Essays in Labor Economics

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschaftswissenschaften

durch

die Rechts- und Staatswissenschaftliche Fakultät der
Rheinischen Friedrich-Wilhelms-Universität Bonn

vorgelegt von

Jakob Wegmann

aus Bad Soden am Taunus

2025

Dekan:	Prof. Dr. Martin Böse
Erstreferent:	Prof. Dr. Thomas Dohmen
Zweitreferentin:	Prof. Dr. Amelie Schiprowski
Tag der mündlichen Prüfung:	25. September 2025

Acknowledgements

I truly enjoyed my PhD and I want to thank everyone who was part of the journey.

Firstly, I want to thank my supervisors Amelie Schiprowski and Thomas Dohmen. They are excellent role models for supervision. Both inherently trust PhD students, are curious in so many domains and are simply very smart. The amount of freedom I enjoyed during my PhD was unique. At the same time, Amelie and Thomas were always available, and I learned a great deal from them. Not only did they help with their deep understanding of economics and econometrics, I also acquired numerous other skills that are helpful in academia and the normal world outside academia. Without a single exception, they have always been supportive, kind and generous. The best compliment I can give them is that I feel unprepared for the world outside Bonn, where people might not be inherently good.

I also want to thank my co-authors that taught me so many different things. Lenard, we completed the Bachelor, the Master and now even the PhD together. My last years would not have been half as enjoyable without you. There is so much I learned from you. Tim, I believe that no one except us is organizing the discussion of politics by topic (at least people react weirdly when I try to explain how our Zulip works). Our long discussions about everything have greatly shaped my understanding of the world and all my research interests. Katrin, my sister in arms in our joint fight against Josua. Without you I would have lost the fight a long time ago. You are a role model terms of rigour, organisational skill, openness and kindness. Pia and Federica, ever since we started working on our parental leave project last year, I have always looked forward to our Zooms and work weeks as much as for no other research project! Clara, working together on our overambitious project is so interesting, inspiring and thought-provoking. The really intense research stays in London were an important part of my PhD. Hannah and Fabian, working and spending time with you is an absolute joy. Thank you for all your time and energy, our research is still my favorite small talk topic. Lastly, I also want to thank my senior co-authors Jonas, Ingo and Marco who were all generous with their time and advice and who have all the skills I aspire to be able to match one day.

I also want to thank all the other PhD students who made my last years so pleasant, among them Christina, Bettina, Marvin, Oleksii, Antonia and Alex. And in particular, I want to thank Anna, the tough experience of the course phase of the

PhD will always bind us together. Also, I am very grateful for the researchers at IZA for being so welcoming and providing such an inspiring place for labour market research. I am particularly thankful to Maria, Christian, Gökay, Ingo and Simon.

Without all the great support I received, the whole PhD would have been impossible. I want to thank Simone Jost and Ilona Krupp who corrected all my mistakes in my interaction with the university administration, Hannes Fauser and Laura Ehrmantraut at Destatis who were going the extra mile so many times to support my research, and Holger Gerhardt for all TeXnical assistance.

Also, I want to thank my family. Judith and Bernd in particular who were always there for me. Griseldis and Bernhard, your place felt like a second home. And also thank you Marlene, it often feels like you are part of my family as well. Lastly, and most importantly, I want to thank Guni. Your love, friendship, creativity, intellect and support makes my life better every day.

Jakob Wegmann
Bonn, August 2025

Contents

List of Figures

List of Tables

Introduction

One of the central tasks society assigns to economists is to improve tax and social transfer systems. This requires not only thorough theoretical thinking about the optimal design of policies to reach the policy objective, but also deep institutional knowledge and the empirical analysis of existing policies. A key finding is that reactions to policies are strongly context-dependent, meaning that theoretical considerations cannot reliably describe human behavior and smart design solutions are often less effective than expected. In the field of income taxation, for example, taxpayers often do not understand the concept of marginal taxes (e.g. Rees-Jones, 2018), tax payers react to imaginary tax incentives (e.g. Lardeux, 2023) and small information treatments have strong effects on political preferences (e.g. Stantcheva, 2021).

Therefore, designing efficient incentive schemes requires policymakers to understand the context-specific preferences of households and take into account the fact that households may not understand complex policy rules or incorrectly assess their financial situation. In this thesis, I address these issues by empirically investigating behavioral aspects relevant to the design of policy rules and the understanding of people of policies and their own finances.

Chapter 1 "Withheld from Working More? Withholding Taxes and the Labor Supply of Married Women" (joint work with Lenard Simon and Tim Bayer) shows that withholding taxes, i.e., the tax-prepayments to the income tax, can have substantial impact on labor supply. This shows that supposedly irrelevant technical decisions of politicians and bureaucrats impact the decision making of households. We investigate the importance in the context of married couples in Germany where the withholding tax liability can be redistributed between spouses. We exploit a reform that reduced the withholding tax for some married women more than for others, while inducing no differences in income taxes. Using administrative data for the full population of German taxpayers, we estimate an elasticity of labor income with respect to the withholding tax eight years after the reform of 0.10. Additional evidence from a self-conducted survey suggests imperfect understanding of the tax system and limited pooling of resources within the household as the main mechanisms. As the majority of couples shift parts of the withholding tax liability from the husband to the wife, our results suggest that the increased withholding tax liability of married women contributes to their low labor supply. This highlights the need

for governments to be aware of the distortion of labor supply incentives when the design of withholding taxes does not match actual income tax incentives.

Chapter 2 "Who cares? Parental Leave Benefits and Household Division of Childcare" (joint work with Federica Meluzzi and Pia Molitor) is motivated by the fact that low take-up of parental leave by fathers is a key driver of the unequal career costs of parenthood for men and women. We study whether the design of parental leave policies can effectively increase fathers' participation. The paper is the first to provide evidence on how the monetary generosity of parental leave benefits shapes the intra-household allocation of childcare. We exploit exogenous variation in benefits generated by a sharp kink in the German parental leave schedule relative to pre-birth income and combine it with novel administrative data covering millions of households between 2014 and 2018. Using a double Regression Kink Design — an extension of the RKD methodology by Card et al. (2015) — we estimate how each parent responds to changes in both their own and their partner's replacement rates. We find that a reduction in the mother's benefit leads to a partial substitution in parental inputs, decreasing her leave duration while increasing the father's take-up, so that the household's total leave duration remains unchanged. Cross-spousal substitution is stronger when fathers face lower opportunity costs and in regions with more egalitarian gender norms. These findings underscore how financial incentives interact with cultural and economic constraints in shaping household behavior, with implications for the design of family policies aimed at promoting gender equality.

Chapter 3 "On the Extent, Sources, and Consequences of Reporting Bias in Survey Wages" (joint work with Marco Caliendo, Katrin Huber and Ingo Isphording) studies the reliability of self-reported wages in survey. We compare self-reported survey wages from the German Socio-Economic Panel (SOEP) with administrative wages from the social security records (IEB) for the same individual via a novel and unique data linkage (SOEP - ADIAB). We describe a modest but economically significant extent of reporting bias, with SOEP respondents understating administrative wages by about 7.3 percent. Individual, household-, and especially firm characteristics vary systematically with the reporting bias. Replicating common empirical exercises using wages as dependent and independent variable, we find that the choice of data matters, in particular, when estimating the relationship between wages and subjective outcome variables which are directly related to income (satisfaction measures) or between wages and variables that are more strongly correlated with the reporting bias (e.g. gender). In these cases, relying on survey wages can significantly overstate the correlation. Our findings underscore the critical importance of exercising caution when analyzing survey-based measures of individual income in empirical research. Depending on the research question at hand, it might be worthwhile to complement survey data by administrative measures.

References

Card, David, David Lee, Zhuan Pei, and Andrea) Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83(6): 2453–83. [2]

Lardeux, Raphaël. 2023. "Behavioral Cross-Influence of a Shadow Tax Bracket: Evidence from Bunching Where Income Tax Liabilities Start." *Journal of Public Economics* 226: 104950. https://doi.org/10.1016/j.jpubeco.2023.104950. [1]

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–78. https://doi.org/10.1093/restud/rdx038. [1]

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *Quarterly Journal of Economics* 136 (4): 2309–69. https://doi.org/10.1093/qje/qjab033. [1]

Chapter 1

Withheld from Working More? Withholding Taxes and the Labor Supply of Married Women*

1.1 Introduction

Many people do not understand how income taxes work. They often do not know the difference between marginal and average tax rates (Liebman and Zeckhauser, 2004; Gideon, 2017; Rees-Jones and Taubinsky, 2020) and which tax rates apply to them (Fujii and Hawley, 1988; Chetty, Friedman, and Saez, 2013; Blaufus et al., 2015; Aghion et al., 2017; Rees-Jones, 2018; Farhi and Gabaix, 2020; Lardeux, 2023; Leite, 2024). Due to this limited understanding, households' reactions are therefore influenced by how they interpret their incentives (Stantcheva, 2021). These interpretations can be influenced by supposedly irrelevant technical decisions of politicians and bureaucrats.

We study the decision about the role of income tax collection, a so far overlooked cornerstone in understanding labor supply responses to income taxes. Almost all countries require employers to withhold monthly tax prepayments which are then credited against the final income tax liabilities of their employees. The collection of these withholding taxes provides governments with a constant stream of revenue and increases tax compliance (Schepanski and Shearer, 1995; Slemrod,

* We thank Hannes Fauser, Stefan Bach, Mareen Bastiaans, Antoine Bozio, Maurice Brandt, Clément Carbonnier, Thomas Dohmen, Jan Feld, Holger Gerhardt, Hannah Illing, Alex Rees-Jones, Amelie Schiprowski, Joel Slemrod, Elena Stancanelli, as well as the participants of Berlin Workshop on Empirical Public Economics, RGS Doctoral Conference, Trier Workshop on Labour Economics, SOLE, ESPE, EDP Jamboree, IZA Summer School, IIPF, EEA, MannheimTaxation Conference, EALE, the Verein für Socialpolitik, SSES, Mannheim Alumni Symposium, and the Seminar at Frauenhofer Institute. Funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2126/1-390838866. Support by the Deutsche Forschungsgemeinschaft (DFG) through CRC TR 224 (Project A05) is gratefully acknowledged.

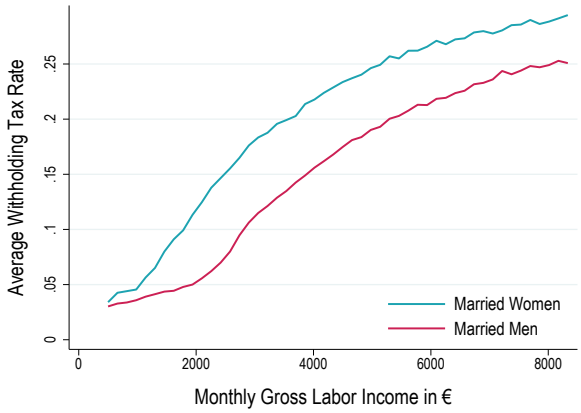


Figure 1.1. Average Withholding Tax Rate by Gender

Notes: The figure displays the average realized withholding tax rate by gender for married couples in Germany. The values correspond to population means as they are based on the universe of observations from the German Taxpayer Panel (TPP) for the year 2010. The figure illustrates that married couples shift a substantial share of the withholding tax liability from men to women through the choice of withholding tax schedules.

2019; Bagchi and Dušek, 2021; Bagchi, 2025). There is, however, no clear best practice of how withholding taxes should be designed. Typically, withholding tax rates do not reflect true effective income tax rates. Many governments often collect more withholding taxes than employees owe in income taxes (Engström et al., 2015; Rees-Jones, 2018; Gelman et al., 2022; Hauck and Wallossek, 2024).[1] In those cases, governments issue a refund to taxpayers after the end of the tax year. Conversely, when the government collects less in withholding taxes than employees owe in income taxes, taxpayers are required to make an additional tax payment to the government. This link between withholding taxes and income taxes adds a layer of complexity that might obscure that withholding tax rates have no effect on the total amount of income taxes paid in a year. Besides liquidity constraints and time discounting, how income taxes are collected should therefore have no effect on how much employees decide to work.

It is difficult to identify the effects of withholding taxes, as they are typically closely tied to the underlying income tax system. It is, therefore, usually not possible to use reforms of the income tax system to draw conclusions about the role of withholding taxes. However, the German income tax system offers an institutional setting that allows for such an analysis. We illustrate the core feature of the institutional setting in Figure 1.1, which displays average withholding tax rates by gender and labor income in Germany. Conditional on labor income, married women are on

1. This overwithholding often reaches substantial magnitudes. For example, nearly a third of the amount of all personal income tax payments is returned as tax refunds in the US (Gelman et al., 2022).

average subject to higher withholding tax rates than married men. This is a result of the fact that married couples can shift parts of the withholding tax liability from one partner to the other by choosing certain withholding tax schedules. This implies that spouses with identical income can be subject to different withholding tax rates even when the overall household income is the same. Importantly, the decision on withholding tax schedules does not affect the final income tax rate. However, a married couple can minimize its joint withholding tax liability by shifting some part of the withholding tax liability from the spouse with higher labor income, i.e., the primary wage earner, to the spouse with lower labor income, i.e., the secondary wage earner. This explains the pattern in Figure 1.1: Married women are typically the secondary wage earner and hence face, on average, a higher withholding tax rate conditional on labor income.[2]

In this paper, we use the German withholding tax system to study how withholding taxes impact labor income. At the core of the paper, we leverage a withholding tax reform that reduced withholding taxes more for some women than for others, allowing us to estimate the effect of withholding taxes on labor income. To understand the underlying behavioral responses, we complement the analysis with a toy model and self-collected survey data. We then compare withholding tax implementations in different countries to discuss the trade-offs policymakers face when implementing withholding taxes.

We conduct our analysis using full-population administrative tax records from the German Taxpayer Panel for the years 2006 to 2018. To obtain exogenous variation in withholding taxes, we exploit a tax reform in 2010 that cut withholding taxes for married women differently across withholding tax schedules. Using a Difference-in-Differences design with continuous treatment intensity, we are able to investigate how married women react to a cut in withholding taxes while keeping income tax liabilities constant.

We estimate an elasticity of labor income with respect to the marginal net-of-withholding tax rate eight years after the reform of approximately 0.10. This effect is more pronounced for women who have greater flexibility in adjusting their labor supply, such as those working part-time or those without children. The estimated elasticity is lower than typical elasticities estimated in response to changes in income tax (Neisser, 2021). Nevertheless, the effect is substantial as we show using a back-of-the-envelope calculation: Aligning the marginal withholding tax rate with the marginal income tax rate for all women whose marginal withholding tax rate is higher than the household's income tax rate would lead to a 1 % increase in labor income for this group.

With the help of a toy model, we discuss three factors that can explain why withholding taxes affect labor supply decisions: First, individuals might not fully

2. Conditional on income, there is no other reason for withholding tax rates of husbands and wives to differ as all tax credits and deductions are applied on the household level.

understand or internalize the relationship between income and withholding taxes. Therefore, they might think that the net income on their monthly payslip equals their actual net income after income taxes and consequently use withholding taxes as a proxy for income taxes in their labor supply decisions. Second, even when households understand the difference between income and withholding taxes, the distribution of withholding taxes between partners might impact labor supply decisions when households do not act as a unit. Lastly, households might also react to withholding taxes if they have strong time preferences for income during the year, for example due to budget constraints. However, the last channel is unlikely to drive our results as suggested by the small time discounting factors found in the literature (e.g., Falk et al., 2018).

As evidence on the knowledge about the interlinkage between withholding and income taxes and on the organization of household finances is much scarcer, we conduct a pre-registered online survey among 506 married and employed German individuals. First, we find that more than 80 % of the surveyed individuals wrongly think that the choice of withholding tax schedules affects the final income tax liability. This suggests that individuals with the same income tax liability, but with differing withholding tax rates, might perceive their income tax liability differently and consequently make different labor supply decisions. Second, we investigate the impact of the system of withholding tax schedules on the organization of household finances in Germany. As seen in Figure 1.1, couples often choose withholding tax schedules that shift parts of the withholding tax liability from men to women. In a unitary household model, this would have no impact on real outcomes as all resources are shared. However, if a woman only has limited access to the income of the husband, the observed pattern of assignment of withholding tax schedules lowers her own disposable net income. We find suggestive evidence that in fact there is no equal access of both partners to all household resources.

These findings speak to important considerations regarding the role of withholding taxes in tax systems with joint taxation. Joint income taxation treats married couples as one unit and therefore sets one joint income tax liability for the couple. In contrast, withholding taxes are always levied on the individual level so that there is no clear optimal withholding tax regime in the presence of joint taxation. We therefore compare the implementations of withholding taxes for married couples in France, US and Germany and study the potential effects of two recently proposed reforms in France and Germany.[3]

Studying these reforms reveals two important lessons: First, we find that policymakers do not attempt to align the perceived incentives with the real incentives as the reforms do not reduce the gap between marginal withholding tax rates and

3. The reform in France will be implemented in September 2025. The German reforms was agreed upon by the governing parties in the coalition agreement in 2021, but it did not pass parliament before the coalition broke in fall 2024.

marginal income tax rates. Instead, they set lower marginal withholding tax rates for secondary earners. In light of our empirical results, this suggests that politicians deliberately use withholding taxes to increase the labor supply of married women.

Second, average withholding tax rates are also substantially altered by the reforms. In countries with tax withholding and joint taxation, policymakers face the challenge of determining how the marriage bonus should be distributed among the spouses throughout the year as they have to set an individual average withholding tax rate for every taxpayer. For example, the gap in average tax rates presented in Figure 1.1 reflects the consequence of the implemented withholding tax system for married individuals in Germany. As we show that couples in Germany do not fully pool their income and as an effect do not fully pool the marriage bonus, this gender gap in withholding taxes might lead to lower work incentives for women. Women might overestimate their individual income tax liability, which can decrease their incentives to work as we show in our toy model. Interestingly, implemented approaches to distributing the marriage bonus vary substantially across countries. The two recently discussed reforms in France and Germany both increase the share of the marriage bonus that the secondary earners receives. This might lead to higher work incentives for married women.

Our paper is closely linked to the behavioral public finance literature studying the interaction between inattention, tax complexity and behavioral responses to taxation. It is well documented empirically, mostly with surveys, that a large share of taxpayers does not understand how income taxation works (Chetty, Friedman, and Saez, 2013; Abeler and Jäger, 2015; Blaufus et al., 2015; Aghion et al., 2017; Gideon, 2017; Rees-Jones, 2018; Rees-Jones and Taubinsky, 2020; Stantcheva, 2021). Relatedly, the view that the complexity of income tax systems matters for labor supply responses and thereby for the optimal design of tax systems is established in the theoretic literature (Liebman and Zeckhauser, 2004; McCaffery and Slemrod, 2006; Bernheim and Taubinsky, 2018; Gabaix, 2019; Farhi and Gabaix, 2020; Moore and Slemrod, 2021). However, the empirical literature on labor supply effects of limited understanding of tax systems is still emerging (Chetty, Friedman, and Saez, 2013; Feldman, Katuščák, and Kawano, 2016; Kostøl and Myhre, 2021; Lardeux, 2023; Leite, 2024). Specifically withholding taxes have been studied by Becker, Fooken, and Steinhoff (2019) and Koch (2024). We contribute to the behavioral public finance literature in two ways. First, we document widespread misunderstanding of the withholding and income tax system in Germany. Second, we highlight the role of income tax collection for the behavioral reaction to income taxes by providing real-world evidence on the impact of withholding taxes on labor supply.

Also, withholding taxes have been studied in other contexts. First, it has been shown that there exists a positive relationship between withholding tax rates and savings and consequently a negative relationship between withholding tax rates and consumption (Shapiro and Slemrod, 1995; Feldman, 2010; Jones, 2012; Messacar,

2018; Gelman et al., 2022; Caldwell, Nelson, and Waldinger, 2023). Second, the consequence and motivation for overwithholding have been investigated. Besides a potential government preference for bolstering the saving rate (Thaler, 1994) and an increase in tax compliance (Schepanski and Shearer, 1995), the motivation for systematic overwithholding can also stem from the fact that taxpayers like getting tax refunds and thus change tax filing behavior discontinuously at the point of exact withholding, e.g., by claiming more tax deductions (Engström et al., 2015; Rees-Jones, 2018). However, our results imply that governments face a trade-off if they prefer overwithholding when designing withholding taxes. When withholding taxes distort labor supply decisions, overwithholding decreases labor supply.

Finally, previous literature has shown that labor supply of women can be detrimentally affected by the design of tax systems. This holds true in particular for income tax systems with joint taxation of married couples (LaLumia, 2008; Selin, 2014; Bick and Fuchs-Schündeln, 2017, 2018; Herold and Wallossek, 2024). As our results imply that income tax collection also plays a role for labor supply decisions, gender gaps in withholding taxes can therefore provide an additional explanation for the observed gender gaps in labor market outcomes. By presenting the trade-offs between different withholding tax regimes in the presence of joint taxation of married couples, we furthermore contribute to the question of how an optimal withholding tax schedule should be designed when also taking into account potential negative effects on labor supply.

The rest of the paper is structured as follows: Section 1.2 presents in detail the German withholding tax system. Section 1.3 discusses the data as well as the empirical strategy and Section 1.4 presents the results. In Section 1.5.1, we introduce a toy model presenting potential mechanisms which we subsequently investigate with the help of a survey in Section 1.5.2. Section 1.6 then discusses the policy implications for the optimal design of withholding taxes and Section 1.7 concludes.

1.2 Institutional Setting

In this section, we first provide context for our study by explaining the German joint taxation system and subsequently present the German withholding tax system for married couples. Thereafter, we describe the reform of withholding taxes that we use to identify causal effects.

1.2.1 Income Taxation of Married Couples

In Germany, married couples have two distinct options for filing their income taxes. They can opt to file them separately, treating their finances as if they were still single individuals, or they can choose to file jointly. Choosing the latter option enables

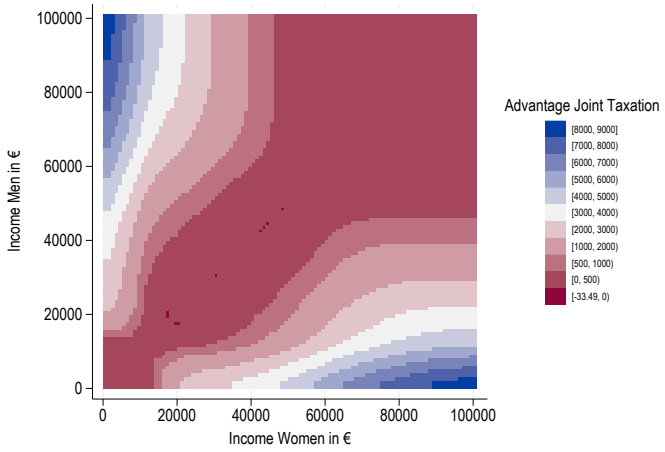


Figure 1.2. Marriage Bonus of Joint Taxation

Notes: The figure illustrates the system of joint income taxation in Germany. It plots the joint taxation benefits depending on the intra-household income distribution in the year 2010. We assume that both spouses contribute to the public health care and pension systems and do not claim any further deductions. The negative values are a consequence of the *Solidaritätszuschlag*, a surtax that tax filers below a specific income tax liability are not subject to.

couples to take advantage of potential tax benefits associated with joint taxation.[4] Under joint income taxation, the individual income tax schedule is applied to half of the joint taxable income for each couple and the resulting tax liability is then doubled. Due to the progressivity of the German income tax system, this creates joint taxation benefits whenever the spouses in a couple would have faced differing marginal income tax rates under separate taxation. For a fixed household income, a couple receives more joint taxation benefits the more unequal the intra-household distribution of income.

We illustrate this feature in Figure 1.2, where we plot the marriage bonus induced by joint taxation. If both spouses contribute equally to the household income, there are no benefits from joint taxation. However, the more unequally the income is distributed within the household, the more the couple benefits from joint taxation. For example, if one spouse earns €70,000 and the other spouse earns €10,000, the couple receives a marriage bonus of €3,900 a year which equals to almost 5 % of the household's gross income. The underlying reason is that the couple's joint income

4. In fact, for the vast majority of couples, choosing joint taxation is at least weakly better than choosing separate taxation. Apart from special cases of couples with a negative marriage bonus arising from the "Solidaritätszuschlag", only couples in which one partner obtains a substantial amount of income replacement benefits (e.g., parental leave benefits, unemployment benefits or short-time work allowances) can be better off by choosing separate taxation. The reason for that is that those payments, while not being taxable, can increase the marginal income tax rate of the couple ("Progressionsvorbehalt").

tax liability is €15,800 under joint taxation but would be €19,700 under separate taxation.[5]

1.2.2 Withholding Taxes of Married Couples

The German government wants to enable couples to profit from the joint taxation benefit already during the year. Therefore, couples have the choice to reduce their withholding tax liability. Married couples can influence both the sum of their withholding tax liabilities and the allocation of their joint withholding tax liability to each spouse. They can effectively choose between three different withholding tax schedules which are illustrated in Figures 1.3 and 1.4 and explained in the following.[6] These withholding tax schedules assign each partner a certain so-called withholding tax class (*Lohnsteuerklasse*), which determines the personal withholding tax payments.

Symmetric schedule. After marriage, each couple in which both spouses receive labor income gets assigned the same withholding tax schedule as a default, which we will call the symmetric schedule. This withholding tax schedule is symmetric since it assigns each spouse the same default withholding tax class (officially called "IV"). In this withholding tax class, the monthly withholding tax payments are calculated as if the individual was single, only taking into account the own individual income. Hence, the withholding tax is the same as the income tax for a couple without joint taxation benefits. If a couple realizes joint taxation benefits, the remitted withholding taxes of both spouses will exceed their joint final income tax liability and the couple will receive a tax refund after filing an income tax return. We illustrate this in Figure 1.3 for a couple in which the husband earns €50,000 and the wife earns €30,000. Being in the symmetric withholding tax schedule causes the couple to receive the joint taxation benefit of €387 as a lump-sum tax refund after filing their income taxes.

5. Typically, the focus of economists lies on the fact that the secondary earner within the couple faces, in the presence of joint taxation benefits, a higher marginal income tax rate under joint income taxation than under separate income taxation. For example, Bick and Fuchs-Schündeln (2017) have identified that this phenomenon as one key policy that explains the low labor market participation of women. We illustrate how the marginal tax rate depends on partner income in the German joint taxation system in Figure 1.A.1.

6. In our analysis, we leave out the fourth, least commonly chosen withholding tax schedule. This "schedule with a factor" (*IV mit Faktor*) was introduced in 2010. The tax office takes into account the past income of both spouses and calculates the exact advantage of joint taxation for both spouses individually. Thereby, the tax office can set the withholding tax for both individuals at a level that allows the household to profit from the advantage of joint taxation during the year. More details on the effects of this schedule on marginal and average withholding tax rates can be found in Section 1.6 where we discuss different implementations of withholding tax schedules that account for joint taxation benefits. There are no official statistics on the use of the schedule with a factor and we do not observe the choice in the data. The German government documents that in 2018 only around 40,000 couples (less than 0.6 % of all married dual-earner couples) used this schedule (*Kleine Anfrage Bundestag,* 2019).

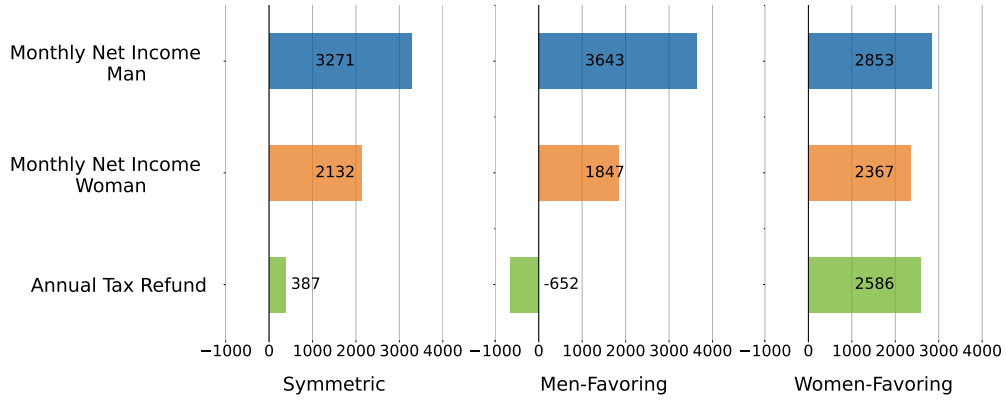


Figure 1.3. Example Illustrating the Different Withholding Tax Schedules

Notes: The figure illustrates how the different withholding tax schedules affect the monthly net incomes of both spouses and the annual tax refund in the year 2010. Net incomes are calculated for a household in which the husband earns €50,000 and the wife earns €30,000 gross per year. The assessed annual income tax liability of the household is €14,774 under the assumption that the couple does not claim any deductions. The figure shows how the different withholding tax schedules shift the withholding tax liability from one partner to the other and how they can affect the annual refund from the final income tax.

To avoid this overpayment of withholding taxes during the year, a couple can decide to switch from the symmetric schedule to a withholding tax schedule that aims at reducing the monthly withholding tax payments to account for the joint taxation benefits already during the tax year.[7]

Men- or women-favoring schedule. The most popular alternative withholding tax schedules are the men-/women-favoring withholding tax schedules.[8] In those schedules, one spouse is assigned the favorable withholding tax class ("III"), while the other spouse is assigned the unfavorable withholding tax class ("V"). The spouse in the favorable withholding tax class is taxed as if the spouse was the single earner, while the withholding tax for the unfavorable withholding tax class is calculated as if the spouse was contributing a third of the household income (Spangenberg, Färber, and Späth, 2020). This leads to a lower withholding tax liability for the spouse in the favorable withholding tax class as compared to being in the default withholding tax class. Simultaneously, the withholding tax liability of the spouse in the unfavorable withholding tax class is higher than in the default withholding tax class and therefore also higher than without marriage. The second column in Figure 1.3 shows that, in the presence of joint taxation benefits, this decreases the joint withholding tax payments during the year if the primary earner is assigned to the favorable withholding tax class. In this setting with the husband earning more than

7. Switching away from the symmetric schedule requires the stated consent of both spouses. For switching back, however, unilateral action suffices.

8. These terms are not official but our own creations. In particular, the law does not explicitly refer to genders.

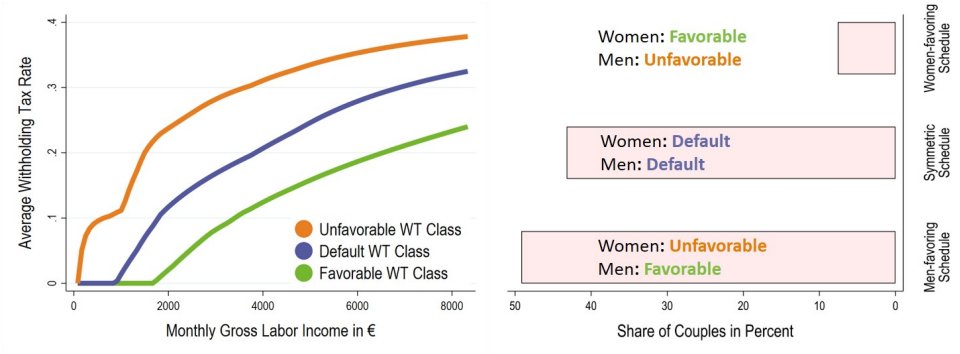


Figure 1.4. Illustration of Different Withholding Tax Schedules

Notes: The figure illustrates the frequency and implications of the different withholding tax schedules. On the left-hand side, the average withholding tax rate by withholding tax class in 2009 is shown. Compared to the default withholding tax class, being in the unfavorable withholding tax class leads to a much higher and being in the favorable withholding tax class to a much lower average withholding tax rate. On the right-hand side, the possible withholding tax schedules and their frequency are shown for the year 2009 for couples where both partners have labor income. Approximately 50 % of these couples choose the men-favoring schedule, in which the man is assigned the favorable withholding tax class and the woman the unfavorable withholding tax class. Around 45 % of the couples choose the symmetric schedule, which keeps both spouses in the default withholding tax class. Finally, less than 10 % of the couples choose the women-favoring schedule.

the wife, choosing the men-favoring schedule shifts the timing of the realization of the joint taxation benefit for the couple forwards and eliminates the lump-sum tax refund at the end of the year. In this concrete example, it even leads to the household paying too little in withholding taxes during the year which obliges them (in the absence of other deductions) to make an additional tax payment at the end of the year.

Conversely, if this couple had chosen the women-favoring schedule, which in this case puts the primary earner into the unfavorable withholding tax class and the secondary earner into the favorable withholding tax class, they would have paid even higher withholding taxes than under the default symmetric schedule and would have received an even larger tax refund at the end of the year. However, the women-favoring schedule is rarely chosen when the income of the husband is higher than the income of his wife.

Effect on tax rates. The shift of withholding tax liability from the primary to the secondary earner cannot only reduce the joint withholding tax liability but also has large effects on the withholding taxes paid by each spouse. The left-hand side of Figure 1.4 displays the average withholding tax rate by withholding tax class. Compared to the default withholding tax class, the average withholding tax in the favorable withholding tax class is lower while the average withholding tax in the unfavorable withholding tax class is higher. These differences are substantial. An individual earning €4,000 monthly gross income pays on average around 20 % in

withholding taxes in the default withholding tax class. The average withholding tax liability of the same individual increases to around 30 % when being in the unfavorable withholding tax class and reduces to around 10 % when being in the favorable withholding tax class. Consequently, the marginal withholding tax rates differ substantially between the different withholding tax classes (see Figure 1.A.2).

Choice of the different schedules. The right-hand side of Figure 1.4 shows the frequencies with which the different withholding tax schedules are chosen and which withholding tax class these schedules allocate to each spouse. Approximately 50% of the couples pick the men-favoring schedule that shifts the withholding tax liability from men to women and around 45% stick with the symmetric schedule. Less than 10% of the couples pick the women-favoring schedule with lower withholding tax rates for women than for men.

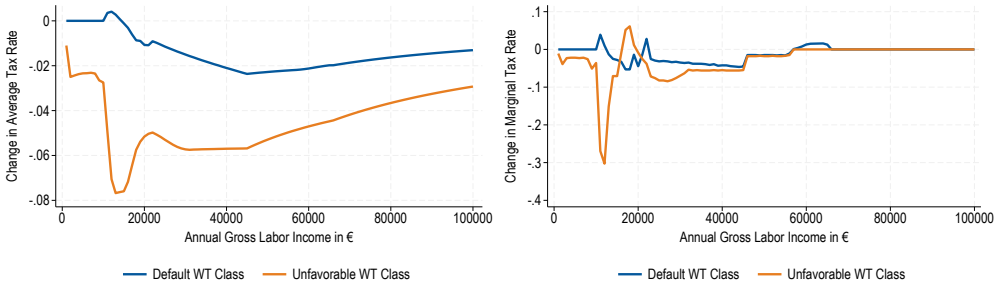
Importantly, while the different choices of withholding tax schedules have strong effects on the amounts of withholding tax payments, they do not affect the final income tax liability of the couple. Couples cannot decrease their final income tax liability by choosing a certain withholding tax schedule, but can only change the timing of the income tax payments throughout the year. However, when taking into account discount rates and liquidity constraints, couples can have benefits or costs from delaying their income tax payments.[9]

1.3 Empirical Strategy and Data

In this paper, we study the causal effect of withholding taxes on labor supply. Identification of this effect would be straightforward if withholding tax schedules were randomly assigned to each couple. However, the choice of withholding tax schedules is highly endogenous. Hence, simply comparing the outcomes of individuals in the different withholding tax schedules can potentially lead to a biased estimate of the effect of withholding taxes on labor supply.

We circumvent this problem by making use of a withholding tax reform in 2010 in Germany, which we outline in Section 1.3.1. The reform disproportionally reduced the withholding tax liability of individuals in the unfavorable withholding tax class compared to individuals in the other two withholding tax classes. In Section 1.3.2, we present the data we use to analyze the reform, in Section 1.3.3 we discuss the sample restrictions, in Section 1.3.4 we show descriptive statistics and in Section 1.3.5 we present the empirical strategy together with the identifying assumptions.

9. Additionally, the choice of withholding tax schedules impacts the size of wage replacement payments like parental leave benefits, unemployment benefits or short-time compensations. As we document in Section 1.F, this is a not well-known fact (also see Spangenberg, Färber, and Späth (2020) and Illing, Reutzel, and Wegmann (2024)).



(a) Average Tax Rates **(b)** Marginal Tax Rates

Figure 1.5. Reform on Withholding Taxes by Withholding Tax Class between 2009 and 2010

Notes: The figure plots the effect of the withholding tax reform 2010 on average and marginal withholding tax rates depending on the withholding tax class.

1.3.1 Withholding Tax Reform of 2010

Background. For the causal identification of the effect of withholding taxes on labor supply, we make use of a German tax reform in 2010 that changed the tax deductability of the mandatory health care insurance contributions. Prior to 2010, health insurance contributions were only taken into account in the calculation of withholding taxes for the default and favorable withholding tax class, but not for the unfavorable tax class. This changed in 2010, resulting in a substantial decrease in the withholding tax liability for individuals in the unfavorable tax class. The reform is described in detail in Stöwhase (2011a).

At the same time, the reform enabled all taxpayers to deduct a much larger share of their health care insurance contributions from the income tax. As the contributions to health care insurance are automatically deducted in the calculation of the withholding tax, the reform was equivalent to a cut in withholding taxes across all withholding tax schedules.

However, as the deductability was newly introduced for the individuals in the unfavorable withholding tax class, the reform as a whole reduced, conditional on income, the withholding tax for those individuals much more than for the individuals in the other withholding tax schedules.

Reform effect. Figure 1.5 shows how average and marginal withholding tax rates changed from 2009 to 2010 by withholding tax class and annual gross labor income. Figure 1.A.3 plots the absolute changes in the withholding tax rates. For example, for individuals with an annual gross labor income of €25,000, the average withholding tax rate decreased by about 5 % (€1300 for each year) for individuals in the unfavorable withholding tax class and by about 1 % (€280) for individuals in the default withholding tax class. The marginal withholding tax rate decreased by 8 % for individuals in the unfavorable withholding tax class and by 3 % for individuals in the default withholding tax class. For higher income levels, the change in

the marginal tax rate is identical for both groups. In other years, there are only minor changes to the withholding tax schedule. Figure 1.A.4 shows the withholding tax schedules for the years between 2006 and 2018 and for an annual individual income of €25,000, approximately the modus income of married women in 2009. The pattern is essentially the same along the whole income distribution.

