Public Policy Complexity, Individual Behavior, and Inequality

Luisa Wallossek

LMU Munich
2024

Public Policy Complexity, Individual Behavior, and Inequality

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

2024

vorgelegt von
Luisa Wallossek

Referent: Prof. Dr. Dominik Sachs
Korreferent: Prof. Dr. Joachim Winter
Promotionsabschlussberatung: 10. Juli 2024

Tag der mündlichen Prüfung: 02. Juli 2024
Namen der Berichterstatter: Dominik Sachs, Joachim Winter, Youssef Benzarti

To whom it may concern.

Acknowledgments

This dissertation would not have been possible without the support of many people.

I thank Dominik Sachs for relentless support, motivation, and for teaching me intellectual rigor. Thank you for pushing me further than I believed I could go and for always seeing the big picture. I am a much better researcher because of you. I thank Joachim Winter for invaluable guidance and consistent support throughout this dissertation. Your expertise and approachability have been a constant source of encouragement and inspiration. I thank Youssef Benzarti for fantastic mentoring and enthusiasm for my research agenda. Your advice and support over the past two years have been invaluable, and I deeply appreciate the mentorship and opportunities you've provided.

I am very grateful for having fantastic co-authors with whom I was lucky to collaborate for this dissertation. Thanks to Elena Herold for co-authoring my job market paper. You are a cool co-author – not a regular co-author. Thank you for pulling this off together! Thanks to Tobias Hauck for being a fantastic co-author and sharing enthusiasm about anything tax. Working with you has been a pleasure and I am grateful for your continued support over the past years. Thanks to Youssef Benzarti for co-authoring my first publication and for boosting my excitement about tax filing and complexity. Thanks to Tabea Bucher-Koenen and Joachim Winter for a supportive and insightful collaboration, and the patience throughout the research process.

This journey would have been much less fun without the many people I met along the way at MGSE. Special thanks to Emilio Esguerra and Leonie Bielefeld for countless espressi, endless hours of gossip, and for making Wokistan great again. Your support has significantly contributed to maintaining my sanity over the past year. Thanks to Leonie Koch for sharing code and frustration about data, academia, and the world. Thanks to Julian Heid for sharing an office and navigating the maze of administration (#TeamDean). Thanks to Cristina Rujan, Friederike Reichel, Jelena Todorović Bojović, and Sarah Kiesl-Reiter for refreshing coffee breaks and fruitful discussions about academia and the real world.

Over the past years I received outstanding support from the LMU economics department. I thank junior and senior faculty for engaging discussions, feedback, and encouragement. Special thanks to Mirko Wiederholt for guidance around the pitfalls of the academic job market, to Jakob Miethe for being a great mentor, and to Martina Magli for sharing valuable advice for academia and life, as well as many coffee breaks. I also thank Silke Englmaier for extensive support over the past years, and Sibel Avsar, Caroline Benkert, and Sabine Rospleszcz for assistance with administrative matters. I want to extend my thanks to the faculty and students at the UC Santa Barbara economics department for being incredibly welcoming and for making Santa Barbara a second academic home for me.

During my PhD, I was lucky to travel a lot and spend extended periods abroad. I thank Youssef Benzarti for inviting me twice to UC Santa Barbara, and Emmanuel Saez for inviting me to UC Berkeley. I am grateful for funding from the German Research Foundation (DFG) through the GRK 1928 "Microeconomic Determinants of

Labor Productivity" that allowed these and other travels.

For most chapters of this dissertation, I use administrative data. I thank the research data center of the federal statistical office, particularly Hannes Fauser, as well as the Bavarian statistical office and the research data center of the German pension insurance, particularly Tatjana Mika, for providing the data and for support throughout my projects.

I am extremely grateful to my friends and family for their continuous love and patience. Thank you for your support - schon all die lange Johr!

Finally, and most importantly, my deepest thanks go to Levin. Thank you for your unconditional love and support, for believing in me every single day, and for always having my back. I am incredibly lucky to be your partner.

Luisa Wallossek
Munich, March 2024

Contents

List of Figures

List of Tables

Preface

While policies are typically designed to achieve a certain goal, the effective outcome may be very different because of how individuals behave in the policy setting. If individuals behave differently than expected, this can hinder the policy's effectiveness or even lead to unintended consequences. A key question in public policy design is therefore how responsive individuals are to financial incentives. Here, complexity is a key factor: in more complex settings, individuals are less responsive to incentives, impeding optimal behavior (Abeler and Jäger 2015).

This dissertation centers around the question how the design of public policies contributes to inequality. Throughout the five chapters of this dissertation, I focus on two dimensions of inequality. The first dimension is income inequality, one of the major concerns of today's societies. Countries around the world use redistributive policies to reduce income inequality and support those with low income via the tax and transfer system. However, those with low income are often unresponsive to incentives from public policy. A large literature documents incomplete take-up of tax and welfare benefits (for an overview see Currie 2006).

Incomplete take-up is higher in more complex settings, and increasing salience and simplicity raises take-up of benefits for eligible taxpayers (Bhargava and Manoli 2015). At the same time, the (very) rich take advantage of the complexity of the tax system: they legally avoid taxes by optimizing under complex rules and exploiting loopholes in the tax code (Saez and Zucman 2019). In addition, rich taxpayers also illegally evade taxes (Alstadsæter, Johannesen, and Zucman 2019; Guyton, Langetieg, Reck, Risch, and Zucman 2021), hoping it remains undetected in the complex tax system. This differential behavior over the income distribution is in line with the finding that financial literacy is correlated with income (Lusardi and Mitchell 2014).

Importantly, the degree of complexity of policy design is endogenous to policymakers, allowing them to enhance equity by reducing complexity. Another way for policymakers is to automate the payment of certain benefits and transfers, rather than requiring active behavior from taxpayers in order to receive benefits they are eligible for. However, policymakers can also deliberately explore individuals' passive behavior in complex settings. Specifically, policy design commonly features default settings that aim at nudging individuals into certain choices.

The second inequality dimension that studied in this dissertation is gender inequality. Albeit remarkable progress in female labor force participation and wages has been made in the past, labor markets today are still characterized by sizeable and persistent gender gaps. For instance, women's earnings in the EU are still 36% lower than men's (Eurostat 2023). Improving gender equality is one of the main policy goals of the 21^st century. The United Nations (2024) want to achieve gender equality as one of their sustainable development goals and the European Commission (2024) is working towards a gender-equal Europe with its gender equality strategy.

Many public policies relate to gender and gender inequality. These policies span a broad

range and include the taxation of couples, divorce law, parental leave regulations, and the provision of public child care. Some of these policies are specifically targeted at increasing gender equality; yet, they do not necessarily achieve their goals. For instance, parental leave and child-care policies have only very limited impact on gender equality in the labor market (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2023). At the same time, many public policies remain in place that are impeding gender equality by dis-incentivizing female labor supply. While explicit gender biases in the tax and transfer system are rare, particularly income taxation commonly features an implicit gender bias (OECD 2022b) by imposing high marginal tax rates for secondary earners, who are predominantly women.

CHAPTER 1, which is joint work with Elena Herold, analyzes the role of marriage for the gender earnings gap by asking: What happens to earnings upon marriage? We show that in addition to a child-related earnings gap (the "child penalty"), there is a marriage earnings gap. Women's earnings drop when they get married and this cannot be explained by the arrival of children. Combining administrative and survey data from Germany we find that women's earnings decrease continuously in the first years after marriage, before the marriage gap stabilizes at a constant level of about 20%. We then decompose the marriage earnings gap and show that married women reduce their hours or stop working all together, but do not a face a wage penalty. Using cross-country survey data, we estimate the marriage gap for more than 20 countries and find that Germany is no outlier: the average gap in Europe is of similar size than the gap in Germany.

We then provide explanations for the marriage earnings gap, with a focus on the role of incentives provided by the government. Marriage is a public policy with rules defined by policymakers. These rules often imply changes in labor supply incentives upon marriage. To understand to what extent the marriage earnings gap results from policy design, we explore variation in incentives in the German setting. We first show that joint taxation of married spouses contributes to the marriage earnings gap: women reduce their earnings more when facing larger increases in marginal tax rates under joint taxation. This explains 30% of the marriage earnings gap. At the same time, we do not find that spousal earnings respond to changes in labor supply incentives from a change in the divorce law.

While tax incentives contribute to the marriage earnings gap, the gap is not only explained by financial incentives. Women experience a drop in earnings when getting married, even if they are primary earners and their marginal tax rate decreases with joint taxation. At the same time, male earnings do not decrease, even if they are secondary earners. This gender difference in responsiveness to tax incentives suggests that gender norms play a role. To support this, we first document a simple correlation using survey data: the marriage earnings gap is larger for women who share more traditional views when asked about gender norms. We then leverage quasi-exogenous variation from the German separation and reunification. We compare individuals who grew up in East vs. West Germany during the country's separation, characterized by very different gender norms, and got married after reunification under identical legal institutions. We find that the marriage earnings gap is larger for women who grew up in the West with more traditional gender norms.

In CHAPTER 2, which is joint work with Tobias Hauck, we study incomplete take-up

in the tax filing context. We show that exempting taxpayers from the duty to file an income tax return weakens effective redistribution by increasing effective tax rates at the bottom of the income distribution. Many countries exempt taxpayers from their tax filing duty, typically employees with wage income only for whom third party reporting via employers ensures that they do not owe taxes. Since tax filing is known to be tedious and costly for taxpayers (e.g., Benzarti 2020), making it optional seems like an innocuous policy that benefits individuals.

Using German administrative income tax data, we show that allowing for tax non-filing comes with significant monetary costs for individuals: non-filers pay too much taxes because of tax over-withholding. We provide suggestive evidence that leaving money on the table by not filing an income tax return cannot be explained by filing costs, implying that taxpayers face other barriers, i.e., informational or behavioral frictions. Since non-filing is most common for low-income taxpayers, it disproportionately increases effective tax burden at the bottom of the income distribution. While seemingly innocuous, optional non-filing weakens the effective progressivity. While it is well established that redistribution is dampened by tax evasion at the top of the income distribution (e.g., Alstadsæter, Johannesen, and Zucman 2019), we are the first to show that optional non-filing dampens effective redistribution by inflating effective tax rates at the other end of the distribution. Non-filing thus effectively acts like reverse evasion: it impedes effective redistribution since low-income taxpayers overpay taxes because of their passive, but legal behavior. If the tax schedule maps the intended degree of redistribution, the straightforward policy implication is to realign effective and statutory taxation. There are several policies that can help reach this, including automatic refunds and the provision of pre-populated forms.

In a complex income tax system, taxpayers do not only refrain from tax filing even when it is beneficial for them, they also make mistakes when filing. Chapter 3 shows that taxpayers often itemize deductions although itemizing has no benefit for them. This behavior is non-optimal since it is strictly dominated by not itemizing deductions, which has also zero benefits but at the same time zero costs. Using German administrative income tax data, I document this type of mistake for 29 percent of tax returns. Taxpayers itemize although the sum of itemized deduction does not exceed the standard deduction that they are entitled too anyways, and/or itemize although their tax liability is already zero.

This non-optimal tax filing behavior reveals information about tax literacy on the individual level. By revealed preferences, taxpayers who make an itemizing mistake expect the benefits to exceed the costs of itemizing. This implies that they know that itemizing can reduce their tax liability. At the same time, they lack tax knowledge to understand that in their specific case there is no benefit. I document that itemizing increases with income, while low-income taxpayers are most likely to make a mistake, suggesting that tax literacy increases with income. This is in line with patterns of financial literacy, which is higher for higher income individuals (Lusardi and Mitchell 2014).

Chapter 4, which is co-authored with Tabea Bucher-Koenen and Joachim Winter, studies individual behavior in a less complex context. We show that the power of automatic pension enrollment is limited in a setting where the choice architecture is simple. Since old-age poverty is one of the major concerns in aging societies, policymakers often want to help individuals to save more for retirement. One tool that has gained a lot

of attention in recent years, both in economic research and in policy making, is automatically enrolling employees in a pension plan they would otherwise have to actively opt in to. The literature commonly studies complex settings, typically in the context of employer provided pensions, such as 401(k) plans in the US. The common result is that automatic enrollment is a powerful tool that can nudge large shares of individuals into enrolling in pensions (e.g., Madrian and Shea 2001). For this project, we exploit a natural experiment, the introduction of automatic enrollment in the public pension insurance for low-income employees in Germany. Different from most private and firm pension plans, individuals can only choose between enrollment yes or no, allowing us to cleanly estimate the effect of the default, absent other potential confounding factors such as switching costs, choice overload, or misperceiving the default as an investment recommendation.

We find that the power of the default is limited in this setting: automatic enrollment increases the enrollment share but leaves the majority of individuals without enrollment. Put differently, most individuals actively opt out. The effect of the default differs across demographic groups and those who are most likely to stick to it are not necessarily those for whom enrollment is most beneficial. For instance, young individuals and those with non-German citizenship are more likely to stick to the default. The main takeaway is that understanding both the limits and heterogeneity in the power of defaults in pension settings is key for designing policies aiming to enhance old-age savings.

Finally, CHAPTER 5, co-authored with Youssef Benzarti, provides novel estimates on measuring income tax complexity and the cost of filing taxes over time and in different countries. To assess complexity and filing costs from an individual perspective, we first ran a survey among US taxpayers. We find that the majority of taxpayers perceive tax filing as (too) complex and wish for a simplified tax filing system. In addition, taxpayers experience that tax complexity and filing costs have been increasing over time, making the system today more complex than in the past. This subjective measure is in line with our second objective measure for complexity: we use the number of words in the tax code as a proxy for tax complexity and compliance costs. This measure shows that tax complexity has been increasing over the past decades, both in the US and in other countries (our sample includes Canada, France, Germany, Morocco, and Switzerland). This result is supported by alternative measures for complexity in the US context, including the number of forms filed per capita, operating costs of the Internal Revenue Service, and Google searches related to tax filing.

Income tax complexity is costly for taxpayers. The average taxpayer in our survey spends four hours on filing their taxes every year, and US taxpayers typically pay for software that allows them to self-file their taxes. In addition to individual costs, we find that taxpayers believe that complexity fosters evasion and makes the income tax system less fair. The majority of taxpayers would be willing to pay for a simpler tax system where filing is less cumbersome, i.e., via pre-populated tax returns. We document broad support for simplification among US taxpayers – across political affiliation. Our results suggest leeway for political action on reducing tax complexity.

Chapter 1

The Marriage Earnings Gap*

1.1 Introduction

Women have lower earnings than men. When combining labor force participation, hours worked, and hourly wages, the average gender earnings gap in Europe is 36% (Eurostat 2023). It is well established that the earnings gap widens when women become mothers (e.g. Angelov, Johansson, and Lindahl 2016; Bertrand, Goldin, and Katz 2010; Kleven, Landais, and Søgaard 2019), often referred to as "child penalty." In this paper, we show that women's and men's earnings also diverge after another event: marriage.

Although most people marry eventually (United Nations 2019), the role of marriage for earnings, distinct from having children, is not fully understood. Using survey data for a large set of European countries, we document that women's labor force participation drops when they get married and that this marriage gap persists when accounting for the child penalty. For the main part of the paper, we focus on Germany, where we can precisely estimate the marriage earnings gap with administrative data and decompose it into participation, hours worked, and wages. In addition, the German setting allows us to analyze the extent to which financial incentives from public policy design explain the marriage earnings gap.

To estimate the marriage earnings gap, we employ an event-study approach that accounts for both the time relative to marriage and the time relative to childbirth. Although the timing of marriage and childbirth is often correlated, variation in the exact timing between the two events allows us to isolate the marriage earnings gap. We use monthly panel data from the German public pension insurance and link this administrative data with annual survey data from the largest German household survey (SOEPRV). We observe monthly earnings, as well as the month of marriage and childbirth for a subsample of the data. Monthly data is preferred over annual data here because marriage and childbirth rarely occur in the same month but more often in the same year: 17% of married women with children marry and give birth to their first child in the same year, but only 1% do so in the same month. However, we show that in the medium run, yearly and monthly data yield almost identical results for the marriage earnings gap. The linked survey data adds information that is key for understanding gender gaps but is typically not covered by administrative data, including hours worked and attitudes towards gender norms.

Using this linked data set, we estimate a significant and persistent marriage earnings gap. Female and male earnings are on similar trajectories until marriage when female earnings drop while male earnings stabilize. Female earnings decrease continuously in the first years after marriage before the gap stabilizes at a constant level: five to ten

*This chapter is based on co-authored work with Elena Herold.

years after marriage women experience a marriage earnings gap of 20%. The marriage earnings gap persists for spouses who are already cohabiting before marriage, ruling out that the gap is explained by couples moving in together in the year of marriage.

We then decompose the marriage earnings gap into the extensive margin (participation), intensive margin (hours worked), and hourly wages. We first first document the extensive margin marriage gap: After marriage, the share of working women decreases by 10%. Employing time-use data from the linked survey, we then estimate the intensive margin marriage gap: Conditional on working, married women reduce their working hours by 18%. For women, spending fewer hours on paid labor market work correlates with spending more hours on unpaid household work, which increases by 16%. By contrast, men do not change their time-use behavior after getting married. In terms of wages, we do not find a marriage wage penalty for women, nor a marriage wage premium for men.

We then analyze the extent to which the marriage earnings gap is explained by incentives from public policy. Individuals get married for various reasons, including love, norms, family planning, and financial reasons – but all enter the same default marriage contract, which is designed by the government. By defining the rules for the contract, policy makers can use marriage as a public policy tool to incentivize or disincentivize spousal labor supply. We study the impact of labor supply incentives from two main areas of marriage-related public policy: joint income taxation and divorce law. We exploit heterogeneity in tax rate changes and a divorce law reform that cuts ex-spousal alimony payments for secondary earners.

We first analyze the role of labor supply incentives from joint taxation. Married couples in Germany are typically taxed jointly. Under joint taxation, spouses face the same marginal tax rate (MTR) because income taxes are defined as a function of household income rather than individual income. With a progressive income tax schedule, this commonly implies a higher MTR for the secondary earner and a lower MTR for the primary earner (similar to joint taxation in other countries such as the US). Joint taxation thus disincentivizes the labor supply of secondary earner – in different sex couples, those are predominantly women. To analyze the effect of tax incentives on the marriage earnings gap, we use administrative income tax data for the full population of German taxpayers from the Taxpayer Panel (TPP). Together with Koch (2024), we developed a novel approach for linking spouses based on their tax identifiers in periods before they are married and file jointly.[1] To the best of our knowledge, we are the first to provide this linkage in the TPP data. The linkage is necessary to analyze changes in MTR under joint taxation, which is typically not possible because only one spouse can be traced back pre marriage.

We exploit cross-sectional heterogeneity in MTR changes under joint taxation created by the non-linear German income tax schedule to estimate the income elasticity. Building on the standard approach for estimating elasticities (e.g., Gruber and Saez 2002), we instrument for the change in MTR by using a hypothetical mechanical change in MTR based on pre-marriage income as an instrument. The obvious threat to identification here is that the MTR change is endogenous to the income level and within household

[1] Because individual tax identifiers were only introduced in 2010, the data only allows to link spouses who marry in 2011 or later.

income distribution, both of which are likely correlated with unobservables that impact labor supply, such as career preferences. To account for this, we include fixed effects for bins of individual income and income share when estimating the elasticity. We show that our results are robust to using different bin sizes. Intuitively, the identifying assumption is that, while having a high or low income and marrying someone with a high or low income is endogenous, the exact amount of own and spousal income can be considered as good as random.

We find that tax incentives from joint taxation increase the marriage earnings gap for women. We estimate an elasticity of gross labor income with respect to the net-of-marginal tax rate that ranges from 0.3 to 1.1 in the first years after marriage. A back-of-the envelope calculation indicates that tax incentives can explain 26% of the marriage earnings gap for women. For men, we do not find an effect of tax incentives on their earnings. Qualitatively, this result is in line with the finding that married women have a high elasticity of taxable income, while married men have a low(er) elasticity (Blau and Kahn 2007; Hermle and Peichl 2018). We also show that women experience a marriage earnings gap even if they are primary earners and thus face an MTR decrease from joint taxation, while men do not reduce their earnings if they are secondary earners. These results suggest that spousal earnings around marriage do not only respond to financial incentives but are impacted by something else, i.e., gender norms, which we exploit in more detail later.

Second, we analyze the impact of labor supply incentives from post-divorce alimony payments by exploiting a reform of the German divorce law in 2008. If eligible for post-divorce alimony, secondary earners receive monthly income transfers from their ex-spouse post divorce. With alimony, marriage acts as insurance for secondary earners who work less during marriage – in most cases women. They face a risk of losing lifetime income, both from working less during marriage and from accumulating less human capital which may translate into lower earnings post divorce. Alimony insures against this risk by increasing the expected lifetime income via transfers from the ex-spouse in case of a divorce. The 2008 reform drastically cut the eligibility and duration of ex-spousal alimony. We exploit this exogenous variation to analyze how the indirect labor supply incentives from alimony affect the marriage earnings gap. Analyzing the earnings of the full population of German taxpayers who got married in the years around the reform (about two million individuals), we find no change in the marriage earnings gap post reform, neither for new marriages nor for those who were married already when the reform was enacted.

Taken together, these two results draw a nuanced picture of the role of labor supply incentives provided by marriage related public policy. Direct tax incentives from joint taxation impact the marriage earnings gap, while indirect incentives from alimony do not have an effect. One potential explanation for this is the difference in immediacy: income is taxed immediately, but alimony payments accrue only in the future. However, Artmann, Fuchs-Schündeln, and Giupponi (2023) find that female labor supply responds to changes in pension entitlements, suggesting that our sample is in general forward looking and responds to incentives that materialize only in the future. Another potential explanation is that alimony is relevant only if spouses divorce. Limited responsiveness to alimony may be explained by spouses underestimating their divorce probability (Berresheim and Koll 2023).

To test the role of gender norms, we leverage quasi-exogenous variation in gender norms from the German separation and reunification. Having been divided for more than 40 years, East and West Germany had very different institutions, cultures, and gender norms, particularly with respect to women's participation in the labor force. While the West German Federal Republic of Germany (FRG) was characterized by male breadwinner marriages with many stay-at-home wives, the East German Democratic Republic (GDR) had a high female labor force participation and many dual-earner couples. Differences in preferences and norms from East and West are persistent even long after reunification (Alesina and Fuchs-Schündeln 2007; Boelmann, Raute, and Schönberg 2021), allowing us to exploit the quasi exogenous variation in gender norms: We compare the marriage earnings gap for individuals born before reunification and who married post reunification. Being born in the East vs. the West, they grew up with different gender norms and different legal institutions. When they married, they faced the exact same legal institutions in reunified Germany.

We show that, while the east-west income difference for women exhibits no time trend before marriage, an event-time related east-west gap opens at marriage. After marriage, women born in the east have relatively higher income, i.e., a smaller marriage earnings gap, than women born in the west. This finding highlights the relevance of gender norms for the marriage earnings gap. Using stated views on gender norms from the survey data, we also document a larger marriage earnings gap for women with more traditional views, providing further suggestive evidence for the role of gender norms. These results complement our finding of gender differences in responsiveness to tax incentives under joint taxation, which suggest that the observed spousal labor supply around marriage is affected by an interplay of financial incentives from public policy and gender norms.

The first main contribution of this paper is to estimate the marriage earnings gap net of the child penalty. We analyze how earnings change when individuals get married and show that female earnings drop after marriage. We add to a broad literature on marriage and income that mostly studies income and wage differences between married and unmarried individuals (e.g., Hill 1979; Juhn and McCue 2016, 2017; Korenman and Neumark 1991). An earlier literature attributed wage differences to changes in productivity after marriage, based on the hypothesis of household specialization because of comparative advantages (e.g. Becker 1985, 1991). More recent research has questioned the causal impact of marriage on productivity (e.g., Killewald and Gough 2013; Killewald and Lundberg 2017; Pilossoph and Wee 2021), and the role of comparative advantages (Siminski and Yetsenga 2022). Our findings align with these results: While we do not find evidence for a marriage wage penalty for women, we show that women face a marriage earnings gap because they work less after getting married.

We also contribute to a growing literature that documents a persistent gender earnings gap emerging after the birth of the first child, commonly referred to as the *child penalty* (e.g., Andresen and Nix 2021; Angelov, Johansson, and Lindahl 2016; Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019; Kleven, Landais, and Søgaard 2019).[2] We contribute to the child penalty literature by documenting that in addition to childbirth, women's earnings also drop after marriage. Although childbirth and marriage often

[2]Our paper also relates to research on expectations about the labor market outcomes of mothers (Boneva, Golin, Kaufmann, and Rauh 2022; Kuziemko, Pan, Shen, and Washington 2018).

coincide closely, the role of marriage has been largely overlooked in the child penalty literature. A very recent exception is concurrent work by Kleven, Landais, and Leite-Mariante (2023), who estimate extensive margin "marriage penalties" together with child penalties and find small extensive margin marriage gaps for high-income countries. This result can be reconciled with ours: using linked administrative and survey data from Germany, a high-income country, we show that the total marriage earnings gap is much higher than the extensive margin gap since married women also respond at the intensive margin by reducing their hours worked. Relatedly, Berniell et al. (2022) account for marriage when estimating extensive margin child penalties.[3]

The second main contribution of our paper is to analyze the mechanisms behind the marriage earnings gap. We empirically analyze how spousal earnings respond to labor supply incentives from marriage-related public policy. Here, we contribute to a literature that studies how taxes and transfers related to marital status impact the labor supply of married women. The role of joint income taxation for spousal labor supply has been emphasized by structural estimations (Bick and Fuchs-Schündeln 2018; Borella, De Nardi, and Yang 2023), cross-country comparisons (Bartels and Shupe 2023; Kalíšková 2020), as well as case studies of the introduction (Kalíšková 2014; LaLumia 2008) and abolition (Selin 2014) of joint taxation for spouses. We also contribute to a literature that studies how divorce law impacts individual and spousal decision making, specialization, and labor supply (e.g., Foerster 2022; González and Özcan 2013; Stevenson 2007, 2008; Voena 2015). Our paper contributes to these two strands of literature by empirically analyzing how spousal earnings respond to two different types of labor incentives in the same setting. While we find that spouses do not respond to indirect labor supply incentives from post-divorce alimony, the earnings gap is responsive to direct labor supply incentives from income tax incentives.

We show that in addition to labor supply incentives from taxation, gender norms and culture are important mechanisms behind the marriage earnings gap. Here, we contribute to a large literature on the relevance of gender norms for gender gaps in the labor market (for an overview, see Bertrand 2011; Blau and Kahn 2017). More specifically, we also contribute to a literature that exploits differences in maternal labor supply based on gender norms and culture between East and West Germany (Boelmann, Raute, and Schönberg 2021; Jessen 2022). Related to our findings, Giommoni and Rubolino (2022) show that gender norms constitute a friction for spouses' responses to tax incentives.

The remainder of the paper is set up as follows. Section 1.2 provides information on the institutional background for marriage and divorce law. In section 1.3, we describe the different data sets we use for our analyses and provide some basic descriptive statistics for the different samples. We explain our empirical approach and estimate the marriage earnings gap in section 1.4, also decomposing the marriage earnings gap and providing results for other countries. We then analyze mechanisms behind this gap in section 1.5. Section 1.6 concludes.

[3]Intuitively, not accounting for marriage when estimating child penalties confounds both the event time coefficients and the counterfactual that the coefficients are divided by to obtain the child penalty estimate. As a result, the child penalty estimates do not necessarily change (see Berniell et al. 2022) when marriage biases the event time coefficients. We discuss the details in section 1.4.

1.2 Institutional Background

Marriage as a Public Policy Marriage is a legal contract between two partners. The government defines the rules of this contract, making marriage a public policy. Typically, marriage is incentivized by granting spouses access to several advantages. These advantages can be monetary (e.g., joint taxation benefits, lower inheritance taxes, survivors' pensions) or non-monetary (e.g., simplified access to citizenship, alleviated parenthood regulations). At the same time, marriage typically also with legal and financial obligations for spouses, both during marriage and after divorce (i.e., property division, ex-spousal alimony). For post-divorce obligations, divorce law acts as a default marriage contract.[4] We study two institutional features of marriage as mechanisms behind the marriage earnings gap: joint taxation and ex-spousal alimony.

Joint Taxation Under joint taxation, income taxes are a function of household income rather than individual income. Joint taxation for married spouses exists in many countries, including the United States and several European countries. In Germany, spouses can choose between joint and individual taxation and the majority choose the former. Under joint taxation, spouses are taxed as if both earn exactly the mean household taxable income. Filing jointly is almost always beneficial for spouses with unequal income. This is different to other countries, such as the US or Switzerland, where joint taxation may lead to a tax penalty. Under joint taxation, both spouses face the same marginal tax rate (MTR), which, in most cases, implies a decrease in MTR (labor supply incentive) for the primary earner and an increase in MTR (labor supply disincentive) for the secondary earner, compared to individual taxation. The size of the incentives depends on the level of household income and the within household share of income (see Figure 1.A.2). We exploit this heterogeneity in subsection 1.5.1, where we discuss the MTR incentives in more detail.

Ex-Spousal Alimony Upon divorce, there are two types of financial obligations for ex-spouses. First, wealth that accrued during marriage is typically divided according to a pre-defined rule. The German divorce law is a community property law, where any wealth acquired during marriage is split equally between divorced spouses. Second, divorcees may face alimony obligations for their ex-spouse during post-divorce periods. Under the German divorce law, this applies if the economically weaker ex-spouse cannot financially support themselves adequately (e.g., because of childcare, illness, age, or unemployment).[5] Alimony payments then amount to 3/7 of the difference in net income between ex-spouses, with the alimony-paying spouse being entitled to a fixed minimum of their own net income. In 2008, a reform drastically cut alimony eligibility, as a result of which secondary earners can expect to receive alimony for a shorter period – and often none at all – from their ex-spouse after divorce. We exploit this exogenous variation in subsection 1.5.2, where we also discuss the reform in more detail.

[4] Spouses can deviate from the default contract in certain domains by signing a prenuptial agreement, but rarely do so: in the sample we use for our main analysis, only 8% have a prenuptial agreement.

[5] Alimony payments are meant to support the ex-spouse. In addition, divorced parents typically face child support obligations if they are the non-custodial parent, i.e. if their child lives with their ex-spouse. Children are eligible for child support irrespective of parental income.

1.3 Data

For our main analysis, we utilize two data sets: German administrative pension data linked to survey data from the SOEP (SOEPRV) and German administrative income tax data (TPP). We explain the different data sets below. In addition, we use international SHARE survey data to estimate the marriage gap for a large set of European countries. We describe the SHARE data in section 1.C.

1.3.1 SOEPRV

The SOEPRV data is jointly provided by the Research Data Centre of the German Pension Insurance (FDZ-RV) and the German Institute for Economic Research (DIW) (Goebel et al. 2022). It combines survey data from the Socioeconomic Panel (SOEP) and administrative data from pension accounts (VSKT). The SOEP is the largest German household panel, and has been conducted annually since 1984. It provides household and individual-level data for a representative sample of the German population. The data set includes a unique identifier to link individuals to administrative VSKT data. The record linkage is available for respondents who participated in the 2018 and 2020 SOEP waves and consented to the linkage. Lüthen et al. (2021) provide a detailed description of the record linkage process.

The linked administrative data is a version of the VSKT pension data, provided specifically for the SOEP linkage. The VSKT is a unique administrative data set that combines monthly earnings from the entire biography of individuals with precise marriage and divorce information for a subset of individuals. Because public pension entitlements are split upon divorce, the German Pension Insurance has information on marriage and divorce as soon as individuals get divorced and split their pension entitlements. For this subsample, we use the marriage info from the administrative data. For everyone else, we use the marriage info from the survey data. We benefit from the combination of precise administrative data and rich survey data and use the SOEPRV data to exploit earnings around marriage.

Variables The SOEPRV contains the full set of variables available from the SOEP as well as a set of variables from the VSKT. Most relevant for our analysis, the VSKT data covers monthly gross wage earnings, as well as marriage and divorce dates for divorced individuals. For women, the data also includes the year of childbirth. All variables are on a monthly level. The Research Data Centre of the German Pension Insurance (FDZ-RV) (2022) provides a detailed description of available variables.

Our main outcome is monthly gross earnings. Earnings are measured in so-called *earnings points* (EP), with 1 EP (1/12 EP) corresponding to the average annual (monthly) gross wage income in a given calendar year. Because EPs are normalized by average annual income, they account for overall income growth and are comparable over time.[6] While our main outcome is earnings from the administrative data, we use survey data to analyze other outcomes, i.e., time use and hourly wages. We also use SOEP variables that reveal individual characteristics, such as marriage information, or information on children, if these are not covered in the administrative data.

[6] EPs are top-coded at about 2, with no information on the exact income of individuals with income exceeding this cutoff.

	All	Women	Men
Age at marriage	27	26	28
Share married East Germany	0.22	0.22	0.23
Year of marriage	1988	1989	1988
Share divorced	0.18	0.20	0.17
Years married	13	13	13
Share children (ever)	0.92	0.93	0.92
Gross monthly income t-12	1898	1667	2150
Source VSKT-VA	0.18	0.20	0.17
N	6,283	3,276	3,007

Data: SOEPRV, raw data (unweighted), mean values. The sample contains individuals that are married at some point, aged 18 to 45 at marriage, and observable a minimum of four years before and eight years after marriage. If available, we take the marriage date from the administrative VSKT data. If not, we take the marriage date from the SOEP survey data. Income is measured in the year before marriage and reported in 2015 Euros.

Table 1.1: Descriptive Statistics SOEPRV Sample

Sample The SOEPRV data covers a total of about 12,000 individuals for whom we can link administrative and survey data. The linked data includes mainly employees, since self-employed individuals, civil servants, and farmers are typically not included in the pension records. Our main sample includes 6291 individuals for whom we observe their first marriage, either in the VSKT (if divorced) and/or in the SOEP (if reported in the survey). We include individuals aged 18 to 45 at the time of marriage, for whom we can observe at least four years prior to and eight years after marriage. Table 1.1 provides descriptive statistics. We use this sample for our main analysis in subsection 1.4.2.

1.3.2 Taxpayer Panel

The *Taxpayer Panel* (TPP) is an annual panel of income tax payers for the years 2001 to 2018 provided by the Research Data Center of the German Federal Statistical Office and the statistical offices of the Länder (RDC, 2022). Our sample covers the full population of German income taxpayers born between 1951 and 1998. The TPP contains income tax return data for tax filers as well as employer provided information for non-filers (similar to the W-2 form in the US, available only since 2012). Taxpayers are either individuals or joint filing married spouses. Upon joint filing, one spouse (typically the woman) is added to the other spouse's (typically the man's) spell. Germany introduced individual tax identifiers only in 2010, allowing us to link spouses also before marriage if they married in 2011 or later. To the best of our knowledge, together with Koch (2024), ours is the first study to link German income tax data for spouses before marriage.

We use the TPP data set to analyze the impact of financial incentives on the marriage earnings gap. The advantage of the TPP data set here is twofold. First, it provides the exact taxable income, which is crucial for analyzing tax incentives in subsection 1.5.1. Second, the large sample size from the full population data allows us to test for changes in the marriage earnings gap between year-of-marriage cohorts around the 2008 alimony

Men age at marriage	36
Women age at marriage	34
Year of marriage	2013
Share East Germany (women)	0.13
Share East Germany (men)	0.13
Share children pre marriage (women)	0.22
Share children pre marriage (men)	0.22
Share children (ever)	0.82
Share working t-1 (women)	1.00
Share working t-1 (men)	1.00
Mean gross labor income t-1 (women)	35,881
Mean gross labor income t-1 (men)	49,941
Median gross labor income t-1 (women)	33,531
Median gross labor income t-1 (men)	43,777
Female income share t-1	0.43
MTR t-1 (women)	0.2985
MTR t-1 (men)	0.3376
Hypothetical joint MTR t-1	0.3241
N	192,596

Data: TPP, This table shows the descriptive statistics for the TPP subsample of spouses that can be linked pre-marriage and that we use for our analysis in subsection 1.5.1. For details on the sample restrictions, see subsection 1.5.1. This is the sample we use for estimating Equation 1.5 and 1.6. Numbers are mean values unless specified differently, income in 2015 Euros. N refers to the number of couples.

Table 1.2: Descriptive Statistics TPP – MTR Sample

reform in subsection 1.5.2.

Variables Data on the full taxpayer population is only provided upon individual request for research projects and made available only with a limited number of variables. Our customized data set contains detailed information on labor income and total income, as well as demographics such as marital status, children, age, state, or religion. Joint filing spouses are treated as one taxpayer unit, but the data reports income, etc. on the individual level. All variables are at the annual level.

Sample In subsection 1.5.1 and 1.5.2, we draw two subsamples of newlyweds for our analyses. To assess the impact of joint taxation on the marriage earnings gap (subsection 1.5.1), we draw a subsample of newlyweds who married between 2011 and 2015, whom we observe three years before and after getting married. For this sample, we link the spouses before marriage, which is crucial for analyzing changes in MTR from joint taxation.

For our analysis of the reduction of ex-spousal alimony (subsection 1.5.2), we draw a subsample of newlyweds who married between 2004 and 2012, i.e., +/- 4 years around the 2008 reform date. For this sample, the TPP data does not allow for linking spouses

before marriage, since individual taxpayer IDs were only introduced in 2010. Since the event study estimation does not require to link individuals over periods, we can estimate the marriage earnings gap for this sample as well (we provide more details on this in subsection 1.5.2). Table 1.3 shows the descriptive statistics.

Men age at marriage	33
Women age at marriage	30
Year of marriage	2009
Share married East Germany	0.16
Share divorced	0.16
Years married	5
Share children (ever)	0.82
Men gross labor income t-1	30,848
Women gross labor income t-1	23,844
N	2,165,423

Data: Taxpayer Panel (TPP), mean values. This table shows the descriptive statistics for the TPP subsample we use for our analysis in subsection 1.5.2: individuals who married between 2004 and 2012. *Share divorced*: share of individuals with observed divorce (data until 2018). *Years married*: length of marriage if divorced. *N*: number of individuals in the subsample.

Table 1.3: Descriptive Statistics TPP – Alimony Sample

1.4 Estimating the Marriage Earnings Gap

1.4.1 Framework for Estimating the Marriage Earnings Gap

Naive Approach: Not Accounting for Children To estimate the marriage earnings gap, we first employ an event study approach that closely follows the standard event study approach for child penalties (see e.g., Kleven, Landais, and Søgaard 2019): We regress the outcome of interest Y_{it}^g on a full set of dummies for event time t defined relative to the event of marriage. We refer to this as a *naive approach* because it does not account for child-related earnings changes. Since the timing of marriage is often correlated with the timing of having a first child (see Figure 1.A.3), children are a confounder this naive estimation does not account for. We estimate the following event study regression for the naive approach:

$$Y_{it}^g = \sum_{j \neq r} \beta_j^g \cdot \mathbb{1}[j = t] + \zeta_{cal}^g + \eta_{age}^g + \epsilon_{it}^g, \tag{1.1}$$

where Y_{it}^g denotes the labor market outcome of individual i of gender g at event time t. We estimate Equation 1.1 separately for women and men. The event dummies capture the dynamics of Y_{it}^g at event time $t = j$, yielding the coefficients of interest β_j^g. Event time t is defined relative to the event of marriage at $t = 0$, omitting the reference period r. For estimations with monthly data, we measure t in months and omit $r = t - 12$, i.e. the period 12 months before the month of marriage. For estimations with yearly data,

we measure t in years and omit $r = t - 1$, i.e. the period before the year of marriage. Year-of-age fixed effects η^g_{age} capture potential life-cycle effects that are correlated with age k, and calendar-year fixed effects ζ^g_{cal} account for common time trends. Since we estimate Equation 1.1 separately for women and men, the fixed effects account for potential gender differences in life-cycle and time trends.

Main Specification: Accounting for the Child Penalty For our main specification, we account for the child penalty by adding a full set of event dummies for event time relative to childbirth to our event study specification:

$$\begin{aligned} Y^g_{it} = &\sum_{j \neq r} \beta^g_j \cdot \mathbb{1}[j = t] \\ &+ \sum_k \gamma^{1,g}_k \cdot \mathbb{1}[k = t^{child1}] + \sum_l \gamma^{2,g}_l \cdot \mathbb{1}[l = t^{child2}] + \sum_m \gamma^{3,g}_m \cdot \mathbb{1}[m = t^{child3}] \\ &+ \zeta^g_{cal} + \eta^g_{age} + \epsilon^g_{it}, \end{aligned} \tag{1.2}$$

where t^{child1}, t^{child2}, and t^{child3} denotes the event time relative to the birth of the first, second, and third child, respectively. The coefficients $\gamma^{1,g}_k$, $\gamma^{2,g}_l$, and $\gamma^{3,g}_m$ capture dynamics in Y^g_{it} that correlate with the event time relative to the respective birth.[7] Note that we do not omit any time dummy for $\sum_k \gamma^{1,g}_k$, $\sum_l \gamma^{2,g}_l$, and $\sum_m \gamma^{3,g}_m$. Because not all married individuals have children, the omitted group is individuals who never have a (first, second, third) child. Our coefficient of interest is still β^g_j, now denoting the change in earnings relative to marriage that is not correlated to the event time with respect to having a first, second, or third child. Following the same steps as for Equation 1.1, we estimate Equation 1.2 separately for gender g = w, m, denoting women and men, respectively.

Although the first child is often born relatively soon after marriage, there is large variation in the exact timing (Figure 1.A.3), which allows us to isolate the estimates for β^g_j from $\gamma^{1,g}_k$. A unique feature of the SOEPRV data is that it provides monthly data for the exact date of marriage and childbirth for a subsample of married individuals.[8] With monthly data, marriage and childbirth rarely happen in the same period: Only 1% of married women who have children at some point in their life have their first child in the month of marriage. With yearly data, the variation is mechanically smaller, i.e., more women have their first child in the period of marriage. However, we show that the results remain virtually unchanged when using annual data rather than monthly.

Estimating the Marriage Earnings Gap To estimate the marriage earnings gap, we follow the standard approach for child penalties from Kleven, Landais, Posch, Steinhauer, and Zweimüller (2019). We first estimate Equation 1.1 and 1.2 separately for women and men. This yields gender-specific absolute estimates $\hat{\beta}^g_t$, where g = w, m for women and men, respectively.

[7] Berniell et al. (2022) and Kleven, Landais, and Leite-Mariante (2023) employ a similar approach to estimate child penalties while accounting for event time with respect to marriage.

[8] Those for whom the administrative VSKT data provides marriage info (this is conditional on divorce) and those who married while already in the SOEP sample. We have yearly dates for everyone else, i.e., those without divorce who married before entering the SOEP sample.

We then rewrite the absolute estimates in relative terms:

$$MG_t^g \equiv \hat{\beta}_t^g / E[\tilde{Y}_{it}^g | t], \tag{1.3}$$

where $\tilde{Y}_{it}^g | t$ denotes the counterfactual outcome, i.e., the predicted outcome without the event dummy estimates. The child penalty literature then typically defines the total penalty as $P_t \equiv (\hat{\beta}_t^m - \hat{\beta}_t^w)/E[\tilde{Y}_{it}^w | t]$, which is interpreted as the percentage by which women fall behind men after the event (see, e.g., Kleven, Landais, and Søgaard 2019). This measure does not account for a gender gap in the counterfactual incomes, which is why we only report MG_t^g and do not compute P_t. We provide a detailed explanation in Appendix 1.B.

We follow the same procedure for the estimates from Equation 1.1 and Equation 1.2. It is important to note that not accounting for the child penalty in Equation 1.1 not only confounds $\hat{\beta}_t^g$, but also $E[\tilde{Y}_{it}^g | t]$. The marriage gap estimates (MG_t^g) from Equation 1.1 and Equation 1.2 are thus not directly comparable. This is because accounting for event time with respect to childbirth affects both the numerator (the coefficient for event time with respect to marriage) and denominator (the counterfactual absent the contribution of these estimates).

1.4.2 Main Result: There is a Marriage Earnings Gap

We first estimate the marriage earnings gap with monthly event time for the subsample of individuals for whom we know the exact monthly date of their marriage and the birth of their first child (if they have children) from the SOEPRV data. With monthly data, the two events almost never happen in the same period, and there is large variation in the exact timing of both, as shown in Figure 1.A.3. This allows us to isolate both events. We start with the sample with monthly info as a proof of concept before moving to the full sample, where we show that annual event time yields similar results.

Figure 1.1 plots the results from estimating Equation 1.1 and Equation 1.2 with monthly data. Immediately after marriage, women's monthly gross earnings start decreasing. The gap widens over time before stabilizing at a constant level after some years. Averaging over five to ten years after marriage, the naive estimate $\hat{\beta_w}$ is 1128 € (Figure 1.1a). A large part of this is explained by the child penalty, but the remaining drop is sizable: After accounting for the child penalty, women's monthly gross earnings still decrease by 251 € per month, equivalent to 3012 € in annual earnings. This absolute estimate translates into a relative marriage earnings gap of $MG_t^w = 20\%$ (Figure 1.1b).

We then turn to the full SOEPRV sample, where we can estimate the marriage earnings gap with annual event time. Mechanically, it is more common that marriage and childbirth happen in the same period if time is measured in years rather than months: 17% of married women have their first child in the year of marriage (compared to 1% giving birth in the month of marriage). However, there is still substantial variation (Figure 1.A.3c), allowing us to estimate Equation 1.1 and 1.2 with annual data.

Figure 1.2a plots the absolute estimates for the marriage earnings gap using annual data. Similar to the results with monthly data, women's and men's earnings are on very similar trajectories before marriage.[9] After marriage, a gap opens: Women's earnings decrease, while men's earnings increase slightly. The immediate drop in earnings here

[9] We find earnings increase for $t < 0$, a pattern similar to child penalty estimates for German-

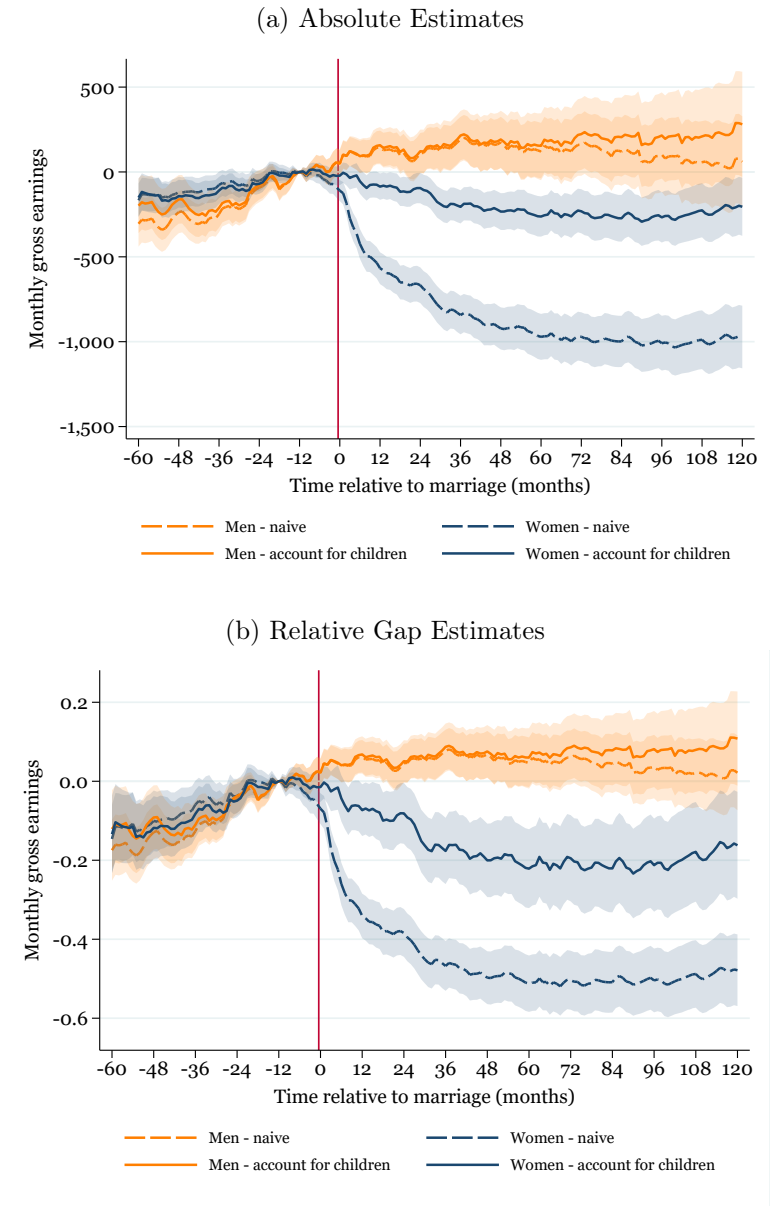
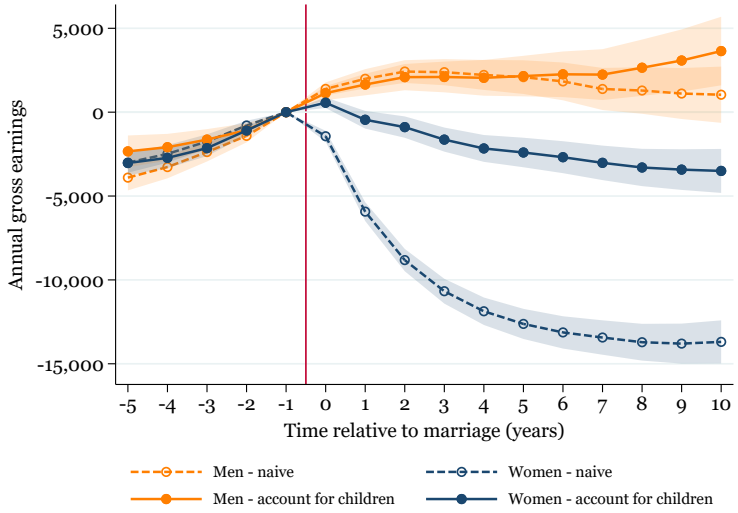


Figure 1.1: Marriage Earnings Gap – Monthly Event Time

Notes: Monthly data from SOEPRV. This figure shows the event study estimates from estimating Equation 1.1 (dashed lines) and Equation 1.2 (solid lines) for monthly gross earnings for the SOEPRV subsample for which we observe the exact month of marriage and childbirth (if an individual has children). Plots show point estimates and 95% confidence intervals. Panel a shows the absolute estimates $\hat{\beta}_t^g$ in 2015 Euros, directly retrieved from estimating Equation 1.2. Panel b shows relative earnings gap estimates for MG_t^g as defined in Equation 1.3.

(a) Absolute Estimates



(b) Relative Estimates

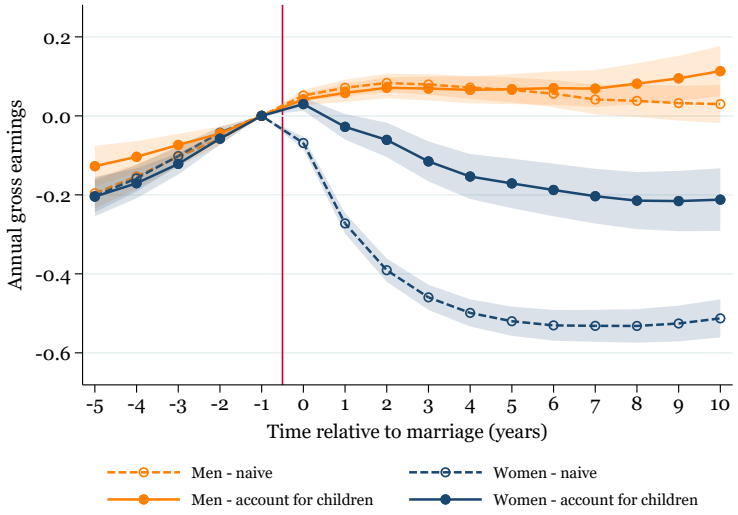


Figure 1.2: Marriage Earnings Gap – Annual Event Time

Notes: Data from SOEPRV, aggregated over years. This figure shows the event study estimates from estimating Equation 1.1 (dashed lines) and Equation 1.2 (solid lines) for annual gross earnings for the full SOEPRV sample. Plots show point estimates and 95% confidence intervals. Panel a shows the absolute estimates $\hat{\beta}_t^g$ in 2015 Euros, retrieved directly from estimating Equation 1.2. Panel b shows relative earnings gap estimates for MG_t^g as defined in Equation 1.3.

is less pronounced than for the monthly estimation. Over time, however, the results are very similar: Averaging over the period five to ten years after marriage, women's annual earnings drop by 3060 €, compared to 3012 € for the subsample with monthly event time (results are also very similar for the naive estimation with 13 404 € vs. 13 536 € with monthly event time).

We then translate the absolute estimate into a relative estimate for the marriage earnings gap. Figure 1.2b plots the resulting estimates for MG_t^g for gender g over event time t. Averaging MG_t^w over five to ten years after marriage, women's earnings drop by 20%. This is almost exactly the same result as for the estimation with monthly event time (20%).

The MG_t^g estimates from Equation 1.1 and Equation 1.2, i.e., with and without accounting for the child penalty, are not directly comparable. Not including event time for children in the naive estimation not only confounds $\hat{\beta}_t^g$, but also $E[\tilde{Y}_{it}^g|t]$, the denominator of $MG_t^g = \hat{\beta}_t^g / E[\tilde{Y}_{it}^g|t]$. A similar argument can be applied to earnings gaps related to children (child penalties). When comparing absolute estimates, marriage clearly confounds the child penalty estimate: The child penalty estimate for annual earnings $\hat{\beta}_t^w$ is 2261 € smaller when accounting for the marriage earnings gap, equivalent to a reduction of 13% (Figure 1.A.5a).[10] When transforming the absolute estimates to the relative child penalty estimates, the difference is much less pronounced (Figure 1.A.5b). This is in line with the results of Berniell et al. (2022), who find no significant difference for child penalties with and without accounting for marriage.

The marriage earnings gap is not explained by a change in the cohabitation status at the event of marriage. Figure 1.A.6 compares the marriage earnings gap for spouses that cohabited already in the year prior to marriage ($t = -1$) and those that did not. The marriage earnings gap remains, and is even larger, for those who are already cohabiting, ruling out that the drop in $t = 0$ is driven by spouses moving in together in that period.

1.4.3 Decomposition of the Marriage Earnings Gap

Earnings are defined as the product of the extensive margin (working yes or no), the intensive margin (hours worked), and hourly wages. In this section, we decompose the change in earnings into its three components. Administrative data informs about earnings and the extensive margin. Linking administrative and survey data, we can add information on hours and back out hourly earnings for a subsample of individuals. This information is only available in survey data. We therefore restrict the analysis of hours and wages to those who are already in the survey population in the years around marriage. Since the outcome from survey data is available on the annual level, we estimate all gaps with annual data here. When the outcome is measured monthly (earnings and extensive margin), we average the monthly data over years.

---

speaking counties (e.g., Boelmann, Raute, and Schönberg 2021; Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019). Intuitively, this means the timing of marriage is correlated with income growth that age and calendar-year fixed effects do not capture. Figure 1.A.4 shows that accounting for career age, defined as time since entering the labor market, reduces income growth for $t < 0$ without visibly impacting the marriage earnings gap estimates.

[10] The absolute difference is not exactly the same as the difference for the marriage earnings gap in Figure 1.2a. This can be explained by the fact that the underlying samples differ: For the marriage earnings gap, everyone in the sample is married, and some have children, whereas for the child penalty,

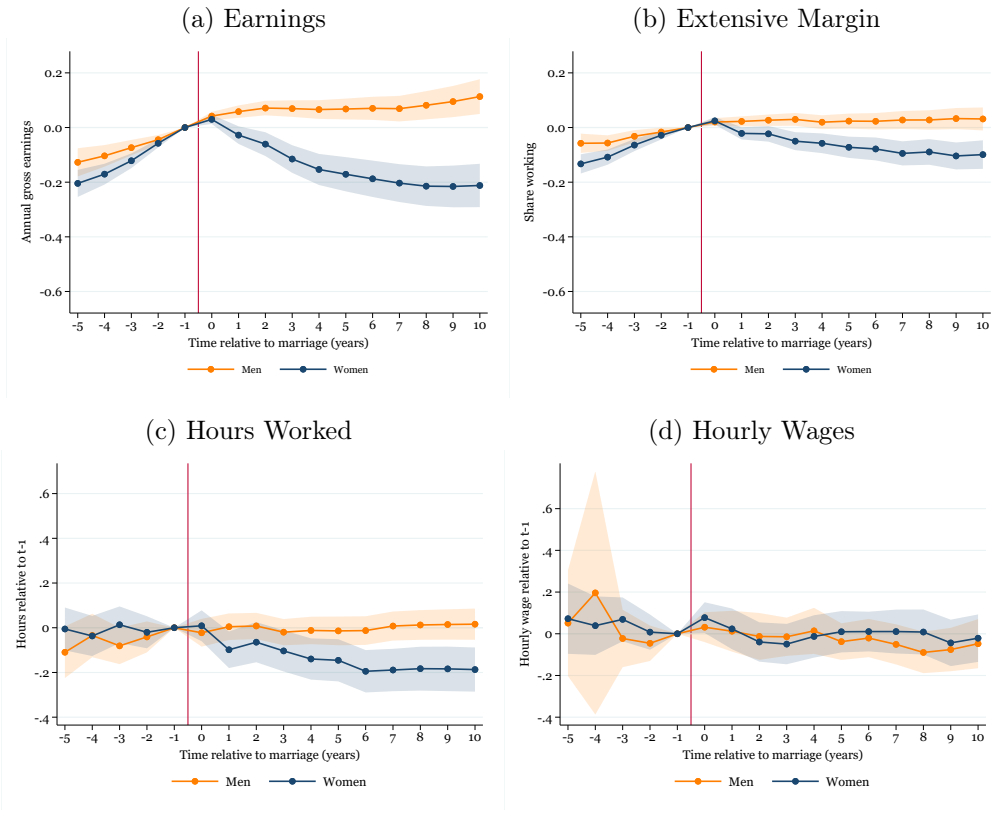


Figure 1.3: Decomposition Marriage Earnings Gap

Notes: Data from SOEPRV. This figure decomposes the overall marriage earnings gap (panel a) into extensive margin (panel b), intensive margin (panel c), and hourly wages (panel d). All panels plot MG_t^g estimates for women and men from estimating Equation 1.2 for different labor market outcomes. For all panels, event time is defined in years. We exclude $t = -1$, i.e., the year before the year of marriage. Panel c plots results for the intensive margin only, conditional on working in a given t. Hourly wages in panel d are defined as earnings divided by hours worked.

Extensive Margin For the extensive margin marriage gap, we estimate Equation 1.2, with the outcome being a dummy for whether individual i is working in month t. We define working as being in either regular employment, marginal employment (so-called mini-jobs), or self-employment, as reported in the administrative data. Figure 1.3b plots the extensive margin marriage gap. We show that women are significantly less likely to work after marriage. Averaging over years five to ten after marriage, women are 10% less likely to be working. In absolute terms, this is equivalent to a reduction of 6 percentage points.

Intensive Margin We define the intensive margin as hours worked, conditional on working. Administrative data rarely informs about hours. We overcome this by using time-use data from the SOEP.[11] To study how the hours worked change after marriage, we restrict the sample to individuals who work in a given t (we take this info from the administrative data). In addition, we now restrict the sample to those already in the SOEP during the period of interest since the outcome of interest is retrieved from the SOEP data. This excludes individuals who married before they entered the survey population. We then estimate Equation 1.2 with the outcome Y_{it}^{g} being hours. Figure 1.3c plots the results: Five to ten years after marriage, women reduce their hours by 18% conditional on working. In absolute terms, this is equivalent to a daily reduction of 1.2 hours.

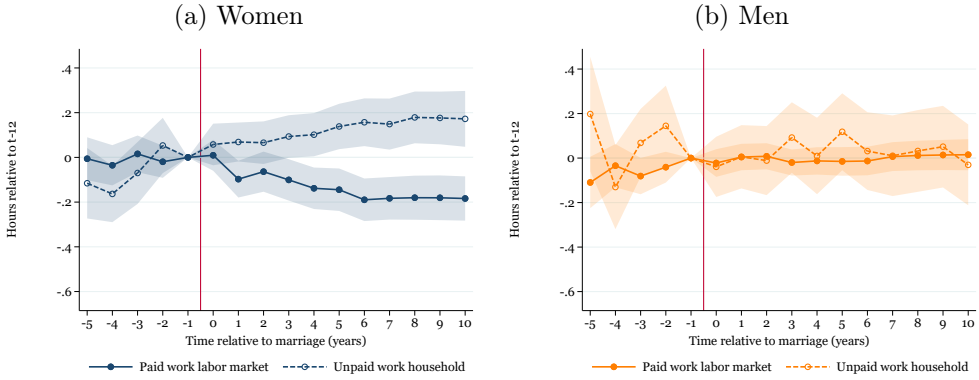


Figure 1.4: Hours Spent in the Labor Market vs. Household

Notes: Annual data from SOEPRV. This figure shows the event study estimates from estimating Equation 1.2 for time use, i.e., the daily hours spent on a given task. This is intensive margin only, conditional on working in a given t. Time-use data stems from the SOEP questionnaire. We estimate Equation 1.2 for annual event time and exclude $t = -1$, i.e., the year before the year of marriage.

Analyzing time-use data, we show that the decrease in hours spent on paid work in the labor market for women goes along with an increase in hours spent on unpaid household production (Figure 1.4a). Five to ten years post marriage, women spend .5% hours more in the household. For men, time-use does not change considerably after marriage (Figure 1.4b).

everyone has children, and some are married.

[11] Jessen (2022) validates the SOEP data on time use by comparing it against time use reported in diary data, specifically designed to measure time use.

Hourly Wages We impute wages by dividing earnings by hours worked. The measure for earnings comes from the administrative data that provides monthly gross earnings (the same measure we use for our main analysis in subsection 1.4.2). The measure for hours worked comes from survey data that provides information on daily time use (the same measure we use for the intensive margin analysis).

Figure 1.3d plots the results from estimating Equation 1.2 for hourly wages. Women's wages do not decrease post marriage. This indicates that the decrease in female earnings after marriage does not stem from decreased productivity. The results are qualitatively the same when using income information from the survey data, i.e., relying on the same data source for both hours and income (see Figure 1.A.7).

Taken together, our decomposition results draw a nuanced picture of how labor market outcomes change after marriage: The marriage earnings gap for women is driven by how much they work and not by how much they are paid. Seeing women shift their hours from paid work in the labor market to unpaid work at home is in line with predicted specialization behavior from the early literature on household specialization (e.g., Becker 1985, 1991), while finding no wage decrease for women is not. Here, our findings are in line with more recent work, which has questioned the marriage wage penalty for women (e.g., Killewald and Gough 2013; Killewald and Lundberg 2017; Pilossoph and Wee 2021).

1.4.4 Cross-Country Evidence

The marriage earnings gap is not unique to Germany. To show this, we use cross-country survey data for more than 20 European countries and Israel (SHARE JEP).[12] With the SHARE JEP data, we can estimate Equation 1.2 for the extensive margin. To validate the SHARE JEP in our context, we first compare the estimates for Germany to our estimates from the administrative German data from our main analysis. Although the data sets are very different and the sample comprises on average earlier marriages than the SOEPRV sample (see Table 1.1 and Table 1.C.1), the results are virtually the same: the extensive margin marriage gap for Germany is 10% when using the SHARE data (see Figure 1.C.1i) and 10% when using the administrative data from SOEPRV.

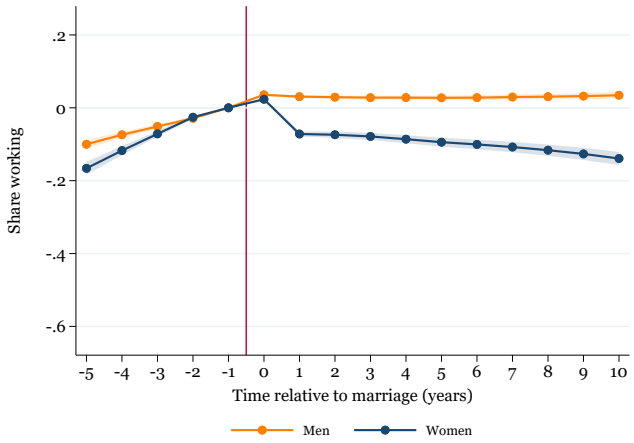
The average gap, estimated for women from all countries, is 11% (Figure 1.5a). This is a lower bound for the total marriage earnings gap, since it only considers the extensive margin. While the average marriage gap across European countries is similar to the marriage gap in Germany, we document substantial heterogeneity between countries (Figure 1.5b). The point estimates range from 33% in Ireland to no long-term gap in northeast countries such as Latvia, Finland, or Estonia. We provide the event study plots for all countries in Appendix 1.C.

