfolio value is the investor average of the average monthly portfolio value as of end of the month. Average percentage gain, average percentage loss and average gain per trade are the investor average of the average gain (loss) of share packages sold by the investor. Sale in December, Sale in December: Gain and Sale in December: Loss show how many of the sold share packages have been sold in December.

Figure 22: Number of share packages sold around the time discontinuity: Gains



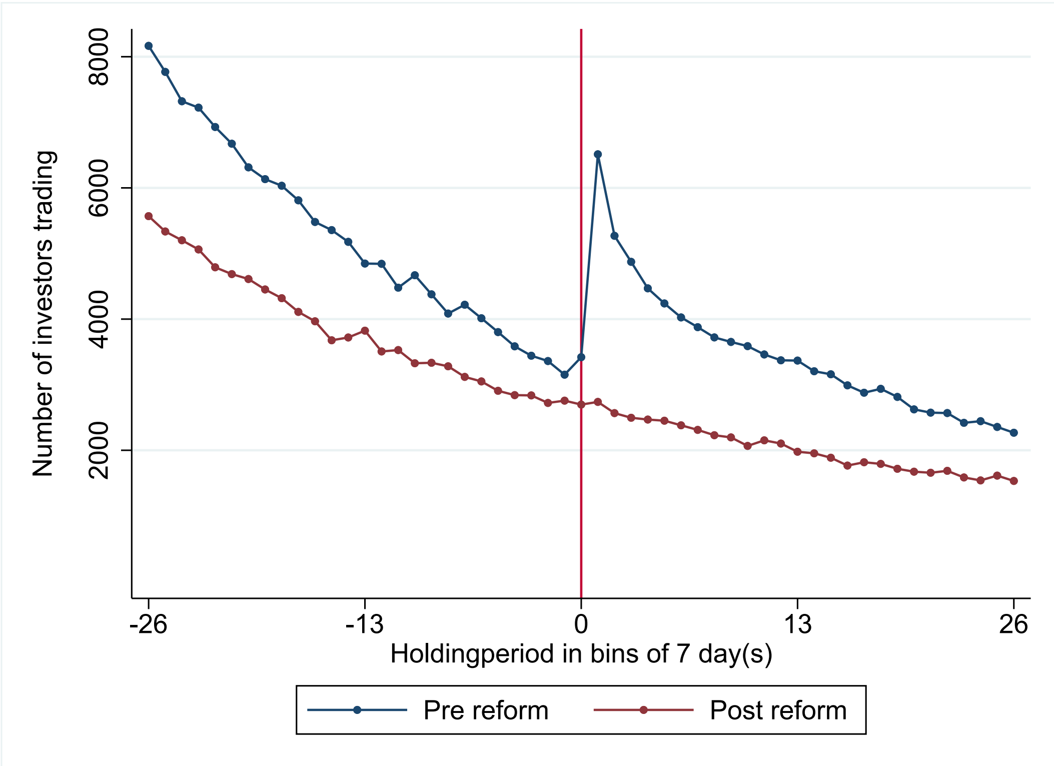
Notes: This figure displays the number of share packages which were sold with a gain in dependency of the holding period. Each dot represents the number of share packages sold in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last week in which gains were taxable. The dotted blue line represents sold share packages for which the purchase was made prior to 2009. The dotted red line represents sold share packages for which the purchase was made after 2009. The vertical red line at x-axis value zero marks the last week in which gains were taxable. Pre reform estimates are based on 44110 investors and 296135 share packages. Post reform estimates are based on 30875 investors and 206263 holding period share packages.

Figure 23: Number of share packages sold around the time discontinuity: Losses



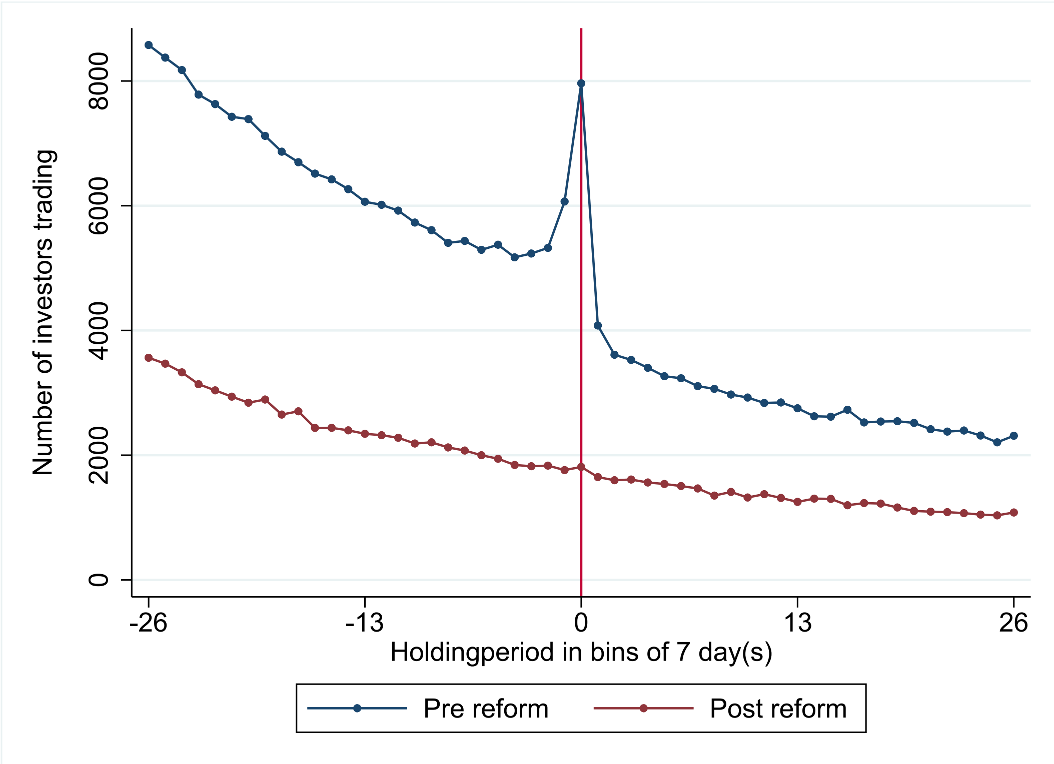
Notes: This figure displays the number of share packages which were sold with a loss in dependency of the holding period. Each dot represents the number of share packages sold in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last week in which losses could be used to offset gains. The dotted blue line represents sold share packages for which the purchase was made prior to 2009. The dotted red line represents sold share packages for which the purchase was made after 2009. The vertical red line at x-axis value zero marks the last week in which losses could be used to offset taxes. Pre reform estimates are based on 43008 investors and 339970 share packages. Post reform estimates are based on 23757 investors and 126280 share packages.

Figure 24: Number of distinct investors trading around time discontinuity: Gains



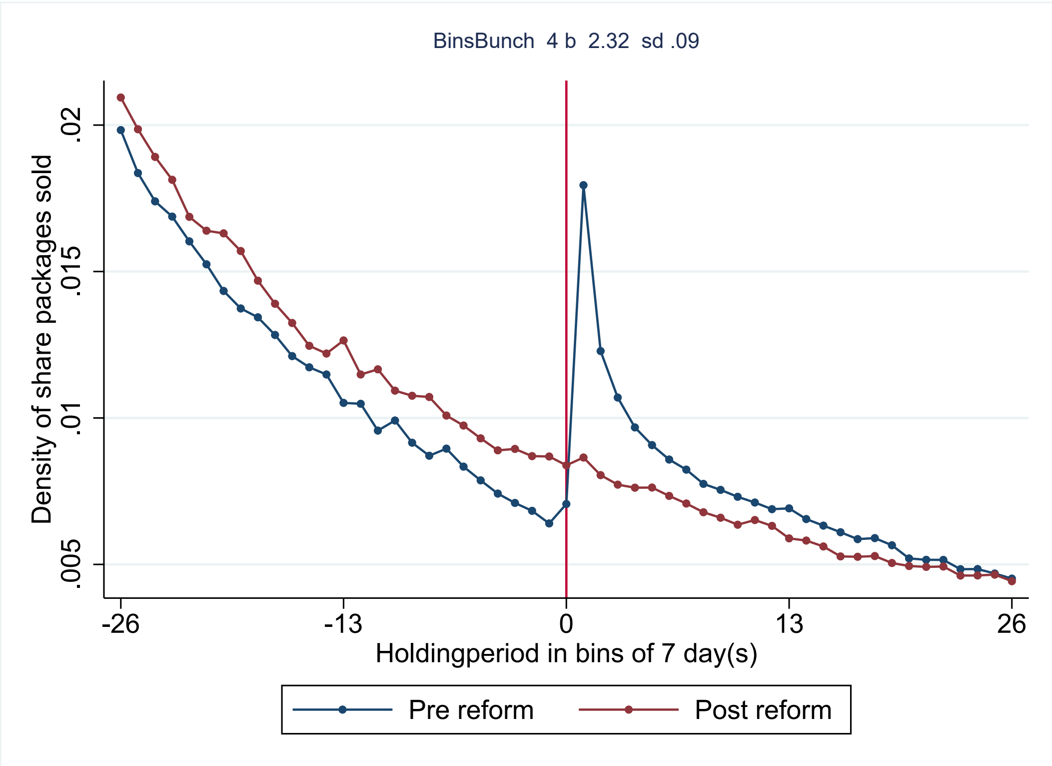
Notes: This figure displays the number of investors who sold an appreciated share package with the respective holding period. Each dot represents the number of investors who sold a share package in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last week in which gains were taxable. The dotted blue line represents the number of investors who sold share packages for which the purchase was made prior to 2009. The dotted red line represents the number of investors who sold share packages for which the purchase was made after 2009. The vertical red line at x-axis value zero marks the last week in which gains were taxable. Pre reform estimates are based on 44110 investors and 230352 share packages. Post reform estimates are based on 30875 investors and 155603 share packages.

Figure 25: Number of distinct investors trading around time discontinuity: Losses



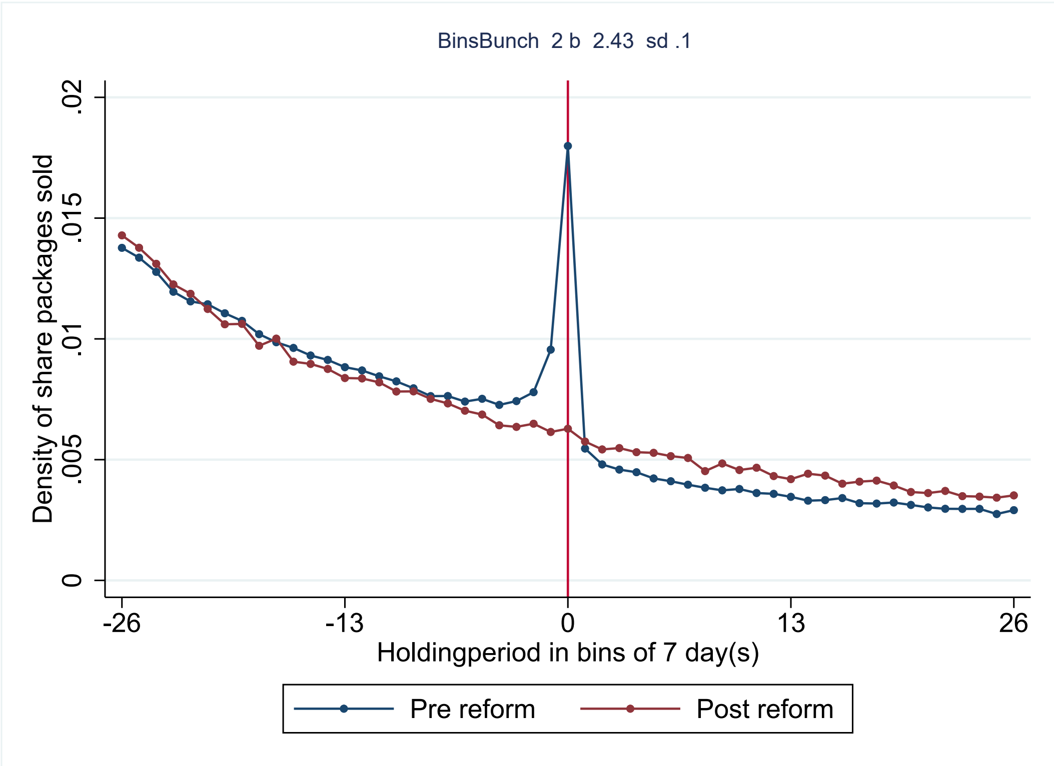
Notes: This figure displays the number of investors who sold a depreciated share package with the respective holding period. Each dot represents the number of investors who sold a share package in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last week in which losses could be used to offset gains. The dotted blue line represents the number of investors who sold share packages for which the purchase was made prior to 2009. The dotted red line represents the number of investors who sold share packages for which the purchase was made after 2009. The vertical red line at x-axis value zero marks the last week in which losses could be used to offset gains. Pre reform estimates are based on 43008 investors and 339970 share packages. Post reform estimates are based on 23757 investors and 126280 share packages.

Figure 26: Difference in bunching: Gains



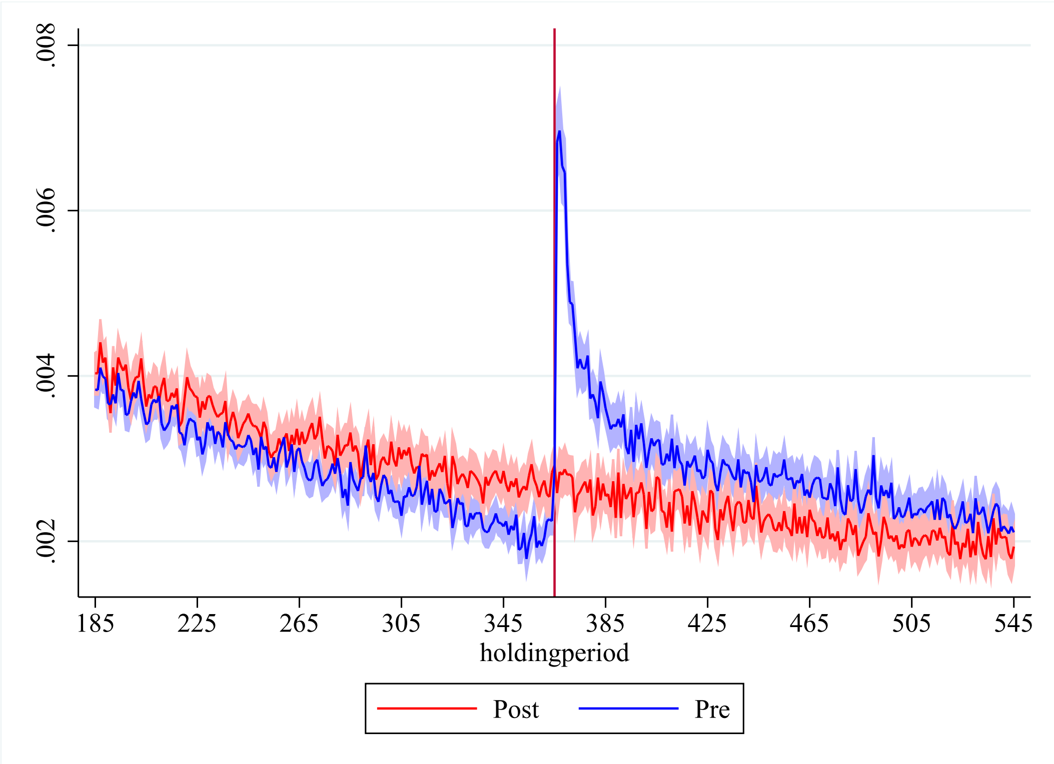
Notes: This figure displays the share of all purchased share packages with a gain in dependency of the holding period. Each dot represents the share of all purchased share packages with a gain which were sold in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last week in which gains were taxable. The dotted blue line represents the share of all share packages with a gain purchased prior to 2009 which were sold. The dotted red line represents the share of all share packages with a gain purchased after 2009 which were sold. The vertical red line at x-axis value zero marks the last bin in which gains were taxable. **BinsBunch** denotes the bunching window which in this case includes the 4 bins right after the last week in which losses could be used to offset gains. **b** represents the excess mass and **sd** the standard errors which are bootstrapped on the investor level. Pre reform estimates are based on 57944 investors and 589254 share packages. Post reform estimates are based on 43584 investors and 405628 share packages. These numbers include share packages of shares which have not been sold in the 26 weeks after the cutoff.

Figure 27: Difference in bunching: Losses



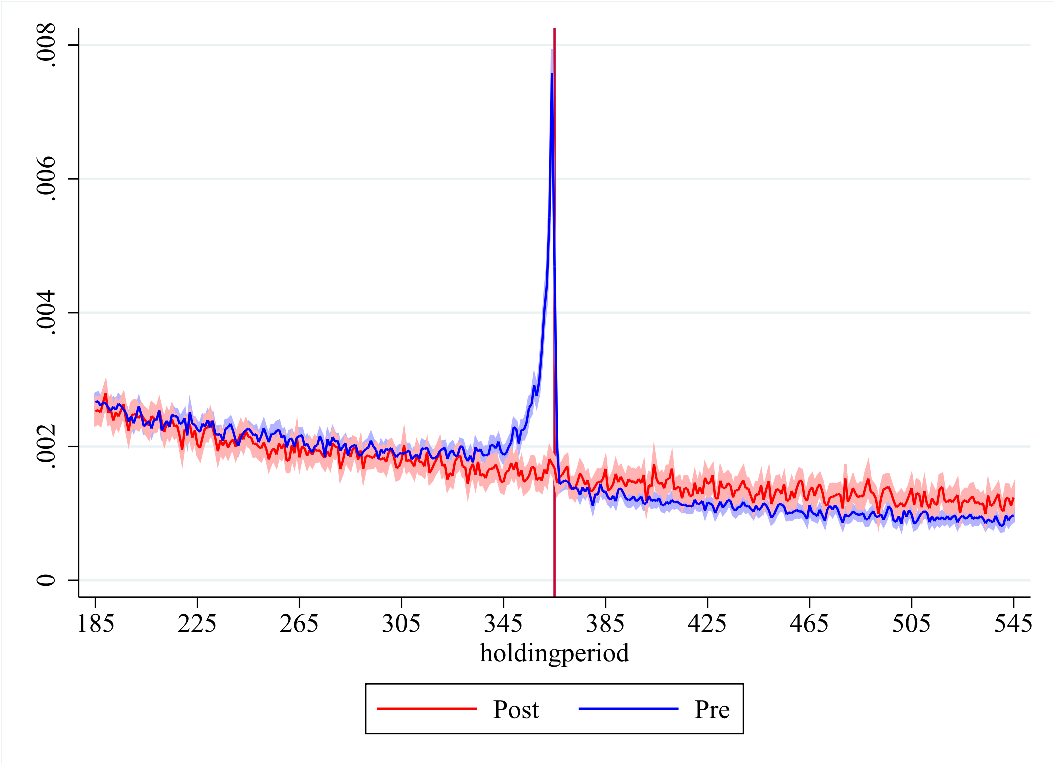
Notes: This figure displays the share of all purchased share packages with a loss in dependency of the holding period. Each dot represents the share of all purchased share packages with a loss which were sold in a 7 days bin of the holding period. Data is shown for 26 weeks before and 26 weeks after the last bin in which losses could be used to offset gains. The dotted blue line represents the share of all share packages with a loss purchased prior to 2009 which were sold. The dotted red line represents the share of all share packages with a loss purchased after 2009 which were sold. The vertical red line at x-axis value zero marks the last bin in which losses could be used to offset gains. **BinsBunch** denotes the bunching window which in this case includes the last week in which taxes were taxable and the week before. **b** represents the excess mass and **sd** the standard errors which are bootstrapped on the investor level. Pre reform estimates are based on 66396 investors and 941351 share packages. Post reform estimates are based on 43196 investors and 351090 holding period share packages. These numbers include share packages of shares which have not been sold in the 26 weeks after the cutoff.