Anticipation and salience. The reform was passed into law on July 23, 2009. There was no public debate about the effect the reform has on withholding taxes. It was not discussed in parliament and there were no newspaper articles discussing changes in withholding tax rates. In fact, we do not find any discussion about the withholding tax reform anywhere on the internet. This strongly suggests that the reform was neither anticipated nor salient. This assessment is corroborated by looking at Google Trends (see Appendix Figure 1.A.5). No striking movements are visible before the dates of the reform announcement and introduction. This implies that couples are not expected to have changed their withholding tax schedules around the reform date in response to the reform or adjusted their labor supply already prior to the reform. However, people were searching more for the term *withholding tax calculator* ("Lohnsteuerrechner") in January 2010 directly after the reform, suggesting that some people have perceived a change in their monthly wage income and tried to understand it. Furthermore, the reform's non-salience implies that spouses in the unfavorable withholding tax class might be unaware that their eventual income tax liability, regardless of it being perceived individually or jointly with their spouse, was not changed to the same extent. The only feature concerning withholding taxes that was indeed salient is the lower monthly withholding tax, i.e., a higher net income on the payslips.[10]

1.3.2 Data Set

Our study is based on the German Taxpayer Panel (TPP). The German Taxpayer Panel is an administrative dataset that contains information on the whole population of taxpayers in Germany for the years 2001 to 2018.[11] It includes information on various characteristics such as income, gender, age, number and age of children, withholding tax class and other tax-related information.[12] The TPP consists of around 63 million records on individuals for whom tax information is available for at least two years. The waves of the TPP for the years 2001 to 2011 were created

10. In addition, households might eventually also realize that they get lower tax refunds or have to pay higher additional tax payments in the upcoming year. However, it remains unclear whether they would connect this to the change on their payslip, particularly because tax refunds or additional tax payments occur on the couple level.

11. RDC of the Federal Statistical Office and Statistical Offices of the Federal States (2022)

12. At the time of the reform, same-sex couples were not yet allowed to benefit from joint taxation and were not allowed to choose their withholding tax classes. Thus, our sample contains only opposite-sex couples.

Table 1.1. Sampling Steps

Sampling stage	Restriction	Households 2009	Household × year observations
1.	Married dual-earner couples 2009	4,545,622	51,980,518
2.	Symmetric or men-favoring schedule 2009	4,015,516	46,216,194
3.	At most 60 years old	3,684,206	38,148,740
4.	Self-employed income 2009 < €1000	3,526,104	36,506,879
6.	Employed income 2009 < €100,000	3,389,714	35,068,697
7.	Civil servant	3,106,409	32,458,708
8.	Couple observable in years between 2006 and 2009	2,583,000	27,716,719
9.	No wage replacement (unemployment benefit, short-time work,...) between 2006 and 2009	1,735,414	18,854,273
10.	Change in annual income smaller than 15% between 2006 and 2009	671,624	7,253,108
11.	Women older than 40 years	574,163	6,084,411
12.	No change in withholding tax schedule between 2006 and 2009	490,288	5,196,307

from the annual income tax statistics, which include data from the tax returns of about 27 million German taxpayers who filed their income taxes. Starting in 2012, the annual federal statistics on wages and income tax replaced the income tax statistics that had been used previously, and the TPP has been continued using data from this statistic. As a result, from 2013 on, the TPP also includes data on about 12 million taxpayers who did not file their income taxes but who did face withholding taxes. In our analysis, we use the TPP for the years 2006 to 2018. In Appendix Section 1.H, we describe in detail how we calculate withholding taxes.

1.3.3 Sample Restrictions

In the following, we outline the restrictions that we impose on our sample. Their stepwise impact on the sample size is presented in Table 1.1.

Basic restrictions. We restrict our sample to married dual-earner couples, i.e., couples in which both spouses received labor income in 2009, the year before the withholding tax reform was implemented. This restriction ensures that these individuals are actually treated at the time of the reform.

We focus on couples in the two most common withholding tax schedules: the men-favoring and the symmetric schedule. We do so for two reasons. First, as shown in Section 1.2.2, the vast majority of dual-earner couples, around 95 %, has chosen either the men-favoring or symmetric schedule. Second, we deem the couples in those two schedules to be more comparable. In most couples in the women-favoring schedule, only the woman is earning labor income while the husband is

self-employed or is claiming pensions. Hence, these couples are very different from the couples in the other two schedules.

We also exclude observations where at least one partner is older than 60, as early-retirement schemes allow individuals to save taxes to reallocate income to later tax years, during which they no longer work. For computational reasons, we also exclude observations where one partner earns more than €100,000.[13] To ensure that labor income is the main source of work income, we exclude couples in which, in the year 2009, at least one spouse received income of more than €1,000 from self-employment. We also exclude couples with individuals earning no more than €4,800 per year. This condition ensures that we exclude individuals in marginal employment, who could earn at most €400 per month at the time of the reform and are exempt from the income tax. We also exclude civil servants as they were subject to a different withholding tax reform.

Financial crisis. The withholding tax reform of 2010, which we use for our identification, partially coincides with the financial crisis in Germany. We see in our data that couples in the men-favoring schedule experienced more extreme variations in labor income during the crisis years. Therefore, to make the couples in the two schedules more comparable, we exclude couples which were especially affected by the crisis. We do so by excluding couples in which at least one spouse received unemployment benefits or short-time work compensation in 2009 and by removing all couples in which at least one spouse had a change in annual labor income of more than 25 % from any one year to the next during the pre-reform years.

Stability of withholding tax schedule choice. A threat to identification is that households might change their withholding tax schedule and thereby be subject to a different withholding tax schedule. There are three potential risks: (i) Households might anticipate the reform and change their withholding tax schedule before the reform, (ii) households might change their withholding tax schedule as a reaction to the reform, and (iii) households might change their withholding tax schedule for other reasons in which case the estimated reform effect would be watered down. In principle, the choice of the withholding tax schedule is not fixed and can be changed by the couple every year.

However, as we observe empirically, typically couples decide on their withholding tax schedule once at marriage, often change it at the birth of the first child and then stick to it for the rest of the working life. Figure 1.B.1 plots the withholding tax around the birth of the first child. We see that the withholding tax schedule is

13. As explained in Section 1.3.5, we add interacted cell fixed effects in €5,000 steps, so keeping higher incomes scales up the complexity of the regression model. The main result is almost unchanged when including all incomes, which is not surprising as within the cells for high incomes there is almost no variation in the choice of withholding tax schedules and no variation in our measure of treatment intensity, the change in the marginal withholding tax.

changed around the birth of the first child, but then remains stable. Figure 1.B.2 depicts the transitions between the different withholding tax schedules of couples in the three different withholding tax schedules for all couples who were married in 2009. Typically, less than 1% of the couples change their withholding tax schedule in a given year and there is no evidence for an increase in withholding tax schedule changes before or after the reform.

In our preferred specification, we apply two sample restrictions to ensure that the withholding tax schedule is stable and that the reform is exogenous to the withholding tax schedule choice. First, we restrict the sample to households where the wife is older than 40 years in 2009. As we observe empirically in the tax data, almost no woman has the first child at the age of above 40, so households are much less likely to switch withholding tax schedules. Second, we restrict the sample to households that did not change their withholding tax schedule in the pre-reform periods between 2006 and 2009. As the reform was not salient (see Section 1.3.1), couples are not expected to change their withholding tax schedule before or after 2009 as a reaction to the reform. However, with this conservative assumption we ensure that the withholding tax schedule is stable and that the reform is exogenous to the withholding tax schedule choice for all households. We show in robustness checks that our results are robust to the removal of both restrictions.

1.3.4 Descriptive Statistics

Income growth. Figure 1.6 displays the development of our main outcome variable, the difference in log labor income in year t relative to 2009. The coefficients thereby approximate growth rates in income relative to 2009. In the years before 2009, average income growth is similar for women in the unfavorable and default withholding tax classes. Between 2009 and 2011, the effects of the financial crisis are visible and income increases only moderately. After 2010, the labor income growth rate is higher for women in the unfavorable withholding tax class.

Difference between groups. While in the years until 2009 the income growth looks very similar, it is important to note that the income distribution of the two groups is very different. Figure 1.7 shows in a heatplot the share of couples in the two withholding tax schedules depending on the income of the wife and the husband. It is visible that the choice of the withholding tax schedule strongly depends on the distribution of income within the household. The larger the income of the husband relative to the wife, the more likely the couple is to choose the men-favoring withholding tax schedule. For example, at the cell with the highest number of observations (wife earning between €15,000 and €20,000, husband earning between €35,000 and €40,000) 77% of the couples are in the men-favoring schedule

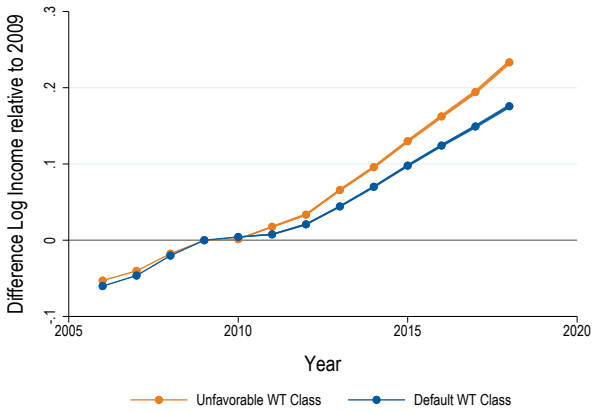


Figure 1.6. Income Growth Relative to 2009

Notes: The figure plots the log income difference in year t relative to 2009. The sample restrictions summarized in Table 1.1 are applied. The values correspond to population means as they are based on the universe of observations from the German Taxpayer Panel (TPP).

while 23% are in the symmetric schedule.[14] However, the two groups do not only differ in income variables. As Table 1.2 documents, the groups are also different in other observables. Specifically, couples in the men-favoring schedule are more likely to be have a child and substantially less likely to live in Eastern Germany. We account for these differences with our estimation strategy that we present in the following.

1.3.5 Estimation Strategy

We focus our analysis on comparing women in the unfavorable withholding tax class, who received a large withholding tax cut, to women in the default withholding tax class, who only experienced a modest withholding tax cut.

Treatment intensity. A naive approach would simply compare the evolution of incomes over time between women in the men-favoring and symmetric schedule using a difference-in-differences design. However, as previously shown in Figure 1.5, individuals' exposure to the reform is not only determined by their withholding tax class but also by their own pre-reform labor income. Therefore, depending on the own pre-reform labor income, there are large differences in the absolute and relative changes in withholding taxes induced by the reform.

To account for these differences in the intensity of treatment and to be able to calculate the elasticity of labor income with respect to withholding taxes, we

14. Buettner, Erbe, and Grimm (2019) document that the men-favoring and women-favoring schedules are not chosen symmetrically. This is not visible in Figure 1.7, as couples in the women-favoring schedule are not part of our sample.

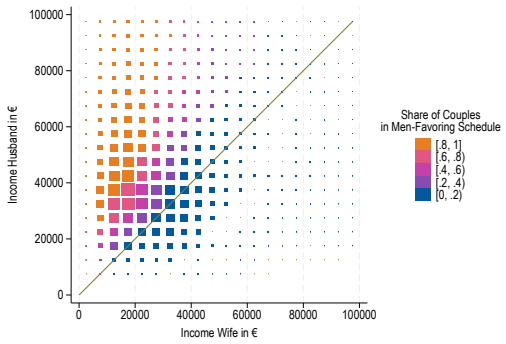


Figure 1.7. Heatplot of the Number of Observations and Share of Couples in Men-Favoring Schedule

Notes: The figure displays the number of observations and the exploited treatment variation by income cells. Each dot represents observations that lie in an interval of €5,000 wife and husband income. For example, the cell at the top-right corner contains women and men with incomes between €95,000 and €100,000. The larger the dot size, the more observations are in the respective cell. The color displays the share of couples in each cell who are in the men-favoring as opposed to the symmetric withholding tax schedule at the time of the reform. Incomes below €5000 are not displayed as these individuals are not subject to income taxes. The size of each bin represents the number of observations in our sample. The sample restrictions summarized in Table 1.1 are applied, so in particular couples in the women-favoring schedule are excluded.

perform our analysis using a continuous treatment variable. The continuous treatment variable measures the log change in the marginal net-of-withholding-tax rate of the woman induced by the reform.[15] We construct the treatment variable for each couple by taking the labor income of the woman in 2009 and calculating the percent change of her marginal net-of-withholding-tax rate resulting from using the tax schedule of 2010 compared to using the one of 2009. The reform-induced change in the marginal net-of-withholding-tax rate is plotted in Figure 1.8.

Identifying assumptions. As the core estimate, we want to measure the elasticity of labor income with respect to the withholding tax rate to obtain an estimate that is comparable to the elasticity of taxable income (ETI). Naturally, our estimation strategy is closely related to the estimation strategies that identify the effects of ETI. Jakobsen and Søgaard (2022) provide a comprehensive framework of the empirical challenges. They differentiate between *between-income tax variation* and *within-income tax variation*. *Between-income tax variation* is exploited by the typical identification strategy for estimating elasticities in the taxation literature that makes use of differential reform effects across income, for example comparing high-income individuals who were subject to a tax reform with low income individuals who were not. In contrast, *within-income tax variation* exploits variation between groups with

15. Using the change in the marginal net-of-withholding-tax rate instead of the marginal tax rate is standard in the literature on income tax elasticities (see for good overviews Saez, Slemrod, and Giertz, 2012; Jakobsen and Søgaard, 2022).

Table 1.2. Descriptive Statistics for the Year 2009

	Men-Favoring	Symmetric
Income Wife	19651.74	33321.58
	(8470.72)	(13402.3)
Income Husband	49737.3	39453.28
	(17046.99)	(15233.01)
Female Income Share	0.29	0.46
	(0.09)	(0.11)
Age Wife	44.63	44.69
	(4.47)	(4.97)
Age Husband	46.57	46.39
	(4.43)	(4.8)
Eastern Germany	0.08	0.36
	(0.27)	(0.48)
Has a Child	0.67	0.31
	(0.47)	(0.46)
Number of Children	1.42	0.76
	(0.88)	(0.86)
Catholic Wife	0.4	0.23
	(0.49)	(0.42)
Catholic Husband	0.37	0.2
	(0.48)	(0.4)
Public Servant Wife	0.12	0.12
	(0.32)	(0.33)
Public Servant Husband	0.2	0.15
	(0.4)	(0.36)
N	243,996	246,334

Notes: The table displays descriptive statistics for the year 2009 for couples who picked either the men-favoring or symmetric withholding tax schedule. The sample restrictions summarized in Table 1.1 are applied. Eastern Germany comprises the area of the former German Democratic Republic and West Berlin.

the same income level. This applies to our case where we compare women in the unfavorable withholding tax class to women in the default withholding tax class conditional on income.

Settings with *within-income tax variation* require two specific assumptions: First, conditional on income, the income growth of the two groups would be identical in the absence of the withholding tax reform. We can test this assumption for the periods before treatment, where the income trends of the two groups should be similar and placebo reforms should display no effect.[16] Second, our approach requires the

16. Similarly, *between-income tax variation* requires the assumption that, in the absence of changes in the tax schedule, the income growth in each year is identical along the income distribution. This often is a threat to identification due to mean reversion and differential secular income trends (see Weber (2014) and Jakobsen and Søgaard (2022)).

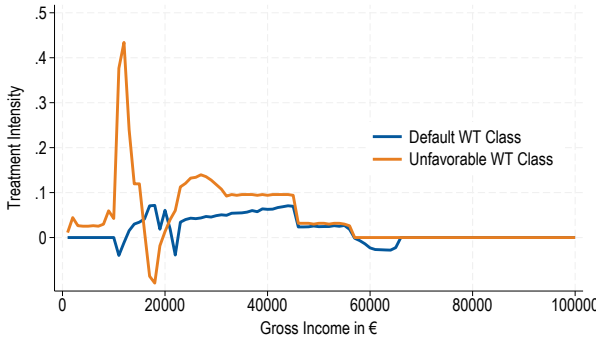


Figure 1.8. Reform-Induced Treatment Intensity

Notes: This figure displays the change in the net-of-withholding tax rate induced by the reform in 2010 depending on an individual's 2009 gross income.

assumption of effect homogeneity as both groups are treated (Callaway, Goodman-Bacon, and Sant'Anna, 2024). This implies, that both groups would react to an identical reform identically. A natural test for this assumption would be a withholding tax (or income tax) reform that hits both groups with the same treatment intensity. Unfortunately, there were no other major reforms in our observed period (see Figure 1.A.4).

Regression equation. Using the treatment intensity, we are able to estimate a difference-in-differences equation which yields us an estimate for the elasticity of outcome y with respect to the withholding tax:

$$\ln\left(\frac{y_{i,t}}{y_{i,2009}}\right) = \sum_{\substack{t=2006 \\ t\neq 2009}}^{2018} \beta_t \left[\ln\left(\frac{1-\text{MWTR}_{i,2010}}{1-\text{MWTR}_{i,2009}}\right) * \mathbb{1}(\text{Year}_t)\right] + \delta f(y_{i,2009}, y_{-i,2009}, \text{child}_{c,2009}, \text{east}_{c,2009}) \times \mathbb{1}(\text{Year}_t) + \theta X_{i,-i,c,2009} + \epsilon_{i,t}, \tag{1.1}$$

Consequently, β_t measures the percent change in labor income in year t relative to 2009 if the marginal net-of-withholding-tax rate of a woman increases by one percent. Further, $X_{i-,i,c,2009}$ controls for characteristics of the woman i, her partner $-i$ and the couple c as a whole in 2009. For both spouses individually we control for age and age squared and for the couple as a whole for the number of children and the region of residence.

We add dense couple-level cell fixed effects $f(y_{i,2009}, y_{-i,2009}, \text{child}_{c,2009}, \text{east}_{c,2009})$ interacted with year dummies. We do this by binning wives' and husbands' incomes in 2009 in steps of €5,000 and interacting these bins with each other. In our preferred specification, we also

interact these cells with indicators for whether a child lived in the household in 2009 and whether the household was located in Eastern Germany in 2009.

These controls are crucial for our identification: First, the controls for own income ensure that we only exploit variation in treatment intensity conditional on income. As the treatment intensity within a withholding tax class varies along the income distribution, a specification without the income controls would also pick up variation within withholding tax classes. Second, the interaction with the controls for partner income further ensures that we compare similar households. Controlling only for the income of the wife would not ensure that households have a similar distribution of income and are subject to the same income tax liability. We also illustrate the mechanics behind these cell controls graphically with a more detailed explanation in Appendix 1.C.

Lastly, interacting the income controls with the child and region indicator makes the comparison between groups even cleaner. As Table 1.D.1 shows, the child and region indicators are, besides income, the main determinants of different withholding tax class choices. By adding these characteristics as additional controls, we then only compare the evolution of income for households where prior to the reform both partners earn a similar amount, have at least one or no child and live in the same region. By interacting these controls with time fixed effects, we allow for different time trends for each cell.

1.4 Empirical Results

In this section, we present our empirical results showing that individuals adjust their labor income in response to changes in withholding taxes. We present event-study Diff-in-Diff estimates, show a robustness analysis and heterogeneities.

1.4.1 Main Results

First, we present in Figure 1.9 the results of the event-study Diff-in-Diff estimation as laid out in Equation 1.1. The independent variable is the treatment intensity of the reform measured by the log change in the marginal net-of-withholding-tax rate between 2009 and 2010. We thereby compare women in the favorable withholding tax class to women in the default tax class who were treated differently by the reform. The dependent variable is the log difference in income between year t and 2009. Hence, the coefficients can be interpretated as elasticities of labor income with respect to the withholding tax rate in the respective year.

Interpretation of pre-reform estimates. As discussed in Section 1.3.5, one implication of the parallel trend assumption is that we should not see economically significant effects of the reform before it takes place. The estimates of this placebo test are displayed in Figure 1.9. The pre-reform estimates are not statistically signifi-

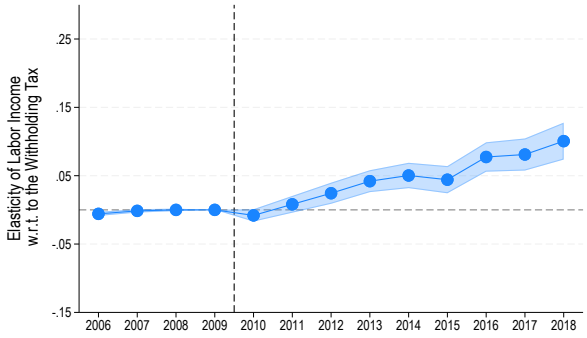


Figure 1.9. Event-Study Diff-in-Diff Estimates

Notes: The figure plots the estimates for the elasticity of labor income with respect to the marginal net-of-withholding tax estimated based on Equation 1.1 for the whole sample. The dependent variable is the log income difference in year t relative to 2009. The independent variable is the log difference in the marginal net-of-withholding-tax rate of the woman induced by the reform of the withholding tax in 2010. The regression includes controls for potentially time-varying characteristics of the couple. Cell fixed effects control for binned own and spousal pre-reform labor income interacted with dummies for parenthood, residence in Eastern Germany, and years. Confidence intervals are plotted at the 95 % level and based on heteroscedasticity-robust standard errors clustered at the household level. The sample excludes households in which at least one member experienced a drop in income by more than 15 % from one year to the next before 2010 to ensure that no individuals directly hit by the financial crisis are part of the sample. This explains the smaller standard errors before the reform. The regression is based on 5,483,974 observations with an adjusted R-squared of 0.095.

cantly different from zero at the 5 % level for most years and generally economically insignificant.[17]

Interpretation of post-reform estimates. We find significant positive estimates for the elasticity of labor income with respect to the withholding tax rate. The effect size gradually increases over time and reaches an elasticity of approximately 0.10. These results imply that a one percent higher change of the marginal net-of-withholding tax rate in 2010 results in an around 0.10 percent higher labor income eight years after the reform. More intuitively, the results indicate that a woman whose marginal withholding tax rate was reduced from 30 % to 23 %, so whose marginal net-of-withholding tax rate increased by 10 percent from 70 % to 77 %, increased her labor income by approximately 1.0 percent eight years after the reform.

Comparison to ETI literature. There are no established elasticities of labor income with respect to the withholding tax that we can relate our estimated elastic-

17. There are two reasons why our sample restrictions (see Section 1.3.3) result in smaller error bands before the reform compared to after the reform. First, we condition that each household is observable in all pre-reform years but after the reform households need not be observable in every year. This results in a larger sample size in the pre-reform years. Second, we exclude from the analysis households in which at least one individual was affected by the financial crisis.

ity to. However, estimates from the literature on the elasticity of taxable income (ETI) with respect to the income tax provide a useful benchmark. As Neisser (2021) stresses in a meta-analysis, estimates for the ETI are strongly context-dependent with an average elasticity of 0.29. Hence, our measured elasticity estimate is smaller than the typical estimate for a change in income taxes. This is in line with our expectation. As the reform changed only tax prepayments, only households with strong time preferences should react which would result in an elasticity very close to zero. Instead, as we discuss in Section 1.5.1, we expect that information frictions and incomplete pooling of household resources drive labor market responses of individuals.

Back-of-the-envelope calculation. To illustrate the magnitude of the effect, we estimate with a simple back-of-the-envelope calculation the expected impact of a reform that abolishes the choice of withholding tax schedules and instead assigns each individual the marginal income tax rate of the respective couple to reflect the true work incentives of both spouses under joint taxation.[18] We illustrate the effects on the marginal tax rate for both examined withholding tax classes in Figure 1.D.1. Such a reform has diverse effects, as the change depends on the partner's income and the withholding tax class before the reform. Focusing only on women who would experience a decrease in withholding taxes, the average decrease in withholding taxes of such a reform would be 5.5 %, resulting in an increase of the marginal net-of-withholding tax rate of 8.5 %. Using the elasticity of 10 %, this translates to an increase in labor income of about 1 %. However, these reform effects would be counteracted by women whose marginal withholding tax rate is lower than the income tax. For them, the marginal tax rate would increase by 7.5 % on average and, consequently, the marginal net-of-withholding tax rate would decrease by 9 %. As a result, labor income for this group would fall by about 1 %.

Gradual increase. The gradual increase of the observed effect size is in line with our expectation: First, it may take time for individuals to learn about the change in withholding taxes as the reform was non-salient. Shapiro and Slemrod (1995) find that one month after a much-debated cut in withholding taxes only a third of the respondents self-report that they noticed the change in withholding taxes even though employers were asked to actively inform their employees about the withholding tax change. As the reform we investigate was not noticed by the public (see Section 1.3.1), it can only have an effect if people realize the cut in withholding taxes independently. For example, individuals might realize that the monthly take-home pay changes or they might consult withholding tax calculators when considering a change in working hours or a job change.

18. A similar withholding tax system is in place in France, where both spouses are assigned the income tax rate of the couple in the previous year. For more details see Section 1.6.

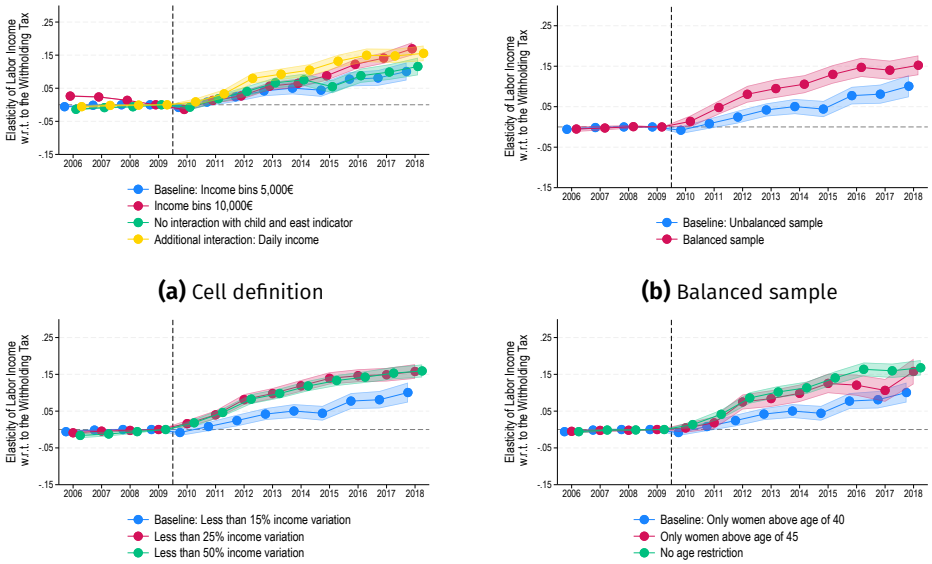
Second, due to labor market frictions the response time of individuals might be heterogeneous. We think that a substantial part of the treatment effect occurs when women obtain an offer from their employer to increase their working hours and then evaluate their marginal gains from doing so using the now lower withholding taxes as their reference point. It is difficult to compare the finding of the gradual increase to the existing ETI literature, as it is uncommon in the ETI literature to report the development of coefficients over time. However, our finding is in line with Gudgeon and Trenkle (2024) who show that low-income workers in Germany react slowly to changes in the German tax schedule.

Furthermore, it is striking that the treatment effect estimates never exhibit a reversing trend. One could expect that taxpayers might update their beliefs on the income tax system and, in particular, the interlinkage between withholding taxes and income taxes, after seeing their income tax returns sometime in the middle of the year following the tax year. If they by this learned about the true relationship between withholding taxes and income taxes, this should incentivize them to revoke potential reactions to the withholding tax reform. The fact that we cannot see such reactions might indicate that they do not gain understanding of the tax system by filing their income tax declaration and receiving the final tax statement. This is in line with our survey evidence (see Section 1.5.2) where we show that the general understanding of withholding taxes is low.

1.4.2 Robustness

In Figure 1.10, we explore the robustness of our empirical results to different restrictions on bin size and sample composition. We plot the estimates before the reform to investigate whether the pre-trends are robust to sample choices and the post-estimates to gauge the sensitivity of the estimates to changing the restrictions.

Other cell definitions. In Panel a, we vary the composition of the cell fixed effects. The cell fixed effects are chosen such that the identification of the reform stems from similar couples. If we allow the comparison of less similar couples by increasing the bin size of income controls to €10,000, the pre-reform estimates suggest a slightly different trend and the post-reform estimates are slightly larger. However, removing the interaction with the dummies for having a child in the household and living in Eastern Germany barely changes the estimates. Particularly interesting is the introduction of an additional interaction with daily income. In principle, we only observe annual income, but for a subset of observations that file their commute we also observe the annual number of working days. This shrinks the overall number of observations from 5,481,626 to 3,275,653. For these observations, we generate 10 percentiles of daily income and add them to the interacted fixed effects in equation 1.1. The estimated elasticity is slightly larger for this particularly comparable group.



(a) Cell definition

(b) Balanced sample

(c) Alternative income restrictions

(d) Alternative age restrictions

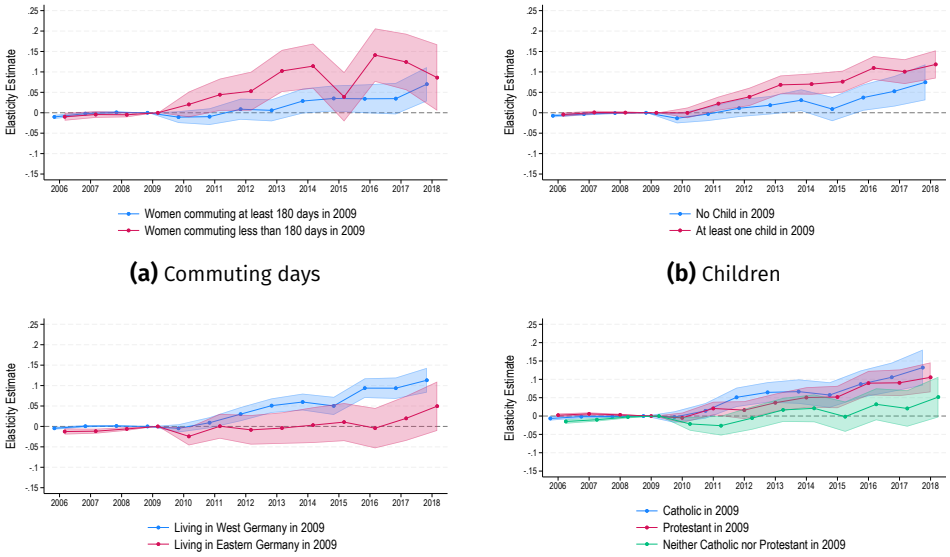
Figure 1.10. Robustness

Notes: See notes of Figure 1.9.

Balanced sample. For the years before 2012, we only observe couples that file taxes, afterwards we observe all individuals with taxable work income in Germany (see Section 1.3.2). To ensure that our results are not driven by differences in tax filing behavior or by leaving the sample for other reasons, Panel b shows the estimate for a balanced sample which slightly increases the main result but has no effect on pre-trends.

Financial crisis. To limit the impact of the financial crisis on our results, we have applied a restriction on the variation of the pre-reform incomes (see Section 1.3.3). In the baseline, we only consider couples in which both spouses had no year-over-year income change larger than 15 % prior to the reform. Relaxing this restriction to 25 % or 50 % in Panel c results in slightly higher estimates.

Stability of the withholding tax choice. Finally, as discussed in Section 1.3.3 we want to ensure that our results are not driven by changes in the withholding tax class after 2009. Withholding classes are often changed after childbirth, so in our main specification we concentrate on women older than 40. In Panel d, we show that setting this age restriction to 45 or abolishing it has only small effects on the regression results.



(a) Commuting days **(b)** Children

(c) West/Eastern Germany **(d)** Religious affiliation wife

Figure 1.11. Effect Heterogeneity

Notes: See notes of Figure 1.9. Eastern Germany comprises the area of the former German Democratic Republic and West Berlin.

1.4.3 Heterogeneity

Figure 1.11 displays the estimated effects for various subgroups. For each subgroup, we estimate the effect in separate regressions with a sample split.

Scope for behavioral changes. The effect is slightly stronger for subgroups who have more scope to adapt their labor supply. As we do not observe working hours in the data set, we proxy working part-time by the number of commuting days. If a women is commuting less than 180 days, we categorize her as a part-time worker. We find that the estimated effect is slightly larger for women working part time (Panel a).[19] In line with this finding, we also observe that the effect is smaller for the group without a child in the household in 2009 as they are more likely to work already full-time before the reform.[20] This is in line with previous studies investigating the behavioral reaction to income tax changes that typically find small behavioral reactions of men and substantial effects for women. This is typically explained by the larger scope for behavioral changes (Keane, 2011).

19. We only observe commuting days for individuals who claim commuter allowance which applies to 60% of the women in the sample.

20. It is worth noting that our analysis only includes dual-earner couples and therefore excludes couples where one partner solely provides childcare.

Other heterogeneities. First, we study the effect separately for Western and Eastern Germany (Panel c). We find that the effect is smaller in Eastern Germany. The reason for that could be that women in Eastern Germany generally work more and therefore have less potential to increase their working hours. Second, we observe a larger effect for Catholic and Protestant households than for other households (Panel d).[21]

1.5 Mechanisms: Theoretic Model and Survey

For a better understanding of our empirical results, we present a simple toy model to illustrate potential mechanisms that rationalizes why withholding taxes can matter for labor supply decisions. The core idea is that withholding taxes are only irrelevant for labor supply decisions if individuals fully internalize tax refunds and are indifferent to the timing of cash flows. We then provide evidence for the relevance of these mechanisms using survey data.

1.5.1 Theoretic Model

We set up a simple labor supply model in Appendix 1.E and derive the optimal labor supply decision as a function of three parameters: the understanding of the tax system (α), the share of the tax refund that the individual expects to receive (γ) and the factor with which the individual values money today in comparison to when the tax refund is paid out (ω). We discuss each of these parameters in the following and summarize their expected labor supply effects in Table 1.3.

Understanding of the difference between income and withholding taxes: α. The parameter α describes the degree to which individuals understand the difference between income and withholding taxes. If $\alpha = 0$, the individual does not realize that withholding taxes are only a prepayment to the income tax and that the payslip is not informative about the actual income tax. The higher α, the more the tax refund affects the labor supply decision. If $\alpha = 1$, the individual fully understands the interlinkage between withholding taxes and income taxes.

Share of tax refund that the individual expects: γ. If a married couple is overwithheld, the resulting tax refund is paid out to a single bank account as they are taxed jointly. The share of the couple's tax refund that each of the spouses receives is thus at the couple's discretion and the result of intra-household bargaining. The share of the tax refund that is expected by an individual is expressed in the parameter γ. The higher γ, the larger the share of the tax refund that the individual receives or that is - in case the tax refund is remitted to a joint bank account of

21. Most other religions and denominations do not collect their church tax through the German income tax system and their religious affiliation is thus not observable.

Table 1.3. Parameters Determining Relevance of Tax Refund

Parameter	Description	Effect on labor supply
α	Degree of understanding of withholding tax system	Lower understanding increases effects of withholding tax change
γ	Share of tax refund that individual expects to receive	Lower share of expected tax refund increases effects of withholding tax change
ω	Time preference	Lower discount factor increases effects of withholding tax change

Notes: The table displays the parameters of the theoretic model.

the couple - attributed to the individual. As both spouses face the same average income tax rate under joint taxation, $\gamma = 1$ reflects a distribution of the household's tax refunds between partners such that this equality of average income tax rates is upheld. Assuming that the average withholding tax rate of both spouses is equal, this implies that each spouse receives a share of the tax refund that is equal to their share of labor income.[22] If $\gamma < 1$, the individual receives a smaller share of the tax refund and ends up with a tax liability higher than implied by the income tax. Conversely, if $\gamma > 1$, the individual receives a larger share of the tax refund and ends up with a tax liability lower than implied by the income tax.

Time preference: ω. The parameter ω captures the degree to which the tax refund is discounted. If $\omega = 1$, the individual is indifferent between obtaining income today and receiving the same sum as a tax refund after the tax year. There are three reasons why ω could be smaller than 1. First, it might simply reflect time preferences. Second, discounting might be impacted by borrowing constraints. If monthly income is essential for the individual to pay for occurring costs of living, then the individual has a strong preference for obtaining the payment today. Third, rational investors can invest money received earlier. So the larger the real interest rate, the lower ω should be for the rational investor. This is especially relevant in countries like Germany or the US, where the government does not usually pay interest on overwithheld taxes.

Optimal Labor Supply. Solving our theoretical model in Appendix 1.E results in the following first order condition:

$$z_{it} = n_{it}\left[\tau_{it}^{WT} - f(\alpha\gamma\omega)(\tau_{it}^{WT} - \tau_{it}^{IT})\right]^{\varepsilon}, \tag{1.2}$$

22. To make the underlying idea more generalizable to other withholding tax systems, we have not explicitly modelled the German system of withholding tax schedules which we use for identification in this paper. In withholding tax systems that shift the withholding tax liability between partners, spouses must potentially have access to the labor income of the spouse in the favorable tax class above and beyond the tax refund to ensure that $\gamma = 1$.

where z_{it} denotes the taxable income and n_{it} the individual's counterfactual income in the absence of taxation. ε is the parameter of interest of our study, the elasticity of taxable income with respect to the marginal net-of-tax rate. $\tau_{it}^{WT} = 1 - T_t'^{WT}\left(z_{it}; x_{it}\right)$ denotes the individual marginal net-of-tax rate with respect to the withholding tax and $\tau_{it}^{IT} = 1 - T_t'^{IT}\left(z_{it}; x_{it}\right)$ the individual marginal net-of-tax rate with respect to the income tax in year t.

The main finding is that the optimal labor supply depends on the weight that an individual assigns to the tax refund. Assuming a positive tax refund, which implies that $\tau_{it}^{WT} > \tau_{it}^{IT}$, labor supply is impacted positively by the three parameters α, γ and ω. Intuitively, all factors that reduce the weight an individual assigns to the tax refund increase the relevance of withholding taxes for the labor supply. If one of the three parameters is 0, labor supply only depends on the withholding tax τ and is independent of the income tax rate $\hat{\tau}$. In this case, the size of the tax refund is irrelevant for the individual. In contrast, if all three factors are equal to 1, the withholding tax rate τ has no impact on the labor supply decision. Thus, in this case the only factor determining labor supply is the income tax. It is important to note that the three factors interact. For example, even a fully-informed individual without time discounting might still show a strong reaction to the size of withholding taxes if the dynamics within the household are such that she is not expecting the full tax refund that is attributed to her under equal average tax rates.

Regarding the time preferences ω, it is well documented in the literature that individuals have time discounting factors smaller than 1 (O'Donoghue and Rabin, 1999; Falk et al., 2018). For the other two factors, it is unclear whether they can influence behavioral reactions to withholding taxes.[23] Therefore, we study them with the help of self-collected survey evidence.