Figure 1.5b reveals an East-West divide with smaller marriage gaps in the East.[13] On average, women in East-European countries experience a marriage gap of 4% as compared to 18% in the West. This pattern is in line with cultural differences with respect to gender norms in the labor market context. Shaped by communist regimes,

[12] We provide more details for the cross-country analysis in Appendix 1.C.

[13] We consider BG, CZ, EE, HR, HU, LT, LV, PL, RO, SI, and SK as Eastern and AT, BE, CH, DE, DK, EL, ES, FI, FR, IE, IL, IT, LU, NL, PT, and SE as Western.

(a) Event Study for All Countries



(b) Long-Term Gap Estimates by Country

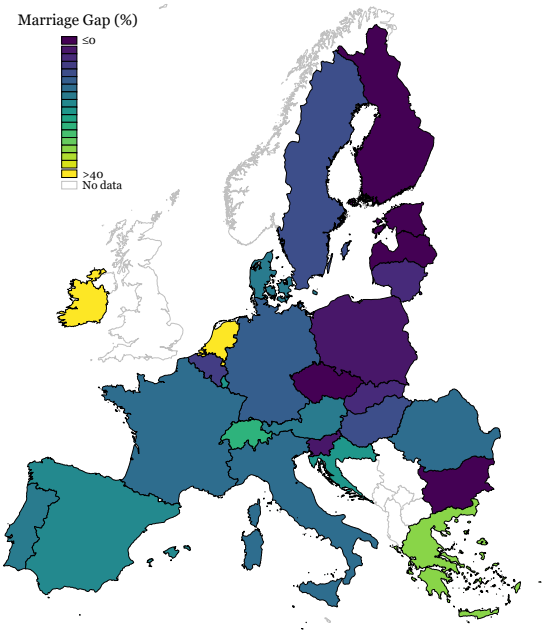


Figure 1.5: The Extensive Margin Marriage Gap Across Europe

Notes: Annual data from SHARE. This figure shows the event study estimates from estimating Equation 1.2 for the extensive margin (working yes/no) with annual event time. For the event time dummies, we exclude $t = -1$, i.e. the year before the year of marriage. Panel a plots the estimate for MG_t^g when including individuals from all countries in the SHARE data, except Germany. Panel b plots the extensive margin marriage gap estimate for women (MG_t^w) for each country.

East European countries commonly share a dual-earner norm while Western European countries often share a male-breadwinner norm. However, also within East and West, the size of the marriage gap varies between countries. One descriptive pattern is that the gap is larger in countries where marriage is associated with tax benefits (15%) as compared to countries with no benefit (8%).[14]

1.5 Mechanisms

The cross-country comparison suggests labor supply incentives from public policy and gender norms as potential mechanisms behind the marriage earnings gap. To study the effect of these mechanisms, we focus on Germany again. Using German administrative data, we explore variation in both financial incentives and norms. To determine how labor (dis-)incentives set by the institutional framework of marriage impact spousal earnings, we analyze the direct impact of joint taxation of spouses, as well as the indirect effect of alimony payments, on spousal earnings. We then study the role of gender norms by exploring quasi-exogenous variation from the German separation and reunification.

1.5.1 Direct Labor Supply Incentives from Joint Taxation

Mechanism Under joint taxation, the spousal income tax liability is a function of the total income of both partners, i and j. In Germany, spouses can choose between individual taxation (T^{single}) and joint taxation (T^{joint}). The vast majority choose joint taxation. When filing individually, the total tax liability is the sum of the individual liabilities:

$$T^{single}(y_i, y_j) = T^{single}(y_i) + T^{single}(y_j).$$

When filing jointly, spouses are taxed as if both earned exactly 50% of the combined household income:

$$T^{joint}(y_i, y_j) = 2 \times T^{single}\left((y_i + y_j)/2)\right).$$

Under the progressive German income tax schedule, there is no tax penalty for jointly filing spouses[15]:

$$T^{joint}(y_i, y_j) \leq T^{single}(y_i, y_j) \text{ for all } y_i, y_j.$$

Since the German income tax code is characterized by linearly increasing MTR for large parts of the income distribution, joint taxation typically implies a tax benefit as long as $y_i \neq y_j$. $T^{joint}(y_i, y_j) = T^{single}(y_i, y_j)$ for $y_i \neq y_j$ only if both spouses' incomes y_i

[14] We follow the definition from Christl, De Poli, and Ivaškaitė-Tamošiūnė (2023), who identify a marriage tax benefit as paying less taxes if married, as compared to cohabiting. They document a marriage tax benefit for BE, CZ, DE, GR, HR, IE, IT, LV, LU, NL, PL, SI, SK, and ES, and no benefit for AT, BG, FI, HU, LT, PT, RO, and SE. Since this includes EU countries only, we exclude non-EU countries here.

[15] There are few special cases where individual taxation is beneficial. For example, for certain types of wage-replacement benefits or loss carryforwards, individual taxation can be beneficial under certain circumstances, typically cases where spouses have similar income levels with small or no tax saving potential from income splitting.

and y_j fall in the same income tax bracket with a flat MTR. This applies only to very low-income couples, with both y_i and y_j below the basic allowance threshold (MTR = 0), and very high income couples with both y_i and y_j in either of the two top tax brackets (MTR = 0.42 and 0.45, respectively).

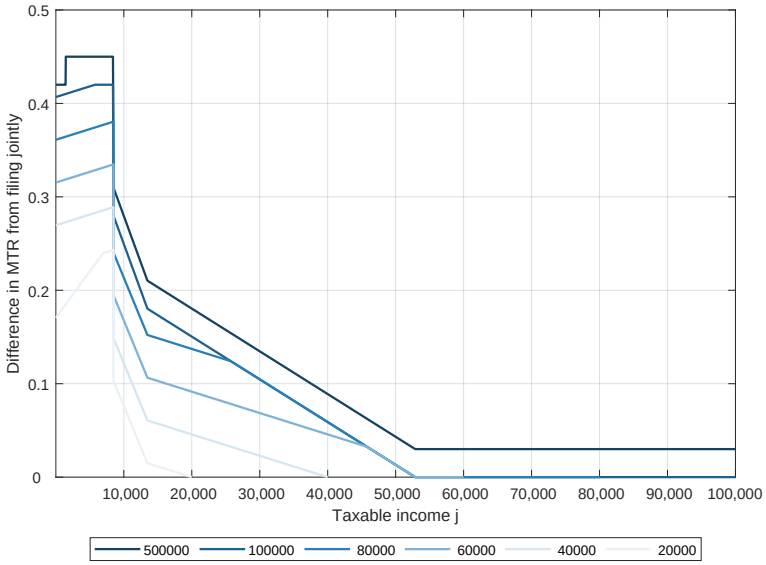


Figure 1.6: MTR Changes for Secondary Earners

Notes: This figure illustrates the change in MTR under joint taxation (ΔMTR_j) over the income of the secondary earner j. ΔMTR_j depends not only on individual income y_j but also on spousal income y_i. For illustration, we plot ΔMTR_j for five different income levels of the primary earner. Secondary earners with the same income y_j have the same x-coordinate; secondary earners married to spouses with the same income y_i are on the same line. Figure 1.A.9 plots the resulting income responses for $\varepsilon_{y_j,1-MTR_j} = 0.3$.

In terms of MTR, both spouses face the same tax rate when filing jointly. This typically decreases the MTR for the primary earner (i) while increasing the MTR for the secondary earner (j):

$$MTR^{single}(y_j) \leq MTR^{joint}(y_i, y_j) \leq MTR^{single}(y_i) \tag{1.4}$$

The MTR for secondary earner j remains unchanged only if y_j and $(y_j + y_i)/2$ fall in the same bracket with a flat MTR (true for very low and very high income only) or if both spouses have exactly the same income. For all other cases, j faces an increase in MTR: $\Delta MTR_j = MTR^{joint}(y_i, y_j) - MTR^{single}(y_j) > 0$, implying a disincentive for their individual labor supply. As depicted in Figure 1.6, the size of the disincentive depends on both spouses' income. In general, ΔMTR_j increases with the relative income difference among spouses, effectively incentivizing spousal specialization. The residual heterogeneity in ΔMTR_j can be attributed to one notch and two kinks between the five different income tax brackets of the German income schedule.

The behavioral response, i.e., the extent to which j reduces their income due to

ΔMTR_j, depends on the elasticity of taxable income with respect to the net-of-tax rate $\varepsilon_{y_j,1-MTR_j}$: $\Delta y_j = -\varepsilon_{y_j,1-MTR_j} \cdot y_j \cdot \Delta MTR_j/(1 - MTR_j)$. For illustration, Figure 1.A.9 plots the simulated Δy_j, assuming an elasticity of $\varepsilon_{y_j,1-MTR_j} = 0.3$. We estimate $\varepsilon_{y_j,1-MTR_j}$ for our setting in the next step.

Testing the Mechanism To analyze the impact of the direct labor disincentives set by joint taxation, we exploit the heterogeneity in ΔMTR_j depicted in Figure 1.6 to estimate the income elasticity for married women and men. We face two challenges for identification here.

The first challenge is that the change in MTR and the change in income are observed simultaneously, with the latter impacting the former and thus introducing a bias in the elasticity estimate. This is a standard challenge when estimating elasticities based on changes in the tax rate (e.g., after a tax reform). The standard solution to this is to instrument for the observed change in MTR by using a hypothetical mechanical tax rate change as the instrument (for details and a literature overview, see Neisser 2021). Following Gruber and Saez (2002), the "most standard approach" (Neisser 2021, p. 3373) for constructing the instrument is to use the income from the period before the change in MTR. When analyzing a tax reform, this is typically the pre-reform period. In our setting, we take the pre-marriage income in $t = -1$ and simulate the change in MTR spouses would have faced had they already been filing jointly in $t = -1$ ($MTR^{hyp}_{i,-1}$). Put differently, this is the MTR spouses would have faced in $t > -1$ if their income had not changed since $t = -1$. For this, we link spouses in the year before marriage. We describe how we do this in the data section below.

Two other challenges are specific to our setting: First, joint taxation itself is not exogenous. Spouses might actively decide to get married because of the lower tax liability they would face when filing jointly. Second, how much the net-of-tax-rate changes under joint taxation is also endogenous as it depends on the individual income and within-household share of income. Income and income share are likely to be correlated with unobservables that may impact spousal labor supply, such as gender norms or career preferences. We address both challenges by including fixed effects for bins of individual income and within-household income share.[16] The identifying assumption here is that while individual income and within-household income distribution can be endogenous, the exact income and share of income are quasi exogenous. To provide an illustrative example, this means an individual may select into a marriage in which they contribute about two-thirds of the household income. However, they cannot precisely anticipate whether their share of income will be 0.64 or 0.66 in $t = -1$ (and similarly for their absolute income). Here, we exploit the substantial heterogeneity in ΔMTR_j for similar income levels and shares (Figure 1.6). In addition, having the full taxpayer population provides us with a large enough sample.

For the 2SLS estimation, we estimate the following equation separately for women and

[16] Binning both is required because the exact income and income share perfectly define the MTR under joint taxation and would remove all variation.

men ($g =$ w, m) in a given event year t:

$$ln\left(\frac{y_{it}^{g}}{y_{i,-1}^{g}}\right) = \varepsilon_{t}^{g} ln\left(\frac{1-\widehat{MTR}_{i,t}}{1-MTR_{i,-1}}\right) + \sum_{t^{child1}} \gamma_{t^{child1}}^{g} + \theta_{income}^{g} + \rho_{share}^{w} + \delta_{age}^{w} + \zeta_{age}^{m} + \eta_{cal}^{g} + \epsilon_{it}^{g}, \tag{1.5}$$

with the first stage defined as:

$$ln\left(\frac{1-MTR_{i,t}}{1-MTR_{i,-1}}\right) = \lambda_{t}^{g} ln\left(\frac{1-MTR_{i,-1}^{hyp}}{1-MTR_{i,-1}}\right) + \sum_{t^{child1}} \gamma_{t^{child1}}^{g} + \theta_{income}^{g} + \rho_{share}^{w} + \delta_{age}^{w} + \zeta_{age}^{m} + \eta_{cal}^{g} + \epsilon_{it}^{g}. \tag{1.6}$$

We denote the gross labor income of individual i in event year t as y_{it}^{g}, with $y_{i,-1}^{g}$ referring to the pre-marriage income in $t = -1$.[17] Our parameter of interest is ε_{t}^{g}, the elasticity of labor income with respect to the net-of-tax rate for gender g. The net-of-tax rate for i is denoted as $1 - MTR_{i,t}$ for period t, and as $1 - MTR_{i,-1}$ for the pre-marriage base year. $1 - MTR_{i,-1}^{hyp}$ is the hypothetical net-of-tax rate i would have faced under joint taxation in $t = -1$. This is our instrument for $1 - MTR_{i,t}$.

We include fixed effects for women's labor income binned in 1,000-€-bins (θ_{income}^{g}) and for women's income share binned in 2-%-bins (ρ_{share}^{w}), both measured pre-marriage in $t = -1$ (we also replicate our results with other bin sizes). In addition, δ_{age}^{w} and ζ_{age}^{m} are fixed effects for the age of women and men in year t, respectively, and account for gender-specific life cycle trends. Calendar year fixed effects η_{cal} account for common time trends.

Lastly, we include event time dummies for time relative to the birth of the first child (t^{child1}) to account for the child penalty. Taken together, we combine the standard approach to empirically recover elasticity estimates from changes in MTR (e.g., Gruber and Saez 2002; Neisser 2021; Saez, Slemrod, and Giertz 2012) with our approach for estimating the marriage earnings gap.

Data To estimate 1.5 and 1.6, we use a unique sample of linked spouses from administrative income tax data (TPP). The sample includes the universe of all taxpayers who got married between 2011 and 2016 and file their taxes jointly. The pivotal advantage of the data is that it yields precise information on spouses' incomes and tax liabilities already before marriage, allowing us to determine the exact MTR for singles and married spouses. Our outcome is gross labor earnings, which we define as the sum of gross wages, self-employed income, and business income. We restrict the sample to working age couples (age 18 to 60) for whom both spouses are observed at least three years before and after the year of marriage. We further drop couples if at least one spouse

[17] The standard approach is to estimate the elasticity of taxable income with respect to the marginal tax rate (e.g., Gruber and Saez 2002; Saez, Slemrod, and Giertz 2012). Our setting does not allow us to use taxable income since, under joint taxation, taxable income is only defined at the household level. To be able to estimate how individual income changes, we use gross labor income, which is observed on the individual level, both for singles and married individuals.

has zero taxable income in $t = -1$ or negative income at any t. Lastly, we require all individuals to have a stable income in the three years preceding their marriage. We drop couples if at least one spouse's income in any $t = -3, -2, -1$ deviates more than 25% from their three-year average income over that period. Having relatively constant income pre-marriage is likely to increase the salience of MTR changes under joint taxation and also mitigates concerns about mean reversion for the elasticity estimation. Table 1.2 provides descriptive statistics of our sample, and Figure 1.A.10 visualizes the distribution of income, MTR, and MTR changes from joint taxation for the sample.

Results Table 1.4 reports the 2SLS elasticity estimates from estimating Equation 1.5 for women (for the first stage results see Table 1.A.2a). Our preferred specification includes fixed effects for absolute female income (θ^w_{income}) and female share of income (ρ^w_{share}) pre-marriage (column 4).

The change in MTR under joint taxation has a significant effect on married women's labor income for all $t > 0$. The higher the increase in MTR, the more women reduce their earnings, which leads to a larger marriage earnings gap. The elasticity estimates range from 0.3 to 1.1 and increase over t. For $t = 0$, we do not find a significant effect in our preferred specification, which is likely explained by measurement error: The income tax data is annual and therefore does not inform about when exactly an individual gets married in $t = 0$. Income in $t = 0$ is thus only partly taxed under joint taxation. For men, the elasticity estimates are statistically insignificant and also economically small. Table 1.A.2 provides the 2SLS results for men.

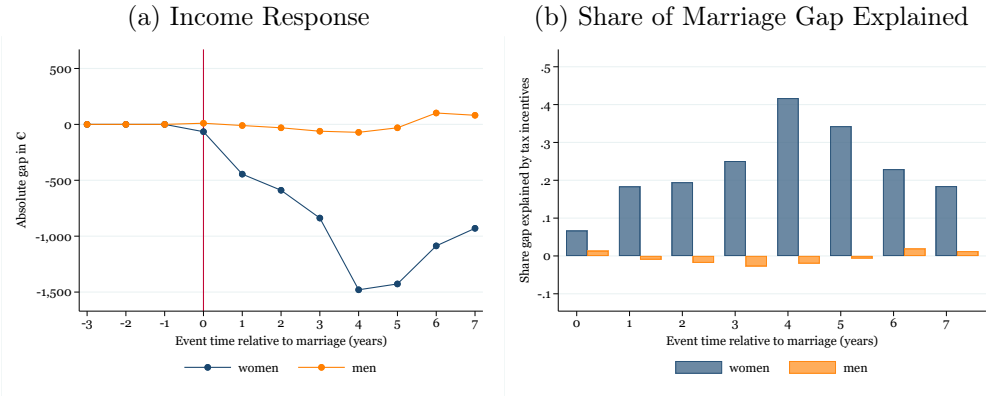


Figure 1.7: Pure tax Effect on Earnings

Notes: Annual gross labor income from TPP in 2015 Euros. Panel a shows the average earnings response to tax changes from joint taxation: $\Delta y^g_t = -\varepsilon^g_t \cdot y^g_{-1} \cdot \Delta MTR^g_{-1}/(1 - MTR^g_{-1})$ computed for every t with parameters from Table 1.4, Table 1.2, and Table 1.A.2. We include results for men, although their elasticity estimates are not statistically significant (Table 1.A.2). Panel b shows the share of the total marriage earnings gap that is explained by the response to tax incentives. For this, we divide the earnings response (panel a) by the absolute marriage earnings gap estimated for this sample (see Figure 1.A.11a). *Reading example*: In $t = 3$ women's income response to tax changes from joint taxation is -838 €, which explains 25% of the marriage earnings gap in that period.

To quantify the impact of tax incentives for the marriage earnings gap, we provide a back-of-the-envelop calculation for the size of the earnings response to changes in MTR. We compute $\Delta y^g_t = -\varepsilon^g_t \cdot y^g_{-1} \cdot \Delta MTR^g_{-1}/(1 - MTR^g_{-1})$ for $g = women, men$, replacing

	(1) baseline	(2) income	(3) share	(4) both
t=0	0.18***	0.11***	-0.11*	0.05
	(0.02)	(0.02)	(0.04)	(0.05)
t=1	0.42***	0.49***	0.32***	0.34***
	(0.03)	(0.04)	(0.08)	(0.09)
t=2	0.74***	1.07***	0.72***	0.45***
	(0.04)	(0.05)	(0.10)	(0.11)
t=3	0.94***	1.55***	0.94***	0.64***
	(0.04)	(0.05)	(0.11)	(0.12)
t=4	0.98***	1.76***	1.43***	1.13***
	(0.05)	(0.06)	(0.14)	(0.15)
t=5	1.01***	1.95***	1.37***	1.09***
	(0.07)	(0.08)	(0.17)	(0.19)
t=6	0.98***	1.94***	1.05***	0.83***
	(0.09)	(0.11)	(0.23)	(0.25)
t=7	0.95***	1.73***	0.87**	0.71*
	(0.13)	(0.15)	(0.33)	(0.35)
Income (θ^w_{income})		✓		✓
Income share (ρ^w_{share})			✓	✓
Event time first child	✓	✓	✓	✓
Age (δ^w_{age}, ζ^m_{age})	✓	✓	✓	✓
Year (η_{cal})	✓	✓	✓	✓

Notes: Estimates for the elasticity of gross income with respect to the net-of-tax rate, from our 2SLS estimation of Equation 1.5 for women. Table 1.A.2a reports first-stage results from estimating Equation 1.6, while Table 1.A.2 reports the second-stage estimates for men. Each column shows the estimates from a different specification, with differences in the included fixed effects. Column (2) adds fixed effects for female income to the baseline model, binned in 1,000-€-bins, and (3) for the female share of income , using bins of 2 percentage points. Specification (4) is our preferred specification and includes fixed effects for both female income and female share of income. Each row shows the estimates for a different event time, with $t = 0$ denoting the year of marriage.
Significance level: *** 0.01; ** 0.05; * 0.1. Robust standard errors are reported in parentheses.

Table 1.4: Joint Taxation – Elasticity Estimates for Women (2SLS)

all parameters on the right hand side with estimates from the data. We take the mean values for $t = -1$ as reported in Table 1.2 and the elasticity estimates from column 4 of Table 1.4. Figure 1.7 plots the absolute annual gross labor income response over event time. Averaging over $t > 0$, women reduce their annual gross labor income by -971 € because of higher MTR from joint taxation. This implies that tax incentives account for 26% of the total marriage earnings gap in this sample (Figure 1.7b). Our empirical finding is in line with results from structural estimations (Bick and Fuchs-Schündeln 2018; Borella, De Nardi, and Yang 2023), which have highlighted the role of tax incentives for spousal labor supply.

Tax incentives from joint taxation significantly contribute to the marriage earnings gap, but cannot fully explain it. In line with this result, we show that even primary earner women whose MTR decrease under joint taxation face a marriage earnings gap. At the same time, men do not decrease their earnings after marriage, even if they face a higher MTR under joint taxation. This is shown in Figure 1.A.12, which plots the corresponding event study estimates for the two subsamples of spouses with primary and secondary earner women, respectively. Taken together, our results suggests that the responsiveness to tax incentives in the marriage context depends on gender (norms) – a mechanism we exploit in more detail in subsection 1.5.3.

1.5.2 Indirect Labor Supply Incentives from Ex-Spousal Alimony

Mechanism With ex-spousal alimony, a reduction in earnings during marriage has a smaller impact on the expected lifetime income of secondary earners. Marriage then acts as an insurance mechanism against the financial risk of reducing earnings. The risk of reducing earnings during marriage and its subsequent consequence of lower post-divorce earnings is mitigated by the future income support provided by the other spouse. In that way, alimony payments are an indirect labor disincentive only coming into effect in the case of divorce.

Testing the Mechanism To test whether spousal earnings respond to this indirect labor supply incentive, we exploit exogenous variation in alimony eligibility. In 2008, the German government introduced a major divorce law reform, cutting the legal entitlement for ex-spousal alimony payments and, thus, financial obligations post-divorce. The reform marked a shift in how court decisions on alimony were made, increasing conditions to meet alimony requirements while shortening payment periods. Besides changing the general understanding of post divorce obligations, the reform also changed specific rules for alimony eligibility. Before the reform, divorced resident parents are eligible for alimony payments until their youngest child turns 15. Post reform, eligibility applies as long as the youngest child is three years old at most. From then on, the resident parent is expected to provide for their living expenses by working full time. The change in the alimony payments is depicted in Figure 1.A.13.

The reform, announced in 2006 (link first draft), enacted in December 2007, and implemented from January 2008, impacted all marriages: future marriages contracted after reform, existing marriages contracted pre reform but not (yet) divorced, and past marriages contracted and divorced pre reform.[18]

[18] For already divorced marriages, alimony payments are not adjusted automatically post reform but have to be renegotiated.

The reform introduces exogenous variation in the insurance value of marriages with heterogeneity in the timing of this shock. Determined by the year of marriage, spouses have more or less (or no) leeway to adjust their earnings when learning about the reduced insurance value. We exploit this natural experiment to analyze the causal impact of alimony on spousal earnings during marriage, comparing the marriage earnings gap across different year-of-marriage cohorts. We first estimate our main event study specification from Equation 1.2 for each cohort separately to compare how the marriage earnings gap evolves around reform.

After providing graphical evidence, we then estimate a modified version of Equation 1.2 for all cohorts together, adding a dummy indicating whether an individual is treated by the reform. We allow the alimony regulation to affect individuals in two ways. First, we analyze the effect of the alimony regime in place at the time of marriage ($t = 0$). Treatment is a function of the marriage cohort: individuals are treated by the reform if they get married post-reform. To estimate this effect, we extent Equation 1.2 by interacting the event time dummies with a dummy treated marriage cohort$_i$, which is 1 for individuals who get married ≥ 2008, and 0 else:

$$\begin{aligned} Y_{it}^{g} = &\sum_{j \neq r} \beta_j^g \cdot \mathbb{1}[j = t] + \sum_{j \neq r} \delta_j^g \cdot \text{treated marriage cohort}_i \times \mathbb{1}[j = t] \\ &+ \sum_k \gamma_k^{1,g} \cdot \mathbb{1}[k = t^{child1}] + \zeta_{cal}^g + \eta_{age}^g + \theta_c^g + \upsilon_{it}^g. \end{aligned} \tag{1.7}$$

Second, we analyze the effect of the status quo alimony regime in place during a given period on the labor supply in that same period. Treatment is then a function of calendar time: individuals are treated in the calendar years ≥ 2008. To estimate this effect, we extend Equation 1.2 by interacting the event time dummies with a dummy treated calendar year$_t$, which is 1 if the individual is observed in calendar years ≥ 2008, and 0 else:

$$\begin{aligned} Y_{it}^{g} = &\sum_{j \neq -1} \beta_j^g \cdot \mathbb{1}[j = t] + \sum_{j \neq -1} \delta_j^g \cdot \text{treated calendar year}_t \times \mathbb{1}[j = t] \\ &+ \sum_k \gamma_k^{1,g} \cdot \mathbb{1}[k = t^{child1}] + \zeta_{cal}^g + \eta_{age}^g + \theta_c^g + \upsilon_{it}^g. \end{aligned} \tag{1.8}$$

The coefficient of interest for Equation 1.7 and Equation 1.8 is δ_j^g, which can be interpreted as a difference-in-marriage-gap estimate. δ_j^g captures the difference between event time estimates at event time $t = j$ of individuals treated by the reform versus those who are not, following the respective definition: *married post reform* for Equation 1.7 and *observed post reform* for Equation 1.8. Event time t is measured in years relative to the event of marriage, omitting the reference period $r = -1$, i.e., the year prior to the year of marriage. Because of a limited number of available variables, we only account for event time with respect to the first child. We denote η_{age}^g, ζ_{cal}^g, and θ_c^g as fixed effects for individuals' age, calendar year, and year-of-marriage cohort, respectively.[19]

[19] Note that we include treated marriage cohort$_i$ and treated calendar year$_t$ only as interaction terms since they are perfectly collinear with θ_c^g and ζ_{cal}^g, respectively.

Data As in subsection 1.5.1, we use administrative income tax data (TPP) for the full population of German taxpayers. Since we want to analyze how the marriage earnings gap changes around the 2008 reform, we restrict the sample to newlyweds who married between 2004 and 2012, i.e., +/- 4 years around the reform. We further restrict the sample to individuals aged 18 or older, who are younger than 40 when they get married. The sample contains individuals who got a divorce, but we drop their post-divorce periods. We proceed the same way with individuals who became widowed. We further restrict the sample to individuals we observe at least three years before or after marriage.[20] Table 1.3 illustrates the sample descriptives.

In the TPP, women are typically added to their husbands' tax records upon joint taxation. Since German tax authorities introduced individual identifiers only in 2010, the TPP data does not allow for a linkage of women across pre-marriage and marriage years for the years around the reform. We overcome this hurdle by defining single filing women aged 18 to 39 who drop out of the panel data entirely as an indication of marriage. We then define event time relative to the drop-out, with the last observed year defining $t = -1$. Hence, we have a panel of men who are linked over time, and women for whom we observe pre- and post-marriage periods but cannot link both. Note that our event study specification (see Equation 1.2) does not require an individual panel identifier, allowing us to estimate the marriage earnings gap even when we cannot link individuals across event time. Kleven (2023) shows that child penalties can be estimated on repeated cross-sectional data. In our case, the data is not a repeated cross-section but actual panel data that lacks uninterrupted individual identifiers for women.

We provide descriptive statistics for the sample of women who drop out of the data (what we define as $t = -1$) and the sample of women who appear in $t = 0$ on their husbands' tax records. Table 1.A.3 shows that both samples have similar characteristics with respect to observable characteristics. The samples are not identical for two reasons. First, we do not observe all women in $t = -1$, because not all women pay income taxes in the year before marriage and/or file an income tax return.[21] Second, dropouts are not necessarily an indicator of marriage: Women may stop paying taxes (or stop filing).

Results Figure 1.A.14 compares MG_t^w estimates for women across different year-of-marriage cohorts around the reform in 2008. Comparing the MG_t^w cohort-by-cohort allows us to differentiate between two different effects. First, we show that married women's earnings are not affected by the alimony regime in place at the time of marriage, at $t = 0$: There is no difference between the earnings trajectories of women married before the reform (2006 and 2007) and those who married after the reform (2008 and 2009).

Second, we show that women's earnings do not respond to a change in the alimony regime at later points of their marriage, at $t > 0$. Figure 1.A.15 provides pairwise comparisons for all marriage cohorts from 2004 to 2012. Panels 1.A.15a to 1.A.15c compare pre-reform marriage cohorts. If already married women increased their labor

[20] Ideally, we would want to restrict the sample to individuals we can observe three years before *and* after marriage. However, the data structure only allows for linking periods $t < 0$ and $t >= 0$ for men, which is why we implement the weaker restriction of observing individuals at least three years before *or* after marriage. We explain the data structure in more detail in the following paragraph.

[21] Before 2012, non-filers are not included in the TPP data. For most German taxpayers, filing is optional, and many do not choose to file (Hauck and Wallossek 2023).

supply as a response to the reform, we would expect to see a change in MG_t^w for post reform t, i.e., for $t \geq$ the respective vertical line. For these cohorts, we can also rule that selection into marriage is affected by the reform, simply because they married before the reform is enacted (and even before it is announced for cohorts 2004-2006).

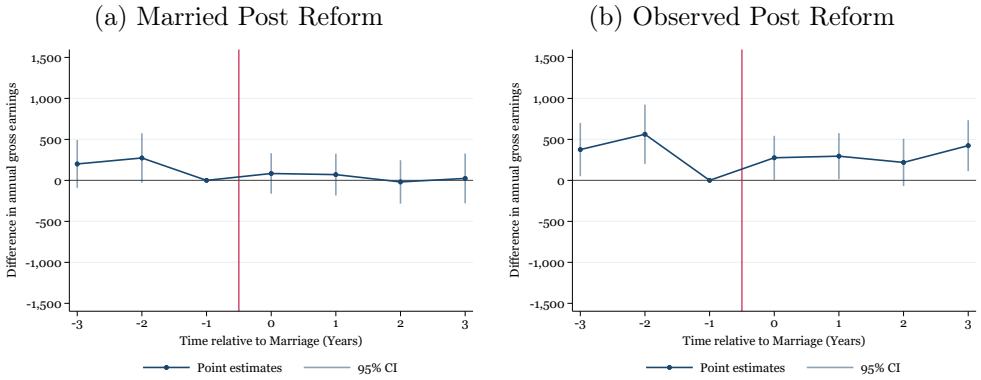


Figure 1.8: Difference Estimates

Notes: Annual data from TPP. This figure plots the difference estimates for women (δ_y^w) from Equation 1.7 (panel a) and Equation 1.8 (panel b). Panel a shows how annual gross labor earnings change at given event time t for women who married post reform, compared to those who married pre reform. Panel b plots the difference for women who, in given period t, are observed in a post-reform calendar year, compared to those observed pre reform.

The regression results from estimating Equation 1.7 confirm the graphical evidence from Figure 1.A.15. The coefficients for the interaction terms δ_y^w are not statistically different from 0 when comparing women who married post- vs. pre-reform (Figure 1.8a), suggesting that women do not adjust their earnings in response to the reform. We also find no clear effect when comparing women who are observed post- vs. pre-reform (Equation 1.8, results are plotted in Figure 1.8b). Our findings are in line with results from Bredtmann and Vonnahme (2019), who also do not find an effect of the alimony reform on spousal labor supply, time-use, or divorce probability using survey data from SOEP.

1.5.3 Gender Norms

Mechanism After studying the role of institutions, we now move to another potential mechanism: gender norms. Norms have shown to be an important driver behind gender gaps in different domains (e.g., Bertrand 2011). They may also explain (part of) the marriage earnings gap when impacting spousal labor supply. We want to test whether the marriage earnings gap is larger under more traditional gender norms.

Testing the Mechanism We first document the correlation between the marriage earnings gap and gender norms. For this descriptive analysis, we use responses to value-based questions from the SOEP to split our sample into individuals with progressive vs. traditional gender norms. We then estimate Equation 1.2 for each subsample. The obvious caveat of this analysis is that norms are likely to be formed endogenously.

To address this, we apply a second analysis where we exploit a natural experiment

with quasi-exogenous variation in gender norms: the division of East and West Germany after World War II. From 1945 until 1990, two German countries existed: the German Democratic Republic (GDR, East Germany) and the Federal Republic of Germany (FRG, West Germany). Under the communist regime in the East, female labor market participation was high, going along with the norm of dual-earner spouses. In the capitalist West, women were less likely to work and the country had a strong male breadwinner norm.[22] East-West differences have been shown to persist even long after reunification. This is true for many dimensions, including preferences and norms (Alesina and Fuchs-Schündeln 2007; Boelmann, Raute, and Schönberg 2021).

To disentangle differences in culture from differences in legal institutions, we compare individuals who were born when Germany was divided but got married after reunification. They grew up in different countries with different norms and cultures, but at the time of marriage, they faced identical legal institutions with identical financial incentives, as discussed in subsection 1.5.1 and 1.5.2. We estimate a modified version of Equation 1.2,:

$$\begin{aligned} Y_{it}^g = &\sum_{j \neq r} \beta_j^g \cdot \mathbb{1}[j = t] + \sum_{j \neq r} \delta_j^g \cdot \mathbb{1}[j = t] \times east_i \\ &+ \sum_{C=1}^{3} \sum_k \gamma_k^{C,g} \cdot \mathbb{1}[k = t^{childC}] + \sum_{C=1}^{3} \sum_k \lambda_k^{C,g} \cdot \mathbb{1}[k = t^{childC}] \times east_i \\ &+ \kappa_{i,-1}^g + \phi_{i,-1}^w + \zeta_{cal}^g + \eta_{age}^g + \epsilon_{it}^g. \end{aligned} \tag{1.9}$$

The main difference to Equation 1.2 is that we interact event time dummies with a dummy variable $east_i$ equal to 1 if i was born in East Germany and 0 if i was born in West Germany. We add the interaction for all events, i.e., marriage as well as the birth of the first, second, and third child. This allows us to account for differences in the child penalties between East and West (Boelmann, Raute, and Schönberg 2021; Jessen 2022). In addition, we add fixed effects for pre-marriage income and female income share, measured in the year before marriage ($t = -1$), and denoted by $\kappa_{i,-1}^g$ and $\phi_{i,-1}^w$, respectively.[23] This accounts for underlying differences in income and gender gaps between East and West.

Data We use the same SOEPRV data we use for our main analysis in section 1.4. We measure individual gender norms with a set of value-based questions from the SOEP. These questions contain statements on gender roles that respondents are asked to agree or disagree with, based on a 1-7 Likert scale (see section 1.D for the questions used). We classify an individual as having traditional gender norms if their average response to these questions is < 4, and as progressive if their average is > 4.

For our East/West analysis, we impose two additional sample restrictions: First, we restrict the sample to individuals who were born in Germany < 1989 and get married

[22] For details on differences in gender norms and culture in the two countries, see e.g., Boelmann, Raute, and Schönberg (2021).

[23] For absolute earnings ($\kappa_{i,-1}^g$), we bin gross earnings in $t = -1$ in decentiles of the income distribution for each gender g: $x = 0$, $0 < x \leq p10$, $p10 < x \leq p20$, ..., $p90 < x \leq p100$. For the female earnings share ($\phi_{i,-1}^w$), we include these bins: $x = 0$, $0 < x \leq 0.1$, $0.1 < x \leq 0.2$, $0.2 < x \leq 0.3$, $0.3 < x \leq 0.45$, $0.45 < x \leq 0.55$, $0.55 < x \leq 0.75$, $0.75 < x < 1$, and $x = 1$.

$>$ 1990. Since the place of birth is not reported in the administrative data, we rely on information from survey data here: SOEP respondents are asked about their place of residence in 1989. We take this as a proxy for the place of birth. We exclude individuals who were born abroad. For individuals who did not answer the survey question on their place of residence in 1989, we proxy for the place of birth by using the first reported spell in the administrative data, an approach that has been established in the context of German administrative data (see e.g., Boelmann, Raute, and Schönberg 2021; Findeisen, Lee, Porzio, and Dauth 2021). Second, we include only individuals for whom we observe their spouses in the SOEPRV data. This restriction is necessary because we include fixed effects for the within-couple income share in $t = -1$.

Results Splitting the sample into individuals with traditional vs. progressive gender norms, we find a clear correlation between norms and post-marriage earnings. Figure 1.A.16 plots the results from estimating Equation 1.2. The marriage earnings gap for individuals with progressive gender norms is 10% (Figure 1.A.16a), while the gap is 33% for those with traditional gender norms (Figure 1.A.16b).

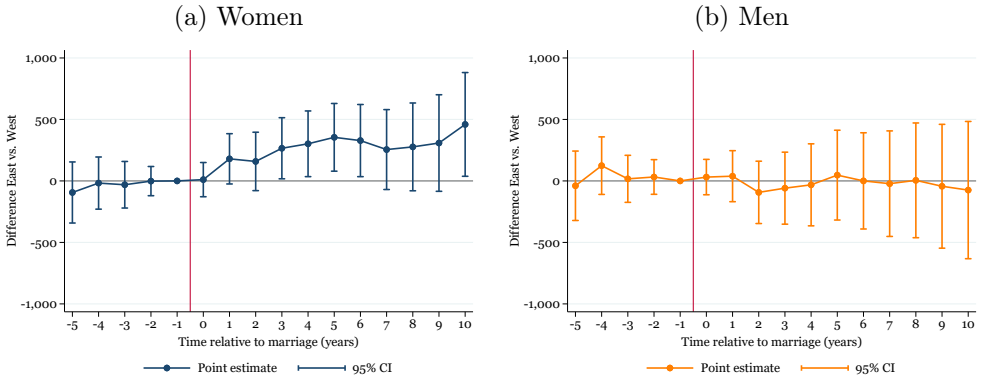


Figure 1.9: The Marriage Earnings Gap – East vs West

Notes: Annual data from SOEP–RV. This figure plots the $\hat{\delta}_j^g$ estimates from Equation 1.9 over event time. This is the difference in earnings for East vs West that is attributed to the given event time $t = j$. Panel a plots the results for women ($\hat{\delta}_j^w$) while panel b plots the results for men ($\hat{\delta}_j^m$). Estimates account for the child penalty and allow child penalty estimates to differ between East an West.

These descriptive findings are supported by the results from leveraging quasi-exogenous variation in gender norms from the German separation. Figure 1.9 plots the $\hat{\delta}_j^g$ estimates from Equation 1.9, i.e., the difference in earnings between East and West that is specific to event time $t = j$ and adds to the overall East-West gap. For men, the earnings difference between those born in the East vs West does not change with event time (Figure 1.9b). For women, the East-West gap is stable for pre-marriage periods as well (Figure 1.9a). However, for $t > 0$, earnings for women born in the East are relatively higher. Put differently, women born in the West, with more traditional gender norms, reduce their earnings more after marriage than do women born in the East. Figure 1.A.17 plots the marriage earnings gap estimates for East and West. Women born in East Germany face no significant drop in earnings after marriage, while women born in the West face a marriage earnings gap of 30%. These results strongly support

the relevance of gender norms for the marriage earnings gap.

1.6 Conclusion

In this paper, we show that there is a marriage earnings gap: Women's earnings drop after marriage, and this is not explained by the child penalty. The marriage earnings gap results from the extensive and intensive margin of labor supply: Married women are less likely to work and, conditional on working, work fewer hours, but they are not paid lower wages. Different from children, marriage is a social construct. In modern societies, it is also a legal contract with rules defined by the government. Policy makers can use marriage as a public policy tool to (dis-)incentivize spousal labor supply. We find that labor supply incentives significantly contribute to the marriage earnings gap: A back-of-the envelop calculation suggests that increased marginal tax rates under joint taxation explain 26% of the gap for women. Given that child-related policies have been shown to have only very limited impact on the child penalty (Kleven, Landais, Posch, Steinhauer, and Zweimüller 2023), this is an important insight for policy makers. However, our findings also highlight that not all labor supply incentives from marriage related policies are equally effective. We do not find an effect on earnings from a divorce law reform that aimed at increasing married women's labor supply by reducing their alimony eligibility. Our results also emphasize the relevance of gender norms: Women who grew up in East Germany with more progressive gender norms experience a smaller marriage earnings gap.

Appendix to Chapter 1

1.A Additional Figures and Tables

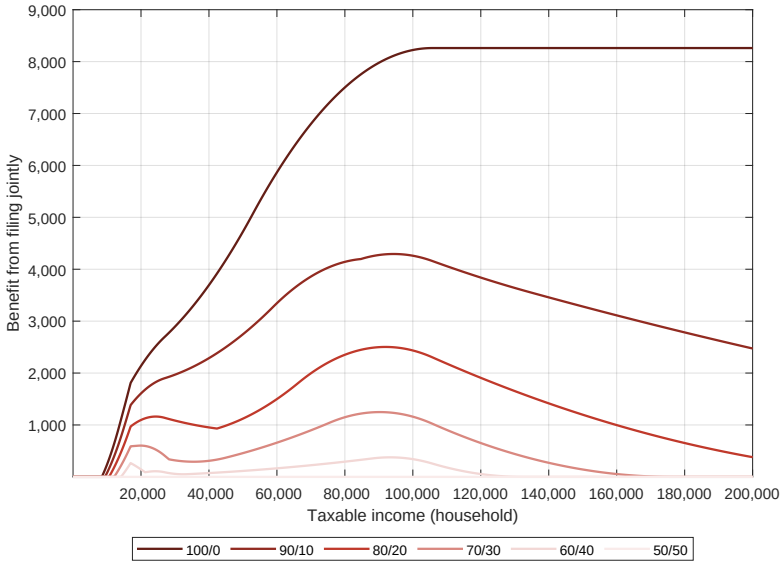
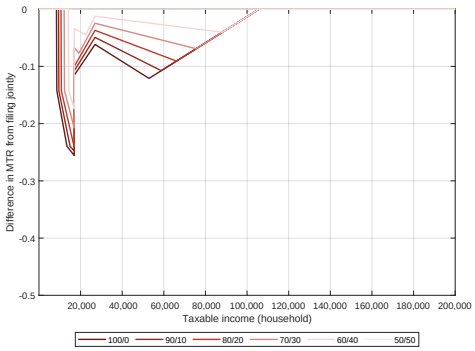


Figure 1.A.1: Monetary Benefits from Joint Taxation

Notes: This figure shows the annual monetary benefit from joint taxation (as compared to individual taxation) for the tax year 2015 as a function of the total taxable income within the household, i.e. the sum of both spouses' individual taxable income. We simulate the benefits for 5 different within-household income shares. 100/0 refers to the extreme case of a single-earner couple, with one spouse earning 100% of the joint income and the other spouse earning 0. The cases 90/10, 80/20, 70/30, and 60/40 refer to cases where both spouses have positive, but different income, with the primary earner earning 90%, 80%, 70%, and 60%, respectively. 50/50 refers to the case of spouses earning exactly the same. In this case, joint taxation has zero monetary benefits, since both spouses face the same ATR when filing individually. All values are in Euro. *Reading example*: A single-earner couple (1:0 case) with an annual income of 40,000 € gains about 3,700 € from filing jointly every year.

(a) Primary Earner



(b) Secondary Earner

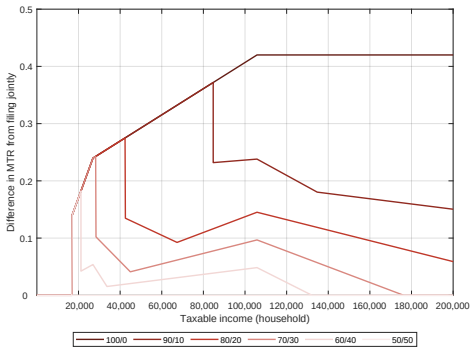


Figure 1.A.2: Labor Supply (Dis-)Incentives Under Joint Taxation

Notes: This figure shows the absolute change in the marginal tax rate (MTR) under joint taxation (as compared to individual taxation) for the tax year 2015 as a function of the total taxable income within the household, i.e. the sum of both spouses' individual taxable income. We simulate the benefits for 5 different within-household income shares. 100/0 refers to the extreme case of a single-earner couple, with one spouse earning 100% of the joint income and the other spouse earning 0. The cases 90/10, 80/20, 70/30, and 60/40 refer to cases where both spouses have positive, but different income, with the primary earner earning 90%, 80%, 70%, and 60%, respectively. 50/50 refers to the case of spouses earning exactly the same. Panel (a) shows that joint taxation reduces the MTR for primary earners i in many cases, increasing their incentive for labor supply. To be precise, joint filing reduces primary earners' MTR whenever $MTR(y_i) > MTR((y_i + y_j)/2)$. Panel (b) shows that joint taxation increases the MTR for secondary earners j in many cases, reducing their incentive for labor supply. To be precise, joint filing increases secondary earners' MTR whenever $MTR(y_j) < MTR((y_i + y_j)/2)$. In the 50/50 case of spouses earning exactly the same ($y_i = y_j$), joint taxation does not impact their MTR. *Reading example*: In a single-earner couple (1:0 case) with an annual income of 40,000 €, joint taxation reduces the MTR for the primary earner by 8 percentage points. For the secondary earner, the MTR increases by 27 percentage points.

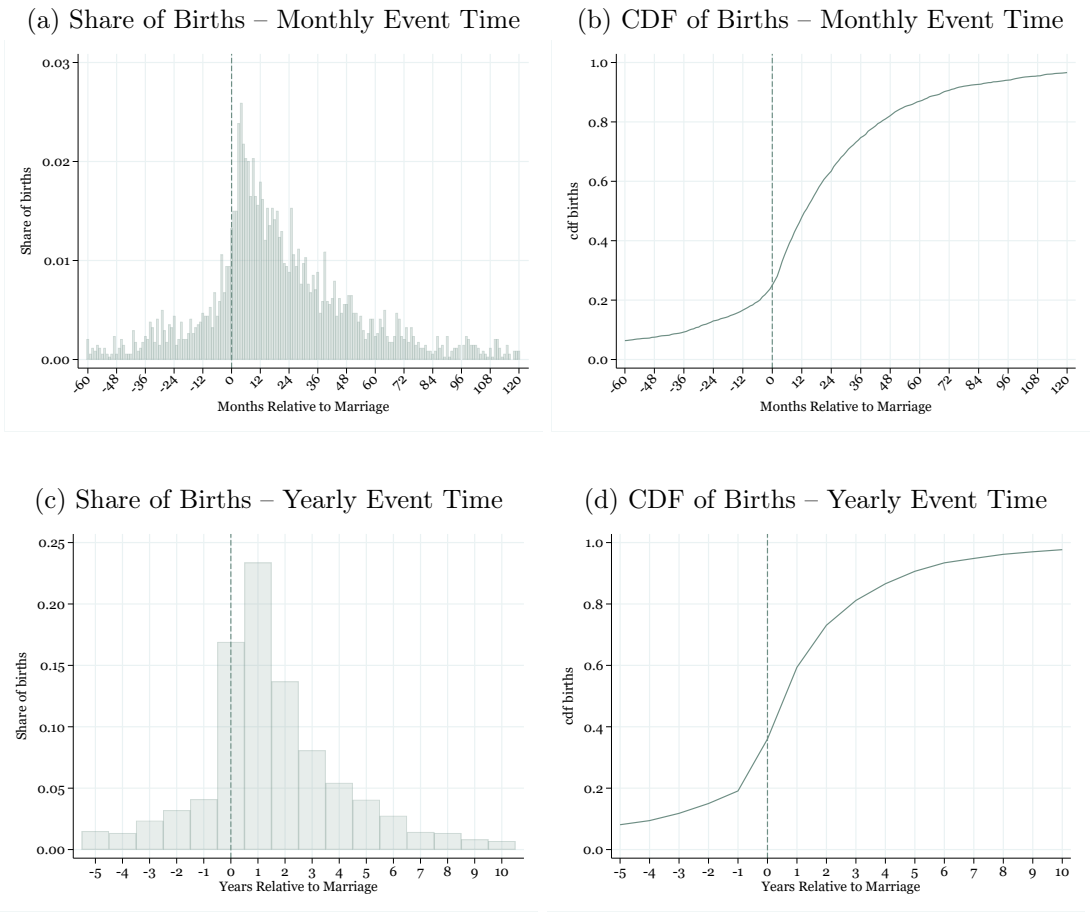
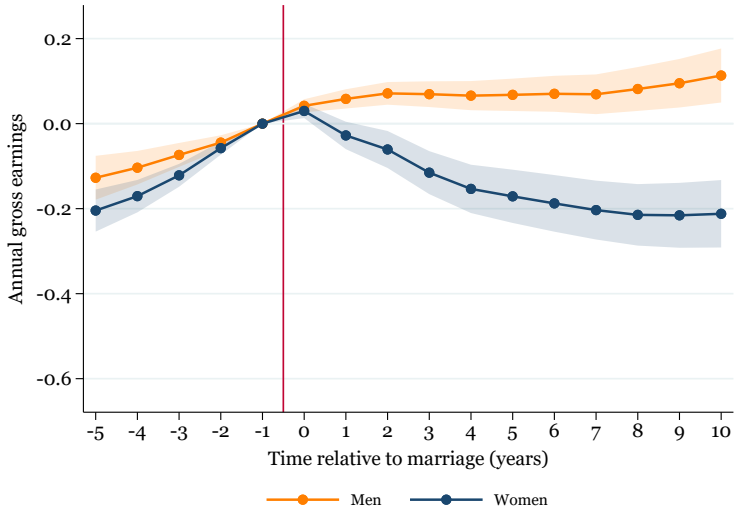


Figure 1.A.3: Timing Birth First Child Relative to Marriage

Notes: Data from SOEPRV. All panels visualize the distribution of the timing between birth of the first child and marriage for women who get married and have at least one child at some point in their (observed) life. The x-axis denotes time relative to marriage. Women who have their first child before marriage are on the left side of the dashed line, women who have their first child after marriage are on the right side. Panel a and c plot the share of births in a given period and panel b and d plot the corresponding cumulative density function (CDF). Panel a and b show monthly data for the subsample where we observe the exact month of marriage and childbirth. Panel c and d show yearly data for the full sample. *Reading example*: Panel a shows that the most common month (mode) to give birth to the first child is 4 months after marriage, implying that these women get married at about 5 months pregnant. Panel b shows that 24% of women have their first child before the month of marriage.

(a) Main Specification



(b) Account for Career Age

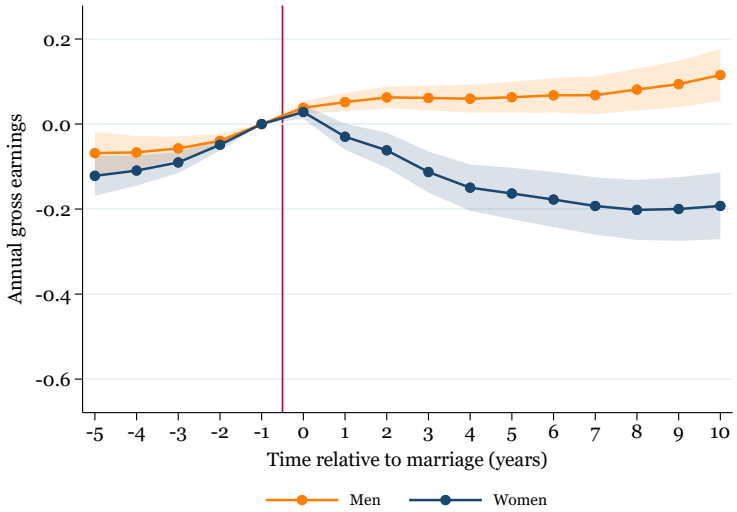
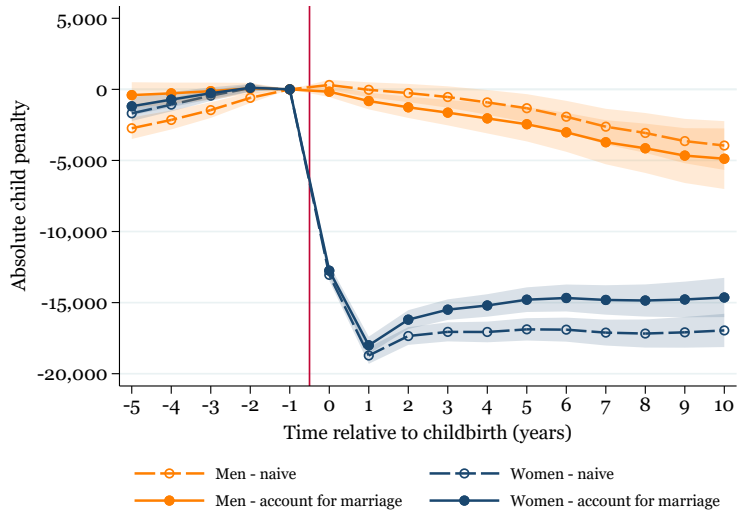


Figure 1.A.4: Pre-Marriage Income Growth and Labor Market Experience

Notes: Annual data from SOEPRV. Figure plots MG_t^g for gender g over event time t from estimating Equation 1.2 with 95% confidence intervals. Panel a shows results for the main specification. Panel b shows results from estimating a modified version of 1.2 that include fixed effects for the number of years since the first employment of individual i (including regular employment and self-employment).

(a) Absolute Estimates



(b) Relative Gap Estimates

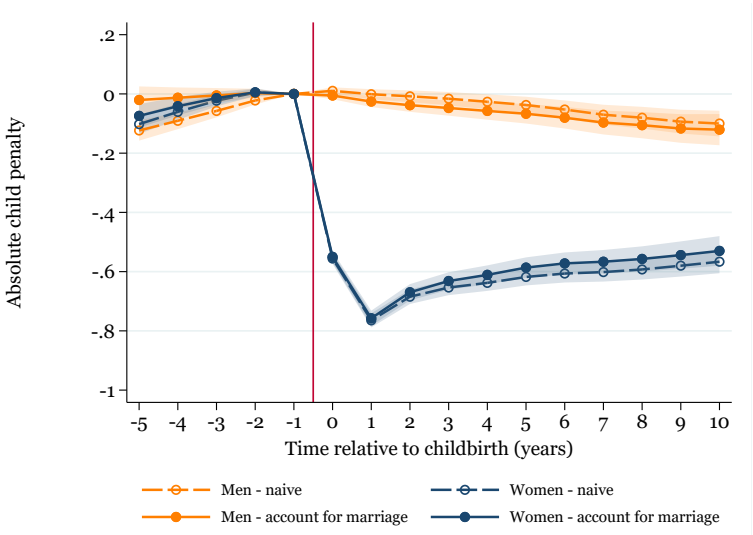


Figure 1.A.5: Child Penalties Accounting for Marriage

Notes: Annual data from SOEPRV. This figure shows the event study estimates from estimating the child gap (or child penalty) in monthly gross earnings. Plots show point estimates and 95% confidence intervals. Dashed lines plot naive estimations from estimating a modified version of Equation 1.1 but over event time relative to the birth of the first child. Solid lines plot estimations from estimating a modified version of Equation 1.2 but over event time relative to the birth of the first child while also including event time dummies for marriage. Panel a shows the absolute estimates $\hat{\beta}_t^g$ in 2015 Euros. Panel b shows relative earnings gap estimates for CG_t^g. The sample is not the same as for our main analysis: every individual in the sample here is a parent, but not all a married. Whereas in our main sample it's the opposite: everyone is married, but not all are parents.

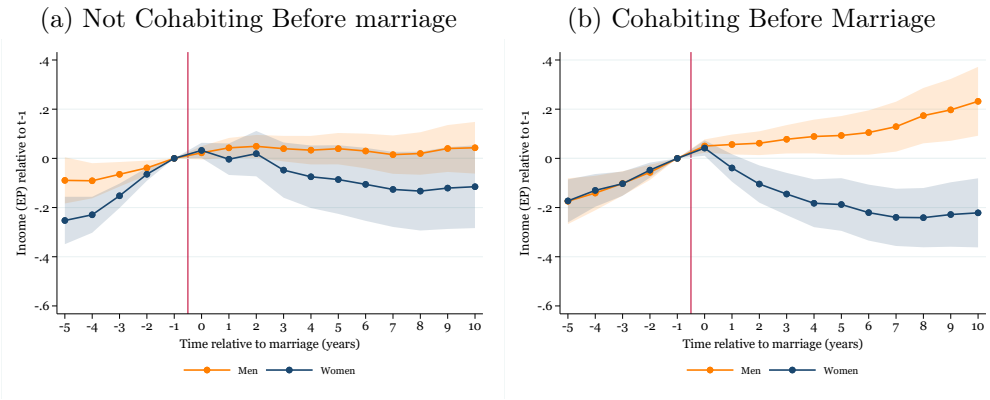


Figure 1.A.6: Marriage Earnings Gap by Cohabitation Status Before Marriage

Notes: Annual data from SOEPRV. This figure shows the event study estimates from estimating the marriage gap in yearly gross earnings. Plots show point estimates and 95% confidence intervals. Lines plot estimations from estimating Equation 1.2. Panel a shows the result for spouses who lived together one year before marriage and Panel b for spouses who lived separately. The sample is not the same as for our main analysis: every individual in the sample here has survey information on the relationship status prior to marriage.

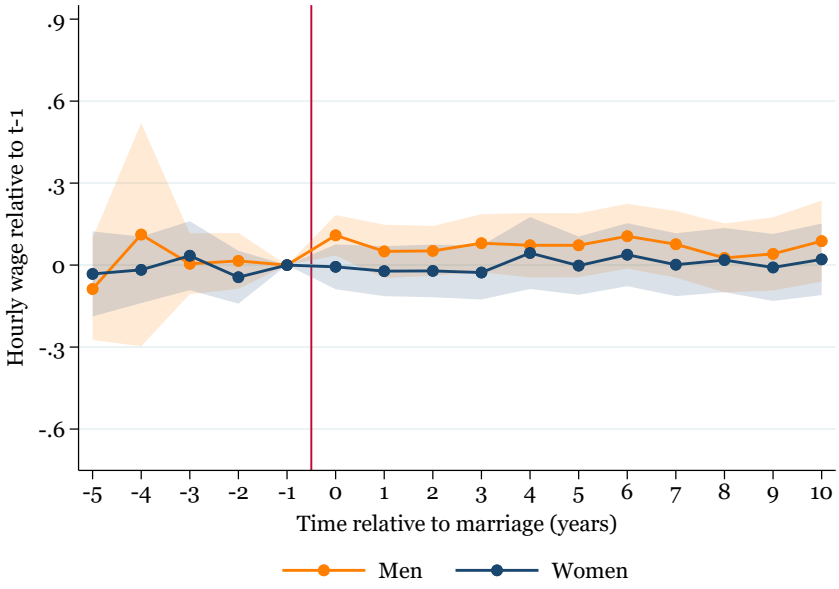


Figure 1.A.7: Hourly Wages – Survey Data Only

Notes: Annual data from SOEP. This figure shows the event study estimates from estimating Equation 1.2 for hourly wages. Hourly wages defined as gross income divided by time spent on paid work in the labor market. Both income and time-use data stem from survey responses in the SOEP data. This is the same analysis as shown in Figure 1.3d but with income from the survey data (same source as hours) rather than from administrative data. The sample is also the same, except for individuals for whom income is reported in the administrative data but not in the survey data (or vice versa).

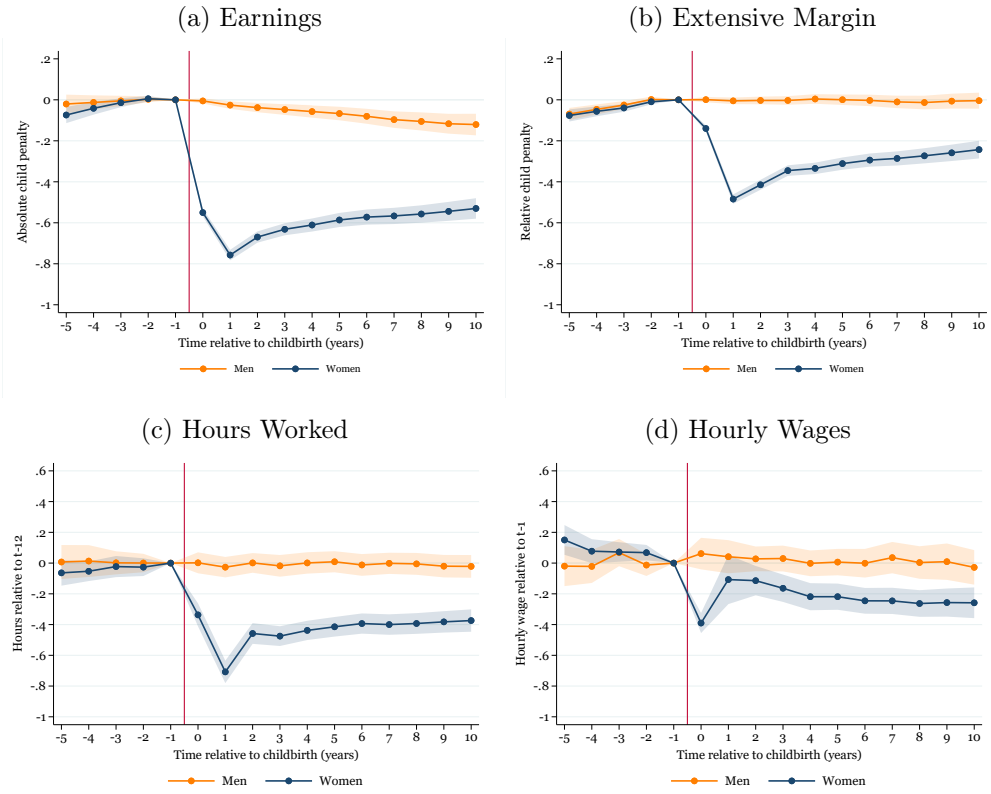


Figure 1.A.8: Decomposition Child Earnings Gap

Notes: Data from SOEPRV. This figure decomposes the overall child earnings gap (panel a) into extensive margin (panel b), intensive margin (panel c), and hourly wages (panel d). All panels plot MG_t^g estimates for women and men from estimating a modified version of Equation 1.2 but over event time relative to the birth of the first child, while also including event time dummies for marriage. For panel panel c and d event time is defined in years because the outcome relies on annual data.

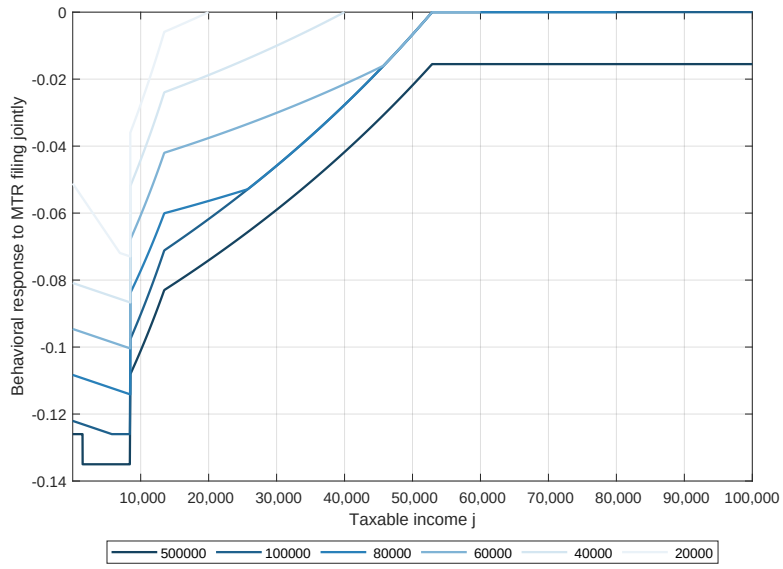


Figure 1.A.9: Simulated Income Changes for Secondary Earners

Notes: This figure illustrates the expected relative income response to ΔMTR_j from Figure 1.6. We define relative income response as $\Delta y_j/y_j = -\varepsilon_{y_j,1-MTR_j} \cdot \Delta MTR_j/(1-MTR_j)$. This figure plots $\Delta y_j/y_j$ over income of the secondary earner j for $\varepsilon_{y_j,1-MTR_j} = 0.3$ for five different income levels of the primary earner. Secondary earners with the same income y_j have the same x-coordinate, secondary earners married to spouses with the same income y_i are on the same line.