Figure 28: Hazard-Rate Regressions: Gains



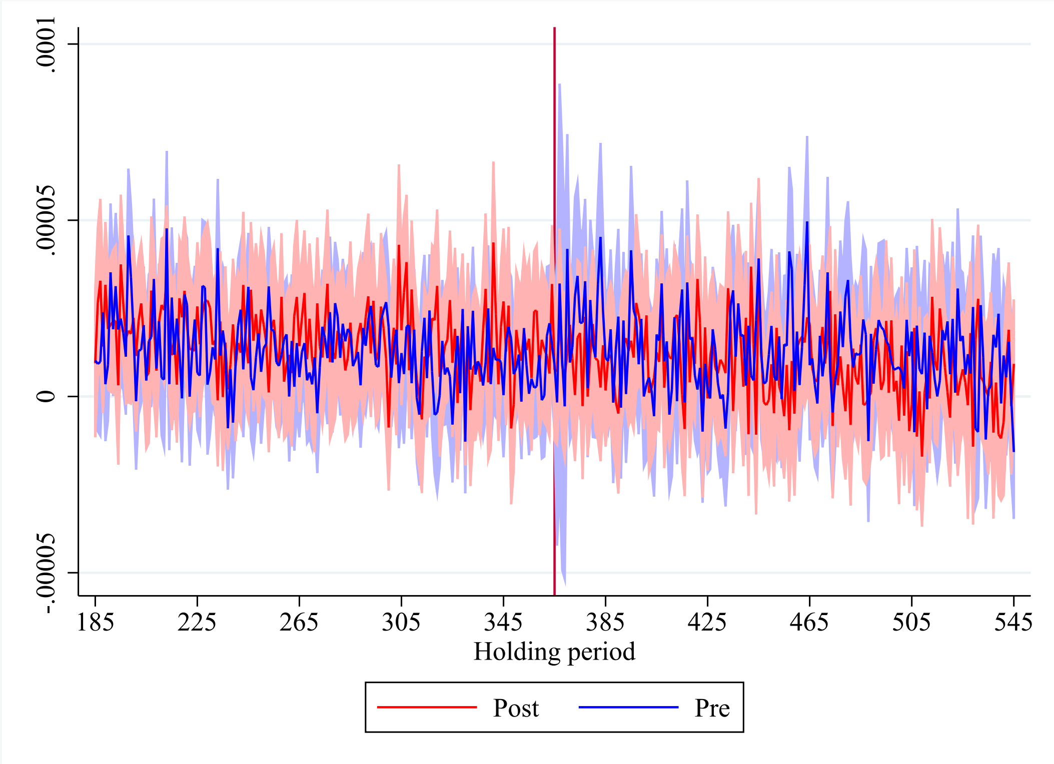
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod for share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. The blue line represents estimates for β_0 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_0 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 63743 investors and 91 million holding period share package observations. Post reform estimates are based on 51244 investors and 72 million holding period share package observations.

Figure 29: Hazard-Rate Regressions: Losses



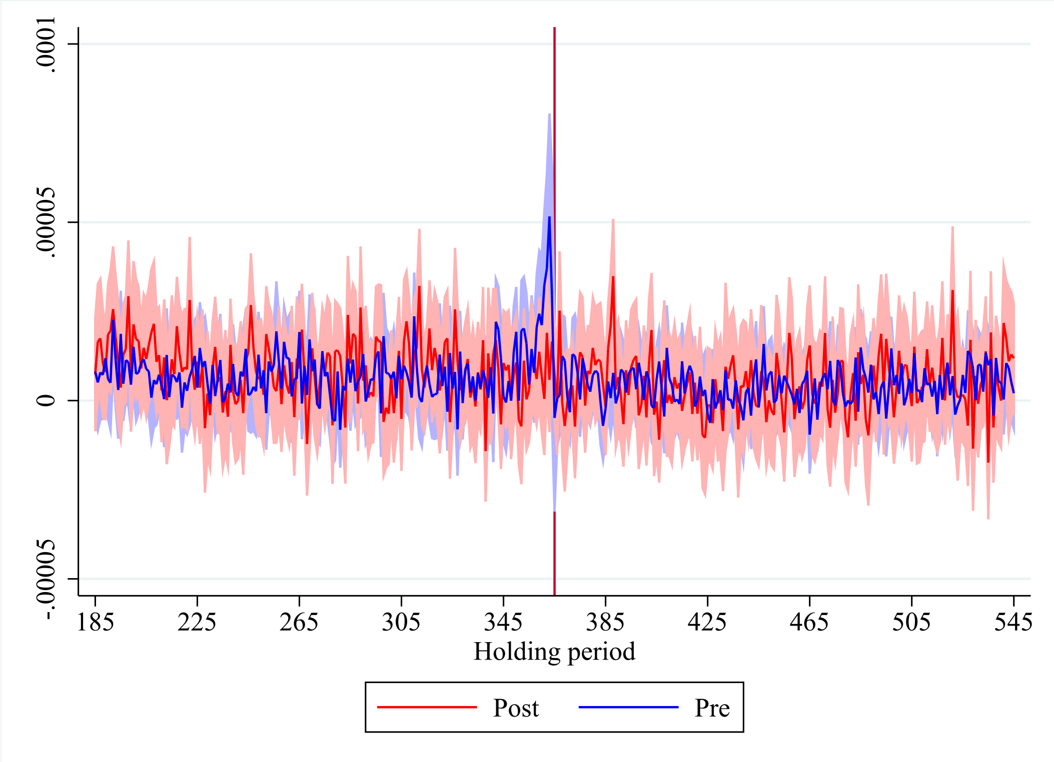
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod for share packages with prices below the purchase price at the respective day. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. The blue line represents estimates for β_0 for share packages which were bought before 2009. The shaded area displays 95 percent confidence intervals. The red line represents estimates for β_0 for share packages which were bought after 2009. The shaded area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which losses could be used to offset gains. Pre reform estimates are based on 70783 investors and 176 million holding period share package observations. Post reform estimates are based on 52290 investors and 76 million holding period share package observations.

Figure 30: Heterogeneity w.r.t. Age: Gains



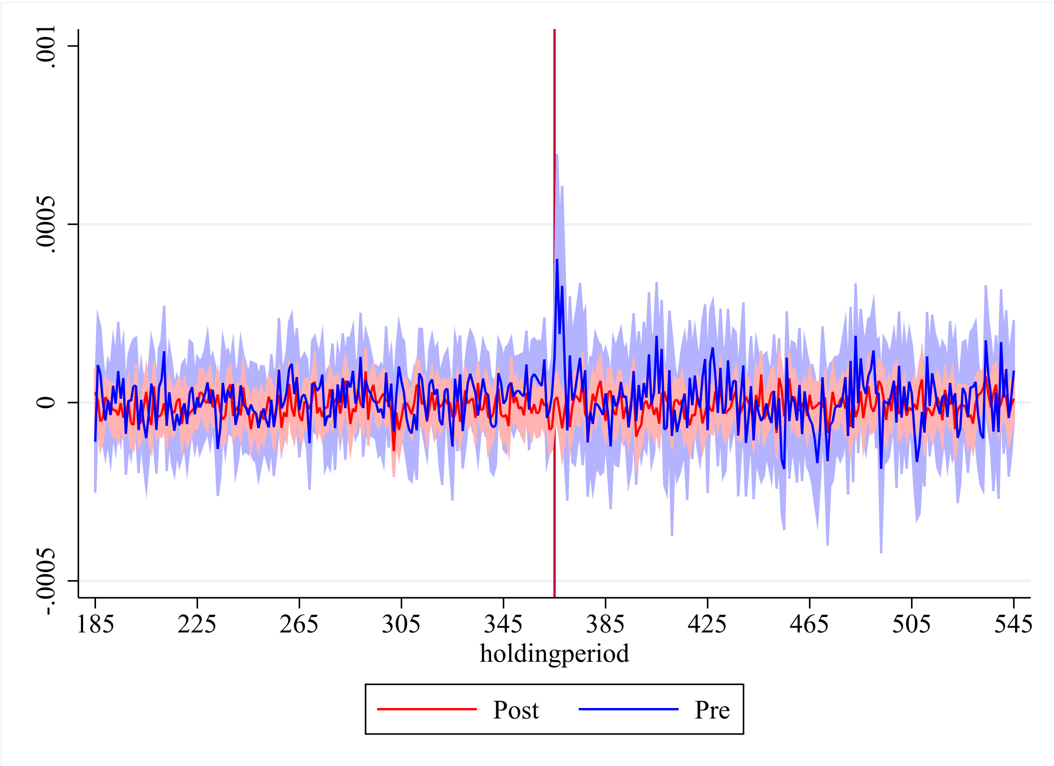
Notes: This figure displays coefficient estimates for investor experience stemming from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices above the purchase price. Coefficients indicate by how much an additional year in age shifts the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Age_{id} + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Where Age is the age of the investor on a given calendardate. $Covariates$ include controls for experience, gender, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 63743 investors and 91 million holding period share package observations. Post reform estimates are based on 51244 investors and 72 million holding period share package observations.

Figure 31: Heterogeneity w.r.t. Age: Losses



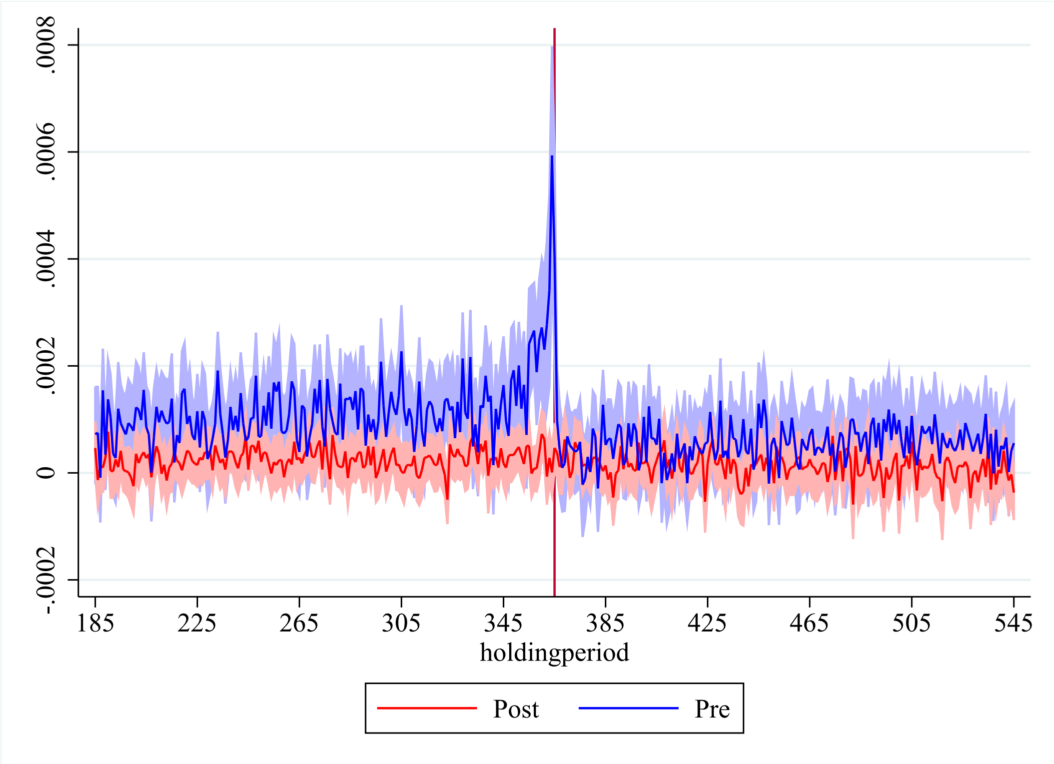
Notes: This figure displays coefficient estimates for investor experience stemming from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices below the purchase price. Coefficients indicate by how much an additional year in age shifts the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Age_{id} + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Where *Age* is the age of the investor on a respective calendar-day date. *Covariates* include controls for experience, gender, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 70783 investors and 176 million holding period share package observations. Post reform estimates are based on 52290 investors and 76 million holding period share package observations.

Figure 32: Heterogeneity w.r.t. Experience: Gains



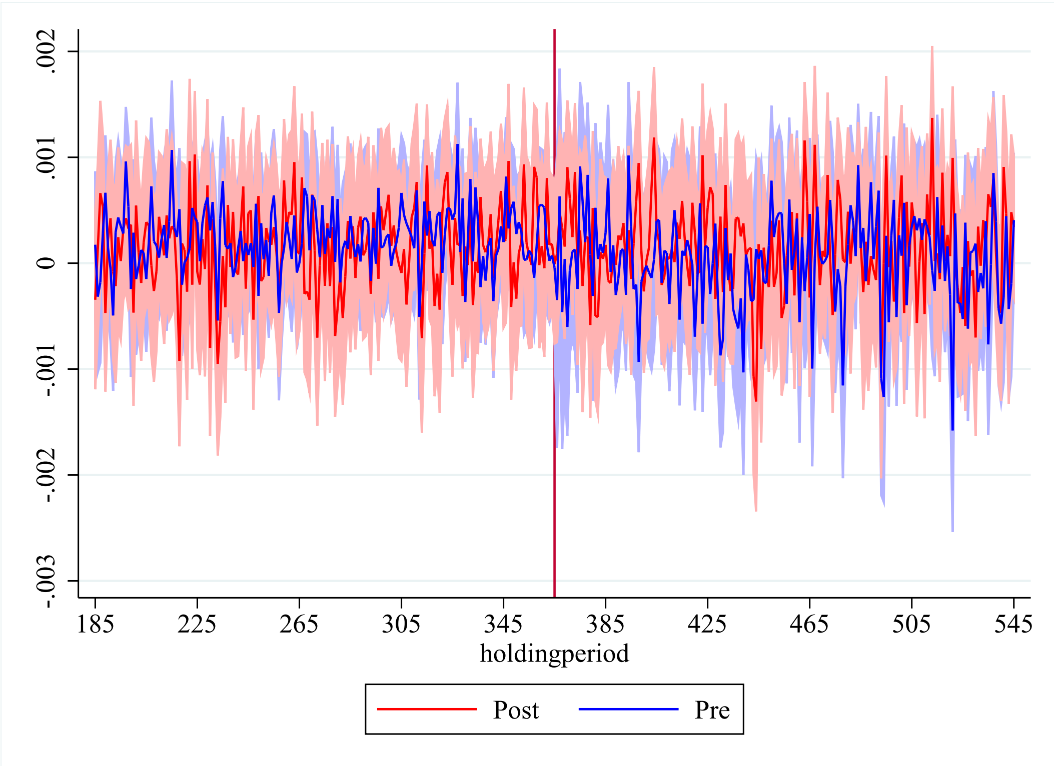
Notes: This figure displays coefficient estimates for investor experience stemming from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices above the purchase price. Coefficients indicate by how much an additional year in experience shifts the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Exp_{id} + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Where Exp is measured by the number of years the investor has a depot at that bank. *Covariates* include controls for age, birthyear (i.e. cohort), gender, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 63743 investors and 91 million holding period share package observations. Post reform estimates are based on 51244 investors and 72 million holding period share package observations.

Figure 33: Heterogeneity w.r.t. Experience: Losses



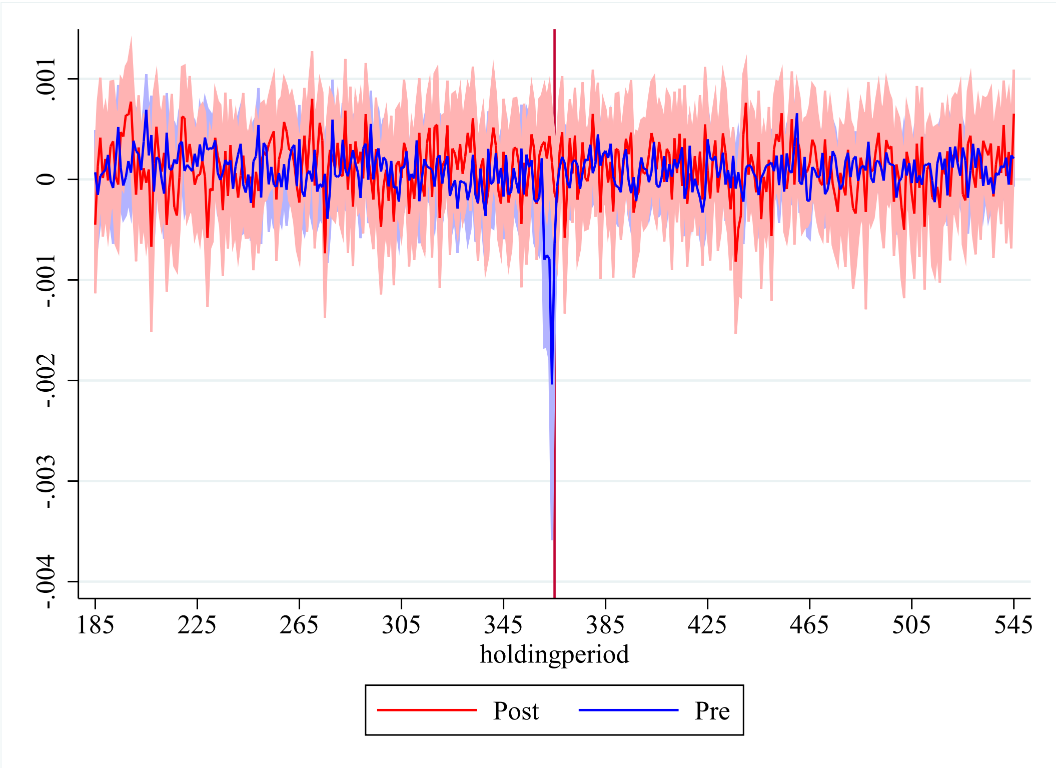
Notes: This figure displays coefficient estimates for investor experience stemming from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices below the purchase price. Coefficients indicate by how much an additional year in experience shifts the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Exp_{id} + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Where Exp is measured by the number of years the investor has a depot at that bank. $Covariates$ include controls for age, birthyear (i.e. cohort), gender, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 70783 investors and 176 million holding period share package observations. Post reform estimates are based on 52290 investors and 76 million holding period share package observations.

Figure 34: Heterogeneity w.r.t. Gender: Gains



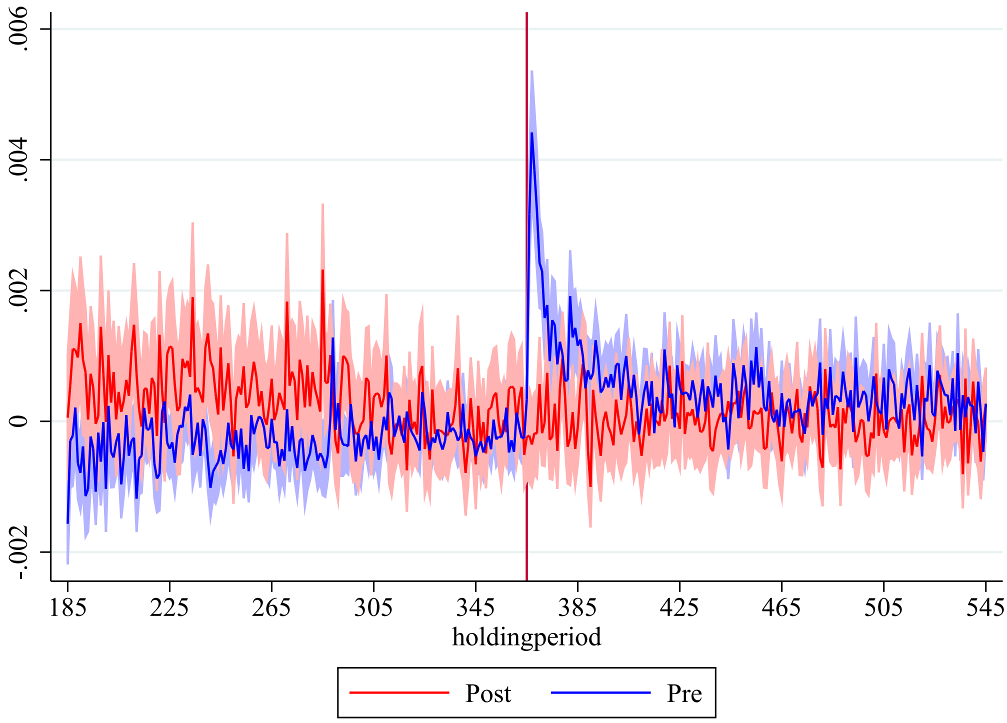
Notes: This figure displays coefficient estimates for a male dummy in the hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices above the purchase price. Coefficients indicate the difference in selling probability of a share-package between men and women. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Male_i + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$ where $Male$ is a dummy variable indicating whether an investor is male or not. $Covariates$ include controls for age, birthyear (i.e. cohort), experience, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 63743 investors and 91 million holding period share package observations. Post reform estimates are based on 51244 investors and 72 million holding period share package observations.