1.5.2 Survey

The goal of the survey is to study the understanding of the interlinkage between withholding taxes and income taxes, as well as the organization of household finances among married German couples.

In this section, we focus on the core results of our final analysis sample consisting of 506 (258 men, 248 women) married respondents in Germany. We provide more details on the setup of the survey in Appendix 1.F: Section 1.F.1 includes information on the implementation and our sample restrictions, Section 1.F.2 provides more-in-depth analyses, and Section 1.F.3 displays additional descriptive figures. Section 1.I displays the original survey questionnaire in German and a translation

23. We know from other settings that the institutional knowledge of individuals is limited (Chetty, Friedman, and Saez, 2013; Gideon, 2017; Rees-Jones and Taubinsky, 2020) and that households do not fully pool resources (see Almås, Attanasio, and Carneiro, 2023, for a recent survey of the literature).

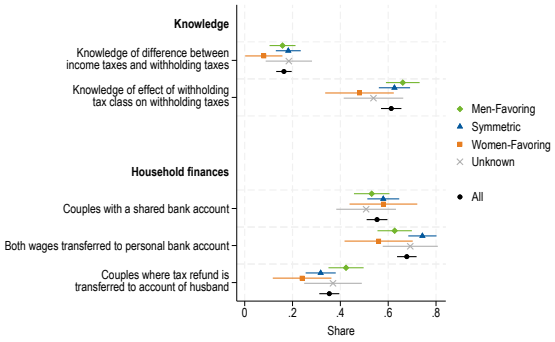


Figure 1.12. Survey on Potential Channels for the Effect of Withholding Taxes

into English. We have pre-registered our survey at the Open Science Foundation (registered as project SGXBP).

1.5.2.1 Understanding of Withholding Taxes

In our model in Section 1.5.1, we introduced the parameter α that captures the understanding of the difference between income and withholding taxes. To learn about α, we elicit whether our survey participants know that withholding taxes, and thus the choice of withholding tax schedules, do not affect a married couple's joint final income tax liability.

We do so by creating a realistic example of gross labor incomes of two spouses (one spouse earning €60,000 per year, the other one €30,000) and then ask the survey participants to select the withholding tax schedule which results in the lowest final income tax liability of the couple. We provide five answer options: the three withholding tax schedules, that it does not matter, and that they do not know the answer.[24] As discussed in Section 1.2.2, irrespective of the choice of the withholding tax schedule, the final income tax liability of the couple is the same.

Row 1 of Figure 1.12 presents the share of individuals who correctly answered the question. We find that only around 16 % of the surveyed individuals know about the irrelevance of the withholding tax schedule for the final income tax liability at the beginning of our survey. Respondents in the men-favoring tax schedule have a better knowledge than respondents in the women-favoring schedule but the knowledge is very limited across all schedules. In Appendix Figure 1.F.2, we also document heterogeneity in additional dimensions. In particular, men (20 %) are better informed than women (13 %) and individuals filing taxes themselves have better knowledge (22 %) than individuals not doing so (13 %).[25]

24. See Question D7 in Appendix Section 1.I for the exact wording of the question.

25. In Appendix Section 1.F.2, we show that (1) knowledge correlates with own tax filing for men and women likewise, (2) men file taxes alone more often than women, but that (3) that gender gap in tax filing cannot explain the gender gap in knowledge to a substantial degree.

Importantly, this observation is not driven by the fact that people are generally unaware of the functioning of withholding taxes. To test the general knowledge we use the fact that in the German withholding tax system the names of the different withholding tax classes are number coded. So for example, the favorable class is called *Steuerklasse 3* while the unfavorable class has the name *Steuerklasse 5*. As row 2 documents, 61 % of the respondents can correctly identify that relative to the default class (*Steuerklasse 4*), the withholding tax is lower in *Steuerklasse 3* while it is higher in *Steuerklasse 5*. This knowledge is lower for individuals in the women-favoring schedule but, as Appendix Figure 1.F.3 documents, homogeneous with respect to gender, withholding tax class, age except for the youngest cohort, and to whether the individual files taxes herself.[26]

Combining the two knowledge questions, we find that 48 % of all respondents know that, and how, withholding tax classes change withholding taxes but not that withholding taxes are tax prepayments and have no impact on the final income tax liability. This is a remarkable finding as it implies that a large share of married couples in Germany might fall for the fallacy that they can save income taxes by choosing a certain withholding tax schedule. Couples who know that the partner in the favorable withholding tax class is subject to lower withholding tax rates and the partner in the unfavorable one is subject to higher withholding tax rates (compared to the symmetric schedule and to individual taxation) might then strategically assign their primary earner to the favorable and their secondary earner to the unfavorable class (corresponding to the men- or women-favoring withholding tax schedule) due to the underlying lack of understanding of the interlinkage between income and withholding taxes. This then distorts the relative intra-household distribution of labor income as paid out by the employers.

The low rate of understanding suggests that for many individuals the relationship between withholding taxes and tax refunds is not salient. For these individuals, the withholding tax rate impacts labor supply decisions as people misperceive withholding taxes as informative for their income taxes.

1.5.2.2 Organization of Household Finances

In the model, the parameter γ identifies the share of the tax refund of the household that the individual expects to receive. This is why we also investigate the organization of household finances in our survey.

The basic assumption underlying withholding tax classes is that married couples act as unitary households. If this assumption does not hold, the choice of withhold-

26. After explaining the institutional setting, we also ask people to self-report whether they had understood the system correctly before (see Table 1.F.1). 54 % of the individuals claim that they had understood that the choice of the withholding tax schedule does not impact the final income tax liability, while 95 % claim that they knew that the choice of withholding tax classes impacts their withholding taxes.

ing tax schedule might impact the eventual intra-household distribution of labor income and by that the size of each spouse's budget and their within-household bargaining power. This is particularly relevant for couples choosing the men- or women-favoring withholding tax schedule as their intra-household distributions of take-home income are distorted compared to their distributions of gross income.[27] Sophisticated couples could make implicit transfers from the spouse in the favorable withholding tax class to the spouse in the unfavorable withholding tax class by getting wages of the spouse in the favorable withholding tax class transferred to either a shared bank account or the bank account of the spouse in the unfavorable withholding tax class and thereby undo the shifting of withholding taxes created by the men- or women-favoring schedule. Moreover, they could channel tax refunds to the spouse in the unfavorable tax class to (partly) account for the shifting of the withholding tax liability. We thus asked whether married couples use shared bank accounts and to which bank account potential tax refunds are transferred.[28]

Bank accounts. We test the assumption that married couples act as unitary households by examining the occurrence of shared bank accounts. If a couple does not have a shared bank account, it is very likely that the distortion of the relative intra-household distribution of labor income induced by choosing the men- or women-favoring withholding tax schedule remains largely unchanged as this couple is less likely to have established a compensatory sharing rule. As shown in the first row of the lower part of Figure 1.12, as much as 47 % of the respondents in the men-favoring withholding tax schedule state to not have a shared bank account as a couple. We consider these couples unlikely to account for the distortion of the relative intra-household distribution of labor income arising from the choice of that schedule. In this context, it is noteworthy that shared bank accounts do not seem to be used more often by couples in the men-favoring schedule than by those in other schedules, indicating that the choice of having a shared bank account is not commonly used to strategically counteract this distortion. However, as shown in the following row, couples in the men-favoring schedule less often get both their wages transferred to the respective personal bank account than couples in the symmetric schedule. Given that joint taxation benefits are distributed to both spouses in the latter but not in the former group, this difference is still strikingly small. This indicates that couples in the men-favoring schedule only use bank accounts to a very limited extent to counteract the distortions arising from the shift of withholding taxes between the spouses.

27. Figure 1.13 shows that for the median couple the wife in the men-favoring schedule has on average less take-home pay compared to separate taxation as not only the entire joint taxation benefit but also an additional amount is attributed to her spouse.

28. See Questions D16a and D17c in Appendix 1.I for the exact wording of the questions.

Tax refunds. Furthermore, we document that the distortion is even aggravated by the way couples deal with tax refunds. As the lowest row of Figure 1.12 shows, 42 % of the couples in the men-favoring withholding tax schedule (16 % of those with and 72 % of those without a shared bank account) let tax refunds be transferred to the husband's personal bank account whereas that share is lower for couples in the other withholding tax schedules. In comparison, only 24 % of the women in the men-favoring schedule get the tax refunds onto their personal bank account.

Taken together, we anticipate that the relative intra-household earnings will be skewed in favor of the husband for the majority of couples in the men-favoring schedule. Moreover, the disposable net income of married women is, on average, reduced compared to separate taxation. This could lead them to overestimate their individual income tax liability, which may negatively impact their bargaining power within the household and diminish their perceived work incentives. Consequently, these findings provide suggestive evidence that a large proportion of households does not fully pool their resources. This suggests that the share of a couple's tax refund that the individual expects to receive (parameter γ in the model in Section 1.5.1) might be substantially different from 1.

Given that a large share of married couples do not seem to choose the instruments they have available to distribute joint taxation benefits among themselves, the (default) options offered by governments deserve special attention. In the following, we thus examine the question of how withholding tax systems in the presence of joint taxation are designed in various countries and of how they should ideally be designed.

1.6 Implementation of Withholding Taxes for Married Couples

Joint income taxation treats married couples as one unit and therefore sets one joint income tax liability for the couple. In contrast, withholding taxes are always levied on the individual level. Not accounting for potential joint taxation benefits in the withholding tax system can lead to substantial overwithholding in progressive tax systems, i.e., couples end up having substantially more withholding taxes remitted during the year than they have to pay in income taxes after filing their taxes. Therefore, minimizing the overwithholding of couples requires the implementation of a withholding tax system specifically for married individuals. Consequently, governments have to take a stance on the distribution of the withholding tax liability within the couple (average withholding tax rates) and on the marginal work incentives in the withholding tax system (marginal withholding tax rates). In this section, we compare the current implementations of withholding taxes for married individuals in Germany, France and the US and evaluate potential effects of two recently proposed withholding tax reforms in Germany and France.

Different withholding tax implementations. Interestingly, while countries with joint taxation have a distinct withholding tax design for married individuals, it differs substantially across countries. These differences are indicative of the diverse set of objectives that policymakers seek to achieve through withholding taxes.

In 2018, *France* introduced a linearized withholding tax system in which the withholding tax rate is determined based on the most recent tax return. The tax authority uses the effective average income tax rate paid by the couple based on its past income and then sets this rate as a common withholding tax rate for both spouses so that marginal and average withholding tax rates are identical. In September 2025, the current system is replaced by an individualized withholding tax system. In this system, the withholding tax rate of the secondary earner is solely based on their own individual past income under the assumption of being unmarried. The primary earner's withholding tax rate, however, depends on the past household income and both spouses' past individual incomes as it is determined by the difference between the expected income tax of the household and the withholding tax paid by the secondary earner.[29] In this individualized system, marginal and average withholding tax rates are also identical, albeit now only on the individual level whereas they may differ between spouses.

In the *US*, all working individuals are asked to submit a W-4 form to their employer to decide on their amount of withholding. Households have an incentive to match their withholding to the expected income tax as they can be fined when underwithheld (Jones, 2012; Gelman et al., 2022). For married couples that file their taxes jointly, the most prominently presented option by the tax authorities is a system where for both spouses the withholding tax is determined similarly to the favorable schedule in Germany, i.e., under the assumption that the other spouse has no labor income.[30] To avoid potentially large underwithholding arising from the fact that both spouses are thus assumed to be the sole earner, couples are then asked to manually adapt the withholding of only the primary wage earner. The tax authorities provide tabulations and a tax calculator that suggest - based on the expected income of both spouses - how much in additional withholding taxes the primary wage earner should pay. If the couple claims additional deductions, e.g. the child tax credit, the tax authorities also recommend that only the withholding tax of the primary wage earner should be decreased.

For *Germany*, we compare the current implementation on the basis of the men-favoring withholding tax schedule with a widely-discussed reform option that decreases overwithholding by adjusting the withholding tax rates of both spouses

29. Couples already have the possibility to opt for this system since 2018.

30. Alternatively, couples can pick a withholding tax schedule that is similar to the German default symmetric schedule that, however, mechanically overwithholds substantially when the income gap between spouses is large.

based on past household incomes.[31] More explicitly, the reform option scales down the individual withholding tax rates that both spouses would have faced had they not been subject to joint taxation by a common factor. The factor ensures that the combined withholding tax payments of both spouses equal the expected income tax payment of the couple.

In the absence of year-over-year income changes, all implementations discussed above can completely eliminate overwithholding caused by the presence of joint taxation savings. However, they differ in how they affect both the marginal and the average withholding tax rates of the primary and secondary earner. We evaluate these differences separately with respect to average and marginal withholding tax rates for an example of a couple that is subject to the German income tax system and in which both the husband and the wife earn the respective median income of married men and women in Germany. Importantly, we compare the different implementations using the German income tax rates so that the differences are not driven by country-specific tax rates. Also we assume that the couple does not claim any further deductions which could change the observed effects in the French and US withholding tax systems substantially.

Average withholding tax rates. We start by evaluating the effects on the individual average withholding tax rate by showing how the monthly take-home pay of each spouse compares to the benchmark of separate taxation under the different withholding tax schedules in Figure 1.13. Withholding tax schedules in the upper-right quadrant result in a higher take-home pay for both the man and the woman compared to separate taxation, while those in the upper-left quadrant result in a higher take-home pay for the man and a lower take-home pay for the woman and those in the lower-right quadrant in a higher take-home pay for the woman and a lower take-home pay for the man. All withholding tax schedules that lie on the dotted red line avoid both over- and underwithholding by fully accounting for the joint taxation benefits of the couple in the withholding tax.

The figure shows that the three current withholding tax systems in Germany, France and the US all eliminate overwithholding for the example couple (the men-favoring schedule in the German system even slightly underwithholds the couple) but change the take-home pay of each spouse differently. In the German and French systems, the woman has a lower take-home pay relative to separate taxation after marriage, while the man has a higher take-home pay. This implies that the husband receives all of the joint taxation benefits and an additional transfer from his wife. In the US system, the opposite is true. The woman has a higher take-home pay and the man a lower take-home pay relative to separate taxation. Consequently,

31. In the German coalition agreement of 2021 (*German Coalition Agreement,* 2021), the parties agreed to abolish the current *system with a choice* and to replace it with the *schedule with a factor*. As a result of the break-up of the coalition in November 2024, the reform was, however, not implemented.

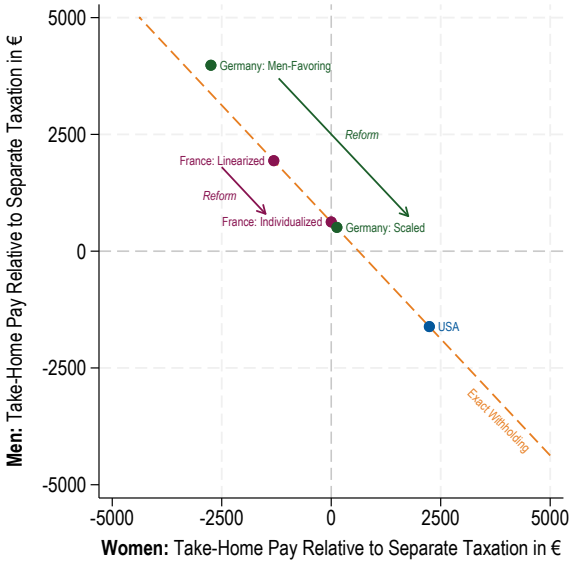


Figure 1.13. Average Withholding Tax: Monthly Take-Home Pay of Spouses After Marriage Relative to Separate Taxation

Notes: The figure illustrates the difference in the monthly take-home pay of joint taxation relative to separate taxation for both spouses in different withholding tax schedules given the German income tax system. All withholding tax schedules on the red dashed line feature precise withholding, i.e., the sum of withholding taxes of the two spouses matches the respective couple's income tax. In withholding tax systems below the red dashed line, households are overwithheld, while households in tax systems above are underwithheld. To the right of the y-axis (above the x-axis), the monthly take-home pay of the wife (husband) is larger than without marriage. The figure is based on a couple in which both the husband and the wife earn the respective median income of married men (€49,000) and women (€22,860) in Germany in 2010. We assume that the couple does not claim any credits or deductions when filing their income taxes.

the joint taxation benefit and a transfer from the husband are assigned to the wife. Both recently discussed withholding tax reforms in Germany and France similarly and dramatically reduce the effects of marriage for both spouses. As a result of the reforms, the husband would face a higher average withholding tax rate and therefore would receive a lower take-home pay, while the wife would face a lower average withholding tax rate and consequently would receive a higher take-home pay.[32]

Marginal withholding tax rates. An additional policy variable is the extent to which the individual marginal withholding tax rates mirror the couple's marginal income tax rate. In Figure 1.14, we illustrate how the individual marginal with-

32. While the illustrations in Figure 1.13 allow us to compare the effects that the different withholding tax schedules have on the monthly take-home pay for the example couple, it is impossible to gauge from this figure the effect that a reform of the withholding tax system would have for all couples. Therefore, to assess the consequences of a reform of the current withholding system in Germany, we compare the effects on the monthly take-home pay for married primary and secondary earners in the men-favoring schedule along the income distribution in Figure 1.G.1.

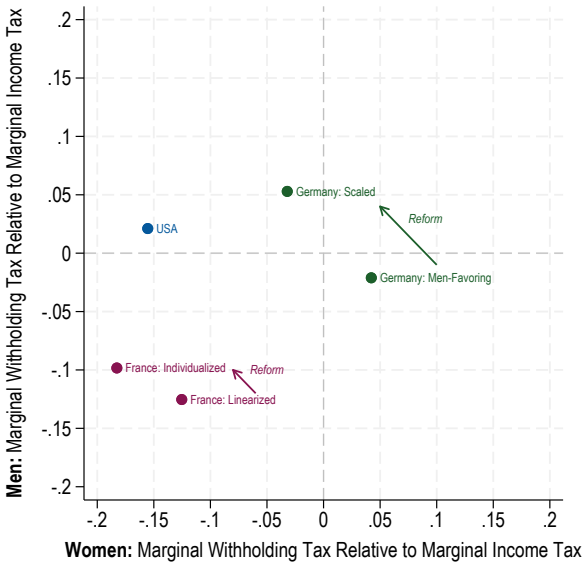


Figure 1.14. Reform Effects: Marginal Withholding Tax Rates of Spouses After Policy Reform

Notes: The figure illustrates the difference between the marginal withholding tax rates of the spouses and the marginal income tax rates of the couple. To the right of the y-axis (above the x-axis), the marginal withholding tax rate of the wife (husband) is larger than the marginal income tax of the couple. The figure is based on a couple in which both the husband and the wife earn the respective median income of married men (€49,000) and women (€22,860) in Germany in 2010. We assume that the couple does not claim any credits or deductions when filing their income taxes.

holding tax rates differ from the couple's marginal income tax rate. Withholding tax schedules in the upper-left quadrant result in a higher marginal withholding tax rate for the man and a lower marginal withholding tax rate for the woman compared to the couple's marginal income tax rate, while those in the lower-left quadrant result in a lower marginal withholding tax rate compared to the couple's marginal income tax rate for both spouses.

The figure shows that for the example couple the men-favoring schedule in Germany comes close to precisely implementing the marginal income tax rate of the couple in the spouses' marginal withholding tax rates but the marginal tax rate is slightly too low for men while it is somewhat too high for women.[33] The reform to the scaled schedule flips this relationship by decreasing the marginal withholding tax rate for women and increasing it for men.[34]

33. The men-favoring schedule is designed such that it mirrors well the work incentives for a couple in which the primary earner earns two thirds of the household income. This is approximately the case for the median couple which explains why the difference between marginal withholding and income tax rates is so small.

34. To assess the consequences of a reform of the current withholding system in Germany for all household income constellations, we compare the effects on the marginal withholding tax rate for

The current French implementation is characterized by lower marginal withholding tax rates compared to the couple's marginal income tax rate for both spouses. The reform results in a lower marginal withholding tax rate for the wife and a higher marginal withholding tax rate for the husband. The US implementation is marked by a lower marginal withholding tax rate compared to the couple's marginal income tax rate for the wife, but in a slightly higher marginal withholding tax rate compared to the couple's marginal income tax rate for the husband.

Implications for policymakers. Figures 1.13 and 1.14 illustrate the trade-offs that policymakers face when designing withholding tax systems for married couples in the presence of joint taxation benefits. Not accounting for joint taxation benefits in the withholding tax system can lead to substantial overwithholding. Although high overwithholding provides governments with a temporary, interest-free loan, it can have (unintended) consequences on households' labor supply decisions.

Figure 1.13 shows that all examined withholding tax schedules are able to eliminate overwithholding, but differ in how they distribute the benefits among the primary and secondary earner. The distribution of the joint taxation benefits among the spouses is relevant because, as we have shown in our survey, households often do not fully pool their income. Receiving a higher share of the joint taxation benefit might then increase the individual work incentives. The recently discussed withholding tax reforms in Germany and France both increase the take-home pay of the wife and decrease the take-home pay of the husband. This can potentially increase the work incentives of the wife.

The design of withholding taxes also affects the perceived marginal work incentives. This is of particular importance for secondary earners as the couple's marginal income tax rate is typically higher than the counterfactual marginal income tax rate of the secondary earner under individual taxation. Previous research (e.g., Bick and Fuchs-Schündeln, 2017) has argued that this higher marginal tax rate caused by joint taxation can lower the work incentives of secondary earners. However, with the help of the withholding tax system, policymakers can decide how closely the individual marginal withholding tax rates reflect the couple's marginal income tax rate. In light of our findings regarding the labor supply effects of withholding taxes, they can potentially attenuate the negative work incentives for secondary earners. Figure 1.14 shows that the example couple choosing the withholding-tax-minimizing schedule in the current withholding tax system in Germany approximately faces the "correct" marginal work incentives as each spouse's marginal withholding tax rate is fairly similar to the couple's marginal income tax rate.[35]

---

married primary and secondary earners in the men-favoring schedule along the income distribution in Figure 1.G.3.

35. The fit between the income and withholding tax rate depends on the chosen withholding tax schedule of the household. For many married couples in Germany, the chosen withholding-tax-

In contrast, the current French and US implementations "deceive" the individuals about their marginal work incentives as the marginal withholding tax does not equal the couple's marginal income tax. Given our findings, deceiving secondary earners by lowering their marginal withholding tax rates below the couple's marginal income tax rates could to some extent circumvent the negative work incentives of joint taxation for secondary earners and thereby increase the labor supply of women. The two recently proposed reform options do just that and thereby increase the work incentives for secondary earners. For primary earners, the figure shows that in both the US implementation and the German reform implementation the marginal withholding tax rates are above the couple's marginal income tax rates. Moreover, the two reform options both increase the marginal withholding tax rate of the primary earner compared to the respective current situation, which could decrease their work incentives. However, as the labor market attachment of primary earners is high and their labor supply relatively inelastic, we would expect that these higher marginal withholding taxes do not substantially affect their labor supply.

1.7 Conclusion

We show that withholding taxes impact labor supply decisions. Examining a reform that reduced withholding taxes more for some women than for others allows us to estimate the elasticity of labor income with respect to the marginal net-of-withholding tax rate. We estimate an elasticity that is larger than what would be expected from the response of a rational, fully-informed individual, since only budget constraints and other time preferences should affect the labor supply decision of such an individual. At the same time, our estimate is smaller than typical estimates for an income tax cut (Neisser, 2021). Based on a model that relates withholding taxes to income taxes and tax refunds, we identify two main mechanisms underlying our findings: imperfect understanding of the relationship between withholding and income taxes, and limited resource pooling within households. Using a self-designed survey, we provide empirical support for both mechanisms. First, fewer than 20 % of respondents are aware that withholding taxes can be fully credited against income taxes. Second, we find evidence that married couples do not fully pool their financial resources. These findings suggest that changes in withholding taxes may have real effects on the individual disposable incomes of spouses within joint taxation systems.

The novel finding that withholding taxes affect labor supply could help to bridge the gap between two parallel strands of the existing literature on behavioral re-

minimizing schedule results in marginal withholding tax rates that are close to the respective couple's marginal income tax rate for both spouses (see Figure 1.G.2).

sponses to income taxation. On the one hand, it is well documented that individuals respond to income taxation (see Neisser, 2021, for a comprehensive review). On the other hand, people often exhibit a limited understanding of income taxes (see for example Feldman, Katuščák, and Kawano, 2016; Stantcheva, 2021). This raises the question of how individuals can respond to taxes that they often do not fully understand. Our results suggest that withholding taxes can constitute a central cornerstone in understanding how people learn about the tax rates they face. In particular, withholding taxes could help to explain the empirical findings from the "schmeduling" literature (Liebman and Zeckhauser, 2004; Rees-Jones and Taubinsky, 2020). This literature documents that a large proportion of taxpayers "iron", i.e., people assume that their marginal tax rate is equal to their average tax rate for every euro of their realized labor income. We posit that this misperception may stem from the salience of withholding taxes, which are prominently displayed on monthly payslips and often represent taxpayers' primary point of contact with the income tax system. People may consequently misinterpret the information they receive on their payslips by inferring their marginal labor supply incentives from simply dividing their monthly take-home pay by their monthly gross labor income. Future research can provide valuable insights into the cognitive processes and heuristics that shape individuals' responses to income taxation, which can then be used to design more efficient tax systems.

Apparently, the potential to influence the decision making of households with the design of withholding taxes is already recognized within the political sphere. In the US, the W-4 form to claim withholding taxes was substantially simplified in 2020, which should reduce the overwithholding of households. In France and Germany, reforms of the withholding tax systems, which reduce the withholding tax burden for secondary earners at the expense of primary earners, have been proposed. Even though withholding taxes only marginally affect resources at the household level, reforms are difficult to implement when they redistribute income from one spouse's paycheck to the other spouse's paycheck. For example, the German proposal substantially decreases withholding taxes for the secondary earner while increasing them for the primary earner. In combination, this increases the monthly take-home pay for the secondary earner at the expense of the primary earner. This kind of reform affects disposable income on individual level if households do not pool resources, for which we show evidence in this paper.

An additional interesting aspect of the proposed reforms is that they do not aim to align marginal withholding tax rates with marginal income tax rates. Instead, they reduce marginal withholding tax rates for secondary earners while increasing them for primary earners, suggesting that withholding taxes are used as tools to increase the labor supply of women.

Appendix 1.A Institutional Setting

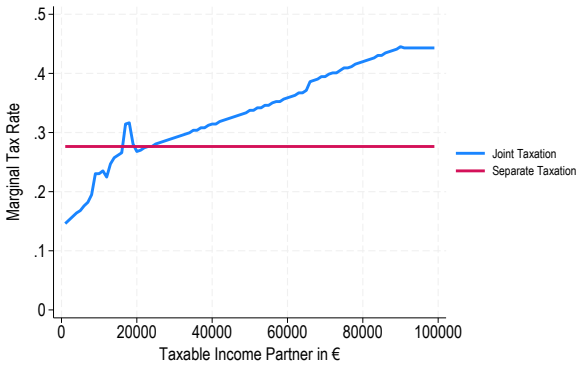


Figure 1.A.1. Marginal Income Tax for Income of 24,000 €

Notes: The figure shows the marginal income tax rate depending on the income of the partner for an individual earning €24,000 under both joint and separate taxation. When a couple is taxed jointly, the marginal income tax rate for the individual increases in partner income as the marginal income tax is a function of household income. If the couple is taxed separately, the marginal income tax rate of each spouse does not depend on partner income.

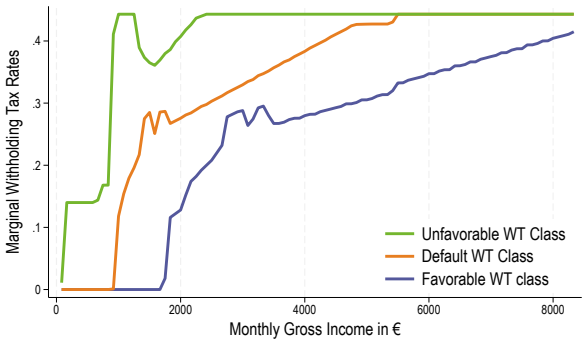


Figure 1.A.2. Marginal Withholding Tax Rates 2009

Notes: The figure plots the marginal withholding tax rates by withholding tax class in 2009.

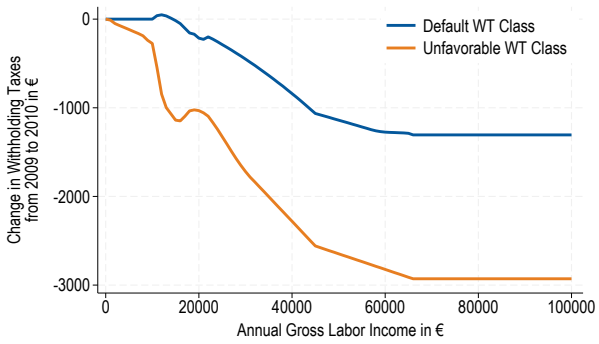


Figure 1.A.3. Absolute Size of Reform on Withholding Taxes by Withholding Tax Class Between 2009 and 2010

Notes: The figure plots the effect of the withholding tax reform 2010 on average and marginal withholding tax rates depending on the withholding tax class.

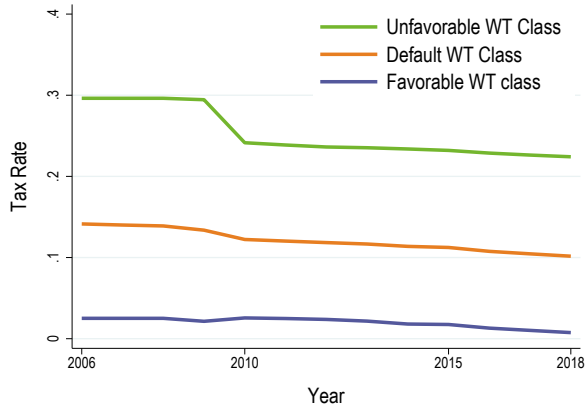
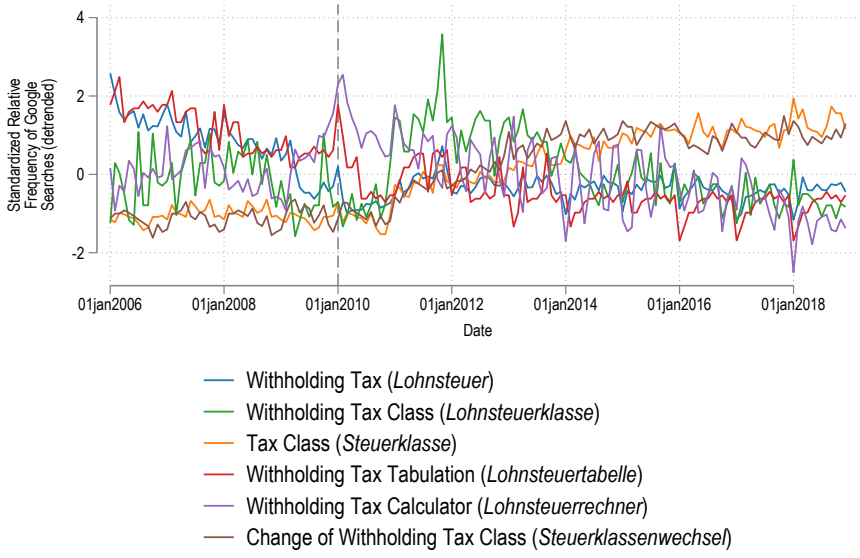
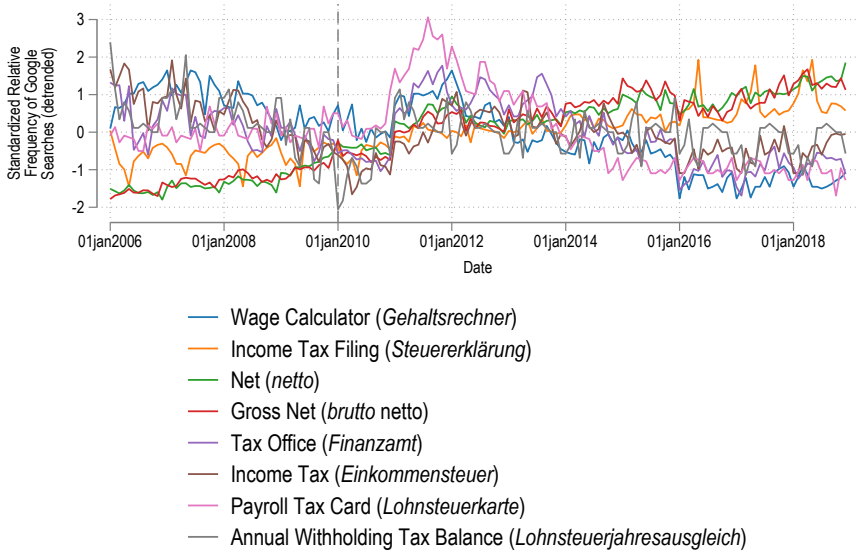


Figure 1.A.4. Development of the Average Withholding Tax Rate

Notes: The figure plots the size of withholding tax payments depending on the withholding tax class for the period from 2006 to 2018. It illustrates for an income of €25,000 that there were no other major reforms changing withholding tax payments except for the 2010 reform that we study in this paper. This holds true along the whole income distribution.



(a) Main search terms



(b) Other search terms

Figure 1.A.5. Trends in Relevant Google Search Terms

Notes: The figure plots the mean-standardized detrended search for terms relevant in the context of withholding taxes on Google.de. For the terms withholding tax reform (*Lohnsteuerreform*), tax prepayment (*Steuervorauszahlung*), other synonyms for a change in the withholding tax class (e.g. *Steuerklassenänderung*, *Steuerklasse ändern/wechseln*), or explicit searches for withholding tax classes (like *Steuerklasse V*), the search volume is so low that the results are not published by Google.
Source: trends.google.de

Appendix 1.B Descriptive Statistics

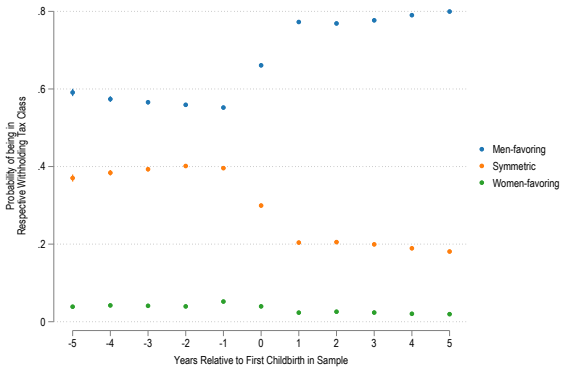


Figure 1.B.1. Choice of the Withholding Tax Schedule Relative to the Birth of the First Child

Notes: The figure displays the share of couples in the three different withholding tax schedules around the birth of the first child for all children born between 2008 and 2018.

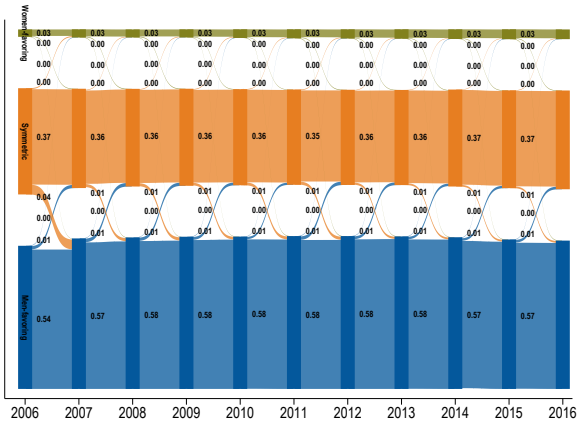


Figure 1.B.2. Changes in the Choice of Withholding Tax Classes Over Time

Notes: The figure displays the share of couples in the three different withholding tax schedules and the transitions between the different withholding tax schedules over time. The figure shows that the choice of withholding tax schedules is relatively stable over time. Couples typically stick with their choice of withholding tax schedules and only a few couples change between different withholding tax schedules. Note that we only consider direct transitions between withholding tax schedules. So we do not include cases where couples do not file their taxes in a specific year and later reenter the dataset with a different withholding tax schedule. The difference in the shares to Figure 1.4 stems from the changed data composition. While this figure has no sample restrictions, the right side of Figure 1.4 documents the share only for couples where both spouses are working.

Appendix 1.C Empirical Strategy

In our empirical strategy, we control for binned own and spousal pre-reform labor income interacted with dummies for pre-reform parenthood and residence in Eastern Germany by adding couple-level cell fixed effects to our regression (see Section 1.3.5). We motivate the underlying reason for the cell controls in the following in more detail.

Controlling for own pre-reform labor income is necessary as the treatment intensity does not only vary across withholding tax classes, but also across labor income. This is illustrated in the lower part of Figure 1.C.1, which displays the change in the net-of-withholding tax rate, i.e., the treatment intensity, induced by the reform. As we only want to use the variation in treatment intensity caused by the different choice of withholding tax classes, it is important to control for own pre-reform income.

Moreover, there are also reasons why it is important to additionally control for spousal pre-reform labor income. First, controlling for joint household income enables us to compare women that face the same income tax liability on the couple level but different changes in their withholding taxes. Second, controlling for the relative within-household labor income allows us to control for the economic importance of own labor income and a couple's labor market related gender norms. Gender norms of the within-household division of labor arguably play a large role in explaining labor market decisions of spouses as well as their choice of withholding tax schedule.[36]

In order to address these above-outlined channels, we follow an empirical strategy that is similar to Carbonnier et al. (2022). The idea is to divide observations into cells to exploit variation in treatment within each cell. In our preferred specification, we classify each individual into one of 1,600 cells based on own and spousal pre-reform labor income in 2009 and dummies indicating parenthood and residence in Eastern Germany. We include the dummies to make sure that we account for the most relevant predictors of the withholding tax schedule choice as shown in Figure 1.D.1. Thereby, we ensure that the compared individuals are more similar in observable characteristics. The cells are created by interacting evenly spaced bins of €5,000 of both own and partner income up to the maximum individual income of €100,000. Each of the resulting 400 cells is then interacted with dummies for parenthood and residence in Eastern Germany. By adding the resulting couple-level cell fixed effects interacted with years as controls we only use the variation in treatment intensity within each cell. We thus compare women with similar own and spousal

36. Our survey shows that couples in the men-favoring schedule hold more traditional gender norms than those in the symmetric schedule (see Figure 1.F.1). Comparing only couples with a similar within-household division of labor income could mitigate this problem because, as we show in Table 1.2, this division is correlated with the choice of withholding tax schedule.