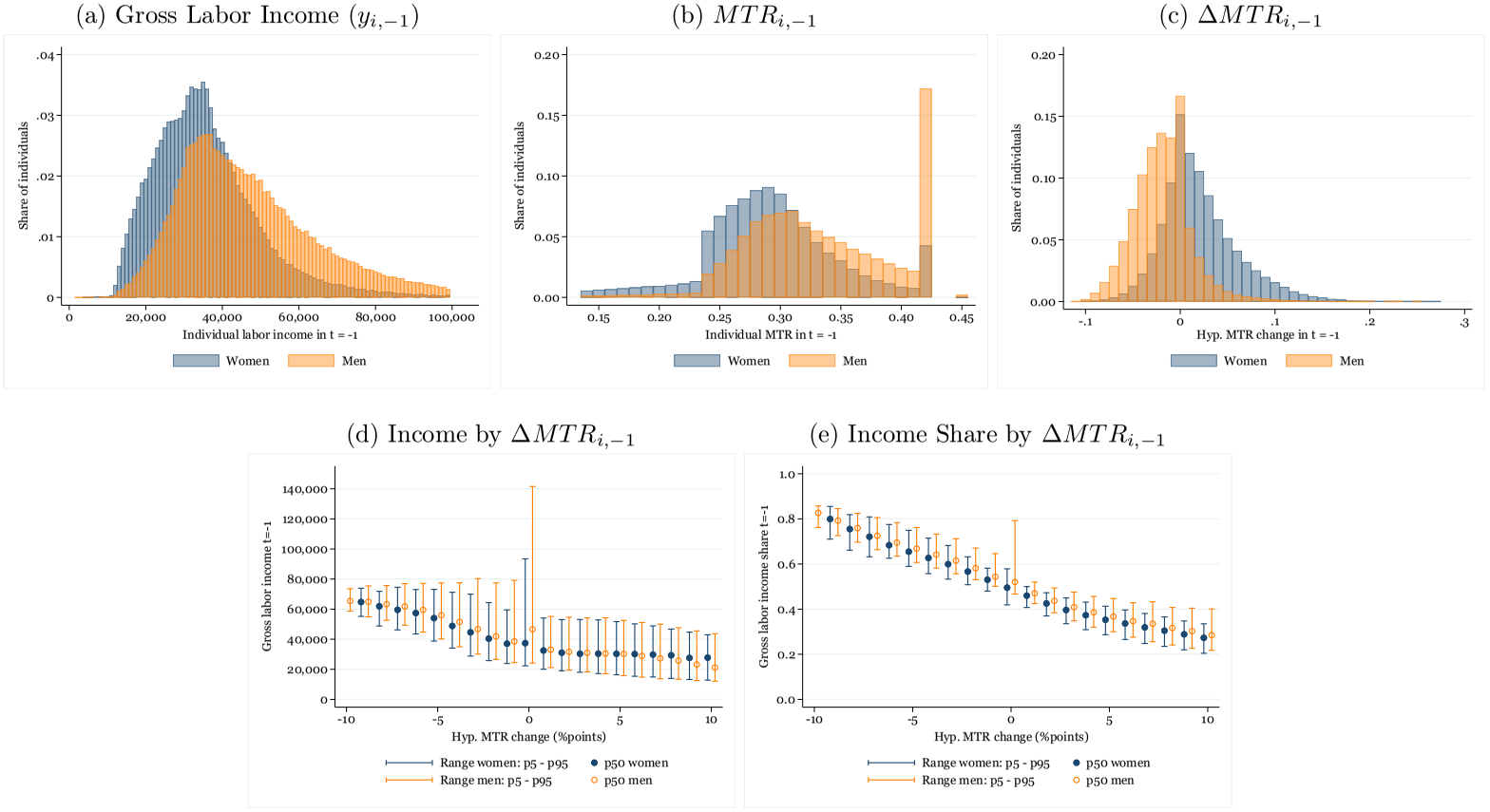


Figure 1.A.10: Descriptive Statistics MTR Sample

Notes: This figure plots descriptive statistics for the MTR sample that we use to estimate Equation 1.5 and 1.6. All panels show results for women and men for the pre-marriage year ($t = -1$). Income in 2015 Euros. Figures a, b, and c show histograms for the distribution of individual gross labor income ($y^g_{i,-1}$), the marginal tax rate ($MTR_{i,-1}$) and the absolute difference in marginal tax rate i would have faced if already filing jointly in $t = -1$ ($\Delta MTR_{i,-1} = MTR_{i,-1} - MTR^{hyp}_{i,-1}$), respectively. Panel d and e depict the variation in absolute income and within couple income share over Diff MTR. The vertical lines show the range for the medium 90% of the distribution (p5 to p95) and the dots show the median value (p50). Outliers in the p95 values at Diff MTR 0 because (very) high income taxpayers are less likely to face a change in MTR under joint taxation: they are in the top MTR brackets with a flat MTR.

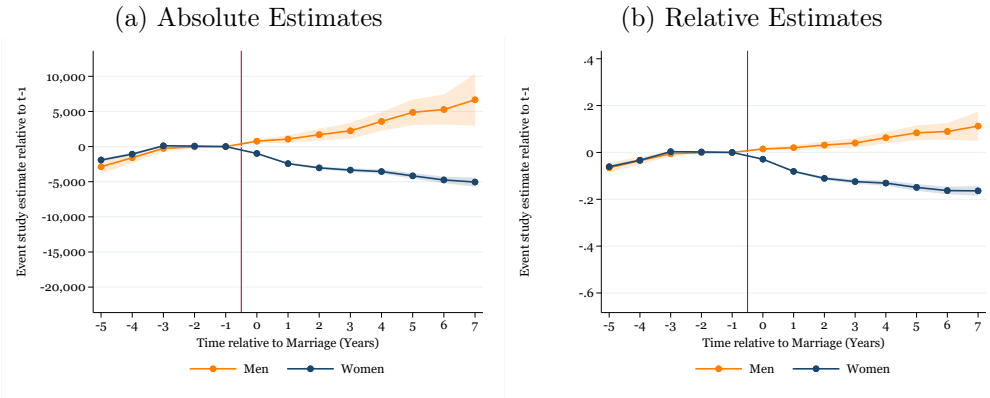


Figure 1.A.11: Marriage Earnings Gap (MTR Sample)

Notes: Annual data from TPP. This figure shows the event study estimates from estimating a modified version of Equation 1.2 for the MTR sample (this is the equivalent of Figure 1.2). Due to the TPP data structure, event time is measured in years and we include event time dummies for the first child only. Plots show point estimates and 95% confidence intervals. Panel a shows the absolute estimates $\hat{\beta}_t^g$ in 2015 Euros, directly retrieved from estimating Equation 1.2. Panel b shows relative earnings gap estimates for MG_t^g as defined in Equation 1.3.

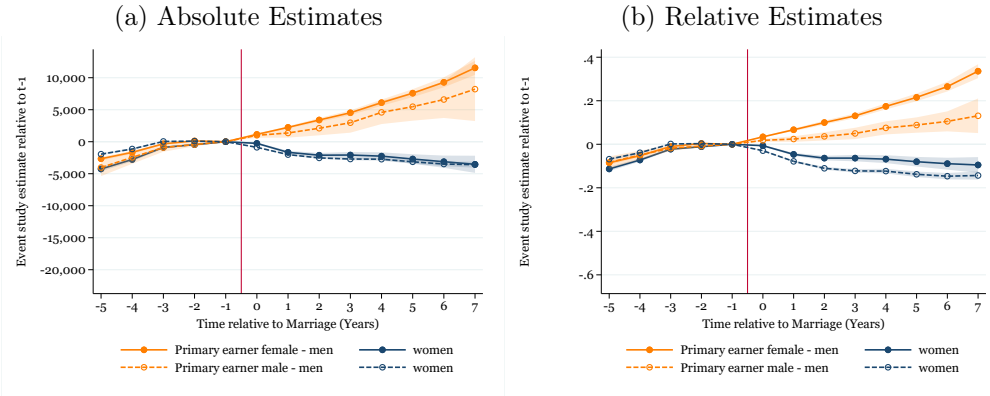


Figure 1.A.12: Marriage Earnings Gap Primary vs. Secondary Earners (MTR sample)

Notes: Annual data from TPP. This figure plots the marriage earnings gap estimates for two subsamples of the full sample from Figure 1.A.11: couples with female primary earners (solid lines) and couples with male primary earners (dashed lines). Based on their income in $t = -1$, primary earners face a labor supply incentive from a lower MTR under joint taxation, while secondary earners face a labor supply disincentive from a higher MTR. We exclude couples where at least one spouse has no change in MTR. These are typically (very) high-income couples, where the MTR remains in the flat top-MTR brackets with joint taxation. Plots show point estimates and 95% confidence intervals. Panel a shows the absolute estimates $\hat{\beta}_t^g$ in 2015 Euros. Panel b shows relative earnings gap estimates for MG_t^g.

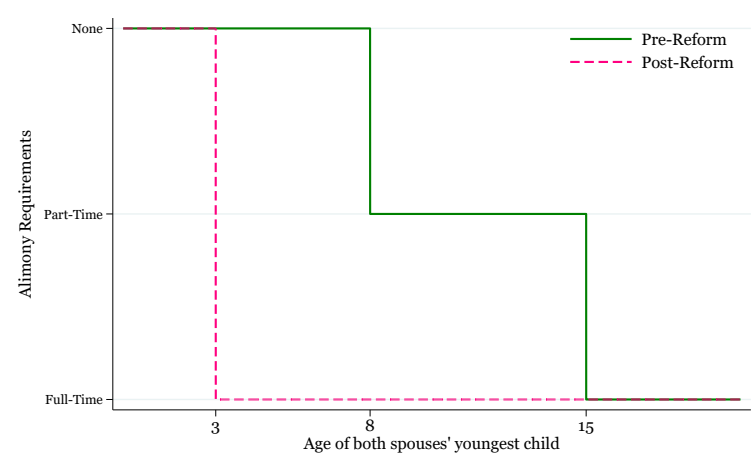


Figure 1.A.13: Illustration 2008 Reform

Notes: Illustration of the alimony payments as a function of child age pre-reform (green line) and post-reform (pink dashed line). Pre-reform, divorced resident parents are fully entitled to alimony payments from their ex-spouse until the youngest child turns 8. For children 8 to 14, the parent is expected to work part-time to be eligible. After that, the parent is no longer eligible. Post reform, eligibility ends as soon as the youngest child turns 3, expecting the caregiving parent to work full-time from this point.

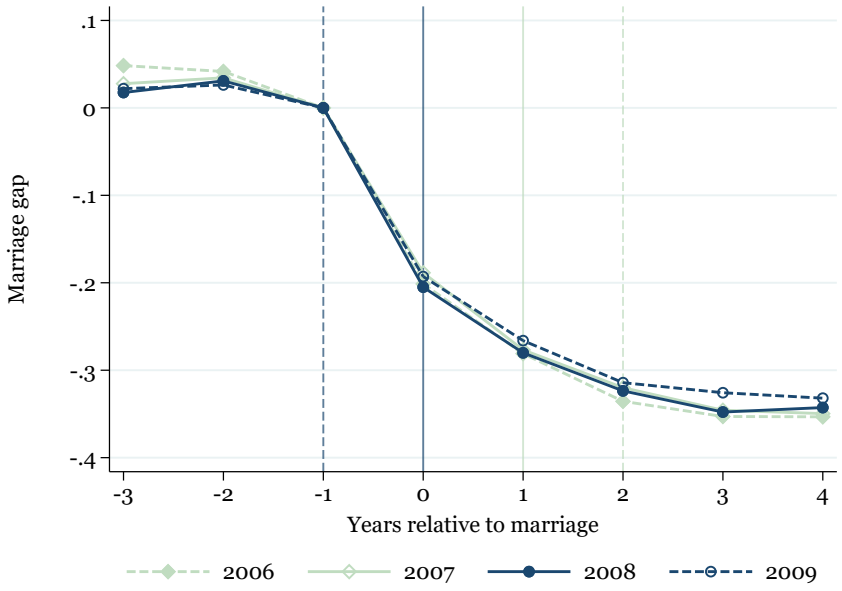


Figure 1.A.14: The Marriage Earnings Gap Around the Reform

Notes: Annual data from TPP. This figure shows the event study estimates MG_t^w from estimating Equation 1.2 for women of different year-of-marriage cohorts. 2006 and 2007 are the last two cohorts that got married under the old regime, 2008 and 2009 are the first two cohorts that got married under the new regime. Vertical lines indicate the event time t at which a cohort is hit by the 2008 reform. For the 2008 cohort, this is $t = 0$, for the 2007 cohort, this is $t = 1$, and so on. Figure 1.A.15 provides pairwise comparisons of the marriage earnings gap estimates for a larger set of year-of-marriage cohorts (2004 to 2011).

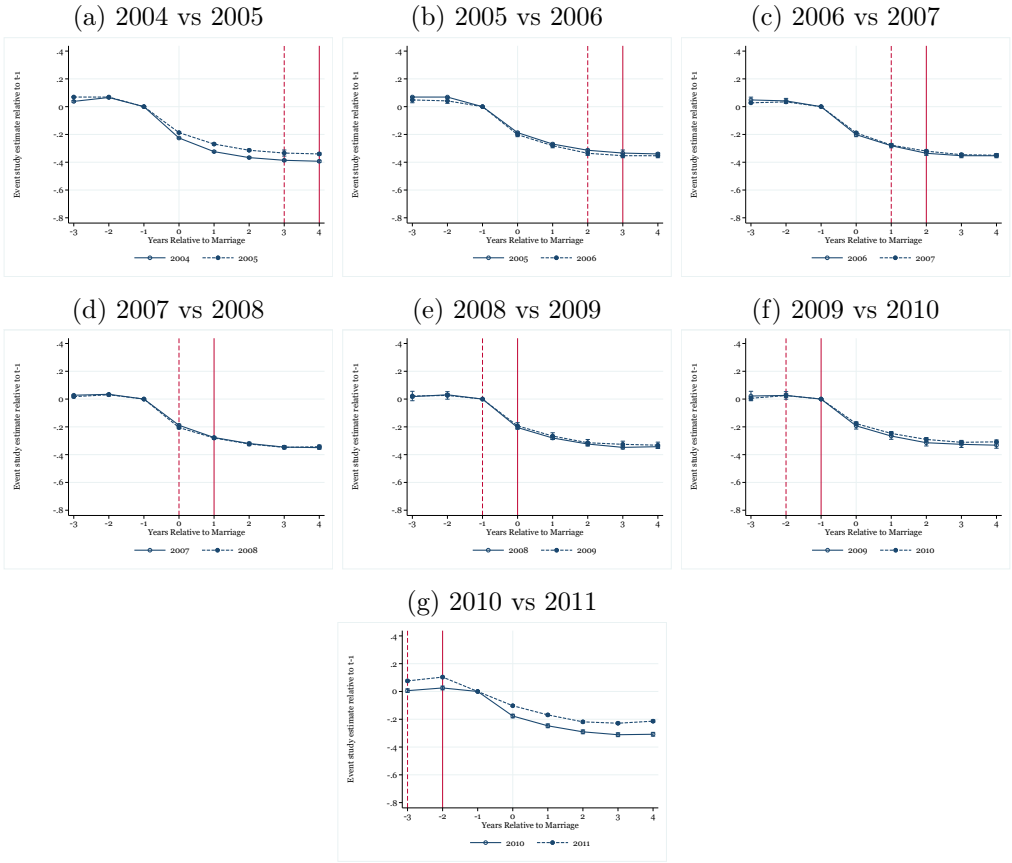


Figure 1.A.15: The Marriage Earnings Gap Around the Reform – Many Years

Notes: Annual data from TPP. This figure shows the event study estimates from estimating Equation 1.2 for different year-of-marriage cohorts. Each panel compares the marriage earnings gap for two marriage cohorts that got married in subsequent years. Vertical lines indicate the event time t in the year of the reform (2008) for the respective year-of-marriage cohort.

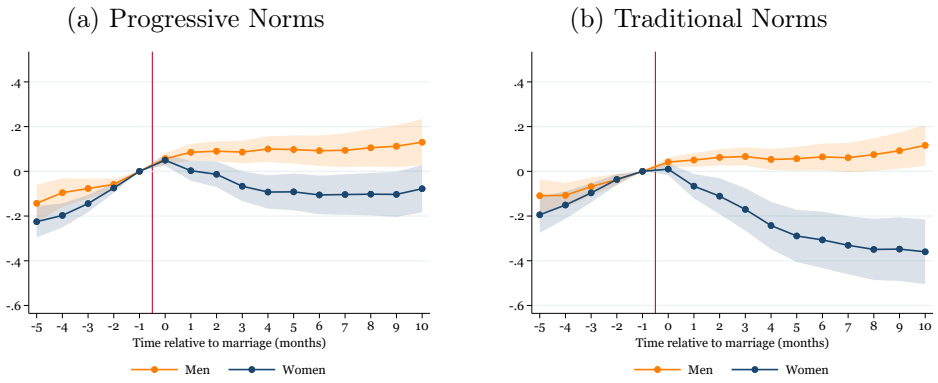


Figure 1.A.16: The Marriage Earnings Gap – Gender Norms

Notes: Annual data from SOEP–RV. This figure shows the event study estimates from estimating Equation 1.2 for individuals who stated their views on gender norms in the SOEP questionnaire. We split the sample into individuals with traditional (panel a) vs progressive (panel b) views on gender roles. See section 1.D for the underlying survey questions.

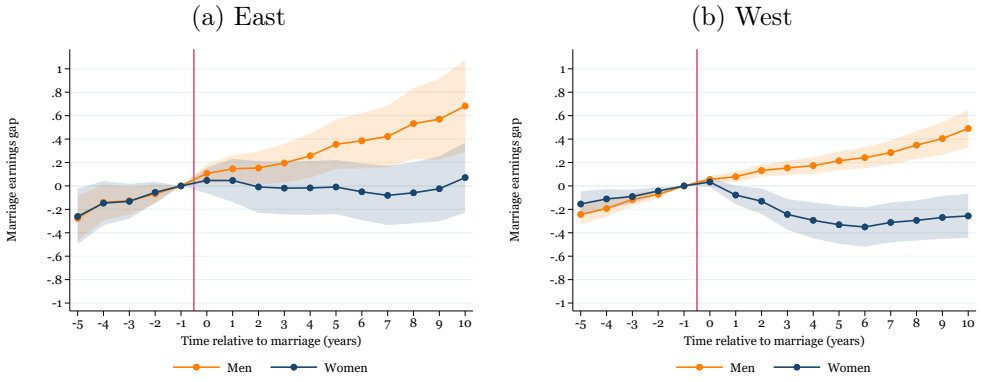


Figure 1.A.17: The Marriage Earnings Gap – East vs West

Notes: Annual data from SOEP–RV. This figure shows the event study estimates from estimating Equation 1.9. The figure plots the marriage earnings gap estimates for individuals who were born before 1989 in East Germany (panel a) vs. in West Germany (panel b) and got married after 1990, in the reunified country. We define the marriage earnings gap as $(\hat{\beta}_j^g + \hat{\delta}_j^g)/E[\tilde{Y}_{it}^g|t]$ for the East, and as $\hat{\beta}_j^g/E[\tilde{Y}_{it}^g|t]$ for the West. Estimates account for the child penalty and allow child penalty estimates to differ between East an West.

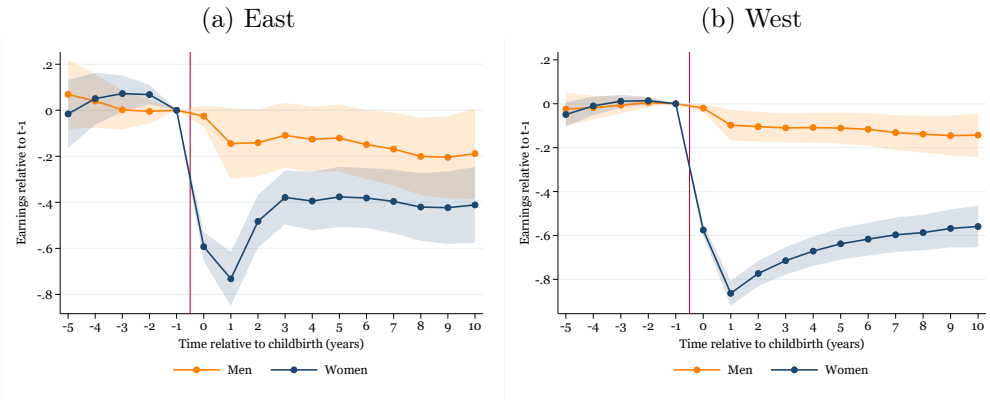


Figure 1.A.18: Child Penalties - East vs. West

Notes: Annual data from SOEPRV. This figure shows the event study estimates from estimating the child earnings gap (or child penalty) in relative earnings for individuals born before the German reunification ($<$ 1989) with childbirth after reunification ($>$ 1990). We split the sample into individuals born in the German Democratic Republic (East, panel a) and individuals born in the Federal Republic of Germany (West, panel b).

(a) Women

	(1) baseline	(2) income	(3) share	(4) both
t=0	0.70***	0.66***	0.35***	0.32***
	(0.00)	(0.00)	(0.00)	(0.00)
t=1	0.55***	0.50***	0.21***	0.19***
	(0.00)	(0.00)	(0.00)	(0.00)
t=2	0.51***	0.45***	0.17***	0.15***
	(0.00)	(0.00)	(0.00)	(0.00)
t=3	0.48***	0.42***	0.16***	0.14***
	(0.00)	(0.00)	(0.00)	(0.00)
t=4	0.47***	0.41***	0.15***	0.13***
	(0.00)	(0.00)	(0.00)	(0.00)
t=5	0.45***	0.40***	0.14***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)
t=6	0.45***	0.39***	0.13***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)
t=7	0.43***	0.37***	0.12***	0.11***
	(0.01)	(0.01)	(0.00)	(0.00)
Income (θ^w_{income})		✓		✓
Income share (ρ^w_{share})			✓	✓
Event time first child	✓	✓	✓	✓
Age (δ^w_{age}, ζ^m_{age})	✓	✓	✓	✓
Year (η_{cal})	✓	✓	✓	✓

(b) Men

	(1) baseline	(2) income	(3) share	(4) both
t=0	0.60***	0.51***	0.32***	0.30***
	(0.00)	(0.00)	(0.00)	(0.00)
t=1	0.42***	0.34***	0.19***	0.18***
	(0.00)	(0.00)	(0.00)	(0.00)
t=2	0.37***	0.30***	0.16***	0.15***
	(0.00)	(0.00)	(0.00)	(0.00)
t=3	0.35***	0.28***	0.14***	0.14***
	(0.00)	(0.00)	(0.00)	(0.00)
t=4	0.34***	0.28***	0.13***	0.13***
	(0.00)	(0.00)	(0.00)	(0.00)
t=5	0.32***	0.27***	0.13***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)
t=6	0.32***	0.26***	0.12***	0.12***
	(0.00)	(0.00)	(0.00)	(0.00)
t=7	0.31***	0.25***	0.11***	0.11***
	(0.01)	(0.00)	(0.00)	(0.00)
Income (θ^w_{income})		✓		✓
Income share (ρ^w_{share})			✓	✓
Event time first child	✓	✓	✓	✓
Age (δ^w_{age}, ζ^m_{age})	✓	✓	✓	✓
Year (η_{cal})	✓	✓	✓	✓

Notes: First stage results from estimating Equation 1.6 for women (panel a) and for men (panel b). Second stage results are reported in Table 1.4 and Table 1.A.2, respectively. For details on the specification, see table notes of Table 1.4 and 1.A.2.
Significance level: *** 0.01; ** 0.05; * 0.1. Robust standard errors are reported in parentheses.

Table 1.A.1: 2SLS first stage

	(1) baseline	(2) income	(3) share	(4) both
t=0	0.04**	0.03	0.02	0.01
	(0.02)	(0.02)	(0.02)	(0.02)
t=1	-0.01	-0.07**	0.03	-0.01
	(0.02)	(0.02)	(0.03)	(0.03)
t=2	-0.07**	-0.16***	0.03	-0.03
	(0.02)	(0.03)	(0.04)	(0.03)
t=3	-0.16***	-0.26***	0.01	-0.06
	(0.02)	(0.03)	(0.04)	(0.04)
t=4	-0.17***	-0.23***	-0.03	-0.07
	(0.03)	(0.03)	(0.05)	(0.05)
t=5	-0.20***	-0.22***	-0.01	-0.03
	(0.04)	(0.04)	(0.06)	(0.06)
t=6	-0.20***	-0.22***	0.17	0.10
	(0.05)	(0.05)	(0.09)	(0.09)
t=7	-0.18*	-0.14	0.07	0.08
	(0.08)	(0.09)	(0.15)	(0.15)
Income (θ^w_{income})		✓		✓
Income share (ρ^w_{share})			✓	✓
Event time first child	✓	✓	✓	✓
Age (δ^w_{age}, ζ^m_{age})	✓	✓	✓	✓
Year (η_{cal})	✓	✓	✓	✓

Notes: Estimates for the elasticity of gross income with respect to the net-of-tax rate, from estimating Equation 1.5 for men. This is the same analysis as reported in Table 1.4 for women. For more details see table notes of Table 1.4.
Significance level: *** 0.01; ** 0.05; * 0.1. Robust standard errors are reported in parentheses.

Table 1.A.2: Joint Taxation: Elasticity Estimates for Men (2SLS)

	$t=-1$ (drop out)	$t=0$ (joint filing)
Share children (pre marriage)	0.22	0.34
Share East Germany	0.18	0.16
Share no religion	0.34	0.36
Share protestant	0.27	0.25
Share catholic	0.32	0.32
Share other	0.07	0.06
N	1,929,430	1,726,221

Notes: TPP data. The table compares characteristics of women that drop out of the tax data (what we define as $t=-1$) and those of women that show up on their husbands tax records in $t=0$. All values are means. Samples are not identical as described in subsection 1.5.2.

Table 1.A.3: Sample Characteristics Women: Dropout vs. Observed

1.B Methodology Penalty Estimates

The child penalty literature typically plots the relative gap $G_t^g \equiv \hat{\beta}_t^g / E[\tilde{Y}_{it}^g|t]$ [24] but states a total penalty point estimate as the definite outcome. This total penalty is defined as $P_t \equiv (\hat{\beta}_t^m - \hat{\beta}_t^w)/E[\tilde{Y}_{it}^w|t]$, which is interpreted as the percentage by which women fall behind men after the event (see, e.g., Kleven, Landais, and Søgaard 2019). This measure compares the difference in the absolute income change of women and men in the numerator $(\hat{\beta}_t^m - \hat{\beta}_t^w)$ relative to women's counterfactual earnings in the denominator $(E[\tilde{Y}_{it}^w|t])$. P_t is thus unaffected by men's counterfactual earnings $E[\tilde{Y}_{it}^m|t]$ and does not account for differences in female and male counterfactual earnings.

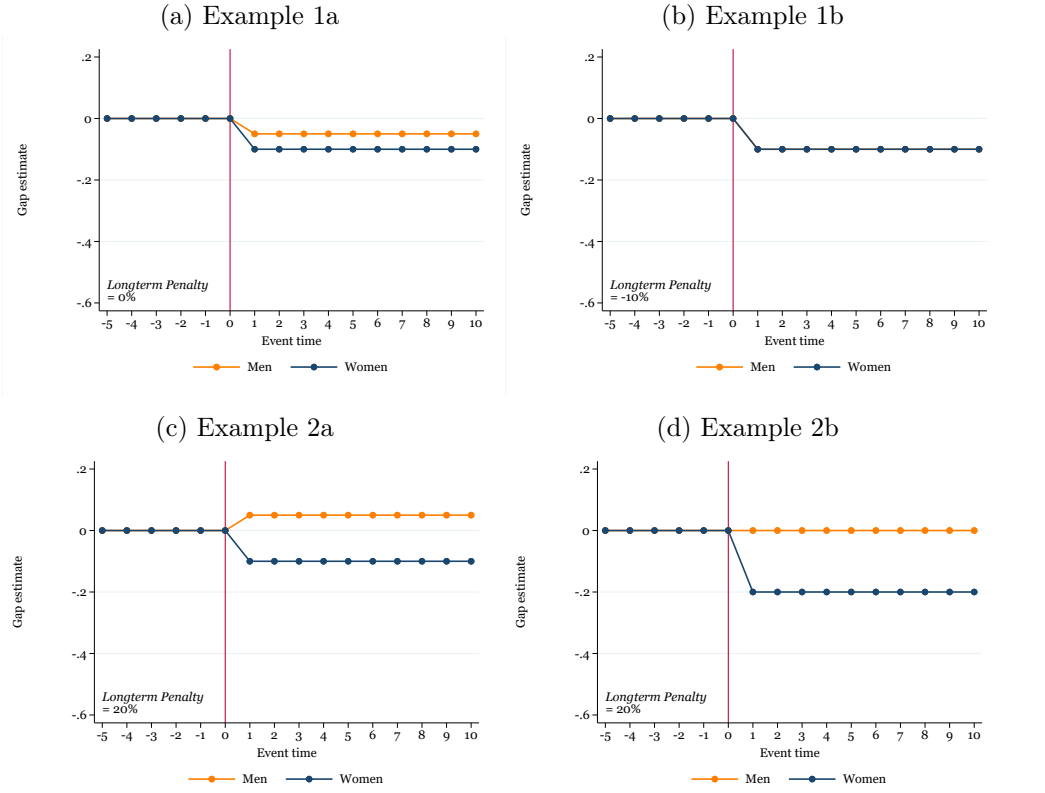


Figure 1.B.1: Illustration Gap Estimates

Notes: This figure illustrates the estimates for G_t^w (orange) and G_t^m (blue) for the 4 illustrative examples described in the text. For all panels, $E[\tilde{Y}_{it}^m|t] = 2{,}000$ Euro and $E[\tilde{Y}_{it}^w|t] = 1{,}000$ Euro for $t \leq 0$.

As a result, the interpretation of P_t is not straightforward, particularly if the absolute female and male counterfactual earnings in t are different. We illustrate this with four simple examples, assuming constant, but different incomes for women and men that are only affected by t for $t > 0$, with counterfactual incomes $E[\tilde{Y}_{it}^m|t] = 2,000$ and $E[\tilde{Y}_{it}^w|t] = 1,000$ for all t.

[24] We denote the gap by G_t^g here, while the child penalty literature typically denotes this by P_t^g (see e.g., Kleven, Landais, Posch, Steinhauer, and Zweimüller 2019). This is equivalent to our MG_t^g definition for marriage.

Since P_t is a relative measure based on absolute differences in event-time estimates of women and men that does not consider differences in the counterfactual income, differences in G_t^m and G_t^w do not directly translate into P_t – and vice versa. This can be illustrated by considering two extreme cases. First, assume that women and men face the same absolute drop in earnings in $t = 1$, e.g., $\hat{\beta}_t^w = \hat{\beta}_t^m = -100$ for $t > 0$. In this case, women's earnings are 10 percent lower than their counterfactual earnings in $t > 0$, while the relative gap is 5 percent for men, as illustrated in Figure 1.B.1a. This scenario results in a penalty of $P_t = (-100 - (-100))/1,000 = 0$. Put differently, the point estimate for the penalty here is zero, although, in relative terms, women reduce their earnings twice as much as men with $G_t^w = -.1$ and $G_t^m = -.05$ for all $t > 0$.

Now consider the other extreme case where women and men face the same relative drop in earnings in $t = 1$, e.g., $\hat{\beta}_t^w = -100$ and $\hat{\beta}_t^m = -200$ for $t > 0$, resulting in a 10% drop for both genders as shown in Figure 1.B.1b. While the relative gap is the same, this scenario results in a penalty of $P_t = (-200 - (-100))/1,000 = -10\%$, indicating that female earnings increase relative to male earnings (or decrease less). These examples show that since male earnings enter the numerator but not the denominator, P_t is not necessarily informative of the relative gap $G_t^g = \hat{\beta}_t^g / E[\hat{Y}_{it}^g | t]$.

In addition, the P_t estimate expresses changes in female earnings always relative to changes in male earnings. For given $E[\hat{Y}_{it}^w | t]$ the P_t estimates are the same as long as the absolute difference $\hat{\beta}_t^m - \hat{\beta}_t^w$ is the same. As a result, P_t is not informative about which gender experiences a change in earnings and of what size. To illustrate this, assume first that women's absolute earnings decrease by 100 for $t > 0$, while men's earnings increase by 100 ($\hat{\beta}_t^w = -100$ & $\hat{\beta}_t^m = 100$). In this case, women's earnings are 10 percent lower than their counterfactual earnings, while men's earnings increase by 5 percent post event (Figure 1.B.1c). This results in a penalty of $P_t = (100 - (-100))/1,000 = 0.2$. Now assume that men's earnings remain unchanged while women's earnings decrease by 200 ($\hat{\beta}_t^w = -200$ & $\hat{\beta}_t^m = 0$). This again results in a penalty of $P_t = (0 - (-200))/1,000 = 0.2$, while the relative gap estimates are very different as illustrated in Figure 1.B.1d. Specifically, the relative gap for women is now twice as large, while men face no change in earnings.

Since this paper focuses on relative earnings changes post marriage, we decided to not use the P_t estimates.

1.C International Comparison

In this section, we reproduce our results for 25 European countries plus Israel.

Data For this cross-country analysis, we use survey data from the SHARE Job Episodes Panel (SHARE JEP, Brugiavini, Orso, Genie, Naci, and Pasini 2022). The SHARE JEP is a retrospective panel, covering the entire biography of respondents from the Survey of Health, Ageing and Retirement in Europe (SHARE). SHARE JEP focuses on older individuals (respondents are aged 50+), but provides information on their labor market status, marriage status, and number of children for every year in a respondent's life. This allows us to estimate Equation 1.2 for the extensive margin. The survey is run in a total of 26 countries: Austria, Belgium, Bulgaria, Croatia, Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Ireland, Israel, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland.

	All	Women	Men
Age at marriage	25	23	26
Year of marriage	1971	1971	1972
Share divorced	0.17	0.17	0.17
Years married	14	13	15
Share children (ever)	0.91	0.92	0.91
N	4,498	2,417	2,081

Data: SHARE, raw data for Germany (unweighted), mean values. Individuals that are married during the sample period, aged 18 to 45 at marriage and observable at minimum 4 years before and 8 years after marriage.

Table 1.C.1: Descriptive Statistics Germany SHARE

Marriage Gap We closely follow our main analysis and estimate Equation 1.1 separately for men and women. Our outcome Y_{it}^g is as a binary dummy of labor market participation that is 1 if respondent i worked in event year t and 0 else, i.e., the extensive margin (working yes/no). Event time is measured in years relative to marriage.

	All		Never		Before		During		After	
	w	m	w	m	w	m	w	m	w	m
AT	1,987	1,390	147	120	322	194	1,488	1,046	30	30
BG	1,104	790	86	68	26	19	989	699	3	4
BR	3,207	2,525	326	277	89	81	2,758	2,131	34	36
CH	1,479	1,203	186	160	63	52	1,218	969	12	22
CY	703	481	23	13	9	5	671	461	0	2
CZ	2,917	1,941	109	84	142	76	2,642	1,752	24	29
DE	2,488	2,129	195	213	214	151	2,036	1,712	43	53
DK	2,055	1,703	171	168	399	332	1,456	1,170	29	33
EE	2,836	1,753	143	97	292	149	2,362	1,483	39	24
ES	2,918	2,290	257	212	79	71	2,577	1,995	5	12
FI	964	813	51	57	64	67	836	682	13	7
FR	2,459	1,761	170	133	153	115	2,095	1,478	41	35
GR	2,280	1,694	257	188	30	26	1,992	1,473	1	7
HR	1,283	994	66	50	50	39	1,159	898	8	7
HU	903	571	51	32	18	14	828	520	6	5
IE	441	327	27	20	9	10	403	295	2	2
IL	1,098	819	155	131	42	31	893	650	8	7
IT	2,785	2,328	220	185	61	46	2,498	2,093	6	4
LT	1,230	695	72	45	48	19	1,094	621	16	10
LU	642	528	51	45	32	31	548	441	11	11
LV	1,011	569	54	30	69	34	877	490	11	15
MT	581	458	110	91	2	4	467	363	2	0
NL	1,124	937	135	101	25	17	956	807	8	12
PL	2,928	2,294	127	114	146	88	2,646	2,081	9	11
PO	643	512	46	37	25	19	570	455	2	1
RO	1,166	848	52	41	52	44	1,046	749	16	14
SE	1,972	1,666	118	137	350	308	1,469	1,194	35	27
SI	1,972	1,381	86	73	208	118	1,671	1,184	7	6
SK	1,067	877	97	105	41	23	928	746	1	3

Notes: Number of married individuals (w = women, m = men) for each country in the SHARE data. *All*: all individuals who get married at some point in their life with $18 \leq$ age ≤ 45 at their first marriage. The other columns refer to subsamples based on whether and when the individual has children. We differentiate between individuals who never have children (*Never*), those who have their first child before their first marriage (*Before*), during that marriage (*During*), or after divorce (*After*).

Table 1.C.2: Sample Size SHARE

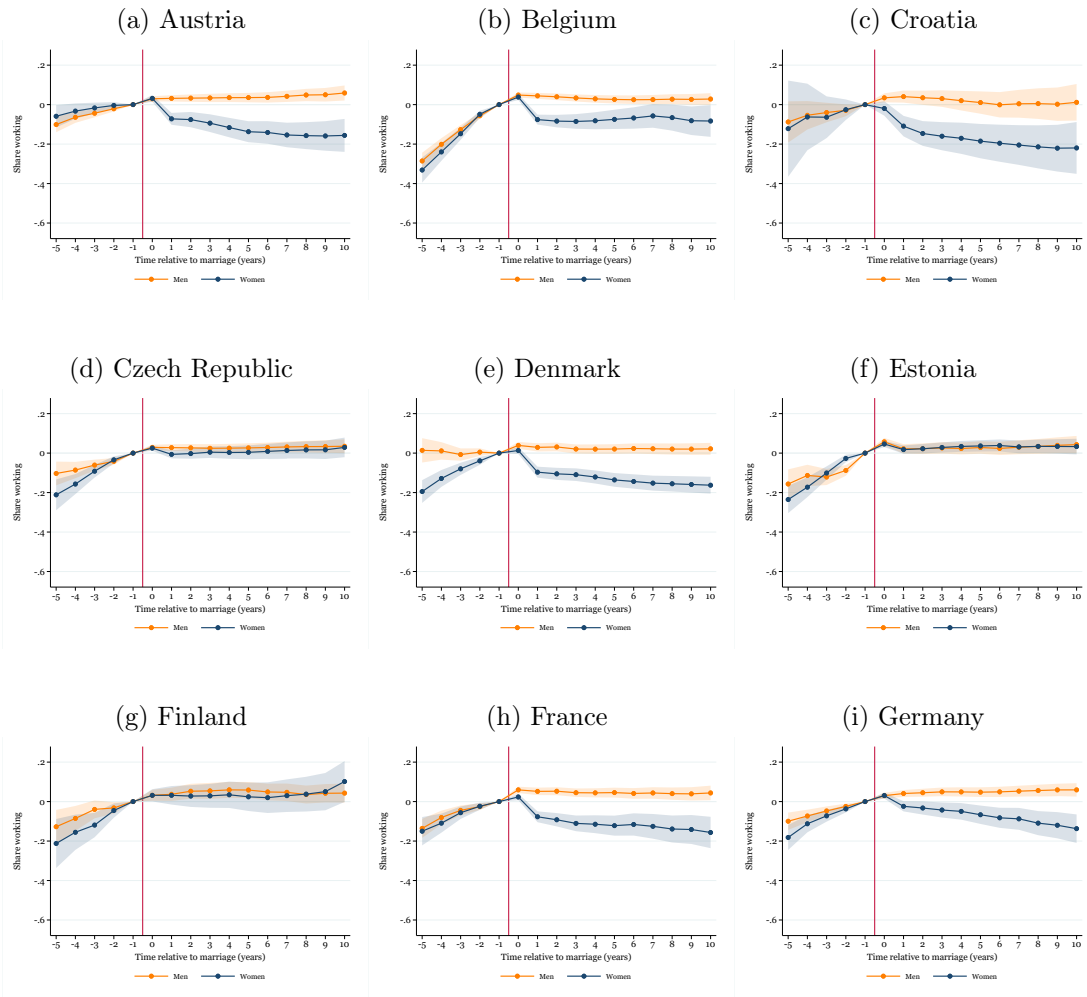


Figure 1.C.1: Extensive Margin Marriage Gap – International Comparison

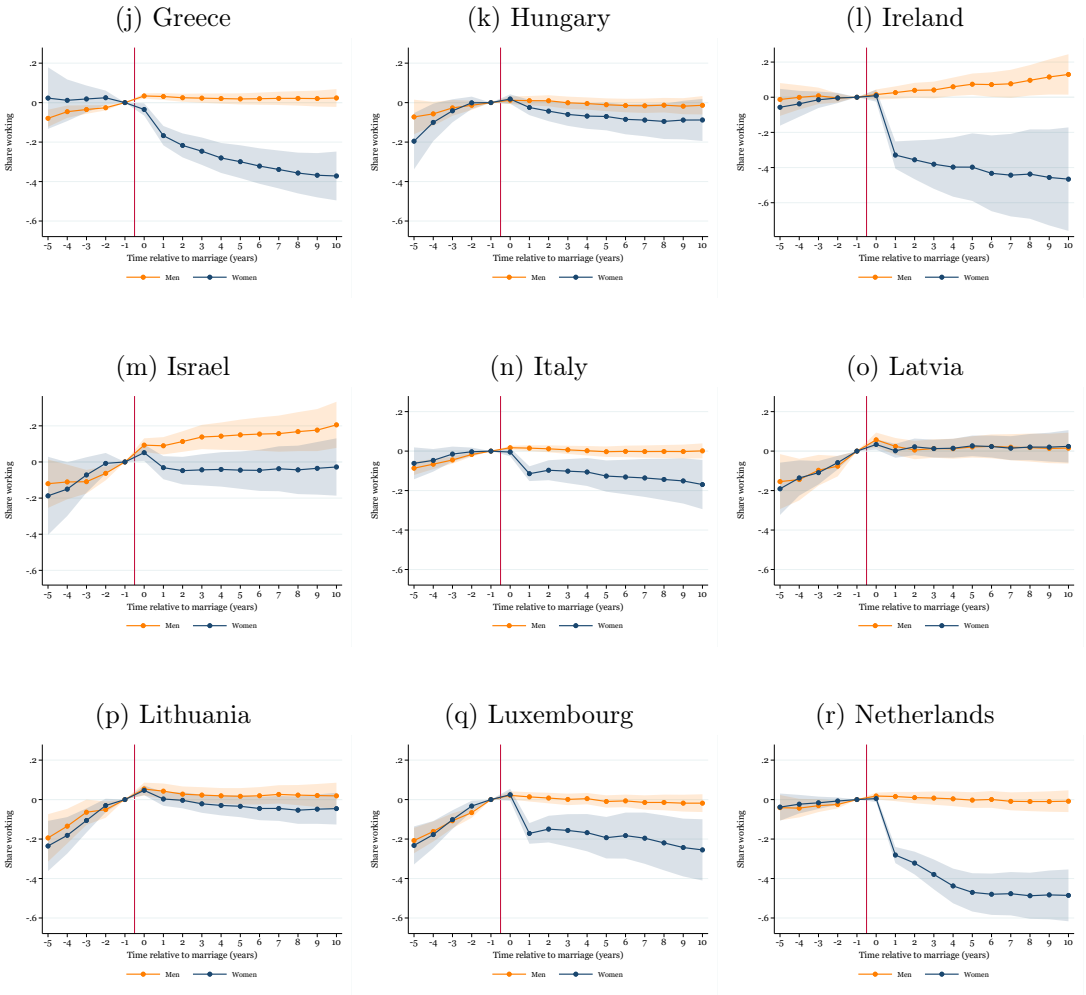


Figure 1.C.1: Extensive Margin Marriage Gap – International Comparison (cont.)

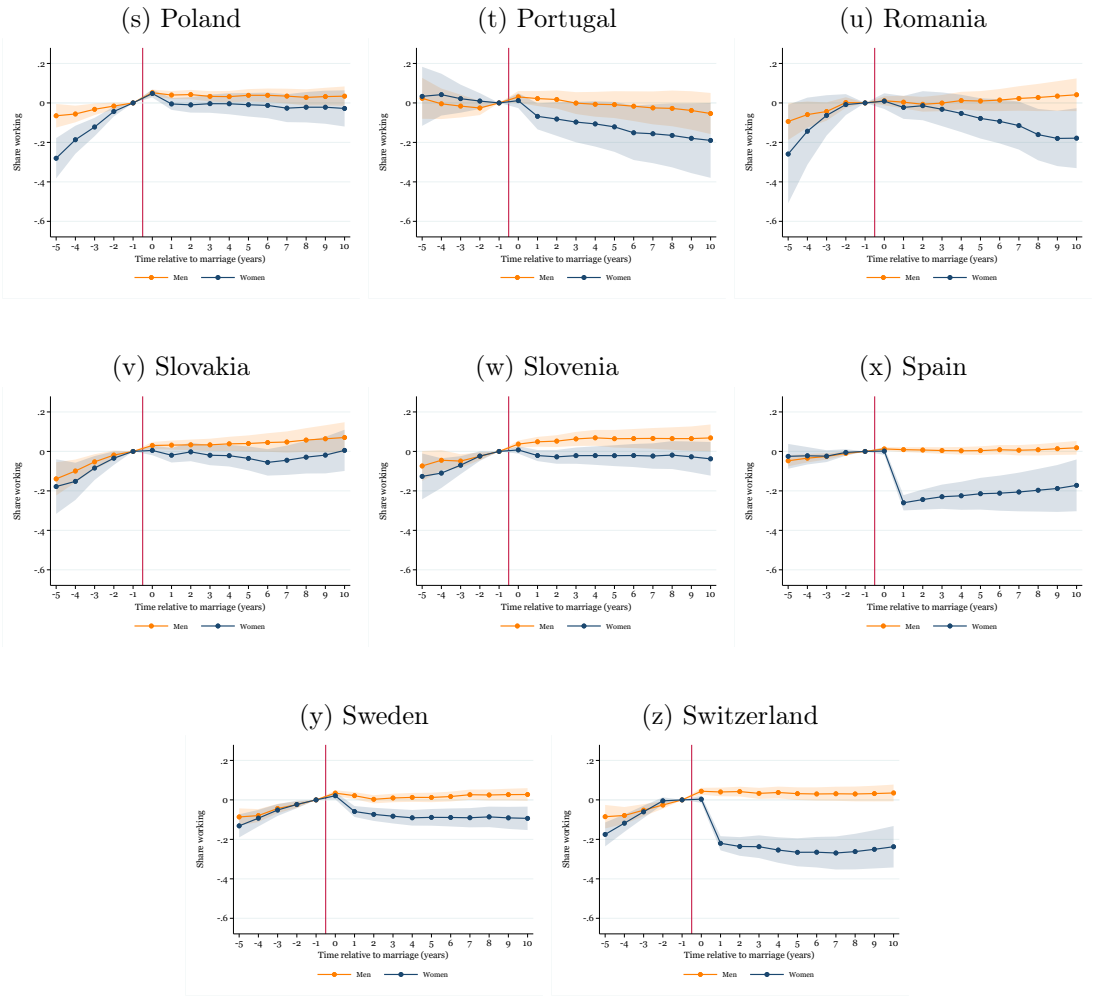


Figure 1.C.1: Extensive Margin Marriage Gap – International Comparison (cont.)

Notes: Extensive margin marriage earnings gap (outcome = working yes/no). Results from estimating Equation 1.2 with the SHARE JEP data. All panels plot the marriage earnings gap estimates $\widehat{MG}_t^q$ for the extensive margin. *Time relative to marriage*: years before / after the first marriage.

1.D Gender Norms Questions SOEP

Below are the SOEP questions used for our analysis on the gender norms of the survey participant in subsection 1.5.3:

1. **Gender Norm Question 1:** Child under 6 years suffers when mother works
2. **Gender Norm Question 2:** Child under 3 years suffers when mother works
3. **Gender Norm Question 3:** Women Should Rather Care About Family Than Career
4. **Gender Norm Question 4:** Best if man and woman work the same amount so they can share the responsibility

Chapter 2

Optional (Non-)Filing and Effective Taxation*

2.1 Introduction

Income inequality is one of the major concerns of today's societies and many countries rely on progressive income tax schedules to reduce it. Effective taxation, however, often diverges from statutory tax schedules. One mechanism that has received a lot of attention in both economic research and political debate is tax evasion at the top of the income distribution (e.g., Guyton, Langetieg, Reck, Risch, and Zucman 2021; Sarin 2023). Tax evasion by rich taxpayers reduces their effective tax rates and thus weakens the effective progressivity of a tax system (e.g., Alstadsæter, Johannesen, and Zucman 2019). However, lower effective tax rates at the top of the distribution are only one mechanism that weakens tax progressivity.

In this paper, we focus on the other end of the distribution and quantify the impact of optional non-filing on effective taxation of low-income taxpayers. Under optional non-filing, certain taxpayers are not required to file an income tax return. More than thirty countries worldwide have a legal tax non-filing options for taxpayers, typically for employees where automatic wage tax withholding ensures that they do not owe additional income taxes (no under-withholding).[1] However, non-filers may face over-withholding, and pay more taxes than intended by the tax schedule. With a progressive tax schedule, over-withholding can occur due to fluctuations in income between payroll periods. In such cases, the projected annual income underlying the withholding in a given payroll period does not align with the realized annual income.

To the best of our knowledge, we are the first to empirically quantify the effective tax rates under non-filing and to emphasize its role for effective income tax progressivity. Analyzing tax (non-)filing in Germany, we show that optional non-filing acts like "reverse evasion": It increases effective tax rates at the bottom of the income distribution, because low-income taxpayers are most likely to refrain from optional filing while facing substantial tax over-withholdings. Although fundamentally different from evasion, non-filing also weakens the effective tax progressivity because low-income individuals pay more taxes than they have to.

Data on non-filers is often scarce, because, by definition, there is no tax return data for them. We overcome this by using administrative German income tax data that combines tax return data for filers with employer provided data for non-filers (Research Data Centre of the Federal Statistical Office and the statistical offices of the Länder

*This chapter is based on co-authored work with Tobias Hauck. A version of this chapter has been published as an open access article: Tobias Hauck and Luisa Wallossek (2024). "Optional (non-)filing and effective taxation." *Journal of Public Economics*, 238, 105187. DOI: 10.1016/j.jpubeco.2024.105187.

[1]For an overview of international non-filing regulations, see International Bureau of Fiscal Documentation (2016).

(RDC) 2018, 2020). Based on annual gross income and individual characteristics, we calculate the statutory tax liability for non-filers and compare it to their effective tax liability to quantify over-remittances at the individual level. German taxpayers are given the option not to file if automatic withholding guarantees no under-withholding, typically in cases where taxpayers have only wage income. This group of optional filers constitutes approximately half of the German taxpayer population. While non-filing can also refer to non-compliance with filing obligations as a form of tax evasion (e.g. Congressional Research Service 2023; Erard and Ho 2001), we focus on optional non-filers and exclude those who have to file.

Our empirical analysis provides three major insights on tax non-filing and its distributional impact on the effective taxation of income. First, we show that non-filing is very common and particularly so at the lower end of the income distribution. Overall, 61% of optional filers do not file. At the bottom of the income distribution, the non-filing share is as high as 90%.

Second, we quantify absolute tax over-remittances that result from non-filing by comparing the effective tax withholding observed in the data to the statutory tax liability as defined in the tax schedule. Non-filers over-remit taxes at all income levels and over-remittances are substantial with an annual value of at least 951 million €. On the individual level, the average non-filer over-remits 119 €. We show that this mean value is not driven by a few extreme cases. One third of the non-filers overpay taxes, representing 2.6 million taxpayers. In addition, we provide suggestive evidence that filing costs are unlikely to explain non-filing given the observed over-remittances.

Third, we compute the effective average tax rates (ATRs) for non-filers and show that they effectively face a different tax schedule than filers. Comparing this effective tax schedule to the statutory tax schedule, we emphasize the role of non-filing for effective redistribution. On average, non-filers' effective ATRs are 1.2 percentage points higher than their statutory ATRs. Non-filing increases effective ATRs at all income levels, but especially at the lower end of the income distribution, where non-filing is most common. Similar to other progressive income tax schedules, the German tax schedule features a zero statutory ATR for lowest-income taxpayers with income below a basic allowance threshold. Although they should pay zero taxes, non-filers in this income range effectively face positive ATRs, averaging close to 2% in many income bins. As a result, the effective progressivity of the income taxation is weakened compared to the statutory tax schedule.

Optional non-filing does not only impact effective taxation in Germany. To illustrate this, we use aggregated Austrian tax data and document similar over-remittances under automatic withholding. In addition, we simulate effective taxation for low-income US non-filers and show that automatic withholding mechanisms lead to qualitatively similar divergences between effective and statutory ATRs.

We then turn to the question of efficiency and analyze whether increasing effective tax rates via non-filing is an efficient way to generate tax revenue. Increasing statutory marginal tax rates (MTR) for a given income range, the textbook case, raises tax revenue from taxpayers in that range as well as from higher-income taxpayers. Higher effective MTRs for non-filers have no such mechanical effect on other taxpayers. Hence, most of the revenue potential is not realized when increasing only effective MTRs.

We benchmark the status quo against two hypothetical tax schedules to illustrate that increasing effective tax rates via non-filing generates less tax revenue than a comparable increase in statutory tax rates.

Our results have clear policy implications: Allowing for non-filing impacts the effective taxation and policy makers should account for that when designing the income tax system. To correct for over-withholding, the straightforward policy is to automatically refund over-withholdings to optional filers, while allowing for the possibility to file an income tax return for taxpayers who want to itemize deductions. Automatic refunds realign effective and statutory taxation and restore the level of effective tax progressivity as defined in the tax schedule without imposing any filing costs on non-filers. The role of non-filing is even more pronounced whenever economic support payments or tax credits such as the EITC are administered via the tax schedule and require filing from eligible taxpayers.

Our work relates to an evolving literature on optional tax filing. We contribute to this literature by documenting the substantial role of optional non-filing for effective taxation, which has not been discussed before. Literature on non-filers often focuses on unclaimed refunds from social welfare or other payments that are administered via the income tax code (e.g., Goldin, Homonoff, Javaid, and Schafer 2022; Guyton et al. 2017; Ramnath and Tong 2017). Here, we also add to a broader literature on incomplete take-up of social welfare programs (e.g., Bhargava and Manoli 2015; Currie 2006; Finkelstein and Notowidigdo 2019). We contribute to this literature by documenting that taxpayers leave money on the table in a context where two common explanations for incomplete take-up can be ruled out: there is no social stigma linked to tax filing, and filing costs are limited since the German tax system requires only minimal tax filing to correct for over-withholding. Another key difference compared to other settings is that policy makers can address the incomplete take-up of tax filing without necessitating changes in individual behavior. Tax authorities already have all relevant information at hand to refund over-withheld taxes for those who do not file.

In addition, research on optional filing shows that taxpayers forgo additional tax refunds to avoid compliance costs from filing (Benzarti 2020) and are more likely to file when expected returns from filing increase, with positive effects on economic outcomes (Ramnath and Tong 2017). In a concurrent paper, Goodman, Lim, Sacerdote, and Whitten (2023) analyze the accuracy with which the IRS could pre-populate tax returns for US taxpayers. They estimate significant tax over-remittances for optional US non-filers, supporting the international relevance of non-filing for effective taxation. Beyond providing point estimates for the average, we show that non-filing impacts effective progressivity by inflating effective ATRs differently over the income distribution.

We also contribute to a literature on effective taxation and tax progressivity. Recent studies document that rich taxpayers often face low effective ATRs (e.g., Advani, Hughson, and Summers 2023; Saez and Zucman 2019, 2020). They evade (e.g., Alstadsæter, Johannesen, Le Guern Herry, and Zucman 2022; Guyton, Langetieg, Reck, Risch, and Zucman 2021) and avoid (e.g. Roller and Schmidheiny 2016) taxes with tangible implications for inequality: When rich taxpayers pay less taxes, this weakens the effective redistribution (Alstadsæter, Johannesen, and Zucman 2019; Roller and Schmidheiny 2016). We contribute to this literature by highlighting the role of optional non-filing for effective taxation at the bottom of the income distribution. By increasing the effective

ATR for low-income taxpayers, optional non-filing acts as reverse evasion and further dampens the effective tax progressivity.

2.2 Institutional Background

Progressive Income Tax Schedule Progressive income tax schedules are characterized by higher average tax rates for higher taxable income. They are very common: In 2020, 34 out of 37 OECD countries have a progressive income tax schedule with Germany being one of them (OECD 2021a). Like many other progressive tax systems, the German income tax schedule features a basic allowance: Annual taxable income up to 8,354 € is tax free.[2] Above this basic allowance, marginal tax rates increase with income from initially 14% up to 45% for taxable income exceeding 250,730 €.

Optional Filing Many countries have non-filing options for some taxpayers.[3] This includes the US, where non-filing can be optional for low-income employees with wage earnings below their basic allowance threshold.

German taxpayers fall into two categories: compulsory filers, who have to file an income tax return, and optional filers, who are free to choose whether or not to file. Filing is optional for taxpayers for whom withholding leads to exact or too high (over-withholding). When optional filers decide to file an income tax return, we refer to them voluntary filers, whereas non-filers abstain from filing an income tax return. Whenever taxpayers earn income from a source where automatic withholding does not take place (e.g., self-employment or business income), they are required to file an income tax return. Wage and capital income do not trigger compulsory filing, as taxes on these incomes are automatically withheld at source.[4]

Additional reasons which can lead to under-withholding and therefore trigger a filing duty include receiving wage replacement benefits like unemployment insurance payments that exceed 410 €, having a second wage income subject to income taxes, or spouses electing to engage in a joint withholding scheme. Third party reporting allows as to identify these cases and drop compulsory files from our sample. We discuss the details of this in Section 2.3.

Automatic Withholding Automatic income tax withholding refers to employers withholding taxes for their employees' wages and directly transmitting it to the tax authority. It is "almost universal" (Brockmeyer and Hernandez 2019, p. 1), key for taxing income at high tax rates (Kleven, Kreiner, and Saez 2016), effective in preventing evasion (Kleven, Knudsen, Kreiner, Peterson, and Saez 2011), and correlated with economic development (Jensen 2022). German employers typically withhold income taxes for their employees on a monthly basis. Employers extrapolate the annual gross income (multiplying the monthly gross wage income with twelve) and derive a corresponding

[2] If not indicated otherwise, numbers from the German income tax code refer to 2014.

[3] List of countries with non-filing options compiled from International Bureau of Fiscal Documentation (2016): Argentina, Austria, Belarus, Bulgaria, Chile, China, Costa Rica, Croatia, Czech Republic, Dominican Republic, Ecuador, El Salvador, Estonia, Guatemala, Iran, Israel, Japan, Korea (Rep.), Lithuania, Luxembourg, Macedonia, Madagascar, Moldova, New Zealand, Nicaragua, Peru, Philippines, Ukraine, Romania, Russia, Slovak Republic, Turkey, United States.

[4] Different from all other income, capital income is subject to a flat tax rate of 25%. See Bartels and Jederny (2015) for a discussion of the German dual taxation system.

taxable income. When doing so, they take into account one twelfth of the basic allowance of 8,354 €, one twelfth of the annual standard deductions (1,000 € for work related expenses and 36 € for special expenses), and the corresponding deductible social security contributions. Then, employers withhold income taxes according to the ATR that applies to the extrapolated taxable income. Similar withholding practices are common in other countries (we discuss the case of Austria and the US in subsection 2.4.6).

Over-Withholding Over-withholding refers to effective annual withholdings exceeding the statutory tax liability that the tax schedule defines for a taxpayer's annual income. It occurs because the tax schedule is a function of annual income, but withholding takes place for each payroll period, i.e., monthly. If monthly gross wage income fluctuates, this can lead to over-withholding via two main mechanism.

First, employers consider 1/12 of the annual standard deductions for automatic withholding every month. If a taxpayer is employed for $x < 12$ months, only $x/12$ of the annual standard deductions are considered for automatic withholding, although all employees are eligible for the full 1,036 €. Likewise, only a fraction of the basic allowance (8,354 €) is considered. Second, the extrapolated and the realized annual income do not coincide if wages are not constant over twelve months. This drives a wedge between statutory and effective ATRs. The tax schedule is progressive and tax liability is a convex function of taxable income. For fluctuating monthly income, the average of the applied ATRs from the extrapolated annual income is always higher than the ATR that applies to the true annual income. As a result, over-withholding is common, while under-withholding is not possible for optional filers.

For illustration, consider this simple example: A taxpayer is employed for 3 months with a monthly gross wage income of 2,000 € and 0 income else (e.g., a graduate starting their first job in October). For each month employed, the employer extrapolates an annual gross income of $12 \times 2{,}000$ € = 24,000 € and withholds taxes at the corresponding ATR (approximately 10%). The true annual gross income is 6,000 €, which implies a statutory ATR of zero. As a result, the employer withholds $3 \times 0.1 \times 2{,}000$ € = 600 € while the statutory tax liability is 0 €. In general, over-withholding occurs when taxes are withheld for employees who do not have constant monthly wages throughout a year.[5]

Minimal Tax Filing To get a refund for over-withheld taxes, only *minimal filing* is needed. The two-page form (Figure 2.A.1b and 2.A.1c) requires some personal information and copy-pasting six values from the wage tax certificate that all employees receive automatically (comparable to form W-2 for US employees, Figure 2.A.1a). This minimal filing corrects for over-withholding by taking into account the full standard deductions and basic allowance, and applying the correct statutory ATR for the resulting taxable income. Throughout the paper, over-remittances refer to this minimal filing scenario and we abstract from further possibilities to reduce the tax liability, i.e. by itemizing deductions. We provide an upper bound estimate for additional refunds that non-filers could realize when filing and itemizing deductions in Appendix 2.G.

[5]Employers can adjust for over-withholding from monthly wage fluctuations if they employ the taxpayer for the full year. For employers with less than ten employees, this is optional. Unfortunately, according to the Federal Ministry of Finance, there is no data on the number or share of conducted end-of-year adjustments.

2.3 Data Set and Samples

Data Set We use cross-sectional, administrative income tax data on German taxpayers, provided by the Research Data Centre of the Federal Statistical Office and the statistical offices of the Länder (RDC 2018, 2020) – the *Lohn- und Einkommensteuerstatistik* (LESt). The LESt is a 10% stratified random sample of taxpayers in Germany. It covers a broad range of variables related to individual income taxation, including different sources of income, deductions, wage-replacement benefits, demographic information, and taxes paid.

Data on Non-Filers For non-filers, the data mainly stems from employer-provided end-of-year wage tax certificates, comparable to form W-2 in the US. This includes the annual gross wage income and withheld income taxes, as well as basic demographic characteristics, such as gender, age and state of residence. Based on this information, we can simulate the statutory tax rates for most non-filers and compare it to their effective taxation, the withheld income taxes. Data on non-filers is available for the two most recent LESt years, 2014 and 2010. We use the 2014 LESt for our main analysis. As a robustness check, we replicate our findings for 2010 in section 2.D. Aggregated data on tax filing status, available for a longer time range, shows that both 2010 and 2014 are typical years in terms of tax filing (see Figure 2.A.2). For filers, the LESt also includes data from their tax filing forms and final tax assessment.[6]

Identifying Optional (Non-)Filers Our analysis focuses on optional (non-) filers, i.e., taxpayers who are not required to file an income tax return. Filing is optional when income taxes are withheld at source and under-withholding is ruled out. This is typically the case if taxpayers have only wage income and potentially capital income, receive no wage replacement benefits, do not have several jobs at a time, and do not opt for joint withholding with their spouse (see section 2.2 for details). We can identify these taxpayers because the LESt data covers relevant information to determine whether filing is optional or compulsory.

The LESt data reports all income sources that trigger compulsory filing, such as self-employed income. Information on capital income is limited in the LESt data, but this is no concern for the identification of optional non-filers, because capital income does not trigger compulsory filing.[7] A false classification of taxpayers with additional income sources as optional filers is only possible if they evade taxes by not reporting this additional income at the extensive margin, e.g., hiding all self-employed income. Although administrative data does not allow to rule out that these cases exist, we argue that such extensive margin tax evasion is unlikely to drive our results (see Appendix 2.B for details). Information on wage-replacement benefits are shared between social insurance institutions and tax offices and reported in the LESt data. Spouses with

[6] A taxpayer is recorded as a filer if their tax assessment is concluded within 2 years and 9 months after the tax year, i.e. September 30, 2016 for the tax year 2014. Optional filers have up to four years to file, which is a potential confounder to our results if such late filing is particularly common among low income optional filers with over-remittances. There is no data on the frequency of late filing behavior, but anecdotal evidence from practitioners suggests that the number of late filers is negligible.