Figure 35: Heterogeneity w.r.t. Gender: Losses



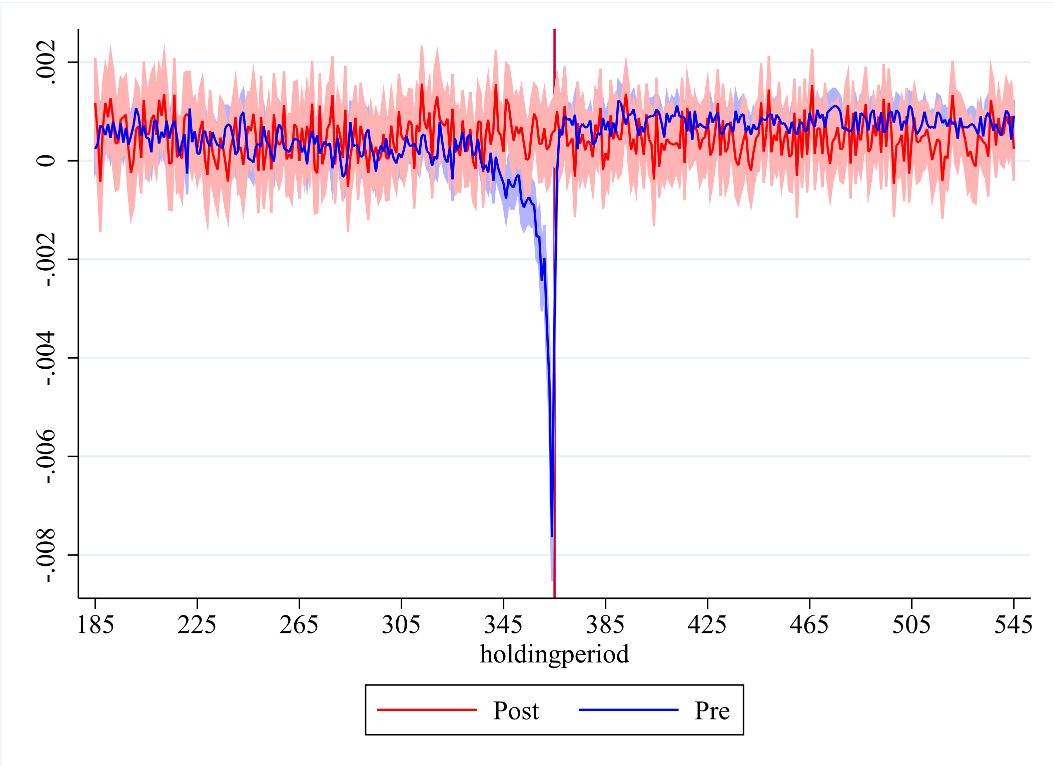
Notes: This figure displays coefficient estimates for a male dummy in the hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices below the purchase price. Coefficients indicate the difference in selling probability of a share-package between man and woman. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Male_i + Covariates_{ijd}\gamma + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Where $Male$ is a dummy variable indicating whether an investor is male or not. $Covariates$ include controls for age, birthyear (i.e. cohort), experience, income category, wealth category, working in the financial sector, having a doctoral degree, and being self-employed. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 70783 investors and 176 million holding period share package observations. Post reform estimates are based on 52290 investors and 76 million holding period share package observations.

Figure 36: Heterogeneity w.r.t. Price-Change Magnitude: Gains



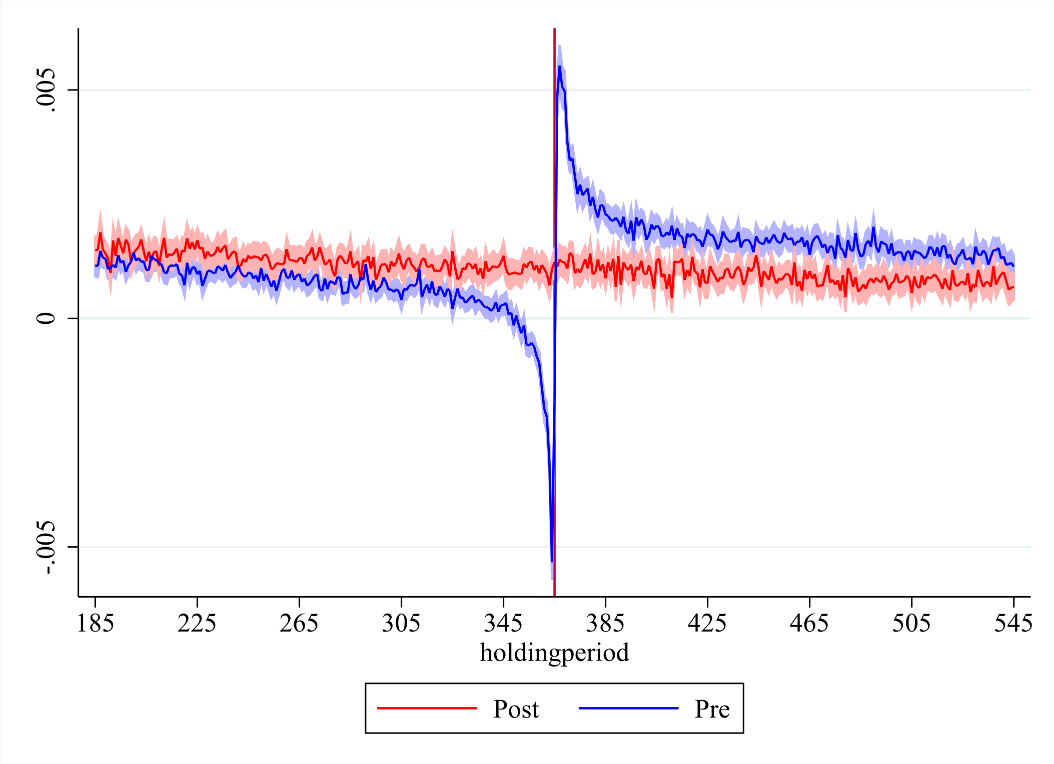
Notes: This figure displays coefficient estimates for the size of a gain from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices above the purchase price. Coefficients indicate by how much an additional percentage point increase in the price increases the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Change_{ijd} + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Where $Change_{ijd}$ is measured as $\frac{p_{ijtd} - p_{ij0d}}{p_{ij0d}}$. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 63695 investors and 88 million holding period share package observations. Post reform estimates are based on 51185 investors and 71 million holding period share package observations.

Figure 37: Heterogeneity w.r.t. Price-Change Magnitude: Losses



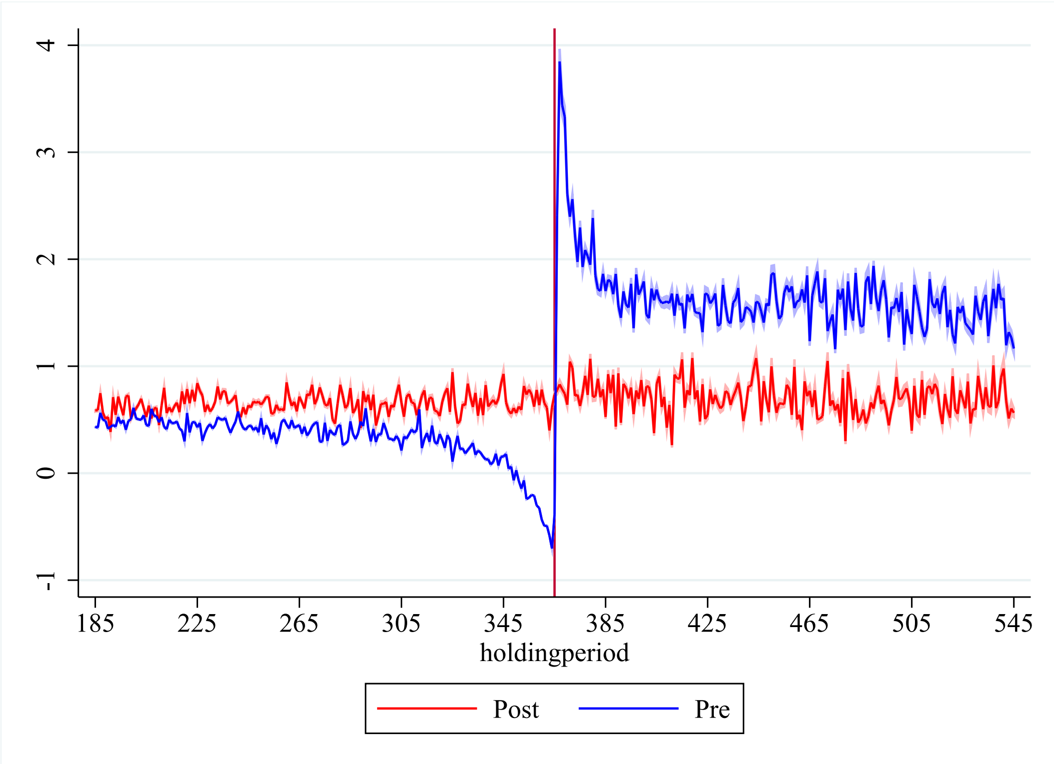
Notes: This figure displays coefficient estimates for the size of a gain from hazard-rate regressions for each day of the holdingperiod. Included are share packages with prices below the purchase price. Coefficients indicate by how much an additional percentage point decrease in the price changes the probability that a share-package is sold on this holding-period day. Note since the change for losses is negative, negative values mean that share packages with higher losses are sold with a higher probability. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Change_{ijd} + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Where $Change_{ijd}$ is measured as $\frac{p_{ijtd}-p_{ij0d}}{p_{ij0d}}$. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded blue area displays 95 percent confidence intervals. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded red area displays 95 percent confidence intervals.The vertical red line at day 365 marks the last day in which gains were taxable. Pre reform estimates are based on 70752 investors and 172 million holding period share package observations. Post reform estimates are based on 52238 investors and 76 million holding period share package observations.

Figure 38: Disposition Effect around time discontinuity



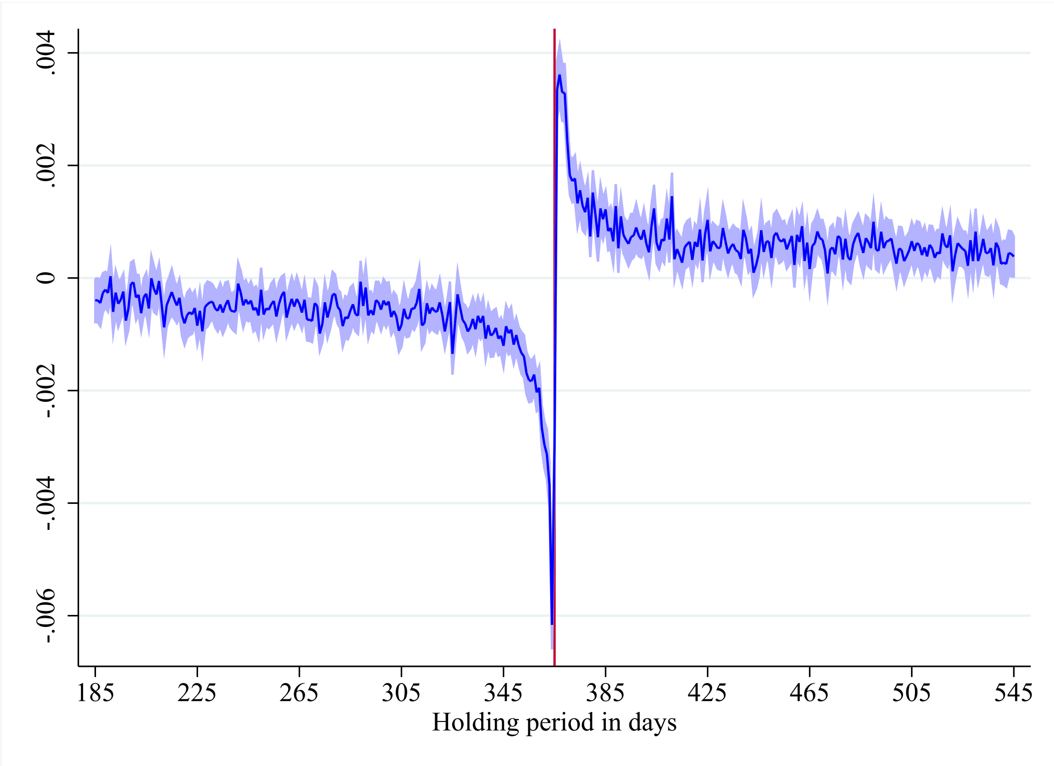
Notes: This figure displays estimates for the average difference in selling probability between gains and losses on each day of the holdingperiod. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Gain_{ijd} + \varepsilon_{ijd}$. The blue line represents estimates for β_1 for share packages which were bought before 2009. The shaded area displays 95 percent confidence interval. The red line represents estimates for β_1 for share packages which were bought after 2009. The shaded area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which gains were taxable prior to 2009. Pre reform estimates are based on 72059 investors and 267 million holding period share package observations. Post reform estimates are based on 55698 investors and 148 million holding period share package observations.

Figure 39: Disposition Effect: Gain Coefficients relative to Loss Coefficients



Notes: This figure displays estimates for the relative difference in selling probability between gains and losses on each day of the holdingperiod. That is the the coefficient of the gain dummy is divided by the constant. Standard errors are calculated using the delta method. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Gain_{ijd} + \varepsilon_{ijd}$. The blue line represents estimates for β_1/β_0 for share packages which were bought before 2009. The shaded area displays 95 percent confidence interval. The red line represents estimates for β_1/β_0 for share packages which were bought after 2009. The shaded area displays 95 percent confidence intervals. The vertical red line at day 365 marks the last day in which gains were taxable prior to 2009. Pre reform estimates are based on 72059 investors and 267 million holding period share package observations. Post reform estimates are based on 55698 investors and 148 million holding period share package observations.

Figure 40: Disposition Effect around time discontinuity: DiD model



Notes: This figure displays difference in difference estimates for the average difference in selling probability between gains and losses on each day of the holdingperiod. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \beta_1 Pre + \beta_2 \mathbb{1}(Gain_{ijd}) + \beta_3 Pre \times \mathbb{1}(Gain_{ijd}) + \varepsilon_{ijd}$. The blue line represents estimates for β_3. The shaded area displays 95 percent confidence interval. The vertical red line at day 365 marks the last day in which gains were taxable prior to 2009. Estimates are based on 87571 investors and 455 million holding period share package observations.

References

Abeler, J. and S. Jaeger (2014). Complex Tax Incentives. *American Economic Journal: Economic Policy 7*(3), 1–28.

Abramitzky, R. and V. Lavy (2014). How responsive is investment in schooling to changes in redistributive policies and in returns? *Econometrica 82*(4), 1241–1272.

Aghion, P., U. Akcigit, M. Lequien, and S. Stantcheva (2017). Tax simplicity and heterogeneous learning. NBER working paper no 24049.

Agrawal, D. R. and D. Foremny (2019). Relocation of the rich: Migration in response to top tax rate changes from spanish reforms. *The Review of Economics and Statistics*. forthcoming.

Agrawal, D. R. and W. H. Hoyt (2016). Commuting and taxes: Theory, empirics, and welfare implications. *The Economic Journal*, accepted.

Alesina, A. and G.-M. Angeletos (2005). Fairness and redistribution. *The American Economic Review 95*(4), 960 – 980.

Alesina, A., A. Miano, and S. Stantcheva (2018). Immigration and redistribution. NBER working paper no 24733.

Alesina, A., S. Stantcheva, and E. Teso (2018). Intergenerational mobility and support for redistribution. *American Economic Review 108*(2), 521 – 554.

Almunia, M. and D. Lopez-Rodriguez (2018). Under the radar: The effects of monitoring firms on tax compliance. *American Economic Journal: Economic Policy 10*(1), 1–38.

Ayers, B. C., C. E. Lefanowicz, and J. R. Robinson (2003). Shareholder taxes in acquisition premiums: The effect of capital gains taxation. *The Journal of Finance 58*(6), 2783–2801.

Ayers, B. C., O. Z. Li, and J. R. Robinson (2008). Tax-induced trading around the taxpayer relief act of 1997. *The Journal of the American Taxation Association 30*(1), 77–100.

Ballard, C. L., S. Gupta, et al. (2018). Perceptions and realities of average tax rates in the federal income tax: Evidence from michigan. *National Tax Journal 71*(2), 263–294.

Barber, B. M. and T. Odean (2000). Trading is hazardous to your wealth: The common stock investment performance of individual investors. *The journal of Finance 55*(2), 773–806.

Barber, B. M. and T. Odean (2001). Boys will be boys: Gender, overconfidence, and common stock investment. *The Quarterly Journal of Economics 116*(1), 261–292.

Barber, B. M. and T. Odean (2004). Are individual investors tax savvy? evidence from retail and discount brokerage accounts. *Journal of Public Economics 88*(1-2), 419–442.

Barber, B. M. and T. Odean (2013). The behavior of individual investors. Volume 2 of *Handbook of the Economics of Finance*, pp. 1533 – 1570. Elsevier.

Bargain, O., K. Orsini, and A. Peichl (2014). Comparing labor supply elasticities in Europe and the US: New results. *Journal of Human Resources 49*(3), 723–838.

Bargain, O. and A. Peichl (2016). Own-wage labor supply elasticities: variation across time and estimation methods. *IZA Journal of Labor Economics 5*(1), 10.

Bastani, S. and H. Selin (2014). Bunching and non-bunching at kink points of the Swedish tax schedule. *Journal of Public Economics*. Forthcoming.

Bastani, S. and D. Waldenstroem (2019). Salience of inherited wealth and the support for inheritance taxation. mimeo, online at `http://spencerbastani.com/Bastani_Waldenstrom_inheritancetax_experiment.pdf`.

Bastian, J. and K. Michelmore (2018). The long-term impact of the earned income tax credit on children's education and employment outcomes. *Journal of Labor Economics 36*(4), 1127–1163.

Benzarti, Y. (2017). How taxing is tax filing? using revealed preferences to estimate compliance costs. NBER working paper no 23903.

Bertrand, M., E. Duflo, and S. Mullainathan (2004). How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics 119*(1), 249–275.

Best, M. C., A. Brockmeyer, H. J. Kleven, J. Spinnewijn, and M. Waseem (2015). Production versus revenue efficiency with limited tax capacity: Theory and evidence from pakistan. *Journal of Political Economy 123*(6), 1311–1355.

Best, M. C. and H. J. Kleven (2018). Housing market responses to transaction taxes: Evidence from notches and stimulus in the u.k. *The Review of Economic Studies 85*(1), 157–193.

Bhargava, S. and D. Manoli (2015). Psychological frictions and the incomplete take-up of social benefits: Evidence from an irs field experiment. *American Economic Review 105*(11), 3489–3529.

Bierbrauer, F. and P. Boyer (2018). Politically feasible reforms of non-linear tax systems. *Working Paper*.

Bierbrauer, F., P. Boyer, and A. Raj (2018). On the political economy of the income-tax threshold. *Working Paper*.

Bitler, M., H. Hoynes, and E. Kuka (2017). Child poverty, the great recession, and the social safety net in the united states. *Journal of Policy Analysis and Management 36*(2), 358–389.

Blattman, C., J. C. Jamison, and M. Sheridan (2017). Reducing crime and violence: Experimental evidence from cognitive behavioral therapy in liberia. *American Economic Review 107*(4), 1165–1206.

Blau, F. D. and L. M. Kahn (2007). Changes in the labor supply behavior of married women: 1980-2000. *Journal of Labor Economics 25*(3), 393–438.

Blaufus, K., J. Bob, J. Hundsdoerfer, D. Kiesewetter, and J. Weimann (2013). Decision heuristics and tax perception–an analysis of a tax-cut-cum-base-broadening policy. *Journal of Economic Psychology 35*, 1–16.

Blaufus, K., S. Eichfelder, and J. Hundsdoerfer (2014). Income tax compliance costs of working individuals: Empirical evidence from germany. *Public Finance Review 42*(6), 800–829.

Blaufus, K., F. Hechtner, and A. Moehlmann (2017). The effect of tax preparation expenses for employees: Evidence from germany. *Contemporary Accounting Research 34*(1), 525–554.

Blaufus, K. and R. Ortlieb (2009). Is simple better? a conjoint analysis of the effects of tax complexity on employee preferences concerning company pension plans. *Schmalenbach Business Review 61*(1), 60–83.

Blesse, S. and F. Heinemann (2019). Citizens' trade-offs in state merger decisions: Evidence from a randomized survey experiment. Technical report, ZEW Discussion Papers.

Blom, A. G., C. Gathmann, and U. Krieger (2015). Setting up an online panel representative of the general population: The German Internet Panel. *Field Methods 27*(4), 391–408.

Blom, A. G., J. M. Herzing, C. Cornesse, J. W. Sakshaug, U. Krieger, and D. Bossert (2017). Does the recruitment of offline households increase the sample representativeness of probability-based online panels? evidence from the german internet panel. *Social Science Computer Review 35*(4), 498–520.

Blouin, J., L. Hail, and M. H. Yetman (2009). Capital gains taxes, pricing spreads, and arbitrage: Evidence from cross - listed firms in the u.s. *The Accounting Review 84*(5), 1321–1361.

Blouin, J. L., B. J. Bushee, and S. A. Sikes (2017). Measuring tax-sensitive institutional investor ownership. *The Accounting Review 92*(6), 49–76.