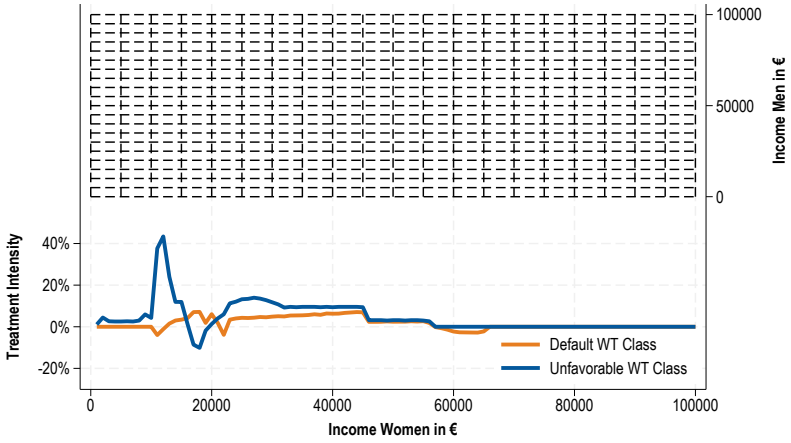


Figure 1.C.1. Illustration of Income Cell Approach

Notes: This figure illustrates the idea behind using income cell fixed effects. The lower part of the figure displays the change in the net-of-withholding tax rate induced by the reform in 2010. The upper part of the figure illustrates the income cell approach. We create bins for the income of women and men, interact them with each other and interact the resulting income cells with our sample years. By adding these interacted cells to our regression equation, we only exploit variation in the treatment within the cells.

pre-reform income characteristics. The remaining variation in treatment that we exploit then only comes from the different choices of withholding tax schedules.

We illustrate how the cell approach helps to tackle endogeneity concerns in Figure 1.C.1. Along the x-axis, the cells help to control for own pre-reform labor income so that differences in treatment intensity are only induced by the choice of withholding tax schedule, not by the income level. Along the y-axis, differences in relative within-household labor income are accounted for. Two women with the same own labor income but different withholding tax classes can still be very different with regards to other relevant factors such as the economic importance of own labor income which is driven by the size of their partners' income. A woman earning €30,000 with a partner earning €20,000 has very different work and tax incentives compared to the same woman with a husband earning €80,000. Using the cell approach therefore ensures comparing more similar couples.

Given the arguments brought forward so far, though, controlling for both own and spousal income separately would be sufficient. However, not only relative within-household labor income but also absolute household labor income might play a role. Couples with higher absolute labor income might tend to choose other withholding tax schedules but also react differently to changes in the net-of-withholding-tax rate. Thus, the bin approach controls for differences in absolute household labor income along the diagonal of the upper part of Figure 1.C.1.

The variation that we can exploit with the cell approach is illustrated in Figure 1.7. It shows for each of the income cells the share of couples who are treated in

a binary sense, i.e., the share of couples being in the men-favoring withholding tax schedule at the time of the reform conditional on being in the men-favoring or symmetric withholding tax schedule. The size of each cell represents the number of observations, meaning that cells with larger dots contain a larger share of the observations in our sample. The plot shows that for the largest shares of couples the husband earns between €20,000 and €50,000 and the wife between €10,000 and €40,000 and that within those cells there is a considerable amount of variation in the choice of withholding tax schedules.

Appendix 1.D Results

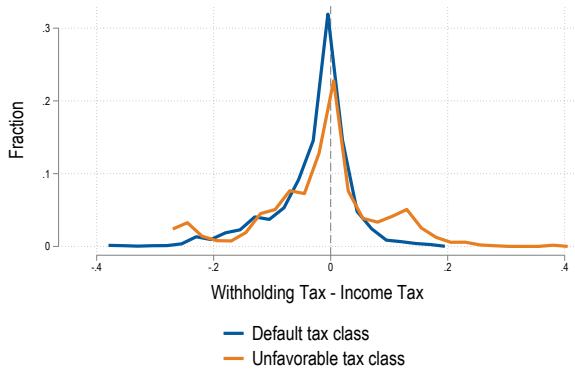


Figure 1.D.1. Difference Between Marginal Withholding and Marginal Income Tax Rates for Married Women

Notes: The figure displays the deviations of married women's marginal withholding tax rates from their marginal income tax rates. The deviations are shown separately for women who chose the default withholding tax class and for women who chose the unfavorable withholding tax class. A positive coefficient implies that the marginal withholding tax rate is larger than the marginal income tax rate. To ease interpretation, we ignore all non-standard deductions. All calculations are based on a 10 % sample of German administrative tax records from the year 2010 (RDC of the Federal Statistical Office and Statistical Offices of the Federal States, 2010), using the German tax code.

Table 1.D.1. Determinants of the Choice of Withholding Tax Schedules

	Choice of Men-Favoring Schedule
Eastern Germany	−0.221***
	(0.011)
Female Income Share	−0.017***
	(0.001)
Income Wife (1000 Euro)	−0.005***
	(0.001)
Has a Child	0.113***
	(0.011)
Number of Children	0.058***
	(0.006)
Catholic Wife	0.005
	(0.01)
Catholic Husband	0.027***
	(0.01)
Age Wife	0.003**
	(0.001)
Age Husband	0.005***
	(0.001)
Constant	0.891***
	(0.054)
N	490,330
Adj. R^2	0.51

Notes: The table displays which characteristics of a couple are predictive for the choice of the men-favoring schedule instead of the symmetric schedule. The coefficients stem from the regression of a dummy indicating the men-favoring schedule on various characteristics of couples in the year 2009, just before the withholding tax reform. Heteroscedasticity-robust standard errors are displayed in brackets.

Appendix 1.E Theoretic Model

To better understand the underlying mechanism, we present a simple toy model to illustrate potential mechanisms that can rationalize why withholding taxes can matter for labor supply decisions. The core idea is that withholding taxes are only irrelevant for labor supply decisions if individuals fully internalize tax refunds and are indifferent to the timing of cash flows. We provide evidence for the relevance of these mechanisms in Section 1.5.2.

1.E.1 Tax Refund and Determinants of its Relevance for Labor Supply

The tax refund that an individual receives after the tax year is specified as

$$R_t\left(z_{it};x_{it};w_{it}\right) = T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - T_t^{IT}\left(z_{it};x_{it}\right), \tag{1.E.1}$$

where $T_t^{WT}\left(z_{it};x_{it};w_{it}\right)$ denotes the withholding tax liability that depends on taxable income z_{it}, a set of other tax-relevant variables x_{it} (following the notation of Jakobsen and Søgaard, 2022) and the individual's withholding tax class w_{it}. $T_t^{IT}\left(z_{it};x_{it}\right)$ denotes the eventual income tax liability, which does not depend on the individual's withholding tax class w_{it}.

In a model without behavioral factors, an individual's tax liability $T_t\left(z_{it};x_{it}\right)$ is equal to their withholding tax liability minus the tax refund

$$\begin{aligned} T_t\left(z_{it};x_{it}\right) &= T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - R_t\left(z_{it};x_{it};w_{it}\right) \\ &= T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - [T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - T_t^{IT}\left(z_{it};x_{it}\right)]. \end{aligned} \tag{1.E.2}$$

Importantly, $T_t^{IT}\left(z_{it};x_{it}\right)$ is equal to $T_t\left(z_{it};x_{it}\right)$ which demonstrates that the withholding tax liability is not a determinant of the total tax liability.

In the following, we expand Equation 1.E.2 to investigate how the introduction of different behavioral factors that influence the perception of future tax refunds can impact labor supply decisions: the understanding of the tax system (α), the share of the tax refund that the individual expects to receive (γ) and the factor with which the individual values money today in comparison to when the tax refund is paid out (ω). We model these factors as a weight that individuals assign to the tax refund and discuss them extensively in Section 1.5.1.

Adding the three parameters α, γ and ω as weights to the tax refund in Equation 1.E.2, the individual tax liability is given as follows:

$$T_t\left(z_{it};x_{it}\right) = T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - f(\alpha\gamma\omega)[T_t^{WT}\left(z_{it};x_{it};w_{it}\right) - T_t^{IT}\left(z_{it};x_{it}\right)]. \tag{1.E.3}$$

1.E.2 Derivation of Optimal Labor Supply

To derive optimal labor supply in our model with withholding taxes, tax refunds and the three parameters α, γ and ω we specify the standard quasi-linear utility function with respect to consumption c_{it} and taxable income z_{it}

$$u_{it}\left(c_{it}, z_{it}\right) = c_{it} - \frac{n_{it}}{1+\frac{1}{\varepsilon}}\left(\frac{z_{it}}{n_{it}}\right)^{1+\frac{1}{\varepsilon}}, \tag{1.E.4}$$

that each individual maximizes subject to the following budget constraint

$$c_{it} \leqslant z_{it} - T_t\left(z_{it}; x_{it}\right). \tag{1.E.5}$$

The parameter n_{it} expresses the individual's counterfactual income in the absence of taxation while ε is the parameter of interest, the elasticity of taxable income with respect to the marginal net-of-tax rate.

The individual tax liability given by Equation 1.E.3 gets plugged into the budget constraint given by Equation 1.E.5. Utility maximization then results in the following first-order condition with $\tau_{it}^{WT} = 1 - T_t^{\prime WT}\left(z_{it}; x_{it}\right)$ denoting the individual marginal net-of-tax rate with respect to the withholding tax and $\tau_{it}^{IT} = 1 - T_t^{\prime IT}\left(z_{it}; x_{it}\right)$ denoting the individual marginal net-of-tax rate with respect to the income tax in year t

$$z_{it} = n_{it}\left[\tau_{it}^{WT} - f(\alpha\gamma\omega)(\tau_{it}^{WT} - \tau_{it}^{IT})\right]^{\varepsilon}. \tag{1.E.6}$$

Appendix 1.F Survey

In this section, we provide more information on the implementation of the survey (1.F.1), discuss additional findings from the survey concerning the filing behavior of couples and gender norms, and present results from an information treatment (1.F.2).

1.F.1 Implementation

We pre-registered our survey with the Open Science Foundation (registered as project SGXBP) and subsequently ran it on the micro job platform Clickworker between December 2022 and April 2023. We prescreened the participants so that they all speak German, are between 20 and 60 years old, married, and employed. We remove 73 respondents from our sample who fail at least one of two attention checks.[37] Furthermore, we restrict the sample to respondents with employed spouses. This makes the sample more comparable to the sample for our main analysis with observational data.[38] Our final sample then consists of 506 respondents (258 men, 248 women).

The survey questions can be divided broadly into four main categories. First, we directly inquire about the participants' understanding of withholding taxes in Germany. Second, we ask for information on the intra-household division of tax-planning and financial decisions. Third, we elicit participants' preferences on changing their weekly working hours and check whether an information treatment, which informs the participants about the withholding tax system in Germany, changes these preferences. Lastly, we elicit respondents' gender norms. Table 1.F.1 documents descriptive statistics of the survey answers.

1.F.2 Detailed Survey Analysis

Filing of taxes. One way to gauge which couples are particularly affected by the intra-household distortion of earnings and thus by adverse labor supply incentives for women is to examine the role of the filing of taxes in the income tax declaration made in the calendar year following the respective tax year. We asked respondents about their tax filing behavior and concentrate on those who file their income taxes

37. The attention checks can be found in the questions A2 and D15 in Appendix 1.I.

38. We also exclude respondents from our analysis who are in a same-sex marriage, where one of the two partners is non-binary or when the gender is not stated. This is for two reasons: First, there is an option for spouses in a same-sex marriage to keep that marriage secret from their employers by choosing withholding tax class I instead of III, IV, or V. This might then influence their knowledge of withholding taxes in an unforeseeable way. Second, same-sex couples were not yet allowed to benefit from joint taxation and were thus not allowed to choose their withholding tax classes at the time of the 2010 reform.

Table 1.F.1. Survey Descriptives

	N	Mean	SD	Min	Max
Demographics					
Women	506	0.49	0.50	0	1
Age	506	39.39	8.67	22	57
At least one child	500	0.61	0.49	0	1
High school degree	506	0.79	0.41	0	1
Work-related Variables					
Income	443	46,896	23,532	5,000	100,000
Income partner	438	43,790	23,904	5,000	100,000
Working hours	506	34.45	9.74	5	65
Working hours partner	506	34.08	10.53	0	70
Gender Norms					
Husband should have last word at home	506	1.89	1.43	1	7
Both partners should work the same	506	2.63	1.71	1	7
Husband should take care of financial security	506	2.80	1.86	1	7
Gender norm index	506	0.00	1.00	-1.13	3.58
Household Finances					
Men-Favoring WT schedule	506	0.35	0.48	0	1
Tax consultant decided on WT class	506	0.09	0.28	0	1
Husband decided on WT class alone	506	0.09	0.29	0	1
Wife decided on WT class alone	506	0.05	0.22	0	1
Shared bank account	506	0.55	0.50	0	1
Income tax filing	501	0.96	0.19	0	1
Husband's wage transferred to shared bank account	506	0.27	0.45	0	1
Tax refund on shared bank account	506	0.44	0.50	0	1
Feel financial constraint at end of month	483	0.44	0.50	0	1
Tests of Knowledge					
Correct identification of existing WT schedules	506	0.39	0.49	0	1
Understood difference IT/WT before survey	506	0.16	0.37	0	1
Understood impact WT schedule on WT	506	0.61	0.49	0	1
WT impacts parental leave benefits	506	0.43	0.50	0	1
WT impacts unemployment benefits	506	0.30	0.46	0	1
Self-reported Knowledge					
Difference IT/WT before survey	506	0.54	0.50	0	1
- Among those who failed knowledge test	423	0.52	0.50	0	1
Impact WT class on WT before survey	506	0.95	0.22	0	1
- Among those who failed knowledge test	196	0.91	0.29	0	1
Overall Sample Size	506				

Notes: The Table displays descriptive statistics for the answers to the survey (see Appendix 1.I for the full questionnaire). Income and working hours were asked in brackets and the respective mean was assigned to every individual for this tabulation. The gender norm index is calculated by summing up the item responses to the three questions (inverting the answers to Question 2) and mean-standardizing this sum.

jointly as a married couple as it is the case for our analysis sample in the administrative data.[39]

Looking at heterogeneities by gender, we find that, among these respondents, 56 % of men but only 37 % of women state that they usually do the majority of the tax declaration alone. This difference in tax filing behavior is driven by couples in the men-favoring withholding tax schedule. Of all men in the men-favoring withholding tax schedule, 65 % do the tax declaration mostly alone, while this only applies to 35 % of the women in that schedule. In the symmetric schedule, however, the gender difference is much lower with 50 % of the men and 46 % of the women claiming to do the tax declaration mostly alone, respectively. This shows that a more gender-equal exposure to the income tax system correlates with a less distortive distribution of withholding taxes in favor of men.

As documented in Figure 1.F.2, women less often than men know that withholding taxes do not have an influence on the final income tax liability. This gender gap in knowledge about the tax system could be linked to the amount of time and effort spent dealing with it by preparing tax declarations. Moreover, we see that those respondents that do most of the tax declaration alone also exhibit a larger knowledge about the irrelevance of withholding taxes for the final income tax liability at the beginning of the survey. For women, knowledge increases from 10 % to 17 % when they deal with the tax declaration mostly alone, for men from 16 % to 25 %.

A possible conclusion from these findings is that couples in which predominantly the husband deals with taxes are also more affected by the incentive distortions arising from the shifting of the withholding tax liability from husbands to wives. This may indicate a self-manifesting role of the household division of tasks, whereas this division itself might be linked to gender norms.

Gender norms. As Stöwhase (2011b) and Buettner, Erbe, and Grimm (2019) show with administrative tax records, German married couples are more likely to choose the men-favoring withholding tax schedule when the husband outearns the wife than to choose the women-favoring schedule when the wife outearns the husband.[40] This phenomenon could be attributed to a gender norm that prescribes the husband to be the main breadwinner (Bertrand, Kamenica, and Pan, 2015). As a consequence, couples with such a norm should be more likely to choose the men-favoring withholding tax schedule.

We investigate this by asking the respondents three questions, each with seven ordered answer options, to elicit their norms regarding gender roles in house-

39. This applies to 82 % of our respondents. A joint tax declaration has to be signed by both spouses but no other participation in filing the declaration is needed. See Question D17 in Appendix 1.I for the exact wording of the question.

40. Moreover, they also more often choose the men-favoring schedule when the wife outearns the husband than they choose the women-favoring schedule when the husband outearns the wife.

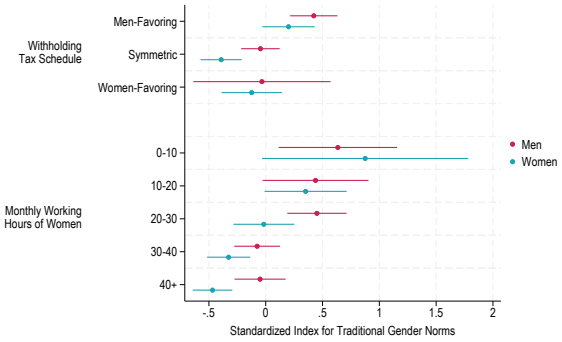


Figure 1.F.1. Gender Norms Index by Gender

Notes: The figure plots standardized index values for gender norms by gender for different withholding schedules and bins of monthly working hours of the wife. A higher value of the gender norms index is associated with more traditional gender norms, i.e., a desired larger role for husbands than for wives with regards to decision-making in the household and market work.

holds.[41] From the answers to these questions, we create a standardized index of the traditionality of gender norms where a higher value means that the respondent wants to have a larger role for husbands than for wives with regard to decision-making in the household and to market work.

As shown in Figure 1.F.1, respondents in the men-favoring have more traditional gender norms than those in the symmetric withholding tax schedule. This holds true for both men and women and indicates that those most affected by distortions of labor supply incentives are also those who favor a traditional division of market and non-market work. This is particularly relevant as the figure also shows that women who hold more traditional gender views are, as expected, also those that have the highest margin for adjusting their working hours as they tend to have fewer working hours than women with more progressive gender norms.

Information treatment. We also implemented an information treatment, consisting of an explanation of the German withholding tax system, in particular of the relationship between the choice of withholding tax classes and monthly take-home pay and of the irrelevance of the withholding tax choice for the income tax. We then assessed whether treated participants - those 423 out of the 506 interviewed individuals who did not know before our provided explanation that the choice of the withholding tax schedule is irrelevant for the income tax - want to change their withholding tax classes and adapt their working hours with these new information at

41. See Question D18 in Appendix Section 1.I for the exact wording of the questions. All three questions have been asked in this form in previous waves of the German Socio-Economic Panel (GSOEP).

Table 1.F.2. Information Treatment

	N	Mean	SD	Min	Max
Success Inform. Treatment					
Would change working hours now	171	0.04	0.20	0	1
Would have changed working hours in the past	171	0.15	0.35	0	1
Would change withholding tax class	158	0.05	0.22	0	1
No Success Inform. Treatment					
Would change working hours now	232	0.12	0.32	0	1
Would have changed working hours in the past	232	0.20	0.40	0	1
Would change withholding tax class	213	0.20	0.40	0	1
Observations	403				

Notes: The table displays the answers to the information treatment for the sample of those survey participants who did not know before our provided explanation that the choice of the withholding tax schedule is irrelevant for the income tax. Also it excludes the individuals who answered "do not know" to the respective questions before or after the treatment.

hand.[42] After the explanation of the withholding tax system, we asked again about the optimal withholding tax choice. As Table 1.F.2 documents, only 42 % (179 individuals) of the 423 individuals who could have learned about the system with the help of our explanation indeed learned that the choice of withholding tax schedules is irrelevant for the income tax. Thereby, the statistical power is too small for a meaningful analysis of the information treatment. Suggestively, the treatment has a negative impact on the intention of changing working hours today or in the past.

42. Questions D10 and D11 explain the system while Question D12 tests the understanding of the irrelevance of the withholding tax choice for the income tax. Questions D13a-c ask about intended behavioral consequences.

1.F.3 Additional Survey Figures

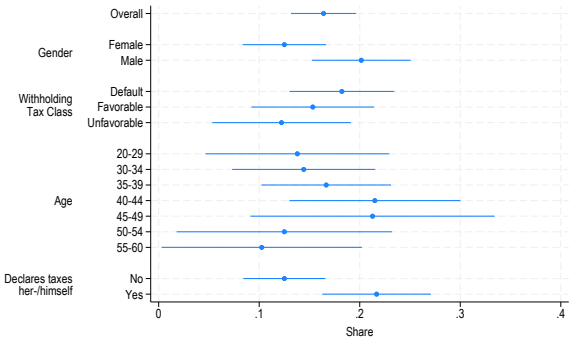


Figure 1.F.2. Knowledge of Interlinkage between Withholding Tax and Final Income Tax Liability by Subgroups

Notes: The figure plots the overall and subgroup-specific shares of surveyed individuals who correctly identify that the choice of withholding tax class does not impact the final income tax liability given an example of the labor incomes of two spouses (one spouse earning €60,000 per year, the other one €30,000). See Question D7 in Appendix 1.I for the exact wording of the question.

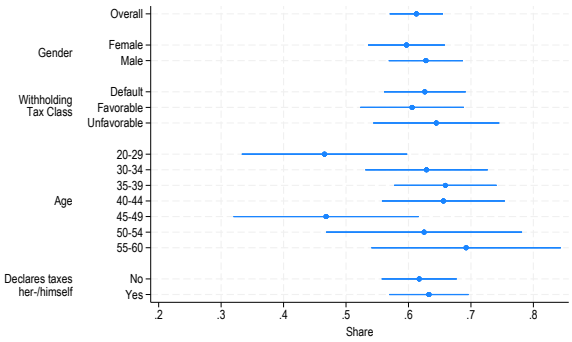
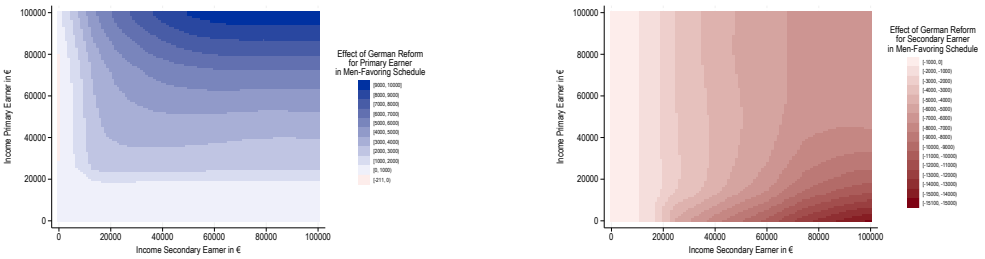


Figure 1.F.3. Knowledge of Interlinkage Between Withholding Tax Classes and Monthly Payslip

Notes: The figure plots the overall and subgroup-specific shares of surveyed individuals who correctly identify that and in which way the choice of withholding tax classes impacts the monthly net wage received from one's employer. Respondents are classified as being knowledgeable if they both answer correctly what happens qualitatively with respect to monthly wage transfers from their employers when changing from the default withholding tax class to (1) the favorable withholding tax class and (2) the unfavorable withholding tax class. See Questions D8 and D9 in Appendix 1.I for the exact wording of the questions.

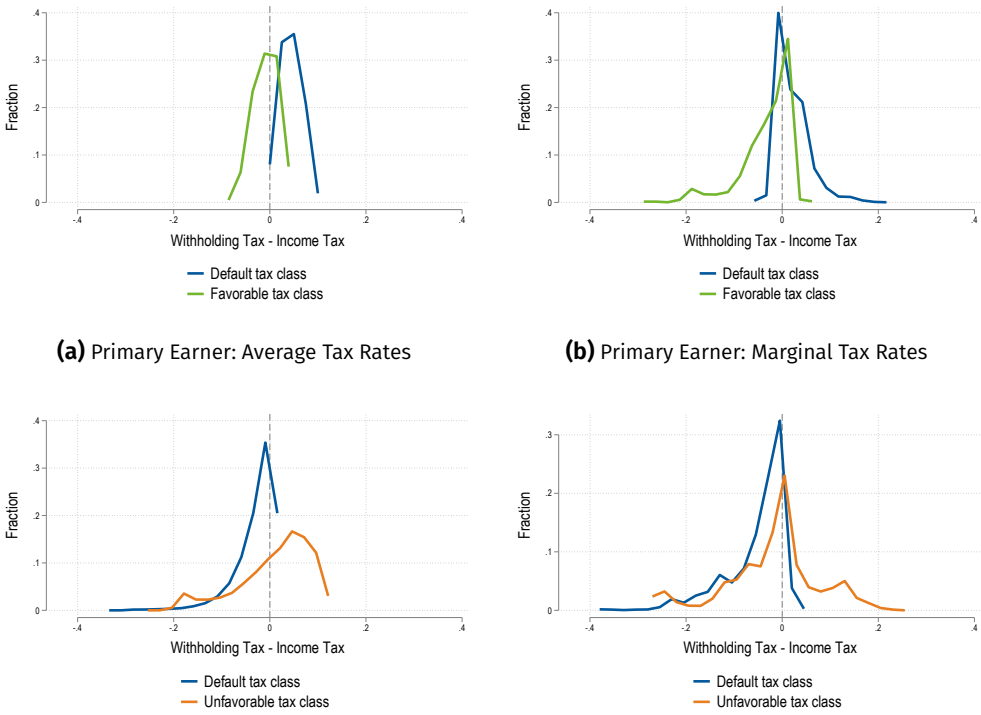
Appendix 1.G Implementation of Withholding Taxes for Married Couples



(a) Primary Earner **(b)** Secondary Earner

Figure 1.G.1. Effect of German Reform on Take-Home Pay for Both Spouses

Notes: The figures display the change in the annual take-home pay resulting from the reform option for primary and secondary earners in the men-favoring schedule. The effects are displayed for all possible income combinations below €100,000 of individual annual labor income. With "primary earner" we denote the individual in the household with higher labor income and with "secondary earner" we denote the individual in the household with lower labor income. We assume that the couple does not claim any credits or deductions when filing their income taxes. For almost all primary earners, the withholding tax substantially increases due to the reform while it decreases for secondary earners.



(a) Primary Earner: Average Tax Rates

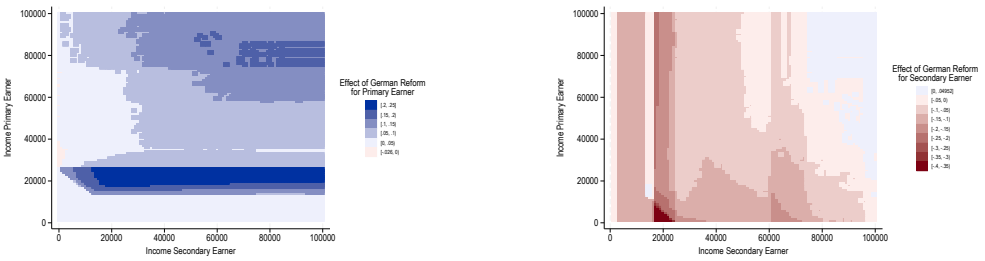
(b) Primary Earner: Marginal Tax Rates

(c) Secondary Earner: Average Tax Rates

(d) Secondary Earner: Marginal Tax Rates

Figure 1.G.2. Difference between Withholding and Income Tax Rates

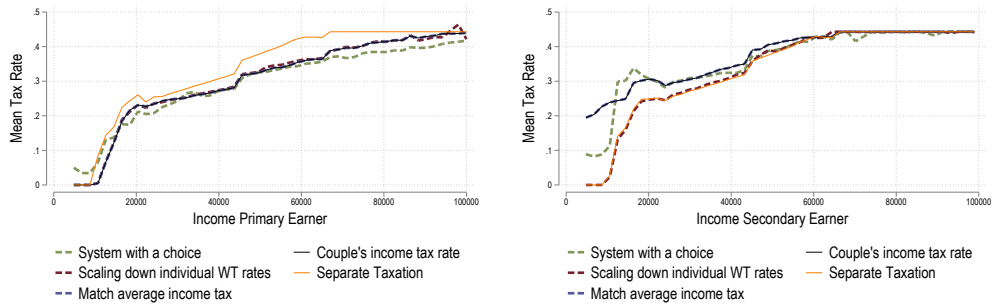
Notes: The figures display histograms of the deviations of marginal and average withholding tax rates from the couple's income tax rates. The deviations are shown separately for individuals that chose the default withholding tax class and for individuals that chose the favorable/unfavorable withholding tax class. A positive x-axis value implies that the withholding tax rate is larger than the income tax rate. We display the deviations in tax rates separately for primary earners in Panel (a) and (b) and for secondary earners in Panel (c) and (d). All calculations are made using the German tax code and are based on a 10 % sample of German administrative tax records from the year 2010 (RDC of the Federal Statistical Office and Statistical Offices of the Federal States, 2010). With "primary earner" we denote the individual in the household with higher labor income and with "secondary earner" we denote the individual in the household with lower labor income. We assume that the couple does not claim any credits or deductions when filing their income taxes.



(a) Primary Earner **(b)** Secondary Earner

Figure 1.G.3. Effect of German Reform on Marginal Tax Rates for Both Spouses

Notes: The figures display the change in the marginal withholding tax rate resulting from the reform option for primary and secondary earners in the men-favoring schedule. The effects are displayed for all possible income combinations below €100,000 of individual annual labor income. With "primary earner" we denote the individual in the household with higher labor income and with "secondary earner" we denote the individual in the household with lower labor income. We assume that the couple does not claim any credits or deductions when filing their income taxes.



(a) Primary Earner: Marginal Tax Rates **(b)** Secondary Earner: Marginal Tax Rates

Figure 1.G.4. Marginal Withholding Tax Rates in Different WT Regimes for Married Couples

Notes: The figures display the *long-term* mean marginal withholding tax rates induced by different withholding tax systems. Additionally, the couple's mean income tax rates and the mean income tax rates under separate taxation are displayed. We display the tax rates separately for primary earners in Panel A and for secondary earners in Panel B. All calculations are based on a 10,% sample of German administrative tax records from the year 2010, using the German tax code (RDC of the Federal Statistical Office and Statistical Offices of the Federal States, 2010). With "primary earner" we denote the individual in the household with higher labor income and with "secondary earner" we denote the individual in the household with lower labor income. When interpreting the figures it is important to keep in mind that along the x-axis individuals have partners with different income. Typically, individuals with higher income also have a partner with higher income. Moreover, in contrast to the remainder of the paper, the figures also include couples where only one partner has wage income. Hence, the panels for the primary earner include more households than the panels for the secondary earner. To ease the interpretation of the figures, we ignore all non-standard deductions.

Appendix 1.H Calculation of Withholding Taxes

Our aim is to calculate the withholding tax as precisely as possible, because our treatment intensity is based on the withholding tax in 2009 and 2010. In Subsection 1.H.1, we discuss how we calculate the withholding tax (*Lohnsteuer*") in detail and in Subsection 1.H.2 how we deal with other withheld taxes and social security contributions. In Subsection 1.H.3, we compare the estimated withholding tax with the withholding tax observed in the data and show that overall we approximate the withholding tax liability well.

1.H.1 Details on the Calculation of Withholding Taxes

The goal of the German government when designing the withholding tax system was to match income tax payments as closely as possible. All standard deductions for the income tax are thus implemented for the withholding tax as well. For the calculation of withholding taxes, we follow the program flowchart (*Programmablaufplan*) published by the Federal Ministry of Finance. To ensure the correctness of the code we test our calculations against the officially published withholding tax tabulations for employers (*Prüftabelle Allgemeine Lohnsteuer* and *Prüftabelle besondere Lohnsteuer*). Also, together with the program flowchart, the ministry publishes random test cases. Our calculator also passes all of these. The calculation depends on various, not always obvious, input parameters that we discuss in this section.

1.H.1.1 Income and Withholding Tax Class

We observe every taxpayer's yearly income and withholding tax class.

1.H.1.2 Payments for Insurance

The contributions for insurance are partly deductible from the income tax and consequently they are also automatically deducted from the withholding tax.

Individuals fully insured in the public social security system. Contributions to the public pension, health and care insurance are partly deductible from the income tax. The exact same rules are applied to the automatic deduction from the withholding tax.

Additional contributions specific to the health insurance provider. Additional contributions specific to the health insurance provider (*Kassenindividuelle Zusatzbeiträge*) were introduced in 2010, but only from 2015 onwards they were taken into account for the calculation of withholding taxes.

Individuals without children. Individuals aged 23 and older without children pay an increased contribution rate to the public care insurance. We take that into account.

Saxony. In Saxony, the contribution rate to the public care insurance is higher than in the rest of Germany. We consider that.

Private health insurance. Conditional on specific characteristics (most importantly occupation, age and income) individuals have the possibility or obligation to opt out of the public health and care insurance system. Public and private health care insurance are slightly differently taken into account in the calculation of withholding taxes and women in the men-favoring schedule in the private health care insurance experienced a smaller cut in withholding taxes in 2010. Unfortunately, we cannot differentiate between privately and publicly insured individuals in the data and assume that everybody is publicly insured. So for this group we potentially underestimate the effect of withholding taxes, and we estimate a lower bound of the elasticity.

No contributions to public pension. Civil servants are excluded from the public pension system and typically hold private health insurance without contributing to the public health insurance. We exclude them from the analysis, as they are treated differently by the reform in 2010.

1.H.1.3 Proportional Tax Allowances for Elderly Retired Persons

As we only consider individuals in working age (between 20 and 60), we do not take into account the proportional tax allowances for elderly non-retired persons (*Altersentlastungsbeiträge*).

1.H.1.4 Variation of Income during the year

The withholding tax is calculated by the employer every month and does not depend on the income of previous months. Consequently, if the income of an employee fluctuates between the months or a thirteenth or even fourteenth salary is paid out as christmas pay (*Weihnachtsgeld*) or holiday pay (*Urlaubsgeld*), the withholding tax liability is higher than in cases of stable income due to tax progressivity. For individuals with the symmetric withholding tax schedule, employers with more than 10 employees have to offset the difference between the withholding tax withheld during the year and the expected income tax of the individual under the assumption that the couple is not filing together (*Lohnsteuerjahresausgleich*). This rule is in place as couples who have chosen the symmetric withholding tax schedule do not have to file taxes (Hauck and Wallossek, 2024) and thereby get compensated if they were overwithheld due to income variation within the year. For individuals in the men- or women-favoring tax schedule, employers are not allowed to offset the overwithheld amount during the year.

1.H.1.5 Non-Observable Factors

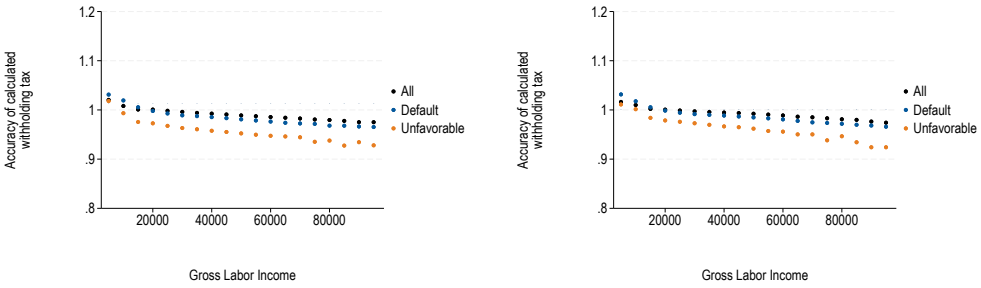
In case of additional income, in particular severance pay and death benefits (*Sterbegeld*), the withholding tax calculation differs. As these extraordinary sources of income are not separable from normal income in the tax data, we cannot implement the calculation. Individuals have under very special circumstances the possibility to decrease their withholding tax. First, the *Hinzurechungsbetrag* allows individuals with more than one job to minimize their withholding tax liability by pooling the withholding tax liability at one employer. Second, individuals who have regular deductions from taxes can request a deduction in the withholding tax (*Freibetrag*). This deduction typically applies to disabled people who have specific deductions from the income tax that they can also claim for the withholding tax. Unfortunately, in the data we do not observe any variables that are informative on whether individuals have filed a form to the financial authorities to use any of these tools to adapt their withholding tax.

1.H.2 Other Withheld Taxes and Social Security Contributions

Solidarity surcharge. The solidarity surcharge (*Solidaritätszuschlag*) is a surtax on the income tax that is subject to withholding. Different to the withholding tax, it also depends on the number of children. We calculate the solidarity surcharge and treat it as part of the withholding tax.

Church taxes. The German government collects income for the church. The church tax is a surtax on the income tax, typically between 8 % and 9 %. The church tax is collected as part of the withholding tax and also fully credited against the church tax liability when a household files income taxes. The church tax rate is a function of the religious denomination of the individual, the religious denomination of the spouse and the church parish the individual belongs to. As there is no comprehensive data set of church taxes, we approximate the withheld church tax empirically for each individual. For that, we divide for each individual and year the observed withheld church tax with the observed withholding tax and assume that the church tax rate from 2009 remains constant for the individual.

Social security contributions. In Germany, pensions, health care, and unemployment insurance are primarily financed by social security contributions which are a function of labor income. These social security contributions are withheld every month and credited against the final social security contributions at the beginning of the next year. So they might be perceived as withholding taxes. However, we decided to exclude social security contributions from our definition of withholding taxes, because they are not taxes and are therefore not informative about the misconception of withholding taxes.



(a) 2009 **(b)** 2010

Figure 1.H.1. Quality of Withholding Tax Estimation

Notes: The Figure displays the results of a regression of the calculated withholding tax on the actually withheld withholding tax for the women in our estimation sample. The assumptions to estimate the withholding tax are discussed in Section 1.H.1 and 1.H.2.

1.H.3 Quality of Withholding Tax Calculations

Figure 1.H.1 displays the quality of the withholding tax calculation along the income distribution for 2009 and 2010 by regressing the calculated withholding tax on the actually withheld withholding tax for the women in our estimation sample. Along the whole income distribution, the overall estimate is close to 1. Deviations mostly likely come from the fact that we do not observe whether individuals contribute to the private or the public health insurance. Also, the withholding taxes are less precisely estimated for women in the unfavorable withholding tax class, probably because for them the employer does not offset the within-year fluctuation in the withholding tax at the end of the year (see Section 1.H.1.4). But also for them, the estimate is always larger than 0.9.

Appendix 1.I Survey Questions

This section documents the survey questions. Section 1.I.1 includes the original questions in German. Depending on the answer to Question A1a, the gender of the interviewed, and A1b, the gender of the partner, the personal pronouns were adapted in all questions and explaining texts. Section 1.I.2 provides a translation into English.

1.I.1 German Version

Guten Tag!