[7] Capital income is taxed with a flat rate of 25% and taxes are automatically withheld at source. Taxpayers are typically not required to report capital income for their income taxation because of this automatic withholding.

joint withholding and taxpayers with multiple jobs can be identified based on their withholding scheme (*Steuerklasse*).

Sample 1 For our analyses, we use different samples of filers and non-filers drawn from the LESt data. To study the prevalence of optional tax non-filing in subsection 2.4.1, we first restrict our sample to optional filers and exclude all compulsory filers (Sample 1). This sample contains 258,139 non-filers. When applying statistical weights, they represent 9 million German taxpayers. We provide descriptive statistics and additional details on the sample restrictions in Appendix 2.B.

Sample 2 To be able to simulate the statutory taxation, we further restrict our sample in subsection 2.4.3 and 2.4.4 to optional filers between 16 and 63 years of age (working-age population) for whom we can impute their taxable income based on the employer-provided information (Sample 2). This excludes civil servants and employees with wage income of more than 48,600 € (about the 90$^{\text{th}}$ percentile of the of optional filers' income distribution). For these taxpayers, tax-deductible social insurance contributions cannot be inferred based on their gross wage income, because they do not face compulsory enrollment. We also exclude taxpayers for whom the withheld taxes are not in line with tax-determining individual characteristics reported in the LESt data.[8] Sample 2 contains 205,678 non-filers, representing 8 million taxpayers.

2.4 Results on Non-Filing, Over-Remittances, and Effective ATRs

2.4.1 Prevalence of Non-Filing

Non-filing is common and there is a clear correlation between non-filing and annual income. Among optional filers, who can choose whether or not to file, 61% do not file. These non-filers have an average annual gross wage income of about 18,000 €, compared to 35,000 € for voluntary filers (Table 2.B.1). Figure 2.1 shows that the non-filing share decreases from 90% for gross wage income of around 10,000 € to around 30% for 50,000 € and higher.[9] While the non-filing share varies significantly over the income distribution, we document that it is remarkably persistent across age groups, gender, family status, and region (see column 2 of Table 2.A.2).

[8] This excludes taxpayers for whom withheld taxes are "too high" as compared to the statutory tax liability computed based on their individual characteristics. While there cannot be under-withholding for optional filers (see section 2.2), these deviations may occur because individual characteristics that determine the tax liability can change throughout the year, but the data contains only end-of-year information. We discuss this in more detail in section 2.B and show that over-remittances are still substantial when including those cases.

[9] We analyze non-filing over gross income instead of taxable income, since the latter is endogenous to tax filing.

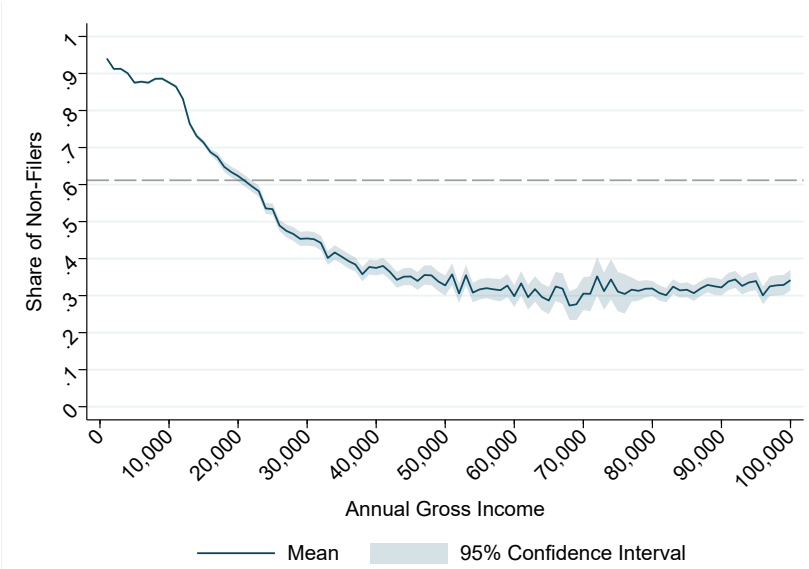


Figure 2.1: Prevalence of Non-Filing over Gross Income

Notes: Share of non-filers among optional filers over annual gross wage income. Dashed grey line: Average share of non-filers (61.15%) across all income levels. Statistics refer to taxable units, which may be either an individual or married spouses in case of joint filing. For jointly filing spouses, we consider the average gross income.

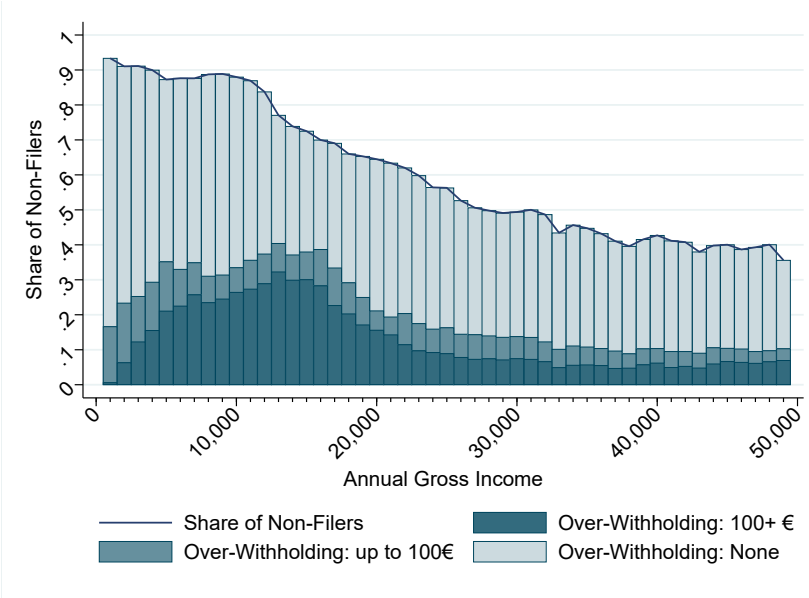


Figure 2.2: Effective Taxation of Non-Filers

Notes: This figure shows the non-filing share (dark blue line, same as in Figure 2.1), decomposed in three subgroups of non-filers (stacked bars). For a given 1,000-€-bin of annual gross wage income, the figure plots the share of non-filers, who over-remit no income taxes (light blue bars), up to 100 € (medium blue bars), and more than 100 € (dark blue bars). None: No over-withholding because taxes are withheld correctly. This 0 € threshold is allowing for a 5-€-tolerance, i.e. including all non-filers with an estimated over-remittance of 0 +/- 5 €. Reading example: 88% of optional filers with an annual gross wage income of 10,000 € are non-filers. 54% face no over-remittance, 7% over-remit up to 100 € and 26% over-remit more than 100 €.

2.4.2 Prevalence of Over-Withholding

After documenting non-filing for large parts of the taxpayer population, we now study the prevalence of over-withholding among those who do not file, to rule out that the average is driven by a small number of non-filers with extreme values of over-remittances. Under over-withholding, $T_i^E > T^S(y_i)$, where T_i^E denotes income taxes effectively withheld for taxpayer i and $T^S(y_i)$ denotes the statutory taxes that the tax schedule defines for their taxable income y_i. We observe T_i^E for non-filers in the LESt data and follow the tax code to compute y_i and $T^S(y_i)$. Starting from the annual gross income, we simulate the statutory income taxation taking into account standard deductions, social insurance contributions, and special allowances if applicable. We restrict the sample to non-filers for whom the taxable income can be inferred from their gross income (sample 2). Appendix 2.B discusses the sample selection and Appendix 2.C explains the income tax schedule simulation. We then compute tax over-remittances as $T_i^E - T^S(y_i)$ and study their prevalence for non-filers.

First, we document that tax over-withholding is common and affects about one third of the non-filing taxpayers. This share is stable across different demographic groups (see column 5 of Table 2.A.2). In absolute terms, applying statistical weights provided by the RDC (2020), non-filers with over-withholdings represent 2.6 million German taxpayers.[10] For the other two thirds of the non-filer population, withholding is exact. They may be eligible for a tax refund if they have tax deductible expenses that exceed the standard deductions and can be itemized. This information is not observable for non-filers, but we provide an upper bound estimate based on filers' deductions (see Appendix 2.G).

Second, we show that over-remittances are not only common among non-filers, they are also often non-negligible in size. For the 2.6 million non-filers with over-remittances, the average tax over-remittance amounts to 361 € with a median of 183 € (see Table 2.A.1 for more percentile estimates).

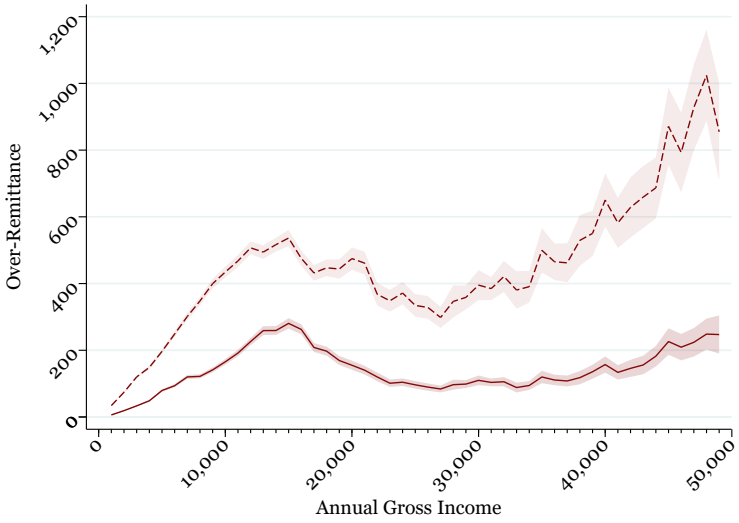
Third, we document that low-income taxpayers are not only most likely to be non-filers, they are also most likely to overpay non-negligible amounts. To show this, we decompose the group of non-filers in 3 subgroups: non-filers with no over-remittances, with small over-remittances of up to 100 €, and with larger over-remittances of more than 100 €. Figure 2.2 plots the decomposed non-filing share over the income distribution. Over-remittances of more than 100 € are most common for low levels of annual gross income.

2.4.3 Tax Over-Remittance Through Non-Filing

After documenting that over-withholding is common, we now analyze the amount and distribution of tax over-remittances in more detail. Figure 2.3 plots the average tax over-remittance for non-filers over the income distribution. We show that non-filers pay too much income taxes at all income levels. Averaging over all non-filers, including those with no over-withholding, the mean over-remittance is 119 € (solid red line in Figure 2.3a). In relative terms, this corresponds to 1.2% of the annual income for the average non-filer. Figure 2.3b shows that relative over-remittances are highest for lowest

[10] The cross-sectional data does not inform about how frequently individuals over-remit taxes under non-filing, but we can use the repeated cross-section to provide suggestive evidence that it is not a once-in-a-lifetime phenomenon (see Hauck and Wallossek (2021) for a discussion).

(a) Absolute Over-Remittances



(b) Relative Over-Remittances

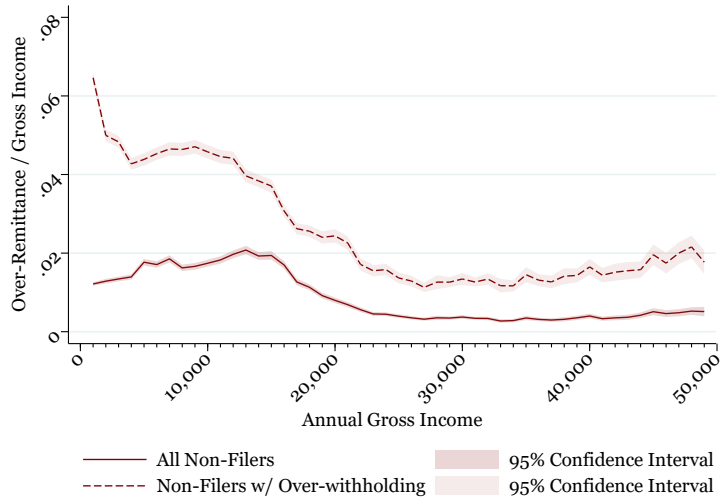


Figure 2.3: Tax Over-Remittances of Non-Filers

Notes: This figure shows the average tax over-remittances for non-filers over annual gross wage income (1,000-€-bins). Panel (a) plots absolute over-remittances in €, panel (b) plots the over-remittances in relative terms, as a share of annual gross income, respectively. The solid red lines refer to all non-filers, the dashed red lines to the subpopulation of non-filers with tax over-withholding, excluding those for whom statutory and effective taxation are aligned. We define tax over-remittances as the amount of income taxes that non-filers pay in excess to the statutory income taxes defined in the tax schedule. A value of 0 indicates no over-remittances, i.e. effective and statutory taxation coincide. Strictly positive values indicate that, on average, non-filers over-remit taxes at all income levels. Figure 2.4 plots the resulting effective ATR. Reading example: Non-filers in the 10,000 € income bin pay 166 € more in taxes than intended by the income tax schedule. Conditional on over-withholding, non-filers in this income bin over-remit 435 €. In relative terms, this is equivalent to 1.7% and 4.7% of their annual gross wage income.

income, where the average over-remittance is close to 2% in many income bins.

For many non-filers, the withholding is exact (see Figure 2.2). When excluding those and conditioning on over-withholding, i.e., taxpayers who leave money on the table by not filing, the average tax over-remittance is 361 € (dashed red line in Figure 2.3a). In terms of economic magnitude, this corresponds to approximately one month of social welfare (2014 baseline level for singles: 391 €).

In total, non-filers over-remit at least 951 million € in 2014 and lowest income taxpayers bear significant parts of this. One third of all over-remittances originates from non-filers with annual gross income below the basic allowance threshold. They would receive a full tax refund for all income taxes withheld when filing a tax return.

Our results are only a lower bound for the forgone refund potential of non-filers. The quantified over-remittances are the refunds that follow minimal filing with standard deductions, but itemizing deductions can further increase refunds. We provide an upper bound for the refund potential when considering additional deductions in Appendix 2.G. Furthermore, the data does not allow for quantifying over-withholding of additional surtaxes and potential joint-filing benefits for married non-filers.

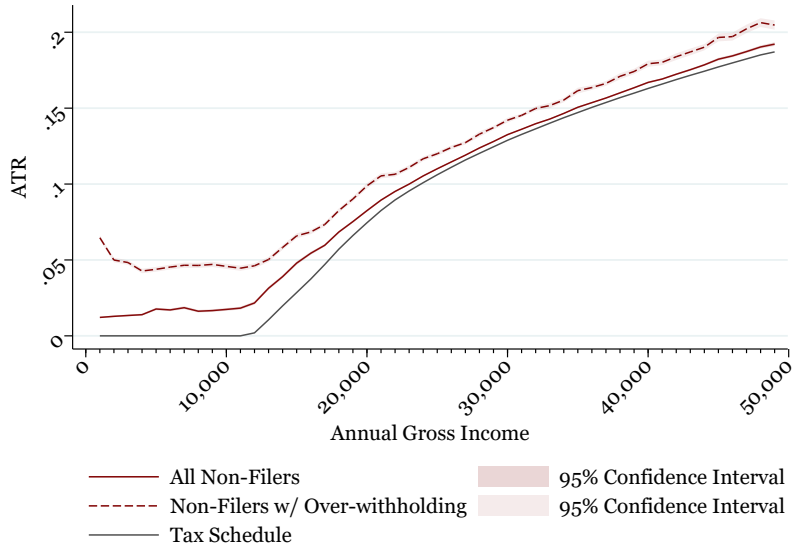


Figure 2.4: Effective vs Statutory ATR for Non-Filers

Notes: This figure shows the statutory (black line) and effective (red lines) average tax rate (ATR) of non-filers over annual gross wage income (1,000-€-bins). The solid red line refers to all non-filers, the dashed red line to the subpopulation of non-filers with tax over-withholding. The black line shows the statutory ATR as defined by the income tax schedule. The deviation between effective and statutory ATR is the additional ATR that non-filers face (see Figure 2.3b). This is the relative value of the absolute over-remittances shown in Figure 2.3, with the tax schedule here being the equivalent to the 0 intercept in Figure 2.3. Reading example: Non-filers in the 10,000 € income bin face an effective ATR of about 1.7% as compared to a statutory ATR of 0%. Conditional on over-withholding, the average effective ATR is 4.6% in this income bin.

2.4.4 Effective ATR Under Non-Filing

The over-remittances of non-filers drive a sizable wedge between their effective ATR (ATR^E) and their statutory ATR as defined in the tax schedule (ATR^S). Effectively, there are two tax schedules in place, one for filers and one for non-filers. Figure 2.4 shows that the latter features higher ATRs: Non-filers face $ATR^E > ATR^S$ at all income levels, with an average deviation of 1.2 percentage points.

The deviation is highest for lowest income non-filers, which weakens the effective redistribution and income tax progressivity. For low income levels up to the basic allowance threshold, the tax schedule intends $ATR^S = 0$. Despite not being liable to pay any income tax, the average non-filer in this income range faces an ATR^E of about 1.5% (solid red line). This average includes non-filers with exact withholding, i.e., $ATR^E = ATR^S = 0$. They typically have no incentives for filing, even if they had tax deductible expenses, since there is nothing to refund. If we exclude them and restrict the sample to non-filers with over-withholding, the average ATR^E is 4.7% for non-filers with $ATR^S = 0$ (dashed red line).

From an equity perspective, non-filing acts as reverse evasion with a qualitatively similar impact on effective taxation: higher effective ATRs for low-income non-filers and lower effective ATRs for rich evaders both weaken the effective tax progressivity.

2.4.5 The Role of Filing Costs

While the income tax return data does not allow for a clear identification of the mechanisms behind the observed non-filing behavior, we can exploit heterogeneity in forgone refunds to provide suggestive evidence for the limited role of filing costs. Figure 2.5 plots the share of non-filers for the distribution of over-withholdings. Two patterns emerge. First, non-filing is indeed less common for higher over-withholdings. Second, albeit this negative correlation, non-filing is strikingly persistent across all levels of over-withholdings: whether taxpayers over-remit 200 € or 2,000 €, about 50% – 70% do not claim their refunds (Figure 2.5a). The pattern is similar when expressing the (forgone) refunds relative to income, as additional ATR under non-filing (Figure 2.5b). We take this limited responsiveness as suggestive evidence that filing costs are not the main driver for the observed non-filing. Intuitively, it seems unlikely that filing a two page form (see Figure 2.A.1 for the form) comes at costs of hundreds of Euros or more.

2.4.6 Over-Withholding for Non-Filers in Other Countries

We build our analysis on German administrative data that allows for quantifying individual tax over-remittances for non-filers. However, the implications of tax non-filing for effective taxation are not limited to Germany. To illustrate this, we discuss effective taxation under automatic withholding with optional non-filing for two countries: Austria and the United States. We show that in both countries, taxpayers for whom tax filing is optional can be subject to over-withholding. Appendix 2.E discusses the underlying computations in more detail.

Austria The Austrian income tax system is similar to the German setting, with optional filing under automatic withholding for taxpayers with wage income only. Similar to the German withholding system, there can be over-withholding, but no under-

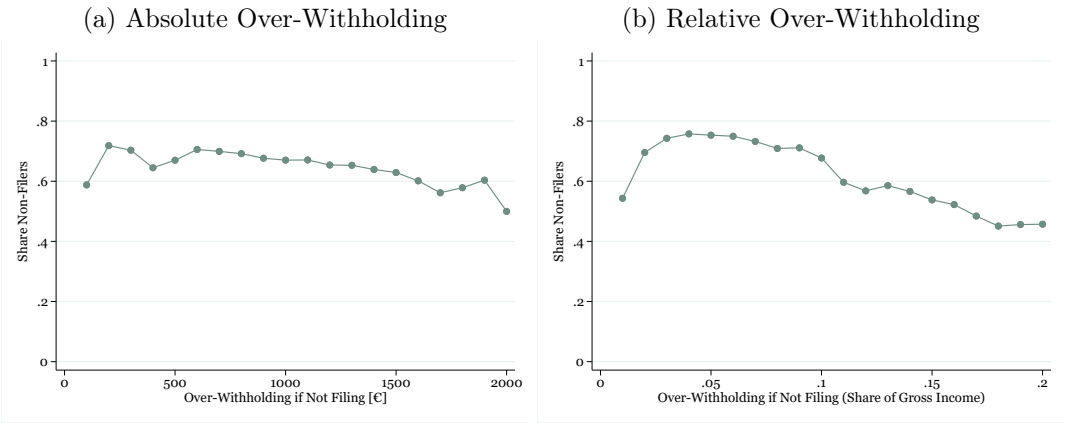


Figure 2.5: Non-Filing Share by Potential Over-Withholding

Notes: This figure shows the non-filing share over potential over-withholding that taxpayers face if not filing (2014, optional filers, weighted data). Panel A plots the non-filing share over potential over-withholding in absolute terms (100-€-bins). Panel B plots the non-filing share over potential over-withholding in relative terms, as share of annual gross income (1-%-bins). This is equivalent to the additional ATR that taxpayers pay because of over-withholding (i.e., the wedge between ATR^E and ATR^S). *Reading example*: 67% of optional filers with an annual over-remittance of 1,000 € do not file an income tax return (Panel A). Optional filers whose ATR^E is 10 percentage points higher than their ATR^S do not file an income tax return in 68% of the cases (Panel B).

withholding for optional filers. Using aggregated administrative Austrian income tax data, we show that low-income employees with $ATR^S = 0$ face over-withholdings that are qualitatively similar to over-withholdings in Germany (we provide details in Appendix 2.E).

In 2017 Austria implemented automatic refunds for non-filers (see Austrian Federal Ministry of Finance (2022) for details). Since this reform, over-withholdings accrue only temporary and are fully refunded after the end of a tax year – realigning ATR^E and ATR^S. The average automatic refund to non-filers in 2017 is 238 € (Austrian Federal Ministry of Finance 2018).

United States The US withholding system differs from the German and Austrian setting. Under automatic wage tax withholding, US taxpayers have substantial discretion over their withholdings. As a result, US taxpayers are typically required to file an income tax return, since there can be both over- and under-withholding, depending on the chosen withholding (see, e.g., Jones 2012). However, tax filing is optional for some US taxpayers: Low-income employees with annual gross income below the applicable standard deduction are typically not required to file. With $ATR^S = 0$, they cannot face under-withholding, while over-withholding can occur when employment is not constant throughout the year – similar to the German setting. Goodman, Lim, Sacerdote, and Whitten (2023) estimate that there is a total of 47 million US non-filers with no filing obligation.

Using the 2014 IRS withholding tables for employers (IRS, 2013), we can simulate monthly withholdings for US non-filers as a function of annual gross income and months employed. While different withholding systems, we show that implications of optional filing are qualitatively similar for lowest-income non-filers in the US and Germany:

although they should pay zero income taxes they can face substantial positive ATR^E when not employed for full 12 months (see Figure 2.E.2 for details). This is in line with findings for the US from Goodman, Lim, Sacerdote, and Whitten (2023). Simulating pre-populated forms for 2019, they estimate that the average optional non-filer over-remits $36 in income taxes. Conditional on over-withholding, they report an average income tax refund potential of $390 = 293€.[11] In addition, US non-filers often forgo further refunds that they are eligible for, such as the EITC (see e.g., Goldin, Homonoff, Javaid, and Schafer 2022; Ramnath and Tong 2017).

2.5 Effective Taxation and Tax Revenue Efficiency

Under optional non-filing, taxpayers pay more taxes than intended by the statutory tax schedule and the government raises more tax revenue. Leaving aside equity concerns, this raises questions of efficiency. Is non-filing an efficient way of reaching the observed effective tax schedule? That is, how much tax revenue is raised when increasing effective tax rates compared to increasing statutory tax rates? To answer this, we benchmark the status quo taxation against two hypothetical tax schedules. These benchmark schedules are designed such that they map the effective marginal tax rates MTR^E (benchmark I) and effective average tax rates ATR^E (benchmark II), as observed in the optional non-filing status quo.

The goal of these benchmarks is to provide the intuition for why increasing effective tax rates under non-filing raises efficiency concerns: Increasing effective tax rates ATR^E and MTR^E under non-filing does not realize the full tax revenue potential, compared to a similar increase in statutory tax rates ATR^S and MTR^S, respectively. The focus for this exercise is on tax revenue. We make no statements on welfare and the benchmark scenarios do not constitute Pareto improvements. Rather, they highlight unexploited tax revenue potential when increasing effective rather than statutory tax rates. For quantifying the unexploited revenue, we rely on a set of simplifying assumptions. When computing behavioral responses, we assume that taxpayers' responsiveness is the same for changes in MTR^S and MTR^E. In addition, we abstract from income effects. For our computations, we use a sample that includes taxpayers irrespective of their filing status (see Appendix 2.B for details on the sample selection and Appendix 2.H for details on the underlying computations).

Effective vs. Statutory Tax Changes In terms of tax revenue, the key difference between increasing the effective vs. statutory MTR is in who is affected by the increase. An increase in MTR^S for a given income range $[y^*; y^*+dy^*]$ raises more tax revenue from individuals with $y^* \leq y_i$ (mechanical effect). Individuals with income $y^* \leq y_i \leq y^*+dy^*$ reduce their taxable income in response to a higher MTR^S. The size of this behavioral effect depends on the elasticity of taxable income, and may stem from labor supply responses or tax avoidance (e.g., Doerrenberg, Peichl, and Siegloch 2017; Neisser 2021; Saez 2001). When abstracting from income effects, individuals with $y^* + dy^* < y_i$ do not respond to the tax change, because they only experience an increase in their ATR^S, but not in their MTR^S (Saez 2001, illustrated in Figure 2.6a).

An increase in MTR^E for the same income range affects only individuals with $y^* \leq$

[11] Conversion to 2019 Euros using purchasing power parities from OECD (2022a).

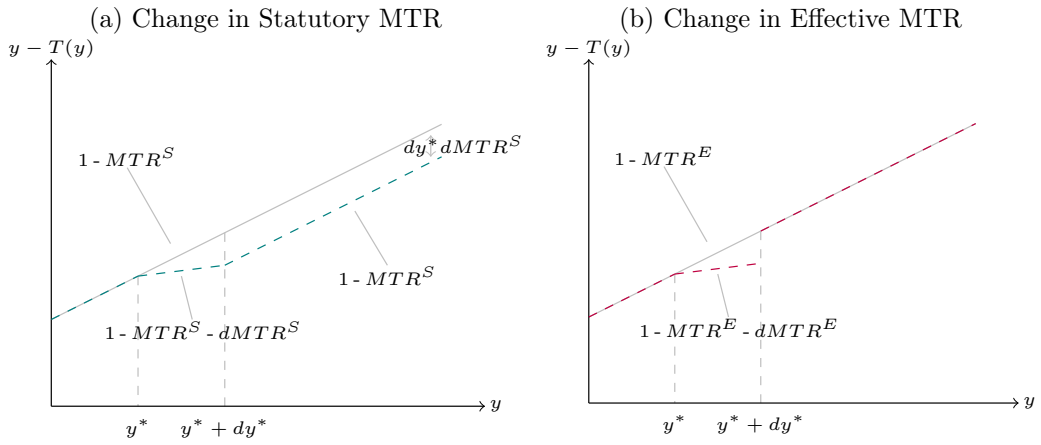


Figure 2.6: Tax Perturbations

Notes: Visualization of the effect of increasing the marginal tax rate for the income range $[y^*; y^* + dy^*]$ by $dMTR > 0$. y: taxable income; $T(y)$: taxes paid on taxable income; $y - (Ty)$: net-of-tax income. Slopes are indicated in the graphs. Panel (a) refers to an increase in the statutory MTR (MTR^S) as defined in the tax schedule and follows closely Saez (2001). Panel (b) refers to an increase in the effective MTR (MTR^E), leaving the tax schedule unchanged. This is a simplified illustration of the MTR^E in the non-filing status quo, where increased MTR^E only affect non-filers with no effects on other, higher-income taxpayers. It is simplified in the sense that non-filing increases MTR^E not only in a strictly defined income range (see Figure 2.H.1a for observed MTR^E).

$y_i \leq y^* + dy^*$ (Figure 2.6b). For the same increase in MTR, the mechanical effect is thus smaller for MTR^E compared to MTR^S, because there is no effect for those with $y^* + dy^* \leq y_i$. At the same time, the distortionary behavioral effect is of the same size when assuming that the elasticity is the same with respect to MTR^S and MTR^E. In our setting, the increase in MTR^E affects only non-filers with no mechanical effect on the tax revenue raised from other taxpayers.

Benchmark I: "Unexploited Tax Revenue from Mechanical Effect" This benchmark demonstrates that increasing effective marginal tax rates via non-filing raises less tax revenue than a tax schedule that maps the observed MTR^E, because the potential for mechanically raising tax revenue is not fully exploited. We illustrate this by focusing on MTRs in the first income tax bracket. The average MTR^E in this bracket is $\overline{MTR_1^E} = 5.43\%$, while $MTR_1^S = 0$.[12]

We compare this against a hypothetical tax schedule that reaches the same $\overline{MTR_1^E}$ via the tax code, with $MTR_1^{S\prime} = MTR_1^{E\prime} = 5.43\%$, while keeping MTR^E unchanged for higher income tax brackets (Figure 2.H.1b plots the resulting MTRs). Without increasing the average MTR^E at any income level, i.e., keeping average effective distortions unchanged, this benchmark mechanically raises 12.9 billion € more in tax revenue than the non-filing status quo. This benchmark is a simplified illustration that concentrates solely on the first tax bracket. Additionally, we abstract from changes in the individual

[12] $MTR^E > ATR^E$ because the minimum positive MTR in the German income tax code is 0.14. As soon as taxes are withheld, the MTR is ≥ 0.14, while the ATR will always be lower than the MTR.

MTR^E in the first bracket, which would increase for filers and non-filers with no over-withholdings, and decrease for non-filers with over-withholdings. With this benchmark tax schedule, most taxpayers would pay more taxes than in the status quo. We abstract from income effects and do not draw implications for social welfare, which would depend on the specific welfare function and how the additional tax revenue is redistributed.

Benchmark II: "Unexploited Tax Revenue from Behavioral Responses" This benchmark demonstrates that increasing effective average tax rates via non-filing raises less tax revenue than a tax schedule that maps the observed ATR^E. This is because the benchmark schedule can reach the same ATR with less distortions, i.e., lower MTR, yielding additional tax revenue potential from behavioral responses. We define the benchmark schedule $T''(y)$ as the function that maps best the observed effective taxation (see section 2.H for details) and derive the corresponding $MTR^{E\prime\prime} = MTR^{S\prime\prime}$ for this schedule. Figure 2.H.1c shows that, without increasing effective average tax rates for a given income bin, this tax schedule has lower effective MTRs ($MTR^{E\prime\prime} < MTR^E$). Intuitively, increased ATR for lower income level raise tax revenue from higher income taxpayers as well, which allows to decrease their MTR.

We then compute behavioral responses to the change in MTR^E on the individual level. Assuming an elasticity of taxable income with respect to the net-of-tax rate of $\varepsilon_{y,1-MTR} = 0.2$, for statutory and effective MTR, we find that tax revenue increases by 297 million €. Most of this effect comes from filers' responses alone: tax revenue increases by 193 million € even when assuming that non-filers do not respond to tax incentives at all. Although we keep the average $ATR^E(y)$ constant, this is no Pareto-improvement: for given y, filers and non-filers without over-withholding face higher a ATR^E than in the status quo, while non-filers with over-withholding face a lower ATR^E.

Taken together, the two benchmark scenarios illustrate that increasing effective taxation via non-filing is not the most efficient way for generating tax revenue. Comparable increases in statutory taxation could either generate more revenue for the same average distortion (benchmark I) or the same revenue in a less distortive way (benchmark II).

2.6 Policy Implications

If the statutory tax schedule maps the intended degree of redistribution, the regressive effects from non-filing on the effective taxation can be considered as unintended redistribution. The coherent policy implication in this case is to automatically refund over-remittances for optional filers to realign statutory and effective taxation. This benefits non-filers without imposing any costs on them. Intuitively, automatic refunds generate equity gains without efficiency losses: They generate welfare gains for low-income non-filers by reducing their effective ATR while at the same time, lower effective MTRs create labor supply incentives. Aligning effective and statutory taxation increases effective progressivity and also strengthens horizontal equity between filers and non-filers at a given income level. It should be noted however, that the system of automatic refunds is only applicable in cases of optional filing where exact withholding or over-withholding takes place. Whenever taxpayers have discretion about the degree of withholding, i.e. under-withholding is possible, automatic refunds are no viable policy recommendation.

Importantly, non-filers benefit from automatic refunds irrespective of the reason for their non-filing. Particularly, because automatic refunds do not require any action from non-filers, potential filing costs will not dampen the progressive effect of automatic refunds for low-income taxpayers. Furthermore, automatic refunds also benefit voluntary filers who no longer need to file an income tax return if they only want to correct for over-withholding.

While non-filers benefit from automatic refunds, drawbacks for other groups of taxpayers and the government are limited. As concerns voluntary filers, automatic refunds can be combined with an option to file an additional tax return if taxpayers want to itemize deductions (as is done in Austria for instance). This allows voluntary filers to realize the same refunds as in the status quo. Assuming that filing requirements are reduced with automatic refunds, because more information is automatically provided by tax authorities, voluntary filers would additionally face lower filing costs. At the same time, automatic refunds reduce the benefit from voluntary filing since over-withholdings are automatically refunded. It is thus unclear, ex-ante, whether there would be more or less voluntary filers under automatic refunds. For compulsory filers, automatic refunds for non-filers do not affect their filing or alter their absolute filing costs or benefits.

With respect to tax authorities and the more broader government, automatic refunds are unlikely to jeopardize the governmental budget: Over-remittances from non-filing taxpayers constitute only about 0.15% of Germany's 644 billion € tax revenue in 2014 (Federal Statistical Office 2022a). Otherwise, the government could for example increase the top MTR to offset the loss in tax revenue. We provide two back-of-the-envelope calculations for such budget neutral reforms in Appendix 2.F. The additional costs for tax authorities associated with refunding over-remittances should be minimal. Information relevant to the computation is readily available and of high quality, as it is provided by third parties. Therefore, tax authorities simply need bank account information for each taxpayer to automatically transfer refunds.

One problem that automatic refunds cannot solve is temporary over-withholding throughout the year, i.e., before over-remitted taxes are refunded. Such temporary over-withholdings can still have negative implications, particularly for low-income taxpayers who are more likely to face liquidity constraints (Caldwell, Nelson, and Waldinger 2023; Jones 2012).

If automatic refunds are not feasible due to administrative or legal constraints, tax authorities can automatically send out pre-populated forms to optional filers. Pre-populated forms increase the salience of over-withholding and reduce the costs associated with filing (see e.g., Benzarti 2021; Goodman, Lim, Sacerdote, and Whitten 2023). The automatic provision of pre-populated forms is popular and used in countries worldwide (OECD 2021b).[13]

From a more general perspective, our results highlight the importance of non-filing for the effective reach of tax and transfer policies: Policy makers should take non-filing behavior into account when designing tax policy. This is not only relevant for income taxes, but for any policy that is administered via the income tax code and only granted

[13]In Germany, such forms are available only upon request (European Commission 2019). However, the German government has agreed to introduce pre-populated forms more broadly as per their 2021 coalition treaty (Bundesregierung 2021).

upon filing. This includes social welfare or income support programs such as the EITC or economic relief payments to increase taxpayers' available income – a common policy in the times of (economic) crises.

2.7 Conclusion

We show that, while seemingly innocuous, optional tax non-filing for employees has a sizable impact on effective taxation. Under optional non-filing, lowest income earners are most likely not to file, while at the same time often being subject to over-withholding. Non-filing thus harms mostly those with lowest income who have substantially higher effective average tax rates than intended by the tax schedule.

So far, the deviation between effective and statutory taxation and its implications for progressivity have been studied mostly in the light of tax evasion of rich taxpayers. We add a new perspective to this by highlighting the role of optional non-filing. Non-filing is fundamentally different from evasion: low income taxpayers face legal tax over-remittances because of their passive behavior. However, the result is qualitatively the same: Both non-filing and evasion weaken the effective tax progressivity. In this sense, non-filing acts like "reverse evasion".

Appendix to Chapter 2

2.A Additional Tables and Figures

		(A) All Non-Filers		(B) With Over-Withholding	
		(1) All	(2) y < threshold	(3) All	(4) y < threshold
Over-Remittance	total	949,512,506	317,948,477	951,685,618	317,774,299
	mean	118.66	79.41	360.04	247.40
	p25	-0.40	0.00	42.00	45.00
	p50	0.00	0.00	182.90	137.00
	p75	40.09	38.00	518.60	348.00
	p90	425.28	277.00	959.56	639.00
	p95	760.16	498.00	1,265.00	846.00
N	raw	205,678	103,761	68,138	33,350
	weighted	8,001,646	4,004,070	2,643,277	1,284,469

Notes: Over-remittances are defined as $T_i^E - T^S(y_i)$ for individual i and listed in €. (A): All non-filers in the sample. (B): Only those non-filers with over-remittance, defined as a deviation of more than 5 € from the statutory tax schedule. The total over-remittance in (B) is smaller than for (A), since (A) includes a +/- 5 € tolerance around 0, and the (small) negative values decrease the total amount. $y < threshold$: Individuals with an annual gross wage income below the basic tax allowance threshold. Table 2.A.1 shows the aggregated numbers for over-remitted taxes. On average, non-filers over-remit 118 €, including those for whom withholding is correct. Conditional on over-withholding, the average over-remittance is 360 €. In total, this leads to an overall sum of over-remitted taxes of at least 949 million €. Columns 2 and 4 of Table 2.A.1 show that one third of the total over-remittances originate from taxpayers at the lower end of the income distribution, below the basic allowance threshold. Taxpayers in this income range over-remit at least 317 million € and account for 33.5% of all over-remittances in our sample, even though their statutory tax payment as is zero.

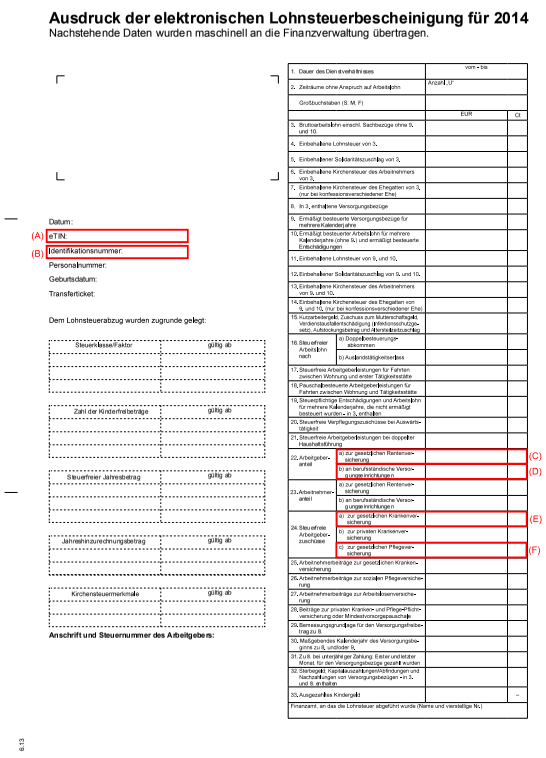
Table 2.A.1: Taxes Over-Remitted Through Non-Filing - Lower Bound Estimates

	(A) Sample 1		(B) Sample 2			
	(1) N	(2) NF/N	(3) N	(4) NF/N	(5) OR/NF	(6) OR/N
all	683,718	0.61	424,147	0.68	0.33	0.22
female	278,817	0.64	206,898	0.66	0.29	0.19
male	373,891	0.65	217,249	0.69	0.37	0.26
children	158,498	0.53	81,728	0.65	0.36	0.23
no children	525,220	0.63	342,419	0.68	0.32	0.22
east	166,113	0.66	121,878	0.74	0.34	0.25
west	517,605	0.60	302,269	0.66	0.33	0.22
age: <20	47,082	0.90	41,552	0.90	0.17	0.15
age: <30	197,376	0.65	186,624	0.67	0.37	0.25
age: <40	91,633	0.56	78,004	0.62	0.38	0.24
age: <50	78,888	0.57	63,696	0.64	0.33	0.21
age: <60	59,721	0.55	45,764	0.65	0.31	0.20
age: 60+	13,101	0.69	8,507	0.71	0.29	0.21

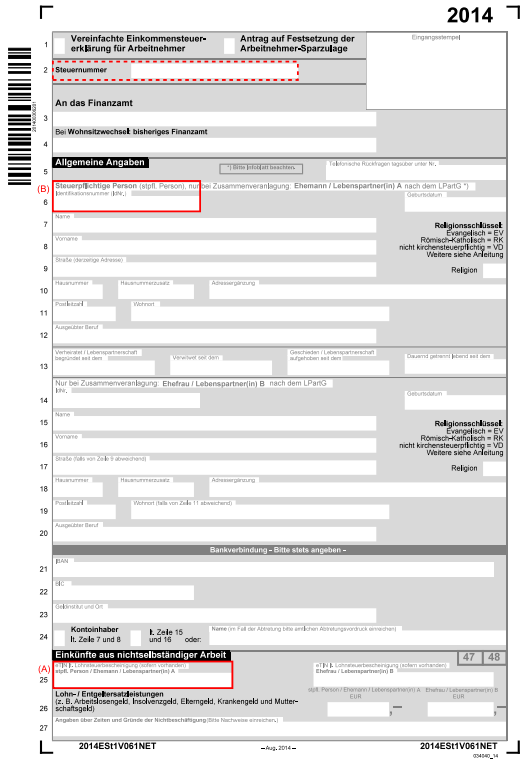
N = sample size optional filers, NF = non-filers, OR = non-filers with over-remittances.
This table shows the prevalence of non-filing and tax over-remittance for different demographic groups. Panel (A) refers to sample 1 with all optional filers, panel (B) refers to sample 2, for which we can compute over-withholdings (see section 2.B for details on the sample restrictions). Columns (1) and (3) indicate the number of taxpayers within each sample. Columns (2) and (4) show the share of non-filers among these taxpayers. Column (5) shows the share of non-filers that overpay taxes. Column (6) shows the share of taxpayers that are non-filers and overpay taxes. This is equivalent to multiplying columns (4) and (5). Means are weighted, sample sizes are unweighted. East/west refers to Eastern vs Western German states. For joint filing spouses in sample 1 we report their mean age and do not include them in the female/male subsamples. For data privacy protection, the income tax data does not provide age information for high-income taxpayers, which is why the age subsamples in sample 1 do not add up to 100%. This is no concern for sample 2, where we do not include high-income taxpayers. *Reading example:* 66% of women in sample 2 are non-filers. 29% of those non-filing women overpay taxes. In total, this implies that 19% of all women in the sample are non-filers who overpay taxes.

Table 2.A.2: Non-Filing and Over-Withholding by Demographic Groups

(a) Yearly Tax Information



(b) Filing Form I



(c) Filing Form II

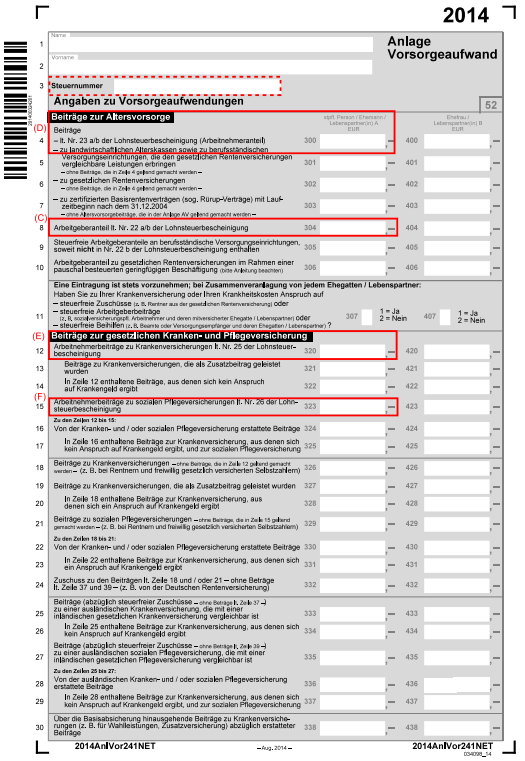


Figure 2.A.1: Documents for Minimal Filing

Notes: (a): Blank example of the yearly tax information that employers provide automatically (similar to W2 in the US). (b) and (c): Relevant forms for minimal filing for 2014. Downloaded from urlhttps://lstn.niedersachsen.de/download/81213 (a), https://www.formulare-bfinv.de/printout/034040_14.pdf (b) and https://www.formulare-bfinv.de/printout/034098_14.pdf (c); last accessed: 2021-08-31; red cells added manually.
In order to correct for over-withholding, taxpayers have to fill in the cells that are marked red in (b) and (c). For the solid red lines, required information can be copied from the employer-provided form (a), letters indicate correspondent cells. (A): Electronic tax identification number. (B): Identification number. (C) & (D): Public pension insurance. (E): Public health insurance. (F): Public nursing care insurance.
Additionally, taxpayers have to indicate personal details (name, address, bank account etc.) as well as the responsible tax authority (lines 4 – 24 in the left panel). The tax number (dashed line) is provided by the local financial authority or left blank when filing for the first time. Additional fields in (b) are for further details on contributions to insurances (lines 4 – 10: pension insurance, lines 11 – 30: health insurance) that are typically not relevant for minimal filing. The right column in (b) is used for joint-filing spouses, again not relevant for minimal filing.

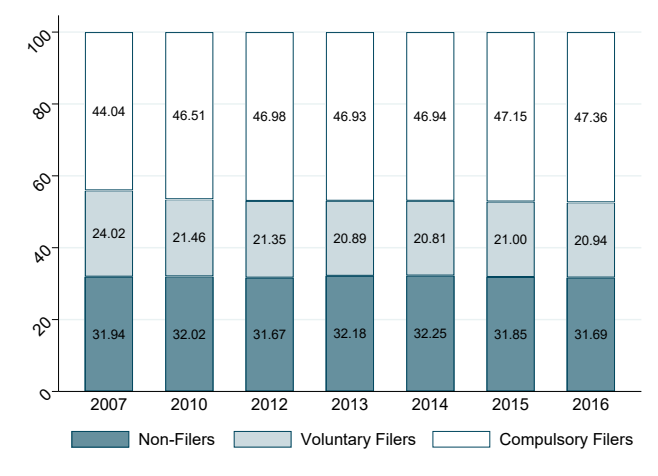


Figure 2.A.2: Filing Status Over Time – Aggregated Data

Notes: Data provided by the German Federal Statistical Office upon request. The data set is the same as the one used in the empirical analysis in this paper (Lohn- und Einkommensteuerstatistik). Prior to 2012, this statistic was published every three years. Since 2012, the statistic is published annually. Differences to 100% are due to rounding.

2.B Sample Restrictions

Sample 1. This sample contains optional filers, i.e. voluntary filers and non-filers. We use this sample to study the prevalence of non-filing over the income distribution in subsection 2.4.1. Table 2.B.1 provides descriptive statistics for the sample. The sample contains 683,718 optional filers with 258,139 being non-filers. One observation represents one taxable unit, which can be either a single tax payer or, for filers, jointly filing spouses. We report descriptives for the weighted sample, using the weights provided by the statistical office.

		Optional Filers	Voluntary Filers	Non-Filers
Income	mean^{+}	24,447.75	34,524.08	18,046.71
	p25	8,231.00	22,624.00	4,479.00
	p50	21,951.00	32,559.00	13,023.00
	p75	35,465.00	43,225.00	27,200.00
	p90	48,539.00	56,855.00	40,063.00
	p99	86,952.50	97,373.00	75,206.00
N	raw	683,718	425,579	258,139
	weighted	14,863,136	5,773,958	9,089,178

Notes: Statistics refer to taxable units, which may be either an individual or married spouses in case of joint filing. Statistics are based on weighted data if not indicated differently. *Income*: Annual gross wage income in €. For jointly filing spouses, the average gross income is taken into account. *Married*: Share of married taxpayers. *East*: Share of taxpayers that live in Eastern states of Germany. *Children*: Share of taxpayers with at least one child that is relevant for the tax authority. $^{+}$ indicates mean/share difference between voluntary filers and non-filers significant at the 0.1% - level (two-sided t-test).

Table 2.B.1: Descriptive Statistics - Sample 1

The vast majority of optional filers are single taxpayers, because most married couples choose a tax withholding schedule that triggers compulsory filing. Spouses who do not choose this schedule face the same withholding as single taxpayers and face three filing options. First, they can file jointly. For those couples, we consider their average income and age in Table 2.B.1. Second, they can file their taxes individually. Third, since filing is optional, they can be non-filers.

One potential concern here is that the sample may include taxpayers who are not only employed but also informally self-employed, i.e. without reporting their self-employed income. In these cases, we would compute over-remittances based on the reported wage income, whereas, in reality, these taxpayer evade taxes. Relying on administrative income tax data, we mechanically cannot rule out that these cases of evasion exist. However, we argue that such evasion behavior is unlikely to drive our results for two reasons. First, to be falsely classified as an optional filer, the self-employment must be unknown to the tax authorities. This implies not only evading 100% of the income taxes for self-employed income, but also limits the business activities to the informal sector, because any issuing of invoices etc. requires a tax ID. The taxpayer would then have to do this for several subsequent years, because the data is published only after waiting for tax assessments concluded within 2 years and 9 months after the tax year. Second, the informal sector in Germany is rather small: for the years 2019 to 2021,

		Optional Filers	Voluntary Filers	Non-Filers
Income	mean^{+}	18,968.17	27,068.54	15,093.15
	p25	6,363.00	18,746.00	4,101.00
	p50	17,497.00	28,325.00	11,409.00
	p75	30,221.00	36,200.00	24,140.00
	p90	38,491.00	42,248.00	34,514.00
	p99	47,155.00	47,894.00	46,151.50
Age	mean^{+}	33.71	35.26	32.97
	p50	30	32	29
Married	share^{+}	4.61	0.64	6.51
East	share^{+}	21.96	17.94	23.88
Children	share^{+}	15.13	16.47	14.49
N	raw	424,147	218,469	205,678
	weighted	11,829,433	3,827,786	8,001,646

Notes: Individual taxpayers with gross income up to 48,600 € only. No civil servants. Statistics are based on weighted data if not indicated differently. *Income*: Annual gross wage income in €. *Married*: Share of married taxpayers. This includes only individually filing spouses. If married taxpayers file, they almost always file jointly with their spouse, which is why <1% of voluntarily filing individual taxpayers are married. *East*: Share of taxpayers that live in Eastern states of Germany. *Children*: Share of taxpayers with at least one child that is relevant for the tax authority. $^{+}$ indicates mean/share difference between voluntary filers and non-filers significant at the 0.1% - level (two-sided t-test).

Table 2.B.2: Descriptive Statistics - Sample 2

German customs report an annual amount of loss in taxes of around 50 million € from undeclared work.[14] For reference, this corresponds to about 5% of our over-remittance estimates for 2014.

Sample 2. This sample is a subsample of sample 1. We use this sample for analyzing tax over-remittances and the resulting effective tax rates for non-filers in subsection 2.4.2, 2.4.3, 2.4.4, and 2.4.5. It includes individual taxpayers for whom we can calculate the statutory tax liability based on their gross income. This excludes high-income taxpayers who face no compulsory enrollment in the public health insurance (see section 2.C for details). Since this is a sample of individual taxpayers, we also drop jointly filing spouses among voluntary filers, but include married taxpayers that are either non-filers or voluntarily file individually.

We further exclude taxpayers for whom withheld taxes are not in line with tax-relevant individual characteristics reported in the data. We define this group as taxpayers for whom the computed statutory tax liability is smaller than the amount of taxes withheld ($\Delta \leq$ -5 €). While under-withholding is not possible for optional filers, such deviation can occur if tax-relevant characteristics change throughout the year. Because the LESt data provides only end-of-year information, we cannot observe such changes. In these cases, our computed statutory tax liability can deviate from the true statutory tax li-

[14] German report: https://www.bundesfinanzministerium.de/Content/DE/Downloads/Zoll/zoll-jahresstatistik-2021-jahresstatistikbroschuere.pdf?__blob=publicationFile&v=4 (last accessed: 2023-01-29), relevant numbers on p.19.

ability and withheld taxes can thus seem "too low". In addition, there may also be mistakes in the data set, causing such deviation. While, by definition, we cannot rule out that there are also cases in which simulated taxes are "too high" for similar reasons, we argue that the potential measurement error does not constitute a thread for implementing automatic refunds for two reasons.

First, financial authorities have more precise information about the relevant characteristics than provided in the LESt data. Most relevant, they observe changes throughout the year, while the LESt data contains only the end-of-year values. If automatic refunds were implemented, changes in these characteristics throughout the year could not cause "too high" refunds. Second, we argue that, in terms of magnitude, this measurement error is of subordinate relevance for the total sum of over-remittance. While, by definition, the data does not allow to identify cases in which simulated taxes are "too high", we can identify cases in which simulated taxes are "too low" (this is the group we are excluding from our sample). Assuming that the extent of both types of measurement error is comparable, including taxpayers for whom the simulated statutory taxes are "too low" for the estimation of over-remittances is equivalent to excluding taxpayers for whom the simulated statutory taxes are "too high". If we include the first group in our sample of non-filers, our results remain fairly robust: the sum of over-remittance in 2014 is 805 million € instead of 949 million €.

Sample 2 contains 424,147 optional filers with 205,678 not filing an income tax return (Table 2.B.2). Applying the weights provided by the statistical office, these non-filers represent 8 million taxpayers. Table 2.B.2 provides descriptive statistics for this sample.

Sample 3. For investigating the implications of non-filing for efficiency, we rely on a sample that is not restricted by filing type or income and thus deviates from the sample used before. We now focus on taxable income rather than gross wage income because we include taxpayers with income other than wage income. We restrict the sample to taxpayers with positive taxable income $y > 0$. For non-filers, we preferably use our self-computed taxable income from section 2.4 and rely on the reported taxable income for non-filers that are not in sample 2. For filers, we use the taxable income from their tax returns, considering the mean income for joint filing spouses. Lastly, we drop taxpayers for whom the imputed statutory tax liability is more than 5 € smaller than the observed remittances (same restriction as for sample 2).

2.C Computing Statutory Taxes for Non-Filers

To compute non-filers' over-remittances, we compare their effectively withheld income taxes T_i^E against the statutory taxes, as defined in the income tax schedule. Starting from the annual gross wage income, we subtract the standard deductions and special allowances for single parents or elderly employees, if applicable. Based on demographics, we then determine social insurance contributions, which are partly tax deductible, and subtract those. This yields y_i, the annual individual taxable income for taxpayer i. Using the statutory tax schedule, we then compute $T^S(y_i)$, the statutory tax liability of individual i. Absent over-withholding, T_i^E and $T^S(y_i)$ coincide. For non-filers with over-withholding, $T^S(y_i)$ is the final tax liability they would face had if they filed their income taxes (under minimal tax filing).

This corrects for over-withholding through employers by smoothing fluctuating incomes, i.e. it performs the annual adjustment of wage tax, and considers both the full standard deductions and the full basic allowance. The resulting estimates can be interpreted as the lower bound for tax over-remittances through non-filing, or potential monetary benefits from tax filing, for three reasons. First, we exclude civil servants and employees with an annual wage income exceeding 48,600 € from our analysis. Since non-filing and potentially over-withholding is also present in this group (but cannot be quantified with the data at hand), this represents a lower bound for the total sum of over-remittances.

We exclude these taxpayers from our sample, because we cannot precisely determine their taxable income and the corresponding statutory income taxes that are required for quantifying over-remittances. Taxpayers with an annual gross income exceeding 48,600 € in 2014 (about the 90$^{\text{th}}$ percentile of wage earners) as well as civil servants are free to choose whether or not to enroll in the public health insurance and there is no information on the enrollment status for non-filers. We thus cannot infer social insurance contributions based on the annual gross income for this group, which is required for computing the taxable income. Additionally, we exclude joint filing married spouses and restrict the sample to the working age population ($16 \leq$ age ≤ 63). The sample restrictions are explained in section 2.B.

Second, our estimates do not take into account the full refund potential. We do not account for over-withholding of additional surtaxes (church tax and solidarity tax) and only consider the standard deductions of 1,000 € for work related expenses and 36 € for special expenses. Given these low standard values, and ample deduction possibilities, taxpayers might realize even larger benefits when itemizing.[15] Doerrenberg, Peichl, and Siegloch (2017) provide a detailed introduction into deduction possibilities in Germany. Third, for married non-filers, we calculate their filing counterfactual assuming individual filing. However, given the system of joint taxation in Germany, married spouses can typically reduce their tax liabilities further, when filing jointly. section 2.G provides an upper bound estimate considering additional deductions.

[15]For example, in 2014 57% of the German population were members of the Roman Catholic or Protestant Church, which automatically levy a state-dependent 8-9% surtax on the income tax, known as "church tax". Even though this church tax qualifies as a special expense, it is not taken into account by the employer when withholding income taxes. Given the data at hand, we are not able to compute this additional benefit.

2.D Main Results for 2010

We replicate our main findings for 2014 with the 2010 LESt data to rule out that our results are specific to the 2014 tax year. Figure 2.D.1a shows that the prevalence of non-filing (panel a), non-filers' over-remittances (panel b) and the resulting ATR^E (panel c) are similar to 2014. The average 2010 non-filer over-remits 103 € (Table 2.D.1), compared to 118 € in 2014. This is equivalent to a 15% increase from 2010 to 2014, which is proportional to the overall economic development over this period: the German gross national income increased by 14% between 2010 and 2014 (Federal Statistical Office 2022b).

		(A) All Non-Filers		(B) With Over-Withholding	
		(1) All	(2) y < threshold	(3) All	(4) y < threshold
Over-Remittance	total	603,472,488	240,758,764	604,502,377	240,584,508
	mean	103.38	81.12	318.89	238.33
	p25	-0.34	0.00	41.90	48.00
	p50	0.00	0.00	164.00	139.00
	p75	36.01	52.00	441.54	331.00
	p90	353.00	286.00	829.04	602.00
	p95	648.73	487.00	1,105.24	798.00
N	raw	167,061	83,296	55,399	29,741
	weighted	5,837,580	2,967,949	1,895,650	1,009,474

Notes: This table replicates the results from Table 2.A.1 for 2010. Over-remittances are defined as $T_i^E - T^S(y_i)$ for individual i and listed in €. (A): All non-filers in the sample. (B): Only those non-filers with over-remittances. See also table notes from Table 2.A.1. The average over-remittance is 15% smaller in 2010 as compared to 2014 (103 € vs 118 €). This is proportional to the overall economic development: The German gross national income increased by 14% between 2010 and 2014 (Federal Statistical Office 2022b). Because of fewer non-filers in the 2010 sample, there is a larger divergence in aggregated over-remittances.

Table 2.D.1: Taxes Over-Remitted Through Non-Filing – 2010

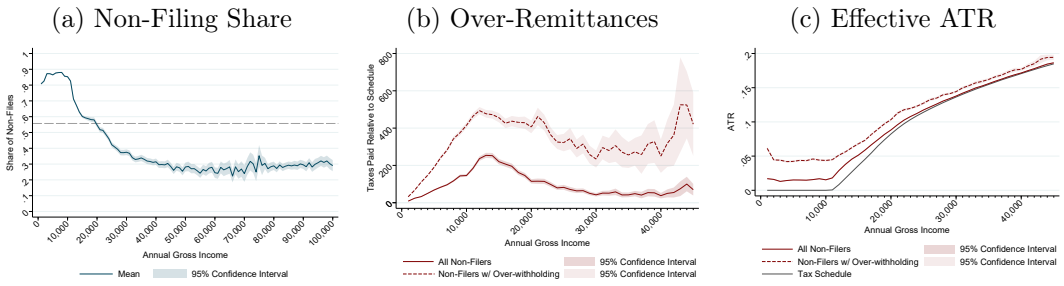


Figure 2.D.1: Main Figures – 2010

Notes: This figure replicates our main findings with 2010 data. Panel (a) is the equivalent of Figure 2.1, (b) is the equivalent of Figure 2.3, and (c) is the equivalent of Figure 2.4 For more information, see figure notes from Figure 2.1, Figure 2.3, and Figure 2.4 in the main text.

2.E Details: Over-Withholding for Non-Filers in Other Countries

Austria. To analyze over-withholding for Austrian employees, we use aggregated data on tax withholdings (Fischer and Milz 2017). The provided data is aggregated over income bins of 2,000 € and contains average amounts of taxes withheld for each bin. We focus on employees with gross income below the basic allowance threshold. The sample includes both non-filers and filers. For filers, the deviation between effective ATR (ATR^E) and statutory ATR (ATR^S) is only temporary because they receive a refund upon filing. For non-filers, the withheld ATR is their final effective ATR. Because the data does not differentiate by filing status, we report ATR^E before potential filing.

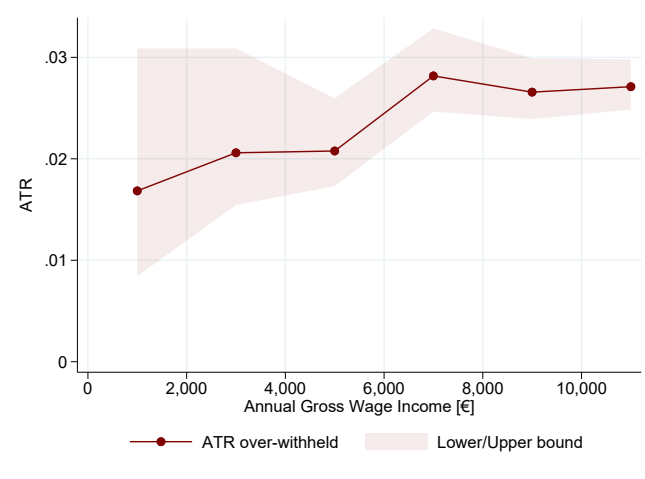


Figure 2.E.1: Over-Withholding Austrian Employees 2014

Notes: The figure shows the effective ATR withheld under automatic withholding for low-income Austrian employees in 2014 – before potential filing. *Reading example*: The average withholding in the income bin 4,000 - 5,999 € is 104 €. Dividing by the middle of the income bin (5,000 €) yields an average ATR^E of 2.1%. Dividing the 104 € by the bin limits of 5,999 € and 4,000 € yields the lower and upper bound of 1.7% and 2.6%, respectively.

Figure 2.E.1 plots the effective ATR withheld under automatic withholding for low-income Austrian employees in 2014 – before potential filing. Within this income range, employees face withholdings of 2 to 3 percentage points of their gross income. We compute the effective ATR (ATR^E) by dividing this average by the middle of the given income bin (dots within each bin). Depending on the within-bin income distribution, the average ATR^E can be higher or lower than this. The lower/upper bound provides the potential range for the average ATR^E by dividing over-withholdings by the respective income bin limits. We cap the upper bound in the first bin because ATR^E converges to ∞ for income converging to 0. We focus on employees whose annual gross wage income is below the basic allowance, implying $ATR^S = 0$ under the assumption that employees do not have other sources of income. Since the data does not inform about other income and/or later filing, we cannot provide exact quantitative measures for over-withholdings here. The goal of this exercise is to show that the Austrian tax system implies over-withholdings that are qualitatively similar to the German case.

United States. We simulate monthly wage tax withholding for a single US taxpayer with one withholding allowance in 2014, following the withholding table in the Employer's Tax Guide provided by the IRS (Department of the Treasury, Internal Revenue Service (IRS) 2013). We assume that they have wage income only and that monthly income is constant over the m months employed in 2014. When withholding income taxes, employers extrapolate the annual income based on the monthly income. Over-withholding can occur for taxpayers who are employed for $m < 12$ months, because the extrapolated annual income is higher than the true annual income and standard deductions are only considered partly – very similar to the German case.