Blouin, J. L., J. S. Raedy, and D. A. Shackelford (2003). Capital gains taxes and equity trading: Empirical evidence. *Journal of Accounting Research 41*(4), 611–651.

Blumkin, T., B. J. Ruffle, and Y. Ganun (2012). Are income and consumption taxes ever really equivalent? Evidence from a real-effort experiment with real goods. *European Economic Review 56*(6), 1200–1219.

Blundell, R. and T. MaCurdy (1999). Labor supply: A review of alternative approaches. handbook of labor economics. *Handbook of Labor Economics 3*, 1559–1695.

Boccanfuso, J. and A. Ferey (2019). Inattention and the taxation bias. mimeo, online at `https://docs.google.com/viewer?a=v&pid=sites&srcid=ZGVmYXVsdGRvbWFpbnxmZXJleWFudG9pbmV8Z3g6NTRlZThiM2UwOWJmYTRhNQ`.

Boenke, Tim, J. B. and C. Schroeder (2017). Fiscal federalism and tax administration: Evidence from germany. *German Economic Review 18*(3), 337–409.

Bogart, W. T. and W. M. Gentry (1995). Capital gains taxes and realizations: Evidence from interstate comparisons. *The Review of Economics and Statistics*, 267–282.

Brennan, G. and J. M. Buchanan (1980). *The Power to Tax*. Cambridge University Press.

Brown, J. R., A. Kapteyn, E. F. Luttmer, and O. S. Mitchell (2017). Cognitive Constraints on Valuing Annuities. *Journal of the European Economic Association 15*(2), 429–462.

Brown, K. M. (2013). The link between pensions and retirement timing: Lessons from california teachers. *Journal of Public Economics 98*, 1–14.

Brusco, S., L. Colombo, and U. Galmarini (2014). Tax differentiation, lobbying, and welfare. *Social Choice and Welfare 42*(4), 977–1006.

Cappelen, A. W., I. K. Haaland, and B. Tungodden (2018). Beliefs about behavioral responses to taxation. mimeo.

Castanheira, M., G. Nicodème, and P. Profeta (2012). On the political economics of tax reforms: survey and empirical assessment. *International Tax and Public Finance 19*(4), 598–624.

Chang, T. Y., D. H. Solomon, and M. M. Westerfield (2016). Looking for someone to blame: Delegation, cognitive dissonance, and the disposition effect. *The Journal of Finance 71*(1), 267–302.

Chetty, R., J. N. Friedman, T. Olsen, and L. Pistaferri (2011). Adjustment costs, firm responses, and micro vs. macro labor supply elasticities: Evidence from danish tax records. *The quarterly journal of economics 126*(2), 749–804.

Chetty, R., J. N. Friedman, and E. Saez (2013a). Using differences in knowledge across neighborhoods to uncover the impacts of the EITC on earnings. *American Economic Review 103*(7), 2683–2721.

Chetty, R., J. N. Friedman, and E. Saez (2013b). Using differences in knowledge across neighborhoods to uncover the impacts of the eitc on earnings. *American Economic Review 103*(7), 2683–2721.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and taxation: Theory and evidence. *American Economic Review 99*(4), 1145 – 77.

Chetty, R. and E. Saez (2005). Dividend taxes and corporate behavior: Evidence from the 2003 dividend tax cut. *The Quarterly Journal of Economics 120*(3), 791–833.

Chetty, R. and E. Saez (2013). Teaching the tax code: Earnings responses to an experiment with EITC recipients. *American Economic Journal: Applied Economics 5*(1), 1–31.

Cole, S., A. Paulson, and G. Shastry (2014). Smart money? the effect of education on financial outcomes. *Review of Financial Studies 27*(7), 2022–2051.

Constantinides, G. M. (1984). Optimal stock trading with personal taxes: Implications for prices and the abnormal january returns. *Journal of Financial Economics 13*(1), 65 – 89.

Correia, S. (2015). Singletons, cluster-robust standard errors and fixed effects: a bad mix. *Federal Reserve Board of Governors, mimeo*.

Cremer, H., F. Gahvari, and J.-M. Lozachmeur (2010). Tagging and income taxation: theory and an application. *American Economic Journal: Economic Policy 2*(1), 31–50.

Cruces, G., R. Perez-Truglia, and M. Tetaz (2013). Biased perceptions of income distribution and preferences for redistribution: Evidence from a survey experiment. *Journal of Public Economics 98*, 100 – 112.

Dai, Z., E. Maydew, D. A. Shackelford, and H. H. Zhang (2008). Capital gains taxes and asset prices: Capitalization or lock-in? *The Journal of Finance 63*(2), 709–742.

Dammon, R., H. Zhang, and C. Spatt (2001). Diversification and capital gains taxes with multiple risky assets.

Daunfeldt, S.-O., U. Praski-Staahlgren, and N. Rudholm (2010). Do high taxes lock-in capital gains? evidence from a dual income tax system. *Public Choice 145*(1), 25–38.

de Bartolome, C. (1995). Which tax rate do people use: Average or marginal? *Journal of Public Economics 56*(1), 79–96.

de Quidt, J., J. Haushofer, and C. Roth (2018). Measuring and bounding experimenter demand. *American Economic Review 108*(11), 3266–3302.

Deutsche Bundesbank (2013). Statistik ueber wertpapierinvestments [statistic about stock investments]. Technical report.

Dhar, R. and N. Zhu (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science 52*(5), 726–740.

Ding, P. (2017). A paradox from randomization-based causal inference. *Statistical science 32*(3), 331–345.

Doerrenberg, P. and A. Peichl (2018). Tax morale and the role of social norms and reciprocity. evidence from a randomized survey experiment. Cesifo working paper no. 7149.

Doerrenberg, P., A. Peichl, and S. Siegloch (2017). The elasticity of taxable income in the presence of deduction possibilities. *Journal of Public Economics 151*, 41–55.

Dolls, M. and N. Wehrhoefer (2018). Attitudes towards euro area reforms: Evidence from a randomized survey experiment. CESifo working paper series no. 7141.

Dorn, D. and G. Huberman (2005). Talk and action: What individual investors say and what they do. *Review of Finance 9*(4), 437–481.

Dowd, T. and R. McClelland (2019). The bunching of capital gains realizations. *National Tax Journal 72*(2), 323–358.

Dowd, T., R. McClelland, and A. Muthitacharoen (2015). New evidence on the tax elasticity of capital gains. *National Tax Journal 68*(3), 511–544.

Dube, A., T. W. Lester, and M. Reich (2010). Minimum wage effects across state borders: Estimates using contiguous counties. *Review of Economics & Statistics 92*(4), 945–964.

Duncan, D., V. Nadella, A. Bowers, S. Giroux, and J. D. Graham (2014). Bumpy designs: Impact of privacy and technology costs on support for road mileage user fees. *National Tax Journal 67*(3), 505–530.

Durante, R., L. Putterman, and J. van der Weele (2014). Preferences for redistribution and perception of fairness: An experimental study. *Journal of the European Economic Association 12*(4), 1059–1086.

Economist (2005). Simplifying tax systems. the case for flat taxes. The Economist article, online at `https://www.economist.com/special-report/2005/04/14/the-case-for-flat-taxes`.

Economist (2013). Tax reform in america. simpler, fairer, possible. The Economist article, online at `https://www.economist.com/leaders/2013/07/13/`

`simpler-fairer-possible`.

Eissa, N. and H. W. Hoynes (2004). Taxes and the labor market participation of married couples: The earned income tax credit. *Journal of Public Economics 88*(9), 1931–1958.

Eissa, N. and H. W. Hoynes (2006). Behavioral responses to taxes: Lessons from the eitc and labor supply. In J. Poterba (Ed.), *Tax Policy and the Economy*, Volume 20, pp. 73–110.

Eissa, N. and J. B. Liebman (1996). Labor supply response to the earned income tax credit. *The Quarterly Journal of Economics 111*(2), 605–637.

Engelmann, D., E. Janeba, L. Mechtenberg, and N. Wehrhoefer (2018). Preferences over taxation of high income individuals: Evidence from survey and laboratory experiments. mimeo.

Evers, M., R. D. Mooij, and D. V. Vuuren (2008). The wage elasticity of labour supply: a synthesis of empirical estimates. *De Economist 156*(1), 25–43.

FAST (2010). *Faktisch anonymisierte Lohn- und Einkommensteuerstatistik by the Federal Statistical Office*. Wiesbaden, Germany.

FAZ (2004). Der steuermann. Frankfurter Allgemeine Zeitung (FAZ) article, online at `https://www.faz.net/aktuell/wirtschaft/reformer-merz-der-steuermann-1192866.html`.

FAZ (2005). Kirchhof-modell: Die reichen profitieren am meisten. Frankfurter Allgemeine Zeitung (FAZ) article, online at `https://www.faz.net/aktuell/wirtschaft/wirtschaftspolitik/steuern-kirchhof-modell-die-reichen-profitieren-am-meisten-1260403.html`.

Feenberg, D. and E. Coutts (1993). An introduction to the taxsim model. *Journal of Policy Analysis and Management 12*(1), 189–194.

Feldman, N. E., P. Katuščák, and L. Kawano (2016). Taxpayer confusion: Evidence from the child tax credit. *American Economic Review 106*(3), 807–35.

Feldman, N. E. and B. J. Ruffle (2015). The impact of including, adding, and subtracting a tax on demand. *American Economic Journal: Economic Policy 7*(1), 95–118.

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), 551–72.

Feldstein, M. (1999). Tax avoidance and the deadweight loss of the income tax. *Review of Economics and Statistics 81*(4), 674 – 680.

Feldstein, M., J. Slemrod, and S. Yitzhaki (1980). The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains*. *The Quarterly Journal of Economics 94*(4), 777–791.

Feng, L. and M. S. Seasholes (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance 9*(3), 305–351.

Finkelstein, A. (2009). E-ztax: Tax salience and tax rates. *Quarterly Journal of Economics 124*(3), 969–1010.

Fisher, R. A. (1935). *The Design of Experiments*. Oliver and Boyd, Ltd.

Fochmann, M., N. Müller, and M. Overesch (2018). Less cheating? the effects of prefilled forms on compliance behavior.

Forbes (2017). Trump wants a simple tax system – history says he won't get it. Forbes article, online at `https://www.forbes.com/sites/taxanalysts/2017/09/11/trump-wants-a-simple-tax-system-history-says-he-wont-get-it/#4c3b63e27c1c`.

Foremny, D. and N. Riedel (2014). Business taxes and the electoral cycle. *Journal of Public Economics 115*, 48 – 61.

Frydman, C. and A. Rangel (2014). Debiasing the disposition effect by reducing the saliency of information about a stock's purchase price. *Journal of Economic Behavior & Organization 107*, 541 – 552.

Frydman, C. and B. Wang (2020). The impact of salience on investor behavior: Evidence from a natural experiment. *The Journal of Finance 75*(1), 229–276.

Fuchs, V. R., A. B. Krueger, and J. M. Poterba (1998). Economists' views about parameters, values, and policies: Survey results in labor and public economics. *Journal of Economic Literature 36*(3), 1387–1425.

Fuest, C., A. Peichl, and T. Schaefer (2008). Does Tax Simplification Yield more Equity and Efficiency? An empirical analysis for Germany. *CESifo Economic Studies 54*(1), 73–97.

Fuest, C., A. Peichl, and S. Siegloch (2018a, February). Do higher corporate taxes reduce wages? micro evidence from germany. *American Economic Review 108*(2), 393–418.

Fuest, C., A. Peichl, and S. Siegloch (2018b). Do higher corporate taxes reduce wages? micro evidence from germany. *American Economic Review 108*(2), 393–418.

Fujii, E. T. and C. B. Hawley (1988). On the accuracy of tax perceptions. *The Review of Economics and Statistics 70*(2), 344–347.

Gale, W. (2001). Tax simplification: Issues and option. mimeo.

Galli, E. and P. Profeta (2009). Tax complexity with heterogeneous voters. *Public Finance and Managemente 9*(2), 1–35.

Gelber, A. M. and J. W. Mitchell (2012). Taxes and time allocation: Evidence from single women and men. *Review of Economic Studies 79*(3), 863–897.

German Federal Bureau of Statistics (2008a). Einkommens- und verbrauchsstichprobe: Einkommen, einnahmen & ausgaben [income and consumption sample statistic: Earnings, income & expenditures]. Technical report.

German Federal Bureau of Statistics (2008b). Einkommens- und verbrauchsstichprobe: Vermoegen & schulden [income and consumption sample statistic: Assets & liabillites]. Technical report.

German Federal Bureau of Statistics (2019). Vgr des bundes - einnahmen und ausgaben des staates: Deutschland, jahre, staatliche teilsektoren, einnahme- und ausgabearten. Technical report.

Gideon, M. (2017). Do individuals perceive income tax rates correctly? *Public Finance Review 45*(1), 97–117.

Glogowsky, U. (2016). Behavioral responses to wealth transfer taxation: Bunching evidence from germany. Technical report.

Goettinger Tageblatt (2018). Friedrich Merz' erster tweet – ein bierdeckel, online at: `http://www.goettinger-tageblatt.de/Nachrichten/Politik/Deutschland-Welt/Friedrich-Merz-erster-Tweet-ein-Bierdeckel`. Technical report.

Goldin, J. and T. Homonoff (2013). Smoke gets in your eyes: Cigarette tax salience and regressivity. *American Economic Journal: Economic Policy 5*(1), 302–36.

Gordon, R. H. and W. Kopczuk (2014). The choice of the personal income tax base. *Journal of Public Economics 118*, 97 – 110.

Gravelle, J. G. and T. L. Hungerford (2012). The challenge of individual income tax reform: An economic analysis of tax base broadening. Congressional research service. washington, dc.

Grinblatt, M. and M. Keloharju (2001). What makes investors trade? *The Journal of Finance 56*(2), 589–616.

Grinblatt, M. and M. Keloharju (2004). Tax-loss trading and wash sales. *Journal of Financial Economics 71*(1), 51 – 76.

Grinblatt, M. and T. J. Moskowitz (2004). Predicting stock price movements from past returns: the role of consistency and tax-loss selling. *Journal of Financial Economics 71*(3), 541 – 579.

Grogger, J. (2003). The effects of time limits, the eitc, and other policy changes on welfare use, work, and income among female-headed families. *Review of Economics & Statistics 85*(2), 394–408.

Hanlon, M. and S. Heitzman (2010). A review of tax research. *Journal of Accounting and Economics 50*(2), 127 – 178.

Hargaden, E. (2015). Taxpayer responses over the cycle: Evidence from irish notches. *University of Tennessee mimeo.*

Hausman, J. A. (1985). Taxes and labor supply. *Handbook of Public Economics 1*, 213–263.

Heckman, J. J. (1993). What has been learned about labor supply in the past twenty years? *American Economic Review 83*(2), 116–121.

Heim, B. T. (2007). The incredible shrinking elasticities married female labor supply, 1978-2002. *Journal of Human Resources 42*(4), 881–918.

Hettich, W. and S. Winer (1999). *Democratic Choice and Taxation*. Cambridge University Press.

Hettich, W. and S. L. Winer (1988). Economic and political foundations of tax structure. *American Economic Review 78*(4), 701–712.

Hines, J. R. (2016). High tax heresy. Working paper, available online: `https://pdfs.semanticscholar.org/2fc7/1dfd009ab42b5899815fe75cda368bdc178a.pdf`.

Hines, J. R. (2019). *Sensible taxes and practical policitcs*. Oxford University Press. Forthcoming 2020.

Hoopes, J., P. Langetieg, S. Nagel, D. Reck, J. Slemrod, and B. Stuart (2016). Who sold during the crash of 2008-9? evidence from tax-return data on daily sales of stock. NBER Working Paper 22209.

Hotz, J. V. and J. K. Scholz (2003). The earned income tax credit. In R. A. Moffitt (Ed.), *In Means-Tested Transfer Programs in the United States*, pp. 141–198.

Hotz, J. V. and J. K. Scholz (2006). Examining the effect of the earned income tax credit on the labor market participation of families on welfare. *NBER Working Paper 11968*.

Hoynes, H. W., D. Miller, and D. Simon (2015). Income, the earned income tax credit, and infant health. *American Economic Journal: Economic Policy 7*(1), 172–211.

Hoynes, H. W. and A. J. Patel (2015). Effective policy for reducing inequality? the earned income tax credit and the distribution of income. *NBER Working Paper 21340*.

Huang, J. (2001). Taxable or tax-deferred account? portfolio decision with multiple investment goals. In *Mimeo*. MIT Sloan School of Management.

Ito, K. (2014). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review 104*(2), 537–63.

Ivković, Z., J. Poterba, and S. Weisbenner (2005). Tax-motivated trading by individual investors. *American Economic Review 95*(5), 1605–1630.

Jacob, M. (2013). Capital gains taxes and the realization of capital gains and losses: Evidence from german income tax data. *FinanzArchiv: Public Finance Analysis 69*(1), 30–56.

Jacob, M. (2018). Tax regimes and capital gains realizations. *European Accounting Review 27*(1), 1–21.

Jäger, S. (2016). How substitutable are workers? evidence from worker deaths. *Discussion Paper mimeo.*

Jakobsen, K., K. Jakobsen, H. Kleven, and G. Zucman (2018). Wealth taxation and wealth accumulation: Theory and evidence from denmark. Working Paper 24371, National Bureau of Economic Research.

James, S., A. Sawyer, and I. Wallschutzky (1997). Tax simplifications - a tale of three countries. *Bulletin for International Fiscal Documentation 51*(11), 493–503.

Jones, M. R. (2014). Changes in eitc eligibility and participation, 2005-2009. *CARRA Working Paper 2014-04*.

Keane, M. P. (2011). Labor supply and taxes: A survey. *Journal of Economic Literature 49*(4), 961–1075.

Keane, M. P. and R. Rogerson (2012). Micro and macro labor supply elasticities: A reassessment of conventional wisdom. *Journal of Economic Literature 50*(2), 464–476.

Kennedy, P. E. (1995). Randomization tests in econometrics. *Journal of Business and Economic Statistics 13*(1), 85–95.

Kerschbamer, R. and D. Müller (2017). Social preferences and political attitudes: An online experiment on a large heterogeneous sample. University of insbruck working paper 2017-16.

Killingsworth, M. R. and J. J. Heckman (1986). Female labor supply: A survey. *Handbook of Labor Economics 1*(4), 103–204.

Kirchhof, P. (2011). *Bundessteuergesetzbuch. Ein Reformentwurf zur Erneuerung des Steuerrechts*. Heidelberg: C.F. Müller Wissenschaft.

Kleven, H. J. (2016). Bunching. *Annual Review of Economics 8*, 435–464.

Kleven, H. J., M. B. Knudsen, C. T. Kreiner, S. Pedersen, and E. Saez (2011). Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark. *Econometrica 79*(3), 651 – 692.

Kleven, H. J., C. Landais, E. Saez, and E. Schultz (2014). Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in denmark *. *The Quarterly Journal of Economics 129*(1), 333–378.

Kleven, H. J. and E. A. Schultz (2014a). Estimating taxable income responses using danish tax reforms. *American Economic Journal: Economic Policy 6*(4), 271–301.

Kleven, H. J. and E. A. Schultz (2014b, November). Estimating taxable income responses using danish tax reforms. *American Economic Journal: Economic Policy 6*(4), 271–301.

Kopczuk, W. (2005). Tax bases, tax rates and the elasticity of reported income. *Journal of Public Economics 89*(11-12), 2093–2119.

Kopczuk, W. (2012). The polish business "flat" tax and its effect on reported incomes: a pareto improving tax reform? mimeo, online at `http://www.columbia.edu/~wk2110/bin/PolishFlatTax.pdf`.

Kopczuk, W. and C. Pop-Eleches (2007). Electronic filing, tax preparers and participation in the earned income tax credit. *Journal of Public Economics 91*(7-8), 1351–1367.

Korniotis, G. M. and A. Kumar (2011). Do older investors make better investment decisions? *The Review of Economics and Statistics 93*(1), 244–265.

Kotakorpi, K. and J.-P. Laamanen (2016). Prefilled income tax returns and tax compliance: Evidence from a natural experiment. University of tampere, school of management, economics.

Kuziemko, I., M. I. Norton, E. Saez, and S. Stantcheva (2015). How elastic are preferences for redistribution? evidence from randomized survey experiments. *American Economic Review 105*(4), 1478–1508.

LaLumia, S. (2009). The earned income tax credit and reported self-employment income. *National Tax Journal 62*(2), 191–217.

Lang, M. H. and D. A. Shackelford (2000). Capitalization of capital gains taxes: evidence from stock price reactions to the 1997 rate reduction. *Journal of Public Economics 76*(1), 69 – 85.