Wir sind Forscher an den Universitäten Bonn und Göteborg und bedanken uns schon jetzt herzlich für Ihre Teilnahme an unserer Umfrage und Ihre damit verbundene Unterstützung unserer Forschung! Ihre Antworten in der Umfrage haben keine Auswirkung auf Ihre persönliche Auszahlung. Wir möchten Sie deshalb darum bitten, alle Fragen ohne Hilfsmittel (Internetrecherche, etc.) zu beantworten.

Wer ist verantwortlich für die Studie?

Kontaktdaten

Welchen Zwecken dient die Studie?

Zweck der Studie ist die Untersuchung ökonomischen Verhaltens. Wie bei ökonomischen Studien üblich, erfolgt daher vorab keine umfassende Aufklärung über den Forschungshintergrund.

Was geschieht mit meinen Daten?

Alle beteiligten Mitarbeiter und Wissenschaftler arbeiten selbstverständlich nach den Vorschriften der DatenschutzGrundverordnung, dem Bundesdatenschutzgesetz und den einschlägigen Landesdatenschutzgesetzen. Die Daten werden auf einem Server der Universität Bonn innerhalb der EU gespeichert. Ihre Daten werden nach erfolgter Auszahlung anonymisiert und anschließend statistisch ausgewertet. Aus den Ergebnissen lassen sich anschließend keine Rückschlüsse auf Sie ziehen.

Welche Rechte habe ich?

Sie haben das Recht, Auskunft über die zu Ihrer Person gespeicherten Daten zu erhalten (Art. 15 DS-GVO). Sollten unrichtige personenbezogene Daten verarbeitet werden, steht Ihnen ein Recht auf Berichtigung zu (Art. 16 DS-GVO). Liegen die gesetzlichen Voraussetzungen vor, so können Sie die Löschung oder Einschränkung der Verarbeitung verlangen sowie Widerspruch gegen die Verarbeitung einlegen (Art. 17, 18 und 21 DS-GVO). Sie haben das Recht, sich mit einer Beschwerde an die zuständige Aufsichtsbehörde für Datenschutz zu wenden. Die hier erklärte Einwilligung können Sie jederzeit mit Wirkung für die Zukunft widerrufen. Sofern Ihre Daten bereits anonymisiert wurden, können Ihnen diese aber nicht mehr zugeordnet werden. Wir können Ihre Angaben also nicht aus dem Ergebnis „herausrechnen“.

Einwilligungserklärung

Hiermit willige ich in die Verarbeitung meiner personenbezogenen Daten für das Forschungsvorhaben ein. Die Einwilligung kann ich jederzeit widerrufen. Ich habe die Hinweise zur Verwendung meiner Daten und zu meinen Rechten in der Datenschutzerklärung zur Kenntnis genommen.

Ich bin einverstanden. (Ja, Nein)

Page Break

Screening

S1 Haben Sie momentan Einkommen aus Lohnarbeit? (Ja, Nein)

S2 Sind Sie verheiratet? (Ja, Nein)

Page Break

A1a Was ist Ihr Geschlecht? (Weiblich, Männlich, Divers)

A1b Was ist das Geschlecht Ihres Ehepartners/Ihrer Ehepartnerin? (Weiblich, Männlich, Divers, Ich habe keinen Ehepartner/keine Ehepartnerin, Keine Angabe)

Page Break

A2 Die nächste Frage betrifft folgendes Problem: In Umfragen wie unserer gibt es manchmal Teilnehmerinnen und Teilnehmer, die die Fragestellungen nicht sorgfältig durchlesen, sondern sich nur schnell durch die Umfrage klicken. Dies führt zu vielen zufälligen Antworten, die die Qualität der Forschungsvorhaben beeinträchtigen. Bitte wählen Sie "Sehr stark interessiert" und "Überhaupt nicht interessiert" als Ihre Antwort auf die kommende Frage, um uns zu zeigen, dass Sie unsere Fragen sorgfältig lesen. Gegeben dieser Information, wie interessiert sind Sie am Thema Steuern?

(Überhaupt nicht interessiert, Fast gar nicht interessiert, Etwas interessiert, Stark interessiert, Sehr stark interessiert)

Page Break

A3 Stellen Sie sich vor, dass Ihr Arbeitgeber Ihnen eine freie Wahl Ihrer wöchentlichen Arbeitsstunden anbietet: Wie würden Sie sich entscheiden? (Ich würde meine Arbeitsstunden erhöhen, Ich würde meine Arbeitsstunden verringern, Ich würde meine Arbeitsstunden unverändert lassen, Weiß nicht)

Page Break

D4 Was ist Ihre momentane Lohnsteuerklasse? (1, 2, 3, 4, 4 mit Faktor, 5, 6, Weiß nicht)

D5 Wer hat über die Steuerklasse entschieden? (Ich, Mein Ehepartner, Mein Ehepartner und ich zusammen, Ein Steuerberater/Eine Steuerberaterin, Eine andere Person, Niemand, Weiß nicht)

Page Break

D_Text Wir wollen nun mehr über Ihr generelles Verständnis der Steuerklassen herausfinden, es geht also in den folgenden Fragen nicht um Ihre eigene Steuerklasse.

Page Break

D6 Existieren die folgenden Steuerklassenkombinationen (Ihr Ehepartner erstgenannt, Sie zweitgenannt)? (Ja, Nein, Weiß nicht)

(4/4, 5/4 , 3/5 , 5/5 , 4/1 , 3/3 , 4/5 , 5/3 , 1/4)

Wenn D4 == "4 mit Faktor":

(4/4, 5/4 , 3/5 , 5/5 , 4/1 , 3/3 , 4/5 , 5/3 , 1/4, 4 mit Faktor/3, 4 mit Faktor/4 mit Faktor, 3/4 mit Faktor, 5/4 mit Faktor, 4 mit Faktor/5)

Page Break

D7 Stellen Sie sich vor, dass Sie 60.000 € und Ihr Ehepartner 30.000 € brutto pro Jahr verdienen und dass Sie eine gemeinsame Steuererklärung machen. Bei welcher Steuerklassenkombination tragen Sie als Paar zusammen die geringste jährliche finale Steuerlast (entspricht der Einkommensteuer)? Alle drei genannten Steuerklassenkombinationen existieren.

(Ich in Steuerklasse 5 und mein Partner in Steuerklasse 3, Ich in Steuerklasse 4 und mein Partner in Steuerklasse 4, Ich in Steuerklasse 3 und mein Partner in Steuerklasse 5, Egal, Weiß nicht)

Page Break

D8 Nehmen Sie nun an, Sie wären in Steuerklasse 4. Was stimmt? Wenn Sie nun von 4 in 3 wechseln, dann bekommen Sie persönlich monatlich...

(...mehr netto von Ihrem Arbeitgeber, ...weniger netto von Ihrem Arbeitgeber, ...gleich viel netto von Ihrem Arbeitgeber, Weiß nicht)

Page Break

D9 Nehmen Sie nun an, Sie wären in Steuerklasse 4. Was stimmt? Wenn Sie nun von 4 in 5 wechseln, dann bekommen Sie persönlich monatlich...

(...mehr netto von Ihrem Arbeitgeber, ...weniger netto von Ihrem Arbeitgeber, ...gleich viel netto von Ihrem Arbeitgeber, Weiß nicht)

Page Break

D10 Bitte nehmen Sie sich ausreichend Zeit, um die folgende Information zu verstehen. In der Tabelle sehen Sie beispielhaft die Lohnsteuer abhängig von den Steuerklassen für ein Paar, bei dem beide Partner brutto 3500 € monatlich verdienen.

	Monatliche Lohnsteuer Partner A	Monatliche Lohnsteuer Partner B
Partner A in Steuerklasse 3 Partner B in Steuerklasse 5	350 €	1 000 €
Partner A in Steuerklasse 4 Partner B in Steuerklasse 4	700 €	700 €
Partner A in Steuerklasse 5 Partner B in Steuerklasse 3	1 000 €	350 €

Sie können sehen, dass die Wahl der Steuerklassen die zu zahlende Lohnsteuer stark beeinflusst. Sind beide Partner in der Steuerklasse 4, so zahlen beide Partner jeweils 700 € Lohnsteuern. Ist ein Partner stattdessen in Steuerklasse 3, so zahlt sie/er 350 € Lohnsteuern. In Steuerklasse 5 werden 1000 € Lohnsteuern fällig. Wie Sie sehen: Ihre individuell gezahlte Lohnsteuer hängt stark von der gewählten Steuerklasse ab. Aber auch die Lohnsteuer Ihres Partners wird stark durch die

Steuerklassenwahl beeinflusst. Waren Ihnen die folgenden Informationen schon bekannt? Bitte antworten Sie ehrlich. Denken Sie daran, dass Ihre Auszahlung in dieser Umfrage nicht von Ihren Antworten auf die Fragen abhängt. (Ja, Nein, Ich verstehe die Aussage nicht)

(Ich wusste, dass die Wahl der Steuerklasse die eigene Lohnsteuer beeinflusst, Ich wusste, dass die Wahl der Steuerklasse die Lohnsteuer meines Partners beeinflusst, Ich wusste, dass es Steuerklassenkombinationen gibt, bei der einer der beiden Partner deutlich mehr und der andere Partner deutlich weniger Lohnsteuern zahlt – selbst wenn beide Partner gleich viel verdienen)

Page Break

D11 Bitte nehmen Sie sich ausreichend Zeit, um auch die folgende Information zu verstehen. Die finale Steuerlast eines Paares wird durch die Einkommensteuer bestimmt. In der Tabelle können Sie sehen, dass Steuerklassen keine Auswirkungen auf die Einkommensteuer, und somit auf die finale Steuerlast eines Ehepaares, haben. Nur die Lohnsteuer wird durch die Steuerklassenwahl beeinflusst:

	Monatliche Lohnsteuer Partner A	Monatliche Lohnsteuer Partner B	Gemeinsame jährliche Einkommensteuerlast
Partner A in Steuerklasse 3 Partner B in Steuerklasse 5	350 €	1 000 €	**16 300 €**
Partner A in Steuerklasse 4 Partner B in Steuerklasse 4	700 €	700 €	**16 300 €**
Partner A in Steuerklasse 5 Partner B in Steuerklasse 3	1 000 €	350 €	**16 300 €**

Die monatlich von Ihnen als Paar gezahlte Lohnsteuer wird am Jahresende mit der Einkommensteuer verrechnet. Wenn also Ihre gezahlte Lohnsteuer höher ist als die zu zahlende Einkommensteuer, bekommen Sie am Jahresende eine Steuerrückzahlung. Und, andersherum, wenn Sie mehr Einkommensteuer zahlen müssen als Sie Lohnsteuer gezahlt haben, müssen Sie eine Steuernachzahlung leisten. Für das Paar in dem Beispiel bedeutet dies, dass es unabhängig von der gewählten Steuerklasse jährlich immer 16 300 € Einkommensteuern zahlt. Steuerklassen haben also keine Auswirkungen auf die finale Steuerlast eines Ehepaares, sondern nur auf die Lohnsteuer. Waren Ihnen die folgenden Informationen schon bekannt? Bitte antworten Sie ehrlich. Denken Sie daran, dass Ihre Auszahlung in dieser Umfrage nicht von Ihren Antworten auf die Fragen abhängt. (Ja, Nein, Ich verstehe die Aussage nicht)

(Ich wusste, dass die gezahlte Lohnsteuer nicht die finale Steuerlast beeinflusst, Ich wusste, dass die Steuerklassenwahl nicht die finale Steuerlast beeinflusst)

Page Break

D12 Stellen Sie sich vor, dass Sie 40.000 € und Ihr Ehepartner 70.000 € brutto pro Jahr verdienen und dass Sie eine gemeinsame Steuererklärung machen. Bei welcher Steuerklassenkombination tragen Sie als Paar zusammen die geringste jährliche finale Steuerlast (entspricht der Einkommensteuer)? Alle drei genannten Steuerklassenkombinationen existieren.

(Ich in Steuerklasse 5 und mein Partner in Steuerklasse 3, Ich in Steuerklasse 4 und mein Partner in Steuerklasse 4, Ich in Steuerklasse 3 und mein Partner in Steuerklasse 5, Egal, Weiß nicht)

Page Break

D13a Steuerklassen haben also keine Auswirkungen auf die finale Steuerlast eines Ehepaares, nur auf die Lohnsteuer. Stellen Sie sich mit diesem Wissen nun vor, dass Ihr Arbeitgeber Ihnen eine freie Wahl Ihrer wöchentlichen Arbeitsstunden anbietet: Wie würden Sie sich entscheiden?

(Ich würde meine Arbeitsstunden erhöhen, Ich würde meine Arbeitsstunden verringern, Ich würde meine Arbeitsstunden unverändert lassen, Weiß nicht)

D13b Steuerklassen haben keine Auswirkungen auf die finale Steuerlast eines Ehepaares, nur auf die Lohnsteuer. Stellen Sie sich mit diesem Wissen nun vor, dass Ihr Arbeitgeber Ihnen in der Vergangenheit eine freie Wahl Ihrer wöchentlichen Arbeitsstunden angeboten hätte. Wie hätten Sie sich entschieden?

(Ich hätte meine Arbeitsstunden erhöht, Ich hätte meine Arbeitsstunden verringert, Ich hätte meine Arbeitsstunden unverändert gelassen, Weiß nicht)

D13c Steuerklassen haben keine Auswirkungen auf die finale Steuerlast eines Ehepaares, nur auf die Lohnsteuer. Wie wirkt sich dieses Wissen auf Ihre bevorzugte Steuerklassenwahl aus?

(Ich würde meine Steuerklasse gerne ändern, Ich würde meine Steuerklasse gerne beibehalten, Weiß nicht)

D14 Beeinflussen Steuerklassen folgende staatliche Leistungen? (Ja, Nein, Weiß nicht)

(Rente, Arbeitslosengeld II/Hartz IV, Arbeitslosengeld I, Elterngeld, Wohngeld, Kurzarbeitergeld)

Page Break

D15 Die nächste Frage betrifft folgendes Problem: In Umfragen wie unserer gibt es manchmal Teilnehmerinnen und Teilnehmer, die die Fragestellungen nicht sorgfältig durchlesen, sondern sich nur schnell durch die Umfrage klicken. Dies führt zu vielen zufälligen Antworten, die die Qualität der Forschungsvorhaben beeinträchtigen. Bitte wählen Sie "Fast gar nicht interessiert" und "Stark interessiert" als Ihre Antwort auf die kommende Frage, um uns zu zeigen, dass Sie unsere Fragen sorgfältig lesen. Gegeben dieser Information, wie interessiert sind Sie am Thema Steuern?

(Überhaupt nicht interessiert, Fast gar nicht interessiert, Etwas interessiert, Stark interessiert, Sehr stark interessiert)

Page Break

D16a Haben Sie als Ehepaar ein gemeinsames Bankkonto? (Ja, Nein, Weiß nicht)

D16b Wohin überweist Ihr Arbeitgeber Ihren monatlichen Lohn? (Auf mein persönliches Bankkonto, Auf das Bankkonto meines Ehepartners, Auf ein Bankkonto, das ich mit meinem Ehepartner teile, Weiß nicht)

D16c Wohin überweist der Arbeitgeber Ihres Ehepartners den monatlichen Lohn? (Auf mein persönliches Bankkonto, Auf das Bankkonto meines Ehepartners, Auf ein Bankkonto, das ich mit meinem Ehepartner teile, Mein Ehepartner ist selbstständig oder arbeitet nicht, Weiß nicht)

Page Break

If D16a == Ja And D16b == Auf mein persönliches Bankkonto

D16d Wie viel Prozent Ihres monatlich von Ihrem Arbeitgeber überwiesenen Lohneinkommens transferieren Sie auf das gemeinsame Konto? (0 % - 20 %, 20 % - 40 %, 40 % - 60 %, 60 % - 80 %, 80 % - 100 %, Weiß nicht)

If D16a == Ja And D16b == Auf das Bankkonto meines Ehepartners

D16e Wie viel Prozent seines monatlich von seinem Arbeitgeber überwiesenen Lohneinkommens transferiert Ihr Ehepartner auf das gemeinsame Konto? (0 % - 20 %, 20 % - 40 %, 40 % - 60 %, 60 % - 80 %, 80 % - 100 %, Weiß nicht)

If D16a == Ja

D16f Haben Sie noch besondere Absprachen für Ihr gemeinsames Konto getroffen? Falls ja, erklären Sie bitte noch genauer, wie Sie Ihr gemeinsames Konto verwalten. Falls Sie keine besonderen Absprachen getroffen haben, lassen Sie das Freifeld gerne einfach frei.

Page Break

D17a Geben Sie und Ihr Partner üblicherweise eine Steuererklärung ab? (Ja. Mein Partner und ich veranlagen gemeinsam, Ja. Mein Partner und ich veranlagen getrennt, Ja. Aber ich weiß nicht, ob wir getrennt oder gemeinsam veranlagen, Nein, Weiß nicht)

Page Break

If D17a == Ja:

D17b Wie machen Sie und Ihr Partner üblicherweise Ihre Steuererklärung? Mehrere Ja-Antworten sind möglich. (Ja, Nein, Weiß nicht)

(Ich mache die Steuererklärung überwiegend alleine, Mein Ehepartner macht die Steuererklärung überwiegend alleine, Wir machen die Steuererklärung gemeinsam, Wir nutzen die Hilfe einer Steuerberaterin/eines Steuerberaters, Wir nutzen die Hilfe eines Steuerprogramms wie etwa WISO, Wir nutzen die Hilfe anderer Personen)

Page Break

If D17a == Ja:

D17c Auf welches Bankkonto werden potentielle Steuererstattungen überwiesen? (Mein Konto, Das Konto meines Ehepartners, Ein gemeinsames Konto, Weiß nicht)

Page Break

If D17a == Nein

D17d Warum geben Sie keine Steuererklärung ab? Mehrere Ja-Antworten sind möglich. (Ja, Nein) (Es ist mir zu viel Arbeit, Ich weiß nicht, wie man das macht, Es lohnt sich für mich kaum, Ich habe Angst, dass ich Steuern nachzahlen muss)

Page Break

D18 Auf einer Skala von 1 bis 7, wie sehr stimmen Sie den folgenden Aussagen zu? 7 bedeutet, dass Sie der entsprechenden Aussage voll zustimmen. 1 bedeutet, dass Sie der entsprechenden Aussage überhaupt nicht zustimmen. (1 Stimme überhaupt nicht zu, 2, 3, 4, 5, 6, 7 Stimme voll zu)

(Der Ehemann sollte zu Hause das letzte Wort haben., Am besten ist es, wenn der Ehemann und die Ehefrau beide gleich viel erwerbstätig sind und sich beide in gleichem Maße um Haushalt und Familie kümmern., Männer sollten sich stärker um die finanzielle Absicherung der Familie kümmern als Frauen.)

Page Break

D19 Wie alt sind Sie? (Jünger als 20, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-60, 61 oder älter)

D20 Was ist Ihr höchster schulischer/akademischer Bildungsabschluss? (Ohne allgemeinen Schulabschluss, Hauptschulabschluss, Mittlere Reife, Fachhochschul- oder Hochschulreife (Abitur), Bachelor, Master/Diplom/Staatsexamen, Promotion)

D21 Haben Sie mindestens ein minderjähriges Kind? (Ja, Nein, Keine Angabe)

Page Break

D22 Haben Sie häufiger das Gefühl, dass das Geld vor der Überweisung des nächsten Gehalts knapp wird? (Ja, Nein, Diese Frage möchte ich nicht beantworten)

Page Break

D23 Wie hoch ist Ihr Bruttoeinkommen aus Lohnarbeit pro Jahr? Für die Beantwortung dieser Frage können Sie gerne in Ihren Unterlagen nachschauen.

(Ich habe kein Lohneinkommen, 1 € - 10.000 €, 10.001 € - 20.000 €, 20.001 € - 30.000 €, 30.001 € - 40.000 €, 40.001 € - 50.000 €, 50.001 € - 60.000 €, 60.001 € - 70.000 €, 70.001 € - 80.000 €, 80.001 € - 90.000 €, 90.001 € - 100.000 €, 100.001 € - 110.000 €, 110.001 € - 120.000 €, Über 120.000 €, Weiß nicht / Keine Angabe)

D24a Wie hoch ist das Bruttoeinkommen Ihres Ehepartners aus Lohnarbeit pro Jahr? Für die Beantwortung dieser Frage können Sie gerne in Ihren Unterlagen nachschauen oder Ihren Ehepartner fragen.

(Mein Ehepartner arbeitet nicht, Mein Ehepartner ist selbstständig, 1 € - 10.000 €, 10.001 € - 20.000 €, 20.001 € - 30.000 €, 30.001 € - 40.000 €, 40.001 € - 50.000 €, 50.001 € - 60.000 €, 60.001 € - 70.000 €, 70.001 € - 80.000 €, 80.001 € - 90.000 €, 90.001 € - 100.000 €, 100.001 € - 110.000 €, 110.001 € - 120.000 €, Über 120.000 €, Weiß nicht / Keine Angabe)

If D24a == Mein Ehepartner ist selbstständig

D24b Wie viel verdient Ihr Ehepartner in selbstständiger Arbeit pro Jahr brutto? Für die Beantwortung dieser Frage können Sie gerne in Ihren Unterlagen nachschauen oder Ihren Ehepartner fragen.

(1 € - 10.000 €, 10.001 € - 20.000 €, 20.001 € - 30.000 €, 30.001 € - 40.000 €, 40.001 € - 50.000 €, 50.001 € - 60.000 €, 60.001 € - 70.000 €, 70.001 € - 80.000 €, 80.001 € - 90.000 €, 90.001 € - 100.000 €, 100.001 € - 110.000 €, 110.001 € - 120.000 €, Über 120.000 €, Weiß nicht / Keine Angabe)

Page Break

D25 Wie hoch ist Ihre durchschnittliche wöchentliche Arbeitszeit in Stunden?

D26 Wie hoch ist die durchschnittliche wöchentliche Arbeitszeit Ihres Ehepartners in Stunden?

Page Break

A27 Haben Sie irgendwelche Anmerkungen zur Umfrage oder zu dem Thema Lohnsteuerklassen?

1.I.2 English Version

Hello and welcome!

We are researchers at the Universities of Bonn and Gothenburg and would like to thank you in advance for taking part in our survey and for thereby supporting our research! Your responses to the survey will not affect your personal payout. We would therefore like to ask you to answer all questions without using any tools (internet research, etc.).

Who is responsible for the study?

Contact details

What is the purpose of the study?

The purpose of the study is to examine economic behavior. As is usual with economic studies, there is no comprehensive explanation of the research background beforehand.

What happens to my data?

Of course, all employees and scientists involved work in accordance with the provisions of the General Data Protection Regulation, the Federal Data Protection Act and the relevant state data protection laws. The data is stored on a server of the University of Bonn within the EU. Your data will be anonymized after the payment has been made and then statistically evaluated. No conclusions can be drawn about you from the results.

What rights do I have?

You have the right to receive information about the data stored about you (Art. 15 DS-GVO). If incorrect personal data is processed, you have the right to rectification (Art. 16 DS-GVO). If the legal requirements are met, you can request the deletion or restriction of processing and object to the processing (Art. 17, 18 and 21 DS-GVO). You have the right to lodge a complaint with the competent supervisory authority for data protection. You can revoke the consent given here at any time with effect for the future. However, if your data has already been anonymized, it can no longer be assigned to you. We can therefore not “remove" your information from the result.

Declaration of consent

I hereby consent to the processing of my personal data for the research project. I can revoke my consent at any time. I have taken note of the information on the use of my data and my rights in the data protection declaration.

I agree. (Yes, No)

Page break

Screening

S1 Do you currently have wage income? (Yes, No)

S2 Are you married? (Yes, No)

Page break

A1a What is your gender? (Female, Male, Diverse)

A1b What is the gender of your spouse? (Female, Male, Diverse, I have no spouse, No answer)

Page break

A2 The next question concerns the following problem: In surveys like ours, there are sometimes participants who do not read the questions carefully, but just click through the survey quickly. This leads to a lot of random answers, which affects the quality of the research projects. Please choose "Very interested" and "Not at all interested" as your answer to the upcoming question to show us that you are reading our questions carefully. Given this information, how interested are you in taxes?

(Not at all interested, Slightly interested, Somewhat interested, Interested, Very interested)

Page break

A3 Imagine that your employer offered you a free choice of your weekly working hours: How would you decide? (I would increase my hours, I would decrease my hours, I would keep my hours the same, Don't know)

Page break

D4 What is your current withholding tax class? (1, 2, 3, 4, 4 with factor, 5, 6, Don't know)

D5 Who decided on the withholding tax class? (Me, My Spouse, My Spouse and I Together, An Accountant, Another Person, Nobody, Don't Know)

Page break

E_Text We now want to find out more about your general understanding of withholding tax classes, so the following questions are not about your own withholding tax class.

Page break

D6 Do the following withholding tax class combinations exist (your spouse named first, you named second)? (Yes, No, Don't know)

(4/4, 5/4, 3/5, 5/5, 4/1, 3/3, 4/5, 5/3, 1/4)

If D4 == "4 with factor":

(4/4, 5/4 , 3/5 , 5/5 , 4/1 , 3/3 , 4/5 , 5/3 , 1/4, 4 with factor/3, 4 with factor/4 with factor , 3/4 with factor, 5/4 with factor, 4 with factor/5)

Page break

D7 Imagine that you earn €60,000 and your spouse €30,000 gross per year and that you file a joint tax return. In which withholding tax class combination do you as a couple have the lowest final annual tax liability (corresponds to income tax)? All three withholding tax class combinations mentioned exist.

(I in withholding tax class 5 and my partner in withholding tax class 3, I in withholding tax class 4 and my partner in withholding tax class 4, I in withholding tax class 3 and my partner in withholding tax class 5, Doesn't matter, Don't know)

Page break

D8 Now suppose you were in withholding tax class 4. Which is correct? If you now switch from 4 to 3, you will personally receive monthly...

(...more net from your employer, ...less net from your employer, ...same amount net from your employer, don't know)

Page break

D9 Now suppose you were in withholding tax class 4. Which is correct? If you now switch from 4 to 5, you will personally receive monthly...

(...more net from your employer, ...less net from your employer, ...same amount net from your employer, don't know)

Page break

D10 Please take enough time to understand the following information. The table shows an example of the withholding tax depending on the withholding tax classes for a couple where both partners earn a gross monthly income of €3,500.

	Monthly Withholding Tax Partner A	Monthly Withholding Tax Partner B
Partner A in Tax Class 3 Partner B in Tax Class 5	350 €	1 000 €
Partner A in Tax Class 4 Partner B in Tax Class 4	700 €	700 €
Partner A in Tax Class 5 Partner B in Tax Class 3	1 000 €	350 €

You can see that the choice of withholding tax class greatly affects the withholding tax you pay. If both partners are in withholding tax class 4, both partners each pay €700 in withholding tax. If a partner is in withholding tax class 3 instead, she/he pays €350 in withholding tax. In withholding tax class 5, €1,000 in withholding tax is due. As you can see, the withholding tax you pay depends heavily on the withholding tax class you choose. But your partner's withholding tax is also strongly influenced by the choice of withholding tax class. Did you already know the following information? Please answer honestly. Remember that your payout in this survey is not dependent on your answers to the questions. (Yes, No, I don't understand the statement)

(I knew that the choice of withholding tax class affects my own withholding tax, I knew that the choice of withholding tax class influences my partner's withholding tax, I knew that there are withholding tax class combinations where one of the two partners pays significantly more and the other partner significantly less pays withholding taxes – even if both partners earn the same amount)

Page break

D11 Please take enough time to understand the following information. The final tax liability of a couple is determined by the income tax. In the table you can see that withholding tax classes have no effect on the income tax and therefore on the final tax liability of a married couple. Only the withholding tax is affected by the withholding tax class selection:

	Monthly Withholding Tax Partner A	Monthly Withholding Tax Partner B	Joint Annual Income Tax Liability
Partner A in Tax Class 3 Partner B in Tax Class 5	350 €	1 000 €	**16 300 €**
Partner A in Tax Class 4 Partner B in Tax Class 4	700 €	700 €	**16 300 €**
Partner A in Tax Class 5 Partner B in Tax Class 3	1 000 €	350 €	**16 300 €**

The withholding tax you pay monthly as a couple is offset against the income tax at the end of the year. So if your paid withholding tax is higher than the income tax to be paid, you will receive a tax refund at the end of the year. And, vice versa, if you have to pay more income tax than you paid withholding tax, you have to make an additional tax payment. For the couple in the example, this means that they always pay €16,300 in income tax annually, regardless of the withholding tax class they choose. Withholding tax classes therefore have no effect on the final tax liability of a married couple, but only on the withholding tax. Did you already know the following information? Please answer honestly. Remember that your payout in this survey is not dependent on your answers to the questions. (Yes, No, I don't understand the statement)

(I knew that the withholding tax paid does not affect the final tax liability, I knew that the choice of withholding tax classes does not affect the final tax liability)

Page break

D12 Imagine that you earn €40,000 and your spouse €70,000 gross per year and that you file a joint tax return. In which withholding tax class combination do you as a couple have the lowest final annual tax liability (corresponds to income tax)? All three withholding tax class combinations mentioned exist.

(I in withholding tax class 5 and my partner in withholding tax class 3, I in withholding tax class 4 and my partner in withholding tax class 4, I in withholding tax class 3 and my partner in withholding tax class 5, Doesn't matter, Don't know)

Page break

D13a Withholding tax classes therefore have no effect on the final tax liability of a married couple, only on the withholding tax. Now, knowing this, imagine that your employer offered you a free choice of your weekly working hours: How would you decide?

(I would increase my hours, I would decrease my hours, I would keep my hours the same, Don't know)

D13b Withholding tax classes have no effect on the final tax liability of a married couple, only on the withholding tax. Now, knowing this, imagine that in the past your employer would have offered you a free choice of your weekly work hours. How would you have decided?

(I would have increased my hours, I would have decreased my hours, I would have left my hours unchanged, Don't know)

D13c Withholding tax classes have no effect on a married couple's final tax liability, only on the withholding tax. How does this knowledge affect your preferred withholding tax class choice?

(I would like to change my withholding tax class, I would like to keep my withholding tax class, Don't know)

D14 Do withholding tax classes affect the following government benefits? (Yes, No, Don't know)

(Pension, unemployment benefit II/Hartz IV, unemployment benefit I, parental benefit, housing benefit, short-time work benefit)

Page break

D15 The next question concerns the following problem: In surveys like ours, there are sometimes participants who do not read the questions carefully, but just click through the survey quickly. This leads to a lot of random answers, which affects the quality of the research projects. Please choose "Slightly interested" and "Very interested" as your answer to the next question to show us that you are reading our questions carefully. Given this information, how interested are you in taxes?

(Not at all interested, Slightly interested, Somewhat interested, Interested, Very interested)

Page break

D16a As a married couple, do you have a joint bank account? (yes, no, don't know)

D16b Where does your employer transfer your monthly wages to? (To my personal bank account, To my spouse's bank account, To a bank account I share with my spouse, Don't know)

D16c Where does your spouse's employer transfer the monthly salary to? (To my personal bank account, To my spouse's bank account, To a bank account I share with my spouse, My spouse is self-employed or does not work, Don't know)

Page break

If D16a == Yes And D16b == To my personal bank account

D16d What percentage of your monthly wage income transferred from your employer do you transfer to the joint account? (0% - 20%, 20% - 40%, 40% - 60%, 60% - 80%, 80% - 100%, Don't know)

If D16a == Yes And D16b == To my spouse's bank account

D16e What percentage of his/her monthly wages transferred from his/her employer does your spouse transfer to the joint account? (0% - 20%, 20% - 40%, 40% - 60%, 60% - 80%, 80% - 100%, Don't know)

If D16a == Yes

D16f Have you made any special arrangements for your joint account? If so, please explain in more detail how you manage your joint account. If you have not made any special arrangements, please feel free to leave the free field empty.

Page break

D17a Do you and your partner usually file a tax return? (Yes. My partner and I file taxes jointly, Yes. My partner and I file taxes separately, Yes. But I don't know if we file our taxes separately or jointly, No, Don't know)

Page break

If D17a == Yes:

D17b How do you and your partner usually file your tax return? Several yes answers are possible. (yes, no, don't know)

(I mostly file the tax return alone, my spouse mostly files the tax return alone, we file the tax return together, we use the help of a tax consultant, we use the help of a tax program such as WISO, we use the help of other people)

Page break

If D17a == Yes:

D17c To which bank account are potential tax refunds transferred? (My Account, My Spouse's Account, A Joint Account, Don't Know)

Page break

If D17a == No

D17d Why don't you file a tax return? Several yes answers are possible. (Yes, No) (It's too much work for me, I don't know how to do it, It's hardly worth it for me, I'm afraid I'll have to pay more taxes)

Page break

D18 On a scale from 1 to 7, how much do you agree with the following statements? 7 means that you fully agree with the corresponding statement. 1 means that you completely disagree with the corresponding statement. (1 Strongly Disagree, 2, 3, 4, 5, 6, 7 Strongly Agree)

(The husband should have the last word at home., It is best if the husband and wife both work an equal amount and both take care of the household and family equally., Men should take more care of the financial security of the family than women.)

Page break

D19 How old are you? (Under 20, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-60, 61 or older)

D20 What is your highest school/academic qualification? (Without general school leaving certificate, secondary school leaving certificate, higher secondary school leaving certificate or higher education entrance qualification (Abitur), bachelor, master/diploma/state examination, doctorate)

D21 Do you have at least one minor child? (Yes, No, Not specified)

Page break

D22 Do you often have the feeling that money is running out before you receive your next salary? (Yes, No, I don't want to answer this question)

Page break

D23 What is your gross income from wage labor per year? You are welcome to consult your documents to answer this question.

(I have no wage income, €1 - €10,000, €10,001 - €20,000, €20,001 - €30,000, €30,001 - €40,000, €40,001 - €50,000, €50,001 - €60,000, €60,001 - €70,000, - €80,000, €80,001 - €90,000, €90,001 - €100,000, €100,001 - €110,000, €110,001 - €120,000, over €120,000, don't know / no answer)

D24a What is your spouse's gross income from wage labor per year? To answer this question, you are welcome to consult your records or ask your spouse.

(My spouse does not work, My spouse is self-employed, €1 - €10,000, €10,001 - €20,000, €20,001 - €30,000, €30,001 - €40,000, €40,001 - €50,000, €50,001 - €60,000, €60,001 - €70,000, €70,001 - €80,000, €80,001 - €90,000, €90,001 - €100,000, €100,001 - €110,000, €110,001 - €120,000, over €120,000, don't know / no answer)

If D24a == My spouse is self-employed

D24b How much does your spouse earn gross per year in self-employment? To answer this question, you are welcome to consult your records or ask your spouse.

(€1 - €10,000, €10,001 - €20,000, €20,001 - €30,000, €30,001 - €40,000, €40,001 - €50,000, €50,001 - €60,000, €60,001 - €70,000, €70,001 - €70,001.1 € - 90,000 €, €90,001 - €100,000, €100,001 - €110,000, €110,001 - €120,000, Over €120,000, Don't know / no answer)

Page break

D25 What are your average weekly working hours?

D26 What are the average weekly working hours of your spouse?

Page break

D27 Do you have any comments on the survey or on the subject of withholding tax classes?