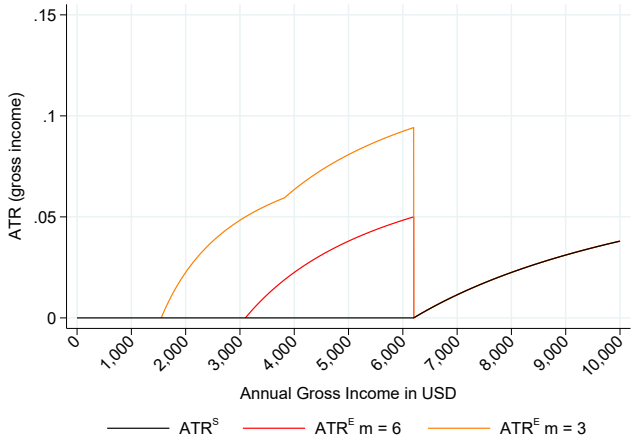


Figure 2.E.2: Over-Withholding US Non-Filers 2014 (simulated)

Notes: The figure shows the simulated effective ATR (ATR^E) that follows automatic monthly withholding for US non-filers employed for 3 ($m = 3$) or 6 months ($m = 6$), as compared to their statutory ATR (ATR^S). Computations based on the Employer's Tax Guide provided by the IRS (Department of the Treasury, Internal Revenue Service (IRS) 2013). *Reading example*: A US taxpayer who is employed for 3 months with annual gross wage of \$6,000 (= monthly gross wage of \$2,000) has $ATR^S = 0$ but faces $ATR^E = 0.09$.

Figure 2.E.2 shows the simulated effective ATR (ATR^E) that follows automatic monthly withholding for US non-filers employed for 3 ($m = 3$) or 6 months ($m = 6$), as compared to their statutory ATR (ATR^S). We choose 3 and 6 months for illustration, but over-withholdings can occur for any $m < 12$, with highest over-withholding potential for $m = 1$ and smallest potential for $m = 11$. Figure 2.E.2 shows that optional non-filers in the US can face substantial tax over-remittances. For example, a US non-filer who is employed for 3 months with annual gross wage of \$6,000 (= monthly gross wage of \$2,000) has $ATR^S = 0$ but faces $ATR^E = 0.09$ (see Figure 2.E.2 for details). These estimates only account for income tax over-withholdings and do not include transfers that low-income taxpayers may be eligible but forgo when not filing (e.g., EITC). By definition, over-withholding only affects lowest-income taxpayers, because non-filing is optional for taxpayers with $ATR^S = 0$ only, i.e. for income up to the standard deduction (\$6,200 for singles in 2014).

2.F Budget Neutral Reform

We provide two back-of-the-envelope calculations for budget neutral reforms that raise additional tax revenue dR from high-income taxpayers to offset the losses in tax revenue from automatic refunds (Reform A and B). Reform A increases the MTR in the top income tax bracket (τ_5) and Reform B increases the MTRs in the top two income tax brackets (τ_4 and τ_5). We take a simplified approach, abstracting from income effects and approximating behavioral responses. For this exercise, we rely on a larger sample that includes compulsory filers as well as taxpayers of all income ranges.

Following the standard approach (e.g., Saez 2001), dR can be decomposed into a mechanical (dM) and a behavioral effect (dB), with $dR = dM + dB$. For Reform A dR is defined as

$$dR = \underbrace{N \cdot (y^m - \bar{y}) \cdot d\tau}_{dM} \underbrace{- N \cdot \varepsilon_{y,1-\tau} \cdot y^m \cdot \frac{\tau_5}{1-\tau_5} \cdot d\tau_5}_{dB} \tag{2.1}$$

that we can rearrange to compute the implied $d\tau$:

$$d\tau = dR \cdot \left[N \cdot (y^m - \bar{y}) \cdot \left(1 - \varepsilon_{y,1-\tau} \cdot \frac{y^m}{y^m - \bar{y}} \cdot \frac{\tau_5}{1-\tau_5} \right) \right]^{-1}. \tag{2.2}$$

For Reform B, $d\tau$ is defined as:

$$\begin{aligned} d\tau = dR \cdot \Bigg[& \left(y^m_{\tau_4} - \bar{y}_{b_4} \right) \cdot N_{\tau_4} + \left(\bar{y}_{b_5} - \bar{y}_{b_4} \right) \cdot N_{\tau_5} + \left(y^m_{\tau_5} - \bar{y}_{b_5} \right) \cdot N_{\tau_5} \\ & - \varepsilon_{y_{\tau_4},1-\tau} \cdot y^m_{\tau_4} \cdot N_{\tau_4} \cdot \frac{\tau_4}{1-\tau_4} - \varepsilon_{y_{\tau_5},1-\tau} \cdot y^m_{\tau_5} \cdot N_{\tau_5} \cdot \frac{\tau_5}{1-\tau_5} \Bigg]^{-1}. \end{aligned} \tag{2.3}$$

For given values of the elasticity of taxable income with respect to the net of tax rate ($\varepsilon_{y^m,1-\tau} = [(1-\tau)/y^m] \cdot [\partial y^m / \partial (1-\tau)]$) we can then compute $d\tau$ with all other parameters being identified by the data (Table 2.F.1). For $\varepsilon_{y,1-\tau} = 0.2$, the required $d\tau$ is 4 percentage points for Reform A and 1 percentage point for Reform B.

	Reform A			Reform B		
(1) Parameters						
τ_5	0.45			0.45		
τ_4	-			0.42		
$\bar{y}_{b_5}$	250,730 €			250,730 €		
$\bar{y}_{b_4}$	-			52,881 €		
N_{τ_5}	76,388			76,388		
N_{τ_4}	-			2,100,109		
$y^m_{\tau_5}$	667,869 €			667,869 €		
$y^m_{\tau_4}$	-			82,654 €		
$y^m_{\tau_5} - \bar{y}_{b_5}$	417,139 €			417,139 €		
$y^m_{\tau_4} - \bar{y}_{b_4}$	-			29,773 €		
$\bar{y}_{b_5} - \bar{y}_{b_4}$	-			197,849 €		
dR	949,512,506 €			949,512,506 €		
dR^m	12,430 €			436 €		
(2) Results		$\varepsilon_{y,1-\tau}$			$\varepsilon_{y,1-\tau}$	
	0	0.2	0.3	0	0.2	0.3
$d\tau$	0.0298	0.0404	0.0491	0.0087	0.0100	0.0108
τ_4^{new}	-	-	-	0.4287	0.4300	0.4308
τ_5^{new}	0.4798	0.4904	0.4991	0.4587	0.4600	0.4608

Notes: Reform A: change MTR in top bracket (τ_5) by $d\tau$. *Reform B*: change MTR in top two brackets (τ_4, τ_5) by $d\tau$ each. The sample includes taxpayers of all income ranges and is not restricted to optional filers. For jointly filing spouses, we consider their average income and tax liabilities.
j: Income tax bracket ($j = 4, 5$). y: Taxable income. τ_j: MTR in bracket j. $\bar{y}_{b_j}$: Income threshold for bracket j. N_{τ_j}: N taxpayers in bracket j. $y^m_{\tau_j}$: Mean income y for taxpayers in bracket j. dR: Change in total tax revenue. dR^m: Change in mean tax revenue (individual level). $\varepsilon_{y,1-\tau}$: Elasticity of taxable income. $d\tau$: Absolute change in MTR required to raise 603 million €. τ_j^{new}: New MTR in bracket j after increase of $d\tau$.

Table 2.F.1: Parameters for Reform A and B

2.G Upper Bound Counterfactual

Our main results from subsection 2.4.3 are a lower bound for forgone refunds of non-filers because we do not account for itemized deductions. However, itemizing deductions accounts for a substantial part of tax refunds. The average voluntary filer realizes a refund of 137 € from over-withholding and 402 € from itemizing deductions.

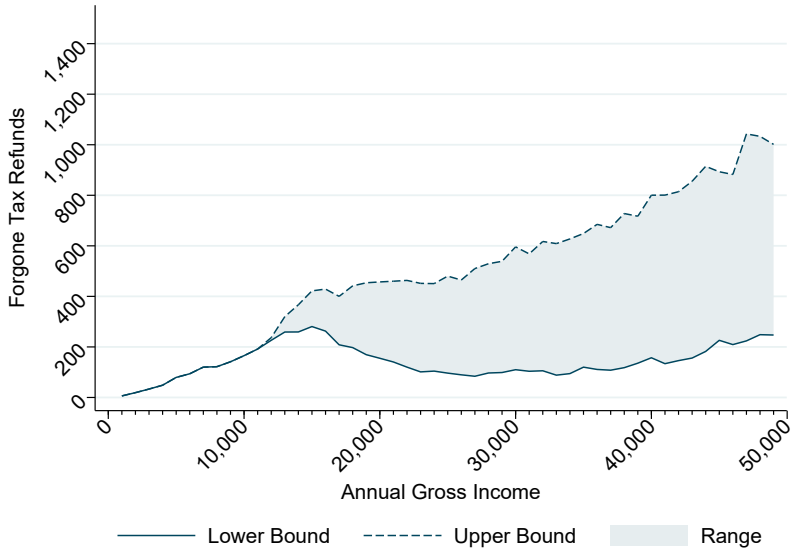


Figure 2.G.1: Forgone Tax Refunds Range for Non-Filers

Notes: Individuals are grouped in 1,000-€-bins. The refund range is bounded from below by the minimal-filing lower bound (subsection 2.4.3) and from above by the realized refunds of voluntary filers in that same income bin. The range widens with income because there is no additional refund potential for incomes below the basic allowance threshold and because voluntary filers with higher income realize higher refunds from itemizing deductions.

We provide an upper bound of forgone refunds for non-filers by assigning each non-filer the average refund that voluntary filers in their gross wage income bin realize upon filing. This is an upper bound as long as there is some positive selection into filing based on refund potential, i.e. as long as those who choose to file can, on average, realize higher refunds. Figure 2.G.1 plots this upper bound for non-filers (dashed line) in addition to our lower bound estimate from the main analysis (solid line).

2.H Details Effective Marginal Tax Rates

Benchmark I: "More Tax Revenue from Mechanical Effect".

This benchmark introduces a tax schedule with $MTR_1^{S\prime} = MTR_1^{E\prime} = 5.43\%$, which maps the average effective MTR for lowest-income taxpayers with income below the basic allowance threshold y_1 in the status quo (tax bracket 1 with $\overline{MTR_1^E}$ as compared to $MTR_1^S = 0$). This increases tax revenue for two reasons. First, the tax liability for taxpayers with $y > y_1$ increases by $MTR_1^{S\prime} \cdot y_1 = 0.0543 \cdot 8{,}354\,€ = 454\,€$. Increasing $MTR_1^{S\prime}$ increases the effective ATR of these taxpayers, but their MTR remains unchanged. Consequently, the revenue gain from this group comes with no distortional costs when abstracting from income effects.

Second, and potentially less intuitive, this also increases the effective ATR for the average taxpayer in bracket 1, although we keep their effective MTR constant. This is driven by the fact that MTRs are more sensitive to over-withholding than ATRs in the status quo. If taxes are withheld, $MTR^E \geq 0.14$, the minimum positive MTR^S, but ATR^E can be close to zero and average to $\overline{ATR_1^E}$ in bracket 1. Implementing $MTR_1^{S\prime} = MTR_1^{E\prime} = 5.43\%$ thus introduces $ATR_1^{E\prime} > \overline{ATR_1^E}$. Combining both groups leads to the overall increase of 12.9 billion € as shown in Equation 2.4.

$$
\begin{aligned}
dT' &= \sum_{y=1}^{\infty} \left(ATR^{E\prime}(y) - ATR^{E}(y)\right) y N_y \\
&= \underbrace{\sum_{y=1}^{y_1} \left(ATR_1^{E\prime}(y) - ATR_1^{E}(y)\right) y N_y}_{\text{Bracket 1}} \; + \; \underbrace{\sum_{y>y_1}^{\infty} ATR_1^{E\prime} y_1 N_y}_{\text{Brackets 2-5}} \\
&= \left(ATR_1^{E\prime} - \overline{ATR_1^E}\right) \overline{y_1} \sum_{y=1}^{y_1} N_y \; + \; ATR_1^{E\prime} y_1 \sum_{y>y_1}^{\infty} N_y \\
&= (0.0543 - 0.0187) \cdot 4{,}077 \cdot 8{,}687{,}419 + 0.0543 \cdot 8{,}354 \cdot 25{,}677{,}500 \\
&= 12.9 \text{ billion}
\end{aligned}
\tag{2.4}
$$

One potential way to use the gained tax revenue dT' is to transfer it back to all 34 million taxpayers (N) in our sample via a lump sum payment defined as $P = dT'/N = 375\,€$. While P is the same for all, net benefits differ over the income distribution. Taxpayers with an income $y_i > y_1$, see an increase in their tax liability by 78 €, whereas for $y_i \leq y_1$, the net benefit is $(\overline{ATR_1^E} - ATR_1')y_i > 0$ with an average refund of 230 €.

The results are an approximation and require a few assumptions. First, we assume that all taxpayers with $y_i > y_1$ do not pay taxes on income below y_1. However, for non-filers this might not be true. Second, by including all taxpayers in this analysis, the measurement of taxable income is less precise than in section 2.4. For some non-filing taxpayers who were not included in the main analysis, we rely on information on taxable income as reported in the data. The measurement of taxable income in the data set may not be fully accurate as discussed in subsection 2.4.3. Furthermore, we keep $\overline{MTR_1^E}$ constant in tax bracket 1 but abstract from variations in the individual

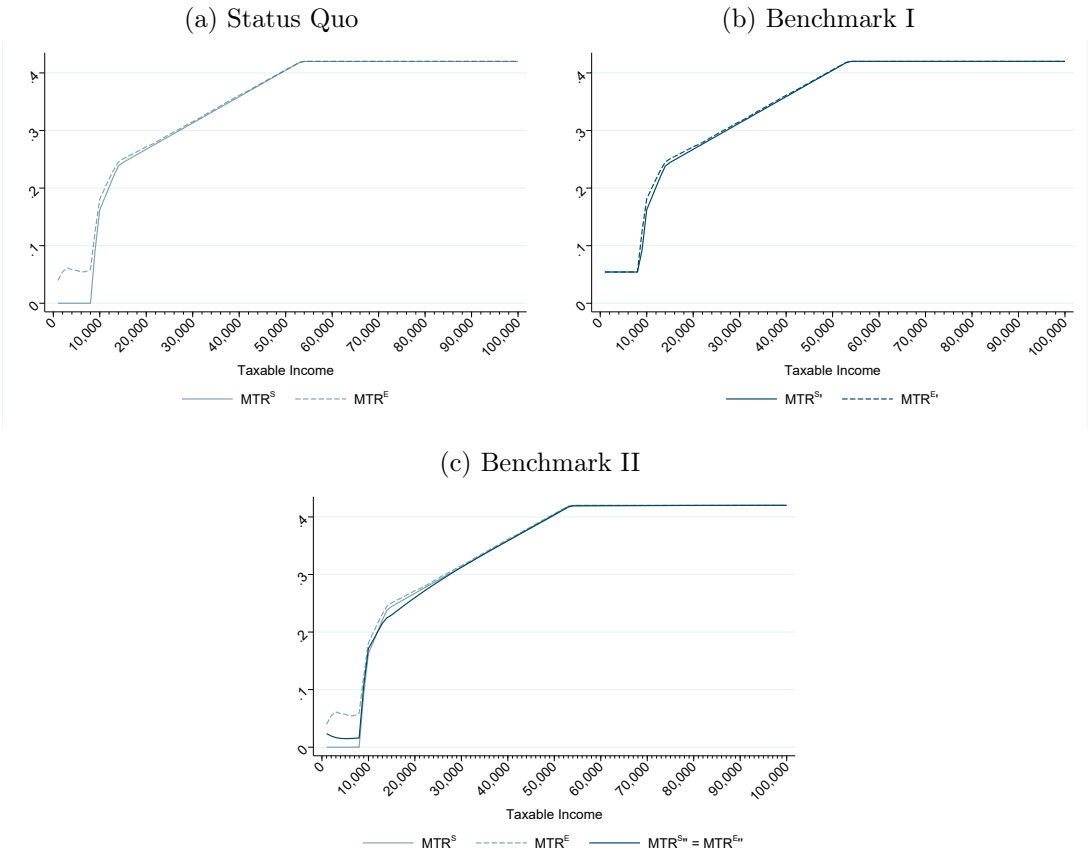


Figure 2.H.1: MTR^S and MTR^E - Status Quo vs. Benchmark I & II

Notes: Taxpayers with taxable income up to 100,000 €, mean income for joint filing spouses. Taxpayers are grouped in 1,000-€-bins. MTR^S: Statutory marginal tax rate as defined in the tax schedule. MTR^E: Effective marginal tax rate that is observed under optional non-filing. MTR^E is the average for all taxpayers at income level y - filers and non-filers.

MTR^E. Consequently, there is no elasticity effect to a change in MTR^E. Lastly, we abstract from income effects.

Benchmark II: "More Tax Revenue from Behavioral Responses".

This benchmark introduces a tax schedule that maps the effective ATR observed in the status quo for any given income level. There is no deviation between statutory and effective taxation, such that $ATR^{S\prime\prime} = ATR^{E\prime}$ and $MTR^{S\prime\prime} = MTR^{E\prime}$.

We define the new tax schedule, $T''(y)$, as a smoothed function of the status quo average effective tax liabilities for taxpayers with taxable income y, $T^E(y)$. We derive $T''(y)$ by fitting a fourth order polynomial function to the observed effective tax liabilities from the status quo over taxable income y in each of the 5 tax brackets b. Estimating the polynomial functions with OLS yields $T''_b(y) = \beta_0 + \beta_1 y + \beta_2 y^2 + \beta_3 y^3 + \beta_4 y^4$ for each b. We set $\beta_0 = 0$ for $b = 1$ to ensure $T''_1(0) = 0$. We define $MTR''(y)$ as the derivative of $T''_b(y)$ and derive $ATR''(y)$ by dividing $T''_b(y)$ by y. As compared to the status quo, the average effective ATR is kept constant for a given income level y. Effective and statutory taxation are aligned now for all income levels: $T''(y) = T^{S\prime\prime}(y) = T^{E\prime\prime}(y)$, with the corresponding $ATR''(y) = ATR^{S\prime\prime}(y) = ATR^{E\prime\prime}(y)$ and $MTR''(y) = MTR^{S\prime\prime}(y) = MTR^{E\prime\prime}(y)$.

Absent behavioral responses, this benchmark keeps the tax revenue constant as compared to the status quo because effective ATRs, on average, do not change.[16] With behavioral responses however, it increases tax revenue because effective MTRs are reduced. Lowest-income taxpayers (bracket 1) face substantially lower distortions with $MTR^{E\prime\prime}(y) < \bar{MTR}^E_1$ for all y in the first bracket. Reaching the same ATR^E via a tax schedule allows for much lower MTR^E as compared to the non-filing status quo, where $MTR^E \geq 0.14$ under any withholding (same mechanism as described for benchmark 1). Intuitively, higher-income taxpayers now pay more taxes on income in lower brackets because the positive MTR^E for lowest income levels is implemented in the tax schedule. Because we keep ATR^E constant at any given income level, this allows for lower MTRs at higher income levels.

In response to lower effective MTRs, taxpayers increase their taxable income. We compute these behavioral responses for each individual taxpayer leading to an aggregated increase in tax revenue of dT'':

$$\begin{aligned} dT'' &= \sum_{i=1}^{N} T''(y_i) - T^E_i + dy_i \cdot MTR''(y_i) \\ &= \sum_{i=1}^{N} T''(y_i) - T^E_i - \varepsilon_{y,1-MTR^E} \frac{dMTR^E_i}{1 - MTR''(y_i)} \cdot MTR''(y_i) \end{aligned} \tag{2.5}$$

Taxpayers adjust their taxable income by $dy_i = \partial y / \partial MTR''$ which increases their tax liability by dy_i times $MTR''(y_i)$.[17] The degree of this behavioral response depends on the elasticity of taxable income with respect to MTR^E: $\varepsilon_{y,1-MTR^E} = (1-\tau)/y \cdot$

[16] The average difference $T''(y_i) - T^E_i$ is close to zero with an annual value of 0.14 €.

[17] We take the simplifying assumption here that $MTR''(y) = MTR''(y + dy)$.

dy_i	$\sum dy_i$	dT_i''	$\sum dT_i''$
(A) Same Elasticity for Filers and Non-Filers $\left(\varepsilon^{F}_{y,1-MTR^E} = \varepsilon^{NF}_{y,1-MTR^E} = 0.2\right)$			
43.93	1,509,612,466	8.66	297,432,805
(B) Different Elasticities for Filers and Non-Filers $\left(\varepsilon^{F}_{y,1-MTR^E} = 0.2, \varepsilon^{NF}_{y,1-MTR^E} = 0.0\right)$			
22.82	784,289,807	5.62	193,009,407

Notes: Quantification of the tax revenue effect of benchmark II as defined in Equation 2.5. All values in €. Panel (A) shows results for $\varepsilon_{y,1-MTR^E} = 0.2$ for all taxpayers, both filers (F) and non-filers (NF). Panel (B) shows results for $\varepsilon^{F}_{y,1-MTR^E} = 0.2$ for filers and $\varepsilon^{F}_{y,1-MTR^E} = 0.0$ for non-filers. Results for $\varepsilon_{y,1-MTR^E} = 0.3$ are shown in Table 2.H.2. Effective tax liabilities are smoothed by fitting a forth order polynomial function (OLS) for each tax bracket b to derive $T_b''(y) = \beta_0 + \beta_1 y + \beta_2 y^2 + \beta_3 y^3 + \beta_4 y^4$. $\beta_0 = 0$ for the first bracket to ensure $T_1''(0) = 0$. The corresponding $MTR''(y)$ is defined as the derivative of $T_b''(y)$ and thus a stepwise third order polynomial function of y within each tax bracket. Applying different polynomial fits for smoothing does not change the magnitude of the effect, see Table 2.H.3 for a stepwise linear and quadratic $MTR''(y)$. Smoothing T^E over y is necessary to obtain $T''(y)$ with meaningful $MTR''(y)$. However, it comes at the cost of not keeping the average effective tax liability 100% constant at each income level. On average, the annual $T''(y)$ is 0.14 € higher for a given income level y. dy_i: Average change in individual taxable income. $\sum dy_i$: Aggregate change in taxable income. dT_i'': Average individual change in tax remittance defined as $T_i'' - T_i^E$. $\sum dT_i''$: Aggregate change in tax revenue.

Table 2.H.1: Benchmark II: Behavioral Responses

dy_i	$\sum dy_i$	dT_i''	$\sum dT_i''$
(A) Same Elasticity for Filers and Non-Filers $\left(\varepsilon^{F}_{y,1-MTR^E} = \varepsilon^{NF}_{y,1-MTR^E} = 0.3\right)$			
65.89	2,264,418,698	12.96	445,405,280
(B) Different Elasticities for Filers and Non-Filers $\left(\varepsilon^{F}_{y,1-MTR^E} = 0.3, \varepsilon^{NF}_{y,1-MTR^E} = 0.0\right)$			
34.23	1,176,434,711	8.40	288,770,183

Notes: Quantification of the tax revenue effect of benchmark II with different assumed elasticity of taxable income. All values in €. Panel (A) shows results for $\varepsilon_{y,1-MTR^E} = 0.3$ for all taxpayers, both filers (F) and non-filers (NF). Panel (B) shows results for $\varepsilon^{F}_{y,1-MTR^E} = 0.3$ for filers and $\varepsilon^{F}_{y,1-MTR^E} = 0.0$ for non-filers. Effective tax liabilities are smoothed by fitting a forth order polynomial function (OLS) for each tax bracket b to derive $T_b''(y) = \beta_0 + \beta_1 y + \beta_2 y^2 + \beta_3 y^3 + \beta_4 y^4$. $\beta_0 = 0$ for the first bracket to ensure $T_1''(0) = 0$. The corresponding $MTR''(y)$ is defined as the derivative of $T_b''(y)$ and thus a stepwise third order polynomial function of y within each tax bracket. Smoothing T^E over y is necessary to obtain $T''(y)$ with meaningful $MTR''(y)$. However, it comes at the cost of not keeping the average effective tax liability exactly constant at each income level. On average, the annual $T''(y)$ is 0.14 € higher for a given income level y. dy_i: Average change in individual taxable income. $\sum dy_i$: Aggregate change in taxable income. dT_i'': Average individual change in tax remittance defined as $T_i'' - T_i^E$. $\sum dT_i''$: Aggregate change in tax revenue.

Table 2.H.2: Benchmark II: $\varepsilon_{y,1-\tau} = 0.3$

$\partial y / \partial(1-\tau)$. Plugging $\varepsilon_{y,1-MTR^E}$ into Equation 2.5, we can quantify the change in tax liability dT_i'' for each individual i in the data set. Aggregating over the taxpayer population N yields the overall effect on tax revenue dT''.

Table 2.H.1 shows the results for $\varepsilon_{y,1-MTR^E} = 0.2$, assuming $\varepsilon_{y,1-MTR^E} = 0.2$ for all taxpayers (Panel A) or $\varepsilon_{y,1-MTR^E} = 0.2$ for filers and $\varepsilon_{y,1-MTR^E} = 0$ for non-filers (Panel B). Table 2.H.2 shows the same results for $\varepsilon_{y,1-MTR^E} = 0.3$ and Table 2.H.3 for the same $\varepsilon_{y,1-MTR^E} = 0.2$ but with different polynomial orders.

Again, the results are an approximation and build on several assumptions. First, we only consider average tax payments at a given income level. It is important to note that

	dy_i	$\sum dy_i$	dT_i''	$\sum dT_i''$
		(A) Same Elasticity for Filers and Non-Filers		
		$\varepsilon^F_{y,1-MTR^E} = \varepsilon^{NF}_{y,1-MTR^E} = 0.2$		
Linear	41.22	1,416,525,933	7.62	261,905,054
Quadratic	43.92	1,509,457,767	8.76	300,878,495
		(B) Different Elasticities for Filers and Non-Filers		
		$\varepsilon^F_{y,1-MTR^E} = 0.2$, $\varepsilon^{NF}_{y,1-MTR^E} = 0.0$		
Linear	20.23	695,093,006	4.59	157,809,423
Quadratic	22.95	788,510,358	5.74	197,136,862

Notes: Alternative quantification of the tax revenue effect of benchmark II as shown in Table 2.H.1. All values in €. Panel (A) shows results for $\varepsilon_{y,1-MTR^E} = 0.2$ for all taxpayers, both filers (F) and non-filers (NF). Panel (B) shows results for $\varepsilon^F_{y,1-MTR^E} = 0.2$ for filers and $\varepsilon^F_{y,1-MTR^E} = 0.0$ for non-filers. *Linear*: Effective tax liabilities are smoothed by fitting a second order polynomial function (OLS) for each tax bracket b to derive $T_b''(y) = \beta_0 + \beta_1 y + \beta_2 y^2$. The corresponding $MTR''(y)$ is defined as the derivative of $T_b''(y)$ and thus a stepwise linear function of y within each tax bracket. *Quadratic*: Effective tax liabilities are smoothed by fitting a third order polynomial function (OLS) for each tax bracket b to derive $T_b''(y) = \beta_0 + \beta_1 y + \beta_2 y^2 + \beta_3 y^3$. Here, $MTR''(y)$ is a stepwise quadratic function of y within each tax bracket. For both the linear and quadratic MTR, $\beta_0 = 0$ for the first bracket to ensure $T_1''(0) = 0$. Smoothing T^E over y is necessary to obtain $T''(y)$ with meaningful $MTR''(y)$. However, it comes at the cost of not keeping the average effective tax liability exactly constant at each income level. On average, the annual $T''(y)$ is 0.10 € lower for a given income level y for the linear MTR and 0.03 € higher for the quadratic MTR. dy_i: Average change in individual taxable income. $\sum dy_i$: Aggregate change in taxable income. dT_i'': Average individual change in tax remittance defined as $T_i'' - T_i^E$. $\sum dT_i''$: Aggregate change in tax revenue.

Table 2.H.3: Benchmark II: Different Polynomial Fit

this is no Pareto-improvement. Under $T''(y)$, some taxpayers in a given income range pay more taxes and others pay less, when compared to their status quo. Second, we only consider behavioral responses at the intensive margin and assume that taxpayers can adjust their taxable income without frictions. Third, we abstract from income effects.

Chapter 3

Lost in Deduction: Taxpayers' Mistakes when Itemizing

3.1 Introduction

Tax filing is complex and commonly believed to distort individual behavior, resulting in non-optimal decisions. It is well established that in complex tax and benefit systems, individuals leave money on the table by not claiming benefits that they are entitled to. Such incomplete take-up spans various public policy settings (e.g., Bhargava and Manoli 2015; Currie 2006), including tax non-filing (e.g., Goodman, Lim, Sacerdote, and Whitten 2023; Hauck and Wallossek 2023; Ramnath and Tong 2017). However, it is typically not possible to ultimately rule out that this passive behavior is in fact rational, since unobserved individual costs for the omitted action may exceed the forgone benefits.

In this paper, I document taxpayer behavior that can be ruled out to be optimal: when filing an income tax return, taxpayers commonly itemize deductions, although itemizing has zero benefits for them. Using German administrative income tax data, I find that 57 percent of taxpayers make such an itemizing mistake at least once over a sample period of nine years. In total, 29 percent of all tax returns feature a mistake.

Taxpayers can reduce their tax liability by claiming deductible expenses when filing an income tax return. Commonly, taxpayers are entitled to a standard deduction, but can also choose to itemize deductions, allowing them to claim more than the standard deduction. Itemizing imposes costs, i.e., time spent for record keeping and entering the information in the tax forms. If the sum of itemized deductions does not exceed the standard deduction threshold, it comes with zero benefits. Likewise, if taxpayers already face a zero tax liability, itemizing has no benefit because it cannot further reduce taxes. In both cases, itemizing is strictly dominated by not itemizing and can be considered a mistake for taxpayers.

I find that itemizing mistakes are common across all demographic groups and along the income distribution, with lowest-income taxpayers being most prone to making mistakes. There is no evidence for learning from past mistakes: taxpayers with a mistake in a given year are even more likely to again make a mistake in the following year, resulting in 3 mistakes during 7 years observed for the average taxpayer. In addition, the prevalence of itemizing mistakes is not particularly high near the standard deduction threshold, indicating that taxpayers do not commonly err by targeting a sum above the threshold but inaccurately predicting their total deductions.

The findings suggest that large parts of the taxpayer population have partial, yet incomplete, tax literacy in the area of tax filing, leading to non-optimal tax behavior. They are aware of the potential benefit of deductions, but fail to assess whether the

benefit applies to their individual case. By documenting a higher rate of mistakes for joint filing spouses whose tax returns are more complex, I provide suggestive evidence that complexity fosters mistakes.

This paper adds to research on income tax filing: taxpayers have been shown to be willing to forgo tax refunds in order to save costs of itemizing (Benzarti 2020) and to be more likely to file when refunds are higher (Ramnath and Tong 2017). TThis evidence for rational filing behavior can be reconciled with the evidence for irrational filing behavior in this paper by taking into account heterogeneity in behavior among different taxpayers. While many tax filers behave irrationally by itemizing deduction when this has no benefit, I also document behavior that suggests rational filing behavior for other filers. In line with results from Benzarti (2020), there is a missing mass of deductions just above the standard deduction threshold.

I document active tax filing behavior that is strictly dominated by remaining passive. This contributes to a literature on incomplete take-up, which can be described as passive behavior that is (likely) dominated by active behavior (e.g., Bhargava and Manoli 2015; Currie 2006; Finkelstein and Notowidigdo 2019). This paper also relates to a literature that documents dominated choices for individuals (e.g., Bhargava, Loewenstein, and Sydnor 2017; Heiss, McFadden, Winter, Wuppermann, and Zhou 2021).

This paper also contributes to a broader literature on income tax complexity that documents taxpayers' confusion and limited ability to optimally respond to tax incentives (e.g., Abeler and Jäger 2015; Feldman, Katuščák, and Kawano 2016; Liebman and Zeckhauser 2004; Rees-Jones and Taubinsky 2020). With a focus on tax filing, I show that taxpayers do not only misperceive the income tax code, but also the more practical rules of the income tax system.

The remainder of this paper is structured as follows: I first provide details on the institutional setting and itemizing mistakes in Section 3.2 and describe the data and sample in Section 3.3. Section 3.4 documents the itemizing behavior and mistakes, before I analyze taxpayer behavior in more detail in Section 3.5. Section 3.6 concludes.

3.2 Institutional Background

Tax Deductions Taxpayers can reduce their taxable income, and thus lower their tax liability, by claiming deductible expenses. This paper focuses on deductions for work-related expenses, such as for commuting or work equipment. A deduction of x Euros reduces the taxable income by x Euros, and the tax liability by x times the marginal tax rate ($0 \leq MTR \leq 0.45$).

Standard Deduction By default, German taxpayers are entitled to a standard deduction of 1,000 Euros if they have wage income in a given calendar year. Wage taxes are withheld at source, and each month 1/12 of the annual standard deduction is automatically factored in. If taxpayers are employed for $1 \leq x < 12$ months in a given year, they are granted the full deduction as soon as they file an income tax return.[1]

[1]If these taxpayers do not file, they overpay taxes that are not refunded to them (Hauck and Wallossek 2023).

Itemizing Taxpayers can itemize deductions to claim expenses that exceed the standard deduction threshold. Itemizing taxpayers are required to keep records for all their expenses throughout the year in order to be able to prove their actual expenses. The costs of itemizing include the time spent for record keeping, as well as the actual time for tax filing, i.e., entering the expenses in the tax filing forms that are shown in Figure 3.A.4.

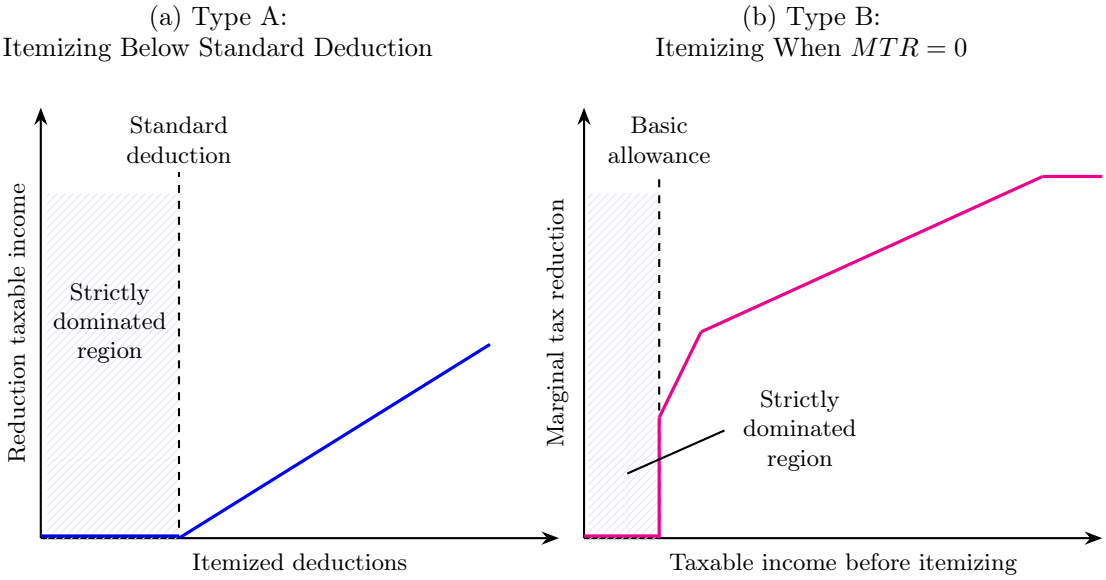


Figure 3.1: Strictly Dominated Itemizing Behavior

Notes: This figure illustrates the effect of itemizing deductions on taxable income (Panel a) and on the final tax liability (Panel b). Panel a plots the reduction in taxable income over the amount of itemized deductions. For any value up to the standard deduction, there is no effect and itemizing is strictly dominated by not itemizing. For any value above the standard deduction, there is a 1:1 reduction in taxable income, i.e., 1 additional Euro reduces the taxable income by 1 Euro. Panel b plots the marginal tax reduction, i.e., the amount by which the final tax liability is reduced when reducing the taxable income by 1 Euro. This is equivalent to the marginal tax rate. Below the basic allowance threshold, for $MTR = 0$, reducing the taxable income has no effect on the tax liability that is already zero. For any value above the basic allowance threshold, there is a reduction of $0 < MTR \leq 0.45$.

Itemizing Mistakes I define an itemizing mistake as itemizing deductions although it does not reduce the tax liability. In these cases, itemizing has zero benefits but non-zero costs, and it is strictly dominated by not itemizing, which has zero benefits and zero costs. Put differently, itemizing in these cases can never be optimal for taxpayers. There are two types of mistakes, that I label Type A and Type B, respectively.

Type A Mistake Taxpayers make a mistake when itemizing with the total amount not exceeding the standard deduction threshold (Type A mistake). In this case, itemizing is strictly dominated by not itemizing, because it has no effect on the taxable income, since the taxpayer is entitled to the standard deduction in any case (Figure 3.1a). Itemizing can also be non-optimal above the standard deduction, since costs may exceed benefits (see Benzarti 2020). I focus on itemizing below the standard deduction, where

I can rule out that the observed behavior is optimal and can thus clearly identify a mistake in the taxpayers' behavior. I briefly discuss filing behavior above the threshold in section 3.5.

Type B Mistake In addition, taxpayers also make a mistake if they itemize while their tax liability is already zero before itemizing (Type B mistake). In this case, itemizing reduces the taxable income, but there is no tax reduction since the tax liability is zero in any case (Figure 3.1b). This applies to taxpayers at the bottom of the income distribution, whose taxable income, before itemizing, does not exceed the standard allowance, resulting in a zero MTR and zero tax liability.

3.3 Data and Sample

Dataset I use the *Taxpayer Panel* (TPP) provided by the Research Data Centre of the Federal Statistical Office and the Statistical Offices of the Länder (RDC 2022). The TPP is an annual panel of the years 2001 to 2018 that covers 5 percent of the total taxpayer population in Germany. I restrict my analysis to years 2010 to 2018, since the variables that inform about itemized deductions are only fully available from 2010 onwards.[2]

Sample Selection To study itemizing behavior, I restrict the sample to tax filers with wage income.[3] The sample consists of optional filers, who voluntarily decide to file, and compulsory filers, who have to file, e.g., because they have additional self-employed income, opted for joint withholding with their spouse, or because they received wage replacement payments such as unemployment benefits. I further restrict the sample to working-age individuals aged 18 to 65. For joint filing spouses, at least one spouse has to have wage income and both have to be in the working-age population in order to be included. Lastly, I drop observations with negative income.

Sample Characteristics The final sample for my analysis contains 1,135,647 taxpayers.[4] Column 1 of Table 3.1 provides the sample characteristics. The average taxpayer in the sample is 43 years old, has a total per capita gross income of 35,682 Euros, and is observed for 7 years. 44 percent are joint filing spouses and 43 percent of individual filers are women.

Variables Most importantly, the TPP data reports the granted deduction ($\geq$ standard deduction threshold), as well as the itemized deductions, irrespective of whether those exceed the standard deduction threshold. This allows me to identify itemizing mistakes. In addition, the TPP data covers a large set of income and income tax related

[2]Two federal states, Hesse and Baden-Wurttemberg, only started reporting itemized deductions in 2010. In addition, the information is only partly reported for years 2001 to 2007 in Hamburg, Bremen, Rhineland Palatinate, and Berlin.

[3]I exclude non-filers for two reasons. First, and most relevant, there is no itemizing option if one does not file. Second, because of the specific sampling procedure for the 5-percent sample, non-filers are undersampled in the TPP data set, not allowing to draw representative conclusions for this population (the sampling procedure is explained here: https://www.forschungsdatenzentrum.de/sites/default/files/tpp_2001-2018_on-site_stp.pdf). For a detailed analysis of non-filing in Germany based on cross-sectional income tax data, see Hauck and Wallossek (2023).

[4]Throughout this paper, I use the term *taxpayer* to refer to the tax unit. This can be either an individual, if not married or if married but filing individually, or a married couple if filing jointly.

Variable	(1) all filers	(2) itemizing	(3) mistake
Age	43	42	42
Female[1]	.43	.42	.43
East	.19	.18	.18
Joint filing spouses	.44	.44	.49
Income	35,682	36,238	34,266
Wage income only	.64	.67	.66
N raw	1,135,647	902,262	579,227
N weighted	26,027,774	22,713,015	14,456,042

Notes: This table summarizes the TPP sample characteristics using statistical weights provided by the Federal statistical office. Column 1 refers to the total sample of tax filers, column 2 refers to the subsample of filers who itemize, and column 3 refers to the subsample of those who itemize and make a mistake when itemizing. For joint filing spouses, I use the within-household mean if variables are reported on the individual level. Income in 2015 Euro.
[1] Refers to individual filers only.

Table 3.1: Sample Characteristics

variables as well as some demographic information. For income, I use the total income ("Summe der Einkünfte") and adjust all income values to 2015 Euros, using consumer price indices. For joint filing spouses, I use the mean income.

3.4 Documenting Itemizing Mistakes

3.4.1 Prevalence of Itemizing

Itemizing is common: 88 percent of taxpayers itemize at least once and 81 percent of all tax returns include itemized deductions. Conditional on itemizing, most taxpayers itemize repeatedly. The average taxpayer itemizes in 83 percent of observed filing years.

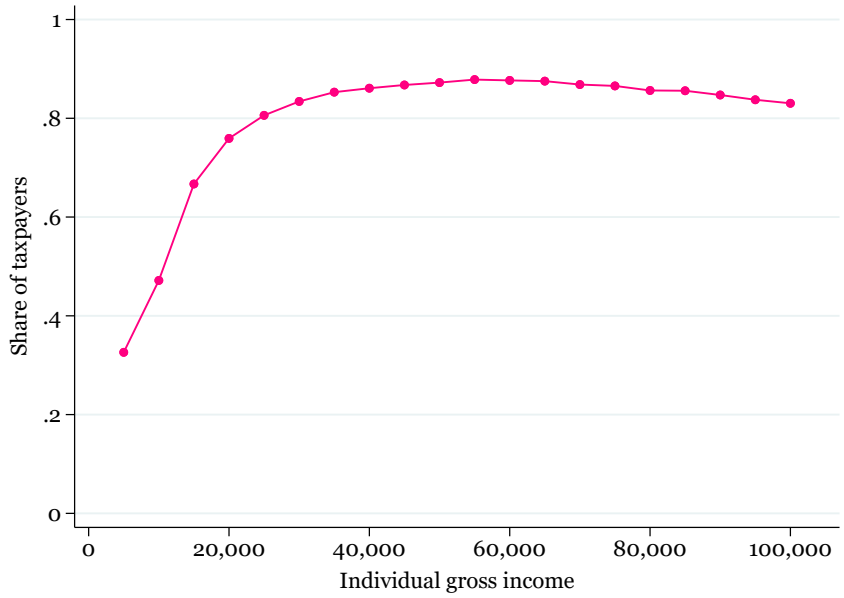
Itemizing is common across different demographic groups: comparing column 1 and 2 of Table 3.1 shows that, overall, the subsample of those who itemize is similar in observable characteristics compared to the the full population of tax filers.

Figure 3.2a shows that itemizing is least common for taxpayers with lowest income. The share of itemizers steeply increases with income at the bottom of the income distribution, before stabilizing above 80 percent for taxpayers with an income of 25,000 Euros or more. As a result, the income distribution of those who itemize is shifted to the right compared to those who do not itemize (Figure 3.A.1a). The income difference remains statistically significant when controlling for other characteristics such as age or gender (Table 3.A.1).

3.4.2 Prevalence of Mistakes

When taxpayers itemize, they often make a mistake by itemizing while it is strictly dominated by not itemizing. 57 percent of itemizing taxpayers have made at least one such mistake, representing 14.5 million taxpayers when employing statistical weights. They itemize although the sum of deductions does not exceed the standard deduction threshold (Type A mistake, 34 percent of returns) and/or although their tax liability is

(a) Itemizing



(b) Itemizing Mistakes

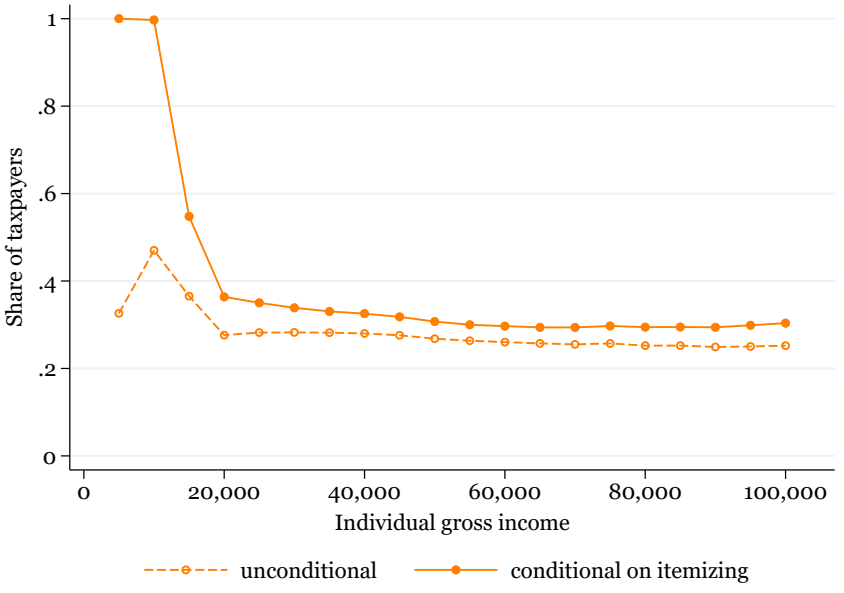


Figure 3.2: Itemizing and Itemizing Mistakes Over the Income Distribution

Notes: Panel (a) plots the share of itemizers among tax filers over the income distribution. Panel (b) plots the share of tax filers who make a mistake when itemizing over the income distribution. The dashed line plots the unconditional share for all tax filers, while the solid line plots the share conditional on itemizing. Income is defined as the sum of all income before itemizing and is binned in bins of 5,000 Euros. For joint filing spouses, income is the mean income. Weighted data.

already zero (Type B mistake, 4 percent of returns). In total, 36 percent of tax returns feature at least one of these itemizing mistakes.

Mistakes Over the Income Distribution Figure 3.2b shows that lowest-income taxpayers are most prone to making a mistake when itemizing (solid line). At the very bottom of the income distribution, itemizing is never optimal, because the marginal tax rate is zero, resulting in no benefit from itemizing (Type B mistake). As a result, 100 percent of itemizing taxpayers in these income bins make a mistake. Even when not conditioning on itemizing, low-income taxpayers are most likely to make a mistake (dashed line). Although the share of mistakes decreases further with higher income it remains strikingly persistent over large parts of the income distribution. Overall, the income distribution of those who make a mistake is shifted to the left as compared to those who make no mistake (Figure 3.A.1a).

3.5 Understanding Itemizing Behavior

Partial Tax Literacy The observed itemizing behavior suggests that many taxpayers have partial, yet incomplete, tax literacy that leads to ill-informed choices. Since they itemize, they clearly know that itemizing is possible and also have the practical knowledge of how to itemize deductions when filing. By reveled preferences, they also now that it can in general reduce the tax liability, because otherwise they would have no reason file. However, at the same time, they lack the tax knowledge to understand that, for their individual situation, itemizing is strictly dominated by not itemizing.

The fact that making a mistake when itemizing requires a certain level of tax literacy may also contribute to the patterns observed over the income distribution. Financial literacy has been shown to correlate with education and income (see Lusardi and Mitchell (2014) for a literature overview), suggesting that tax literacy is on average higher for higher income taxpayers. In line with this, I document that low-income taxpayers are least likely to itemize, but conditional on itemizing, they are most likely to make a mistake. However, I find that mistakes remain common for high income taxpayers as well, suggesting that these mistakes require a certain level of tax literacy. Similarly, I find patterns over the age distribution to be in line with hump-shaped financial literacy over the life cycle (see e.g., Lusardi and Mitchell 2011, 2014). Youngest and oldest taxpayers are least likely to itemize and, conditional on itemizing, most likely to make a mistake (Figure 3.A.2); yet mistakes remain common in the middle of the age distribution as well.

More Mistakes for More Complex Tax Returns I provide suggestive evidence that complexity increases the probability of making a mistake by showing that joint filing spouses are particularly prone to mistakes. 42 percent of tax returns from joint filing spouses feature an itemizing mistake, while only 29 percent do so for single filing taxpayers. Joint filing spouses' tax returns are mechanically more complex because they require information from two individuals. With respect to itemizing, the standard deduction applies to each individual and the itemizing decision of one spouse does not trigger an itemizing obligation for the other spouse. In addition, taxpayers whose tax returns are less complex because they have wage income only are also less likely to make a mistake when itemizing (see Table 3.A.1).

No Evidence for Prediction Mistakes One potential explanation for itemizing a total amount inferior to the standard deduction (Type A mistake) is uncertainty with respect to the sum of all itemized deductions. If taxpayers do not calculate the exact sum of their deductions before entering in the tax form, they might make a mistake when predicting their total deduction value before deciding to itemize. This type of mistake is expected to be more common when close to the standard deduction threshold. Consequently we would expect to see an increased mass of deductions just below the threshold. However, the distribution of itemized deductions (Figure 3.3) does not support this.

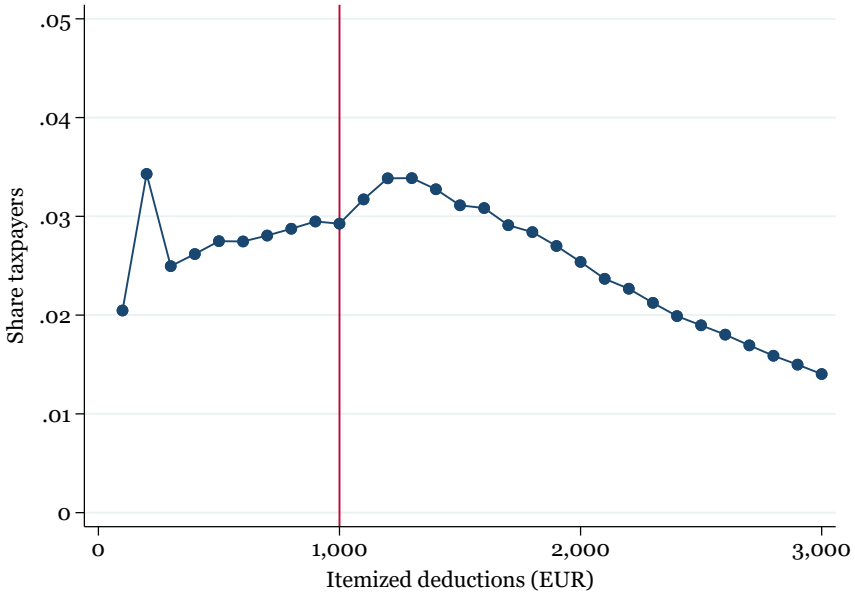


Figure 3.3: Distribution Itemized Deductions

Notes: This figure plots the distribution of itemized deductions for single taxpayers in 2011 to 2018, when the standard deduction was 1,000 Euros. Deductions are binned in 100-Euro bins. Weighted data.

No Evidence for Learning After filing an income tax return, taxpayers receive their final tax bill as soon as the tax authorities determine their final income tax liability. This document lists the itemized deductions as well as the granted deductions, typically on the first page (see Figure 3.A.5). If taxpayers itemized below the standard deduction (Type A mistake), they see that they received the standard deduction although they itemized. Even though taxpayers receive this information after making a mistake, I find no evidence for taxpayers learning from their mistakes. On the contrary, I find that, conditional on itemizing, taxpayers who make a mistake in $t-1$ are 17 percentage points more likely to make a mistake again in t (see regression results in Table 3.A.2). Taxpayers commonly repeat their mistakes: conditional on making a mistake at least once, the average taxpayer shows this behavior 3 times or for 61 percent of their tax returns.

Common Mistakes The peak for deductions in the 100-Euro bin in Figure 3.3 can be explained by taxpayers who itemize exactly the amount of deductions that, under common law, tax authorities typically accept without proof. Itemized deductions of 16 Euro are commonly accepted for account management fees and 103 or 110 Euro for work equipment. The peak in the distribution is driven by taxpayers who itemize exactly 16, 119 (= 103 + 16), or 126 (= 110 + 16) Euros. With no de facto requirement for record keeping, the cost of itemizing these amounts are limited, but the benefit remains zero as long as the sum of deductions remains below the standard deduction threshold. One potential explanation for this behavior is that taxpayers know about the common law, but do not understand that the benefit is zero if not exceeding the standard deduction threshold. An alternative explanation is that tax preparation software automatically fills in values for deductions, which is supported by anecdotal evidence. While values of 16, 119, and 126 are common, they do not drive my results. In total, less than 10 percent of mistakes are common mistakes (see Figure 3.A.3).

Missing Mass Above the Standard Deduction Above the standard deduction threshold, itemizing is optimal if the benefit exceeds the cost. The larger the total value of itemized deductions, the larger the benefit. Costs of itemizing above the threshold have a fixed cost component: no matter how much their deductions exceed the threshold, the taxpayer has to itemize the first 1,000 Euros. As a result, itemizing amounts just above the threshold is more likely to be dominated by the total costs. Benzarti (2020) documents a missing mass above the standard deduction threshold in the US, showing that taxpayers are willing to forgo refunds to avoid high costs of itemizing. The distribution of deductions in Figure 3.3 shows a similar pattern above the threshold, suggesting that, while many taxpayers make mistakes when itemizing, others behave rationally and refrain from itemizing because costs are too high.

3.6 Conclusion

I show that large parts of the taxpayer population itemize deductions even when it is strictly dominated by not itemizing. The prevalence of these itemizing mistakes suggests a significant lack of applied tax literacy, contributing to an ongoing discourse on income tax complexity and the challenges taxpayers face in navigating the income tax system. If policymakers want to help taxpayers in making optimal filing decisions, a straightforward approach would be to enhance taxpayers' tax literacy by adding information on tax forms.

My results also provide useful insights for future research on individual behavior in complex tax systems. Taxpayers' responsiveness to tax incentives is often limited and lacking tax literacy is commonly believed to play a key role. However, tax literacy is often hard to measure on the individual level with administrative data. The itemizing mistakes I document in this paper provide a direct measure of incomplete tax literacy. Employing this measure can be informative for future research and potentially help understanding heterogeneity in taxpayers' responsiveness to tax incentives.

Appendix to Chapter 3

3.A Additional Figures and Tables

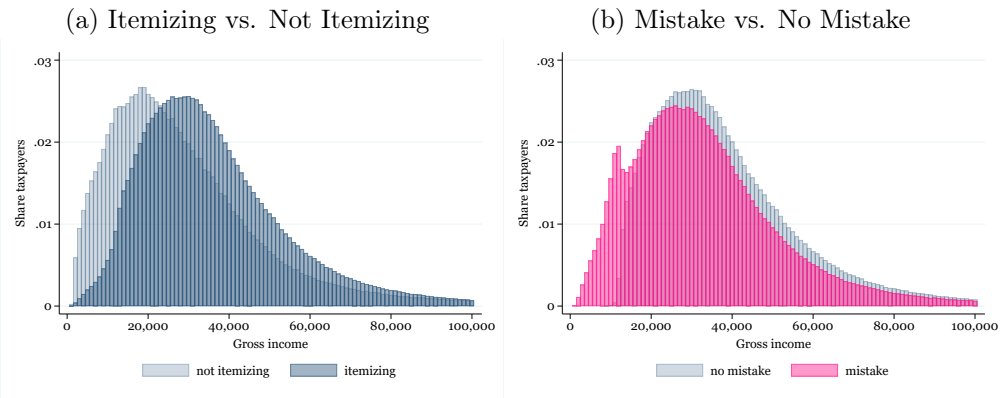


Figure 3.A.1: Income Distribution

Notes: This figure plots the share of taxpayers over the income distribution for different subsamples. Panel a plots compares those who itemize (dark blue) with those who don't (light blue). Panel b compares those who make a mistake (pink) with those who don't (light blue). Both groups in Panel b are conditional on itemizing. If the MTR is 0, itemizing is always a mistake, which explains the increased mass of taxpayers with a mistake for very low income levels. Income is defined as the sum of all income before itemizing and is binned in bins of 5,000.€ For joint filing spouses, income is the mean income. Weighted data.

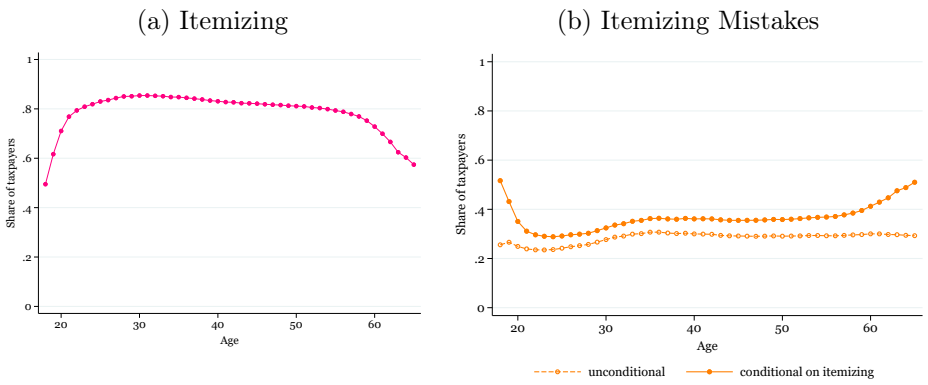


Figure 3.A.2: Itemizing and Itemizing Mistakes Over the Age Distribution

Notes: Panel (a) plots the share of itemizers among tax filers over the age distribution. Panel (b) plots the share tax filers, who make a mistake when itemizing, over the age distribution. The dashed line plots the unconditional share for all tax filers, while the solid line plots the share conditional on itemizing. For joint filing spouses, age is the rounded mean age. Weighted data.

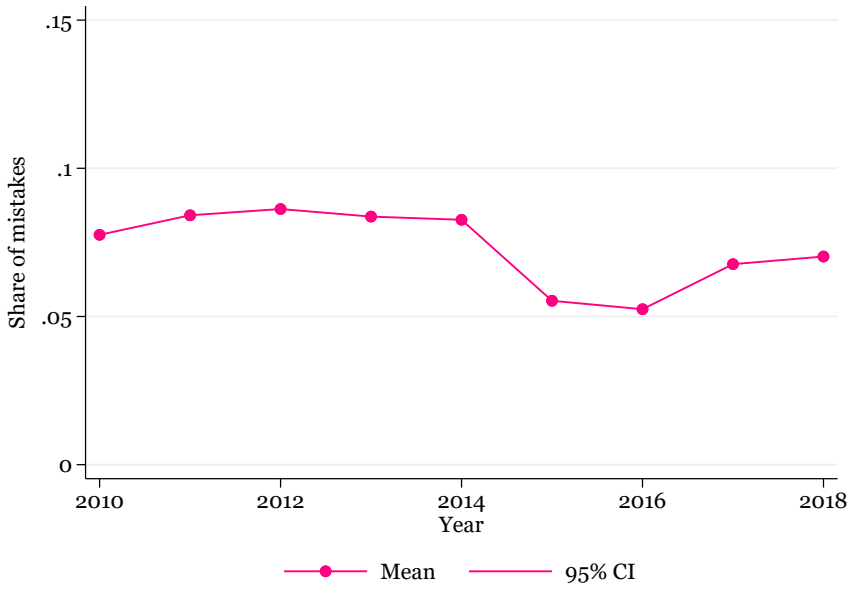


Figure 3.A.3: Share of Common Mistakes Among All Mistakes

Notes: This figure plots the share of common mistakes among all itemizing mistakes over time. A common mistake is defined as itemizing with the total sum of itemized deductions being exactly 16, 119, or 126 Euro. Weighted data.

(a) Page 1: Wage Earnings

(b) Page 2: Deductions I

Figure 3.A.4: Tax Filing Forms Wage Income and Deductions

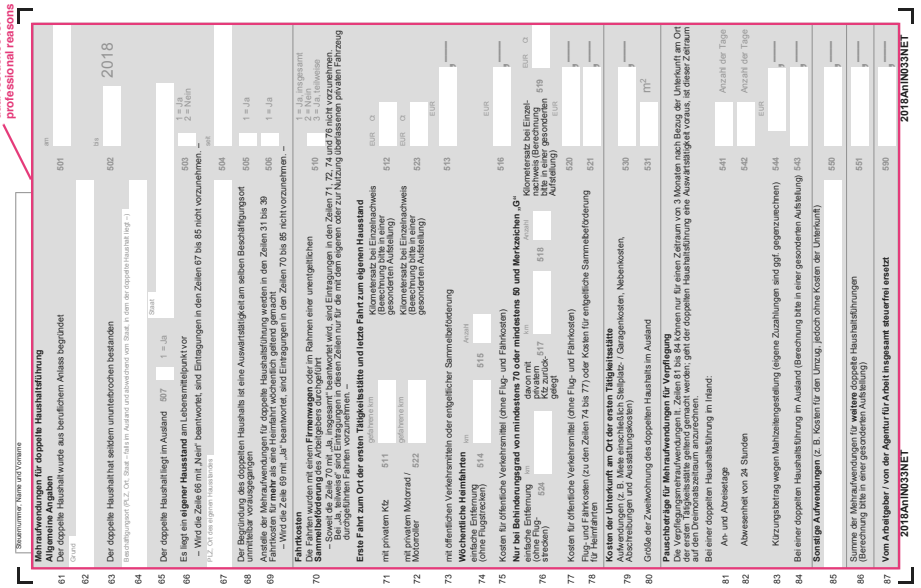


Figure 3.A.4: Tax Filing Forms Wage Income and Deductions (cont.)

Notes: This figure shows the tax filing forms for wage earnings. Page 1 covers information related to the wage earnings, such as gross wage earnings, taxes withheld, etc. Page 2 to 4 cover work-related deductions. Explanations added by the author.

	Itemizing in t		Mistake in t	
	(1)	(2)	(3)	(4)
Joint spouses	0.039***	0.022***	0.085***	0.118***
	(0.001)	(0.003)	(0.002)	(0.005)
Single male	0.019***	-0.026***	-0.084***	-0.061***
	(0.001)	(0.003)	(0.001)	(0.005)
Total income	0.000***	0.000***	-0.000***	-0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
Wage income only	0.122***	0.051***	-0.013***	-0.018***
	(0.001)	(0.001)	(0.001)	(0.001)
Age	0.018***	0.039***	-0.008***	-0.017***
	(0.000)	(0.000)	(0.000)	(0.000)
Age squared	-0.000***	-0.000***	0.000***	0.000***
	(0.000)	(0.000)	(0.000)	(0.000)
East	0.003***	-0.012***	-0.007***	0.005
	(0.001)	(0.003)	(0.001)	(0.004)
Constant	0.382***	-0.180***	0.511***	0.604***
	(0.005)	(0.007)	(0.007)	(0.010)
Fixed effects	No	Yes	No	Yes
R-squared	0.034	0.672	0.025	0.616
N	6,895,125	6,790,208	5,139,146	5,039,206

Notes: This table shows the $\hat{\beta}^J$ estimates from estimating linear regressions of the form $y_{it} = \alpha + \sum_J \beta^J X_{it}^J + \delta_i + \epsilon_{it}$, where y_{it} denotes the filing behavior of individual i in year t, X_{it}^J denotes taxpayer characteristic J as listed in the table, and δ_i is a taxpayer fixed effect. For columns (1) and (2) the outcome is a dummy that is 1 if i itemizes in t and 0 else. For columns (3) and (4) the outcome is a dummy that is 1 if i makes an itemizing mistake in t and 0 if itemizing without a mistake. Columns (1) and (3) report results from estimating the regression without taxpayer fixed effects, while columns (2) and (4) include fixed effects. Since taxpayers may change their marital status during the sample period, moving from single to married or vice versa, *joint spouses* and *single male* are not omitted when including taxpayer fixed effects.
Significance level: *** 0.01; ** 0.05; * 0.1. Robust standard errors are reported in parentheses. Weighted data, N = raw sample size.