Lergetporer, P., G. Schwerdt, K. Werner, M. R. West, and L. Woessmann (2018). How information affects support for education spending: Evidence from survey experiments in germany and the united states. *Journal of Public Economics 167*, 138 – 157.

Lergetporer, P., K. Werner, and L. Woessmann (2018). Educational inequality and public policy preferences: Evidence from representative survey experiments. CESifo working paper no. 7192.

Leuz, C., S. Meyer, M. Muhn, E. Soltes, and A. Hackethal (2017). Who falls prey to the wolf of wall street? investor participation in market manipulation. NBER Working Paper 24083.

Lichter, A., M. Löffler, and S. Siegloch (2015). The economic costs of mass surveillance: Insights from stasi spying in east germany. *IZA Discussion Paper 9245*.

Liebman, J. B. and R. J. Zeckhauser (2004). Schmeduling. mimeo, Harvard Kennedy School.

Loos, B., S. Meyer, and M. Pagel (2020). The consumption effects of the disposition to sell winners and hold losers. NBER Working Paper 26668.

Martinez, I. Z., E. Saez, and M. Siegenthaler (2018, May). Intertemporal labor supply substitution? evidence from the swiss income tax holidays. Working Paper 24634, National Bureau of Economic Research.

McClelland, R. and S. Mok (2012). A review of recent research on labor supply elasticities. *CBO Working Paper 43675*.

Meghir, C. and D. Phillips (2008). Labour supply and taxes. *IFS Working Papers 8*.

Meltzer, A. H. and S. F. Richard (1981). A rational theory of the size of government. *Journal of political Economy 89*(5), 914–927.

Meyer, B. and D. Rosenbaum (2001). Welfare, the earned income tax credit, and the labor supply of single mothers. *The Quarterly Journal of Economics 116*(3), 1063–1114.

Meyer, B. D. (2010). The effects of the earned income tax credit and recent reforms. In J. R. Brown (Ed.), *Tax Policy and the Economy*, Volume 24, pp. 153–180.

Moffitt, R. A. (2013). The great recession and the social safety net. *The Annals of the American Academy of Political and Social Science 650*(1), 143–166.

Müller, D. and S. Renes (2017). Fairness views and political preferences - evidence from a large online experiment. University of insbruck working paper 2017-10.

Mummolo, J. and E. Peterson (2019). Demand effects in survey experiments: An empirical assessment. *American Political Science Review*. forthcoming.

Neisser, C. (2017). The elasticity of taxable income: A meta-regression analysis.

Neumark, D. and K. E. Williams (2016). Do state earned income tax credits increase participation in the federal eitc? *Discussion Paper*.

Nichols, A. and J. Rothstein (2016). The earned income tax credit. In R. A. Moffitt (Ed.), *Economics of Means-Tested Transfer Programs in the United States*, Volume 1, pp. 137–218.

NPR (2015). Lots of candidates want to simplify tax code; here's what they get wrong. National Public Radio (NPR) online article. https://www.npr.org/sections/itsallpolitics/2015/09/28/443203910/lots-of-candidates-want-to-simplify-the-tax-code-heres-what-they-get-wrong?t=1551777844261.

NYT (2015). The tax code can be simpler. but not three pages. New York Times (NYT) article, online at https://www.nytimes.com/2015/11/15/upshot/a-three-page-tax-code-not-exactly-simple.html.

Odean, T. (1998). Are investors reluctant to realize their losses? *The Journal of Finance 53*(5), 1775–1798.

OECD (2010a). Tax design considerations. in *Tax Policy Reform and Economic Growth*, oecd publishing, paris. doi: https://doi.org/10.1787/9789264091085-7-en.

OECD (2010b). Tax expenditures in OECD countries. Oecd publishing, paris. doi: https://doi.org/10.1787/9789264076907-en.

Ooghe, E. and A. Peichl (2015). Fair and efficient taxation under partial control. *The Economic Journal 125*(589), 2024–2051.

Paetzold, J. (2019). How do taxpayers respond to a large kink? evidence on earnings and deduction behavior from austria. *International Tax and Public Finance*, 1–31.

Paetzold, J. and H. Winner (2016). Taking the high road? compliance with commuter tax allowances and the role of evasion spillovers. *Journal of Public Economics 143*, 1 – 14.

Pencavel, J. (1986). Labor supply of men: a survey. *Handbook of Labor Economics 1*, 3–102.

Pitt, M. and J. Slemrod (1989). The compliance cost of itemizing deductions: Evidence from individual tax returns. *American Economic Review 79*(5), 1224–32.

Poterba, J. M. (2001). Taxation and portfolio structure: Issues and implications. NBER Working Paper 8223.

Poterba, J. M. (2002). Taxation, risk-taking, and household portfolio behavior. In *Handbook of public economics*, Volume 3, pp. 1109–1171. Elsevier.

Poterba, J. M. and S. J. Weisbenner (2001). Capital gains tax rules, tax-loss trading, and turn-of-the-year returns. *The Journal of Finance 56*(1), 353–368.

Rees-Jones, A. and D. Taubinsky (2016). Measuring "schmeduling". NBER working pape no. 22884r.

Reese Jr, W. A. (1998). Capital gains taxation and stock market activity: Evidence from ipos. *The Journal of Finance 53*(5), 1799–1819.

Rohaly, J. and W. G. Gale (2004). Effects of tax simplification options on equity, efficiency, and simplicity: A quantitative analysis. In H. J. Aaron and J. Slemrod (Eds.), *The Crisis in Tax Administration*. Washington D.C.: Brookings Institution Pres.

Roth, C. and J. Wohlfart (2018). Experienced inequality and preferences for redistribution. *Journal of Public Economics 167*, 251 – 262.

Roth, C. and J. Wohlfart (2019). How do expectations about the macroeconomy affect personal expectations and behavior? Working paper, available at `http://dx.doi.org/10.2139/ssrn.3016052`.

Rothstein, J. (2010). Is the EITC as good as an NIT? Conditional cash transfers and tax incidence. *American Economic Journal: Economic Policy 2*(1), 177–208.

Saez, E. (2010a). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Saez, E. (2010b). Do taxpayers bunch at kink points? *American Economic Journal: Economic Policy 2*(3), 180–212.

Saez, E. (2017). Taxing the rich more: Preliminary evidence from the 2013 tax increase. *Tax Policy and the Economy 31*(1), 71–120.

Saez, E., J. Slemrod, and S. H. Giertz (2012). The elasticity of taxable income with respect to marginal tax rates: A critical review. *Journal of Economic Literature 50*(1), 3–50.

Saez, E. and S. Stantcheva (2016). Generalized social marginal welfare weights for optimal tax theory. *American Economic Review 106*(1), 24–45.

Seida, J. A. and W. F. Wempe (2000). Do capital gain tax rate increases affect individual investors' trading decisions? *Journal of Accounting and Economics 30*(1), 33 – 57.

Seru, A., T. Shumway, and N. Stoffman (2009). Learning by Trading. *The Review of Financial Studies 23*(2), 705–739.

Seru, A., T. Shumway, and N. Stoffman (2010). Learning by trading. *The Review of Financial Studies 23*(2), 705–739.

Shackelford, D. A. and R. E. Verrecchia (2002). Intertemporal tax discontinuities. *Journal of Accounting Research 40*(1), 205–222.

Shefrin, H. and M. Statman (1985). The disposition to sell winners too early and ride losers too long: Theory and evidence. *The Journal of Finance 40*(3), 777–790.

Shoven, J. B. and C. Sialm (2004). Asset location in tax-deferred and conventional savings accounts. *Journal of Public Economics 88*(1-2), 23–38.

Sikes, S. (2018). Capital gains lock-in and share repurchases. mimeo.

Slemrod, J. and W. Kopczuk (2002). The optimal elasticity of taxable income. *Journal of Public Economics 84*(1), 91–112.

Stantcheva, S. (2020). Understanding economic policies: What do people know and how can they learn? mimeo, presentation slides online at `https://scholar.harvard.edu/stantcheva/publications`.

Suarez Serrato, J. C. and O. Zidar (2016). Who benefits from state corporate tax cuts? a local labor markets approach with heterogeneous firms. *American Economic Review 106*(9), 2582–2624.

Tax Policy Center (2019). What are tax expenditures and how are they structured? The Tax Policy Center's Briefing Book. a citizen's guide to the fascinating (though often complex) elements of the federal tax system, online at `https://www.taxpolicycenter.org/sites/default/files/briefing-book/bb_full_2018_1.pdf`.

Tran-Nam, B. (2000). Tax reform and tax simplicity: A new and 'simpler' tax system? *University of New South Wales Law Journal, The 23*(2), 241–251.

Tsankova, T., C. Imbert, M. Luts, and J. Spinnewijn (2019). 'how to improve tax compliance? evidence from countrywide experiments in belgium.

van Rooij, M., A. Lusardi, and R. Alessie (2011). Financial literacy and stock market participation. *Journal of Financial Economics 101*(2), 449–472.

Vox (2017). Why democrats should support radically simpler taxes. Vox article, online at `https://www.vox.com/policy-and-politics/2017/5/8/15442172/democrats-tax-plan-return-free-filing-trump-ambitious`.

Wagner, F. W. (2006). Was bedeutet Steuervereinfachung wirklich? *Perspektiven der Wirtschaftspolitik 7*(1), 19–33.

Warskett, G., S. L. Winer, and W. Hettich (1998). The complexity of tax structure in competitive political systems. *International Tax and Public Finance 5*, 123–151.

Weinzierl, M. (2014). The promise of positive optimal taxation: normative diversity and a role for equal sacrifice. *Journal of Public Economics 118*, 128 – 142.

Weinzierl, M. (2017). Popular acceptance of inequality due to innate brute luck and support for classical benefit-based taxation. *Journal of Public Economics 155*, 54

– 63.

Westfall, P. H. and S. S. Young (1993). *Resampling-based multiple testing: Examples and methods for p-value adjustment*. John Wiley & Sons.

WiWo (2017). Pendler kosten den fiskus fuenf milliarden euro. WirtschaftsWoche (WiWo) article, online at `https://www.wiwo.de/politik/deutschland/berufspendler-pendler-kosten-den-fiskus-fuenf-milliarden-euro/20654060.html`.

Yagan, D. (2015, December). Capital tax reform and the real economy: The effects of the 2003 dividend tax cut. *American Economic Review 105*(12), 3531–63.

Young, A. (2018). Channeling fisher: Randomization tests and the statistical insignificance of seemingly significant experimental results. *The Quarterly Journal of Economics 134*(2), 557–598.

Zwick, E. (2018). The costs of corporate tax complexity. NBER working paper no 24382.

A Appendix to Chapter 2

A.1 The EITC tax schedule

Figure 41 illustrates the EITC tax refund schedule for families with one and two children. The refunds refer to 2009, the last year in our sample.

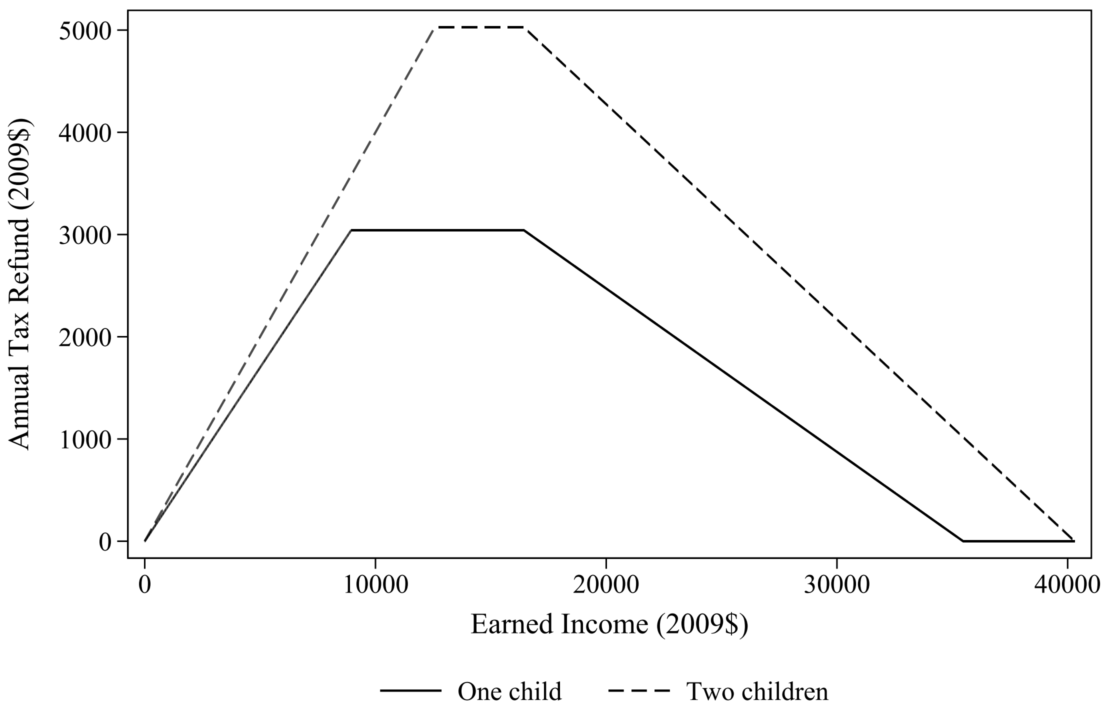


Figure 41: The EITC schedule in 2009

Notes: This graph displays the relationship between the tax refund and household labor income according to the 2009 federal EITC schedule. Tax units with adjusted gross income above the earned income threshold are not eligibile. First EITC kink point for families with one child: USD 8 950; for families with two children USD 12 570. Second kink point at USD 16 420.

A.2 Predicting EITC expansions

Our identification strategy relies on the assumption that the top-up rate in a state is uncorrelated with county or state characteristics. A central concern with this assumption is that the generosity of the state EITC is driven by the business cycle, state-level fluctuations in tax revenue, or changes in minimum wages. To address this concern, we follow Bastian and Michelmore (2018), and predict the level of the state EITC based on current and lagged state-level economic variables in a panel regression. If any of the variables turned out to be statistically significant, this would be reason for concern, as it would cast doubt on the validity of the identifying assumption.

For this purpose, we collected state-level data on the welfare state (top marginal income tax rate, level of minimum wage, monthly welfare benefits), as well as tax revenues, which can be seen as a measure of the business cycle. The data span the years 1995-2009.[71] The regression results are shown in Table 10. Given that statistically insignificant results are more likely when standard errors are clustered, we report here conventional standard errors. None of the regressors is statistically significantly different from zero, which we interpret as strong evidence that changes in the state EITC are not driven by state-level fluctuations in the economy.

Table 10: OLS Results: predictors of State EITC top-up rates

Control Variables:	Top-up Rate
Top Marginal Income Tax Rate	0.1827
	(0.6228)
Lagged Top Marginal Income Tax Rate	-0.8308
	(0.7124)
Minimum Wage	-0.2128
	(0.3834)
Lagged Real Minimum Wage	0.6750
	(0.4262)
Max Monthly Welfare Benefits, Family of 3 (in 100 USD)	-0.4842
	(1.1698)
Lagged Max Monthly Welfare Benefits, Family of 3 (in 100 USD)	-0.4390
	(1.2205)
State Tax Revenue (in 1M USD)	0.0001
	(0.0001)
Lagged State Tax Revenue (in 1M USD)	-0.0001
	(0.0001)
Controls:	
Year FE	Yes
State FE	Yes
N	714

This table displays the results of a panel OLS regression of the state EITC top-up rate on state-level economic variables. All regressions control for year and state fixed effects. Wages, welfare benefits and tax revenues are deflated to 2010 USD. Conventional standard errors are displayed in parentheses.

[71] Sources: minimum wages: St. Louis Fed, welfare benefits: welfare rules databook, tax revenue: Annual Survey of State Government Tax Collections, consumer price index: St. Louis Fed, marginal income tax rates: NBER Taxsim.

A.3 EITC claimants before, during and after the Great Recession

Figure 42 displays the number of EITC claimants around the time of the Great Recession, between 2007 and 2012. This number has been increasing throughout, although the increase was strongest during the Great Recession, between 2008 and 2009.

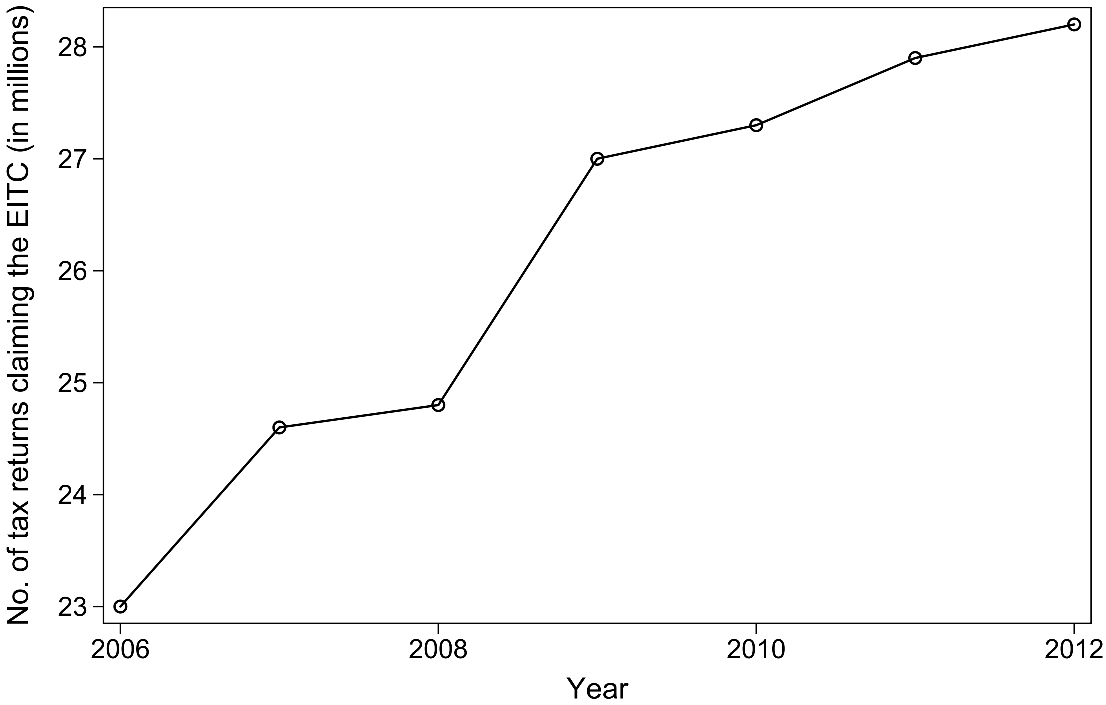


Figure 42: EITC claimers, 2007-2012

Notes: This graph displays the number of people in the US claiming the EITC in a given year. Source: IRS

A.4 Converting zip-code-level data to county-level data

The dataset by Chetty et al. (2013a) provides data at the level of three-digit zip codes. Because the border pair design requires information at the county-level, we convert the data from the zip-code to the county level. The dataset mainly consists of absolute numbers, such as the number of EITC claimants in a given zip code. If a zip code comprises more than one county, we divide the absolute numbers evenly across all counties within a ZIP code. For example, if there are 1000 claimants in zip-code A and A consists of two counties we assign each county 500 claimants. If, on the other hand, a county is part of more than one zip code, we assign this county the sum of the absolute numbers. If the zip code that cuts through a county also covers another county, we split the absolute numbers between these countries before adding up within counties. For example, if zip codes A (1,000 claimants) and B (500 claimants) are completely contained in county X, we assign county X 1,500 claimants. If, however, zip code A also covers another county while B is fully contained in X, we assign county X 500 claimants from A and 500 claimants from B.[72]

For the 3,141 counties in our dataset, we apply the first method — split the numbers between counties within a zip code — to 1,179 counties. For another 1,960 counties, we apply both methods, namely we split numbers between counties as well as aggregate numbers within counties. The remaining two counties coincide with the zip codes.

A.5 Identifying variation

Table 11 displays the amount of variation — measured by the standard deviation — in the most important variables for different samples as well as different fixed effect specifications. Column (1) displays the variation for all counties, whereas Columns (2)-(4) display the variation for border counties only. In the border pair sample, some counties appear more than once if they have more than one neighbor in a different state. Going from left to right, one can see that the amount of variation is reduced as more fixed effects are added. However, even after controlling for pair-by-year fixed effects, there remains substantial variation in top-up rates as well as the outcome variables.

[72]We found splitting the number of claimants evenly between counties the most transparent way of converting zip-code-level data to county-level data. It would also be possible to (dis-)aggregate the numbers based on population measures. However, without further assumptions, this would only be possible for disaggregation (one zip code contains more than one county), but not for aggregation (one county contains more than one zip code).