References

Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7 (3): 1–28. https://doi.org/10.1257/pol.20130137. [9]

Aghion, Philippe, Ufuk Akcigit, Matthieu Lequien, and Stefanie Stantcheva. 2017. "Tax Simplicity and Heterogeneous Learning." National Bureau of Economic Research. [5, 9]

Almås, Ingvild, Orazio Attanasio, and Pedro Carneiro. 2023. "Household Decisions and Intra-Household Distributions." In *Handbook of the Economics of the Family,* 1: 111–49. Elsevier. https://doi.org/10.1016/bs.hefam.2023.01.008. [33]

Bagchi, Sutirtha. 2025. "The Effects of Introducing Withholding on Tax Compliance: Evidence from Pennsylvania's Local Earned Income Tax." *National Tax Journal* 78 (2): 327–67. https://doi.org/10.1086/733662. [6]

Bagchi, Sutirtha, and Libor Dušek. 2021. "The Effects of Introducing Withholding and Third-Party Reporting on Tax Collections: Evidence from the U.S. State Personal Income Tax." *Journal of Public Economics* 204. https://doi.org/10.1016/j.jpubeco.2021.104537. [6]

Becker, Johannes, Jonas Fooken, and Melanie Steinhoff. 2019. "Behavioral Effects of Withholding Taxes on Labor Supply." *Scandinavian Journal of Economics* 121 (4): 1417–40. https://doi.org/10.1111/sjoe.12345. [9]

Bernheim, B. Douglas, and Dmitry Taubinsky. 2018. "Behavioral Public Economics." *Handbook of behavioral economics: Applications and Foundations 1* 1: 381–516. [9]

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households*." *Quarterly Journal of Economics* 130 (2): 571–614. https://doi.org/10.1093/qje/qjv001. [58]

Bick, Alexander, and Nicola Fuchs-Schündeln. 2017. "Quantifying the Disincentive Effects of Joint Taxation on Married Women's Labor Supply." *American Economic Review* 107 (5): 100–104. https://doi.org/10.1257/aer.p20171063. [10, 12, 42]

Bick, Alexander, and Nicola Fuchs-Schündeln. 2018. "Taxation and Labour Supply of Married Couples across Countries: A Macroeconomic Analysis." *Review of Economic Studies* 85 (3): 1543–76. https://doi.org/10.1093/restud/rdx057. [10]

Blaufus, Kay, Jonathan Bob, Jochen Hundsdoerfer, Christian Sielaff, Dirk Kiesewetter, and Joachim Weimann. 2015. "Perception of Income Tax Rates: Evidence from Germany." *European Journal of Law and Economics* 40 (3): 457–78. https://doi.org/10.1007/s10657-013-9389-9. [5, 9]

Buettner, Thiess, Katharina Erbe, and Veronika Grimm. 2019. "Tax Planning of Married Couples and Intra-Household Income Inequality." *Journal of Public Economics* 179: 1040–48. https://doi.org/10.1016/j.jpubeco.2019.104048. [21, 58]

Caldwell, Sydnee, Scott Nelson, and Daniel Waldinger. 2023. "Tax Refund Uncertainty: Evidence and Welfare Implications." *American Economic Journal: Applied Economics* 15 (2): 352–76. https://doi.org/10.1257/app.20210383. [10]

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro H. Sant'Anna. 2024. "Difference-in-Differences with a Continuous Treatment." National Bureau of Economic Research. https://doi.org/10.3386/w32117. [24]

Carbonnier, Clément, Clément Malgouyres, Loriane Py, and Camille Urvoy. 2022. "Who Benefits from Tax Incentives? The Heterogeneous Wage Incidence of a Tax Credit." *Journal of Public Economics* 206: 1045–77. https://doi.org/10.1016/j.jpubeco.2021.104577. [49]

Chetty, Raj, John N. Friedman, and Emmanuel Saez. 2013. "Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings." *American Economic Review* 103 (7): 2683–721. JSTOR: 42920668. [5, 9, 33]

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax Compliance and Loss Aversion." *American Economic Journal: Economic Policy* 7 (4): 132–64. https://doi.org/10.1257/pol.20130134. [6, 10]

Falk, Armin, Anke Becker, Thomas Dohmen, Benjamin Enke, David Huffman, and Uwe Sunde. 2018. "Global Evidence on Economic Preferences*." *Quarterly Journal of Economics* 133 (4): 1645–92. https://doi.org/10.1093/qje/qjy013. [8, 33]

Farhi, Emmanuel, and Xavier Gabaix. 2020. "Optimal Taxation with Behavioral Agents." *American Economic Review* 110 (1): 298–336. [5, 9]

Feldman, Naomi E. 2010. "Mental Accounting Effects of Income Tax Shifting." *Review of Economics and Statistics* 92 (1): 70–86. https://doi.org/10.1162/rest.2009.11892. [9]

Feldman, Naomi E., Peter Katuščák, and Laura Kawano. 2016. "Taxpayer Confusion: Evidence from the Child Tax Credit." *American Economic Review* 106 (3): 807–35. https://doi.org/10.1257/aer.20131189. [9, 44]

Fujii, Edwin T., and Clifford B. Hawley. 1988. "On the Accuracy of Tax Perceptions." *Review of Economics and Statistics* 70 (2): 344–47. https://doi.org/10.2307/1928321. JSTOR: 1928321. [5]

Gabaix, Xavier. 2019. "Behavioral Inattention." In *Handbook of Behavioral Economics: Applications and Foundations 1,* 2: 261–343. Elsevier. [9]

Gelman, Michael, Shachar Kariv, Matthew D. Shapiro, and Dan Silverman. 2022. "Rational Illiquidity and Consumption: Theory and Evidence from Income Tax Withholding and Refunds." *American Economic Review* 112 (9): 2959–91. https://doi.org/10.1257/aer.20191385. [6, 10, 38]

Gideon, Michael. 2017. "Do Individuals Perceive Income Tax Rates Correctly?" *Public Finance Review* 45 (1): 97–117. https://doi.org/10.1177/1091142115615670. [5, 9, 33]

Gudgeon, Matthew, and Simon Trenkle. 2024. "The Speed of Earnings Responses to Taxation and the Role of Firm Labor Demand." *Journal of Labor Economics,* 723831. https://doi.org/10.1086/723831. [28]

Hauck, Tobias, and Luisa Wallossek. 2024. "Optional (Non-)Filing and Effective Taxation." *Journal of Public Economics.* [6, 66]

Herold, Elena, and Luisa Wallossek. 2024. "The Marriage Earnings Gap." *Working Paper.* [10]

Illing, Hannah, Fabian Reutzel, and Jakob Wegmann. 2024. "Gendered Unemployment Benefits: Withholding Taxes and Their Propagated Behavioral Impact." *Mimeo.* [15]

Jakobsen, Katrine Marie, and Jakob Egholt Søgaard. 2022. "Identifying Behavioral Responses to Tax Reforms: New Insights and a New Approach." *Journal of Public Economics* 212: 104691. https://doi.org/10.1016/j.jpubeco.2022.104691. [22, 23, 54]

Jones, Damon. 2012. "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds." *American Economic Journal: Economic Policy* 4 (1): 158–85. https://doi.org/10.1257/pol.4.1.158. [9, 38]

Keane, Michael P. 2011. "Labor Supply and Taxes: A Survey." *Journal of Economic Literature* 49 (4): 961–1075. https://doi.org/10.1257/jel.49.4.961. [30]

2019. *Kleine Anfrage Drucksache 19/12857: Ehegattensplitting Und Gleichstellung Im Deutschen Steuersystem.* [12]

2021. *Koalitionsvertrag Zwischen SPD, Bündnis 90/Die Grünen, FDP.* [39]

Koch, Leonie. 2024. "Income Taxation of Couples and Gender (In)Equality in Earnings." *Working Paper.* [9]

Kostøl, Andreas R., and Andreas S. Myhre. 2021. "Labor Supply Responses to Learning the Tax and Benefit Schedule." *American Economic Review* 111 (11): 3733–66. https://doi.org/10.1257/aer.20201877. [9]

LaLumia, Sara. 2008. "The Effects of Joint Taxation of Married Couples on Labor Supply and Non-Wage Income." *Journal of Public Economics* 92 (7): 1698–719. https://doi.org/10.1016/j.jpubeco.2008.01.009. [10]

Lardeux, Raphaël. 2023. "Behavioral Cross-Influence of a Shadow Tax Bracket: Evidence from Bunching Where Income Tax Liabilities Start." *Journal of Public Economics* 226: 104950. https://doi.org/10.1016/j.jpubeco.2023.104950. [5, 9]

Leite, David. 2024. "Tax Withholding and Labor Supply - Evidence from Bunching at the Withholding Eligibility Threshold." *Working Paper.* [5, 9]

Liebman, Jeffrey B., and Richard J. Zeckhauser. 2004. "Schmeduling." *Working Paper.* [5, 9, 44]

McCaffery, Edward, and Joel Slemrod. 2006. *Behavioral Public Finance.* Russell Sage Foundation. [9]

Messacar, Derek. 2018. "The Effect of Tax Withholding on Pre-Retirement Savings Withdrawals: Evidence from Canada." *Journal of Pension Economics and Finance* 17 (4): 534–53. https://doi.org/10.1017/S1474747217000269. [9]

Moore, Dylan T., and Joel Slemrod. 2021. "Optimal Tax Systems with Endogenous Behavioral Biases." *Journal of Public Economics* 197: 104384. https://doi.org/10.1016/j.jpubeco.2021.104384. [9]

Neisser, Carina. 2021. "The Elasticity of Taxable Income: A Meta-Regression Analysis." *Economic Journal* 131 (640): 3365–91. https://doi.org/10.1093/ej/ueab038. [7, 27, 43, 44]

O'Donoghue, Ted, and Matthew Rabin. 1999. "Doing It Now or Later." *American Economic Review* 89 (1): 103–24. https://doi.org/10.1257/aer.89.1.103. [33]

RDC of the Federal Statistical Office and Statistical Offices of the Federal States. 2010. *Lohn- und Einkommenssteuerstatistik 2010, Scientific Use File.* https://doi.org/10.21242/73111.2010.00.00.3.1.0. [52, 63, 64]

RDC of the Federal Statistical Office and Statistical Offices of the Federal States. 2022. *Taxpayer-Panel 2001-2018.* https://doi.org/10.21242/73111.2018.00.01.2.1.1. [17]

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–78. https://doi.org/10.1093/restud/rdx038. [5, 6, 9, 10]

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *Review of Economic Studies* 87 (5): 2399–438. https://doi.org/10.1093/restud/rdz045. [5, 9, 33, 44]

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. https://doi.org/10.1257/jel.50.1.3. [22]

Schepanski, Albert, and Teri Shearer. 1995. "A Prospect Theory Account of the Income Tax Withholding Phenomenon." *Organizational Behavior and Human Decision Processes* 63 (2): 174–86. https://doi.org/10.1006/obhd.1995.1071. [5, 10]

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. https://doi.org/10.1007/s10797-013-9283-y. [10]

Shapiro, Matthew D., and Joel Slemrod. 1995. "Consumer Response to the Timing of Income: Evidence from a Change in Tax Withholding." *American Economic Review* 85 (1): 274–83. JSTOR: 2118010. [9, 27]

Slemrod, Joel. 2019. "Tax Compliance and Enforcement." *Journal of Economic Literature* 57 (4): 904–54. https://doi.org/10.1257/jel.20181437. [5]

Spangenberg, Ulrike, Giesela Färber, and Corinna Späth. 2020. "Mittelbare Diskriminierung im Lohnsteuerverfahren. Auswirkungen der Lohnsteuerklassen auf Nettoeinkommen und Lohnersatzleistungen." In *Working Paper Forschungsförderung,* vol. 190. Hans-Böckler-Stiftung. [13, 15]

Stantcheva, Stefanie. 2021. "Understanding Tax Policy: How Do People Reason?" *Quarterly Journal of Economics* 136 (4): 2309–69. https://doi.org/10.1093/qje/qjab033. [5, 9, 44]

Stöwhase, Sven. 2011a. "Bürgerentlastungsgesetz Krankenversicherung: Ambivalente Effekte auf die Lohnsteuer." *Zeitschrift für Wirtschaftspolitik* 60 (1): 3–23. https://doi.org/10.1515/zfwp-2011-0102. [16]

Stöwhase, Sven. 2011b. "Non-Minimization of Source Taxes on Labor Income: Empirical Evidence from Germany." *Review of Economics of the Household* 9 (2): 293–306. https://doi.org/10.1007/s11150-009-9077-2. [58]

Thaler, Richard H. 1994. "Psychology and Savings Policies." *American Economic Review* 84 (2): 186–92. [10]

Weber, Caroline E. 2014. "Toward Obtaining a Consistent Estimate of the Elasticity of Taxable Income Using Difference-in-Differences." *Journal of Public Economics* 117: 90–103. https://doi.org/10.1016/j.jpubeco.2014.05.004. [23]

Chapter 2

Who cares? Parental Leave Benefits and Household Division of Childcare*

Joint with Federica Meluzzi and Pia Molitor

2.1 Introduction

After decades of progress in women's labor market outcomes—driven by structural changes and rising educational attainment—convergence in the economic trajectories of men and women has slowed markedly since the late 1990s (Blau and Kahn, 2017; Kuziemko et al., 2018; Olivetti, Petrongolo, and Pan, 2025). A central driver of the remaining gender gap is the asymmetric impact of parenthood, often referred to as the *child penalty* (Angelov, Johansson, and Lindahl, 2016; Lundborg, Plug, and Rasmussen, 2017; Kleven et al., 2019; Cortes and Pan, 2023), which has proven remarkably persistent (Kleven, 2024). At the root of this penalty lies an unequal division of domestic responsibilities: fathers' participation in childcare and parental leave remains limited, shaped by slow-moving social norms and economic incentives.

Family policies—especially parental leave—are often viewed as promising tools to reduce gender inequality. Yet, despite broad reforms across high-income countries, evidence on whether and how these policies foster a more equitable division of labor within households remains limited. A growing body of micro evidence suggests that existing parental leave reforms have yielded only modest effects on gender convergence in labor market outcomes and childcare responsibilities (Olivetti

* For very helpful comments, we would like to thank Sydnee Caldwell, David Card, Stefano Dellavigna, Jonas Jessen, Patrick Kline, Rafael Lalive, Emmanuel Saez, Arne Uhlendorff, as well as seminar participants at IZA and the University of Zurich. We thank the Statistical Office in Germany for the cooperation, in particular Laura Ehrmantraut, Alexander Seidel and Stefan Küpper. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A05) is gratefully acknowledged.

and Petrongolo, 2017). For instance, expansions in leave duration have had limited or even adverse effects on women's long-term labor market outcomes (Lalive and Zweimüller, 2009; Schönberg and Ludsteck, 2014; Dahl et al., 2016; Waights, 2022; Kleven et al., 2024; Bailey et al., 2025; Glogowky et al., 2025). Similarly, the introduction of "daddy quotas"—non-transferable leave earmarked for fathers—has often produced only modest and transitory effects on paternal involvement (Ekberg, Eriksson, and Friebel, 2013; Cools, Fiva, and Kirkebøen, 2015).

In this paper, we draw attention to a distinct and understudied aspect of parental leave design: the monetary generosity of parental leave benefits. As with unemployment insurance, a central policy lever is the replacement rate - that is, the proportion of previous earnings replaced by parental leave benefits. Despite its salience, little is known about how benefit generosity shapes intra-household childcare allocation or labor supply decisions, nor about the optimal design of replacement rates. We ask whether increasing the economic cost of adhering to traditional gender norms - by altering relative replacement rates between partners - can shift behavior within households. Specifically: how would fathers' leave-taking respond if they received more generous replacement rates than mothers?

Our empirical strategy exploits a sharp kink in the German parental leave benefit schedule, where benefits replace 65% of individual pre-birth income up to a monthly cap of €1,800. Assuming that parents cannot precisely manipulate their earnings around the kink, and under a set of formally defined smoothness conditions, this institutional feature generates plausibly exogenous variation in both an individual's own replacement rate and that of their partner, which we rely on for identification. In Germany, following mandatory maternity leave, parents are entitled to up to 12 months of paid, job-protected parental leave—or up to 14 months if both partners share the leave—and may freely allocate this time between themselves. In this paper, we assess how financial incentives shape the intra-household division of this parental leave.

We introduce a methodological innovation: a double regression kink design (double RKD), extending the framework of Card et al. (2015), to estimate how each parent responds to changes in their own and their partner's incentives. This is the first application of the RKD method to study household-level behavioral responses. Our analysis leverages novel administrative microdata on the universe of parental leave recipients in Germany (*Elterngeldstatistik*), which include precise information on benefit amounts, leave duration, labor income, and background characteristics (socio-demographics, religious affiliation, insurance status, and job characteristics)—and, crucially, allow us to link parents within households. The richness and scale of these data—over one million households near the kink between 2014 and 2018—allow us to generate highly precise estimates and explore heterogeneous effects across key dimensions.

We present a set of novel findings. First, households adjust the division of parental leave in response to financial incentives. A reduction in the mother's ben-

efits shortens her leave and increases the father's leave duration—reflecting substitution in parental inputs. The father's response operates on the intensive margin (more months among those already taking leave), with no effect on take-up rates. Total parental leave duration at the household level remains unchanged, suggesting maternal incentives shift allocation rather than overall childcare time. While both parents respond to changes in maternal incentives, the magnitude of these responses is modest: a 10% reduction in the mother's replacement rate leads on average to a 1% decline in her leave duration and an 8.3% increase in the father's duration (relative to baseline durations).

Second, we investigate whether the modest average responsiveness to maternal financial incentives masks underlying heterogeneity driven by differences in paternal opportunity costs. We uncover substantial heterogeneity: households in which fathers earn below the benefit cap respond nearly twice as strongly to changes in mothers' incentives compared to those where fathers earn above the cap. This pattern is consistent with the role of foregone earnings in shaping the willingness of fathers to substitute into childcare. Third, cultural context matters: cross-spousal substitution is at least twice as large among couples in East Germany, where gender norms are historically more egalitarian (Boelmann, Raute, and Schönberg, 2025). These findings highlight the importance of both economic and cultural constraints in mediating the effects of parental leave policies.

Fourth, replacement rates influence not only leave allocation but also labor market activity during leave. In Germany, parents may work up to 30 hours weekly while receiving benefits. We find that lower maternal replacement rates increase mothers' labor earnings during leave, driven by more months worked and likely longer hours—suggesting substitution from full-time caregiving toward part-time employment. Fathers show similar patterns: as their leave duration increases in response to reduced maternal benefits, they too show a greater likelihood of working part-time during leave. This reveals a novel margin through which policy can affect labor market attachment.

Our paper contributes to the growing literature on family policies and gender equality (see Olivetti and Petrongolo (2017), Canaan et al. (2022) for reviews). Prior work has largely focused on leave duration extensions, often finding limited or negative impacts on women's long-term outcomes (Lalive and Zweimüller, 2009; Schönberg and Ludsteck, 2014; Dahl et al., 2016; Canaan et al., 2022; Kleven et al., 2024; Bailey et al., 2025; Glogowky et al., 2025), or on paternity leave expansions, which have increased fathers' take-up but have had modest effects on gender gaps (Ekberg, Eriksson, and Friebel, 2013; Cools, Fiva, and Kirkebøen, 2015; Avdic and Karimi, 2018; Gonzalez and Farré, 2019; Andresen and Nix, 2024).

More closely related to our study is a smaller literature on the role of benefit generosity—particularly replacement rates—in shaping parental leave behavior. Asai (2015) and Bana, Bedard, and Rossin-Slater (2020) find no significant effects of increased benefits on maternity leave duration in Japan and California, respec-

tively, while Ziegler and Bamieh (2023) document limited effects of higher replacement rates on fathers' leave-taking in Austria. However, these studies typically focus on individual responses to own incentives, largely due to data limitations, especially the inability to link partners in administrative records. We contribute to this literature along four key dimensions. First, to our knowledge, we provide the first causal evidence on how replacement rates influence the intra-household allocation of parental leave. Exploiting a sharp kink in the German benefit schedule, our empirical strategy identifies cross-spousal behavioral responses, demonstrating that the division of leave is sensitive to relative financial incentives within the household. These findings complement the work of Ichino et al. (2024), who study how economic incentives affect the division of home production by leveraging variation in relative tax rates between spouses in Sweden. Second, we show that behavioral responses are shaped by both economic factors (e.g., earnings differentials) and cultural context (e.g., regional gender norms). Third, we document a novel form of adjustment: mothers respond to reduced benefits not only by shortening leave, but also by shifting from full-time caregiving to part-time employment during leave—highlighting that benefit design affects both who provides care and how care and work are combined. Building on these findings, we will use the causal estimates of spousal responses to benefit generosity to calibrate a household decision model. This will allow us to conduct counterfactual policy simulations that evaluate how alternative designs of (gender-specific) replacement rates may influence the division of parental leave and reduce long-term child penalties. In this respect, our work is complementary to recent research by Jørgensen and Søgaard (2024), which combines quasi-experimental variation in the Danish parental leave system with a sufficient statistics approach to assess the optimal design of earmarked leave aimed at increasing fathers' participation.

The remainder of the paper is organized as follows. Section 2.2 outlines the institutional background of parental leave and childcare policy in Germany. Section 2.3 introduces the data and presents summary statistics for the analysis sample. Section 2.4 presents the empirical strategy and discusses its validity. Section 2.5 presents baseline RKD estimates of how benefit generosity affects parental leave decisions. Section 2.6 explores heterogeneity in treatment effects by household income composition and regional gender norms. Section 2.7 reports preliminary estimates from the double RKD. Section 2.8 investigates underlying mechanisms. Section 2.9 concludes and outlines directions for future research.

2.2 The Parental Leave System in Germany

In Germany, government-provided paid parental leave has a long-standing tradition, dating back to the 1950s. A major reform in 2007 introduced earnings-related,

universal parental benefits (*Elterngeld*), offering up to 14 months of paid leave per household.[1]

More specifically, mothers are entitled to 14 weeks of mandatory maternity leave at full wage replacement—six weeks before and eight weeks after childbirth. This period is followed by up to 14 months of parental leave benefits, which may be claimed by either parent. These benefits can be freely divided between parents; however, if only one parent takes leave, the maximum duration is capped at 12 months. Unlike in other countries, Germany has no mandatory paternity leave. The minimum benefit-claimable leave duration is two months. Thus, each parent's choice set effectively consists of (i) no leave, (ii) exactly two months of paid leave, or (iii) more than two months, up to a maximum of 12 months. Twin births do not entitle parents to additional leave. Benefits must be used without interruption within 12 for 14 months following childbirth, and deferral is not permitted. However, parents have considerable flexibility in how they use the leave: it may be taken sequentially or concurrently, and parents are allowed to work up to 30 hours per week while receiving benefits. Also, parents can prolong the period of parental leave by working part-time (see next Section for more details).

2.2.1 Parental Leave Compensation

While on parental leave, individuals receive compensation proportional to their individual average net labor income over the 12 months preceding childbirth (the reference income).[2] For most recipients, earnings are replaced at a flat rate of 65%, up to a maximum benefit of €1,800 per month.[3] This ceiling corresponds to a reference income of €2,769, which is above the sample mean of €1,743.31. For individuals with a reference income below €1,200, the replacement rate increases to 67%. The minimum benefit is set at €300 per month, which also applies to individuals who were not employed prior to childbirth. The reference income used for benefit determination includes positive labor income and/or self-employment income, net of standard deductions. These deductions include monthly withholding income tax, solidarity surcharge, church tax (if applicable), and employee contributions to social security. Figure 2.1 illustrates the kinked structure of the parental leave benefit schedule as a function of net monthly income in the year preceding childbirth.

1. Note that parental leave (*Elternzeit*) and parental leave benefits (*Elterngeld*) are distinct components of Germany's institutional framework for new parents. Parental leave grants the right to unpaid, job-protected time off for childcare—typically up to three years per child. Parental leave benefits are wage replacement payments compensating for income loss when parents reduce or stop working to care for a child during the first 12 to 14 months after childbirth. This paper focuses on parental leave benefits; all references to parental leave henceforth imply the receipt of benefits.

2. The net labor income is calculated based on gross income, the withholding tax class, church affiliation and contribution to the social security system. For more details see Section 2.B.

3. Between 2014 and 2018, it was not common practice for employers in Germany to provide top-up payments to supplement government-paid parental leave benefits.

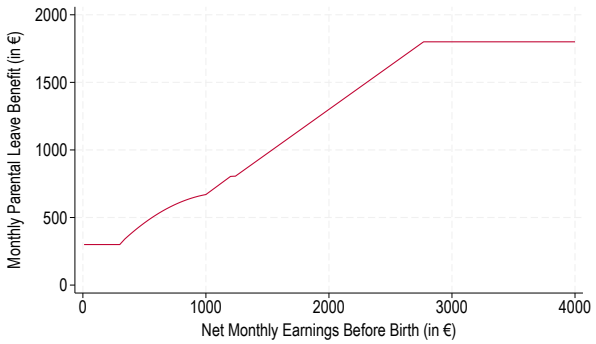


Figure 2.1. Parental Leave Benefits Schedule with Respect to Reference Income

Notes: The benefits schedule represents monthly parental leave payments as a function of net monthly earnings in the 12 months preceding childbirth.

2.2.2 Work During Parental Leave

Parents are allowed to work up to 30 hours per week while receiving parental leave benefits. When working part-time during the leave period, benefits are calculated as 65% of the adjusted reference income, defined as the difference between pre-birth net earnings and earnings during the benefit period.

In July 2015, a new scheme—Parental Benefit Plus (*Elterngeld Plus*)—was introduced to make part-time work during parental leave more attractive and flexible, particularly for parents seeking to combine employment and childcare. Under this scheme, parents working part-time may receive up to half of the basic parental leave benefit they would have received had they not worked. In exchange, the benefit duration is extended, effectively doubling the length of entitlement (e.g., from 12 to 24 months). The total leave duration is capped at 28 months per household when both parents take leave while working.[4] Furthermore, if both parents work between 25 and 30 hours per week for four consecutive months, they each qualify for an additional four months of *Elterngeld Plus*, referred to as the Partnership Bonus (*Partnerschaftsbonus*).

2.2.3 Childcare Availability

Since 2013, all children aged one and older have had a legal entitlement to a place in early childhood education and care in Germany. However, formal childcare use for children under the age of one remained very limited during the sample period (2014–2018). For example, in 2018, only about 2% of children under one were enrolled in formal childcare settings Destatis (2018). By contrast, childcare availability for children aged three to six was nearly universal, with utilization rates

4. For example, a parent entitled to €1,800 per month in basic *Elterngeld* could choose instead to receive €900 per month for 24 months under *Elterngeld Plus*.

around 94–96%. Childcare costs in Germany are determined at the municipal level and vary based on household income, number of children, and hours of care. During the sample period, parental contributions typically ranged from €100 to €400 per month (Geis-Thöne, 2018).

2.3 Data and Sample Description

2.3.1 Data

Our analysis is based on novel administrative microdata on the full population of parental leave benefit recipients in Germany (*Elterngeldstatistik*), provided by the German Federal Statistical Office (Statistisches Bundesamt). To our knowledge, this is the first study to exploit these register-based data for research purposes. The dataset covers all individuals who applied for and received parental leave benefits between 2014 and 2018. Observations correspond to parent-child pairs, and the data allow for reliable linkage of parents within the same household via a shared child identifier. This feature enables household-level analyses that are typically infeasible with more commonly used administrative sources.

A key advantage of this dataset is that it overcomes limitations present in other administrative sources. In contrast to linked employer-employee data or pension data which are typically used to study parental leave in Germany, the data set (i) allows the observation of both parents, (ii) allows the clear identification of child birth even after the first child, (iii) includes self-employed and civil servants, and (iv) covers the universe of all parental leave spells. Also, in contrast to the tax data, the parental leave benefit data allows the isolation of parental leave benefits from other types of transfers and includes the exact reference income that is used to calculate parental leave benefits. Our dataset provides exact information on benefit duration and amount, types of leave taken, part-time employment during leave, and associated monthly earnings. Additionally, the data include all variables necessary to determine benefit eligibility and replacement rates, such as pre-birth gross and net earnings, employment type (e.g., self-employed, civil servant[5]), withholding tax classes, church tax payment status, and health insurance type. Finally, comprehensive socio-demographic information is available, including age, number of dependents, municipality of residence, country of birth, and marital status.

A limitation of the dataset is that pre-birth earnings and related characteristics are observed only for individuals who take up parental leave. This results in near-universal coverage for mothers—whose take-up rate approaches 100%—but only partial coverage for fathers, of whom approximately 65% in our sample take at least two months of leave during the study period. This poses challenges for our empirical analysis, as the absence of reference income data for non-recipients

5. Civil servants are identified by their exemption from public pension insurance contributions.

limits our ability to assess how increased paternal leave benefits affect household decision-making. To overcome this limitation, we are in the process of supplementing the main dataset with administrative microdata from the German Taxpayer Panel, which covers all income tax filers and contains detailed information on earnings, transfers, and individual-level covariates similar to those in our primary data. However, parental links in the Taxpayer Panel can only be established for married couples filing jointly, and the data are recorded annually, complicating precise measurement of leave duration and reconstruction of monthly reference income in the 12 months preceding childbirth. Despite these constraints, the Taxpayer Panel offers key advantages: it captures the characteristics—including income—of individuals who do not take up parental leave, enabling causal estimation of benefit effects for fathers as well as mothers and allowing us to corroborate the validity of our empirical strategy through additional checks. Furthermore, its panel structure facilitates analysis of long-term earnings trajectories following parental leave.

2.3.2 Sample Description

Our analysis focuses on all parental leave benefit recipients in Germany with children born between 2014 and 2018, which includes nearly five million individuals. Due to data constraints, the study is limited to heterosexual couples.

Baseline sample. Our empirical strategy leverages a sharp kink in the parental leave benefit schedule, which is a function of pre-birth net income, to identify the causal effect of benefit generosity on household leave-taking decisions. Similar to a standard regression discontinuity (RD) design, this approach relies on the comparability of individuals on either side of the kink point. Accordingly, we restrict our baseline sample to households where either (i) the mother's or (ii) the father's reference income lies within €500 of the kink threshold (€2,769). Because income data are only available for parental leave recipients, condition (ii) effectively applies only to fathers who take some leave. This results in a sample of 1,127,848 individuals from 563,924 households, representing approximately 15.5% of the full population. Each birth constitutes a separate observation in our dataset.[6]

Descriptive statistics. Table 2.1 presents descriptive statistics for the baseline sample. Among these parents, 98.7% of mothers and 82.1% of fathers take at least some parental leave, resulting in 80.8% of households where both partners share leave. However, leave duration is highly uneven: on average, mothers take more than four times as much leave as fathers (11.2 months versus 2.6 months). Approximately one-fifth of parents engage in part-time employment during their leave period. Since we do not impose the condition that both parents fall within the income bandwidth around the kink, average net monthly income differs by gender:

6. Due to data limitations, multiple births within the same household cannot be linked.

women earn €2,028.9 on average, compared to €2,589.5 for men. This income gap translates into lower average parental leave benefits for mothers relative to fathers (€1,265.6 vs. €1,568.9).

Mothers are more likely than fathers to have had no employment in the reference year and are more often employees rather than self-employed. Women are also more likely than men to be civil servants. Regarding individual characteristics, most individuals are German-born, nearly all reside in Germany, and approximately 14% live in East Germany. The overrepresentation of West German residents in the sample reflects the location of the kink point in the upper part of the income distribution, combined with relatively lower average earnings in East Germany. Men are on average 2.4 years older than women, around 78% of households are married, and for 77% of women and 62.2% of men, the current one is their first child. Between 52.5% and 59.1% of mothers and fathers have a religious affiliation (Catholic or Protestant) and pay church taxes.

Parents with incomes around the kink. In the current version of the paper, due to missing income data for fathers who do not take parental leave, the empirical analysis primarily focuses on estimating the causal effects of mothers' benefit amounts on household decisions. This implies restricting the sample to mothers whose net income falls within a bandwidth of ±€500 around the kink point. Table 2.A.1 reports summary statistics for the sample where the income of both partners lies within this range. Mothers take an average of 11.4 months of parental leave, and in 65.4% of cases, leave is shared with their partner. One-quarter of mothers engage in part-time employment during the leave period. On average, they receive €1,694 per month in parental leave benefits, and their average reference income is €2,706.7. Approximately 36% of mothers in this sample are civil servants, and the vast majority of their income derives from labor earnings. Overall, their observable characteristics are broadly comparable to those in the full sample reported in Table 2.1.

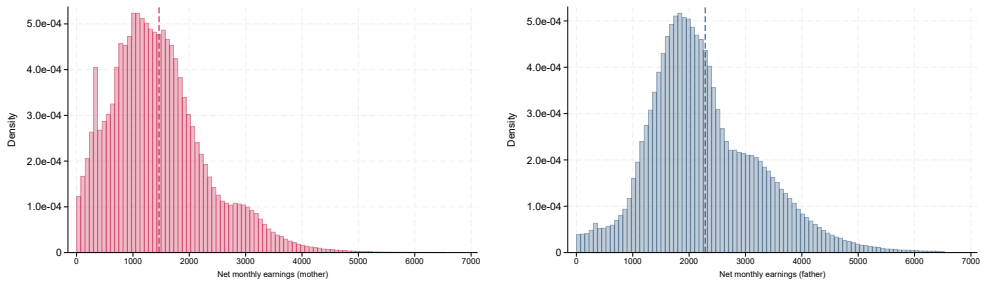
Households sharing leave. Table 2.A.2 reports summary statistics for households in which both partners take at least two months of parental leave. Reflecting strong assortative mating, most couples have identical replacement rates—meaning both partners fall on the same side of the kink. Nevertheless, in 11.7% of households, the mother has a lower replacement rate than the father (i.e., she is above the kink), while in 28.2% of households, the father has a lower replacement rate (i.e., he is above the kink). On average, these households take 13.7 months of parental leave, with over 82% utilizing the full 14-month entitlement. Mothers account for 81.3% of the total leave taken. Regarding demographic and economic characteristics, fathers are on average 2.5 years older than mothers, and mothers contribute approximately 40% of total household income. The average net household income in this group is €4,467.80 per month.

Table 2.1. Descriptive Statistics on PL Claims, Outcomes and Individual Characteristics

	Mothers			Fathers		
	Mean	SD	Obs	Mean	SD	Obs
PL claim characteristics						
Months of paid parental leave	11.1	2.2	563,924	2.1	1.9	563,924
Take-up of parental leave (%)	98.7	11.3	563,924	82.1	38.3	563,924
Both parents take up PL (%)	80.8	39.4	563,924	80.8	39.4	563,924
Months of paid parental leave (>0)	11.2	1.8	556,613	2.6	1.8	463,102
Part-time work during PL (%)	21.0	40.7	556,613	22.5	41.7	463,102
Parental leave Plus (%)	1.6	12.7	556,613	0.7	8.5	463,102
Monthly net PL benefit (€)	1,265.4	584.7	556,613	1,568.7	353.0	463,102
Demographic information						
Age	33.4	3.8	556,613	35.8	4.9	463,102
Resident in Germany (%)	99.7	5.8	556,613	99.8	4.7	463,102
Resident in East Germany (%)	13.7	34.4	554,705	14.2	34.9	462,096
German citizenship (%)	93.0	25.5	556,613	94.3	23.1	463,102
Affiliated to Church (%)	59.1	49.2	523,271	52.5	49.9	454,294
Married (%)	78.1	41.3	563,924	78.4	41.2	563,924
First birth (%)	77.0	42.1	554,089	62.2	48.5	461,744
Number of dependents	0.3	0.6	554,089	0.5	0.7	461,744
Income in reference year						
No income (%)	7.8	26.8	556,613	1.9	13.6	463,102
Labor income (%)	88.4	32.0	563,924	77.5	41.7	563,924
Self-employed income (%)	6.3	24.2	556,613	9.7	29.7	463,102
Both self-employed and labor income (%)	3.6	18.6	563,924	5.0	21.7	563,924
Gross monthly income (€)	3,054.1	1,577.0	556,613	4,074.8	1,390.1	463,102
Net monthly income (€)	2,028.5	1,000.3	556,613	2,588.9	785.4	463,102
Civil servant (%)	25.4	43.5	556,613	14.4	35.1	463,102
Withholding tax class 1 (%)	32.7	46.9	532,557	29.6	45.7	458,255
Withholding tax class 2 (%)	0.8	8.8	532,557	0.1	2.9	458,255
Withholding tax class 3 (%)	12.6	33.1	532,557	23.0	42.1	458,255
Withholding tax class 4 (%)	44.1	49.7	532,557	39.6	48.9	458,255
Withholding tax class 5 (%)	9.8	29.7	532,557	7.7	26.7	458,255

Notes: The table reports summary statistics for the main sample, which includes households where either (i) the mother's or (ii) the father's reference income lies within €500 of the kink threshold (€2,769). Because income data are only available for parental leave recipients, condition (ii) effectively applies only to fathers who take some leave. As a result, fathers' take-up rates are higher in this sample compared to the overall population. Individual-level variables are observed only for parents who claimed parental leave, which accounts for the smaller sample sizes in some entries.

The unconditional distribution of net monthly labor income among German parental leave recipients is shown in Figure 2.2, separately by parent's gender. The bimodal shape of the distribution is the mechanical result of regulation on public health insurance contributions. Above a certain gross income threshold, individuals are no longer required to contribute to statutory health insurance, leading to an ac-



(a) Mothers **(b)** Fathers

Figure 2.2. Distribution of Net Income for Mothers and Fathers Receiving Parental Leave Benefit

Notes: These figures display the unconditional distribution of net monthly earnings (reference income) among parental benefit recipients in Germany, separately for mothers and fathers. The dashed lines represent the median.

cumulation of net incomes around this cutoff. Notably, this bimodality is absent in the gross income distribution.[7]

2.4 Empirical strategy

For each parent $j \in \{F, M\}$ in household h, we examine the relationship:

$$pl_{jh} = \beta_0^j + \beta_1^j b_{jh} + \beta_2^j b_{-jh} + u_{jh} \tag{2.1}$$

where the dependent variable pl_{jh} denotes the number of months parent j spends in paid parental leave, and b_{jh} is the individual's parental leave benefit. The term b_{-jh} denotes the benefit received by the partner. While prior studies have estimated the direct effect of benefit generosity on own leave-taking behavior (β_1)[8], the magnitude of cross-spousal effects (β_2) has not been documented so far. Yet this parameter is crucial: it captures the extent to which parental inputs in childcare are substitutable. When inputs are highly substitutable, changes in (relative) replacement rates lead to larger gains in household disposable income, enabling couples to more readily reallocate time in response to changing economic incentives. Conversely, a small β_2 implies low substitutability and suggests that couples have strong preferences for specific combinations of parental inputs—even at the expense of potential income. This paper fills this gap by estimating the degree of cross-spousal respon-

7. The same pattern emerges for both married and unmarried individuals (Figure 2.A.1), alleviating concerns that the increase in density is due to strategic manipulation of tax classes, which is not possible for unmarried individuals.

8. See, for example, Asai (2015), Bana, Bedard, and Rossin-Slater (2020), Ginja, Jans, and Karimi (2020), and Ziegler and Bamieh (2023).

siveness to benefit generosity and evaluating how couples adjust their allocation of parental leave in response to shifts in financial incentives.

Endogeneity. Estimating the relationship outlined in equation 2.1 is likely to result in biased estimates of both parameters, even with flexible controls for observed characteristics. This bias arises from the non-random assignment of parental benefits: benefit levels are mechanically linked to pre-birth net income, which is itself likely correlated with unobserved determinants of leave-taking (e.g. career attachment, intra-household bargaining power). The empirical challenge therefore lies in finding credibly exogenous sources of variations in parental leave benefits. To this end, we exploit a kink in the parental leave benefit formula, which induces a discrete change in the slope of benefits with respect to net income (Figure 2.1). This institutional feature allows us to implement an extended Regression Kink Design (RKD) to causally estimate the effect of benefit levels on the leave duration of both parents.

Benefit schedule. The benefit schedule around the kink is defined as:

$$B(W) = \begin{cases} 0.65 \cdot W & \text{if } W \leq 2{,}769 \\ 1{,}800 & \text{if } W > 2{,}769 \end{cases}$$

At the kink point, the first derivative of the benefit function with respect to income exhibits a discontinuous drop. In other words, the replacement rate—the proportion of prior earnings replaced by benefits—decreases sharply once net income exceeds the threshold of €2,769, as illustrated in Figure 2.3.[9] This discontinuity in the slope of the benefit function, combined with a large sample of over one million individuals, provides a compelling setting for a regression kink design (RKD). The key identifying assumption underlying our empirical strategy is that individuals should not be able to precisely manipulate their position relative to the kink (i.e., the net income level that determines benefit generosity). Under this assumption, the kink introduces quasi-experimental variation in benefit generosity. Provided that other determinants of the outcome evolve smoothly around the threshold, this variation can be leveraged for causal identification (Card et al. (2015)).

2.4.1 Regression Kink Design (RKD)

Let Y denote the outcome of interest, B the PL benefit level, and W the reference income. Card et al. (2015) show that, under a set of smoothness assumptions, the RK estimand

9. The maximum benefit amount remained constant since 2007 throughout the sample period, so also the threshold remained at the same income.