Table 3.A.1: Itemizing Behavior and Individual Characteristics

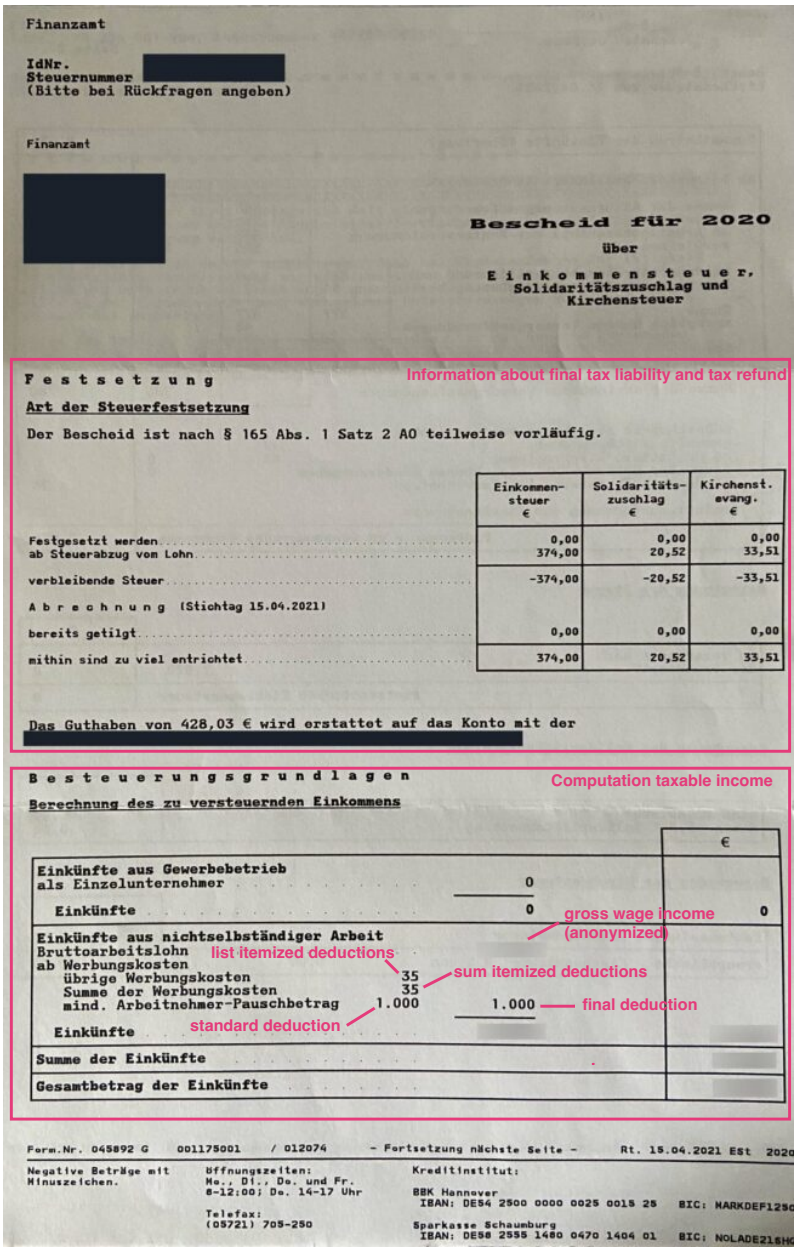


Figure 3.A.5: Example First Page Final Income Tax Bill

Notes: Example of the first page of a final income tax bill. Anonymized picture from a website that offers an online tool for tax filing (https://www.smartsteuer.de/online/steuerwissen/steuerbescheid/, last accessed: 2024-03-05), explanations added by the author. In this example, the taxpayer itemized 35 Euros for "other deductions" and was granted the full standard deduction of 1,000 Euros.

	(1) Itemizing in t	(2) Mistake in t
Mistake in $t-1$	-0.028***	0.170***
	(0.000)	(0.001)
Total income (000s)	0.000***	-0.000***
	(0.000)	(0.000)
Constant	0.972***	0.292***
	(0.000)	(0.001)
Individual FE	Yes	Yes
Time FE	Yes	Yes
R-squared	0.321	0.657
N	4,244,254	4,069,168

Notes: This table shows the $\hat{\beta}$ and $\hat{\gamma}$ estimates from estimating linear regressions of the form $y_{it} = \alpha + \beta mist_{it-1} + \gamma inc_{it} + \delta_i + \mu_t + \epsilon_{it}$, where y_{it} denotes the filing behavior of individual i in year t, $\beta mist_{it-1}$ is a dummy that is 1 if i made an itemizing mistake in $t-1$ and 0 if they itemized with no mistake, inc_{it} is the gross income, δ_i is a taxpayer fixed effect, and μ_t is a time fixed effect. For columns (1) and (2) the outcome is a dummy that is 1 if i itemizes in t and 0 else. For columns (3) and (4) the outcome is a dummy that is 1 if i makes an itemizing mistake in t and 0 if itemizing without a mistake.
Significance level: *** 0.01; ** 0.05; * 0.1. Robust standard errors are reported in parentheses. Weighted data, N = raw sample size.

Table 3.A.2: Itemizing Behavior After Itemizing Mistakes

Chapter 4

Opt-in or Opt-out? The Effect of Defaults on Public Pension Enrollment*

4.1 Introduction

How to improve individual choices is a key question in both economic research and policy making. One way that policy makers can direct choices is by setting defaults. The application that probably attracted most attention in this context is savings for retirement. Interest has been growing over the past two decades since retirement plans increasingly build on defined contributions that typically feature default options both in terms of contributions as well as investment strategy (Benartzi and Thaler 2007). The findings of the literature are very clear: many individuals show passive saving behavior, they are prone to inertia, and therefore stick to the default if enrolled automatically (e.g., Blumenstock, Callen, and Ghani 2018; Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen 2014; Choi, Laibson, Madrian, and Metrick 2004; Madrian and Shea 2001).

This paper shows that in a setting where the choice menu is simple, the effect of the default on enrollment is smaller than what is typically observed in more complex settings. After the introduction of automatic public pension enrollment for low-income employees in Germany, the enrollment share in the first month of employment increases by 23 percentage points. However, a majority of 70 percent opt out immediately and permanently remains not enrolled under automatic enrollment. Using administrative data, we emphasize the heterogeneity of choices that is hidden behind these aggregate numbers. On the one hand, individuals react to disincentives, i.e., they are less prone to the default if their enrollment costs are higher. On the other hand, the power of the default is stronger for individuals with presumably lower financial literacy, potentially nudging them into enrollment although it is not optimal for them.

Enrollment in the German public pension insurance is optional for employees below a certain income threshold – so-called *mini-job* employees. All mini-job employees face the same binary choice: to enroll or not to enroll. There is no choice of the contribution rate level and since the German public pension system is a pay-as-you-go system, there are no funds to choose for investment. This binary choice menu is much simpler than most other settings for automatic enrollment that often include multi-dimensional and continuous options, e.g. 401(k) savings plans.

We examine the effect of auto-enrollment by studying a natural experiment. A reform in 2013 introduces a change in the default enrollment status for new mini-jobs, going from an opt-in to an opt-out regime with automatic enrollment. Building on administrative panel data from the German Pension Insurance, we employ a regression-discontinuity design (RDD) to identify the causal effect of automatic enrollment in the binary choice

*This chapter is based on co-authored work with Tabea Bucher-Koenen and Joachim Winter.

setting. We find that after the reform, enrollment for new mini-jobs increases by 23 percentage points in the first month. Given that the pre-reform enrollment share is about 5 percent, this effect is sizable. However, for the majority of individuals, the change in default does not affect their enrollment with about 70 percent opting-out immediately. The effect decreases with length of employment but remains economically and statistically significant in the medium run: after 12 months in a mini-job, the enrollment share is still about 14 percentage points higher under automatic enrollment.

We find heterogeneous effects of the default across different demographic groups. Automatic enrollment has a stronger effect on younger individuals, non-German citizens, women, and those who live in the Eastern part of Germany. In addition, the effect is less strong for those who have a longer mini-job employment history. These patterns are in line with patterns of financial literacy, which has been shown to be lower for women, for young individuals, and in East Germany (Bucher-Koenen and Lusardi 2011; Bucher-Koenen, Lusardi, Alessie, and Rooij 2017; Lusardi and Mitchell 2014). Furthermore, non-German citizens are likely to be less familiar with the specific context, while individuals with a longer mini-job employment history are likely more knowledgeable. From a policy perspective, it is important to understand both limits and heterogeneity of the power of defaults when designing policies aiming to enhance old-age savings, especially as those most likely to adhere to defaults are not necessarily those for whom enrollment is most beneficial.

To better understand the high prevalence of opt-outs, we analyze the enrollment behavior of individuals observed under both default regimes. We show that a large share of mini-job employees are so-called "never takers": two thirds never enroll, irrespective of the default. Truly passive behavior, defined as always sticking to the default, is observed for less than 15% of the mini-job employees. Furthermore, we show that a significant share of mini-job employees seems to understand and react to enrollment incentives in general by analyzing discontinuities in incentives at waiting period thresholds that are decisive for pension eligibility.

This paper contributes to the literature on the impact of default options on individual behavior and in particular to the literature on auto-enrollment in retirement saving and pension schemes. A significant share of this literature studies choices in employer sponsored defined-contribution plans in the US (401(k) plans). Starting with Madrian and Shea (2001), several studies have investigated the effect of auto-enrollment in 401(k) plans, e.g., Choi, Laibson, Madrian, and Metrick (2004), Bernheim, Fradkin, and Popov (2015) or Beshears, Choi, Laibson, Madrian, and Wang (2016).[1] This literature commonly finds that default setting has a strong impact on individual 401(k) savings behavior. Under automatic enrollment, employees are more likely to contribute to a 401(k) plan and if they contribute, they typically stick to the default contribution rate and invest in the default funds. Strong effects are also found in other settings, e.g., Australia (Butt, Donald, Foster, Thorp, and Warren 2018) or Afghanistan (Blumenstock, Callen, and Ghani 2018).

The power of defaults has been documented in settings with complex choice menus. In this paper, we investigate the impact of default setting for a very simple choice menu.

[1] See Clark and Pelletier (2019) for a more detailed overview of the findings of the default literature in the 401(k) setting.

This allows to disentangle the pure impact of default setting from other factors that are likely at play in situations where individuals face more complex choices. For instance under 401(k) plans, employees do not only choose whether or not to enroll, but also how much to contribute and where to invest their contributions. If the choice problem is multi-dimensional or the number of options is large, choice overload may contribute to the default stickiness (Iyengar and Lepper 2000, Goda, Levy, Manchester, Sojourner, and Tasoff 2020). Madrian and Shea (2001) argue that parts of the observed default stickiness also stem from employees' interpretation of the default as their employer's investment advice. Additionally, switching costs can account for inertia as well (Gabaix 2019, Heiss, McFadden, Winter, Wuppermann, and Zhou 2021).

In the setting we study in this paper, none of these complications arise. The employees' choice menu is simple: to enroll or not. Choice overload is thus unlikely to explain default stickiness. Since the default is set by the federal government and not the employers, the default cannot be misinterpreted as their advice either. In addition, the immediate cost for opting out from default enrollment for mini-jobs in the pension insurance is comparatively low (there is only a standard form that has to be filled in; see Figure A.6 for an example), making switching costs an unlikely explanation preventing individuals from opting out.

This paper also adds to the literature on passive savings behavior. We find that only a minority of individuals show truly passive behavior, which we define as always sticking to the default irrespective of what the default is. The majority of mini-job employees is best described as never takers, who never enroll, irrespective of the default. This is in contrast to findings from Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen (2014), who find that the vast majority of individuals in the Danish context can be described as passive savers, who do not respond to subsidies nor adjust their savings outside their pension accounts when facing changes in contribution rates. One explanation is that passive behavior is not necessarily an individual trait but depends on the default (Goda, Levy, Manchester, Sojourner, and Tasoff 2020).

The remaining part of the paper is structured as follows. We give an overview of the institutional background and the reform that changed the enrollment default for mini-job employees in section 4.2 and introduce the data set we use in section 4.3. We analyze the effect of the default on the public pension enrollment of mini-job employees in section 4.4. In section 4.5 we shed some light on the drivers of the observed individual behavior. We conclude with section 4.6.

4.2 Institutional Setting

4.2.1 German Statutory Pension Insurance

The statutory pension system in Germany is an earnings related pay-as-you-go (PAYG) system with compulsory enrollment for most employees.[2] Both employees and employers make contributions with contribution rates being defined by law on the national level.

[2]Civil servants and most of the self-employed are excluded, but self-employed may enroll deliberately. For certain self-employed, such as physicians or lawyers, occupation-specific plans are available.

Pension Points Contribution payments are translated into pension points that an insuree accumulates over their working life. An enrolled employee earning exactly the average annual income obtains 1 pension point, an employee who earns 50% (150%) of the average annual income obtains 0.5 (1.5) pension points and so on, with contribution being capped for incomes above a certain threshold. To a smaller extent, pension points can also be acquired during other periods such as parental leave or unemployment.

Pension points are monetized upon retirement and determine the level of the monthly pension payment. Their value mainly depends on the average labor market income at a given year, and the contribution rate for the working population as well as on a sustainability factor accounting for societal developments, e.g. demographic changes. It is adjusted on a yearly basis and has been steadily increasing over time (see Table A.1 for an overview of the relevant operands).

Waiting Periods In order to become eligible for a regular pension at the statutory retirement age, a minimum waiting period of five years is required. Eligibility for special pensions is tied to other minimum waiting periods. For instance, to become eligible for early retirement, 35 years of contributions are required. See Table A.2 for details.

4.2.2 Mini-Jobs

Mini-jobs are characterized by very low monthly gross income During our sample period, income is capped at 400 € for years before to 2013 and 450 € since then. Mini-jobs are exempt from income taxes and also from most social security contributions, including contributions to the public health insurance. Mini-jobs can serve as either a primary source of employment or as a supplementary side-job alongside regular employment. However, due to data limitations, this paper concentrates on the former group.

Employers' Contribution Contributing to the public pension insurance is compulsory for mini-job employers, irrespective of the employee's enrollment choice. Employers contribute at a fixed contribution rate τ_{er} of 15% for our sample period.[3] Note that employers always contribute the mandated τ_{er}, so there are no financial incentives for them to encourage or discourage their employees' enrollment. Since they pay contributions to the pension insurance anyways, we argue that administrative costs are also negligible for them. Furthermore, opting-in or opting-out only requires the employer to sign the respective one-page form (see Figure A.5 and A.6), imposing negligible compliance cost on them.

Individual Enrollment Choice For mini-job employees, enrollment in the statutory pension insurance is optional. They face a purely binary choice menu regarding their enrollment. They can either contribute a given τ_{ee} or not contribute at all. The employees' contribution rate is much smaller than the employers', with $\tau_{ee} = 3.9\%$ and $\tau_{er} = 15\%$ in 2013. If an individual enrolls, their total contribution rate is $\tau_{er} + \tau_{ee} = 18.9\%$ of their gross income as compared to $\tau_{er} = 15\%$ if not enrolled (values for 2013, see Table A.1 for changes over time).

[3] This is higher than for regular employment, where employer and employee each contribute at the same contribution rate, e.g. 9.45% in 2013 (see Table A.1). If employees are enrolled, the total contribution rate (employer + employee) is the same as under regular employment.

Default Mini-job employees are not required to make an active enrollment choice. As long as they do not actively choose to do the opposite, they will remain at the default enrollment status which is defined by the German government for all mini-jobs. Prior to 2013, the default was no enrollment and mini-job employees have to actively opt in for enrollment. With the 2013 reform, the default changed to automatic enrollment. Ever since, mini-job employees have to actively opt out of enrollment.

Reform The reform came into effect on January 1, 2013, after the respective law passed the German parliament on December 5, 2012. The new default with automatic enrollment applies to all mini-job employees whose employment starts in 2013 or later. Employees whose mini-job starts pre reform (2012 or earlier) remain under the old opt-in regime post reform. However, the new default applies to them if their income surpasses the former income threshold of 400 € or if they take up a new mini-job. As a second feature of the reform, the allowed income threshold for mini-jobs increases from 400 € to 450 € per month from 2013 onwards. This second feature attracted much public attention, while the change in default setting, key feature for our analysis, was perceived as a minor change.

It is important to note that the actual enrollment options are not affected by the reform and that there was also no change in incentives, neither for employees nor for employers. This makes the reform particularly interesting and suitable for investigating the impact of default setting on individual retirement savings behavior for very low income earners who face a simple choice menu.

Costs of Enrollment Enrollment comes at the monetary cost of contribution payments that are deducted from wage earnings and depend on the mini-job employee's income y_t and the respective τ_{ee} in period t. For instance, an employee with a monthly income of 450 € in 2013 faces monthly costs of 450 €×0.039 =17.50 € when contributing and 0 € else. There is a minimum assessment base y_{min} that ensures a minimum absolute contribution if enrolled. No matter how low their monthly income, the absolute contribution for enrolled individuals can never fall below $(\tau er + \tau ee) \times y_{min}$. For monthly income $y < y_{min}$, enrolled employees have to top up their regular contribution until this minimum absolute contribution is reached. In subsection 4.4.3, we investigate heterogeneity in enrollment behavior related to increased enrollment costs for this group.

In addition to monetary costs, there could be non-monetary costs for enrolling under the opt-in regime, i.e. time costs for filling in the required form. Since opting-in requires filling in a one-page form with easily accessible information only, we argue that compliance costs are negligible (see Figure A.5 for an example). Opting-out under automatic enrollment is equally simple (Figure A.6 shows an opt-out form).

Incentives for Enrollment There are two incentives for enrolling in the statutory public pension insurance. First, individuals acquire more earning-points and thus increase their future pension entitlements on the intensive margin when enrolling in the public pension insurance. Second, mini-job employees increase their insurance record when enrolling, which counts towards their waiting periods that may be decisive for public pension eligibility and thus pension entitlements on the extensive margin.

Pension Points The number of acquired earning-points for a given employment period is defined as

$$EP = \begin{cases} \frac{y}{Y} \text{ if enrolled} \\ \frac{y}{Y} \times \frac{\tau_{er}}{\tau_{ee}+\tau_{er}} \text{ if not enrolled,} \end{cases}$$

where y denotes the individual gross income for the employment period and Y denotes the average annual income for the respective year as defined by the pension insurance (see Table A.1 for details). Enrollment increases EP for the employment period by $\frac{\tau_{ee}}{\tau_{er}+\tau_{ee}}$, which is equivalent to an increase of 26% in 2013. This first incentive increases future pension entitlements on the intensive margin and is thus relevant for individuals who expect to be eligible for a pension in the future. By enrolling, they increase their pension entitlements at the intensive margin.

Waiting Periods While the earnings points determine the pension level, the pension eligibility depends on the individual waiting period. When enrolled, months employed in a mini-job are fully credited, i.e. one month of mini-job employment is equivalent to one additional month for the waiting period. Without enrollment, the credited waiting period depends on the income and is determined by $\frac{EP}{0.0313}$. This second incentive can increase future pension entitlements at the extensive margin for individuals below a relevant waiting period threshold. An extreme example would be an employee who was enrolled for 4 years and 11 months at some point in their life. Since their waiting period is less than 5 years, they are not eligible for any pension payments. However, by enrolling in a mini-job for only one more month, they will become eligible for monthly pension payment during their entire retirement period.[4]

We illustrate the two different enrollment incentives with an exemplary mini-job employee in section 4.B. In subsection 4.5.2, we analyze the individual enrollment behavior with respect to these (dis-)incentives and find that individuals (at least partly) react to those thresholds when deciding about their contributions.

4.3 Data

Dataset Our analysis is based on the VSKT data from the German statutory pension insurance ((Deutsche Rentenversicherung Bund 2017)). The VSKT is a monthly panel data set with information for the entire biography of a subsample of the universe of insurees born between 1949 and 2001. Our main variable of interest records the insuree's insurance status for every month of their working life, including not only enrollment through mini-job employment but also other periods, for instance periods of regular employment, unemployment or parental leave. We analyze the enrollment behavior for mini-job employees for whom their mini-job is their main employment. Demographic characteristics such as date of birth, gender, citizenship or region are available for the day of sampling, December 31, 2016. The same applies for the statistical weights that allow for drawing conclusions from the sample for the entire population.

Sample We restrict the sample to individuals for whom a mini-job is the main employment for at least one month between January 2011 and November 2016. We further

[4] In the PAYG system, the sum of the monthly payment depends on the income. If they earned the average income during the 4 years and 11 months, they acquired 4.9167 EP which is equivalent to a monthly payment of 138.36 € in 2013.

exclude mini-job observations if the reported average monthly income exceeds the applying upper bound threshold $\bar{y}$ for mini-job income.[5] Until 2012, $\bar{y} = 400\,€$, before it increases to $450\,€$ from 2013 onwards.

Furthermore, we restrict the sample to observations for which we can unambiguously observe the starting date of the mini-job employment, which is decisive for determining the applied default regime: all mini-jobs that started in or after 2013 are subject to the new default of automatic enrollment. Since the pension data stems from annual employer spells, the recorded starting date for a mini-job in year t never dates before January 1 of year t. Consequently, when observing a non-stop mini-job employment period that comprises the turn of a year, it is impossible to tell whether the individual remained in the same employment or whether they started a new mini-job on January 1. For our main analysis, we focus on the first month of mini-job employments, for which we can unambiguously identify the starting date, and thus the default.[6]

Sample Characteristics Table 4.1 provides sample characteristics for the final sample in the first, third, sixth, and twelfth month of employment, denoted by $m = 1, 3, 6, 12$. Our final sample comprises 337,109 monthly observations for $m = 1$, representing more than 25 million mini-jobs over the sample period when applying statistical weights. The number of observations decreases significantly with increasing m, because of the short average duration of mini-job employments.

	$m = 1$	$m = 3$	$m = 6$	$m = 12$
Female	0.61	0.63	0.66	0.69
Age	32	33	34	37
West Germany	0.85	0.86	0.86	0.87
German citizenship	0.79	0.81	0.82	0.83
Observations	337,109	201,808	119,554	57,353
Weighted	25,621,597	15,575,701	9,369,103	4,581,213

Notes: Number of new mini-jobs in the period 01/2011– 11/2016. Four points of observation for each employment: in the first ($m = 1$), third ($m = 3$), sixth ($m = 6$) and twelfth ($m = 12$) month of employment. Basic characteristics for the weighted sample. *Female*, *West Germany* and *German citizenship*: share of mini-jobs with this attribute. *Age*: mean age for a mini-job observation.

Table 4.1: Sample Characteristics

[5] The data does not contain information on the exact monthly income but the total income for an observed employment period of x months. For the average monthly income, we divide the total income by x. Mini-job income is allowed to exceed $\bar{y}$ up to 3 times per year if the annual mini-job income does not exceed $12 \times \bar{y}$. The data does not allow for disentangling these cases from reporting errors which is why we exclude those observations.

[6] A mini-job employment is considered to have its start in month t if the recorded starting date lies within that month but is not January 1. Mini-jobs with a recorded starting date of January 1 are only considered to have started in January if it is the first recorded mini-job employment for the individual or if their last mini-job employment ended before December 31 of the previous year.

4.4 Estimating the Impact of the Default

4.4.1 Empirical Strategy

We use a regression-discontinuity (RD) approach to determine the causal effect of the default on the enrollment, with the monthly starting date t of the employment being the running variable. The enrollment default is a deterministic function of t with a discontinuity at the turn of the year 2012/2013. All mini-jobs that start before 2013 are under the opt-in regime (no automatic enrollment) and all mini-jobs that start in 2013 or later are under the opt-out regime (automatic enrollment). We define D_t as a dummy variable for the default with

$$D_t = \begin{cases} 0 \text{ if } t < 01/2013 \text{ (pre reform)} \\ 1 \text{ if } t \geq 01/2013 \text{ (post reform).} \end{cases}$$

We estimate the effect of the default for an individual i that starts their mini-job in month t on their enrollment status in the m^{th} month of tenure in the employment that started in t. Thus, $m = 1$ refers to the first month of employment (t), $m = 3$ to the third month ($t+2$), and so on. We denote the individual enrollment status as E_{it}^m with $E_{it}^m = 1$ if enrolled and 0 else. We estimate the impact of the default on individual i's enrollment choice at four different points in time, months $m = 1$, 3, 6 and 12:

$$E_{it}^m = \alpha + \gamma\, t + \beta D_t + \eta_{it} \tag{4.1}$$

We refer to the effect of the default on enrollment in the first month of employment ($m = 1$) as *instantaneous effect* and to the effect on enrollment at later points in time ($m = 3, 6, 12$) as *medium-run effects*. The coefficient of interest, β, measures the estimated effect of automatic enrollment.

We then include a set of individual characteristics J of individual i in month t captured by the vector X_{Jit} and allow for different trends over time pre and post reform, for $t < 01/2013$ and $t \geq 01/2013$ respectively. The set of characteristics is described in detail in subsection 4.4.2. We use ordinary least squares (OLS) to estimate the full model, including X_{Jit} and allowing for different time trends:

$$E_{it}^m = \alpha + \gamma_{pre}\, t\, \mathbb{1}\,(t < 01/2013) + \gamma_{post}\, t\, \mathbb{1}\,(t \geq 01/2013) + \beta D_t + \sum_J \delta_J X_{Jit} + \eta_{it}\,. \tag{4.2}$$

As discussed in subsection 4.4.4, we replicate the analysis with non-linear Logit regressions. The results are very similar, with the marginal effects from the Logit regressions being close to the OLS estimates.

Identifying Assumptions The main identifying assumption is that assignment in the neighborhood of the cutoff (January 1, 2013) is as good as random, such that any discontinuity in the outcome at the threshold can be attributed to a discontinuity in the treatment variable. Put differently, we require that, absent the change in default, there would be no discontinuity in the enrollment share.

One concern in RD designs is that covariates other than the running variable may be discontinuous at the cutoff (Imbens and Lemieux 2008). To address this concern, we show monthly mean values for a set of covariates in Figure A.2, including age, gender, nationality, state of residence and past experience with mini-jobs. There is no evidence for discontinuities for these characteristics. In addition, panel f of Figure A.2 shows that there is also no discontinuity in the estimated error term $\hat{\eta}_{it}$ from Equation 4.2. We have no reason to suspect a discontinuity for any unobserved characteristics.

A second concern for RD designs is individuals' ability to manipulate the running variable, leading to non-random assignment around the cut-off (Imbens and Lemieux 2008; Lee 2008). In our setting, the running variable is the starting date of the mini-job employment. Clearly, employees as well as employers have leeway over the starting date of an employment contract. If there was manipulation around the reform, we would expect a discontinuity in the density of contracts around the cutoff. Figure A.1 shows that, while there are clear seasonal patterns, the number of new mini-jobs in the months in 2012 and 2013 are comparable to the years before and after. This in line with the fact that there is no incentive for manipulation, neither for employees nor for employers. While the default changes at the cutoff, costs and benefits of being enrolled remain unchanged and there are no changes in the enrollment incentives.

4.4.2 Overall Effect on Enrollment

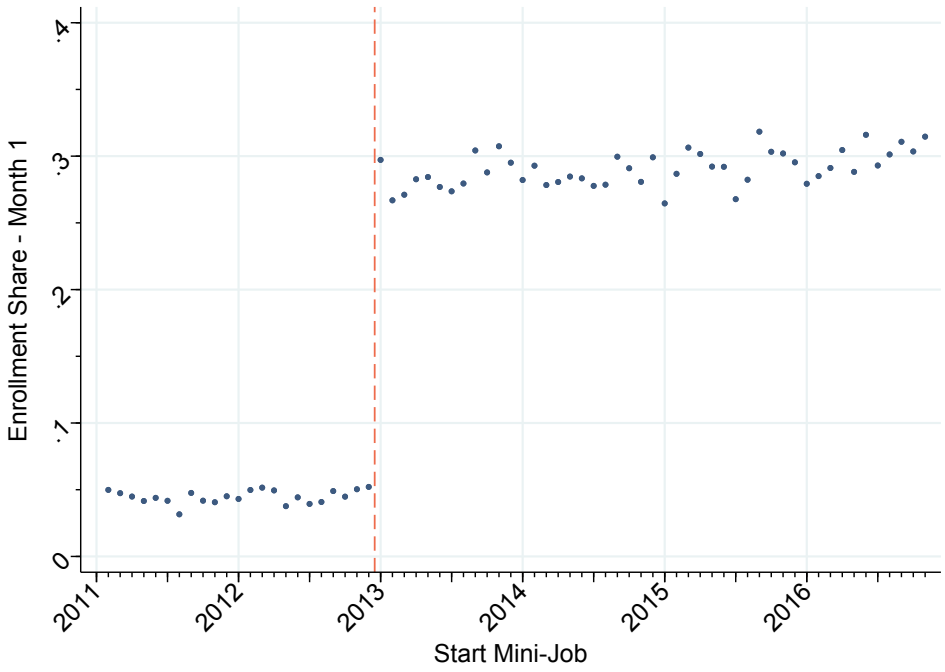


Figure 4.1: Enrollment Share – Unconditional Means

Notes: Weighted data. Scatter plot displays the average enrollment share in the first month of the mini-job for individuals who started their mini-job in a given month t. The corresponding absolute numbers are shown in Figure A.1.

Instantaneous Effect Figure 4.1 provides graphical evidence that introducing an opt-out regime has a positive effect on enrollment. The graph plots the unconditional enrollment share in the first month of a new mini-job ($m = 1$). Under the opt-in regime pre reform, the enrollment share is about 5% for all starting months of employment. Starting with automatic enrollment in January 2013, the enrollment share jumps to a share of about 30% and stabilizes at that level for all post-reform months with a slight increase over time.

While Figure 4.1 shows a clear increase in enrollment post reform, it also shows that the majority of mini-job employees are not affected by the default. Both pre and post reform, most individuals are not enrolled. Put differently, under automatic enrollment, about 70% opt out immediately. Table 4.2 reports the results from estimating Equation 4.1 and 4.2 and confirms the graphical evidence from Figure 4.1. Automatic enrollment significantly increases the enrollment share by about 23 percentage points in the first month of employment. Both magnitude and significance remain unchanged over different specifications reported in columns (1) to (4).

	(4.1)	(1a)	(1b)	(4.2)
D_t	0.2301***	0.2328***	0.2244***	0.2272***
	(0.0031)	(0.0031)	(0.0047)	(0.0047)
t	0.0005***	0.0004***		
	(0.0001)	(0.0001)		
t_{pre}			0.0001	0.0001
			(0.0001)	(0.0001)
t_{post}			0.0005***	0.0004***
			(0.0001)	(0.0001)
cons	0.0385***	0.0477***	0.0426***	0.0517***
	(0.0014)	(0.0027)	(0.0020)	(0.0031)
X_{Jit}	No	Yes	No	Yes
R2	0.0794	0.1093	0.0794	0.1094
N	337109	333707	337109	333707

Notes: Effect on the enrollment E_{it} of individual i in the first month of their mini-job in month t. Coefficients from the OLS estimation of Equation 4.1 to 4.2, robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. t: month of observation, 1/2011 = 1. t_{pre}: pre-reform months. t_{post}: post-reform months. D_t: Dummy for the default, $D_t = 1$ under auto-enrollment and 0 else. *cons*: constant from the OLS estimation. X_{Jit}: vector of individual control variables. Controls include age, age squared, gender, citizenship, east/west and experience with mini-job employment in the past. See Figure A.2 for mean values for the control variables over time. Table A.3 replicates the results using Logit.

Table 4.2: Instantaneous Effect of Automatic Enrollment

Medium-Term Effects To better understand the impact of automatic enrollment on public pension entitlements, we widen the time horizon and analyze the medium-term effect of automatic enrollment. We do so by tracking individuals over 3 different points in the first year of employment, in the third, sixth and twelfth month ($m = 3, 6, 12$). Intuitively, by increasing m we decrease our the sample size, as employment contracts

can end before 3, 6 or 12 months. Furthermore, assuming that at least some individuals are partly inert and take some time to deviate from the default and to actively opt in (under the old default) or opt out (under automatic enrollment), we expect β to decrease as m increases.

Widening the time horizon comes at the cost of loosing precise information. As soon as we track individuals for $m > 1$ months, the employment history will include a turn of the year for at least some individuals. For $m = 3$ for instance, every individual who started their job in November or December is now observed in the next year (January and February, respectively). While the starting date is still observable for those who started their employment in the same calendar year (group 1), we lose this information for everyone else, because we cannot distinguish individuals who remained in the same job over the turn of the year from those who started a new mini-job at the beginning of the new year. For employment periods over the turn of a year pre or post reform, we know the default but there remains uncertainty about the precise length of their current employment (group 2). For individuals whose employment period includes the turn of the year 2012/2013, we can no longer determine the default (group 3).

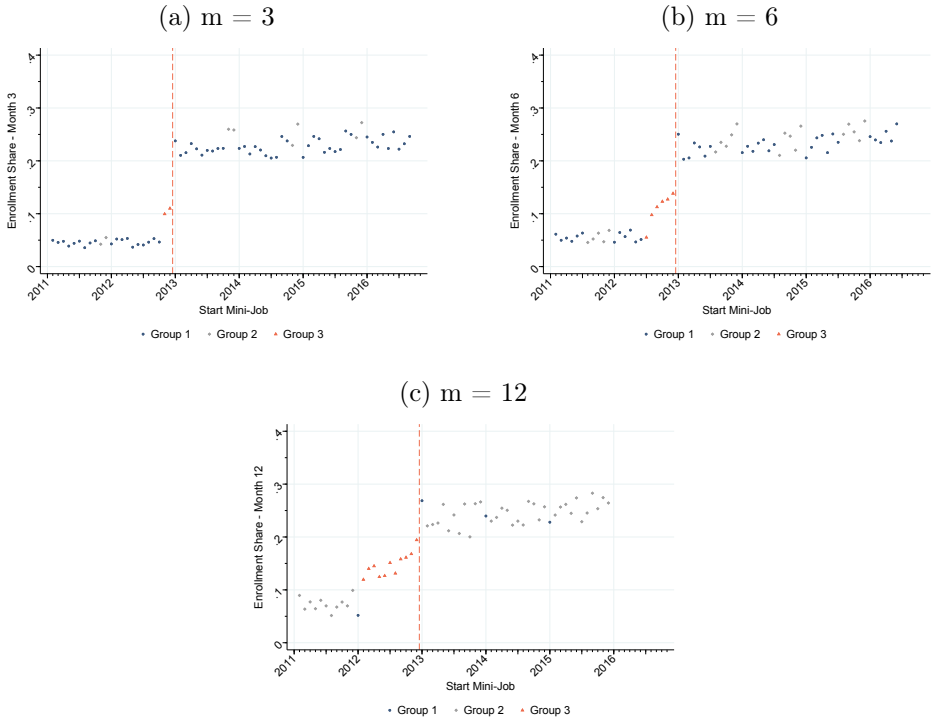


Figure 4.2: Enrollment Share in the Medium Run – Unconditional Means

Notes: Weighted data. Scatter plots display the average enrollment share E_t^m in the m^{th} month of the mini-job for individuals who started their mini-job in month t over that t. *Group 1*: no uncertainty about the default or the starting date of the employment. *Group 2*: no uncertainty about the default but uncertainty about the starting date. *Group 3*: uncertainty about both the default and the starting date. In each panel, a marker at given t contains the same group of individuals, exclusive those who dropped out of their mini-job employment before reaching the m^{th} month of this employment. For example, an individual who is employed for four months only, is only considered for Panel a. The corresponding absolute numbers are shown in Figure A.1

Figure 4.2 shows a clear positive medium-term effect of auto-enrollment: over 20% of individuals that are employed in a mini-job for at least 3, 6 or 12 months (panel a, b and c, respectively), are enrolled for all m and over the entire post reform period. This is in line with some individuals being partly inert and thus opting out only with some delay. Pre-reform enrollment shares increase with m, indicating that those who are employed in a mini-job for a longer time are more likely to actively opt in if not enrolled automatically.

Figure 4.2 also documents a clear increase in enrollment for some months right before the reform date and the number of months is increasing with m. To be precise, there is an increased enrollment share if mini-jobs that started in this pre-reform month are observed post reform after m months (group 3). For $m = 3$, this concerns the last 2 months of 2012, for $m = 6$ the last 5 months and for $m = 12$ all months in 2012 except for January. As described above, we cannot distinguish between individuals who remained in their old mini-job under the old default and those who started a new mini-job under automatic enrollment on January 1, 2013. The increased enrollment shares pre reform are in line with some individuals starting a new mini-job under the new default (see subsection 4.4.4 for more details).

We exclude group 3 for estimating the medium-term impact of automatic enrollment because we do not know their default with certainty. We deviate from the standard RD design here because the discontinuity does no longer occur between adjacent months.

	m = 3		m = 6		m = 12	
D_t	0.1648***	0.1667***	0.1484***	0.1483***	0.1344***	0.1359***
	(0.0061)	(0.0059)	(0.0086)	(0.0083)	(0.0150)	(0.0145)
t_{pre}	0.0001	0.0000	0.0001	0.0001	-0.0006	-0.0001
	(0.0002)	(0.0002)	(0.0004)	(0.0004)	(0.0010)	(0.0010)
t_{post}	0.0005***	0.0004***	0.0007***	0.0007***	0.0008**	0.0008**
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
cons	0.0448***	0.0272***	0.0543***	0.0244***	0.0758***	0.0266**
	(0.0027)	(0.0038)	(0.0042)	(0.0055)	(0.0085)	(0.0103)
X_{Jit}	No	Yes	No	Yes	No	Yes
R2	0.0531	0.1187	0.0462	0.1191	0.0356	0.1187
N	196696	194893	110173	109306	47297	46923

Notes: Effect on the enrollment E_{it}^m of individual i in the m^{th} month of their mini-job in month t. Only observations without uncertainty about the default (excluding group 3 from Figure 4.2). Coefficients from the OLS regression specified in Equation 4.2 with (right columns) and without individual characteristics (left columns). Robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. t: month of observation, 1/2011 = 1. D_t: Dummy for the default, $D_t = 1$ under auto-enrollment and 0 else. *cons*: constant from the OLS estimation. X_{Jit}: vector of individual control variables. Controls include age, age squared, gender, citizenship, east/west and experience with mini-job employment in the past. Intuitively, the sample size N decreases with increasing m because not all individuals remain 3, 6 or 12 months in their mini-job employment and we exclude an increasing number of months with uncertain default information (group 3 from Figure 4.2). See Table A.4 for the same table without excluding group 3. Table A.5 replicates the results using Logit.

Table 4.3: Medium-Term Effects of Automatic Enrollment

Table 4.3 provides the results from the estimation of Equation 4.2 for $m = 3$, 6, and 12 and confirms the graphical evidence in Figure 4.2: automatic enrollment significantly increases enrollment shares in the medium run, but relative to the first month, the effect decreases over time, to about 14 percentage points after 12 months. Again, including individual characteristics (right columns) does not change magnitude or significance of the effects. Table A.4 provides results from estimating the medium-term effect including all observations.

There are two explanations for the difference between the instantaneous and medium-run effect. First, inertia delays the response and individuals take some time to deviate from the default. Using within-subject variation in the default, we investigate the prevalence of such partly inert behavioral types in subsection 4.5.1 and find that between 5 and 8 percent of individuals are either delayed never takers or delayed always takers who take some time to deviate from the default. A second explanation is attrition. Individuals who remain longer in the sample because they have longer employment periods are more likely to make an active choice and deviate from the default, both pre and post reform. With increasing m, short-time employees drop out of the sample and longer-term employees account for larger parts of the sample which mechanically increases the enrollment share over m.

4.4.3 Heterogeneous Effects of the Default

We now analyze the effect of the default in more detail and investigate heterogeneity in the impact of automatic enrollment. To account for potentially heterogeneous effects, we interact the default dummy D_t with individual characteristics X_{Jit}, thereby allowing the treatment effect of the default to be heterogeneous. Adding the interaction terms $D_t \times X_{Jit}$ to Equation 4.2 yields:

$$\begin{aligned} E_{it}^{m} = \alpha + \gamma_{pre}\, t\, \mathbb{1}\,(t < 01/2013) + \gamma_{post}\, t\, \mathbb{1}\,(t \geq 01/2013) \\ + \beta D_t + \sum_J \delta_J X_{Jit} + \sum_J \zeta_J D_t \times X_{Jit} + \eta_{it}, \end{aligned} \tag{4.3}$$

where ζ_J captures heterogeneity in the effect of automatic enrollment for different values of characteristic J.

Table 4.4 displays the results from estimating Equation 4.3. For each characteristic J, the table displays the estimated coefficient $\hat{\delta}_J$ as well as the coefficient from the interaction with the default $\hat{\zeta}_J$. The results show that there is significant heterogeneity in the impact of the default across demographic groups, while the overall effect of the default is persistent when allowing for heterogeneous effects.

Demographic Characteristics Understanding the interplay between automatic enrollment and different demographic characteristics is of great relevance for policy makers, since it allows for understanding the impact of the default for different groups of the population. For example, policy makers might be particularly interested in the default's impact on enrollment for women, who constitute the vast majority of mini-job employees with longer periods of mini-job employment (Table 4.1), while at the same time facing a substantial gender pension gap and an increased old-age poverty rate (OECD

	(1)	(2)	(3)
D_t	0.2640***	0.2546***	0.3235***
	(0.0061)	(0.0061)	(0.0061)
t_{pre}	0.0002	0.0001	0.0001
	(0.0001)	(0.0001)	(0.0001)
t_{post}	0.0006***	0.0006***	0.0004***
	(0.0001)	(0.0001)	(0.0001)
Age	0.0020***	0.0016***	0.0015***
	(0.0001)	(0.0001)	(0.0001)
Age $\times D_t$	-0.0016***	-0.0008***	-0.0009***
	(0.0001)	(0.0001)	(0.0001)
Female	0.0324***	0.0260***	0.0258***
	(0.0015)	(0.0015)	(0.0015)
Female $\times D_t$	0.0006	0.0110***	0.0081**
	(0.0028)	(0.0029)	(0.0028)
Non-German	-0.0287***	-0.0253***	-0.0266***
	(0.0013)	(0.0013)	(0.0013)
Non-German $\times D_t$	0.0202***	0.0137***	0.0112***
	(0.0025)	(0.0026)	(0.0025)
East	-0.0027	0.0031	0.0065*
	(0.0026)	(0.0026)	(0.0026)
East $\times D_t$	0.0370***	0.0274***	0.0320***
	(0.0045)	(0.0045)	(0.0044)
Experience		0.0006***	0.0006***
		(0.0001)	(0.0001)
Experience $\times D_t$		-0.0010***	-0.0011***
		(0.0001)	(0.0001)
HigherCost			-0.0422***
			(0.0016)
HigherCost $\times D_t$			-0.1418***
			(0.0027)
cons	-0.0363***	-0.0325***	-0.0142***
	(0.0030)	(0.0030)	(0.0029)
r2	0.0836	0.0847	0.1171
N	333707	333707	333707

Notes: Effect on the enrollment of individual i in the first month of their mini-job in month t. Coefficients from the regression specified in Equation 4.3, robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. t: month of observation, 1/2011 = 1. D_t: Dummy for the default, $D_t = 1$ under auto-enrollment and 0 else. *Age*: Age of individual i at the time of observation. $Female$: Dummy for the gender of i, $Female = 1$ if female and 0 else. $Non\text{-}German$: Dummy for the citizenship of i, $Non\text{-}German = 1$ for individuals without the German citizenship and 0 else. *East*: Dummy for the state of residence of i, $East = 1$ for East Germany and 0 else. *Experience*: Number of months with mini-job employment in the past. $HigherCost$: Dummy for very low-income employments with higher relative enrollment costs. *cons*: constant.

Table 4.4: Heterogeneous Effects of Automatic Enrollment

2019).[7].

The administrative data contains a set of basic individual demographics (age, gender,[8] citizenship, region) and Table 4.4 documents substantial heterogeneity in the response to the default for different demographic groups along these dimensions. The negative coefficient for the interaction term ζ_{Age} implies that the effect of automatic enrollment ($D_t = 1$) decreases with age. In terms of magnitude, being one year older dampens the effect of $D_t = 1$ by about 0.1 percentage points. The estimate for δ_{Age} however is positive, indicating that age is associated with higher enrollment for $D_t = 0$. The opposite pattern emerges for non-German mini-job employees. Compared to German employees, they are significantly less likely to enroll at the baseline but the effect of the default is larger for them. There are also systematic differences in the enrollment behavior of women and men. Women are in general more likely to enroll and automatic enrollment increases their enrollment more than for men (in absolute terms). The effect of the default is also higher for individuals living in the Eastern part of Germany.

Illustrating the Effect Heterogeneity To better understand the heterogeneity, we illustrate the effect of the default for two different hypothetical mini-job employees, person A and B. Person A is a long term mini-job employee, a 50 year old woman with 10 years of experience as a mini-job employee. Person B is a 25 year old man, who has never before worked in a mini-job. Both are in the first month of a new mini-job, live in the Western part of Germany, are German citizens, and have an income above the minimum assessment threshold y_{min}. We predict their enrollment just before and after the reform, for December 2012 and January 2013, based on the specification of Equation 4.3 with all characteristics as described in column 3 of Table 4.4. The predicted enrollment probability pre reform is 16.1% for A and 2.6% for B. post reform, the predicted enrollment probability rises to 32.3% for A and 31.9% for B. A has a higher enrollment probability pre reform, but the effect of the default is stronger for B, both in absolute and in relative terms.

Salience and Financial Literacy One potential explanation for the heterogeneous effects of the default on enrollment are differences in individuals' understanding of the default. If the default regime and the options for deviation are less salient for certain groups, those groups may be more likely to stick to the default. Similarly, differences in financial literacy might also explain differences in default stickiness.

The administrative data does not include a direct measure for salience or financial literacy. To overcome this, we first use past experience with mini-jobs as a proxy for salience, assuming that exposure to the mini-job system increases institutional knowledge and the salience of the default regime. Column 2 of Table 4.4 shows that a longer employment history with mini-jobs is associated with higher enrollment for $D_t = 0$, but the interaction with the default dummy is negative. This is consistent with the default being more salient for individuals who are more experienced with mini-job employments.

Financial literacy has been shown to differ across demographic groups. Financial liter-

[7]The average pension for women in Germany is 46 percent lower than for men and the income poverty rate at age 66 or older of 10.6% is increased by 43 percent as compared to 7.4% for men (OECD 2019).

[8]Following the gender records in the administrative data, we can only differentiate gender along the binary distinction of female and male.

acy is lower in East Germany (Bucher-Koenen and Lusardi 2011) and women are less financially literate than men (Bucher-Koenen, Lusardi, Alessie, and Rooij 2017).[9] Furthermore, in the German pension context, the specific knowledge of the pension system is arguably lower for non-German citizens who are likely to be less familiar with the setting and may face language barriers. The estimated interaction effects in Table 4.4 are in line with these financial literacy patterns.

Financial Incentives Heterogeneity in the power of the default holds significant importance for policy makers. Specifically, it raises the question whether automatic enrollment nudges "the right" individuals into enrollment. As individuals face different incentives for enrollment, it is to analyze heterogeneity in the costs and benefits of enrollment. For example, non-German mini-job employees may have lower incentives for enrollment if they are less likely to claim an old age pension in Germany, as they may only live in Germany for a limited time period.

While financial incentives are likely to play a role for enrollment, they are not observed on the individual level in the administrative data. However, we can identify a group that experiences higher costs: individuals with income below the minimum contribution threshold y_{min}. As described in section 4.2, enrolled individuals with income below a certain threshold have to top up regular contribution rates to meet an absolute minimal monthly contribution. All else equal, this decreases incentives for enrollment by imposing higher costs. We add a dummy for this group of individuals in column 3 of Table 4.4 and show that their enrollment share is much lower compared to individuals who face the regular τ_{ee}. The estimated interaction effect is by far the most sizable, it's absolute value amounts to almost half of the effect of D_t. This implies that individuals are more likely to deviate from the auto-enrollment default if facing larger financial incentives.

4.4.4 Robustness Checks

Unknown Default When we track the enrollment behavior of individuals for longer than the first month of their mini-job, we lose information whenever the time period contains the turn of a year. Figure 4.2 shows that enrollment behavior changes for those for whom the default is unknown. One explanation is that a sizable share of these individuals actually start a new job with the beginning of the new year and are thus automatically enrolled if the starting point is in the post-reform period.

With a simple back-of-the-envelope calculation, we can quantify the share of mini-jobs employees that would have to start a new mini-job on January 1, 2013, in order to explain the observed pattern. For $m = 3$, we see an average enrollment share of about 5% pre reform and about 23% post reform. For those who started their mini-jobs in December 2012, the enrollment share in the third month (February 2013) is about 9%. Assuming that this increase is only driven by a share x of individuals who started a new mini-job under auto-enrollment on January 1 2013, we need $x \cdot 0.23 + (1-x) \cdot 0.05 = 0.09$. Solving for x leads to $x = 0.22$. Thus, if 22% of these individuals started a new mini-job in January 2013, this would fully explain the increased enrollment share of 9% that we observe for these group in February 2013.

[9] Bucher-Koenen, Alessie, Lusardi, and Rooij (2021) find that a part of the gender gap in financial literacy can be explained by a gap in confidence rather than in knowledge, but the gap is persistent even when accounting for the differences in confidence.

To assess the plausibility of $x = 0.22$, we use additional data from the central agency that is handling administration for mini-jobs (*Minijob-Zentrale*).[10] We use the quarterly data from March 31, 2013 and look at the subsample of individuals that have started a mini-job in the past 180 days.[11] Out of these mini-jobs, the share of those that started within the past 90 days, thus post reform, is 77%. Having $x = 0.22$ for our subsample in February 2013 thus seems to be within a plausible range. Thus, we conclude that the observed patterns for the medium term are likely driven by individuals who actually started a new mini-job in January 2013, which is not observable in the data.

Change in Income Threshold With the 2013 reform, the income threshold for mini-jobs increased from 400 € to 450 €. This constitutes a potential confounder for our analysis if a higher income threshold increases the probability of enrollment. Two potential mechanisms are possible, but we provide suggestive evidence that none of them is at play.

First, individuals might respond to the increased income threshold by adjusting their labor supply at the extensive margin. In particular, individuals with a monthly reservation wage between 400 € and 450 € will supply labor after the reform, but not before. If those individuals are more likely to enroll (irrespective of the default), this would increase the post-reform enrollment share. If extensive margin responses were driving the observed enrollment patterns, we would expect to see a sizable increase in the number of mini-jobs after the reform. Figure A.1 documents that this is not the case. Compared to 2012, the monthly numbers of new mini-jobs is only somewhat higher in 2013 and 2014 and then decreases again in 2015 and 2016.

Second, abor supply at the intensive margin, coupled with liquidity constraints, could potentially lead to increased enrollment post reform. Assume that the 400 € threshold is binding for the labor supply of individual i, i.e., they would like to work more in their mini-job. Let's further assume that with a monthly income of 400 €, liquidity constrains hinder i from enrollment. In this case, the increased income threshold will increase i's labor supply post reform, potentially lifting liquidity constrains, and as a result, i may now enroll. In this scenario, i dopes not enroll because of the change in default but because they are no longer liquidity constrained.

Figure A.3 shows that the income distribution for mini-jobs peaks at 400 € pre reform, but there is also substantial mass below that threshold. This suggests that the threshold is not binding for most employees. Furthermore, many mini-job employees still have a monthly income of 400 € post reform, as indicated by the persistence of the peak at 400 € for post reform years.[12] To provide further evidence, we rerun the main analysis for a subsample of mini-jobs with monthly income $\leq$ 400 € post reform. The estimated effects are of comparable size as shown in Figure A.4 and Table A.6. We thus argue

[10] The data is not publicly available and was provided upon request.

[11] The data provides information on the mini-job tenure on a quarterly level. Tenure is categorized in multiples of 90 days (1– 90 days, 91– 180 days and so on) that do not necessarily coincide with the turn of a month and the information refer to all mini-job employees, while we are focusing on a specific subsample in our analysis.

[12] One explanation for the persisting income concentration at 400 € post reform are adjustment frictions for jobs that started pre reform, Employees (or employers) may also take the 400 € as reference point, because mini-jobs have long been referred to as "400- €-jobs", even though there are no financial reasons to do so. Seibold (2021) documents strong reference point dependence for German employees in the context of retirement age thresholds.

that the increased income threshold cannot explain the increased enrollment after the reform and does not constitute a threat for identification.

Logistic Regression In addition to the OLS estimation in the main analysis, we replicate our findings using logistic regression estimations (Logit). Table A.3 shows the corresponding marginal effects for the instantaneous effect ($m = 1$). The results are very similar to those from the OLS estimation as provided in Table 4.2. The estimated marginal effect of automatic enrollment is 0.24 compared to 0.23 in the OLS estimation. Table A.5 shows that for the medium term ($m = 3, 6, 12$), marginal effects estimated from Logit are somewhat larger than those from OLS as shown in Table 4.3. All effects remain highly significant at the 0.001 level.

4.5 Understanding Individual Behavior

We find that, even though there is a positive effect of automatic enrollment, only a minority of individuals stick to the default while the majority opts out from enrollment immediately. This is in contrast to findings from the existing literature, that typically documents enrollment for the majority of individuals who are enrolled automatically (e.g., Blumenstock, Callen, and Ghani 2018; Cribb and Emmerson 2020; Madrian and Shea 2001). To better understand the comparably low enrollment rates, we study individual enrollment behavior in more detail.

4.5.1 Behavioral Types of Enrollment

When assessing the impact of default setting or other retirement savings policies, individuals are often classified into active and passive savers. This terminology has been heavily influenced by Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen (2014). Following their terminology, active savers are typically defined as individuals whose behavior is in line with standard economic models and who react to changes in savings incentives (e.g., subsidies for pension contributions). Active savers are not expected to change their savings behavior in response to a change in the default as long as incentives remain unchanged. Passive savers on the other hand stick to the default and do not react to changes in savings incentives.

Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen (2014) find that the vast majority (85%) of the Danish population can be described as passive savers, who do not respond to savings subsidies and who save more (less) when the default contribution rates increase (decrease). Only 15% are active savers, reacting to subsidies by shifting savings to the subsidized accounts and setting off changes in automatic contributions by adapting their contributions in other accounts. More recent research has shown that individual behavior is not always consistently active or passive, but may change depending on the setting (see e.g., Butt, Donald, Foster, Thorp, and Warren 2018; Goda, Levy, Manchester, Sojourner, and Tasoff 2020; Goodman 2020).

Definition Behavioral Types To examine the prevalence of active and passive saving behavior in the simple choice framework for mini-jobs, we exploit the panel structure of the data set by comparing the enrollment behavior of individuals observed under both defaults. We define a total of 6 behavioral types, as shown in Table 4.5.

Never takers never enroll, neither under the opt-in nor under the opt-out regime. They

		Opt-in Regime (Old Default)	Opt-out Regime (New Default)
Active	Never Taker	✗	✗
	Always Taker	✓	✓
Passive		✗	✓
Partly Inert	Delayed Never Taker	✗	✓→ ✗
	Delayed Always Taker	✗→ ✓	✓
Inconsistent		✓/✗	✓/ ✗

Notes: Definition of behavioral types based on the enrollment behavior of individuals that are observed under both defaults. ✓: enrollment for all months under the corresponding default. ✗: no enrollment for all months under the corresponding default. ✓→ ✗: enrollment for the first $x < n$ consequent months and no enrollment for the last $n - x$ months under the corresponding default. ✗→ ✓: no enrollment for the first $x < n$ consequent months and enrollment for the last $n - x$ months under the corresponding default. ✓/✗: none of the above patterns.

Table 4.5: Definition of Behavioral Types

always stick to the default when the default is no enrollment and always opt out when they are enrolled by default. For never takers, there is no inertia under the new default, neither inattention nor switching costs prevent them from opting out. *Always takers* on the other hand always enroll irrespective of the default. For both never and always takers, the default does not affect their enrollment status, i.e., they do not contribute more under automatic enrollment. Both groups can be described as active individuals who either actively chose to enroll (always takers) or not to enroll (never takers), but whose choice is not affected by the default.

The third group comprises individuals that always stick to the default: they do not enroll under the opt-in regime and they do not opt out when automatically enrolled. Again following the Chetty, Friedman, Leth-Petersen, Nielsen, and Olsen (2014) terminology, we refer to them as *passive* individuals. They exhibit perfect default stickiness under both the opt-in and the opt-out regime and their behavior can thus be described as true passive behavior. Their enrollment follows the default-setting and a default-changing reform thus influences their future pension entitlements.

The remaining individuals do neither consequently pursue enrollment or no enrollment (active behavior), nor do they always stick to the default (passive behavior). If they stick to the default first but then deviate after some time, we refer to their behavior as *partly inert*. *Delayed never takers* never actively enroll in the pension insurance under the opt-in regime, stick to the default for the first $x < n$ months under auto-enrollment, then opt out and never enroll again for the remaining $n - x$ months. Analogously, we define *delayed always takers* as individuals who do not enroll for the first $x < n$ months under the opt-in regime, then start enrollment and continue enrollment for all future $n - x$ months pre reform and never opt out from automatic enrollment. There are different explanations for partly inert behavior like this. Mini-job employees may be inattentive regarding their default and/or the possibility to deviate from it and may only learn about it after x months. Alternatively, they may be attentive but other factors, like (perceived) switching costs prevent them from immediately engaging in

active behavior. The remaining behavioral patterns are summarized as *inconsistent* behavior.

	Case I	Case II
NeverTakers	62.43	67.04
AlwaysTakers	5.35	5.30
Passive	12.70	8.26
DelayedAlwaysTakers	3.70	1.44
DelayedNeverTakers	4.02	3.80
Inconsistent	11.81	14.15
Total	100.00	100.00
N	47,765	57,628

Notes: Sample: all individuals that are observed under both defaults for at least one month between 01/2011 and 11/2016. Weighted data. Case I and II account for the uncertainty of the default for some post-reform observations as described in the text. They mark the two most extreme cases and thus provide a corridor for the true value, which has to lie between the two extremes. See Table 4.5 for the definition of the behavioral types.

Table 4.6: Prevalence of Behavioral Types

Sample In order to examine the taxonomy of individual behavior under different defaults, we restrict the sample to individuals that we observe at least once under each default regime. As discussed in section 4.4, we cannot unambiguously determine the default for all post-reform observations. For mini-jobs that started pre reform, the old default applies post reform as long as a certain income threshold is not surpassed (see section 4.2 for details). However, in the data, those individuals are not distinguishable from individuals whose mini-job ended on December 31, 2012 and who started a new one, under the new default, on January 1, 2013. Consequently, the default is unknown for post-reform observations of these individuals.

We account for this uncertainty by showing the results for the two most extreme cases. For case I, we assume that all individuals remained in their mini-job and none of them started a new mini-job on January 1, 2013. For case II, we assume that all individuals in question started a new mini-job on January 1, 2013. With these two extreme cases we provide a corridor for the true values.

Results Table 4.6 indicates that the majority of individuals are never-takers, meaning they never enroll, whether under the opt-in or the opt-out regime. In both cases we consider, we observe that approximately two-thirds of individuals fall into this category. Therefore, the majority of individuals exhibit no inertia under the new default setting and opt out immediately after commencing their mini-job with auto-enrollment. Policy makers do not influence their enrollment behavior by altering the default option. The default setting has a consistent effect solely on passive individuals, who constitute only a minority of individuals. As shown by Table 4.6, this purely passive behavior is exhibited by approximately 8 to 13 percent of individuals, i.e., transitioning from an opt-in to an opt-out regime nudges only about 1 in 10 individuals into enrollment. There are a few always takers and individuals whose behavior is in line with delayed opting-in or

opting-out, but they only constitute a small minority.

The differences between case I and case II are driven by two factors. First, since we restrict the sample to individuals with at least one observation under each default, we consider different samples for the two cases. An individual for whom the actual default is unknown for all post-reform observations is considered for the old default in case I and for the new default in case II. In case I, this leads to no observations under the old default for this individual which is why we do not include them in the sample here. This explains the smaller sample size for case I. Second, the same behavior is classified differently under the two cases for some individuals. This can be illustrated with an individual for whom we observe three mini-job periods: One pre-reform period without enrollment, a second period post reform without enrollment and under unknown default and a third period post reform with enrollment under the new default. In case I, this is classified as passive behavior since the individual never enrolls under the old default (period 1 and 2) but always under the new default (period 3). In case II, the behavior is classified as inconsistent because the enrollment status under the new default (period 2 and 3) is not consistent. However, while the precise shares differ between case I and II, the overall patterns remain unchanged.

4.5.2 Additional Evidence: Responsiveness to Incentives

A key finding of our analysis is that most individuals opt out and we show that this is driven by a majority of never-takers. An open question that remains is why are individuals so little prone to automatic enrollment in this setting? One potential explanation is the simplicity of the setting: when facing a simple choice menu, costs and benefits from enrollment may be more salient for individuals. We provide suggestive evidence for this hypothesis by studying enrollment behavior of individuals who are close to reaching a threshold for waiting periods.

As described in section 4.2, periods of mini-job employment are fully credited as waiting period only if individuals enroll. Different amounts of these waiting periods are required to meet with eligibility thresholds for different types of pension. For individuals close to a waiting period threshold, e.g., 35 years for early retirement (see Table A.2 for more details), benefits from enrollment are larger, potentially being decisive for eligibility at the extensive margin for the respective pension type. For instance, an individual with an insurance record of 34 years will become eligible for early retirement if they work for one additional year in a mini-job and enroll in the public pension insurance. Without enrollment, less than 12 months will be considered as waiting periods and they will not reach the 35 year threshold. The same logic applies for the other thresholds. Consequently, we expect higher enrollment for individuals close to a threshold if the incentives are salient to them.

To assess whether mini-job employees understand and react to these incentives, we compare the average enrollment share for mini-job employees over time relative to the event of reaching an eligibility threshold, e.g. for early retirement. For each threshold, we include all individuals that we eventually observe reaching the respective threshold in the data. The results are shown in Figure 4.3. For all thresholds, the enrollment share peaks at, or very close to, the threshold. The observed behavior is in line with individuals responding to the incentives for enrollment that are tied to the waiting period thresholds.

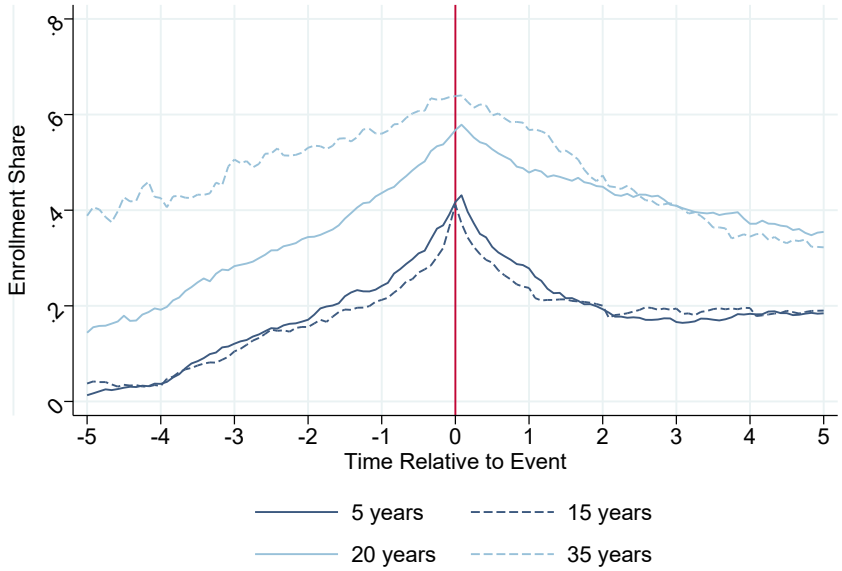


Figure 4.3: Enrollment Relative to Waiting Period Thresholds

Notes: Weighted Data. *Event*: the month an individual reaches the threshold for the respective waiting periods. *Time Relative to Event*: position in the individual employment history relative to reaching the threshold. A value of 0 marks the month the given threshold is reached, -1 indicates that those individuals needed one more year (12 more months) of contribution time before they reached the threshold. *Sample*: For each waiting period threshold (see Table A.2 for details), the sample includes all individuals that have reached that threshold already. For each line, N is thus constant for periods ≤ 0 but decreasing for periods > 0, because not all individual are observed 5 years after they reached the threshold.

There are two potential mechanisms behind the observed effects. First, individuals who are in a mini-job already, may start enrolling when approaching the threshold. Second, individuals may start a mini-job with enrollment when approaching the threshold. In both cases, a certain knowledge of the institutional setting is required to be able to react to the thresholds. If (at least parts of) the mini-job employees understand incentives in the rather complex German public pension system, it seems plausible that they are also aware of their enrollment default. This is one potential explanation for why most mini-job employees opt out of automatic enrollment.

4.6 Conclusion

It is well established that individuals commonly stick to defaults, particularly in the context of old age savings. We show that the power of the default is limited in a pension setting where the choice menu is simple. In this simple setting, many of the explanations for inertia in more complex settings can be ruled out, including switching costs (Gabaix 2019; Heiss, McFadden, Winter, Wuppermann, and Zhou 2021), choice overload (Goda, Levy, Manchester, Sojourner, and Tasoff 2020; Iyengar and Lepper 2000), and (mis-)perceiving a default as investment advice from the employer (Madrian and Shea 2001).

We analyze a natural experiment of introducing automatic enrollment for low-income employees with a binary choice menu in the German PAYG pension system. The setting allows us to rule out the above explanations for inertia, leaving inattention as the main explanation. We find that the introduction of autmatic enrollment significantly increases the take-up. However, compared to the existing literature on auto-enrollment, the effect of changing the regime from opt-in to opt-out is rather small, with the majority of individuals opting out immediately.