Table 11: Variation in key variables

	(1) All Counties	(2) Border Counties	(3) Border Counties	(4) Border Counties
Top-up rates				
SD	6.86	7.56	5.43	4.88
Top-up dummy				
SD	0.44	0.45	0.33	0.29
Share of self-employed near the kink point				
SD	3.83	3.75	1.89	1.42
EITC claimants, self-employed				
SD	3956.62	3391.67	2299.95	2175.05
EITC claimants, non-self-employed				
SD	13245.24	13029.56	8284.27	8238.79
Self-employed claimants near the kink				
SD	684.01	686.86	504.16	460.22
Controls:				
Year FE	No	No	Yes	No
Pair FE	No	No	Yes	No
Pair $\times$ Year FE	No	No	No	Yes
N	43967	36616	36616	36616

This table displays the variation — measured by the standard deviation — in the main variables with various sets of fixed effects. The all-county dataset comprises all counties in the US. The border county dataset comprises counties straddling a state borders only. Columns (1) -(2) display the raw standard deviations. Column (3) shows the residual variation after a transformation of separate year and pair fixed effects. Column (4) shows the residual variation after a transformation of year-by-pair fixed effects

Figure 43 further illustrates the relationship between state-specific top-up rate (horizontal axis) and the degree of bunching (vertical axis) in a binned scatter with ten equally sized bins on each axis. The graph controls for state-specific characteristics of the EITC — a dummy that equals unity if the the refund depends on the number of children, and a dummy that equals unity if a positive refund is given if a person's tax credit exceeds his/her tax liability — as well as pair-by-year fixed effects. The regression line corresponds to the regression coefficient in Table 2, Panel A), Column (4).

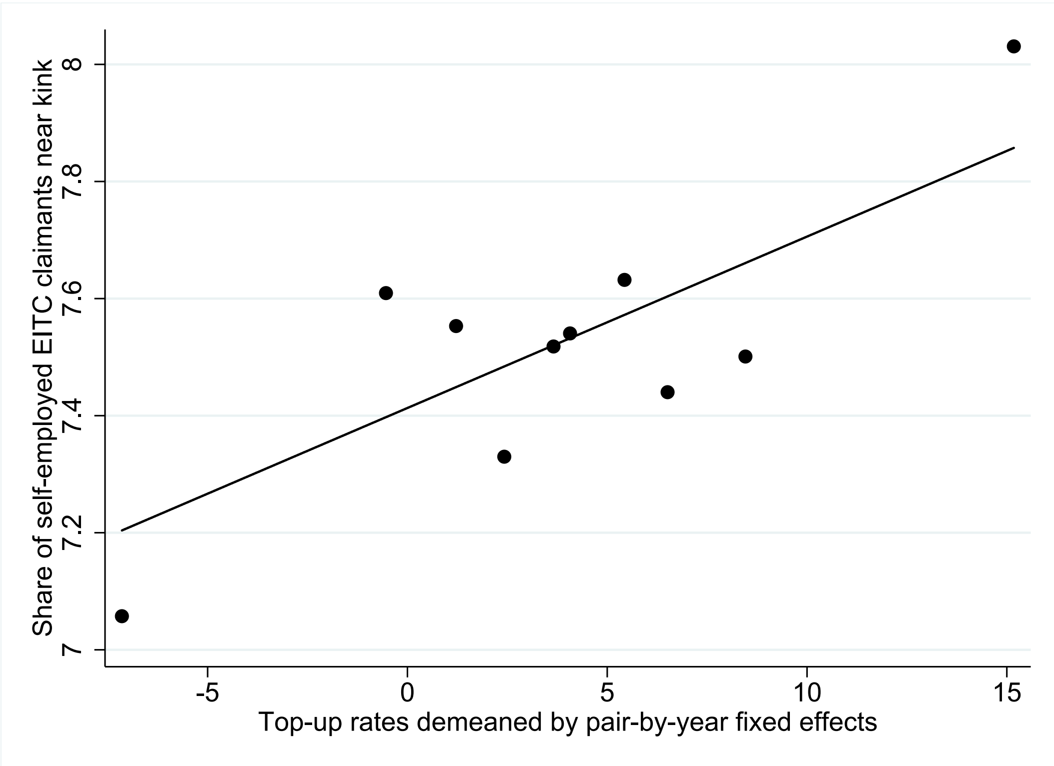


Figure 43: Bunching vs. top-up rates

Notes: This graph displays the relationship between the share of self-employed at the first kink point of the EITC and the state specific top-up rates in a binned scatter, whereby each variable is divided in ten equally sized bins. Both variables have been demeaned by pair-by-year fixed effects, and we control for state-level features of the EITC.

A.6 Assessing inference through permutation tests

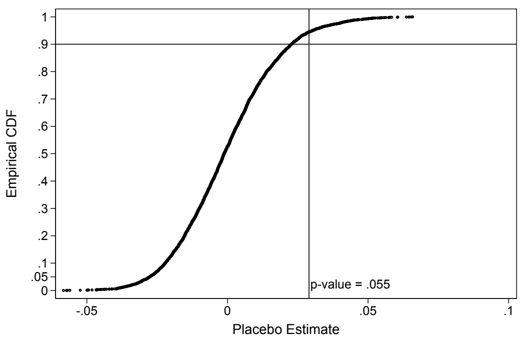
While the border design facilitates estimating a causal effect by providing clear treatment and control counties, it also complicates statistical inference. The error terms can be correlated across space as well as within counties over time, which can lead to an underestimation of standard errors, and an under-rejection of the null hypothesis of no effect (Bertrand et al. 2004). Moreover, in the border pair design, some counties are part of multiple pairs, such that their errors are mechanically correlated. As a first step, to account for correlations in the error term, we applied to all estimates a two-way clustering procedure at the county and pair level. However, this may not eliminate all systematic correlations of the error terms.

To assess the statistical significance of our estimates without relying on assumptions about clustering, we additionally perform permutation tests for the four main outcomes. In these tests, we first obtain an empirical placebo distribution of estimates that would occur under the null hypothesis of there being no effect. In a second step, we compare our estimates to the placebo distribution and obtain a empirical p-value that describes the probability of obtaining a result that is at least as extreme as ours.[73] In a conventional case — namely one in which a treatment is as-signed once — the placebo distribution is obtained by repeatedly randomizing the treatment across observations and estimating the same model in each replication. The complication in our case is that top-up rates within states are path-dependent. States do not randomly set a top-up rate every year, but rather adjust the rate of the previous year. To account for path-dependency, we therefore randomize over 14-year paths in top-up rates. In each replication, we randomly assign each state a path for its top-up rate and estimate the model.

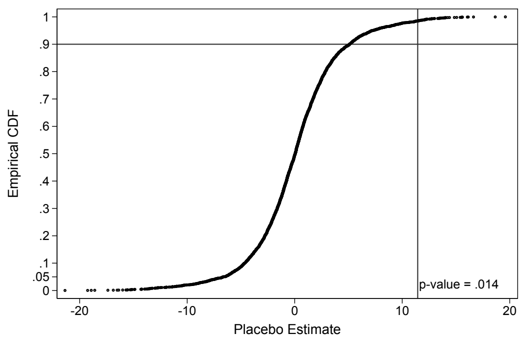
Figure 44 displays the cumulative density function of the placebo distributions based on 5,000 replications, as well as the z-scores of our estimates (vertical lines) from Column (6) in Table 2. The horizontal lines describe the 90-th percentile of the placebo distribution. Statistical significance at the 10% level requires that the intersection of both lines is located South-East of the placebo distribution. This is the case for the outcomes displayed in Panels A-C, where the empirical p-values are 0.055, 0.014, and 0.027, respectively. For the outcome in Panel D — namely the total number of non-self-employed claimants — the p-value is 0.128, which means that this estimate is not statistically significant at the 10% level.

These results confirm the inference drawn from the two-way clustering approach in Table 2. Raises in the top-up rate significantly increase bunching near the kink point, which is the result of an overproportional increase in the number of claimants with an income close to the kink point. As before, we find no statistically significant effect on the total number of non-self-employed EITC claimants.

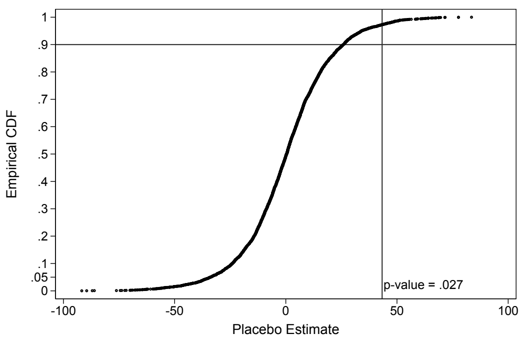
[73]This procedure follows Kennedy (1995) and Chetty et al. (2009).



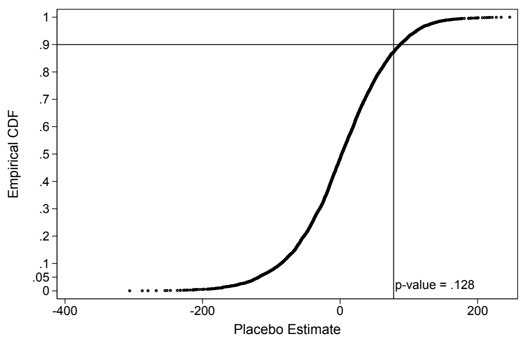
(A) Bunching



(B) Self-employed near kink



(C) Total number of self-employed



(D) Total number of non-self-employed

Figure 44: Permutation tests

Notes: These graphs display the results of permutation tests for each of the four outcomes. Each panel displays the cumulative density from 5,000 replications, as well as the empirical p-values.

B Appendix to Chapter 3

B.1 Additional Figures and Tables

Table 12: Balancing Tests First Experiment

Variable	Redistribution	Avoidance	Control
Gender: Reference category Male			
Sex	0.022	-0.029	0.002
	(0.019)	(0.019)	(0.019)
Marital status: Reference category: Not married			
Married	0.013	0.029	-0.036*
	(0.019)	(0.019)	(0.019)
Unemployed: Reference category: Employed			
Unemployed	-0.011	-0.061	0.063
	(0.074)	(0.074)	(0.074)
Retirement status: Reference category: Not retired			
Retired	0.042*	0.016	-0.042*
	(0.023)	(0.023)	(0.023)
Household size:			
HH-size	-0.007	0.012	-0.008
	(0.009)	(0.009)	(0.009)
Education:			
Education	-0.010	-0.011	0.002
	(0.011)	(0.011)	(0.011)
Household net income: Reference category poor			
2	-0.006	0.031	-0.016
	(0.036)	(0.035)	(0.036)
3	-0.030	0.065*	-0.032
	(0.036)	(0.035)	(0.035)
4	-0.032	0.047	-0.010
	(0.037)	(0.037)	(0.037)
Rich	-0.020	0.041	-0.014
	(0.038)	(0.037)	(0.037)

No income stated	-0.033	0.053	-0.001
	(0.040)	(0.040)	(0.040)
Age category:			
Age	0.017**	-0.004	-0.007
	(0.007)	(0.007)	(0.007)
Political orientation: Reference category: Left-wing			
Conservative	-0.022	-0.056***	-0.033
	(0.021)	(0.021)	(0.021)
Non partisans	-0.051	-0.003	0.050
	(0.037)	(0.037)	(0.037)
Difficulty in declaring taxes: Reference category: No difficulty			
2	0.022	0.099	-0.022
	(0.063)	(0.063)	(0.063)
3	-0.066	0.116*	-0.050
	(0.060)	(0.060)	(0.060)
4	-0.059	0.064	-0.005
	(0.060)	(0.060)	(0.060)
Very difficult	-0.012	0.020	-0.009
	(0.063)	(0.063)	(0.063)
No taxes declared	-0.055	0.086	-0.031
	(0.065)	(0.065)	(0.065)
Taxes not self declared	-0.030	0.089	-0.059
	(0.059)	(0.059)	(0.059)

Notes: Randomization checks for the first experiment. The table shows the coefficients and standard errors (in parentheses) from a series of regressions of the form: $y_i = \beta Covariate_i + \epsilon_i$. Where $Covariate_i$ is the respective covariate listed above. In Column (1) y_i equals 1 if participant i is in the redistribution group and 0 otherwise. In Column (2),y_i equals 1 if participant i is in the avoidance group and 0 otherwise. In Column (3), y_i equals 1 if participant i is in the control group and 0 otherwise. Standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 13: Balancing Tests Second Experiment

Variable	Economic efficency	Special interest	Control
Gender. Reference category Male			
Sex	-0.025	0.021	-0.001
	(0.033)	(0.034)	(0.033)
Marital status: Reference category: Not married			
Married	-0.038	0.032	0.012
	(0.034)	(0.034)	(0.033)
Unemployed: Reference category: Employed			
Unemployed	0.302**	-0.165	-0.134
	(0.118)	(0.120)	(0.118)
Retirement status: Reference category: Not retired			
Retired	0.031	-0.002	-0.026
	(0.041)	(0.042)	(0.041)
Household size:			
HH-size	-0.01	0.009	-0.002
	(0.016)	(0.016)	(0.015)
Education:			
Education	-0.018	0.011	0.009
	(0.019)	(0.019)	(0.019)
Household net income: Reference category poor			
2	-0.069	0.027	0.043
	(0.061)	(0.062)	(0.062)
3	-0.012	-0.008	0.020
	(0.061)	(0.062)	(0.062)
4	-0.138**	0.063	0.075
	(0.063)	(0.064)	(0.063)
Rich	-0.092	-0.004	0.088
	(0.064)	(0.065)	(0.065)
No income stated	-0.014	0.090	-0.086
	(0.068)	(0.069)	(0.067)

Age category:			
Age	-0.004	0.014	-0.008
	(0.012)	(0.013)	(0.012)
Political orientation: Reference category: Left-wing			
Conservative	-0.007	-0.019	0.026
	(0.036)	(0.037)	(0.036)
Non partisans	0.141**	-0.160***	0.022
	(0.059)	(0.060)	(0.059)
Difficulty in declaring taxes: Reference category: No difficulty			
2	-0.025	-0.009	-0.125
	(0.104)	(0.106)	(0.103)
3	0.117	0.021	-0.145
	(0.101)	(0.102)	(0.100)
4	0.135	0.050	-0.191*
	(0.099)	(0.101)	(0.098)
Very difficult	0.206**	-0.091	-0.115
	(0.104)	(0.106)	(0.103)
No taxes declared	0.175	0.021	-0.197*
	(0.108)	(0.110)	(0.107)
Taxes not self declared	0.116	0.012	-0.128
	(0.098)	(0.100)	(0.097)

Notes: Randomization checks for the second experiment. The table shows the coefficients and standard errors (in parentheses) from a series of regressions of the form: $y_i = \beta Covariate_i + \epsilon_i$. Where $Covariate_i$ is the respective covariate listed above. Sample restricted to those participants who were in the control group in the first experiment. In Column (1) y_i equals 1 if participant i is in the economic efficency group and 0 otherwise. In Column (2),y_i equals 1 if participant i is in the special interest group and 0 otherwise. In Column (3), y_i equals 1 if participant i is in the control group and 0 otherwise. Standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 14: Exp 1: Effect on Preferences for Tax Simplification. Ordered Probit

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Redistribution	-0.109*	-0.123**	-0.144**	-0.146**	-0.144**
	(0.060)	(0.061)	(0.062)	(0.062)	(0.062)
Avoidance	-0.041	-0.038	-0.024	-0.024	-0.019
	(0.059)	(0.060)	(0.061)	(0.061)	(0.061)
Constant	-0.120***	0.482***	1.537***	1.499***	1.516***
	(0.043)	(0.150)	(0.216)	(0.227)	(0.228)
N	2190	2132	2109	2109	2109
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions on preferences for tax simplification. Estimated by an ordered probit regressions of preferences for tax simplification on treatment dummies. Tax simplification is measured on a 6 point scale based on the question: "Do you generally think that the income tax system in Germany needs to be simplified?" The experimental groups are: Control group, Redistribution group and Avoidance group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "In Germany there is an ongoing debate on whether the income tax system is too complicated because of many deduction possibilities and allowances." Participants in the Redistribution group receive the following information: "However, it is sometimes also argued that a tax system with many deduction possibilities and allowances has a social-policy and redistributive compensation role. For example, tax deductions can be used to reduce the tax burden of taxpayers who are disadvantaged by circumstances." Participants in the Avoidance group receive the following information: "In this context, one argument is that a tax system with many deduction possibilities and allowances offers more scope for tax avoidance and tax adjustment. For example, tax deductions can be used to reduce one's own tax burden through better knowledge of the tax system or through unjustified specifications in the tax return." Columns (1)-(5) differ in the included sets of covariates. (1): no covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. Robust The scale of the outcome variable is 1 (absolutely not) to 6 (absolutely). Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 15: Exp 1: Effect on Distributional Effects of Complexity. Ordered Probit

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Redistribution	-0.004	-0.007	-0.010	-0.004	-0.004
	(0.058)	(0.059)	(0.059)	(0.059)	(0.060)
Avoidance	0.079	0.092	0.109*	0.110*	0.120**
	(0.059)	(0.060)	(0.061)	(0.061)	(0.061)
Constant	0.463***	0.976***	1.216***	1.229***	1.238***
	(0.046)	(0.142)	(0.195)	(0.211)	(0.211)
N	1998	1946	1931	1931	1931
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions on beliefs about whether people think that deductions work in favor of the rich. Estimated by an ordered probit Regressions of believes on treatment dummies. The outcome is measured on a 6 point scale based on the question: 'Do you think that numerous deductions and allowances contribute to a fairer distribution of income, or do you believe that high-income citizens benefit more from these deductions and allowances?" The experimental groups are: Control group, Redistribution group and Avoidance group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "In Germany there is an ongoing debate on whether the income tax system is too complicated because of many possible deductions and allowances." Participants in the Redistribution group receive the following information: "However, it is sometimes also argued that a tax system with many possible deductions and allowances has an important social-policy role, particularly in relation to income redistribution. For example, tax deductions can be used to reduce the tax burden of taxpayers who are disadvantaged by circumstances" Participants in the Avoidance group receive the following information: "In this context, one argument is that a tax system with many possible deductions and allowances offers greater opportunity for tax avoidance . For example, when individuals have a better knowledge of the tax system or make unjustified declarations, they can reduce their tax burden by taking advantage of certain allowances or deductions." Columns (1)-(5) differ in the included sets of covariates. (1): no covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. Robust The scale of the outcome variable 1 (add to a fair income distribution) to 6 (higher incomes benefit). Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

Table 16: Exp 2: Effect on Preferences for Tax Simplification. Ordered Probit

	(1)	(2)	(3)	(4)	(5)
Experimental Group Reference category: Control					
Economic Efficiency	-0.153	-0.182*	-0.210**	-0.220**	-0.224**
	(0.102)	(0.105)	(0.108)	(0.108)	(0.108)
Special Interest	-0.041	-0.037	-0.023	-0.038	-0.035
	(0.096)	(0.098)	(0.101)	(0.101)	(0.101)
Constant	0.048	0.527***	1.145***	1.066***	1.107***
	(0.068)	(0.155)	(0.205)	(0.218)	(0.219)
N	2187	2134	2114	2114	2114
Demographics	No	Yes	Yes	Yes	Yes
Tax difficulty	No	No	Yes	Yes	Yes
Household Income	No	No	No	Yes	Yes
Political Preference	No	No	No	No	Yes

Notes: The table presents the effects of the randomized treatment interventions of the second experiment on preferences for tax simplification. Estimated by an ordered probit regressions of preferences for tax simplification on treatment dummies and a full set of interactions of the treatment groups of the first and second experiment. Tax simplification is measured on a 6 point scale based on the question: "Now that we have dealt extensively with various aspects of the German tax system in this survey, we would like to ask again: do you generally believe that the income tax system should be simplified in Germany?" The experimental groups are: Control group, Economic efficiency group and Special interest group. Control is omitted, implying that the effects are relative to the Control Group. All participants receive the following information: "We would like to once again address the ongoing debate concerning whether the income tax system is too complicated due to the many possible deductions and allowances." Participants in the Economic efficiency group receive the following information: "One argument that is often used against tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances provides better opportunities to tax individuals in accordance with their ability to pay and is therefore economically more efficient." Participants in the Special interest group receive the following information: "One argument that is often used against tax simplification and that has not been addressed so far is that a tax system with many deductions and allowances offers special interest groups greater opportunity for obtaining exemptions " Columns (1)-(5) differ in the included sets of covariates. (1): no covariates, (2): gender, age, marital status, household size, employment status, retirement status, and education, (3): (2) plus perceived difficulty to declare taxes, (4): (3) plus net household income, (5): (4) plus political preferences. The scale of the outcome variable is 1 (absolutely not) to 6 (absolutely). Robust standard errors are in parentheses *** p<0.01, ** p<0.05, * p<0.1.