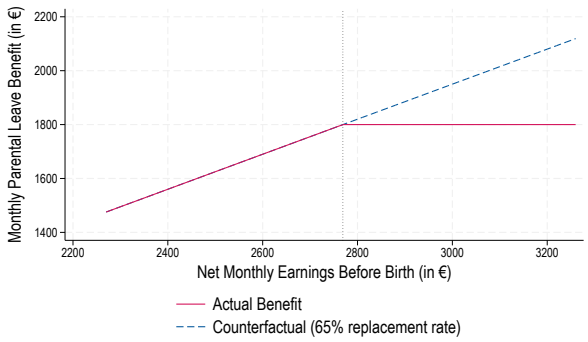


Figure 2.3. Variation in Parental Leave Benefits Around the Kink

Notes: This figure shows the parental leave benefits schedule in a bandwidth of ±500€ around the kink point.

$$\alpha = \frac{\lim_{w \to k^+} \frac{\partial E[Y|W=w]}{\partial w} - \lim_{w \to k^-} \frac{\partial E[Y|W=w]}{\partial w}}{\lim_{w \to k^+} b'(w) - \lim_{w \to k^-} b'(w)}, \tag{2.2}$$

identifies a weighted average of the marginal effects of B on Y. In this setting, all benefit assignments follow a deterministic formula, implying that B is a known function of W, and the denominator of the RKD estimand is a known constant. As a result, identification relies solely on estimating the numerator, which captures the change in the slope of the conditional expectation function of the outcome with respect to the assignment variable at the kink. Following Card et al. (2015) and Card et al. (2017), this can be done by estimating a local polynomial regression of the form:

$$E[Y|W = w] = \mu_0 + \sum_{p=1}^{P} \Big[\gamma_p (w-k)^p + \beta_p (w-k)^p \cdot D \Big] \tag{2.3}$$

$$\text{where } |w-k| \leq h$$

where W is the assignment variable; k is the kink point; $D = \mathbb{1}[W \geq k]$ is an indicator for observations above the kink; and P denotes the order of the polynomial. The coefficient β_1 captures the change in the slope of the conditional expectation function at the kink and thus corresponds to the numerator of (2.2). Estimation is conducted using observations within a symmetric bandwidth h around the kink.

Due to data limitations discussed earlier, our baseline analysis uses the mother's income as the assignment variable. This yields causal estimates of her benefit generosity on both parents' leave durations, via the following specification:

$$pl_{jh} = \mu_0 + \gamma_{1j}(w_{Mh} - k) + \beta_{1j}(w_{Mh} - k) \cdot \mathbb{1}[w_{Mh} \geq k] + \varepsilon_{jh} \tag{2.4}$$

where for each parent $j \in \{M, F\}$, the coefficient β_{1j} captures the causal effect of a change in the mother's benefit level on leave duration.

Identifying assumptions. We extend the identification assumptions outlined in Card et al. (2015) to our setting. To interpret the RKD estimand as a causal effect, two key assumptions must hold:

(1) In the absence of the policy kink, the marginal effect of the assignment variable on the outcome should evolve smoothly for both the individual and their partner.

(2) The distribution of unobserved determinants of Y—both own and partner's—must evolve smoothly with the assignment variable at the kink point.

If these assumptions hold, any observed change in the slope of the outcome with respect to the assignment variable at the kink can be attributed to the change in benefit generosity, rather than to unobserved confounding factors.

2.4.2 Validity of the Empirical Approach

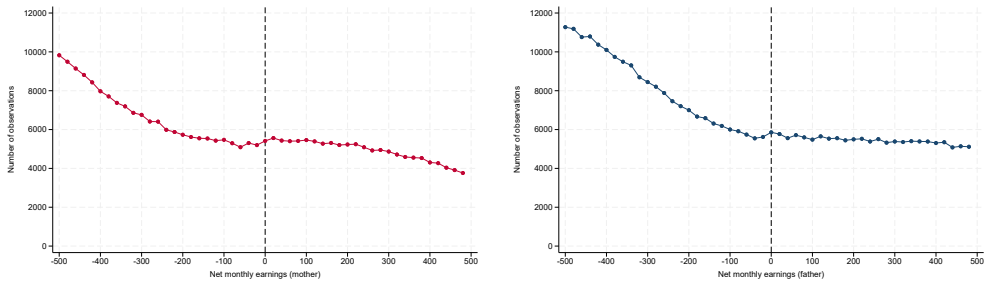
These smoothness assumptions lead to two testable implications. First, the density of the assignment variable should evolve smoothly at the kink. A discontinuity in the density of W may suggest manipulation of net income to just above or below the kink, undermining the quasi-random variation that the RKD exploits. Second, predetermined covariates should not exhibit changes in slope at the kink. We empirically assess both conditions. Figure 2.4 plots the number of parental leave spells by €20 bins of pre-birth net monthly earnings for mothers (Panel a) and fathers (Panel b). We find no visual evidence of bunching or discontinuities around the kink.[10]

To assess the second identifying assumption, Figure 2.6 plots mean values of several predetermined covariates in €20 bins of mothers' income near the kink. All covariates evolve smoothly through the threshold, with no evidence of changes in slope, supporting the validity of the RKD design.[11] Note that under our empirical strategy, this assumption also requires fathers' covariates to evolve smoothly at the mother's kink. Due to data limitations, we defer this analysis to future work using fiscal data with comprehensive income coverage for both partners.

Civil servants. For civil servants there is a visible break, shown in Figure 2.5, which exhibits a visible change in slope in proximity of the kink point—specifically, at an income level roughly €100 below the kink. The share of civil servants increases steadily up to this threshold, then plateaus before declining. This pattern

10. Because earnings are only observed for parental leave recipients, and not all fathers take leave, the analysis for fathers should be considered preliminary. We plan to strengthen this analysis using fiscal data.

11. A preliminary analysis of fathers' covariates around their own kink is presented in Figure 2.A.2.



(a) Mothers **(b)** Fathers

Figure 2.4. Distribution of Parental Leave Spells Around the Kink

Notes: The figure assesses the validity of the RKD assumptions by graphically testing the smoothness of the distribution of the assignment variable at the kink point in the PL benefit schedules. It plots the probability density function of the assignment variable, normalized at the kink, separately for mothers (Panel a) and fathers (Panel b). The assignment variable is binned in intervals of 20€.

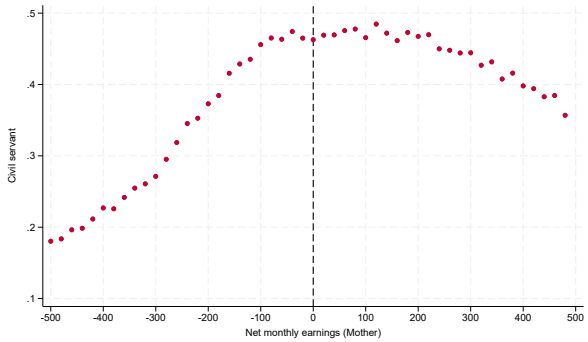
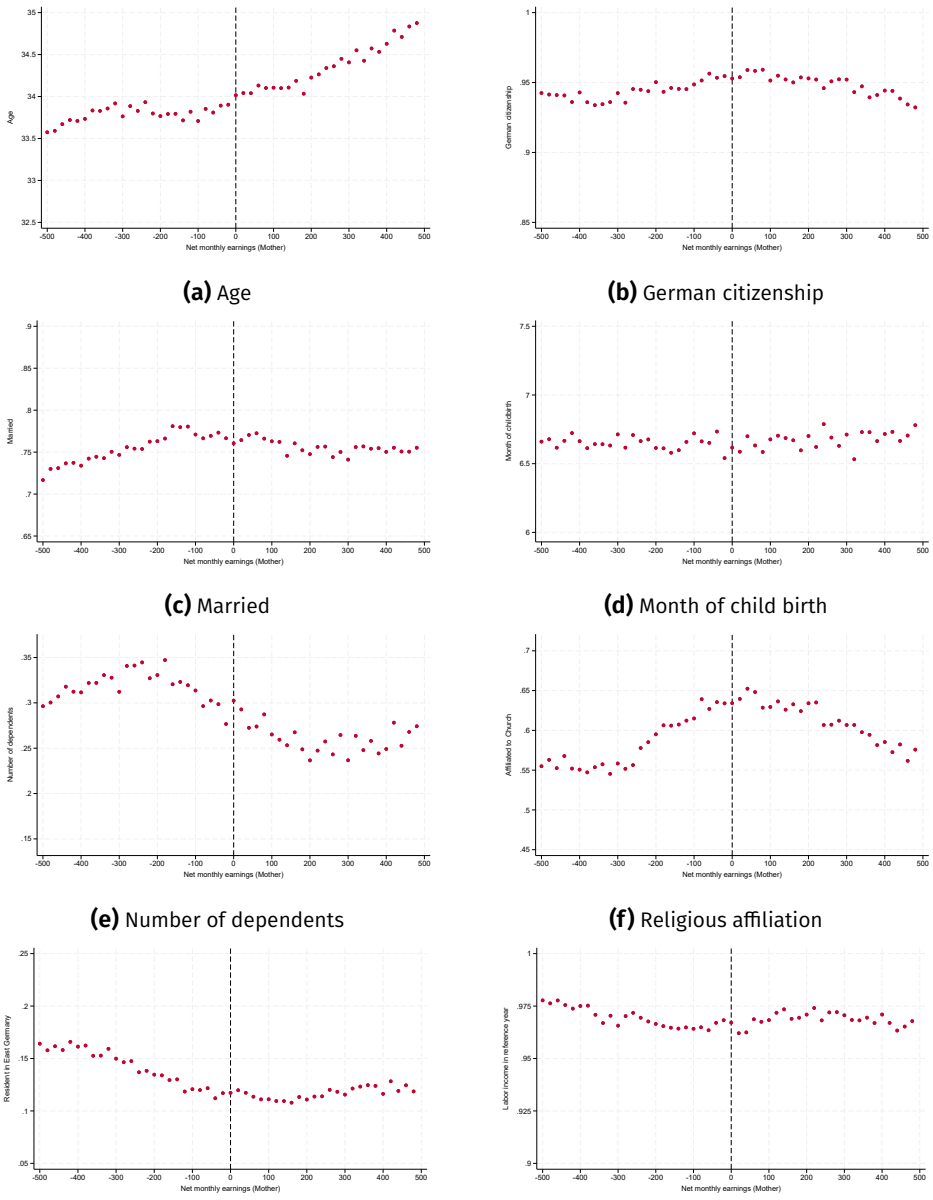


Figure 2.5. Distribution of Civil Servants Around the Kink (Mothers)

Notes: This graph plots the share of civil servants by €20 bins of the assignment variable. The sample includes all mothers with reference income within a bandwidth around the kink point.

aligns with the natural distribution of civil servants along the gross income spectrum, as illustrated in Figure 2.A.3, where their share peaks around €4,000 gross monthly income and declines thereafter.[12] This distribution arises from the structure of public-sector compensation in Germany, where salaries follow rigid, seniority-based pay scales. Most civil servants in our sample are in their mid-30s occupying mid-level grades, making higher salary tiers uncommon. Because civil servant status is likely correlated with job stability and other institutional factors influencing parental leave behavior, we include it as a control in our empirical specifications and conduct subsample analyses by civil servant status.

12. The mapping from gross to net income in Germany depends on withholding tax class, marital status, church tax obligations, and health insurance. A gross income of €4,000 typically corresponds to a net income of approximately €2,600–€2,800, consistent with the observed plateau.



(a) Age **(b)** German citizenship

(c) Married **(d)** Month of child birth

(e) Number of dependents **(f)** Religious affiliation

(g) Resident in East Germany **(h)** Employee status

Figure 2.6. Distribution of Pre-determined Covariates Around the Kink (Mothers)

Notes: The graphs test the validity of the smoothness assumptions of the RK design. The sample includes all mothers with reference income within a neighborhood around the kink point. Each panel displays the mean of a different predetermined covariate, binned in €20 intervals of the assignment variable. The visual evidence indicates that covariates evolve smoothly at the kink, supporting the credibility of the RKD identification assumptions.

2.4.3 Double Regression Kink Design (Double RKD)

One goal of this paper is to move beyond estimating how an individual's benefit level affects their own parental leave uptake, and to also identify cross-spousal elasticities. Building on the standard RKD framework, we extend the approach to jointly estimate the effects of both own and partner benefits on leave duration—a strategy we refer to as a double regression kink design (double RKD). The key idea exploits that both partners face the same kink in the benefit schedule, while their positions relative to the kink depend on their own pre-birth earnings. Provided neither partner can precisely manipulate their income around the kink, this setup yields quasi-experimental variation in both own and partner benefit generosity. Since each parent's leave decision may respond to changes in either their own or their partner's benefit, the double RKD allows us to identify both own- and cross-elasticities within the following unified empirical framework:

$$\begin{aligned} pl_{jh} \;=\; & \delta_0 + \delta_{1j}(w_{jh} - k) + \eta_{1j}(w_{jh} - k) \times \mathbb{1}[w_{jh} \geq k] \\ & + \delta_{2j}(w_{-jh} - k) + \eta_{2j}(w_{-jh} - k) \times \mathbb{1}[w_{-jh} \geq k] + \varepsilon_{jh} \end{aligned} \tag{2.5}$$

The parameters of interest are η_{1j} and η_{2j}, which capture changes in the slope of leave duration at the kink point with respect to own and partner income, respectively. A non-zero value of η_{2j} indicates that a partner's benefit generosity has a causal effect on the other partner's leave behavior. To illustrate the interpretation, suppose $\eta_{1M} < 0$, indicating that a reduction in the mother's benefit leads her to shorten her leave duration. This could give rise to two potential responses by the father: (1) No adjustment, i.e., $\eta_{2F} = 0$, in which case total household leave duration falls; (2) Offsetting response, i.e., $\eta_{2F} > 0$, where the father increases his own leave duration in response. In this case, the total household leave duration may remain unchanged, but the intra-household allocation of leave shifts between parents.

Identifying assumptions. To interpret the estimates from the double RKD as causal, we rely on an extension of the standard RKD identifying assumptions to a setting with two assignment variables—one for each partner. First, in the absence of the kink, the relationship between each parent's leave duration and both their own and their partner's income must evolve smoothly through the kink point. That is, there should be no change in slope in the conditional expectation of pl_{jh} with respect to either w_{jh} or w_{-jh} at k, absent the policy kink. Second, the distribution of both observable characteristics and unobserved determinants of parental leave should be continuous at the kink points in both w_{jh} and w_{-jh}. If these conditions hold, any observed kinks in the relationship between leave duration and income can be causally attributed to benefit changes driven by the policy schedule.

Take-up of fathers. The challenge that we face in implementing the double RKD design is data availability. In particular, we only observe the father's pre-birth

earnings—the running variable required to determine his position relative to the kink—for the subset of fathers who take parental leave, which accounts for approximately 65% of the sample of mothers near the kink. Consequently, the main analysis focuses on a single RKD estimating the effect of the mother's benefit on both her own and her partner's leave duration (equation 2.4), with results presented in the next section. As a supplementary exercise, we restrict the sample to households with both partners' earnings observed and estimate the full double RKD model. Preliminary estimates appear in Section 2.7. However, this subsample may not represent the broader population, as fathers who take leave may differ systematically in preferences, income, or bargaining dynamics, introducing potential selection bias. We plan to address this by integrating fiscal data with full income coverage for both partners in future work.

2.5 Results from simple RKD

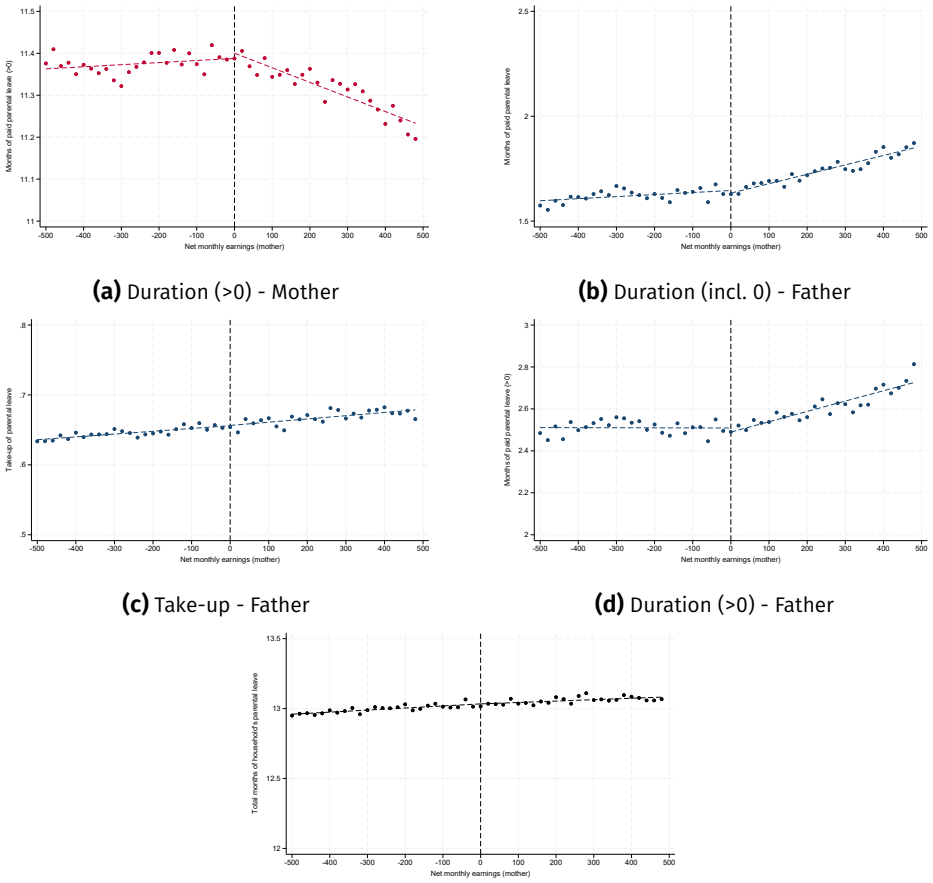
2.5.1 Graphical Evidence

We begin by graphically illustrating the kinked relationship between our main outcomes, i.e. parental leave durations of mothers and fathers, and the mother's running variable at the kink point, shown in Figure 2.7. Unlike the smooth distributions of covariates documented earlier, both parents' leave durations exhibit a clear change in slope at the kink. Specifically, while the duration of the mother's leave remains relatively flat with respect to reference income before the kink—consistent with a constant replacement rate—it drops sharply just after the kink, reflecting the decline in benefit generosity. Conversely, Panel (b) shows that the father's leave duration increases at the mother's kink point, suggesting a substitution in parental inputs following the reduction in the mother's benefits. Panels (c) and (d) indicate that this increase in the father's leave is driven by a longer duration among takers (intensive margin), rather than by an increase in take-up rates (extensive margin). Overall, the total parental leave taken by the household appears unaffected by the mother's kink.

Before interpreting these slope changes as causal effects of benefit changes, we present graphical evidence that they are not driven by shifts in mothers' observable characteristics.[13] Specifically, we regress each partner's parental leave outcomes (take-up and duration) on a comprehensive set of mothers' observable characteristics, including church tax liability, county (*Kreis*) fixed effects, age, employment status (employee vs. self-employed), marital status, child's month of birth, and number of dependents.[14] Figure 2.A.4 plots the predicted parental leave durations—rather

13. Due to data limitations, we currently cannot include fathers' observables but plan to do so in future work using tax return data.

14. We exclude a civil servant indicator here; its role is discussed in detail in the next section.



(a) Duration (>0) - Mother

(b) Duration (incl. 0) - Father

(c) Take-up - Father

(d) Duration (>0) - Father

(e) Duration - Household

Figure 2.7. Effects of Mother's Benefit Level on Households' Decisions

Notes: The graphs display the mean values of parental leave (PL) durations within €20 bins of pre-birth net earnings for mothers, normalized at the kink point. In all panels, the x-axis represents the mother's reference income. The outcome variables are: (a) the duration of the mother's PL in months; (b) the duration of the father's PL in months, including zeros; (c) the probability that the father takes up any PL; (d) the duration of the father's PL, conditional on take-up (i.e., excluding zeros); and (e) the total duration of PL taken by the household in months. The graphs show clear evidence of a kink in both the mother's and father's PL durations. By contrast, neither the extensive margin of father's PL take-up nor the total household PL duration exhibit a visible kink.

than the raw outcomes—binned by the mother's assignment variable. The smooth evolution of these predicted values around the kink suggests that the observed discontinuities in leave behavior are unlikely due to unobserved determinants.

This graphical evidence thus indicates that the reduction in the mother's benefit level at the kink point leads to a decrease in her parental leave duration, while fathers respond by increasing their leave. The next section quantifies the magnitude of these behavioral responses.

Table 2.2. RKD Estimates of the Effects of Mother's PL Benefits on Household Decisions

Variable	$\hat{\alpha}$	se($\hat{\alpha}$)	$\hat{\varepsilon}$	se($\hat{\varepsilon}$)	P-val	Mean	N
Household's PL duration (months)	-0.015	0.014	-0.020	0.019	0.292	13.031	164,979
Mother's PL duration (months)	0.062	0.015	0.098	0.023	0.000	11.366	164,979
Father's PL duration (months, incl. 0)	-0.077	0.015	-0.831	0.164	0.000	1.665	164,979
Prob. of PL take-up	-0.005	0.004	-0.145	0.114	0.203	0.656	164,979
Father's PL duration (months, >0)	-0.095	0.016	-0.677	0.117	0.000	2.537	108,236

Notes: Each row reports estimates from the baseline specification in equation (2.4), using a different outcome variable as indicated in the first column. α is the RK estimate of the average treatment effect of a €100 increase in parental leave (PL) benefits on the number of months spent on parental leave. Elasticities with respect to benefit levels are computed as: $\hat{\alpha} \cdot \frac{\bar{b}}{\bar{Y}}$, where $\bar{Y}$ is the mean outcome at the kink point, and $\bar{b}$ is the average benefit amount at the kink point. Robust standard errors are reported. All estimates are based on the linear specification, using a sample of households in which the mother's reference income falls within a ±€300 bandwidth around the kink point (€2,769).

2.5.2 RKD Estimates

Table 2.2 presents the results from the baseline specification of equation 2.4, estimated using a linear model. The sample includes all households in which the mother's reference income falls within a ±€300 bandwidth around the kink point, irrespective of the father's position in the income distribution. In Column (1), we report estimates of $\hat{\alpha} = \left(\frac{\hat{\beta}_1}{-0.65}\right) \times 100$, where $\hat{\beta}_1$ denotes the estimated change in slope in the relationship between the outcome and the assignment variable at the kink. The denominator reflects the deterministic change in the slope of the parental leave benefit schedule at the kink point. Multiplying this ratio by 100 allows us to interpret $\hat{\alpha}$ as the effect of a €100 increase in mothers' monthly benefit levels on the average duration (in months) of parental leave taken by each parent.

Results confirm the visual evidence discussed earlier. A €100 reduction in the mother's monthly parental leave benefit—equivalent to a 5.5% decrease relative to the average benefit at the kink—induces a reallocation of leave between partners. On average, mothers reduce their leave duration by 1.9 days (0.062 months), while fathers increase theirs by 2.31 days (0.077 months), yielding own- and cross-elasticities of 0.098 and –0.83, respectively. The asymmetry in elasticities reflects the substantial disparity in baseline leave durations across genders. The increase in fathers' leave is driven by longer durations among those who take leave (intensive margin), rather than an increase in take-up rates (extensive margin). Conditional on take-up, fathers extend their leave by approximately three days—about 4% relative to the mean. Importantly, the total amount of leave taken at the household level remains unchanged, indicating that the reduction in mothers' benefits affects only the intra-household allocation of leave, not the overall duration. These estimates are robust to the inclusion of a rich set of maternal covariates, including civil servant

status, religious affiliation, county fixed effects, age, employment characteristics, number of dependents, and month of childbirth.

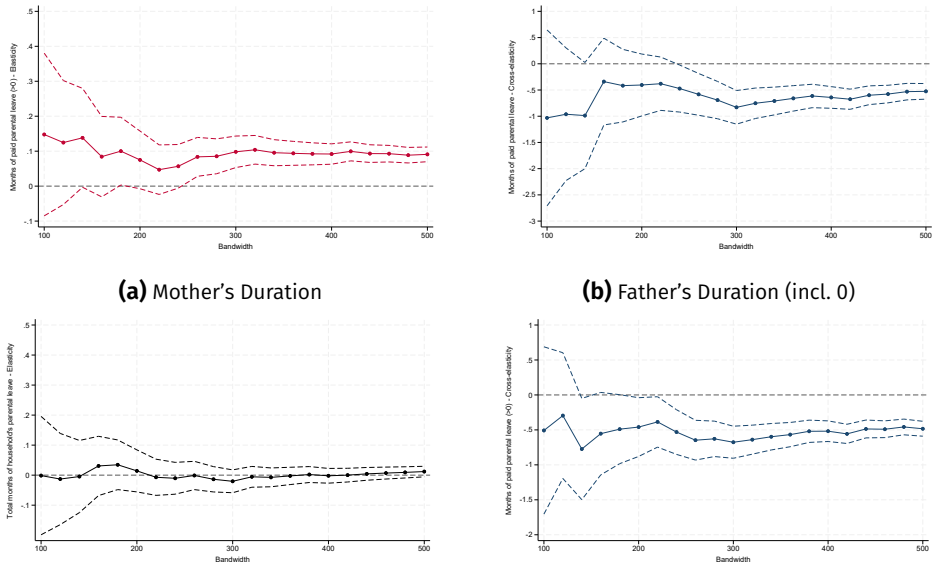
These estimates suggest that while parents respond to financial incentives, the magnitude of their responses is modest. This limited responsiveness may reflect gender differences in the opportunity costs of taking leave, particularly since fathers are older on average and earn higher labor market incomes (Table 2.1). An alternative explanation would be that parents have rigid preferences for a specific combination of parental inputs, potentially rooted in gender norms. In Section 2.6, we explore the first channel by examining heterogeneity in responses according to the father's relative income position.

Sensitivity to bandwidth selection. While the point estimates presented here are specific to a chosen bandwidth, we follow the recommendations of Lee and Lemieux (2010) and assess the robustness of our findings to alternative bandwidth selections. Figure 2.8 presents estimates of own- and cross-elasticities (with respect to the mother's benefit level) using bandwidths ranging from €100 to €500 around the kink. The results demonstrate that our estimates remain stable across this range. Moreover, although regression kink designs are typically data-intensive and can produce imprecise estimates in smaller samples, the richness of our administrative data provides ample statistical power even at narrower bandwidths. Consequently, we obtain precise and stable elasticity estimates, with confidence intervals that exclude zero for a wide array of bandwidth choices.

2.6 Heterogeneity

2.6.1 Father's Position in the Income Distribution

The results in Section 2.5 indicate that both own and cross-spousal responses to more generous PL benefits for mothers are modest in magnitude. A plausible explanation is that parents face asymmetric opportunity costs of taking leave—particularly in households with large earnings gaps between partners. In families where the father earns substantially more than the mother, the marginal incentive created by a change in the mother's benefit may be insufficient to induce reallocation of leave. In contrast, families where the father earns less—or faces a higher replacement rate by falling below the kink—may be more responsive to changes in maternal incentives. We explore this heterogeneity by examining whether cross-spousal substitution varies with the father's position in the income distribution, by splitting the sample based on whether his reference income lies above or below the kink—corresponding to a lower or higher replacement rate, respectively—and re-estimating the RKD within each group. Due to data limitations, this analysis is restricted to households in which both partners take parental leave. Figure 2.9 reports the estimated effects on both maternal and paternal leave durations.



(a) Mother's Duration

(b) Father's Duration (incl. 0)

(c) Household's Duration

(d) Father's Duration (>0)

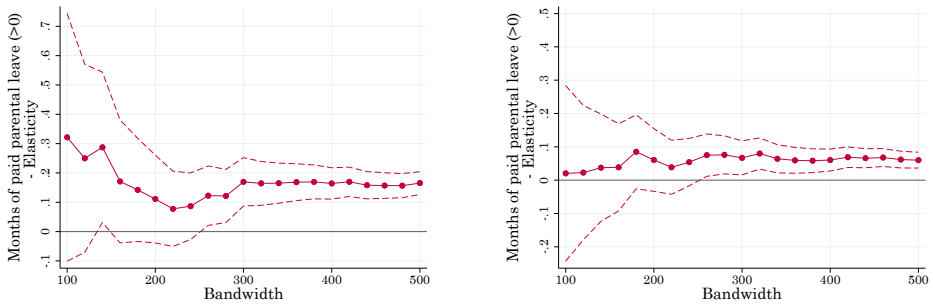
Figure 2.8. Effects of Mother's PL Benefits – Sensitivity to Bandwidth Selection

Notes: Each graph shows estimated elasticities of different outcomes with respect to the mother's benefit level. For each outcome, elasticities are plotted across varying bandwidths, ranging from €100 to €500 around the kink point. The estimates are generally stable across bandwidth choices, supporting the robustness of the results.

The results show that a reduction in the mother's benefit leads to a reallocation of leave from mother to father in both subsamples, but the magnitude differs by the father's income. Specifically, the direct effect of a benefit reduction on the mother's own leave duration is nearly twice as large when the father's reference income is below the kink, with an elasticity of approximately 0.2. Cross-spousal responses are also somewhat larger in this group. This suggests that substitution is more likely when the father faces a lower opportunity cost.

While intra-household reallocation becomes more pronounced when fathers face lower opportunity cost of leave-taking, the overall responsiveness remains limited, and parental leave uptake continues to be highly asymmetric across genders. These patterns suggest that economic incentives, though relevant, are insufficient to fully account for parental leave behavior. Instead, the results underscore the role of non-monetary factors in shaping intra-household labor supply decisions. In particular, they point to preferences for a specific allocation of parental inputs—likely influenced by prevailing gender norms, social expectations, or institutional constraints such as workplace culture and stigma—which limit the degree to which caregiving responsibilities are transferable across partners, even when economic conditions would favor such a shift. We plan to investigate among these potential explanation by collecting novel survey data.

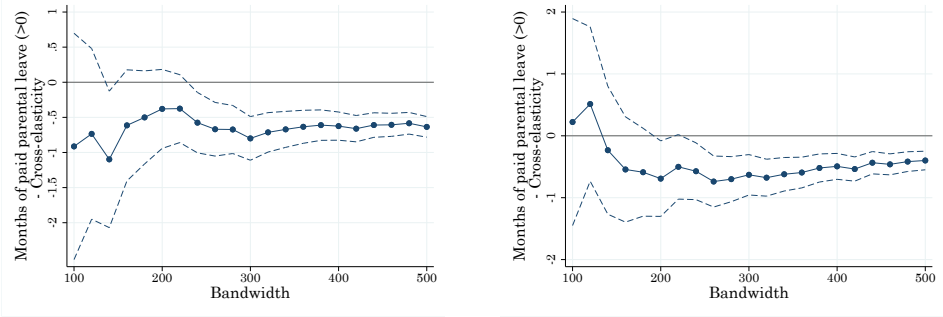
(1) Effects of Mother's PL Benefit on Mother's PL Duration



(a) Father: high replacement rate

(b) Father: low replacement rate

(2) Effects of Mother's PL Benefit on Father's PL Duration



(c) Father: high replacement rate

(d) Father: low replacement rate

Figure 2.9. Effects of Mother's Benefit Level by Father's Replacement Rate

Notes: These graphs present RKD estimates of the effects of the mother's PL benefit on leave duration for mothers (Panel 1) and fathers (Panel 2), separately by the father's replacement rate (high vs. low). The analysis is restricted to households in which the mother's reference income falls within the selected bandwidth around the kink point.

2.6.2 East vs. West Germany

We further examine whether behavioral responses to economic incentives differ between households residing in East versus West Germany. This heterogeneity is particularly relevant given well-documented regional differences in gender norms: East Germany is associated with more progressive attitudes toward gender roles, and women tend to return to work substantially earlier than their counterparts in West Germany (Boelmann, Raute, and Schönberg (2025)). Table 2.3 presents RKD estimates separately for the two subsamples.

The estimates reveal substantial regional heterogeneity in responsiveness. While households in both East and West Germany react to changes in the generosity of the mother's PL benefit, the magnitude of cross-spousal substitution is at least twice as large among couples in East Germany. Despite similar baseline PL durations for mothers across regions, a €100 decrease in monthly benefits reduces maternal leave by approximately 3.5 days in East Germany, compared to 1.65 days in West

Table 2.3. RKD Estimates of the Effects of Mother's Benefits in East and West Germany

Variable	$\hat{\alpha}$	se($\hat{\alpha}$)	$\hat{\varepsilon}$	se($\hat{\varepsilon}$)	P-val	Mean	N
East Germany							
Household's PL duration (months)	-0.008	0.039	-0.011	0.054	0.833	13.221	20181
Mother's PL duration (months)	0.115	0.043	0.182	0.068	0.008	11.325	20181
Father's PL duration (months, incl. 0)	-0.123	0.047	-1.166	0.443	0.009	1.896	20181
Prob. of PL take-up	-0.013	0.011	-0.334	0.297	0.261	0.696	20181
Father's PL duration (months, >0)	-0.125	0.050	-0.829	0.328	0.012	2.724	14047
West Germany							
Household's PL duration (months)	-0.012	0.015	-0.017	0.021	0.404	13.013	144020
Mother's PL duration (months)	0.054	0.015	0.085	0.024	0.000	11.375	144020
Father's PL duration (months, incl. 0)	-0.066	0.016	-0.728	0.176	0.000	1.638	144080
Prob. of PL take-up	-0.003	0.004	-0.091	0.123	0.457	0.653	144080
Father's PL duration (months, >0)	-0.087	0.017	-0.625	0.125	0.000	2.508	94084

Notes: This table reports RKD estimates based on equation (2.4), estimated separately for households residing in East and West Germany. All specifications are linear and use a bandwidth of ±€300 around the kink point (€2,769) in the benefit schedule. Coefficient interpretations follow those in Table 2.2. The sample includes households in which the mother's reference income falls within the selected bandwidth.

Germany. Likewise, the shift in leave toward fathers is more pronounced in East Germany, with implied cross-elasticities of –1.2 and –0.73, respectively.

2.7 Double RKD Estimates

Table 2.4 reports estimates from the double regression kink design (RKD) specified in equation (2.5). Due to the data limitations discussed earlier, this analysis is limited to the subsample of households in which both parents take up parental leave.

The first panel re-estimates the effect of the mother's parental leave benefits, focusing on households where the mother's reference income lies within €300 of the kink. The father's income is included as a control, as specified in equation 2.5. The estimates indicate that a €100 decrease in the mother's benefit reduces her leave duration by approximately 3 days (0.098 months), with an offsetting increase in the father's leave. Importantly, total household leave remains unchanged, consistent with earlier findings. These results closely match those from the single RKD when similarly restricted to leave-taking couples. The second panel turns to the father's benefit schedule, restricting the sample to households in which the father's income lies within €300 of the kink and controlling for the mother's income. In contrast to the maternal estimates, we find no significant effect of the father's benefit level on his own leave duration. Moreover, there is no evidence of compensatory changes in the mother's behavior. These findings suggest limited responsiveness to paternal incentives at the intensive margin.

These findings seem to indicate a marked asymmetry in how financial incentives shape maternal versus paternal leave behavior. However, the null effects for fathers

Table 2.4. D-RKD Estimates of the Effects of PL Benefits on Household Decisions

Variable	$\hat{\alpha}$	se($\hat{\alpha}$)	$\hat{\varepsilon}$	se($\hat{\varepsilon}$)	P-val	Mean	N
Effect of mother's PL benefit							
Total months of household's parental leave	-0.002	0.013	-0.003	0.017	0.877	13.031	108236
Mother's months of paid parental leave (>0)	0.098	0.019	0.156	0.030	0.000	11.366	108236
Father's months of paid parental leave (>0)	-0.100	0.016	-0.712	0.117	0.000	2.537	108236
Effect of father's PL benefit							
Total months of household's parental leave	0.008	0.010	0.010	0.014	0.460	13.031	180150
Mother's months of paid parental leave (>0)	-0.005	0.016	-0.009	0.026	0.741	11.366	180150
Father's months of paid parental leave (>0)	0.013	0.014	0.092	0.101	0.361	2.537	180150

Notes: This table reports estimates from equation 2.5, using the outcome variable indicated in the first column. Panel (1) focuses on households where the mother's earnings lie within €300 of the kink, controlling for the father's income position. Panel (2) restricts the sample to households where the father's earnings lie within €300 of the kink, controlling for the mother's income position. In both panels, the sample is limited to households in which both partners take up parental leave.

should be interpreted with caution. The double RKD estimates identify intensive-margin responses, conditional on fathers taking leave, and may therefore understate the full impact of benefit generosity. If financial incentives primarily operate on the extensive margin, inducing more fathers to take leave in the first place, our design will miss this behavioral margin entirely. Moreover, if higher paternal benefits induce more fathers to take leave, the observed sample of leave-taking fathers near the kink is not a random slice of all fathers—it overrepresents those who take leave even when benefits are relatively low (e.g., those with stronger preferences, more workplace flexibility, or more egalitarian gender norms). These individuals may be less responsive to marginal changes in benefits than the average father. As a result, the sample may exclude those who would be most responsive to the policy (i.e., fathers on the margin of taking leave), leading to attenuation bias. Future work using fiscal data—with comprehensive coverage of all parents regardless of leave-taking status—will allow us to address this limitation more directly.

2.8 Mechanisms

In this setting, parents are allowed to combine their PL benefits with part-time employment, up to 30 hours per week. We investigate how benefit generosity affects the incidence of part-time work during leave. We find that when mothers' PL benefits are reduced, women increase their labor market earnings during leave. This response appears to be driven by greater part-time work along the intensive margin—reflected in longer durations of labor market participation and, possibly, more hours worked. These results suggest that financial incentives influence not only the decision to take leave, but also the extent to which leave is combined with paid work. In particular, reduced benefits appear to induce a substitution of full-time childcare with part-time employment. While direct evidence is limited, this form of labor mar-

ket engagement may ease transitions back to full-time work and strengthen long-run labor force attachment. Conversely, we find significant cross-spousal responses in fathers' labor supply behavior. When mothers' PL benefits are reduced, fathers are more likely to engage in part-time work during their own leave, leading to an increase in their labor market earnings. This pattern suggests that the observed rise in paternal leave duration is accompanied, at least in part, by greater labor market engagement while on leave. In other words, some of the substitution in caregiving responsibilities appears to occur alongside continued—albeit reduced—labor force participation, rather than through a full withdrawal from work, aligning with previous findings by Andresen and Nix (2024).

2.9 Conclusions

The differential impact of parenthood on the careers of men and women—commonly referred to as the child penalty—accounts for a substantial share of the remaining gender gap in labor market outcomes. A central driver of this penalty is the persistently limited involvement of fathers in parental leave and early childcare, a pattern observed across all OECD countries. As a result, family policies—particularly parental leave—are widely viewed as a promising lever for promoting gender equality. Yet, evidence remains limited on whether, and how, the design of these policies can promote a more balanced division of childcare responsibilities within households.