Only few individuals exhibit true passive behavior defined as always sticking to the default, while the majority is best described as never-takers who never enroll. In addition, a significant share of individuals seem to understand and react to enrollment incentives. Using automatic enrollment to nudge low income earners into higher public pension savings for their retirement may thus be less effective than policy makers may hope.

We also observe sizable heterogeneity in the effect of the default, which should be taken into account when designing public policies that aim at shaping individual behavior using default setting. Implementing automatic enrollment for everybody does not necessarily enhance future pension entitlements equally for all groups of individuals. And depending on the target group, automatic enrollment may nudge the "wrong" individuals into enrollment, for example because they are less financially literate.

Appendix to Chapter 4

4.A Additional Figures and Tables

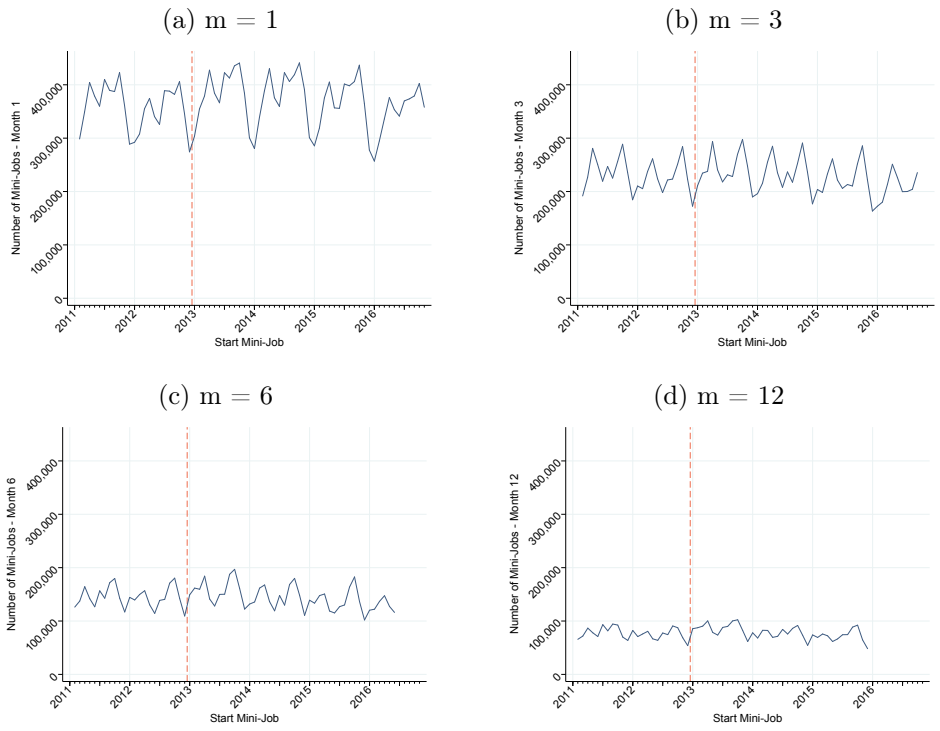


Figure A.1: Number of Mini-Jobs

Notes: Weighted data. Graphs display the number of mini-jobs for which we identify a start in month t in the m^{th} month of that mini-job over t. In each panel, the line at given t contains the same group of individuals, exclusive those who dropped out of their mini-job employment before reaching the m^{th} month of this employment. For example, an individual who is employed for four months, is only considered for Panel a and b.

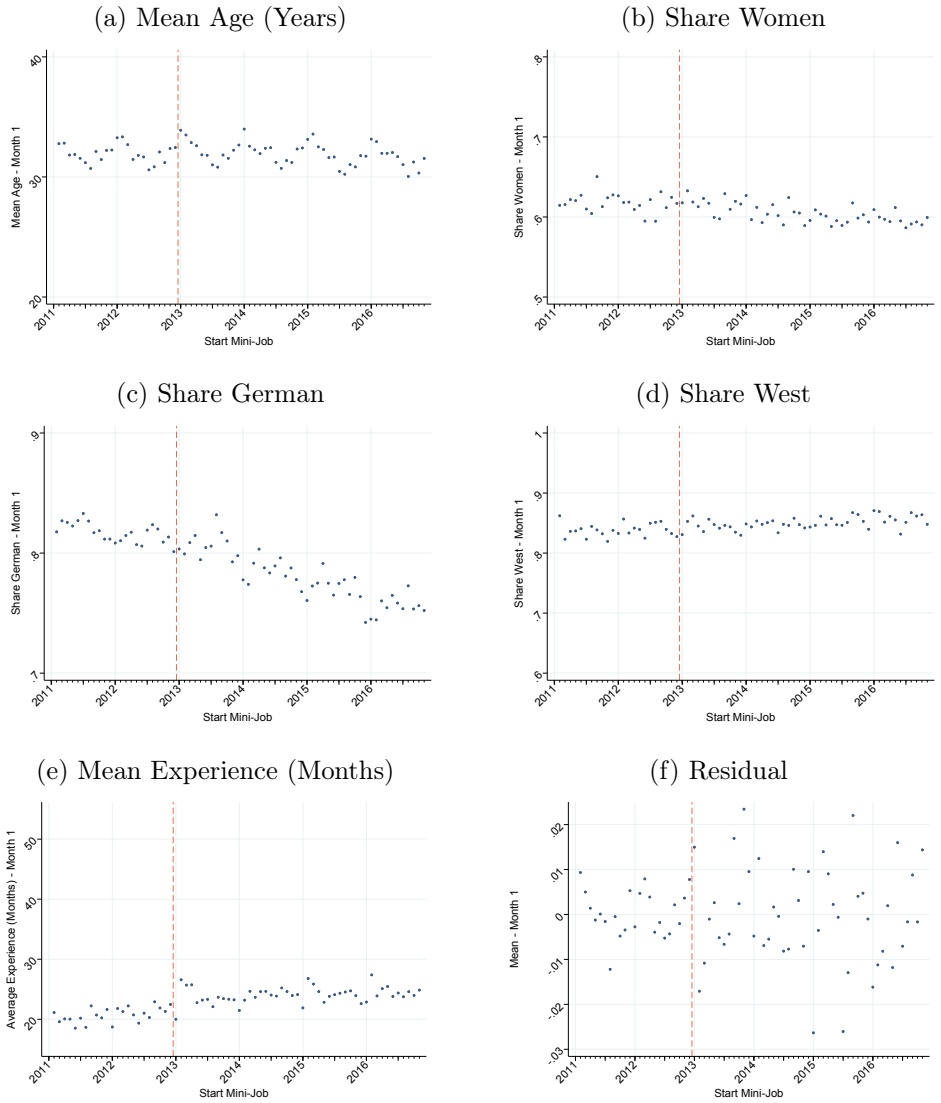


Figure A.2: Control Variables and Residual Over Time

Notes: Weighted data. Scatter plots in panel a to e display the average value for each control variable in the 1^{st} month of the mini-job for individuals who started their mini-job in month t over that t. Panel f displays the average residuals (η_{it}) from Equation 4.2, again for the 1^{st} month of employment.

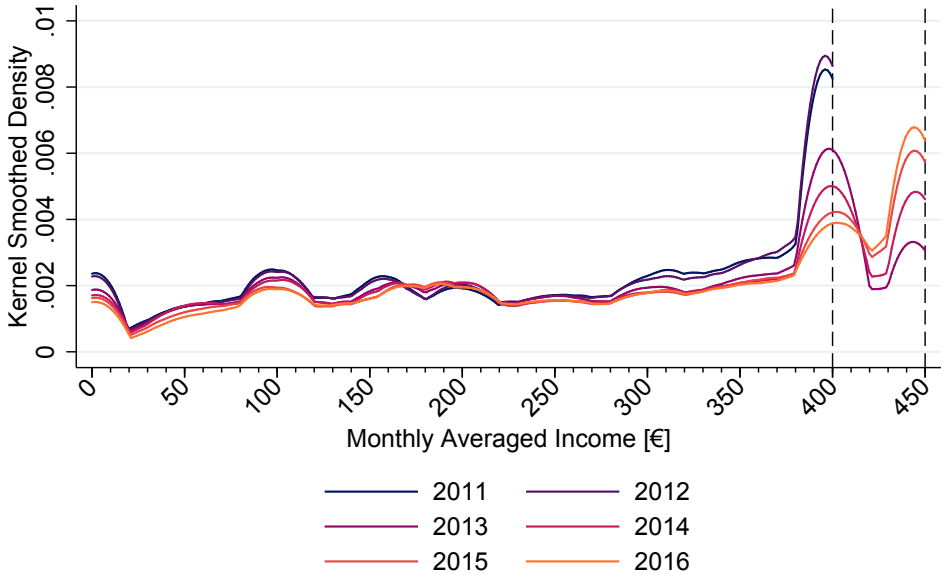


Figure A.3: Income Distribution Mini-Jobs

Notes: Annual income distribution for mini-jobs. Kernel smoothed density, weighted data.

year	Y	$\overline{Y}$		PPV		τ		
		east	west	east	west	τ_{full}	τ_{er}	τ_{ee}
2011	32,100	57,600	66,000	24.37	27.47	19.9	15	4.9
2012	33,002	57,600	67,200	24.92	28.07	19.6	15	4.6
2013	33,659	58,800	69,600	25.74	28.14	18.9	15	3.9
2014	34,514	60,000	71,400	26.39	28.61	18.9	15	3.9
2015	35,363	62,400	72,600	27.05	29.21	18.7	15	3.7
2016	36,187	64,800	74,400	28.66	30.45	18.7	15	3.7

Notes: Y = average annual income in €, valid from July in the given year until June in th following year; $\overline{Y}$ income threshold in € (no contributions for $Y - \overline{Y}$); PPV = pension point value in € for pensions payed in the given year; τ = contribution rate; τ_{full} = contribution rate under full contribution; τ_{er} = contribution rate for mini-job employers; τ_{ee} = contribution rate for mini-job employees (all in %).

Table A.1: Operands of the German Statutory Pension Insurance

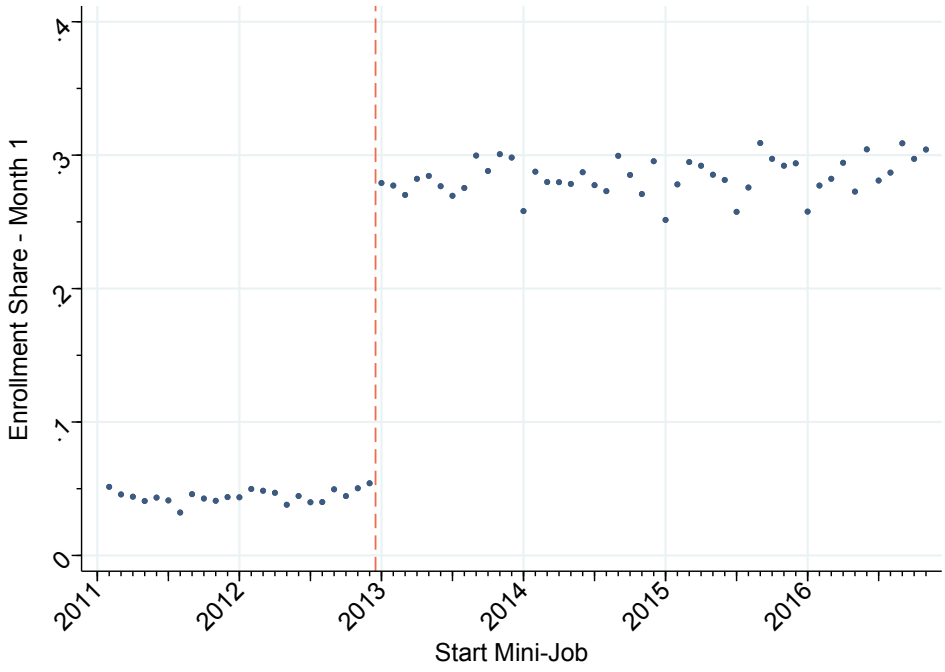


Figure A.4: Robustness Check: Income ≤ 400 € ($m = 1$)

Notes: This figure replicates Figure 4.1, but restricting the post-reform sample to mini-jobs with income up to the pre-reform income threshold of 400 €, as described in subsection 4.4.4. See also figure notes for Figure 4.1.

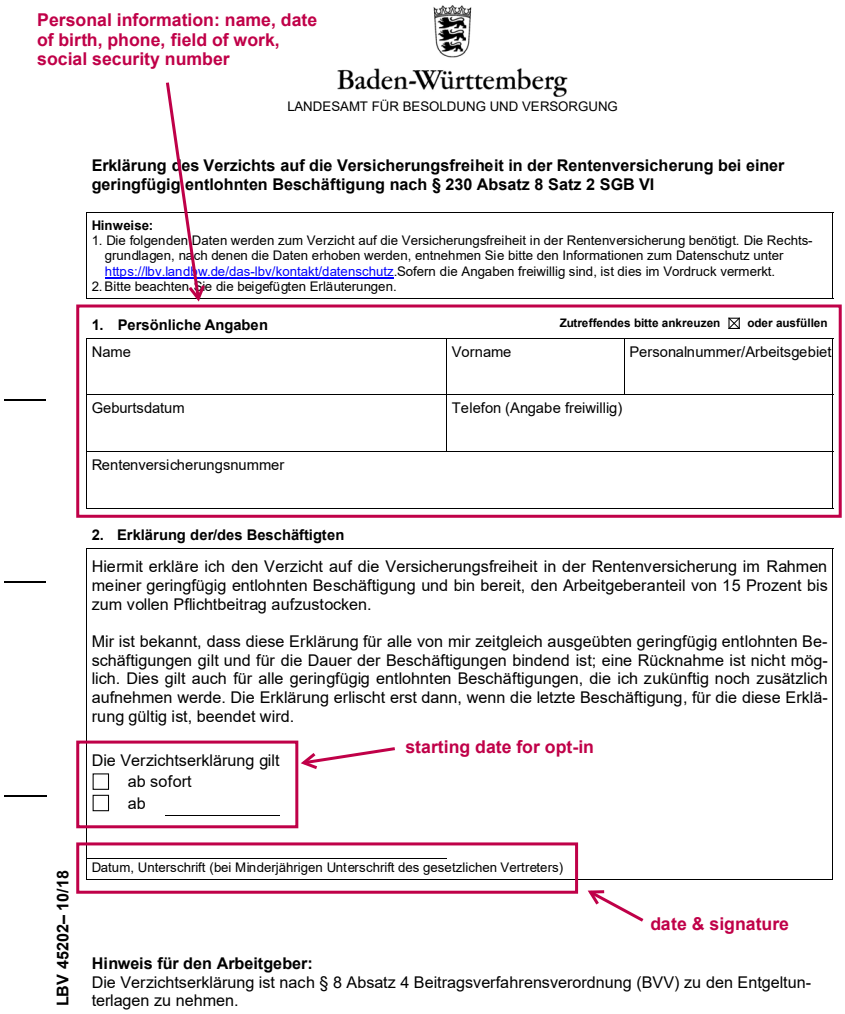


Figure A.5: Example Form for Opting-In for Enrollment

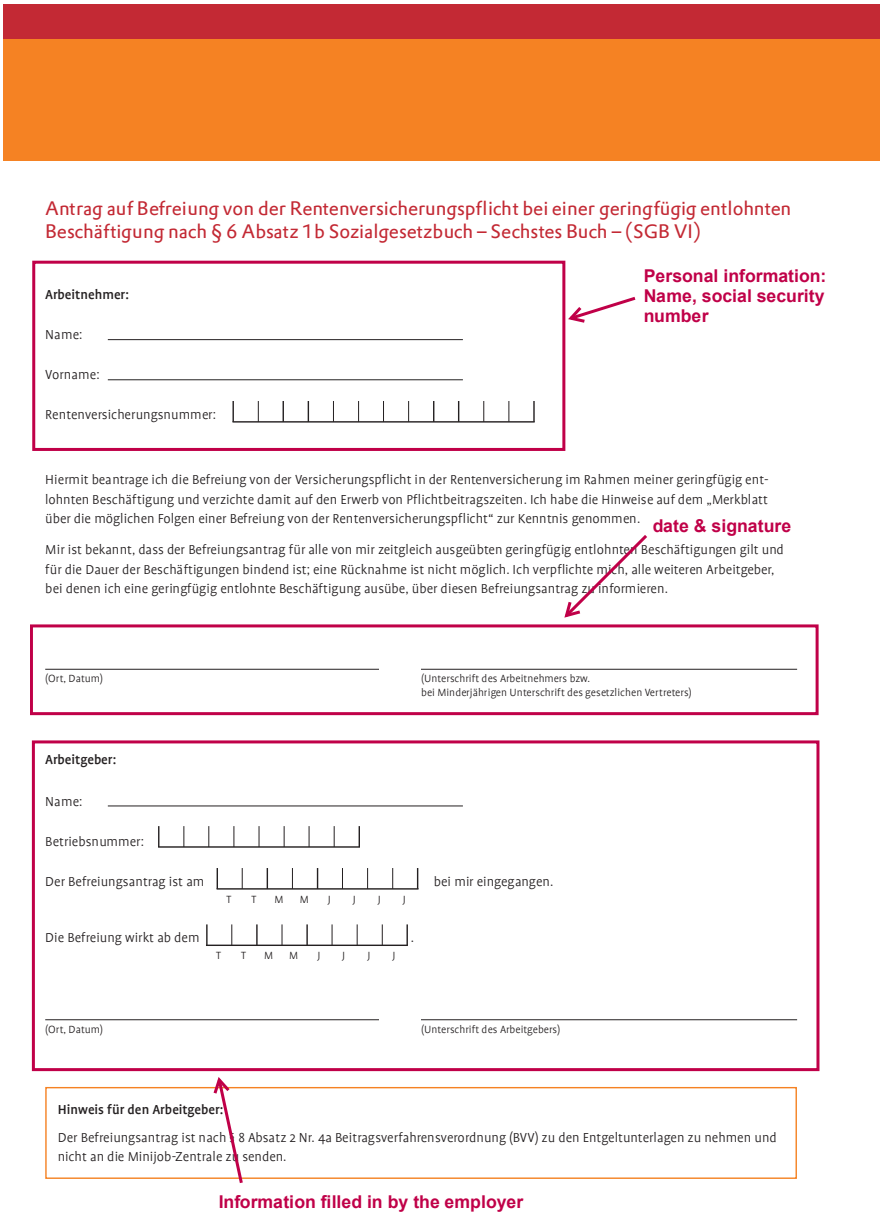


Figure A.6: Example Form for Opting-Out from Enrollment

Threshold	Pension Type	Eligible Group
5 years	- Standard old-age pension	- No further restrictions
15 years	- Old-age pension for women	- Women born before 1952
	- Old-age pension on account of unemployment	- Individuals born before 1952 and above a certain age threshold
20 years	- Reduced earning capacity pension	- Individuals with reduced earnings capacity who have not reached the 5-year threshold.
35 years	- Long service pension	- Individuals above a certain age threshold
	- Old-age pension for people with severe disabilities	- Individuals with severe disabilities and above a certain age threshold
45 years	- Exceptionally long service pension	- Individuals above a certain age threshold

Notes: Incomplete and simplifying, see https://www.bmas.de/EN/Our-Topics/Pensions/old-age-pensions.html [last accessed: 2020-10-31] for more details. Waiting periods include periods of (regular) employment as well as a variety of other situations, including parental leave or unemployment. Which situations are considered as waiting period differs slightly across the differennt thresholds.

Table A.2: Waiting Period Thresholds

	(1)	(1a)	(1b)	(2)
D_t	0.2381***	0.2397***	0.2395***	0.2414***
	(0.0021)	(0.0020)	(0.0047)	(0.0046)
t	0.0004***	0.0003***		
	(0.0001)	(0.0001)		
t_{pre}			0.0005	0.0005
			(0.0005)	(0.0005)
t_{post}			0.0004***	0.0003***
			(0.0001)	(0.0001)
X_{Jit}	No	Yes	No	Yes
N	337109	333707	337109	333707

This table replicates the results from Table 4.2 using Logit instead of OLS. Marginal effects from Logit regressions, robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. See Table 4.2 for explanations of variables etc.

Table A.3: Instantaneous Effect – Logit

	m = 3		m = 6		m = 12	
D_t	0.1738***	0.1776***	0.1725***	0.1767***	0.1744***	0.1747***
	(0.0060)	(0.0058)	(0.0084)	(0.0081)	(0.0137)	(0.0131)
t_{pre}	0.0011***	0.0013***	0.0031***	0.0036***	0.0055***	0.0061***
	(0.0002)	(0.0002)	(0.0003)	(0.0003)	(0.0005)	(0.0004)
t_{post}	0.0005***	0.0004***	0.0007***	0.0007***	0.0008**	0.0008**
	(0.0001)	(0.0001)	(0.0002)	(0.0002)	(0.0003)	(0.0003)
cons	0.0357***	0.0154***	0.0301***	-0.0076	0.0358***	-0.0212**
	(0.0026)	(0.0037)	(0.0038)	(0.0051)	(0.0059)	(0.0077)
X_{Jit}	No	Yes	No	Yes	No	Yes
R2	0.0525	0.1176	0.0447	0.1163	0.0327	0.1170
N	201808	199970	119554	118629	57353	56913

Notes: This table replicates the findings from Table 4.3 using all observations (not excluding group 3 from Figure 4.2). Effect on the enrollment E_{it}^m of individual i in the m^{th} month of their mini-job in month t. Robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. See Table 4.3 for explanations of the variables etc.

Table A.4: Medium-Term Effects – All Observations

	$m = 3$	$m = 6$	$m = 12$
D_t	0.1776***	0.1645***	0.1465***
	(0.0055)	(0.0077)	(0.0140)
t_{pre}	0.0002	0.0003	-0.0008
	(0.0006)	(0.0009)	(0.0022)
t_{post}	0.0003**	0.0006***	0.0007**
	(0.0001)	(0.0001)	(0.0002)
X_{Jit}	Yes	Yes	Yes
N	194893	109306	46923

This table replicates the results from Table 4.3 that include X_{Jit}, using Logit instead of OLS. Marginal effects from Logit regressions, robust standard errors in parentheses. Weighted data. Significance level: *** 0.001; ** 0.01; * 0.05; ^{x} 0.1. See Table 4.3 for explanations of variables etc.

Table A.5: Medium-Term Effects – Logit

	(4.1)	(1a)	(1b)	(4.2)
D_t	0.2312***	0.2406***	0.2295***	0.2379***
	(0.0034)	(0.0034)	(0.0052)	(0.0051)
t	0.0003**	0.0003***		
	(0.0001)	(0.0001)		
t_{pre}			0.0002	0.0001
			(0.0001)	(0.0001)
t_{post}			0.0003**	0.0003**
			(0.0001)	(0.0001)
$cons$	0.0408***	0.0408***	0.0419***	0.0427***
	(0.0014)	(0.0028)	(0.0021)	(0.0032)
X_{Jit}	No	Yes	No	Yes
R2	0.0835	0.1069	0.0835	0.1069
N	281562	278840	281562	278840

Notes: This table replicates the findings from Table 4.2, but restricting the post reform sample to mini-jobs with income up to the pre-reform income threshold of 400 €, as described in subsection 4.4.4. See notes in Table 4.2 for explanation of variables etc.

Table A.6: Robustness Check: Income ≤ 400 € ($m = 1$)

4.B Illustration of Enrollment Incentives

In order to better understand the incentives of enrollment, we illustrate them with an exemplary mini-job employee who worked for 12 months in 2013 with a monthly income of 450 €. That employee acquires 0.1604 EP under enrollment and 0.1273 EP with employer contribution only.[13] They have an incentive to enroll if their return from the increased pension entitlements is higher than their costs from enrollment. By how much they increase their total pension entitlements depends on the length of the retirement period as well as on the earning-points' monetized value upon retirement. In 2013, the equivalent monthly pension value for 0.1604 EP and 0.1273 EP amounts to 4.51 € and 3.58 €, respectively.[14] With a monthly contribution of 17.50 € for one year, the employee thus increases their future monthly pension entitlement by 0.93 € in 2013 numbers. When assuming constant value for the pension points after 2013, the total increase in pension entitlement surpasses the contribution payments after 18.8 years of retirement (17.50/0.93).

In reality, the pension points' value is tied to the development of the market income such that the equivalent monthly pension value from the enrollment in the above mini-job example has increased substantially already since 2013. In 2020, the 0.1604 EP and 0.1273 EP are worth 5.29 € and 4.21 €, respectively. Under enrollment and with a monthly contribution of 17.50 € in 2013, the employee thus increases their monthly pension entitlement by 1.08 € in 2020 numbers. If we again assume constant future value for the pension points, the increase in pension entitlements exceeds the payments from enrollment after 16.2 years.

This oversimplified back-on-the-envelop calculation is abstracting from many relevant factors such as the actual development of the pension point value, life expectancy, risk-aversion, intertemporal discounting, present-bias or outside-options for savings. It is meant to provide a broad idea of how much it pays off for minijob-employees to enroll in the statutory pension insurance.

In our example with 12 months mini-job employment and a monthly income of 450 € in 2013, the mini-job employee obtains 12 months for their waiting period when enrolled but only 4 months when not enrolled ($\frac{0.1273}{0.0313}$ rounded to full months). It is important to note that the waiting period is not income related under enrollment, but depends on the income when not enrolled. A mini-job employee with a monthly income of 200 € for 12 months pays 7.80 € and obtains 0.0713 EP as well as 12 months for their waiting period under enrollment in 2013. Without enrollment, they acquire 0.0566 EP and 2 months ($\frac{0.0566}{0.0313}$ rounded to full months) for their waiting period. The enrollment incentive for waiting periods, relative to the costs of enrollment, is thus stronger for lower incomes.

[13] $Y_t = 33,659$ (Table A.1). Earning-points for enrolled employees are calculated with $EP = \frac{450 \times 12}{33,659}$. For non-enrolled employees, $EP = \frac{450 \times 12}{33,659} \times \frac{15}{18.9}$.

[14] These are the values for western German states, which account for the vast majority of the German population and also for the majority of mini-job employees. For employees in eastern states, the values are 4.13 € and 3.28 €, respectively. The monthly pension values per earning point over time for east and west are showed in Table A.1

4.C Potential information channels

A significant share of mini-job employees shows no inertia at all when being under the opt-out regime. Two thirds are not enrolled a single month (passive savers, see Table 4.6), thus opted-out of the default enrollment immediately after taking up their employment.[15] Being attentive to the default situation is a necessary condition for opting out. A natural question to ask is where the employees' attention comes from, or, more specifically, where they learn about the default and how to opt out.

One potential source of information is media coverage. In Germany, it is very common that at the end of a year newspapers and magazines inform their readers about institutional changes that become effective in the subsequent year. These overviews are typically published in late December and feature information on changes in the tax code, social security or other broad interest topics. We studied these overviews for 2012 from three large news papers that included information on the reform for mini-jobs[16]. However, the focus was clearly on the increase of the income threshold from 400 € to 450 € and much less on the change from opt-in to opt-out. *Bild* (the largest German newspaper) did not even mention the change in default. The *Frankfurter Allgemeine Zeitung* informed about the new default, but without mentioning the possibility to opt out. In *Die Welt*, the information was more exhaustive, explaining some advantages of being enrolled and that mini-job employees can opt out.

We investigate Google search volume for "minijob rentenversicherung" (mini-job + pension insurance) and find a clear peak immediately after the reform. The search volume reaches its highest point in January 2013 and has remained constantly above the pre-reform level ever since. However, we cannot say whether this is driven by employers or by employees (or even other individuals).

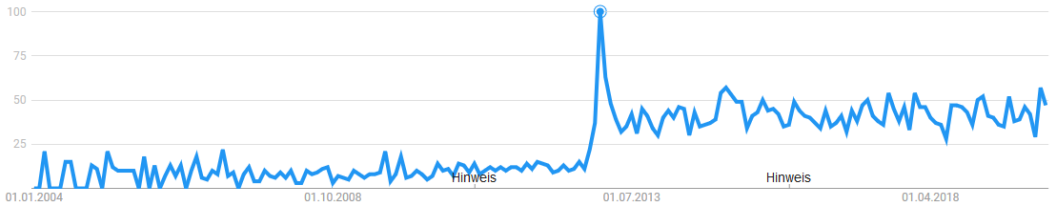


Figure A.0.1: Google Search Volume

Note: Search volume from January 2004 to February 2020. Search term: "minijob rentenversicherung" (mini-job + pension insurance). Highest value for January 2013. Screenshot from https://trends.google.de/trends (last accessed 2020-02-12).

[15] The deadline for opting out for a given month is the Xth of that month.

[16] *Frankfurter Allgemeine Zeitung*: https://www.faz.net/-gqe-75cu8 (last accessed 2019-12-08); *Welt*: https://www.welt.de/print/wams/finanzen/article112300864/Alles-neu-macht-der-Januar.html (last accessed 2019-12-08); *Bild*: https://www.bild.de/geld/wirtschaft/gehalt/hier-gibt-es-2013-mehr-geld-27784884.bild.html (last accessed 2019-12-08)

Chapter 5

Rising Income Tax Complexity*

5.1 Introduction

High and rising tax filing costs and tax complexity is a bipartisan issue: both Republicans and Democrats agree, in principle, that filing taxes imposes a substantial burden on US taxpayers that should be reduced, but generally disagree over what methods to use to reduce them.[1] Similarly, economists and US taxpayers generally agree that these costs are high and increasing. In spite of this general consensus, there has been limited research documenting the evolution of tax compliance costs.

This paper attempts to fill this gap using two distinct and complementary approaches. We first designed and ran a survey of approximately 800 US taxpayers with the goal of eliciting their perceptions of the complexity of the tax code. While survey instruments are imperfect, they allow us to validate some of our findings from the observational data approach we describe below, and address questions that are unanswerable using observational data.

There are many reasons why a tax system might become more complex. In the US, for example, many welfare programs are administered through the tax system, such as the Earned Income Tax Credit. Similarly, government policies aimed at tackling externalities are also implemented via taxes and credits, resulting in a more complex tax system. And political considerations, such as the 10-year budget window and the sun-setting of provisions can lead to a longer tax code. In the survey, we focus on two possible justifications for adding complexity. The first one is that a more progressive tax system often leads to more complexity: a flat tax would be substantially simpler than current tax systems but has the downside of being regressive. Having a more complex tax system allows for a more progressive tax schedule. We show that US taxpayers do not share this view. The majority of our survey respondents believes that the tax code's complexity makes the tax system *less* fair overall. And because the majority of our respondents have been filing taxes for more than a decade, this appears to be representative of a long time trend. The taxpayers' perception is consistent with the findings of Saez and Zucman (2019), for example, who show that tax progressivity has not increased in the US.

The second potential justification for increasing tax complexity is to hinder tax evasion: a more complex tax system requiring taxpayers to provide more information may make it harder to evade taxes. However, the majority of respondents believe that increased complexity *fosters* evasion, further exacerbating the perception of unfairness that com-

*This chapter is based on a published article in the *National Tax Journal* (Benzarti and Wallossek 2024).

[1]See Senator Elizabeth Warren's "Tax Simplification Act" of 2022 (link) and the Trump administration's 2018 Economic Report of the President (link, page 41).

plexity imposes on taxpayers. This is consistent with the observation that while tax complexity has increased, tax evasion has not decreased.

Given that the majority of respondents believe that the tax system has become more complex and that this complexity tends to hurt US taxpayers either via increased evasion or lower progressivity, it would come as no surprise that most taxpayers prefer a substantially simpler tax system. While a large share of taxpayers state that they would be willing to forgo some of their tax refund to make the US tax system simpler, the respondents are equally split between being very willing and not willing at all to do so. Among those that are willing to pay for a simpler tax system, the willingness to pay is on average $130 per year, which is within the range of what a taxpayer would pay when using a tax preparation software (such as TurboTax) and less than what the same taxpayer would pay if visiting a tax preparer.

Relatedly, another way to ease the complexity burden on taxpayers is to reduce filing costs. One approach, which is commonly used in some OECD countries is to pre-populate tax returns prior to sending them to taxpayers who are then given the opportunity to review, amend and then file them. While there are no pre-populated forms for US taxpayers, Goodman, Lim, Sacerdote, and Whitten (2023) show that for more than 40% of returns could be pre-populated. Pre-populating forms has the advantage of reducing record keeping costs and some form filling costs. Opponents of pre-populated forms – in the US – generally argue that they would make taxpayers too complacent, simply agreeing to anything the IRS would write on the pre-populated forms, thus leading to a higher tax burden.

To our knowledge, except for anecdotal evidence, we do not know whether taxpayers would be in favor of receiving pre-populated tax returns. According to our survey, the majority of taxpayers would be willing to pay for receiving a pre-populated tax return. And among those that are willing to pay for them, they would pay $77 on average. Both taxpayers who identify as liberal and conservative are willing to pay for pre-populated forms, suggesting that pre-populated forms would be a popular reform. However, liberal taxpayers are significantly more willing to pay for pre-populated forms than conservative taxpayers.

We also asked US taxpayers how much time it takes them to file their taxes.[2] On average, taxpayers who do not use a tax preparer spend 2.2 hours using tax software, 1.4 hours looking for receipts and forms and 0.8 hours learning about the tax law. Those that use a tax preparer spend 1 hour going to and waiting at the tax preparer, 1.9 hours looking for forms and receipts and 0.6 hours learning about the tax law. These figures are lower than those reported by the IRS or estimated in Benzarti (2021). This could be due to the fact that the sample of US taxpayers in our survey differs from that used by the IRS. And it is consistent with the fact that our respondents believe that it is less tedious to file taxes for themselves than for others.

In our second approach, we proxy for filing costs using the length of tax codes over time and across a few different countries.[3] In general, longer tax codes do not necessarily

[2]This was regularly done by the IRS but has not been updated in recent years.

[3]Word counts have been used as a measure for tax code complexity before (see e.g., Bacher and Brülhart 2013; Hoppe, Schanz, Sturm, and Sureth-Sloane 2021; Moody, Warcholik, and Hodge 2005; Weber 2015).

imply more complex tax codes. A wordier tax code might even be simpler if the additional words describe the law in more detail, reducing uncertainty as to how they will be applied. A better approach would measure tax complexity by using the revealed preferences of a representative sample of taxpayers and comparing them over time and across countries. While previous research has attempted this for specific tax code provisions and a sub-sample of the population over certain periods (Benzarti (2020)), generalizing this approach to compare estimates across countries, time and individuals is challenging, if not impossible. This is why we rely on the tax code word count approach, which is incomplete in its coverage of complexity but is implementable. Moreover, this imperfect proxy has the advantage of capturing part of the complexity of the tax code due to additions to it, rather than existing provisions that are made more complex. For the US, we show that the word count measure correlates with alternative measures of income tax complexity.

Similarly, if tax complexity has been growing over time, this should be reflected in the tax preparation and software industry. The ideal measure for this would be industry profits for the tax preparer and the tax preparation software industry, which is not readily available. Instead, we use the number of employees in this industry and their average wage as a proxy and find that these have been steadily growing in the US since the early 2000's.

Overall, both the observational and survey data clearly confirm the general consensus that tax filing costs have been increasing. This pattern holds in the US, both when using word counts and survey responses, but also in the other countries we consider, namely France, Germany, Switzerland, Canada and Morocco, albeit only using the observational data.

This paper's main contribution is to show that filing costs and complexity have been growing over time and across countries, as well as to document perceptions surrounding tax filing costs and complexity. Our paper contributes to the growing literature that uses observational data to estimate tax filing costs, such as in Pitt and Slemrod (1989), Benzarti (2020), Hauck and Wallossek (2023). With the exception of Benzarti (2021), most of these papers estimate filing costs for a specific provision of the tax code (such as itemizing deductions) and a specific time period.[4] Our paper also contributes to a literature that uses survey evidence to elicit the perception of taxpayers about the tax code, pioneered by Kuziemko, Norton, Saez, and Stantcheva (2015). More recently, Blesse, Buhlmann, and Dörrenberg (2019) implemented a survey of German taxpayers eliciting their preferences over the trade-off between simplifying the German tax code and making it less progressive, which is related to one of the questions we ask our survey participants.

5.2 Perceived Complexity in the US

5.2.1 Data

Complexity describes "the state of having many parts and being difficult to understand or find an answer to" (Cambridge dictionary, link). Throughout the paper, we define tax

[4]Benzarti (2021) relies on the estimates from Benzarti (2020) and extrapolates them to rest of the tax filing schedule.

complexity along the same lines: a complex income tax system is difficult to understand, comply with, and navigate because it has many parts and rules.[5] As a result, it is costly for taxpayers to deal with. We focus on the complexity of the tax code and income tax filing for individual taxpayers.

To gain some understanding into how taxpayers perceive the complexity of the tax code and elicit beliefs over the ramifications of this complexity, we developed and administered a survey of US taxpayers. The survey questions and possible responses are documented in Appendix 5.B.

For most survey questions, we provide respondents with a slider ranging from one extreme option to another, e.g. from "not willing" to "willing". The underlying scale ranges from 0 to 100 (these numbers are not displayed when the survey is being administered). The default position for the slider is 50 and respondents are required to move the slider. Throughout our empirical analysis, we often group respondents in two subgroups, indicating support or opposition for a given statement. We divide the sample based on whether the response is greater or smaller than 50, excluding those who are exactly at 50.

We ran the survey using the platform Prolific. This platform, similar to Amazon MTurk, recruits a pool of participants who are available to participate in surveys. Participants self select into the survey they want to participate in. We pre-screened participants in order to focus on US respondents that are of prime working age and have had some experience with tax filing. In particular, we pre-screened participants on the following characteristics: live in the US, first language is English, older than twenty five years old and full- or part-time employment status. To further ensure that our sample is drawn from tax filers, we noted that anybody who has never filed taxes themselves or using a tax preparer should not participate in the study recruitment message. We also drop participants who responded that they do not usually file taxes (Question 1). Information on age and race is collected by Prolific. As with all surveys administered using an online platform we can expect that survey participants might be selected on being more comfortable with technology than the rest of the population. This might bias the sample towards taxpayers who self-file using tax software.

The survey ran from July 10th, 2023 to July 11th, 2023. The survey was expected to take three minutes and the median completion time was two minutes and forty three seconds. Participants were approximately paid twelve dollars an hour for completing it. Note that, prior to running the survey, we ran a pilot to ensure that there were no issues with the implementation of the survey.[6] We recruited eighty individuals to participate in the pilot and ensured that none of the individuals who participated in the pilot were part of the main survey. We also excluded the pilot observations from the analysis. We did not detect any issues with the pilot and therefore did not make any changes to the survey when running it on the full pool of participants, with the exception of some minor wording changes. Appendix Figure 5.A.12 shows screenshots of the survey instruments as experienced by the survey participants.

Our final sample of survey participants contains 796 US taxpayers. Overall, 85% of sur-

[5] Note that there is no uniform definition of tax complexity in the literature (Hoppe, Schanz, Sturm, and Sureth-Sloane 2021).

[6] The pilot survey was ran on June 23rd, 2023.

vey respondents report that they self-file their tax returns, which may include the usage of paid tax preparation software (e.g., TurboTax) but not paid tax preparers.[7] Figure 5.A.1a shows the distribution of income in this sample using income bins. Overall, all income levels below $150,000 appear to be relatively well represented in our sample. There are also some respondents who reported income levels above $150,000. Given that we have elicited income using bins, the median appears to fall in the $60,000 to $80,000 bin, which matches the US median household income ($\approx$ $70,784).

Figure 5.A.1b shows the distribution of the first year a given taxpayer has filed a tax return. The earliest year is 1970 and the latest year is 2023. The median year is 2006. It is hard to know whether this distribution is representative of the US population, although we believe that our sample is likely to be skewed towards more recent first-time filers. However, there are enough early first-time filers to assess whether taxpayers perceive that the tax code has changed significantly over the past two decades.

Figure 5.A.1c shows the age distribution in our sample. The average age is 41, which is consistent with the distribution of the first year of filing a tax return, and the youngest age is 25, which is due to the fact that we have added a restriction that participants should be older than 25. The oldest taxpayer is 86 years old. Moreover, 82% of of the sample are employed full time (all respondents are employed), 55% are male, 79% are White, 8% are African American and 5% are Asian.

Figure 5.A.1d shows the distribution of self-reported political preferences of respondents on the liberal/conservative spectrum. The sample tends to lean towards liberal, with the average respondent falling to left of center, however conservatives are still well represented.

5.2.2 Results

We first analyze how taxpayers perceive income tax filing with respect to complexity. We find that the majority (77%) of taxpayers consider tax filing to be tedious (Figures 5.A.2b and 5.A.2a). The average taxpayer in our sample spends approximately 4 hours filing taxes: 4.4 hours if self-filing (Figure 5.A.3a, Figure 5.A.3b and Figure 5.A.3c) and 3.5 hours when using tax preparation services (Figure 5.A.4a, Figure 5.A.4b and Figure 5.A.4c). Interestingly, neither income nor political preferences appear to matter when it comes to the perception of how complex the tax system is. The mean response for high-income individuals is 31 (where 0 is the perception that the tax system is very complex and 100 is that it is not complex at all). Whereas for low-income individuals it is 32.[8] Similarly, we find that perceived complexity is similar across political preferences: respondents who identify as liberal report a mean complexity of 31 and those that identify as conservative report only a slightly higher mean of 33. This across the board perception of complexity matches well with the fact that tax complexity is a bi-partisan issue in the US, attracting attention from both the conservative and liberal voters and politicians.

[7] The share of self-filers in our sample is higher than in the overall US population. For 2022, the IRS reports that 44% of individual tax returns were self-prepared (link). One explanation for the deviation is that taxpayers using paid tax preparers may be underrepresented on Prolific, e.g., because they are more likely to be tech savvy or higher income.

[8] We divide the sample into below and above median household income based on the 2021 median income for US households of $70,784, as reported by the US Census Bureau (link).

When asked whether tax complexity has changed over time, the majority of taxpayers say that filing income taxes has become more complex (Figure 5.1a). Moreover, as shown in Figure 5.1b, taxpayers with more experience filing taxes perceive that complexity has increased more over time than taxpayers with a more limited experience. This subjective measure of individual taxpayers is in line with our objective word-count measure, discussed below. The average taxpayer in our sample filed their first income tax return in 2006 (Figure 5.A.1b), providing them with 17 years of individual tax filing experience. Over this period, the number of words of the IRC from the Internal Revenue Code has increased by approximately 70%, as we show in Section 5.3.2. We take these results as evidence that tax filing is costly for the vast majority of US taxpayers and that these complexity costs have increased over time.

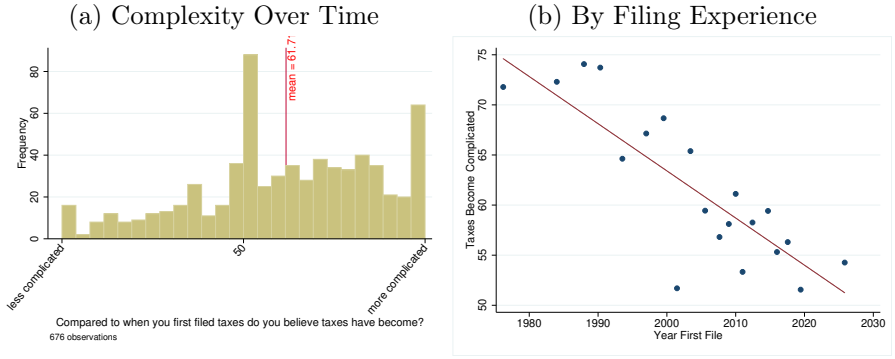


Figure 5.1: Perceived Complexity

Notes: These figures show the respondents' believes about tax complexity. Panel (a) shows the distribution of participants' perception of tax complexity over time. Panel (b) shows a correlation between years of experience filing a tax return and the perception that taxes have become more complicated.

Desired Simplification When asked about the ideal tax filing system, most taxpayers describe it as less complex than the current one. Figure 5.2 shows that 46% of respondents choose the most extreme answer in favor of less complexity. The desire for a simpler tax filing system is almost universal in our sample. Virtually all taxpayers want a simpler tax system, irrespective of income, age, gender, or political affiliation.

In theory, one may be in favor of a more complex tax filing system if it increases the fairness of income taxation, allowing taxpayers to deduct different expenses to account for individual circumstances and/or allowing for more progressivity via more tax brackets and credits for low-income taxpayers, among other reasons. We show that most taxpayers do not share this view. The majority of the respondents believe that complexity is contributing to making the income tax system *less fair* (Figure 5.A.5a). In general, taxpayers share this view across income groups and party affiliation, but the degree to which they believe complexity *reduces* fairness varies by party affiliation as discussed below.

Another argument in favor of a more complex tax filing system is that it may discourage tax evasion by requiring taxpayers to report their income and deductions in more detail. Similarly to the progressivity argument, we find that most taxpayers do not share this view, believing that complexity *encourages* evasion (Figure 5.A.5b).

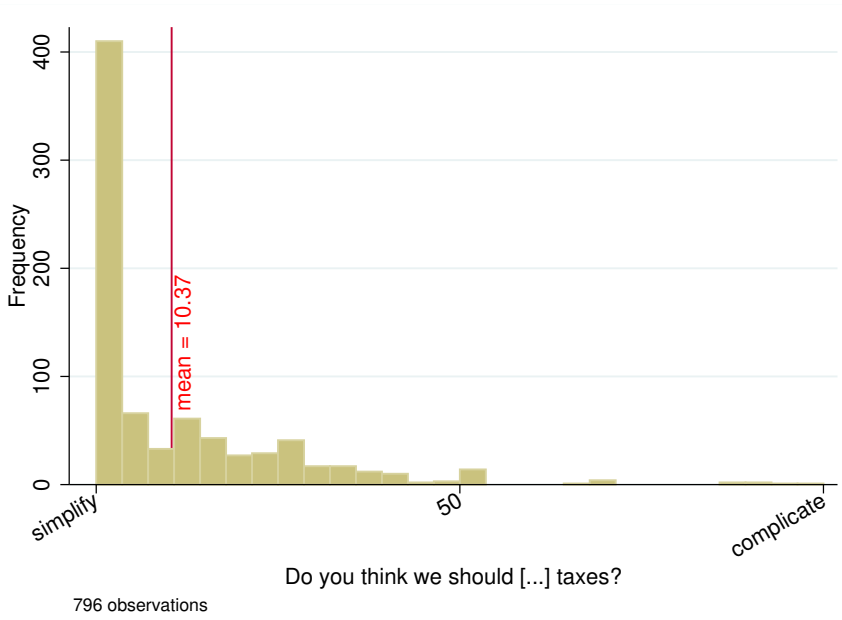


Figure 5.2: Should we simplify taxes?

Notes: This figure shows the distribution of responses to the question of whether taxes should be simplified.

Willingness to Pay for Reduced Complexity Next, we elicit taxpayers' willingness to pay (WTP) for a simplified tax system. We provide respondents with a hypothetical scenario of a simplified tax filing system that offers fewer deduction possibilities than then the current tax system. We explain that this saves time and effort for tax filers, but may lead to a higher tax liability, because there are fewer options to reduce taxable income. We then ask whether taxpayers would be willing to pay more taxes in exchange for such a simplified tax system, and if so, how much more. Note that we ask each of these questions conditional on the response to the question before. We ask for the WTP at the extensive margin (yes/no) only if the respondent stated that they want a less complex tax filing system in the previous question. We then ask for the WTP at the intensive margin (how many dollars) conditional on reporting some WTP at the extensive margin. More specifically, we ask taxpayer how much more in taxes they would be willing to pay if the simplified tax system saves them 50% of their time (or money for tax preparation service users).

We find that the majority of taxpayers are willing to pay more taxes in exchange for a simplified tax filing system (Figure 5.3a), with substantial heterogeneity.[9] The average taxpayer's annual WTP for reducing filing costs is $130 (Figure 5.3c). This is equivalent to about 4% of the average annual tax refund for US taxpayers.[10] This figure is of a similar magnitude as the fees charged by the tax preparation software industry. While these fees vary by the specific tax filing status (what schedules a taxpayer has to file

[9]Figure 5.A.6a shows that income heterogeneity does not matter much for explaining differences in whether taxpayers are willing to pay for simplification.

[10]For the 2022 filing season, the IRS reports an average tax refund of $3,039 (link).

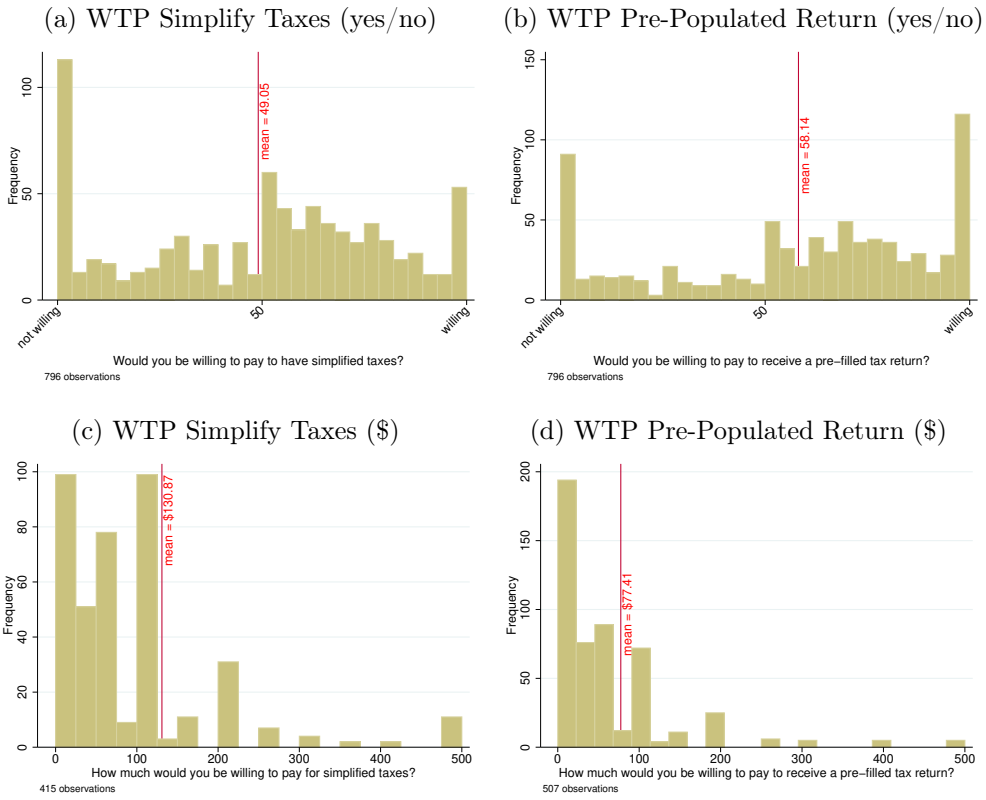


Figure 5.3: Willingness to Pay for Simplifying Taxes and Pre-Populated returns

Notes: These figures show the distribution of survey responses to the questions of whether (panel (a)/panel (b)) and if so, how much (panel (c)/panel (d)) participants would be willing to pay (WTP) for simpler tax filing and pre-populated returns, respectively.

etc.), the fees are commonly around $100 for standard cases. For example, TurboTax charges $129 for a taxpayers who itemize deductions and $69 for those who claim the standard deduction (link with current pricing information).

A Specific Approach to Simplification: Pre-Populated Forms Given the widespread desire and substantial willingness to pay for a simplified tax filing system, we turn to a specific approach to tax simplification: pre-populated forms. Pre-populated forms are ready-made tax return forms, where tax authorities automatically fill in the information that is available to them, such as income and certain deductions. Pre-populated forms reduce filing costs by limiting the amount of information taxpayers have to provide. They are relatively common in other countries (OECD 2021b) and are also being discussed in the US context (Benzarti 2021; Goodman, Lim, Sacerdote, and Whitten 2023).

The majority of taxpayers report that they are willing to pay more taxes if provided with pre-populated forms (Figure 5.3b) with an average annual WTP of $77 (Figure 5.3d).[11] This support for pre-populated forms stands in contrast with the conservative counter-

[11] Figure 5.A.6b shows that income heterogeneity does not matter much for explaining differences in whether taxpayers are willing to pay for pre-populated tax returns.

argument to using them, which is that it is likely to increase tax bills for inattentive taxpayers. This argument is also contradicted by empirical evidence showing that inattentive taxpayers tend to be forgetful and will often leave tax deductions/credits on the table, which are more likely to be claimed when the tax authorities pre-fill them, as shown by Gillitzer and Skov (2018) in Denmark.

Our findings complement the fact that US tax returns could be accurately pre-populated in many cases (Goodman, Lim, Sacerdote, and Whitten 2023) by showing that this would also have support from large parts of the US taxpayer population. Goodman, Lim, Sacerdote, and Whitten (2023) also provide one potential explanation for the observed WTP here: taxpayers for whom pre-populated forms can accurately determine their tax liability, i.e., taxpayers for whom the IRS has all the required information already at hand, often pay for tax preparation services when filing their tax returns. These costs would be dispensable for this group of taxpayers if forms were pre-populated.

Political Preferences Matter (Some) We find that political preferences, with and without controls for income, do not seem to correlate with any of the "factual" questions we ask. Liberals and conservatives report that it takes them similar numbers of hours to file their tax returns. They also perceive tax filing as being equally tedious. And both groups hold similar beliefs over how tedious it is for other individuals to file their taxes.

However, political affiliation does matter when we ask respondents about their opinions as to how to reform the tax system in order to reduce filing and complexity costs. Liberals believe that tax system complexity tends to encourage evasion and decreases fairness more than conservatives. On average both liberals and conservatives are in favor of more tax simplification and adopting pre-populated forms, but liberals are more willing to forgo some of their tax refund in exchange for a simpler tax system and for having access to pre-populated tax returns. These results, which are obtained by running a simple linear regression with linear controls, are reported in Table 5.A.1.

Next, we use observational data from the US and other countries to add external validity to our survey data and put it in context across time and other countries.

5.3 Complexity Over Time and Across Countries

5.3.1 Data

To get an objective measure of the complexity of the income tax code, we count the number of words in the tax code using a simple word count algorithm.[12] The length of the tax code has been previously used as a measure for tax complexity. For example, Bacher and Brülhart (2013) use the number of words to measure Swiss cantonal tax code complexity. Similarly, Moody, Warcholik, and Hodge (2005) and Weber (2015) use it to measure tax code complexity in the US. In a similar vein, Slemrod (2005) measures US state income tax complexity by counting lines of tax forms and pages of instruction booklets.

[12]We begin by tokenizing each income tax code into individual words, employing white space as the separator. We then count the resulting number of these words to determine the overall word count in a given text.

While the number of words is only a proxy for complexity, it has several advantages. First, it accounts for added complexity from additions to the income tax code. Second, it is readily available across time and across countries. Third, it allows for comparisons across different income tax systems from different countries. To further support the use of word counts as a complexity measure in our context, we show, in subsection 5.3.2, that an increase in the number of words in the Internal Revenue Code (IRC) correlates with alternative measures of tax filing complexity in the US.

We collect data on the length of the (income) tax law for six countries on three continents: Canada, France, Germany, Morocco, Switzerland, and the United States. We chose these countries for three main reasons: (1) their tax codes were available online, (2) their tax codes are available in languages we speak and (3) they span a diverse set of countries. We compile panels of annual data for each country. For countries with Federal and state level income taxation, we only consider their Federal tax code. When countries have income tax codes that are separate from the rest of the tax code (Canada, Germany, Switzerland), we only count the income tax code portion. Conversely, in the case of countries with tax codes that do not separate income taxes from the rest (France, Morocco, US), we count the entire tax code. If there is more than one version of the tax code for a given year and country, we always use the version that applies as of December 31 of that year.

Canada For Canada, we use the English version of the Income Tax Act. Historic versions are provided by the Canadian government for years 2004 to 2023 (link). For 2023, we use the latest available version, enacted in June 2023. We count words on the website.

France For France, we use the general tax code (*Code Général des Impôts*). The French government provides historic versions of the tax code dating back as far as January 1, 1979 (link for 2023-01-01). We count words on the website.

Germany For Germany, we use the income tax code (*Einkommensteuergesetz*). The website Juris (link) provides historic versions of the income tax code from 1980 to 2023.[13] For years before the German reunification, i.e. before 1990, we use the income tax law from the Federal Republic of Germany (West Germany). We count words in PDFs we downloaded from Juris.

Morocco For Morocco, we use the French version of the general tax code (*Code Général des Impôts*). The Moroccan government provides the tax code as PDF files for the years 2008, 2009, 2010, 2015, 2021, and 2022 (link 2021 version). We count words in the PDF files.

Switzerland For Switzerland, we use the German version of the Federal income tax code (*Bundesgesetz über die direkte Bundessteuer*). The Swiss government provides historic versions since 1980 (link). We downloaded the tax code as PDF files and then count words in the PDF files.

United States For the US, we use the Internal Revenue Code, which is the Federal tax law. The IRC is provided by the Federal government online for years 1994 to 2021 (link 2021 version). We downloaded the PDF versions of the IRC for each year and

[13] To the best of our knowledge, there is no publicly available archive of the German income tax code.

then count words in the PDF files.

5.3.2 Results

Figure 5.4 plots our word-count based measure of tax complexity, over time. All countries share a common trend: tax code complexity is increasing over time, although there is substantial heterogeneity in levels, i.e., the absolute number of words as shown in Figure 5.A.7. Differences in levels across countries may occur because of language differences as well as because of differences in the national income tax system, but since our analysis is within rather than across countries and over time, this is not an issue. Volatility in the word count within countries over time is explained by changes in the content of the tax code. For example, the German word count increases sharply in 1995, when the country adds child support regulations to its income tax code, expanding the number of paragraphs from 61 to 78. In Switzerland, the word count drops in 2014, when 19 paragraphs with rules for special tax assessment are dropped from the tax code.

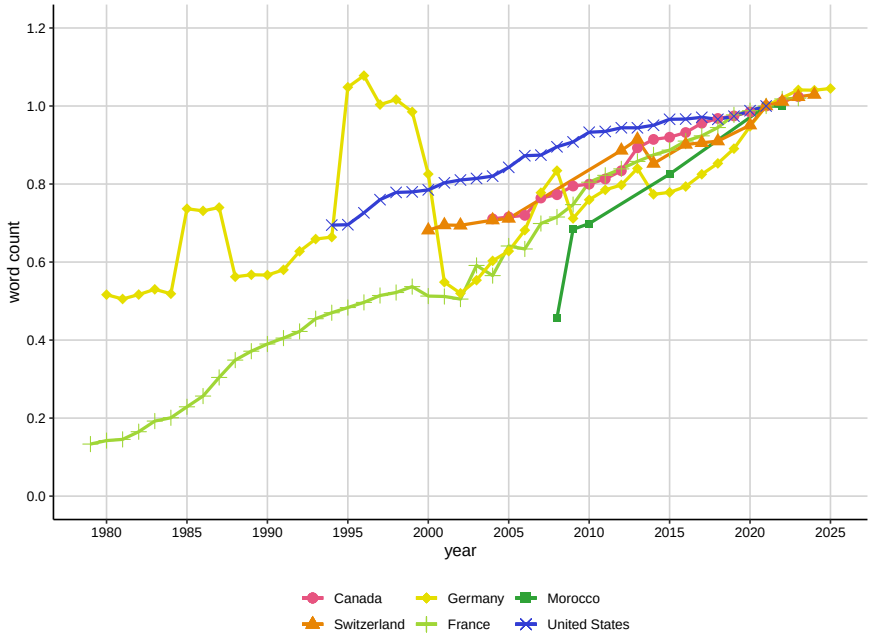


Figure 5.4: Relative Word Count Tax Code Over Time – Cross-Country Comparison

Source: Own calculations. This figure shows how the tax code length develops over time for different countries. To account for differences in levels across countries, we normalize the tax code length, taking 2021 as reference year. We define length of tax code as the number of words in the respective income tax law. For details on the word count, see subsection 5.3.1. Some countries publish the income tax code already in prospective for future years, which is why the x-axis includes years > 2023.

In the US, over the past 30 years, the number of words in the IRC has increased from 3.1 million to 4.3 million (Figure 5.A.7f). In relative terms, this is equivalent to an increase of about 40%. To compare this development with other countries, we normalize the tax code complexity within each country, with 2021 as reference year. Figure 5.4 plots this relative complexity measure over time and shows that other countries have experienced

even larger increases in complexity over time. Notably, the United States is the country with the longest income tax code in our six-country sample.

Alternative Measures To support the use of word counts as a measure for complexity in our context, we examine the trends in alternative complexity measures for the US. During our sample period, the US tax code experienced not only an increase in the number of words but also in the number of sections: between 1954 and 2005 the number of income tax sections increased from 103 to 736 (Moody, Warcholik, and Hodge 2005). In a similar vein, Marcuss et al. (2013) document that the number of subdivisions and cross references in the internal revenue code has increased from less than 50,000 to almost 70,000 between 1991 and 2012. This suggests that more words indicate more complexity because they are used to describe more content.

An increased complexity of the income tax code does not necessarily imply an increase in the complexity of the tax filing process for individual taxpayers. Although taxpayers spend time learning about the tax code, they spend more time filing (see, for example, our survey results in subsection 5.2.2). To assess the complexity of the filing process itself, we provide two measures. First, we analyze how many tax forms taxpayers have filed. Figure 5.A.8 shows that the number of forms filed by the average US taxpayers has increased over time. Second, we analyze the web search behavior for tax filing related queries. Figure 5.A.9 shows that the Google searches for "tax filing", "help filing", "tax preparer" , and "easy tax" have increased over time, indicating that taxpayers increasingly search for help with tax filing. Both measures support that an increased tax code complexity implies an increase in the individual tax filing complexity.

In addition to direct costs from individual tax filing, tax complexity also imposes indirect costs on taxpayers, by making tax administration more costly. Over the past three decades, the operating costs of the IRS have doubled, amounting to $14 billion in 2022 (Figure 5.5a). In per-capita values, this is equivalent to an indirect cost of about 40$ per US inhabitant.

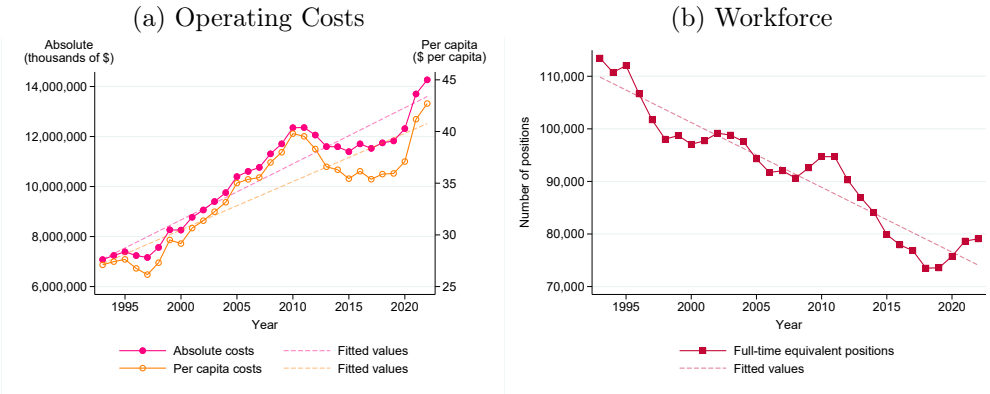


Figure 5.5: IRS Statistics

Source: Data from Internal Revenue Service (2023). Panel a plots the operating costs in absolute terms (pink line, left axis) and as per capita value (orange line, right axis). Per capita value is defined as absolute value divided by US-population. All values are in current US Dollar. Panel b plots the IRS workforce, measured in full-time equivalent positions. Dashed lines show linear predictions.
To download the data visit https://www.irs.gov/pub/irs-soi/22dbs06t31cs.xlsx.

While tax complexity can be costly for individual taxpayers, it contributes to the business model of the tax preparation industry. Revenue from the tax preparation industry provides an alternative measure for the costs of tax filing complexity. We proxy for profits in the tax preparation industry by using the number of employees and the total wages paid in firms with the respective 4-digit NAICS Code (accounting, tax preparation, bookkeeping, and payroll services). Figure 5.A.10 shows that the industry has significantly grown over the past two decades with an increase in employment of about 30%.[14] This lends further support to the fact that tax (filing) complexity in the US has increased over time.

5.4 Discussion

Although some argue that a progressive tax system requires more complexity, allowing for more deduction possibilities does not benefit all taxpayers equally. First, deduction possibilities are likely to vary over the income distribution with high-income taxpayers being more likely to have deductible expenses. Second, conditional on having deductible expenses, the likelihood of claiming those can vary over the income distribution. It is well established that incomplete take-up is pronounced at the lower end of the income distribution. Low-income taxpayers leave money on the table by not claiming EITC benefits they are eligible for (e.g., Bhargava and Manoli 2015; Ramnath and Tong 2017) or not filing an income tax return at all (Goodman, Lim, Sacerdote, and Whitten 2023; Hauck and Wallossek 2023).