B.2 Illustration of treatment conditions

B.2.1 First experiment

Figure 45: Redistribution treatment

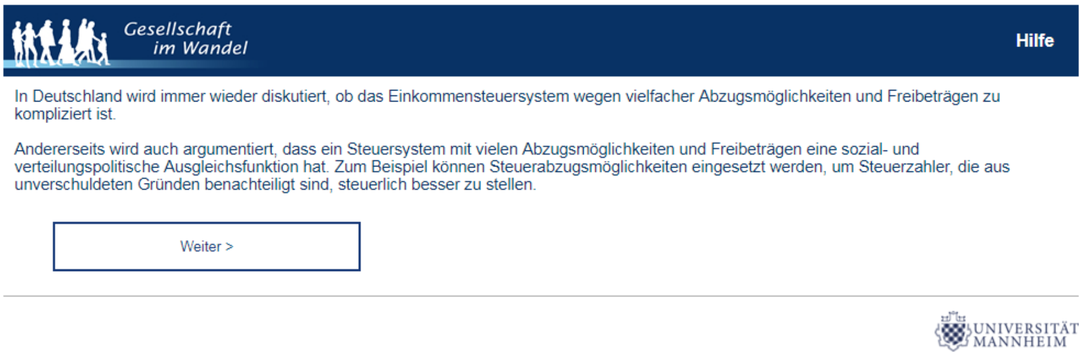
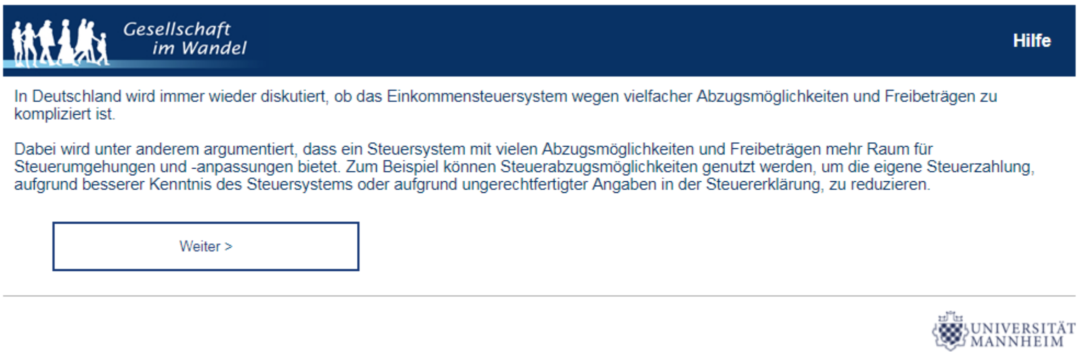


Figure 46: Avoidance treatment



B.2.2 Second experiment

Figure 47: Economic efficiency treatment

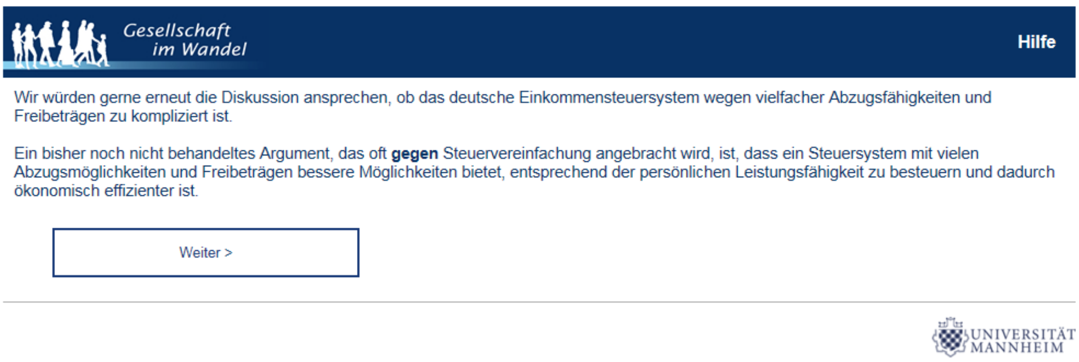
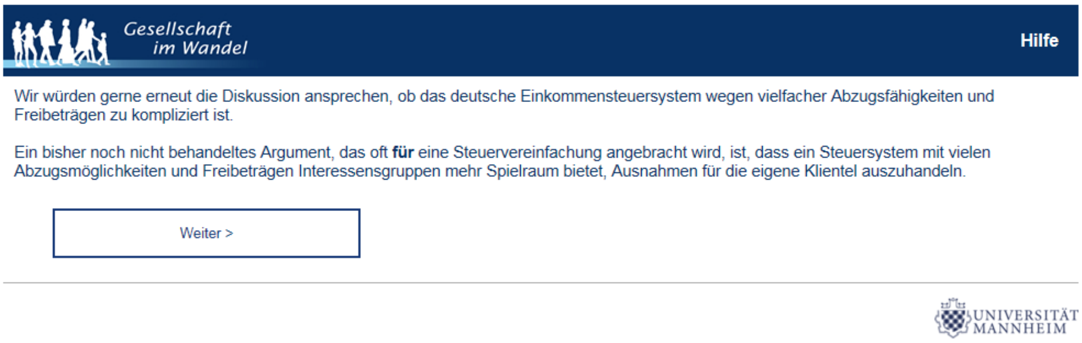


Figure 48: Special interest group treatment



B.2.3 Detailed Questionnaire

This appendix section presents the translated survey questions including reply categories. The order of presentation and the numbering of the question corresponds with the description of the survey structure in section 3.3.2.

- **Introduction:** *In the following, we would like to ask you some questions about the tax system in Germany. We will focus in particular on rules surrounding the income tax and whether they are complicated or easy to understand. Whether a tax system is generally complicated or easy to understand depends in particular on the number of possible deductions and allowances.*
- **Q1:** How difficult is it for you to fill out your tax return?
  1 Very easy;...; *5 Very difficult*; *I do not know because no taxes are declared in my name*; *I do not know because I do not declare taxes myself (rather, my partner or a tax consultant, etc. does this)*; *I do not know*
- **Randomized Experiment 1**: See body of the text (section 3.3.3) and section B.2 above.
- **Q2:** Do you generally think that the income tax system in Germany needs to be simplified?
  1 Absolutely not;...; *6 Absolutely*; *I do not know*
- **Q3:** Do you generally believe that the income tax system in Germany is in need of reform?
  1 Absolutely not;...; *6 Absolutely*; *I do not know*
- **Q4:** Do you think that numerous deductions and allowances contribute to a fairer distribution of income, or do you believe that high-income citizens benefit more from these deductions and allowances?
  1 They contribute to fairer income distribution;...; *6 High-income citizens benefit*; *I do not know*

- **Q5:** Which of the following measures to simplify the income tax system would you like the most? Assume the proposed measures will lead to unchanged tax revenues in each case.
  Same rate for all but no deductions and allowances; *Same rate for all and same deductions and allowances as under current system*; *More progressive tax rates and no deductions and allowances*; *Automatic determination of amounts in income tax declaration*; *No change*; *Other measure [insert text]*; *I do not know*

- **Introduction for Q6-8** Imagine two persons, A and B. Which person do you think should pay more taxes in the following situation?

- **Q6:** In contrast to Person B, Person A has a poor mother in need of elderly care and has to spend a considerable amount of her income for the care of her mother. Person A and B have the same gross income and are very similar in all other respects. (randomize order of answer categories)
  Person A should pay higher taxes; *Person B should pay higher taxes*; *Person A and B should pay equal tax amounts*

- **Q7:** Person A spends a considerable amount of her income on charitable giving. Person B does no such thing. Both Person A and B have the same gross income and are very similar in all other respects. (randomize order of answer categories)
  Person A should pay higher taxes; *Person B should pay higher taxes*; *Person A and B should pay equal tax amounts*

- **Q8:** Person A has to travel a considerable distance to work. Person B lives very close to work. Both Person A and B have the same gross income and are very similar in all other respects.(randomize order of answer categories)
  Person A should pay higher taxes; *Person B should pay higher taxes*; *Person A and B should pay equal tax amounts*

- **Randomized Experiment 2**: See body of the text (section 3.3.3) and section B.2 above.

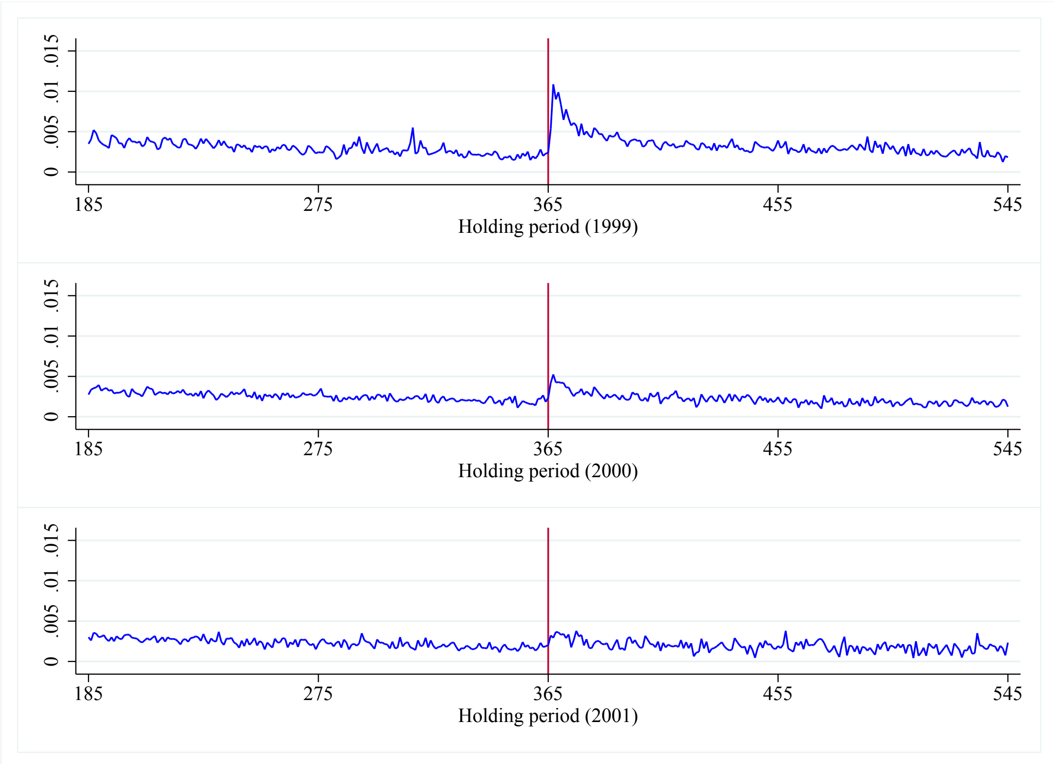
- **Q9:** Now that we have dealt extensively with various aspects of the German tax system in this survey, we would like to ask again: do you generally believe that the income tax system should be simplified in Germany?
  1 Absolutely not;*...*; *6 Absolutely*; *I do not know*

- **Q10:** Which of the following deductions and/or allowances do you usually use when filing your income tax?
  Maintenance of two households; *Home office*; *Commuting allowance*; *Other job*

related expenditures; *Pension expenses*; *Education costs*; *Care relatives*; *Child allowance, childcare*; *Donations*; *Others [insert text]*; *No deductions*; *I do not know*

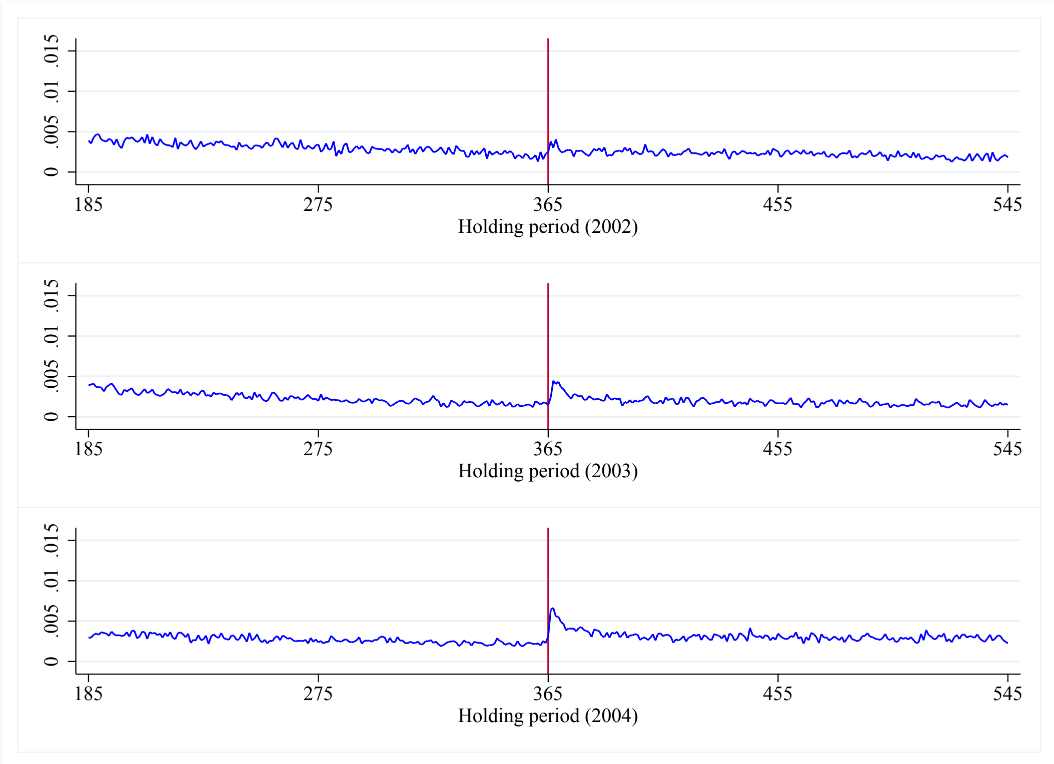
C Appendix to Chapter 4

Figure 49: Hazard-Rate Regressions by Year: Gains, Pre Years 1999-2001



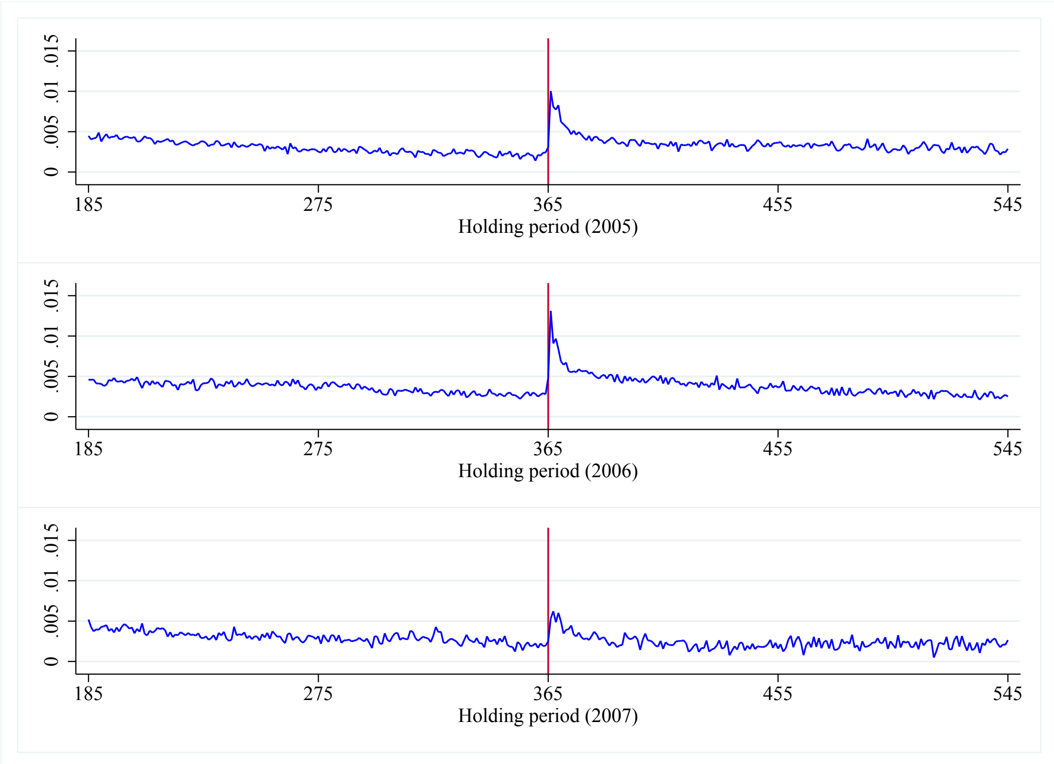
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 1999-2001. Included are share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which gains were taxable. Estimates for 1999 are based on 20516 investors and 11.5 million holding period share package observations. Estimates for 2000 are based on 28274 investors and 7.6 million holding period share package observations. Estimates for 1999 are based on 20695 investors and 4.5 million holding period share package observations.

Figure 50: Hazard-Rate Regressions by Year: Gains, Pre Years 2002-2004



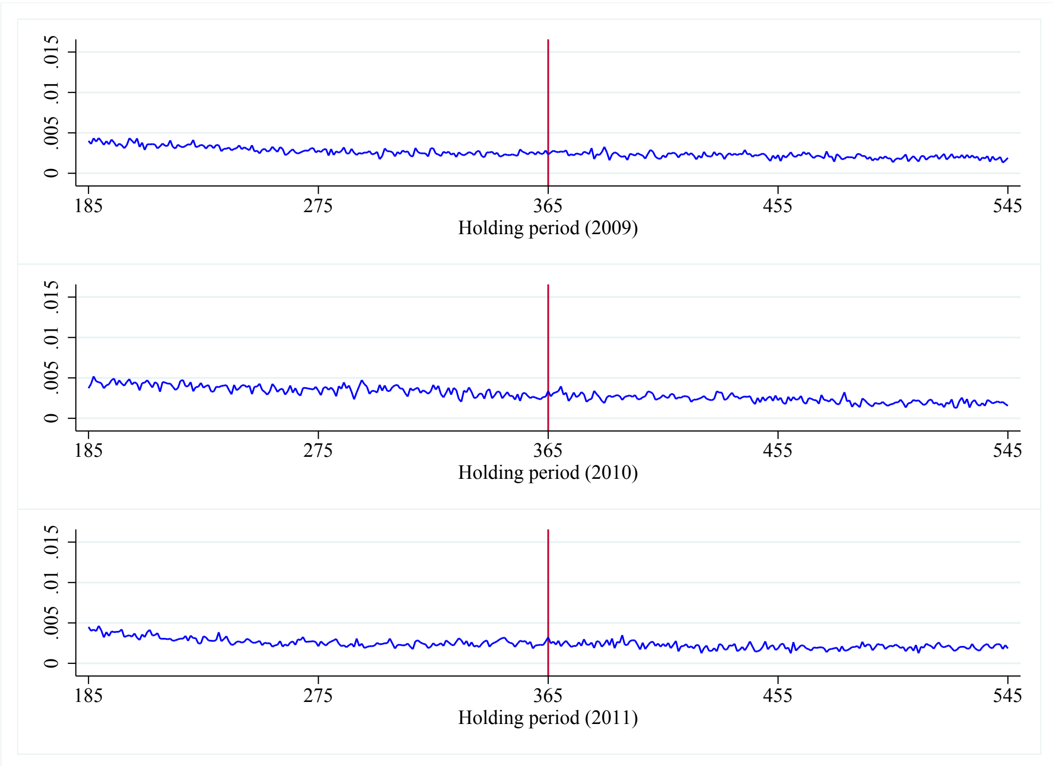
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2002-2004. Included are share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which gains were taxable. Estimates for 2002 are based on 18958 investors and 7.1 million holding period share package observations. Estimates for 2003 are based on 20859 investors and 12.6 million holding period share package observations. Estimates for 2004 are based on 23991 investors and 12.5 million holding period share package observations.