Based on our survey results, we find that most taxpayers share the latter view, believing that complexity decreases the fairness of the tax system (see subsection 5.2.2 for details).[15] This view is supported when comparing the trends in complexity and progressivity of the US income tax system over time. We show that tax complexity has increased over the past decades. Effective average tax rates of the (very) rich however have decreased over the past decades, weakening the effective progressivity of the US income tax system (e.g., Piketty, Saez, and Zucman 2018; Saez and Zucman 2019).

In addition, filing costs have been shown to be regressive in the cross-section: lowest income taxpayers face the highest burden, measured relative to income (Marcuss et al. 2013). Consequently, reducing filing costs by reducing complexity is expected to have a progressive effect. Goodman, Lim, Sacerdote, and Whitten (2023) find that prepopulation is most accurate at the bottom of the income distribution. If policy makers want to increase the effective progressivity of the income tax (filing) system, reducing complexity, particularly via pre-populating forms, seems to be a well-suited approach.

Another argument in favor of a more complex tax (filing) system is that complexity may impede tax evasion, requiring more information from tax filers. In the context of French self-employed, Aghion, Akcigit, Gravoueille, Lequien, and Stantcheva (2023) argue that taxpayers choosing a simpler tax regime can partly be explained by evasion. However, with more opportunities for misreporting and mistakes, complexity may also facilitate evasion. Our survey results show that most taxpayers share the latter view, believing that complexity encourages evasion. This is in line with statements from the

[14] This data was obtained from the Bureau of Labor and Statistics.

[15] We decided to ask about fairness rather than progressivity to avoid economic jargon and facilitate understanding for respondents.

Congressional Research Service (2023, link), the Joint Committee on Taxation (2015, link), and the Office of Tax Policy at the US Department of the Treasury (2006, link), who all argue that complexity fosters evasion.

We can further support this view by showing that when tax complexity increases over time, tax evasion does not decrease. The standard measure for income tax evasion in the US is the federal tax gap, defined as the difference between income taxes owed and income taxes paid. Over the past two decades, it has been relatively stable at around 500 billion in 2021$ (Figure 5.A.11a). Over the same period, the income tax code complexity, measured by the number of words, has steadily increased: Between 2001 and 2016, the number of words has increased by about 18% (Figure 5.A.7f).[16] Figure 5.A.11b documents a positive correlation between tax complexity and tax gap.

5.5 Conclusion

Measuring tax complexity is complex. In this paper, we use one simple proxy for complexity – the number of words in the income tax code – and we show that it correlates with both other measures of complexity and with perceived complexity of the tax code as measured by our survey. Understanding complexity at the individual level can help us understand the welfare implications of tax (filing) complexity. Here, further research could shed light on how the length of the income tax code and the number of forms translate into individual filing costs. Particularly, the interplay between tax complexity and progressivity is not fully understood yet. In addition to the cost side of complexity, there is more to learn about the benefit side as well. We provide suggestive evidence that an increasing tax complexity benefits the tax preparation industry, but it is still not well understood how that industry and tax complexity interact. The Biden administration recently mandated that the IRS implement a free filing option, as an alternative to the different options offered by the tax filing industry. It will be interesting to see how this free filing option is implemented, whether it will compete with the non-free tax filing software and if it will affect tax filing and tax bills.

[16] Tax gap estimates constitute a lower bound for tax evasion (International Monetary Fund 2021), particularly for high income taxpayers (Guyton, Langetieg, Reck, Risch, and Zucman 2021). One reason is limited resources for tax enforcement: since the 1990s, the IRS workforce dropped from more than 110,000 full-time equivalent positions to less than 80,000 (Figure 5.5b) and an analysis from the Government Accountability Office shows that this decline comes largely from a decrease in the positions in enforcement (link GAO report). This implies that the true extent of tax evasion in the US is likely higher than the tax gap estimates we report in Figure 5.A.11a.

Appendix to Chapter 5

5.A Additional Figures and Tables

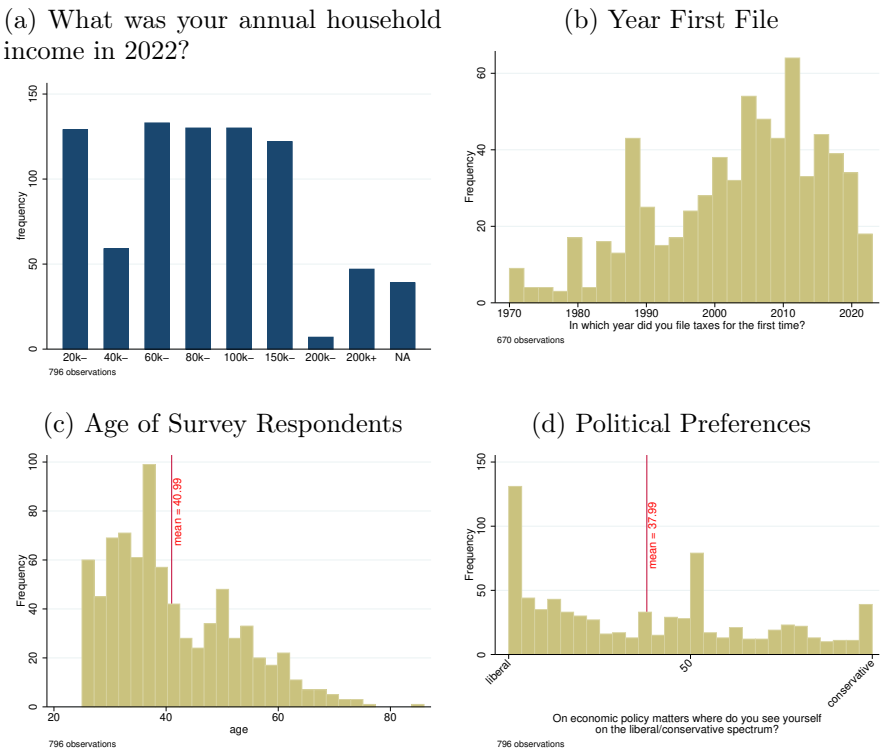


Figure 5.A.1: Survey Participant Demographics

Notes: These figures show some demographics of the survey respondents. Panel (a) shows the distribution of survey answers to the question stated above, where xk- refers to an income bin below $x0,000 and above the previous bin. E.g.: 40k- corresponds to individuals with annual household incomes below $40,000 and above $20,000.. Panel (b) shows the distribution of the first year a given respondent filed their tax return. Panel (c) shows the distribution of the age of each survey participant. Panel (d) plots the distribution of self-stated political preferences.

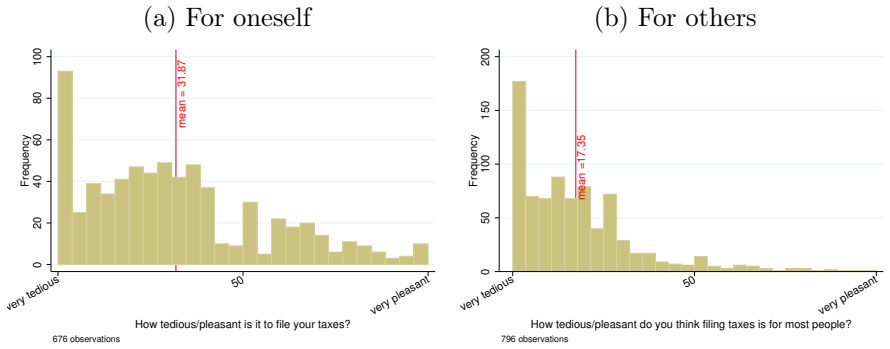


Figure 5.A.2: How tedious/pleasant is it to file taxes?

Notes: These two figures show the distribution of responses to the question of how tedious/pleasant it is to file taxes.

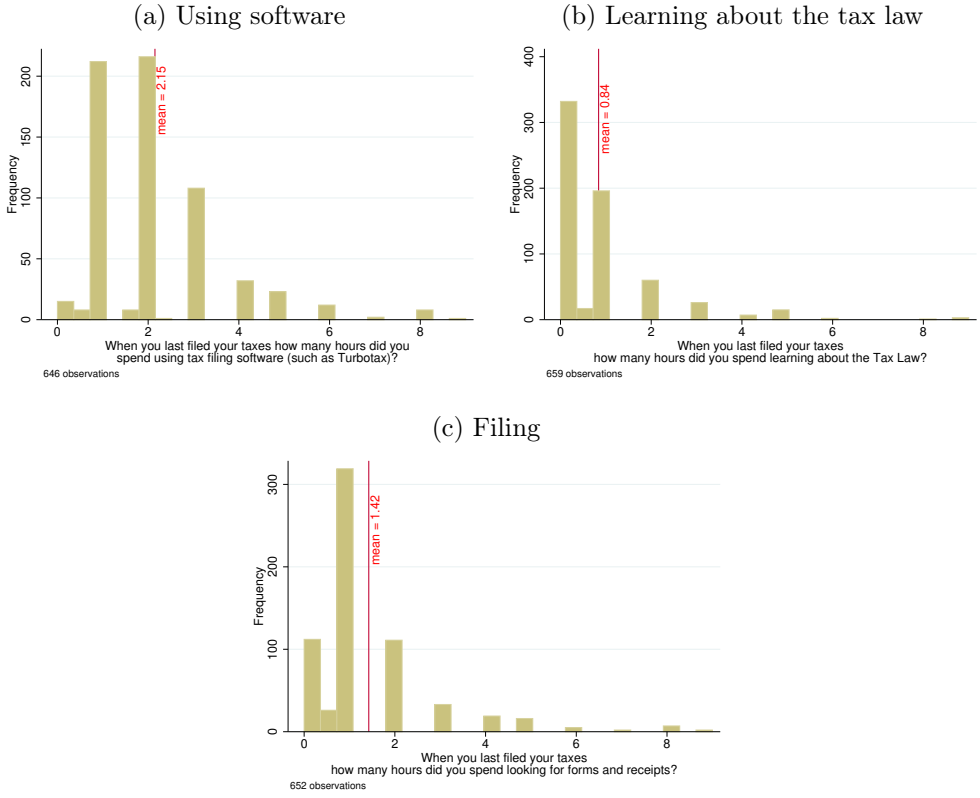


Figure 5.A.3: Hours Spent Filing Taxes if Self File

Notes: These three figures show the distribution of responses to the question of how much time taxpayers who self file spending working on their tax returns.

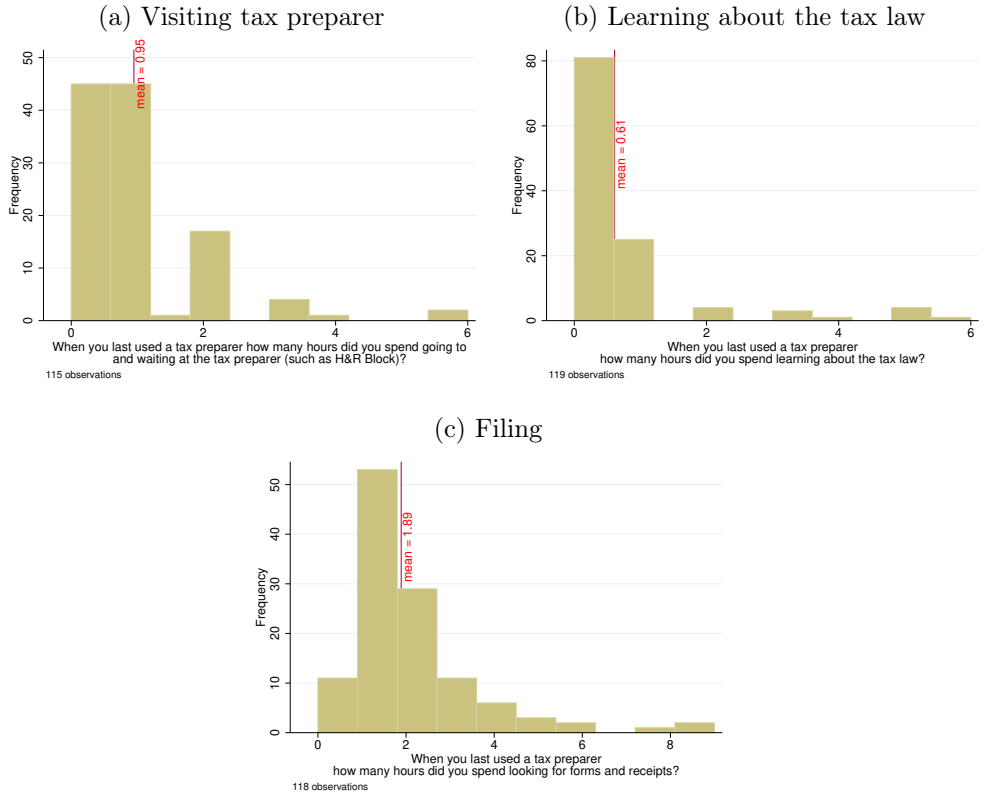


Figure 5.A.4: Hours Spent Filing Taxes if Using Tax Preparer

Notes: These three figures show the distribution of responses to the question of how much time taxpayers who self file spending working on their tax returns.

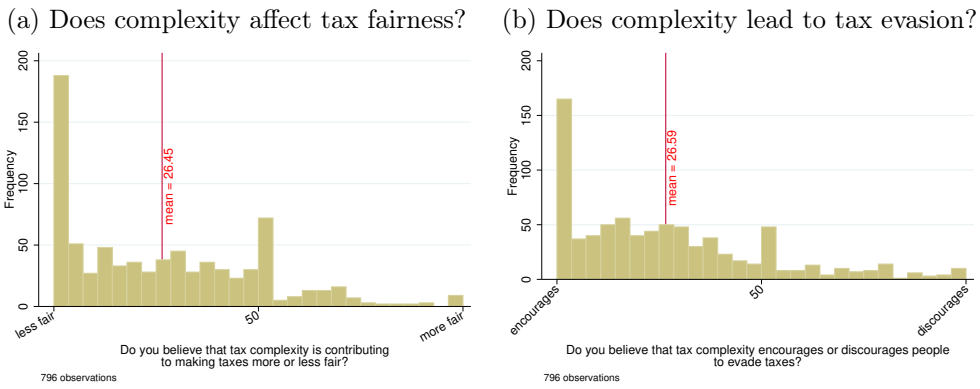
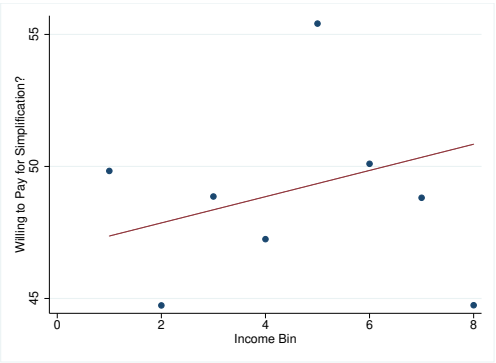


Figure 5.A.5: Complexity, Fairness and Evasion

Notes: These figures show the respondents' attitudes towards tax complexity in the context of fairness and evasion. Panel (a) shows their perception of whether tax complexity leads to a more/less fair tax system and panel (b) shows their perception as to whether complexity leads to more/less tax evasion.

(a) WTP for Simplification



(b) WTP for Pre-Populated Forms

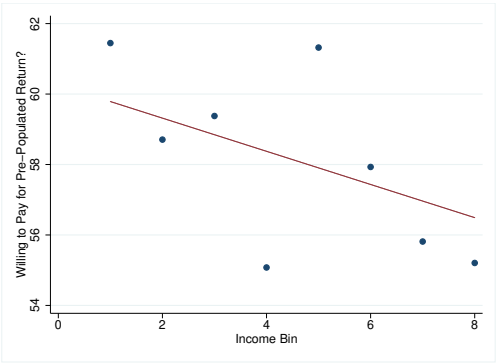


Figure 5.A.6: Heterogeneity

Notes: These Figures show heterogeneity analysis by income and filing experience. Panel (a) shows a correlation between income and whether a taxpayer is willing to pay for simplifying taxes. Panel (b) shows a correlation between income and whether a taxpayer is willing to pay for pre-populated tax returns.

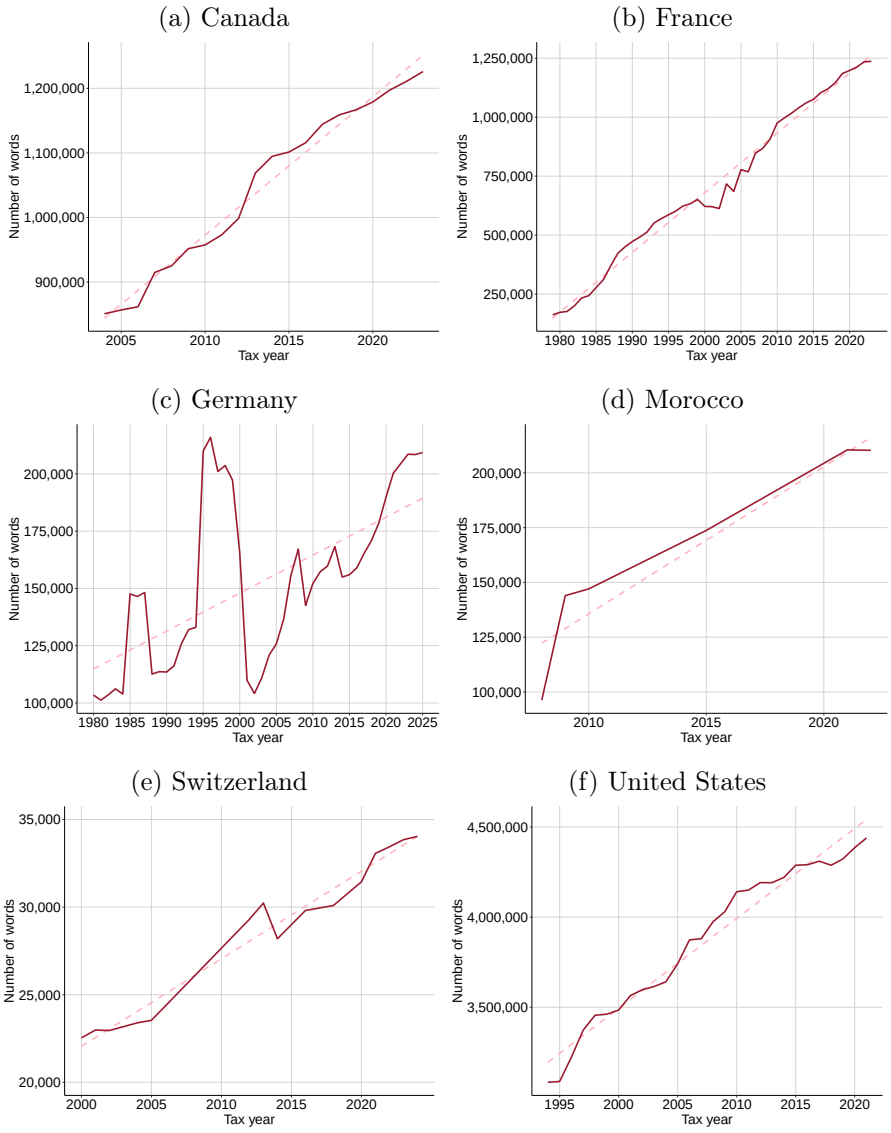


Figure 5.A.7: Word Count Tax Code Over Time – Cross-Country Comparison

Source: Own calculations. The figures show how the length of different tax codes develops over time for Canada, France, Germany, Morocco, Switzerland, and the United States. We define length of tax code as the number of words in the respective income tax law. For details on the word word count, see subsection 5.3.1. Dashed lines show linear predictions.

Figure 5.A.8: Number of Forms Filed Over Time

Source: Data from Benzarti (2020). The figure shows the average number of forms filed per tax return for US tax payers over time. Forms include 1040 and schedules A – F. The dashed line shows the linear prediction.

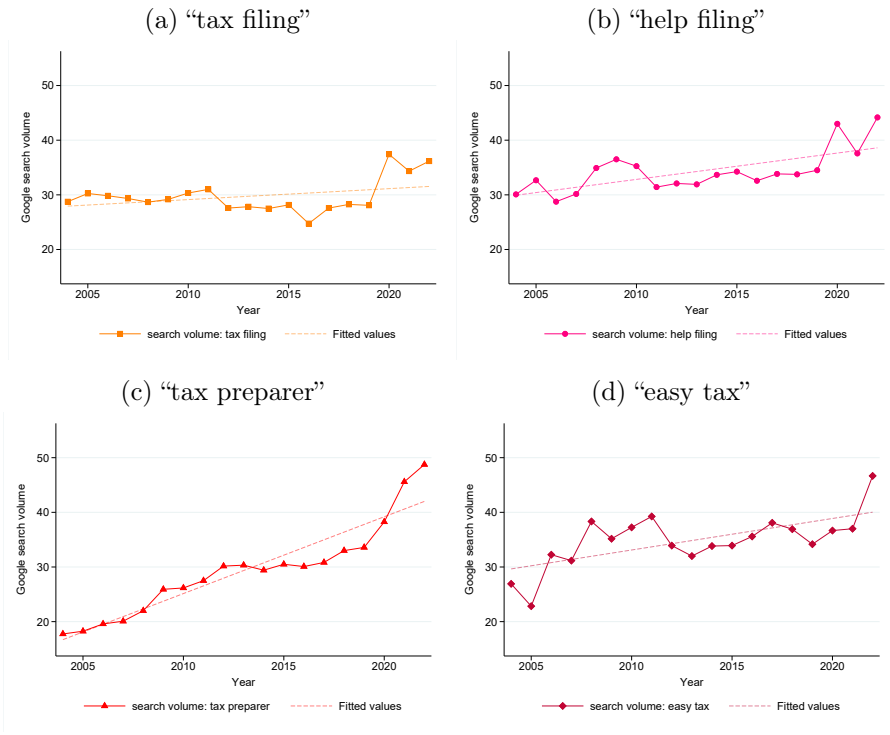


Figure 5.A.9: Google Search Volume

Source: Data from Google trends. Each panel plots the Google search volume for the indicated query in the US between January 2004 and December 2022. The search volume is defined as relative search interest within a given query, ranging from 0 to 100, with 100 indicating the month with peak interest. Plotted values are always > 0 and < 100 because we averaged the monthly data over years to smooth out seasonal search behavior. Dashed lines show linear predictions.
To download the data visit https://trends.google.com/trends/explore?date=all&geo=US and search for the respective key words.

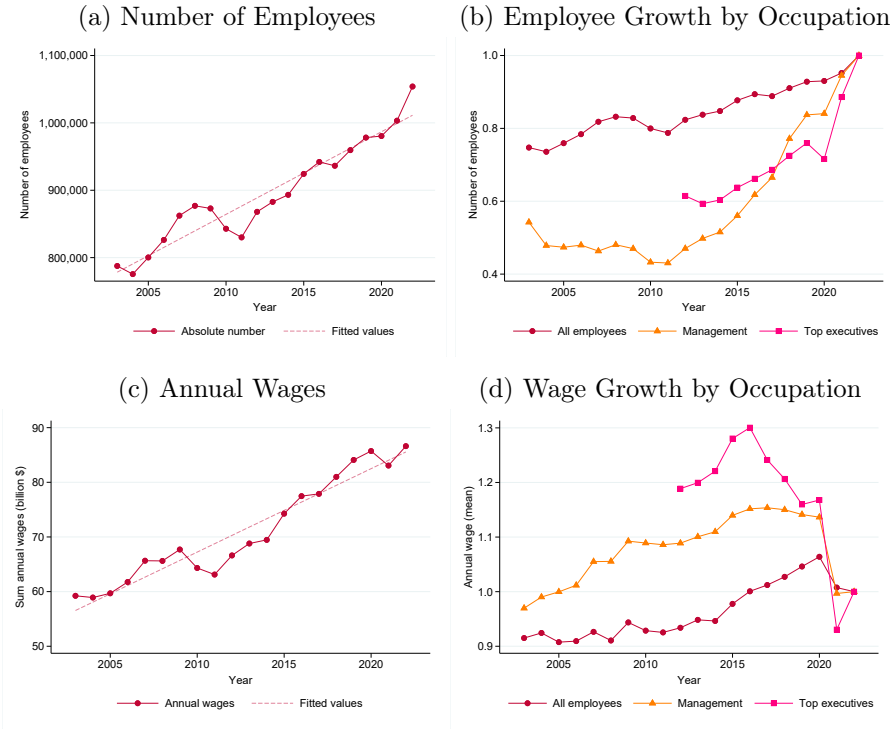


Figure 5.A.10: Tax Preparation Industry

Source: Data from the Occupational Employment and Wage Statistics from the U.S. Bureau of Labor Statistics. All panels refer to NAICS = 541200 "Accounting, Tax Preparation, Bookkeeping, and Payroll Services". All Dollar values are CPI adjusted to 2022 values. Panel a shows the total number of employees over time. Panel b shows the relative growth of the number of employees for all employees (this corresponds to panel a), as well as for the two subgroups of management and top executive employees, with 2022 being the reference year. Data on top executives is only available from 2012 onward. Panel c plots the inflation adjusted sum of annual wages paid to all employees. Panel d shows the growth of the mean inflation adjusted wage for all employees (this corresponds to the sum plotted in panel c), as well as for the two subgroups of management and top executive employees, with 2022 being the reference year.
To download the data visit https://www.bls.gov/oes/special-requests/oesmYYin4.zip and replace YY with the last two digits of a given calendar year 20YY (link for 2022 data). Data is available for 2003 to 2022.

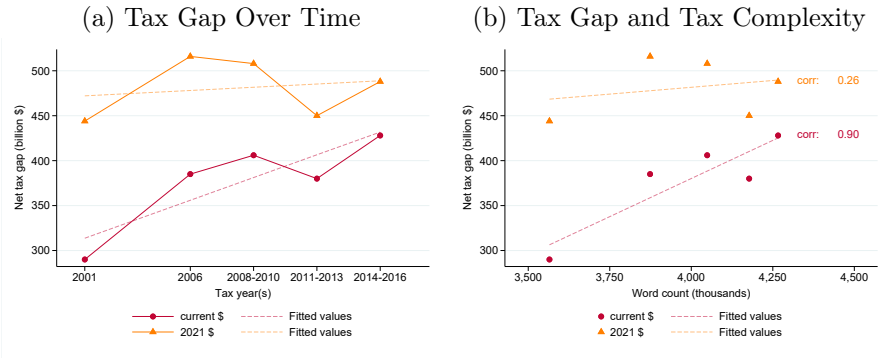


Figure 5.A.11: The US Tax Gap

Source: Tax gap estimates from the IRS as reported by Congressional Research Service (2023). Panel a plots the net federal tax gap for the US over time in current US Dollars and in 2021 US Dollars. Panel b plots the same values over the length of the US Federal tax code. When the tax gap estimate is for > 1 year, we take the average tax code length across the years. Dashed lines show linear predictions and *corr* (panel b) reports the correlation coefficients.

Figure 5.A.12: Survey Visual Examples

Notes: These figures show some visual examples of the survey as experienced by the survey participants.

	(1)	(2)	(3)	(4)	(5)	(6)
Liberal	-0.62	-6.98	-7.02	-1.53	4.32	7.86
	(1.99)	(1.72)	(1.83)	(1.12)	(2.25)	(2.46)
Income controls	yes	yes	yes	yes	yes	yes
Constant	65.4	30.4	31.7	12.7	47.5	53.2
	(2.69)	(2.05)	(2.31)	(1.43)	(2.92)	(3.35)
Observations	676	796	796	796	796	796
R^2	0.026	0.027	0.024	0.019	0.017	0.020

Notes: Each column in this table reports the regression of an outcome variable defined below on an indicator variable for whether a given respondent identifies as liberal and income controls. The outcome variables we consider are answers to the following questions: (1) "Compared to when you first filed taxes do you believe taxes have become more complicated/less complicated?", (2) "Do you believe that tax complexity is contributing to making taxes more or less fair?", (3) "Do you believe that tax complexity encourages or discourages people to evade taxes?", (4) "Do you think we should simplify/complicate taxes?", (5) "Would you be willing to pay to have simplified taxes?" and (6) "Would you be willing to pay to receive a pre-filled tax return?". Robust standard errors in parentheses.

Table 5.A.1: Political Preferences and Attitudes Towards Tax Filing

5.B Survey Questions

1. How do you usually file taxes?
    - I usually file taxes myself (including: using a software such as TurboTax etc.)
    - Someone else files my taxes (e.g., spouse, parent, etc.)
    - I usually pay a tax preparer (such as H&R Block) to file my taxes
    - I don't usually file a tax return
2. (If responded "I usually file taxes myself" in question 1.) When you last filed your taxes, how many hours did you spend
    - Looking for forms and receipts
    - Learning about the Tax Law
    - Using the tax filing software (such as Turbotax)
3. (If responded "I usually pay a tax preparer" in question 1.) When you last used a tax preparer, how many hours did you spend
    - Looking for forms and receipts
    - Going to and wait at the tax preparer (such as H&R Block)
    - Learning about the Tax Law
4. (If responded "I usually file taxes myself" in question 1.) How tedious/pleasant is it to file your taxes?
    - slider from 0 (very tedious) to 100 (very pleasant)
5. How tedious do you think filing taxes is for most people?
    - slider from 0 (very tedious) to 100 (very pleasant)
6. Do you think we should simplify/complicate taxes
    - slider from 0 (simplify a lot more) to 100 (complicate a lot more)
7. (If responded "I usually file taxes myself" in question 1.) Compared to when you first filed taxes, do you believe taxes have become
    - slider from 0 (much less complicated) to 100 (much more complicated)
8. Do you believe that tax complexity is contributing to making taxes more or less fair?
    - slider from 0 (much less fair) to 100 (much more fair)
9. Do you believe that tax complexity encourages or discourages people to evade taxes?
    - slider from 0 (encourages tax evasion) to 100 (discourages tax evasion)
10. Would you be willing to pay to have simplified taxes?

- slider from 0 (not willing at all) to 100 (very willing)

11. (If responded more than 50 in question 10.) How much would you be willing to pay for simplified taxes?

    - Text entry

12. Pre-filled tax returns are forms that are automatically filled in by the IRS based on information they already have. These exist in other countries, but not in the US. Taxpayers can review and verify the pre-filled information and make any necessary adjustments or additions before submitting the form. Would you be willing to pay to receive a pre-filled tax return?

    - slider from 0 (not willing at all) to 100 (very willing)

13. (If responded more than 50 in question 12.) How much would you be willing to pay to receive a pre-filled tax return?

    - Text entry

14. What was your annual household income in 2022?

    - Less than $20,000
    - $20,000 to $40,000
    - $40,000 to $60,000
    - $60,000 to $80,000
    - $80,000 to $100,000
    - $100,000 to $150,000
    - $150,000 to $200,000
    - More than $200,000
    - I don't know
    - I prefer not to answer

15. On economic policy matters, where do you see yourself on the liberal/conservative spectrum?

    - slider from 0 (very liberal) to 100 (very conservative)

Bibliography

Abeler, Johannes and Simon Jäger (2015). "Complex Tax Incentives." *American Economic Journal: Economic Policy*, 7 (3), 1–28. doi: 10.1257/pol.20130137.

Advani, Arun, Helen Hughson, and Andy Summers (2023). "How much tax do the rich really pay? Evidence from the UK." *Oxford Review of Economic Policy*, 39 (3), 406–437. doi: 10.1093/oxrep/grad032.

Aghion, Philippe, Ufuk Akcigit, Maxime Gravoueille, Matthieu Lequien, and Stefanie Stantcheva (2023). *Tax Simplicity or Simplicity of Evasion? Evidence from Self-Employment Taxes in France*. url: https://www.maximegravoueille.com/research/AALGS.pdf (last accessed on 08/23/2023).

Alesina, Alberto and Nicola Fuchs-Schündeln (2007). "Good-bye Lenin (or Not?): The Effect of Communism on People's Preferences." *American Economic Review*, 97 (4), 1507–1528. doi: 10.1257/aer.97.4.1507.

Alstadsæter, Annette, Niels Johannesen, Ségal Le Guern Herry, and Gabriel Zucman (2022). "Tax evasion and tax avoidance." *Journal of Public Economics*, 206, 104587. doi: 10.1016/j.jpubeco.2021.104587.

Alstadsæter, Annette, Niels Johannesen, and Gabriel Zucman (2019). "Tax Evasion and Inequality." *American Economic Review*, 109 (6), 2073–2103. doi: 10.1257/aer.20172043.

Andresen, Martin and Emily Nix (2021). "What Causes the Child Penalty? Evidence from Adopting and Same-Sex Couples." *Journal of Labor Economics*, Forthcoming. Publisher: The University of Chicago Press. doi: 10.1086/718565.

Angelov, Nikolay, Per Johansson, and Erica Lindahl (2016). "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics*, 34 (3), 545–579. doi: 10.1086/684851.

Artmann, Elisabeth, Nicola Fuchs-Schündeln, and Giulia Giupponi (2023). *Forward-Looking Labor Supply Responses to Changes in Pension Wealth: Evidence from Germany*. SSRN Scholarly Paper. doi: 10.2139/ssrn.4447094.

Austrian Federal Ministry of Finance (2018). *Anfragebeantwortung durch den Bundesminister für Finanzen Hartwig Löger zu der schriftlichen Anfrage (3424/J-BR/2018) der Bundesräte David Stögmüller, Kolleginnen und Kollegen an den Bundesminister für Finanzen betreffend Öffentlichkeitsarbeit im BMF 2017*. url: https://www.parlament.gv.at/PAKT/VHG/BR/AB-BR/AB-BR_03170/imfname_687503.pdf (last accessed on 02/02/2022).

— (2022). *Employee assessment without application ("automatic tax equalisation")*. url: https://www.oesterreich.gv.at/en/themen/steuern_und_finanzen/arbeitnehmerveranlagung/Seite.340002.html (last accessed on 04/03/2022).

Bacher, Hans Ulrich and Marius Brülhart (2013). "Progressive taxes and firm births." *International Tax and Public Finance*, 20 (1), 129–168. doi: 10.1007/s10797-012-9218-z.

Bartels, Charlotte and Katharina Jederny (2015). "The Role of Capital Income for Top Income Shares in Germany." *The World Top Incomes Database Working Paper*, (2015/1).

Bartels, Charlotte and Cortnie Shupe (2023). "Drivers of participation elasticities across Europe: gender or earner role within the household?" *International Tax and Public Finance*, 30 (1), 167–214. doi: 10.1007/s10797-021-09711-z.

Becker, Gary S. (1985). "Human Capital, Effort, and the Sexual Division of Labor." *Journal of Labor Economics*, 3 (1), 33–58. url: http://www.jstor.org/stable/2534997.

— (1991). *A treatise on the family*. Enl. ed. Cambridge, Mass: Harvard University Press.

Benartzi, Shlomo and Richard H. Thaler (2007). "Heuristics and Biases in Retirement Savings Behavior." *Journal of Economic Perspectives*, 21 (3), 81–104. doi: 10.1257/jep.21.3.81.

Benzarti, Youssef (2020). "How Taxing Is Tax Filing? Using Revealed Preferences to Estimate Compliance Costs." *American Economic Journal: Economic Policy*, 12 (4), 38–57. doi: 10.1257/pol.20180664.

— (2021). "Estimating the Costs of Filing Tax Returns and the Potential Savings from Policies Aimed at Reducing These Costs." *Tax Policy and the Economy*, 35. Publisher: The University of Chicago Press, 55–85. doi: 10.1086/713493.

Benzarti, Youssef and Luisa Wallossek (2024). "Rising Income Tax Complexity." *National Tax Journal*, 77 (1), 143–173. doi: 10.1086/728586.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov (2015). "The Welfare Economics of Default Options in 401(k) Plans." *American Economic Review*, 105 (9), 2798–2837. doi: 10.1257/aer.20130907.

Berniell, Inés, Lucila Berniell, Dolores de la Mata, María Edo, Yarine Fawaz, Matilde P. Machado, and Mariana Marchionni (2022). "Motherhood, pregnancy or marriage effects?" *Economics Letters*, 214, 110462. doi: 10.1016/j.econlet.2022.110462.

Berresheim, Ursula and David Koll (2023). *Staying Together Forever? Life-Cycle Effects of Overoptimistic Couples*.

Bertrand, Marianne (2011). "New Perspectives on Gender." *Handbook of Labor Economics*. Vol. 4. Elsevier, 1543–1590. doi: 10.1016/S0169-7218(11)02415-4.

Bertrand, Marianne, Claudia Goldin, and Lawrence F Katz (2010). "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics*, 2 (3), 228–255. doi: 10.1257/app.2.3.228.

Beshears, John, James J. Choi, David Laibson, Brigitte C. Madrian, and Sean (Yixiang) Wang (2016). *Who Is Easier to Nudge?* URL: https://scholar.harvard.edu/laibson/publications/who-easier-nudge (last accessed on 01/21/2020).

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor (2017). "Choose to Lose: Health Plan Choices from a Menu with Dominated Option." *The Quarterly Journal of Economics*, 132 (3), 1319–1372. DOI: 10.1093/qje/qjx011.

Bhargava, Saurabh and Dayanand 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. DOI: 10.1257/aer.20121493.

Bick, Alexander and Nicola Fuchs-Schündeln (2018). "Taxation and Labour Supply of Married Couples across Countries: A Macroeconomic Analysis." *The Review of Economic Studies*, 85 (3), 1543–1576. DOI: 10.1093/restud/rdx057.

Blau, Francine D. and Lawrence M. Kahn (2007). "Changes in the Labor Supply Behavior of Married Women: 1980–2000." *Journal of Labor Economics*, 25 (3). Publisher: The University of Chicago Press, 393–438. DOI: 10.1086/513416.

— (2017). "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature*, 55 (3), 789–865. DOI: 10.1257/jel.20160995.

Blesse, Sebastian, Florian Buhlmann, and Philipp Dörrenberg (2019). "Do People Really Want a Simple Tax System? Evidence on Preferences Towards Income Tax Simplification." *SSRN Electronic Journal*. DOI: 10.2139/ssrn.3515915.

Blumenstock, Joshua, Michael Callen, and Tarek Ghani (2018). "Why Do Defaults Affect Behavior? Experimental Evidence from Afghanistan." *American Economic Review*, 108 (10), 2868–2901. DOI: 10.1257/aer.20171676.

Boelmann, Barbara, Anna Raute, and Uta Schönberg (2021). *Wind of Change? Cultural Determinants of Maternal Labor Supply*. CESifo Working Paper No. 9094. URL: https://www.cesifo.org/en/publications/2021/working-paper/wind-change-cultural-determinants-maternal-labor-supply (last accessed on 07/17/2023).

Boneva, Teodora, Marta Golin, Katja Kaufmann, and Christopher Rauh (2022). *Beliefs About Maternal Labor Supply*. IZA DP No. 15788. URL: https://www.iza.org/publications/dp/15788/beliefs-about-maternal-labor-supply (last accessed on 03/11/2024).

Borella, Margherita, Mariacristina De Nardi, and Fang Yang (2023). "Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labour Supply?" *The Review of Economic Studies*, 90 (1), 102–131. DOI: 10.1093/restud/rdac018.

Bredtmann, Julia and Christina Vonnahme (2019). "Less money after divorce – how the 2008 alimony reform in Germany affected spouses' labor supply, leisure and marital stability." *Review of Economics of the Household*, 17 (4), 1191–1223. DOI: 10.1007/s11150-019-09448-z.

Brockmeyer, Anne and Marco Hernandez (2019). *Taxation, Information, and Withholding: Evidence from Costa Rica*. URL: https://www.annebrockmeyer.com/

uploads/1/2/1/4/121485108/taxation_information_and_withholding.pdf (last accessed on 02/24/2022).

Brugiavini, Agar, Cristina E. Orso, Mesfin G. Genie, Rinaldo Naci, and Giacomo Pasini (2022). *SHARE Job Episodes Panel*. DOI: 10.6103/SHARE.JEP.800.

Bucher-Koenen, Tabea, Rob J. Alessie, Annamaria Lusardi, and Maarten van Rooij (2021). *Fearless Woman: Financial Literacy and Stock Market Participation*. NBER Working Paper 28723. DOI: 10.3386/w28723.

Bucher-Koenen, Tabea and Annamaria Lusardi (2011). "Financial literacy and retirement planning in Germany." *Journal of Pension Economics & Finance*, 10 (4), 565–584. DOI: 10.1017/S1474747211000485.

Bucher-Koenen, Tabea, Annamaria Lusardi, Rob Alessie, and Maarten van Rooij (2017). "How Financially Literate Are Women? An Overview and New Insights." *Journal of Consumer Affairs*, 51 (2), 255–283. DOI: https://doi.org/10.1111/joca.12121.

Bundesregierung (2021). *Mehr Fortschritt wagen. Bündnis Für Freiheit, Gerechtigkeit und Nachhaltigkeit. Koalitionsvertrag zwischen SPD, BÜNDNIS 90/DIE GRÜNEN und FDP*. URL: https://www.bundesregierung.de/resource/blob/974430/1990812/04221173eef9a6720059cc353d759a2b/2021-12-10-koav2021-data.pdf?download=1 (last accessed on 04/03/2022).

Butt, Adam, M. Scott Donald, F. Douglas Foster, Susan Thorp, and Geoffrey J. Warren (2018). "One size fits all? Tailoring retirement plan defaults." *Journal of Economic Behavior & Organization*, 145, 546–566. DOI: 10.1016/j.jebo.2017.11.022.

Caldwell, Sydnee, Scott Nelson, and Daniel Waldinger (2023). "Tax Refund Uncertainty: Evidence and Welfare Implications." *American Economic Journal: Applied Economics*, 15 (2), 352–76. DOI: 10.1257/app.20210383.

Chetty, Raj, John N. Friedman, Søren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen (2014). "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark." *Quarterly Journal of Economics*, 129 (3), 1141–1219. DOI: 10.1093/qje/qju013.

Choi, James J., David Laibson, Brigitte C. Madrian, and Andrew Metrick (2004). "For Better or for Worse. Default Effects and 401(k) Savings Behavior." *Perspectives on the Economics of Aging*. Chicago: Wise, David A., 81–126.

Christl, Michael, Silvia De Poli, and Viginta Ivaškaitė-Tamošiūnė (2023). "Does it pay to say 'I do'? Marriage bonuses and penalties across the EU." *Journal of European Social Policy*, 33 (3), 317–336. DOI: 10.1177/09589287231159492.

Clark, Robert L and Denis Pelletier (2019). *Impact of Defaults in Retirement Saving Plans: Public Employee Plans*. NBER Working Paper 26234. DOI: 10.3386/w26234.

Congressional Research Service (2023). "Federal Tax Gap: Size, Contributing Factors, and the Debate over Reducing It (Version 4, Updated)." *In Focus*, (1887). URL: https://crsreports.congress.gov/product/pdf/IF/IF11887.

Cribb, Jonathan and Carl Emmerson (2020). "What happens to workplace pension saving when employers are obliged to enrol employees automatically?" *International Tax and Public Finance*, 27 (3), 664–693. DOI: 10.1007/s10797-019-09565-6.

Currie, Janet (2006). "The Take-Up of Social Benefits." *Public Policy and the Income Distribution*. Ed. by Alan J. Auerbach, David Card, and John M. Quigley. Russell Sage Foundation, 80–148. URL: https://www.jstor.org/stable/10.7758/9781610440202.7.

Department of the Treasury, Internal Revenue Service (IRS) (2013). *Publication 15. Cat. No. 10000W. (Circular E), Employer's Tax Guide. For use in 2014*. Tech. rep. URL: https://www.irs.gov/pub/irs-prior/p15--2014.pdf (last accessed on 01/29/2023).

Deutsche Rentenversicherung Bund (2017). *Datensatz SK 79 für die Versicherungskontenstichprobe gemäß § 1 Abs. 2 RSVwV ab der Erhebung zum 31.12.2016 und für die Sondererhebung über vollendete Versichertenleben sowie für die Sondererhebung zum Projekt SHARE*.

Doerrenberg, Philipp, Andreas Peichl, and Sebastian Siegloch (2017). "The elasticity of taxable income in the presence of deduction possibilities." *Journal of Public Economics*, 151, 41–55. DOI: 10.1016/j.jpubeco.2015.10.001.

Erard, Brian and Chih-Chin Ho (2001). "Searching for ghosts: who are the nonfilers and how much tax do they owe?" *Journal of Public Economics*, 81 (1), 25–50.

European Commission (2019). *German pre-filled tax return*. URL: https://ec.europa.eu/cefdigital/wiki/cefdigital/wiki/display/CEFDIGITAL/2019/07/31/German+pre-filled+tax+return (last accessed on 04/03/2022).

— (2024). *Gender equality strategy. Achievements and key areas for action*. URL: https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en (last accessed on 03/09/2024).

Eurostat (2023). *Gender overall earnings gap*. URL: https://ec.europa.eu/eurostat/databrowser/view/teqges01/default/table (last accessed on 07/28/2023).

Federal Statistical Office (2022a). *71211-0001: Revenue from taxes: Germany, years, types of taxes prior to distribution*. URL: https://www-genesis.destatis.de/genesis/online?sequenz=tabellen&language=en&selectionname=71211-0001 (last accessed on 04/03/2022).

— (2022b). *81000-0003: National accounts - Gross national income, national income: Germany, years*. URL: https://www-genesis.destatis.de/genesis/online?sequenz=tabellen&language=en&selectionname=81000-0003 (last accessed on 04/03/2022).

Feldman, Naomi E., Peter Katuščák, and Laura Kawano (2016). "Taxpayer Confusion: Evidence from the Child Tax Credit." *American Economic Review*, 106 (3), 807–835. DOI: 10.1257/aer.20131189.

Findeisen, Sebastian, Sang Yoon (Tim) Lee, Tommaso Porzio, and Wolfgang Dauth (2021). *Transforming Institutions. Labor Reallocation and Wage Growth in a Reunified Germany.*

Finkelstein, Amy and Matthew J. Notowidigdo (2019). "Take-Up and Targeting: Experimental Evidence from SNAP." *The Quarterly Journal of Economics*, 134 (3). Publisher: Oxford Academic, 1505–1556. doi: 10.1093/qje/qjz013.

Fischer, Florian and Josef Milz (2017). "Arbeitnehmerveranlagung 2014." *Statistische Nachrichten*, (3), 212–225.

Foerster, Hanno (2022). *Untying the Knot: How Child Support and Alimony Affect Couples' Dynamic Decisions and Welfare.* Working Paper.

Gabaix, Xavier (2019). "Behavioral inattention." *Handbook of Behavioral Economics: Applications and Foundations 2.* Vol. 2. Elsevier, 261–343. doi: 10.1016/bs.hesbe.2018.11.001.

Gillitzer, Christian and Peer Ebbesen Skov (2018). "The use of third-party information reporting for tax deductions: evidence and implications from charitable deductions in Denmark." *Oxford Economic Papers*, 70 (3), 892–916. doi: 10.1093/oep/gpx055.

Giommoni, Tommaso and Enrico Rubolino (2022). *The Cost of Gender Identity Norms: Evidence from a Spouse Tax Credit.* Dondena Working Paper No. 153.

Goda, Gopi Shah, Matthew R. Levy, Colleen Flaherty Manchester, Aaron Sojourner, and Joshua Tasoff (2020). "Who is a passive saver under opt-in and auto-enrollment?" *Journal of Economic Behavior & Organization*, 173, 301–321. doi: 10.1016/j.jebo.2019.08.026.

Goebel, Jan, Markus M. Grabka, Carsten Schröder, Hannah Penz, Tatjana Mika, Daniel Brüggmann, Sebastian Ellert, Katharina Werhan, Forschungsdatenzentrum Der Rentenversicherung, and Deutsches Institut Für Wirtschaftsforschung (DIW Berlin) (2022). *SOEP-RV: Linking German Socio-Economic Panel Data to Pension Records.* doi: 10.5684/SOEP.V37-RV.VSKT2020.

Goldin, Jacob, Tatiana Homonoff, Rizwan Javaid, and Brenda Schafer (2022). "Tax filing and take-up: Experimental evidence on tax preparation outreach and benefit claiming." *Journal of Public Economics*, 206, 104550. doi: 10.1016/j.jpubeco.2021.104550.

González, Libertad and Berkay Özcan (2013). "The Risk of Divorce and Household Saving Behavior." *Journal of Human Resources*, 48 (2), 404–434. doi: 10.1353/jhr.2013.0013.

Goodman, Lucas (2020). "Catching up or crowding out? The crowd-out effects of catch-up retirement contributions on non-retirement saving." *Journal of Public Economics*, 188, 104221. doi: 10.1016/j.jpubeco.2020.104221.

Goodman, Lucas, Katherine Lim, Bruce Sacerdote, and Andrew Whitten (2023). "Automated Tax Filing: Simulating a Prepopulated Form 1040." *National*

Tax Journal, 76 (4). Publisher: The University of Chicago Press, 805–838. DOI: 10.1086/726592.

Gruber, Jon and Emmanuel Saez (2002). "The elasticity of taxable income: evidence and implications." *Journal of Public Economics*, 84 (1), 1–32. DOI: 10.1016/S0047-2727(01)00085-8.

Guyton, John, Pat Langetieg, Day Manoli, Mark Payne, Brenda Schafer, and Michael Sebastiani (2017). "Reminders and Recidivism: Using Administrative Data to Characterize Nonfilers and Conduct EITC Outreach." *American Economic Review: Papers & Proceedings*, 107 (5), 471–475. DOI: 10.1257/aer.p20171062.

Guyton, John, Patrick Langetieg, Daniel Reck, Max Risch, and Gabriel Zucman (2021). *Tax Evasion at the Top of the Income Distribution: Theory and Evidence*. NBER Working Paper 28542. DOI: 10.3386/w28542.

Hauck, Tobias and Luisa Wallossek (2021). "Optional (Non-)Filing and Effective Taxation." Publisher: Universitätsbibliothek der Ludwig-Maximilians-Universität München. DOI: 10.5282/UBM/EPUB.77429.

— (2023). *Optional (Non-)Filing and Effective Taxation*. URL: https://www.dropbox.com/s/z9mv19h3fwpb45u/Hauck_Wallossek_Tax_Non_Filing.pdf?dl=0 (last accessed on 02/26/2024).

— (2024). "Optional (non-)filing and effective taxation." *Journal of Public Economics*, 238, 105187. DOI: 10.1016/j.jpubeco.2024.105187.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou (2021). "Inattention and Switching Costs as Sources of Inertia in Medicare Part D." *American Economic Review*, 111 (9), 2737–2781. DOI: 10.1257/aer.20170471.

Hermle, Johannes and Andreas Peichl (2018). *Jointly Optimal Taxes for Different Types of Income*. DOI: 10.2139/ssrn.3275422.

Hill, Martha S. (1979). "The Wage Effects of Marital Status and Children." *The Journal of Human Resources*, 14 (4). Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], 579–594. DOI: 10.2307/145325.

Hoppe, Thomas, Deborah Schanz, Susann Sturm, and Caren Sureth-Sloane (2021). "The Tax Complexity Index – A Survey-Based Country Measure of Tax Code and Framework Complexity." *European Accounting Review*, 1–35. DOI: 10.1080/09638180.2021.1951316.

Imbens, Guido W. and Thomas Lemieux (2008). "Regression discontinuity designs: A guide to practice." *Journal of Econometrics*. The regression discontinuity design: Theory and applications, 142 (2), 615–635. DOI: 10.1016/j.jeconom.2007.05.001.

Internal Revenue Service (2023). *Data Book, 2022*. URL: https://www.irs.gov/pub/irs-pdf/p55b.pdf (last accessed on 10/07/2023).

International Bureau of Fiscal Documentation (2016). *Global Individual Tax Handbook 2016*. Amsterdam: IBFD.

International Monetary Fund (2021). *The Revenue Administration Gap Analysis Program: An Analytical Framework for Personal Income Tax Gap Estimation.* Technical Notes and Manuals 21/02.

Iyengar, Sheena S. and Mark R. Lepper (2000). "When choice is demotivating: Can one desire too much of a good thing?" *Journal of Personality and Social Psychology*, 79 (6), 995–1006. doi: 10.1037/0022-3514.79.6.995.

Jensen, Anders (2022). "Employment Structure and the Rise of the Modern Tax System." *American Economic Review*, 112 (1), 213–234. doi: 10.1257/aer.20191528.

Jessen, Jonas (2022). "Culture, children and couple gender inequality." *European Economic Review*, 150, 104310. doi: 10.1016/j.euroecorev.2022.104310.

Jones, Damon (2012). "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds." *American Economic Journal: Economic Policy*, 4 (1), 158–185. doi: 10.1257/pol.4.1.158.

Juhn, Chinhui and Kristin McCue (2016). "Evolution of the Marriage Earnings Gap for Women." *The American Economic Review: Papers & Proceedings*, 106 (5), 252–256. doi: 10.1257/aer.p20161120.

— (2017). "Specialization Then and Now: Marriage, Children, and the Gender Earnings Gap across Cohorts." *Journal of Economic Perspectives*, 31 (1), 183–204. doi: 10.1257/jep.31.1.183.

Kalíšková, Klára (2014). "Labor supply consequences of family taxation: Evidence from the Czech Republic." *Labour Economics*, 30, 234–244. doi: 10.1016/j.labeco.2014.04.010.

— (2020). "Tax and transfer policies and the female labor supply in the EU." *Empirical Economics*, 58 (2), 749–775. doi: 10.1007/s00181-018-1544-2.

Killewald, Alexandra and Margaret Gough (2013). "Does Specialization Explain Marriage Penalties and Premiums?" *American sociological review*, 78 (3), 477–502. doi: 10.1177/0003122413484151.

Killewald, Alexandra and Ian Lundberg (2017). "New Evidence Against a Causal Marriage Wage Premium." *Demography*, 54 (3), 1007–1028. doi: 10.1007/s13524-017-0566-2.

Kleven, Henrik (2023). *The Geography of Child Penalties and Gender Norms: Evidence from the United States.* NBER Working Paper 30176 (Revised Version, Issue Date: June 2022). doi: 10.3386/w30176.

Kleven, Henrik, Claus Thustrup Kreiner, and Emmanuel Saez (2016). "Why Can Modern Governments Tax So Much? An Agency Model of Firms as Fiscal Intermediaries." *Economica*, 83 (330), 219–246. doi: 10.1111/ecca.12182.

Kleven, Henrik, Camille Landais, and Gabriel Leite-Mariante (2023). *The Child Penalty Atlas.* NBER Working Paper 31649. doi: 10.3386/w31649.

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller (2019). "Child Penalties across Countries: Evidence and Ex-

planations." *AEA Papers and Proceedings*, 109, 122–126. DOI: 10.1257/pandp.20191078.

— (2023). "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation." *American Economic Journal: Economic Policy*, forthcoming.

Kleven, Henrik, Camille Landais, and Jakob Egholt Søgaard (2019). "Children and Gender Inequality: Evidence from Denmark." *American Economic Journal: Applied Economics*, 11 (4), 181–209. DOI: 10.1257/app.20180010.

Kleven, Henrik Jacobsen, Martin B. Knudsen, Claus Thustrup Kreiner, Søren Peterson, and Emmanuel Saez (2011). "Unwilling or Unable to Cheat? Evidence From a Tax Audit Experiment in Denmark." *Econometrica*, 79 (3), 651–692. DOI: 10.3982/ECTA9113.

Koch, Leonie (2024). *Income Taxation of Couples and Gender (In)equality in Earnings*.

Korenman, Sanders and David Neumark (1991). "Does Marriage Really Make Men More Productive?" *The Journal of Human Resources*, 26 (2). Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System], 282–307. DOI: 10.2307/145924.

Kuziemko, Ilyana, Michael I. Norton, Emmanuel Saez, and Stefanie Stantcheva (2015). "How elastic are preferences for redistribution? Evidence from randomized survey experiments." *American Economic Review*, 105 (4), 1478–1508.

Kuziemko, Ilyana, Jessica Pan, Jenny Shen, and Ebonya Washington (2018). *The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?* NBER Working Paper 24740. DOI: 10.3386/w24740.

LaLumia, Sara (2008). "The effects of joint taxation of married couples on labor supply and non-wage income." *Journal of Public Economics*, 92 (7), 1698–1719. DOI: 10.1016/j.jpubeco.2008.01.009.

Lee, David S. (2008). "Randomized experiments from non-random selection in U.S. House elections." *Journal of Econometrics*. The regression discontinuity design: Theory and applications, 142 (2), 675–697. DOI: 10.1016/j.jeconom.2007.05.004.

Liebman, Jeffrey B. and Richard J. Zeckhauser (2004). *Schmeduling*. URL: https://scholar.harvard.edu/sites/scholar.harvard.edu/files/jeffreyliebman/files/Schmeduling_WorkingPaper.pdf (last accessed on 02/24/2024).

Lusardi, Annamaria and Olivia S. Mitchell (2011). "Financial literacy around the world: an overview." *Journal of Pension Economics and Finance*, 10 (4), 497–508. DOI: 10.1017/S1474747211000448.

— (2014). "The Economic Importance of Financial Literacy: Theory and Evidence." *Journal of Economic Literature*, 52 (1), 5–44. DOI: 10.1257/jel.52.1.5.

Lüthen, Holger, Carsten Schröder, Markus M. Grabka, Jan Goebel, Tatjana Mika, Daniel Brüggmann, Sebastian Ellert, and Hannah Penz (2021). "SOEP-RV: Linking German Socio-Economic Panel Data to Pension

Records." *Jahrbücher für Nationalökonomie und Statistik*. Publisher: De Gruyter Oldenbourg. DOI: 10.1515/jbnst-2021-0020.

Madrian, Brigitte C. and Dennis F. Shea (2001). "THE POWER OF SUGGESTION: INERTIA IN 401(k) PARTICIPATION AND SAVINGS BEHAVIOR." *Quarterly Journal of Economics*, 116 (4), 1149–1187. DOI: 10.1162/003355301753265543.

Marcuss, Rosemary, George Contos, John Guyton, Patrick Langetieg, Allen Lerman, Susan Nelson, Brenda Schafer, and Melissa Vigil (2013). "Income Taxes and Compliance Costs: How Are They Related?" *National Tax Journal*, 66 (4), 833–854. DOI: 10.17310/ntj.2013.4.03.

Moody, J. Scott, Wendy P. Warcholik, and Scott Hodge (2005). "The Rising Cost of Complying with the Federal Income Tax." *Special Report - Tax Foundation*, (138). URL: https://files.taxfoundation.org/legacy/docs/sr138.pdf.

Neisser, Carina (2021). "The Elasticity of Taxable Income: A Meta-Regression Analysis." *The Economic Journal*. DOI: 10.1093/ej/ueab038.

OECD (2019). *Pensions at a Glance 2019: OECD and G20 Indicators*. OECD Pensions at a Glance. OECD. DOI: 10.1787/b6d3dcfc-en.

— (2021a). *OECD.Stat - Table I.1. Central government personal income tax rates and thresholds*. URL: https://stats.oecd.org/index.aspx?DataSetCode=Table_I1 (last accessed on 03/02/2022).

— (2021b). *Tax Administration 2021: Comparative Information on OECD and other Advanced and Emerging Economies*. Tax Administration. OECD. DOI: 10.1787/cef472b9-en.

— (2022a). *Purchasing power parities (PPP) (indicator)*. DOI: 10.1787/1290ee5a-en.

— (2022b). *Tax Policy and Gender Equality: A Stocktake of Country Approaches*.

Piketty, Thomas, Emmanuel Saez, and Gabriel Zucman (2018). "Distributional National Accounts: Methods and Estimates for the United States." *The Quarterly Journal of Economics*, 133 (2), 553–609. DOI: 10.1093/qje/qjx043.

Pilossoph, Laura and Shu Lin Wee (2021). "Household Search and the Marital Wage Premium." *American Economic Journal: Macroeconomics*, 13 (4), 55–109. DOI: 10.1257/mac.20180092.

Pitt, Mark M. and Joel Slemrod (1989). "The Compliance Cost of Itemizing Deductions: Evidence from Individual Tax Returns." *The American Economic Review*, 79 (5), 1224–1232.

Ramnath, Shanthi P. and Patricia K. Tong (2017). "The Persistent Reduction in Poverty from Filing a Tax Return." *American Economic Journal: Economic Policy*, 9 (4), 367–394. DOI: 10.1257/pol.20150486.

Rees-Jones, Alex and Dmitry Taubinsky (2020). "Measuring "Schmeduling"." *The Review of Economic Studies*, 87 (5), 2399–2438. DOI: 10.1093/restud/rdz045.

Research Data Centre of the Federal Statistical Office and the statistical offices of the Länder (RDC) (2018). *Lohn- und Einkommenssteuerstatistik 2010, Scientific Use File*. DOI: 10.21242/73111.2010.00.00.3.1.0.

— (2020). *Lohn- und Einkommenssteuerstatistik 2014, Scientific Use File*. DOI: 10.21242/73111.2014.00.00.3.1.0.

— (2022). *Taxpayer-Panel 2001-2018, On-Site, Version 1*. DOI: 10.21242/73111.2018.00.01.1.1.1.

Research Data Centre of the German Pension Insurance (FDZ-RV) (2022). *Codeplan Scientific Use File SOEP-RV VSKT 2020*.

Roller, Marcus and Kurt Schmidheiny (2016). *Effective Tax Rates and Effective Progressivity in a Fiscally Decentralized Country*. CEPR Discussion Paper 11152. URL: https://cepr.org/active/publications/discussion_papers/dp.php?dpno=11152 (last accessed on 05/04/2020).

Saez, Emmanuel (2001). "Using Elasticities to Derive Optimal Income Tax Rates." *The Review of Economic Studies*, 68 (1), 205–229. DOI: 10.1111/1467-937X.00166.

Saez, Emmanuel, Joel Slemrod, and Seth 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. DOI: 10.1257/jel.50.1.3.

Saez, Emmanuel and Gabriel Zucman (2019). *The triumph of injustice: how the rich dodge taxes and how to make them pay*. First Edition. New York, NY: W. W. Norton & Company.

— (2020). "Progressive Wealth Taxation." *Brookings Papers on Economic Activity*, Fall 2019, 437–533. URL: http://www.jstor.org/stable/10.7864/j.ctvz0h9kb.8.

Sarin, Natasha (2023). *The Case for a Robust Attack on the Tax Gap*. URL: https://home.treasury.gov/news/featured-stories/the-case-for-a-robust-attack-on-the-tax-gap (last accessed on 10/01/2023).

Seibold, Arthur (2021). "Reference Points for Retirement Behavior: Evidence from German Pension Discontinuities." *American Economic Review*, forthcoming. DOI: 10.1257/aer.20191136.

Selin, Håkan (2014). "The rise in female employment and the role of tax incentives. An empirical analysis of the Swedish individual tax reform of 1971." *International Tax and Public Finance*, 21 (5), 894–922. DOI: 10.1007/s10797-013-9283-y.

Siminski, Peter and Rhiannon Yetsenga (2022). "Specialization, Comparative Advantage, and the Sexual Division of Labor." *Journal of Labor Economics*, 40 (4), 851–887. DOI: 10.1086/718430.

Slemrod, Joel (2005). "The Etiology of Tax Complexity: Evidence from U.S. State Income Tax Systems." *Public Finance Review*, 33 (3). Publisher: SAGE Publications Inc, 279–299. DOI: 10.1177/1091142105275003.

Stevenson, Betsey (2007). "The Impact of Divorce Laws on Marriage-Specific Capital." *Journal of Labor Economics*, 25 (1), 75–94. DOI: 10.1086/508732.

Stevenson, Betsey (2008). "Divorce Law and Women's Labor Supply." *Journal of Empirical Legal Studies*, 5 (4), 853–873. doi: 10.1111/j.1740-1461.2008.00143.x.

United Nations (2024). *Goal 5: Achieve gender equality and empower all women and girls*. url: https://www.un.org/sustainabledevelopment/gender-equality/ (last accessed on 03/09/2024).

United Nations, Department of Economic and Social Affairs, Population Division (2019). *World Marriage Data 2019*. url: https://www.un.org/development/desa/pd/data/world-marriage-data (last accessed on 10/30/2022).

Voena, Alessandra (2015). "Yours, Mine, and Ours: Do Divorce Laws Affect the Intertemporal Behavior of Married Couples?" *American Economic Review*, 105 (8), 2295–2332. doi: 10.1257/aer.20120234.

Weber, Rick (2015). "The Effect of Tax Code Complexity on Entrepreneurship." *The Journal of Private Enterprise*, 30 (2), 83–102. url: http://journal.apee.org/index.php/2015_Journal_of_Private_Enterprise_vol_30_no_2_parte5.pdf.