Figure 51: Hazard-Rate Regressions by Year: Gains, Pre Years 2005-2007



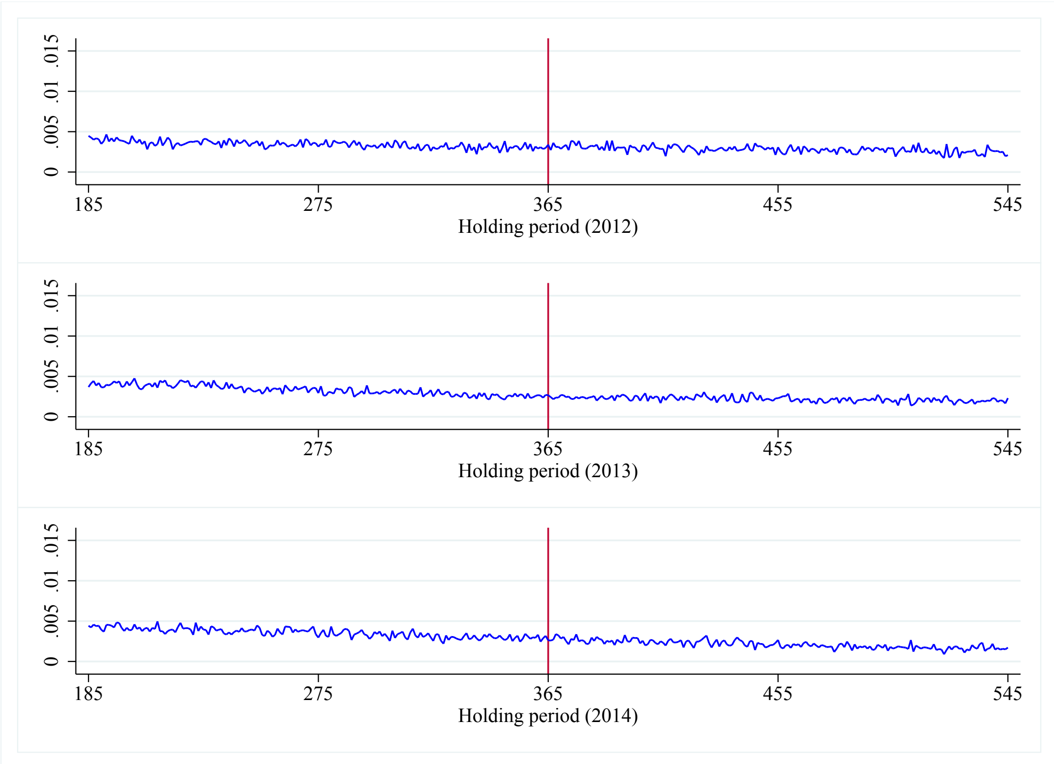
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2005-2007. Included are share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which gains were taxable. Estimates for 2005 are based on 25118 investors and 15.3 million holding period share package observations. Estimates for 2006 are based on 28202 investors and 14.2 million holding period share package observations. Estimates for 2007 are based on 21473 investors and 5.3 million holding period share package observations.

Figure 52: Hazard-Rate Regressions by Year: Gains, Post Years 2009-2011



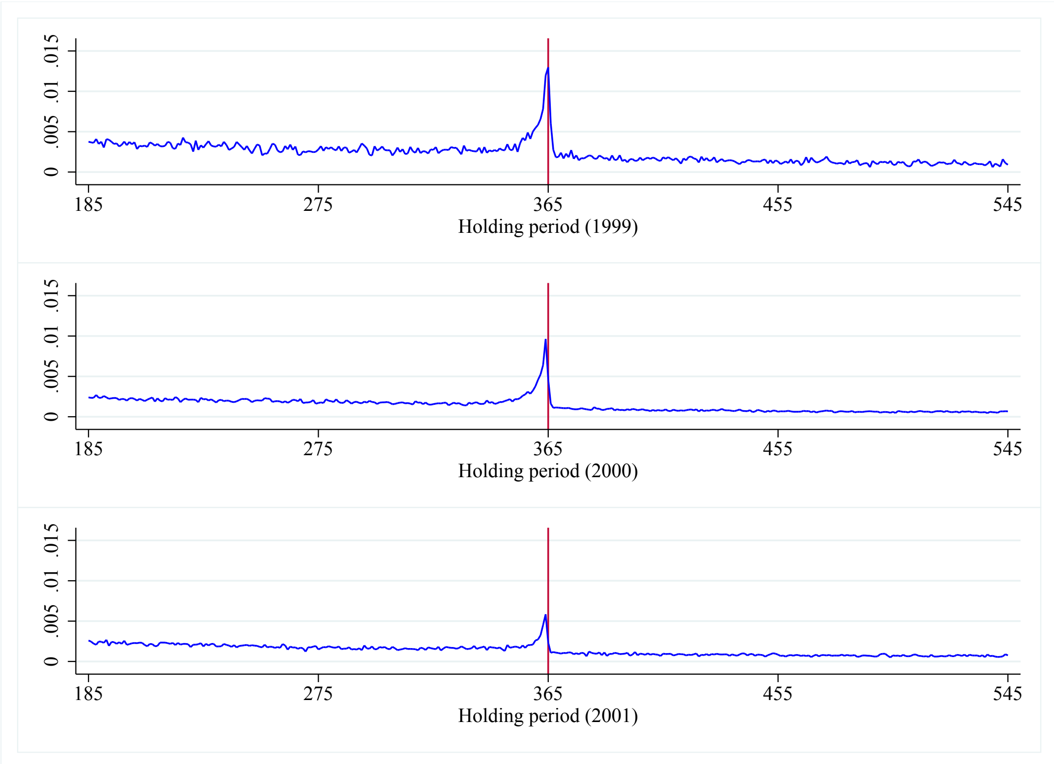
*Notes:*This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2009-2011. Included are share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which gains were taxable. Estimates for 2009 are based on 19304 investors and 11.5 million holding period share package observations. Estimates for 2010 are based on 22902 investors and 10.1 million holding period share package observations. Estimates for 2011 are based on 22129 investors and 10.3 million holding period share package observations.

Figure 53: Hazard-Rate Regressions by Year: Gains, Post Years 2012-2014



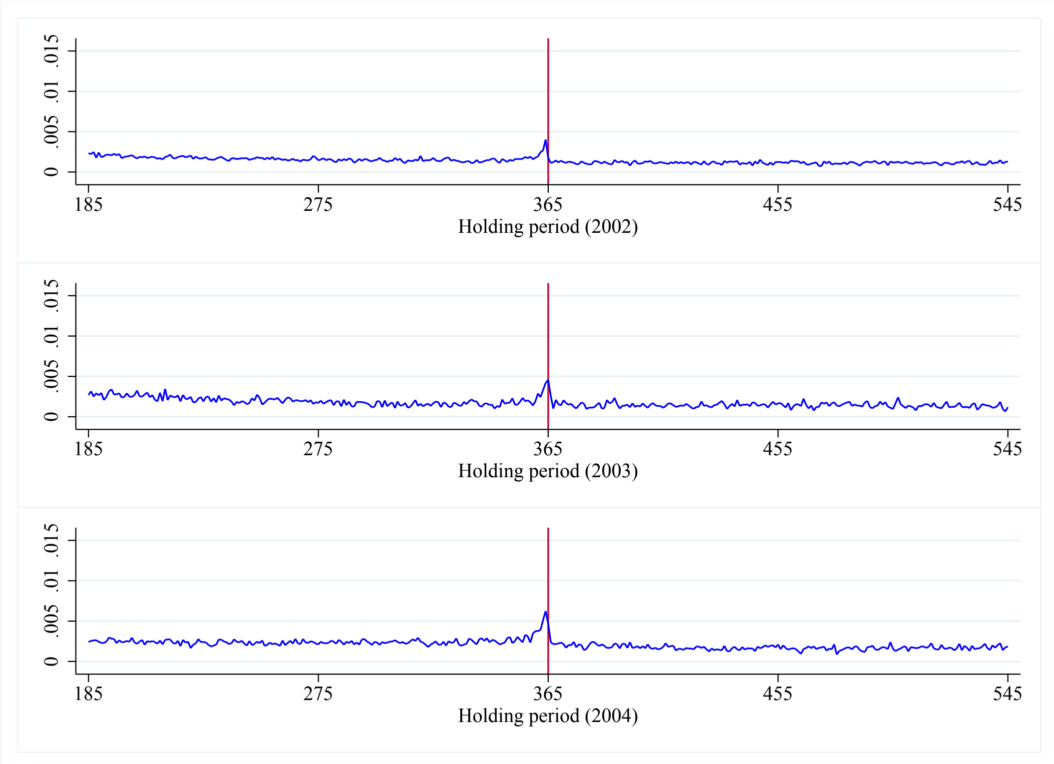
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2012-2014. Included are share packages with prices above the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Gain_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which gains were taxable. Estimates for 2012 are based on 19350 investors and 9.9 million holding period share package observations. Estimates for 2013 are based on 22118 investors and 13.6 million holding period share package observations. Estimates for 2014 are based on 22356 investors and 10.3 million holding period share package observations.

Figure 54: Hazard-Rate Regressions by Year: Losses, Pre Years 1999-2001



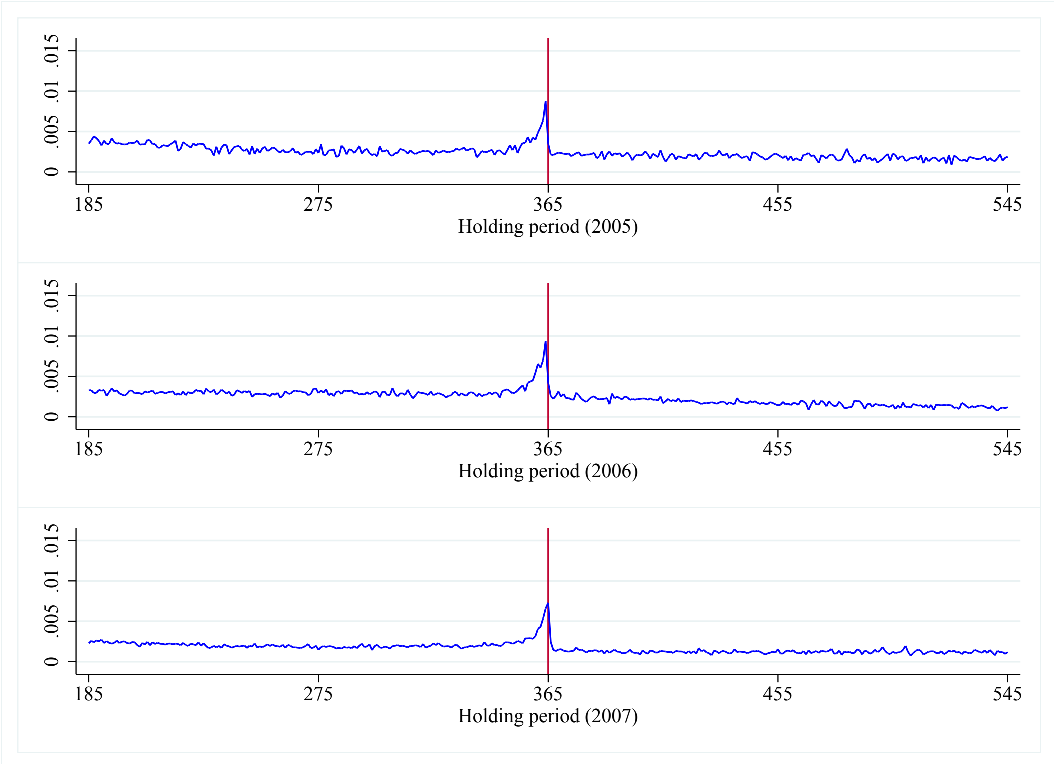
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 1999-2001. Included are share packages with prices below the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which losses could be used to offset gains. Estimates for 1999 are based on 19527 investors and 8.1 million holding period share package observations. Estimates for 2000 are based on 44045 investors and 53.4 million holding period share package observations. Estimates for 1999 are based on 32947 investors and 30.1 million holding period share package observations.

Figure 55: Hazard-Rate Regressions by Year: Losses, Pre Years 2002-2004



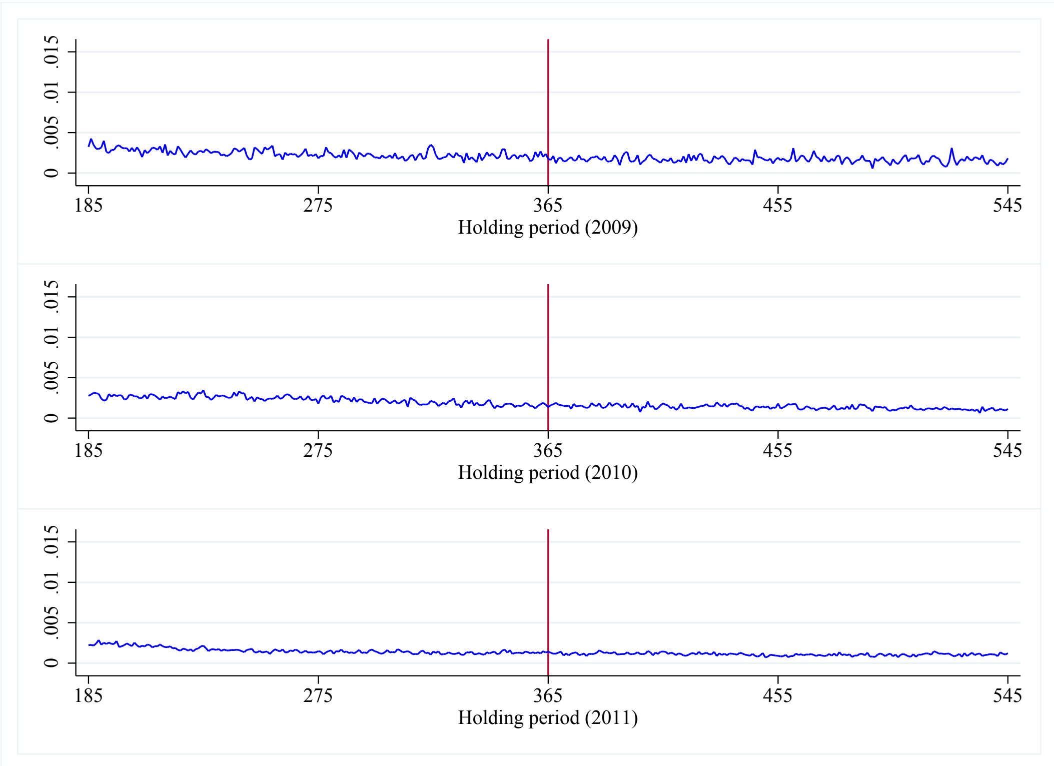
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2002-2004. Included are share packages with prices below the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which losses could be used to offset gains. Estimates for 2002 are based on 26297 investors and 18.6 million holding period share package observations. Estimates for 2003 are based on 17310 investors and 6.6 million holding period share package observations. Estimates for 2004 are based on 23255 investors and 10.5 million holding period share package observations.

Figure 56: Hazard-Rate Regressions by Year: Losses, Pre Years 2005-2007



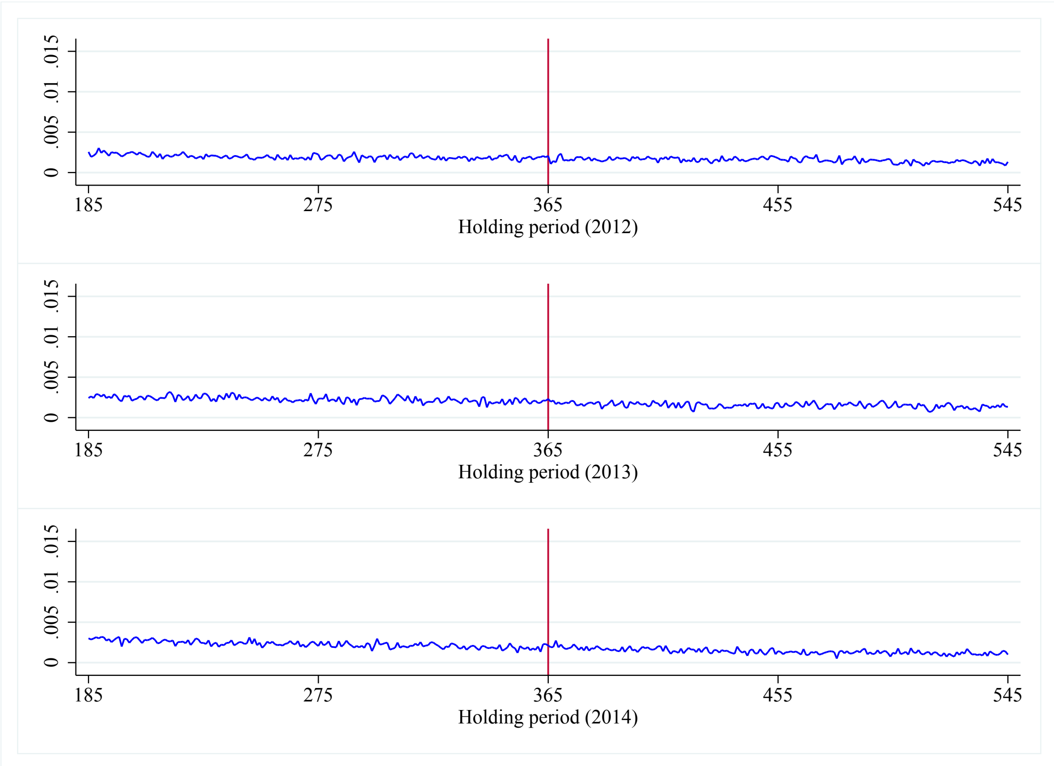
Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2005-2007. Included are share packages with prices below the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which losses could be used to offset gains. Estimates for 2005 are based on 21079 investors and 7.1 million holding period share package observations. Estimates for 2006 are based on 28061 investors and 12.8 million holding period share package observations. Estimates for 2007 are based on 30331 investors and 27.1 million holding period share package observations.

Figure 57: Hazard-Rate Regressions by Year: Losses, Post Years 2009-2011



Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2009-2011. Included are share packages with prices below the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which losses could be used to offset gains (prior to 2009). Estimates for 2009 are based on 15025 investors and 5.2 million holding period share package observations. Estimates for 2010 are based on 23644 investors and 13 million holding period share package observations. Estimates for 2011 are based on 27081 investors and 20.9 million holding period share package observations.

Figure 58: Hazard-Rate Regressions by Year: Losses, Post Years 2012-2014



Notes: This figure displays the hazard-rate regressions estimates for each day of the holdingperiod separately for the years 2012-2014. Included are share packages with prices below the purchase price. Coefficients indicate the probability that a share-package is sold on this holding-period day. Coefficients and on the investor level clustered standard errors stem from a series of regressions of the form $Sell_{ijd} = \beta_0 + \varepsilon_{ijd}$ if $\mathbb{1}(Loss_{ijd}) = 1$. Regressions are estimated for each day of the holding period. The blue line represents estimates for β_0. The vertical red line at day 365 marks the last day in which losses could be used to offset gains (prior to 2009). Estimates for 2012 are based on 19277 investors and 9.9 million holding period share package observations. Estimates for 2013 are based on 19761 investors and 8 million holding period share package observations. Estimates for 2014 are based on 21803 investors and 10.7 million holding period share package observations.

Florian Buhlmann

CV

Education

02/19 – 03/19 **Research stay** *USA*, Massachusetts Institute of Technology (MIT)

09/14 – to date **PhD Program (CDSB)** University of Mannheim Topic: "Essays in empirical taxation and empirical public economics"
Supervised by Johannes Voget and Andreas Peichl

10/12 – 09/14 **Master of Science Economics** University of Bonn

09/13 – 03/14 **Semester Abroad** *France*, ENSAE ParisTech

10/09 – 10/12 **Bachelor of Science Volkswirtschaftslehre** University of Bonn

08/07 – 07/09 **Apprenticeship** Frankfurter Volksbank eG

07/07 **Higher Education Entrance Qualification (A-levels)** Adolf-Reichwein-Schule Neu-Anspach

Publications and Research Projects

Publication *Buhlmann, Florian, Benjamin Elsner and Andreas Peichl (2018) Tax refunds and income manipulation: evidence from the EITC, International Tax and Public Finance*

Seminars and Conference Presentations

2019 *Massachusetts Institute of Technology, 75th IIPF Annual Congress (Glasgow), 112th NTA Annual Conference on Taxation (Tampa)*

2018 *111th NTA Annual Conference on Taxation (New Orleans), ZEW/SFB Conference (Mannheim)*

2017 *110th NTA Annual Conference on Taxation (Philadelphia), 73th IIPF Annual Congress (Tokyo)*

Selected Consulting Projects

2017 **Bertelsmann Stiftung** *Grenzbelastungen im Steuer-, Abgaben- und Transfersystem – Fehlanreize, Reformoptionen und ihre Wirkungen auf inklusives Wachstum* joined with Max Löffler and Andreas Peichl

2017 – 2019 **BMWI** *Short Expertises: Forschungsrahmenvertrag "Ökonomische Bewertung verschiedener Reformoptionen im deutschen Steuer- und Transfersystem"* joined with Holger Bonin, Eric Sommer and Holger Stichnoth

2018 **BMFSJF** *Arbeitsangebotseffekte einer Reform des Kinderzuschlags* joined with Holger Bonin, Eric Sommer and Holger Stichnoth

2016 **European Commission** *Convergence of Unemployment Benefit Schemes in Europe* joined with Mathias Dolls, Carla Krolage and a team of the JRC in Sevilla