This paper provides new causal evidence on how the monetary generosity of parental leave benefits affects the intra-household allocation of childcare. Exploiting a sharp kink in the German parental leave benefit schedule relative to pre-birth income, and applying an extension of the regression kink design, we estimate how both mothers and fathers adjust their leave behavior in response to variation in replacement rates—both their own and their partner's. The analysis leverages novel administrative microdata linking partners within households and covering over one million individuals between 2014 and 2018, with detailed information on earnings, benefit amounts, parental leave, and labor market activity. We find that households respond to relative replacement rates: when maternal benefits decline, mothers shorten their leave while fathers extend theirs, reallocating care within the household without affecting total leave duration. These responses, however, are quantitatively modest—the own elasticity of maternal leave duration with respect to the benefit level is approximately 0.1. Elasticities are larger when paternal opportunity costs are lower and in regions with more egalitarian gender norms, consistent with both economic and cultural constraints shaping behavior. We also document a previously overlooked behavioral margin: reductions in benefit levels lead mothers to shift from full-time caregiving to part-time work during leave, highlighting that benefit design also affects how childcare and work are combined.

We plan to extend this research in two directions. First, we will collect original survey data to better understand the mechanisms underlying limited responsiveness to financial incentives, focusing on gender norms, second-order beliefs, preferences, and expectations about employer behavior. Second, we will incorporate our reduced-form estimates into a structural household model to simulate counterfactual policy designs. These simulations will assess how alternative (and potentially gender-differentiated) replacement rates can promote a more equitable division of parental leave and mitigate long-run child penalties, while accounting for their broader welfare implications.

Appendix 2.A Additional Figures and Tables

Table 2.A.1. Summary Statistics — Parents Individually Around the Kink

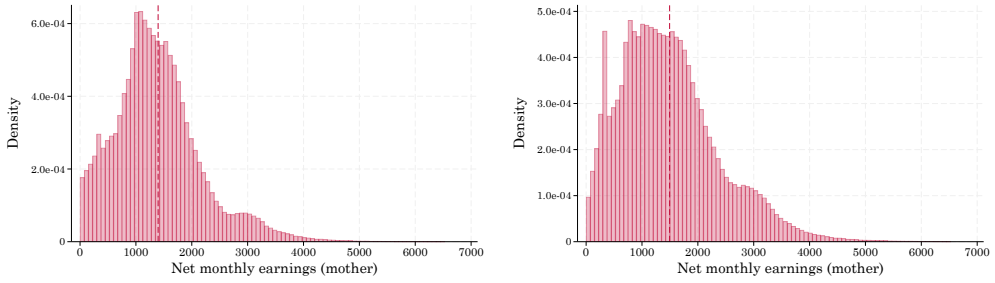
	Mothers			Fathers		
	Mean	SD	Obs	Mean	SD	Obs
PL claim characteristics						
Months of paid PL (>0)	11.4	1.7	291,336	2.6	1.9	338,798
Both parents take-up PL (%)	65.4	47.6	291,336	97.8	14.5	338,798
Part-time during PL (%)	24.8	43.2	291,336	22.7	41.9	338,798
Parental leave Plus (%)	1.6	12.4	291,336	0.7	8.4	338,798
Monthly net PL benefit (€)	1,694.9	115.0	291,336	1,689.8	115.2	338,798
Demographics						
Age	34.0	3.6	291,336	35.6	4.8	338,798
Resident in Germany (%)	99.5	6.8	291,336	99.8	4.8	338,798
East Germany (%)	13.3	34.0	289,972	13.8	34.5	338,000
German citizen (%)	94.6	22.7	291,336	94.6	22.5	338,798
Affiliated to church (%)	59.3	49.1	289,380	53.0	49.9	335,967
Married (%)	75.3	43.1	291,336	80.6	39.5	338,798
First birth (%)	73.9	43.9	291,035	56.2	49.6	338,205
Dependents	0.3	0.6	291,035	0.5	0.8	338,205
Income (reference year)						
No income (%)	0.0	0.7	291,336	0.0	0.6	338,798
Labor income (%)	97.0	17.2	291,336	96.7	17.8	338,798
Self-employed income (%)	6.9	25.3	291,336	9.5	29.3	338,798
Both types of income (%)	3.9	19.3	291,336	6.2	24.1	338,798
Gross monthly income (€)	4,016.1	678.4	291,336	4,183.6	675.8	338,798
Net monthly income (€)	2,706.7	292.1	291,336	2,698.5	296.3	338,798
Civil servant (%)	36.4	48.1	291,336	16.0	36.7	338,798
Tax class 1 (%)	34.5	47.5	291,042	28.5	45.1	338,330
Tax class 2 (%)	0.8	8.7	291,042	0.1	2.9	338,330
Tax class 3 (%)	18.6	38.9	291,042	29.8	45.8	338,330
Tax class 4 (%)	44.4	49.7	291,042	36.9	48.3	338,330
Tax class 5 (%)	1.8	13.2	291,042	4.7	21.2	338,330

Notes: Summary statistics for mothers and fathers where the individual reference income lies within a ±€500 bandwidth around the kink point (€2,769). Since income data are only available for parental leave recipients and not all fathers take parental leave, these statistics only represent fathers who take some leave.

Table 2.A.2. Summary Statistics — Households Sharing Parental Leave

	Mean	SD	Obs
PL characteristics			
Mother has lower replacement rate (%)	11.7	32.1	455,791
Father has lower replacement rate (%)	28.2	45.0	455,791
Household's months of PL	13.7	1.2	455,791
Mother's share of leave (%)	81.3	12.1	455,791
Mother's PL duration	11.1	1.9	455,791
Father's PL duration	2.5	1.6	455,791
Takes 14 months of PL (%)	82.3	38.1	455,791
Baseline and Income			
Father's - Mother's age	2.5	4.0	455,791
Net monthly household income (€)	4,467.8	1,264.7	455,791
Mother's income share (%)	39.7	20.5	455,791
Father's net monthly income (€)	2,587.0	790.6	455,791
Mother's net monthly income (€)	1,880.8	1,040.6	455,791

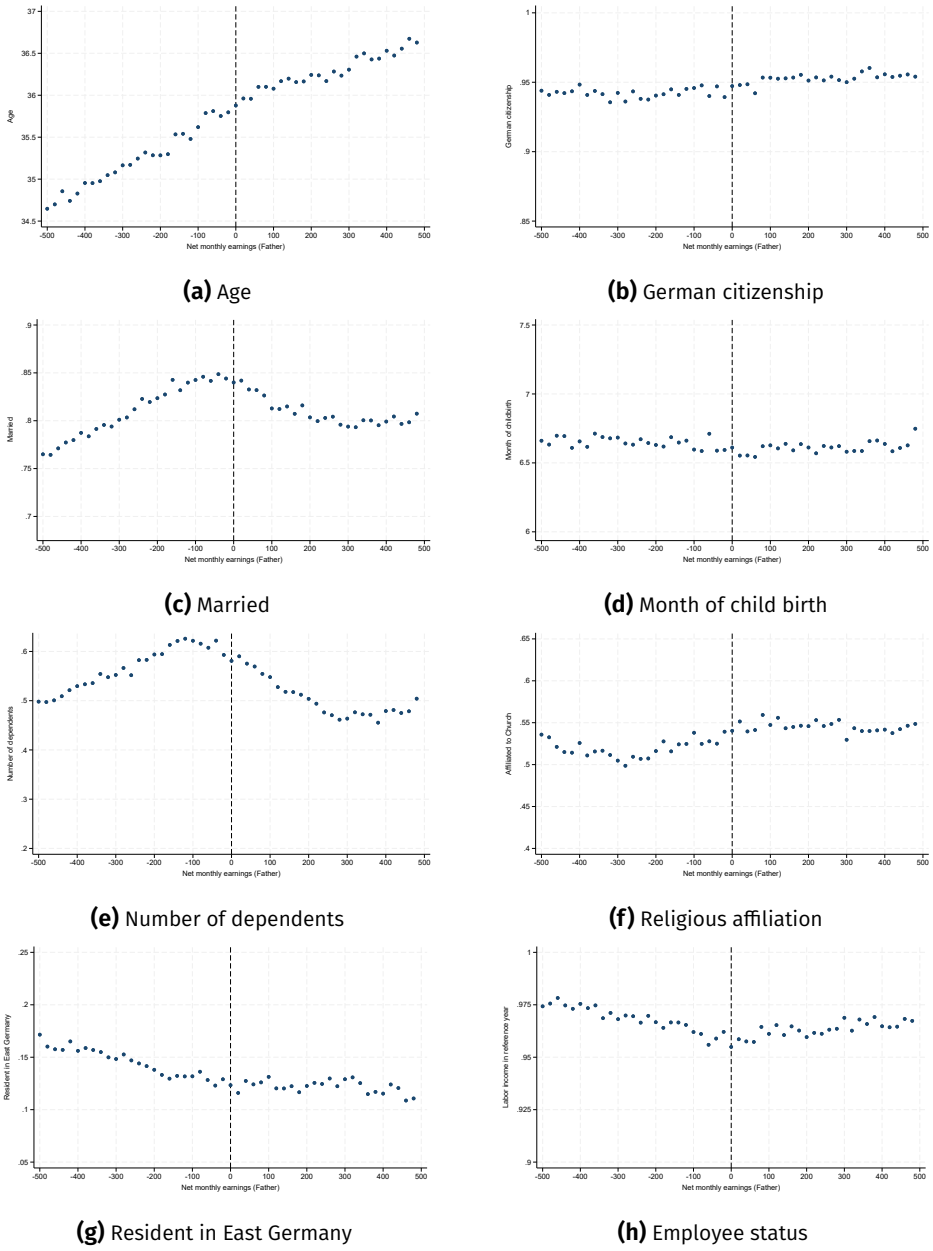
Notes: Summary statistics for households where both partners take up at least 2 months of parental leave.



(a) Non-married **(b)** Married

Figure 2.A.1. Distribution of Net Income for Non-married and Married Mothers

Notes: These figures show the distribution of net monthly earnings during the 12 months preceding childbirth (the reference income) among recipients of parental benefits in Germany. Panel (a) displays the distribution for non-married mothers, who file their income taxes individually. Panel (b) shows the distribution for married mothers, who file jointly with their spouses and are thus subject to tax rates based on household income rather than individual gross income. Vertical lines correspond to the median in the respective distributions.



(a) Age **(b)** German citizenship

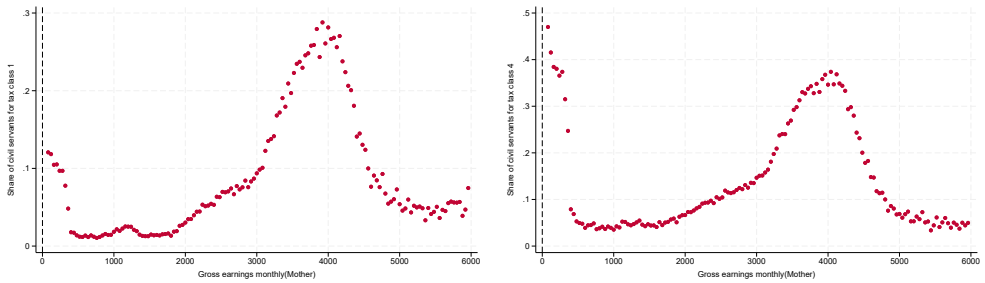
(c) Married **(d)** Month of child birth

(e) Number of dependents **(f)** Religious affiliation

(g) Resident in East Germany **(h)** Employee status

Figure 2.A.2. Distribution of Pre-determined Covariates Around the Kink (Fathers)

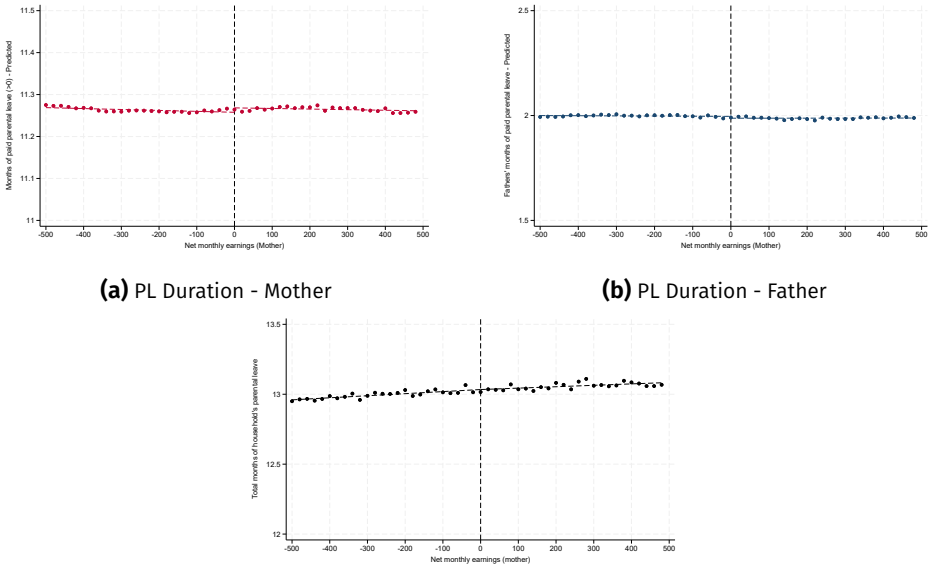
Notes: The graphs test the validity of the smoothness assumptions of the RK design. The sample includes all fathers with reference income within a neighborhood around the kink point. Each panel displays the mean of a different predetermined covariate, binned in €20 intervals of the assignment variable. The visual evidence indicates that covariates evolve smoothly at the kink, supporting the credibility of the RKD identification assumptions.



(a) Withholding Tax Class 1 **(b)** Withholding Tax Class 4

Figure 2.A.3. Distribution of Civil Servants by Gross Income and Withholding Tax Class (Mothers)

Notes: These graphs show the distribution of civil servants along the gross income spectrum for two groups of women, distinguished by their withholding tax class. Panel (a) corresponds to non-married women (Withholding Tax Class 1), while panel (b) represents married women opting for Withholding Tax Class 4 (taxed individually as if single). In both cases, the share of civil servants peaks at around €4,000 of gross monthly income, with a sharp decline beyond that level.



(a) PL Duration - Mother **(b)** PL Duration - Father

(c) PL Duration - Household

Figure 2.A.4. Distribution of Predicted PL Durations Around the Mother's Kink

Notes: The graphs display mean predicted parental leave (PL) durations within €20 bins of pre-birth net earnings for mothers, normalized at the kink point. In all panels, the x-axis represents the mother's reference income. The outcome variables are: (a) predicted PL duration for mothers (conditional on take-up), and (b) predicted PL duration for fathers (conditional on take-up). Predicted values are obtained by regressing PL durations on a set of pre-determined maternal characteristics: an indicator for church tax liability, county (*Kreis*) fixed effects, age, employment status (employee vs. self-employed), marital status, month of birth of the child, and number of dependents.

Appendix 2.B Calculation of parental leave benefits

Since 2013 parental leave benefits are not replacing the actual net income before birth, but rather an estimated net income calculated based on gross income.

Reference Gross income. For individuals in regular employment, the reference income is the average monthly gross income in the 12 months before birth. Months are excluded from the calculation if parental leave was taken in that month, or if the individual was in civil service or compulsory military service. One-off payments such as severance pay or Christmas bonuses are not taken into account. For individuals with self-employed income, the reference income is that of the previous year.

Reference Net income. Parental leave replaces a proportion of the reference net income. The reference net income is calculated by deducting from reference gross income those social security contributions and taxes that are typically withheld by the employer to approximate the actual monthly income that is transferred to the bank account.

Consequently, only the social security contributions that are withheld from the employer are taken into account. So for example, an individual who does not contribute to the unemployment insurance system (like a civil servant) has a higher reference net income and a higher parental leave benefit. In particular, it means that individuals who are voluntarily enrolled in the public health insurance scheme ("gesetzlich freiwillig versichert") do not have their public healthcare payments deducted in the calculation of their reference net income, as these contributions are not withheld by the employer.

The calculation of net income also mirrors the amount of tax withheld. Therefore, church taxes also reduce the reference net income and consequently the parental leave benefit. The most important factor is the chosen withholding tax schedule, as this determines the amount of tax withheld by the employer. Therefore, depending on the chosen tax class, the reference net income and parental leave benefit differ substantially.

References

Andresen, M. E., and E. Nix. 2024. "You Can't Force Me Into Caregiving: Paternity Leave and the Child Penalty." *Economic Journal.* [89, 112]

Angelov, N., P. Johansson, and E. Lindahl. 2016. "Parenthood and the Gender Gap in Pay." *Journal of Labor Economics* 34(3): 545–79. [87]

Asai, Y. 2015. "Parental leave reforms and the employment of new mothers: Quasi-experimental evidence from Japan." *Labour Economics* 36: 72–83. [89, 97]

Avdic, D., and A. Karimi. 2018. "Modern Family? Paternity Leave and Marital Stability." *American Economic Journal: Applied Economics* 10(4): 283–307. [89]

Bailey, M., T. Byker, E. Patel, and S. Ramnath. 2025. "The long-run effects of California's paid family leave act on women's careers and childbearing: New evidence from a regression discontinuity design and US tax data." *American Economic Journal: Economic Policy* 17(1): 401–31. [88, 89]

Bana, S. H., K. Bedard, and M. Rossin-Slater. 2020. "The Impacts of Paid Family Leave Benefits: Regression Kink Evidence from California Administrative Data." *Journal of Policy Analysis and Management* 39(4): 888–929. [89, 97]

Blau, F. D., and L. M. Kahn. 2017. "The Gender Wage Gap: Extent, Trends, and Explanations." *Journal of Economic Literature* 55(3): 789–865. [87]

Boelmann, B., A. Raute, and U. Schönberg. 2025. "Wind of Change? Cultural Determinants of Maternal Labor Supply." *American Economic Journal: Applied Economics* 17(2): 41–74. [89, 109]

Canaan, S., A. S. Lassen, P. Rosenbaum, and H. Steingrimsdottir. 2022. "Maternity Leave and Paternity Leave: Evidence on the Economic Impact of Legislative Changes in High Income Countries." *IZA Discussion Paper No. 15129.* [89]

Card, David, David Lee, Zhuan Pei, and Andrea) Weber. 2015. "Inference on Causal Effects in a Generalized Regression Kink Design." *Econometrica* 83(6): 2453–83. [88, 98–100]

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber. 2017. "Regression Kink Design: Theory and Practice." *Advances in Econometrics* 38: 341–82. [99]

Cools, S., J. H. Fiva, and L. J. Kirkebøen. 2015. "Causal Effects of Paternity Leave on Children and Parents." *Scandinavian Journal of Economics* 117(3): 801–28. [88, 89]

Cortes, P., and J. Pan. 2023. "Children and the Remaining Gender Gaps in the Labor Market." *Journal of Economic Literature* 61(4): 1359–409. [87]

Dahl, G. B., K. V. Løken, M. Mogstad, and K. V. Salvanes. 2016. "What is the case for paid maternity leave?" *Review of Economics and Statistics* 98(4): 655–70. [88, 89]

Destatis. 2018. "Kindertagesbetreuung regional 2018." *Statistische Ämter des Bundes und der Länder.* [92]

Ekberg, J., R. Eriksson, and G. Friebel. 2013. "Parental leave — A policy evaluation of the Swedish "Daddy-Month" reform." *Journal of Public Economics* 97: 131–43. [88, 89]

Geis-Thöne, Wido. 2018. "Familien müssen für die gleiche Betreuung in der Kita unterschiedlich viel zahlen." *IW-Report 50/18.* [93]

Ginja, R., J. Jans, and A. Karimi. 2020. "Parental Leave Benefits, Household Labor Supply, and Children's Long-Run Outcomes." *Journal of Labor Economics* 38(1): 261–320. [97]

Glogowky, U., E. Hansen, H. Lüthen, and D. Sachs. 2025. "The Evolution of Child-Related Gender Inequality in Germany and the Role of Family Policies, 1960-2018." *European Economic Review* 175: 105018. [88, 89]

Gonzalez, L., and L. Farré. 2019. "Does Paternity Leave Reduce Fertility?" *Journal of Public Economics* 172: 52–66. [89]

Ichino, A., M. Olsson, B. Petrongolo, and P. Skogma-Thoursie. 2024. "Taxes, Childcare and Gender Identity Norms." *Journal of Labor Economics.* [90]

Jørgensen, T. H., and J. Søgaard. 2024. "The Division of Parental Leave: Empirical Evidence and Policy Design." *Journal of Public Economics* 238. [90]

Kleven, H. 2024. "The Geography of Child Penalties and Gender Norms: A Pseudo-Event Study Approach." *NBER Working Paper 30176.* [87]

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimuller. 2019. "Child Penalties Across Countries: Evidence and Explanations." *AEA Papers and Proceedings* 109: 122–26. [87]

Kleven, H., J. Posch, A. Steinhauer, and J. Zweimuller. 2024. "Do Family Policies Reduce Gender Inequality? Evidence from 60 Years of Policy Experimentation." *American Economic Journal: Economic Policy* 16(2): 110–49. [88, 89]

Kuziemko, I., J. Pan, J. Shen, and E. Washington. 2018. "The Mommy Effect: Do Women Anticipate the Employment Effects of Motherhood?" *NBER Working Papers 24740, National Bureau of Economic Research, Inc.* [87]

Lalive, R., and J. Zweimüller. 2009. "How Does Parental Leave Affect Fertility and Return to Work? Evidence from Two Natural Experiments." *Quarterly Journal of Economics* 124(3): 1363–402. [88, 89]

Lee, D. S., and T. Lemieux. 2010. "Regression Discontinuity Designs in Economics." *Journal of Economic Literature* 48: 281–355. [107]

Lundborg, P., E. Plug, and A. W. Rasmussen. 2017. "Can Women Have Children and a Career? IV Evidence from IVF Treatments." *American Economic Review* 107(6): 1611–37. [87]

Olivetti, C., and B. Petrongolo. 2017. "The economic consequences of family policies: Lessons from a century of legislation in high-income countries." *Journal of Economic Perspectives* 31(1): 205–30. [87, 89]

Olivetti, C., B. Petrongolo, and J. Pan. 2025. "The Evolution of Gender in the Labor Market." *CEP Discussion Paper No. 2067* 5. [87]

Schönberg, U., and J. Ludsteck. 2014. "Expansions in Maternity Leave Coverage and Mothers' Labor Market Outcomes after Childbirth." *Journal of Labor Economics* 32(3): 469–505. [88, 89]

Waights, Sevrin. 2022. "Parental Leave Benefits and Child Penalties." *SSRN Electronic Journal,* https://doi.org/10.2139/ssrn.4203721. [88]

Ziegler, L., and O. Bamieh. 2023. "What Drives Paternity Leave: Financial Incentives or Flexibility?" *IZA DP No. 15890.* [90, 97]

Chapter 3

On the Extent, Correlates, and Consequences of Reporting Bias in Survey Wages*

Joint with Marco Caliendo, Katrin Huber and Ingo E. Isphording

3.1 Introduction

While economists often disagree on many issues (Kearl et al., 1979; Fuchs, Krueger, and Poterba, 1997), there is a consensus on the importance of wages and personal income. Individual wages are a core component of virtually every model in labor economics and related fields. Economic theory predicts close relationships between wages and individual decisions about labor supply (Pencavel, 1986; Mulligan and Rubinstein, 2008), household investment (Acemoglu, 1997), consumer demand, or savings (Sandmo, 1970). Wage levels are considered an important measure of economic well-being (Deaton, 2008), and are closely linked to other economic outcomes such as education (Becker and Chiswick, 1966), health (Deaton, 2003), and social mobility, and determine tax liabilities and the eligibility for transfer payments.

Applied researchers working on wage-related issues can choose between two main types of data. Register-based administrative data are considered highly reliable, often offering superior measurement quality, larger population coverage, and arguably higher consistency, precision, and more regular updating (Künn, 2015).

* We thank Julian Budde, Jonas Jessen, Nikolas Mittag, Markus Nagler, Michael Oberfichtner, Nico Thurow, Erwin Winkler, participants at the annual meetings of EALE (2023), and Verein für Socialpolitik (2023), and seminars at the University of Potsdam, and the University of Trier for their valuable comments and suggestions. We are grateful to Maximilian Biller and Jessica Fuchs for excellent research assistance. Support by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) through CRC TR 224 (Project A02 and A05) is gratefully acknowledged.

Survey-based self-reported data, on the other hand, provide access to interesting variables that are essential for many modern labor economics applications – personality traits, risk preferences, and cognitive skills are examples – but are typically not available in administrative data (Borghans et al., 2008; Arni et al., 2014; Künn, 2015). Moreover, survey data can be flexibly adapted to the specific needs of researchers (Stantcheva, 2022). As a result, survey data complement or even replace administrative records, depending on the research question. However, the reliability and quality of self-reported wage information has often been questioned (Bound, Brown, and Mathiowetz, 2001).

Against this background, this paper addresses three main research questions. First, we examine the extent to which survey respondents misreport their wages. Second, we analyze which observable characteristics can predict the extent of misreporting. Third, we examine the consequences that applied researchers face when working with these misreported wages. We use the novel and unique record linkage of the SOEP-ADIAB. This dataset combines the German Socioeconomic Panel (SOEP), one of the longest running and most widely used household surveys in the world (Goebel et al., 2019), with high quality administrative data from the Integrated Employment Biographies (IEB) of the Federal Employment Agency, covering all dependent employment in Germany. This unique data linkage provides us with 59,118 linked survey/administrative wage observations on about 5,500 individuals for up to 36 years.

With this novel data setup, we contribute new answers to an established literature that dates back to early studies by Bound and Krueger (1991) and Pischke (1995) and that has received continued attention from researchers (Angel et al., 2019; Antoni, Bela, and Vicari, 2019; Roth and Slotwinski, 2021). The unprecedented scope of data in the SOEP-ADIAB, both in terms of breadth of data and individual time series of linked subjective and administrative wage information, allows us to provide a more detailed characterization of the reporting bias in survey wages than previous studies. The linkage also provides us with a broad portfolio of available control variables at the individual, household, job and firm levels to examine a comprehensive set of predictive factors that goes beyond the existing literature. Finally, we re-examine common empirical relationships using wages as both dependent and independent variables, describing meaningful consequences of choosing self-reported survey wages over administrative records for typical empirical exercises in applied economics.

Our analysis consists of three steps. First, we quantify the extent of reporting bias. Mean wages differ moderately between SOEP and IEB, with SOEP respondents on average underreporting their individual wages by €186, with a mean relative reporting bias of 7.3% . Consistent with previous research (e.g., Bound and Krueger, 1991; Bound, Brown, and Mathiowetz, 2001; Kim and Tamborini, 2014; Christopher R Bollinger et al., 2018), we find evidence of a regression-to-the-mean component in the reporting bias. Low-income individuals in the bottom four ventiles over-

estimate their true wages on average. The reporting bias becomes negative from the fifth ventile and increases steadily with income, both in absolute and relative terms. Rounding contributes only modestly to the reporting bias, and changes in wages appear to be more accurately captured by survey responses than levels. The results are robust to several alternative wage concepts and definitions.

Second, we analyze which factors at the individual, household, job and firm levels can predict reporting bias. At the individual level, only a few characteristics such as gender and extroverted personality play a role in explaining the variation in reporting bias. At the household level, partners' income in particular in interaction with one's gender, predicts the direction and magnitude of the reporting bias, suggesting that social dynamics within households influence the reporting bias, as recently shown by Roth and Slotwinski (2021). The strongest predictors are job and firm level characteristics, with union membership, firm size, average wage level, and workforce composition strongly correlated with reporting bias.

Third, we examine the consequences of using survey rather than administrative wages in commonly analyzed economic relationships. We consider two settings in which wages are the dependent variable: returns to education and the gender wage gap. In the case of returns to education, the choice of data appears to be largely inconsequential; estimates based on SOEP and IEB wages are similar in magnitude, and a two-sample GMM t-test rejects significant differences in all but one specification. In the case of the gender wage gap, the significant relationship between gender and reporting bias leads to a small but significant overestimation of the gender wage gap that links to systematic variation of misreporting by gender. Estimates differ moderately, with the conditional gap shrinking from 10.7% to 7.7% based on IEB data, with a two-sample GMM t-test confirming significance of this difference.

To examine the sensitivity of the relationships with wages as the independent variable, we estimate wage effects on self-reported satisfaction. We show that using self-reported wages leads to a substantial overestimation of the effect of wages on subjective indicators such as satisfaction with personal income. For example, while satisfaction with personal income increases with an additional €1,000 by about 25.4% of a standard deviation when assessed on the basis of the SOEP, this relationship shrinks by a quarter to only 20.2% of a standard deviation when assessed on the basis of IEB data. This bias is exacerbated with extended control variables, but mitigated by relying on inter-personal changes. The difference runs counter to the intuition of attenuation bias by classical measurement error, but rather hints at the role of systematic correlates of the measurement error (Bertrand and Mullainathan, 2001).

Taken together, our results should be read as a cautionary tale for researchers relying on self-reported measures of income and wages. While the reporting bias in wages turns out to be moderate, it is systematically related to several characteristics, and thus can – depending on the application – significantly bias commonly estimated empirical relationships. This is especially true when wages are used as

explanatory variables to explain subjective measures such as well-being, concerns, or attitudes. In these cases, subjective measures of the same individual enter the estimation simultaneously as both dependent and explanatory variables. This may exaggerate problems of omitted variables driving the relationship, e.g. through individual-specific reporting behavior or reported wages reflecting desired or aspired instead of actual wages (Bertrand and Mullainathan, 2001; Prati, 2017).

Despite these implications, representative survey datasets such as the SOEP remain indispensable for applied economics, underpinning many influential studies and reports. Survey data allows to augment potentially more reliable information from administrative sources with information on attitudes, preferences, noncognitive skills, thus allowing for more far reaching explorations of the determinants of wages. However, given the sensitivity of results to misreporting bias, findings should be corroborated with more reliable income measures whenever possible. Where such measures are not available, the results should be interpreted with appropriate caution. We see the need for further research to extend our exemplary analyses of common empirical relationshops – focusing on the returns to education, the gender gap, and the wage-satisfaction relationship – to additional research questions. This will help to get a clearer picture of the importance of the problem of misreporting and will further highlight the potential benefits of linked administrative-survey data.

Our analysis contributes to a long-running literature using linked data sources to assess the reliability of survey wages (see Moore, Stinson, and Welniak, 2000; Bound, Brown, and Mathiowetz, 2001, for a comprehensive review). In a seminal study, Duncan and Hill (1985) describe substantial bias in the comparison of administrative records with employee survey responses from a manufacturing firm. Subsequent studies (e.g., Bound and Krueger, 1991; Bound et al., 1994; Pischke, 1995) extend this cross-sectional single-firm analysis to multiple firms over two waves. Overall, the results confirm the existence of reporting errors and show that they are correlated over time (Bound et al., 1994). Other studies have examined the reliability of survey responses beyond self-reported wages, e.g., hours worked (Duncan and Hill, 1985), occupation (Isaoglu, 2010), pensions and unemployment benefits (Angel et al., 2019; Bollinger and Tasseva, 2023), or loans and debts (Madeira et al., 2022). Several studies have extended the originally mostly US-based literature to test the reliability of German survey data (Antoni, Bela, and Vicari, 2019; Valet, Adriaans, and Liebig, 2019; Schmillen, Umkehrer, and Wachter, 2024). We contribute to these studies by relying on an unusually large sample, both in number of observations and length of individual time series, from a fundamental dataset in economic research. The scope of our data allows for more in-depth analysis, a broader set of potential correlates of reporting bias, and the analysis of changes in reporting bias over time. Compared to previous results, we find established patterns of mean reversion (e.g., Bound et al., 1994; Pischke, 1995), while the extent of mis-

reporting is rather moderate, nonetheless comes with substantial consequences for applied work.

Second, we complement several previous studies by examining which individual-, household-, job- and firm-level factors are correlated with reporting bias. Duncan and Hill (1985) show a correlation between misreporting and job tenure, Kim and Tamborini (2014) find that education, race, and wage level are associated with reporting bias. Antoni, Bela, and Vicari (2019) explore whether reporting bias could be caused by socially desirable reporting and whether the error is affected by interview characteristics. Angel et al. (2019) also find evidence for social desirability as a possible explanation for reporting bias. Finally, Roth and Slotwinski (2021) investigate whether partner income influences wage reporting. Most recently, Meyer, Mittag, and Wu (2024) provide a literature review summarizing patterns of greater measurement error among minorities. Because we are able to draw on the breadth of characteristics covered by the SOEP, in addition to the firm-level variables drawn from the IEB, we contribute to this area of research by jointly estimating the predictive power individual-, household-, job- and firm-level characteristics and examining the relative contribution of explained variance between these groups.

Finally we contribute to a literature that assesses the consequences of misreported wages for the estimation of economic parameters. Already Solon (1992) described the role of misreported income for the estimation of intergenerational mobility. Kim and Solon (2005) found that reporting errors can lead to misleading conclusions about the procyclicality of real wages. Gottschalk (2005) and Gottschalk and Huynh (2010) analyze the effect of misreported survey wages on the probability of estimating nominal wage cuts. Hurst, Li, and Pugsley (2014) show how misreporting may hinder the accurate estimation of self-employed income. Meyer and Mittag (2019) demonstrate substantial effects of misreported wages on the estimation of program participation. Bossler and Westermeier (2020) show that misclassifications of minimum wage eligible workers based on survey wages can bias treatment effects by up to 30%. Roth and Slotwinski (2021) identify the consequences of misreporting due to social norms for estimating the gender wage gap. Stüber, Grabka, and Schnitzlein (2023) find differences in aggregate measures of wage inequality between unlinked survey data and administrative data in Germany. We complement this literature by demonstrating the importance of misreporting for several commonly analyzed economic relationships relying on wage as either dependent or independent variable.

The rest of the paper is as follows. In chapter 3.2 we provide a detailed description of our unique dataset. We continue by showing the extent of the reporting bias in chapter 3.3.1 and provide the empirical analysis of correlating factors in section 3.3.2. Chapter 3.3.3 discusses the implications for economic analyses and chapter 3.4 concludes.

3.2 Data

3.2.1 Data Sources

Our analysis is based on a novel linkage of individual- and firm-level data from German social security records with survey data on individuals and households in the German Socio-Economic Panel (SOEP-ADIAB).[1] Starting in 2019, SOEP respondents were asked for their consent to retrospectively link their survey responses with their employment histories (*Integrierte Erwerbsbiographien*, IEB).

The IEB provide detailed and accurate data on the employment and unemployment histories of individuals in Germany, capturing labor market outcomes such as exact gross wages for each job held. The dataset covers employees subject to social security contributions, workers in marginal employment[2], recipients of social security benefits, the registered unemployed, and participants in active labor market programs. It covers both individual- and firm-level data, such as the size of the establishment, the demographics of the employees, and industry sector. However, it does not cover civil servants or the self-employed. In addition, wage data in the IEB are top-coded at the assessment ceiling for social security contributions in Germany. Notably, top-coding affects only 5.2% of all employment spells between 1975 and 2019, although this rate varies across subgroups and has increased over time.

The German Socio-Economic Panel (SOEP) has been conducted annually since its inception in 1984 and currently includes about 30,000 respondents from 15,000 households. The SOEP contains detailed information on self-reported gross wages. In addition to wages, it collects extensive self-reported data on individual and household characteristics such as personality traits, risk preferences, worries, beliefs, and household composition.

By linking the IEB and SOEP, we can directly compare individuals' administrative and self-reported wages to assess the extent of reporting bias. In addition, we can examine factors that are correlated with reporting discrepancies at the individual, household, job and firm levels. In the following, we discuss details of the sampling process and data linkage underlying our empirical analysis.

1. Our analysis relies on a pilot version of the SOEP-ADIAB linked data, which preceeded the now publicly available SOEP-CMI-ADIAB (Antoni et al. (2023)), and is largely identical. We use the SOEP Core v37 (SOEP (2022)), and the SOEP IS 2019 (SOEP-IS (2020)).

2. Employees in Germany, with the exception of the self-employed and civil servants, and their respective employers are obliged to pay social security contributions. Employees earning up to a certain amount per month (in our observation period this threshold varied between approximately €200 in 1984 and €450 in 2020) are exempt from paying social security contributions. This type of employment is called marginal employment (or "minijobs").

3.2.2 Data Cleaning, Linkage and Sample Restrictions

Data Linkage. We prepare spells of the administrative information in the SOEP-ADIAB linkage data using standard procedures following Dauth and Eppelsheimer (2020). The preparation consists of splitting overlapping spells, creating biographical variables, merging information from the Establishment History Panel (EHP), and cleaning occupational and educational information. To this data, all respondents to the SOEP are prepared for linkage who consented to this procedure. We include both individual respondents from the SOEP core sample and the SOEP IS sample, an independently drawn sample used to pilot new questions and running since 2011 (Richter and Schupp, 2015).

Unlike the annual SOEP data, the IEB data are stored in spell form. Each work spell is stored with a start date and an end date. We overcome this mismatch in the data structure by linking consenting SOEP observations to all administrative spells that either overlap with the SOEP interview or ended in the month prior to the interview. If SOEP interviews coincide with more than one parallel IEB spell, indicating multiple jobs at the same time, we sum wages from all concurrent spells. This linkage results in an unbalanced panel structure with one observation per individual per year. We link additional establishment-level information from the EHP to each respective administrative spell.[3]

Sample Restrictions. Table 3.1 provides details on the number of successful linkages and the subsequent sample restrictions required. Of the 23,525 observations being asked about linkage in 2019, 15,012 individuals consented.[4] Of the consenting respondents, 14,983 can be successfully linked to their social security records. These linked individuals have at least one spell in the IEB data set, i.e., they were in dependent employment, active labor market policies or officially registered as unemployed at least once during their working life. Rows 4 to 9 document how many observations we can observe, conditional on the necessary sample restrictions. Among the 14,983 successfully linked individuals, 7,614 respondents have a concurrent work spell in the IEB data in 2019 and thus serve as the basis for our sample. In the SOEP questionnaire, individuals are asked to report their last month's wage income. Accordingly, we restrict the sample to those individuals who, at the time of the interview, have a work spell that started at least 30 days before the interview

3. In the case of multiple spells, we set establishment-level information to missing, leading to slightly reduced numbers of observation in some analyses.

4. Table 3.B.1 in the Appendix describes differences in covariates by consent. Individuals who consented to be linked and those who refused to be linked do not differ systematically on most basic characteristics, and when they do, the differences are only very small. Similar findings are reported by Sakshaug and Kreuter (2012) for the case of linkage of IEB to the "German Labour Market and Social Security" (PASS) study. Selective consent appears to be small, especially for substantive variables like employment, income and benefit recipiency.

to ensure that this survey and the administrative wage measure the same job and period.

We impose several additional sample restrictions. First, we drop individuals who report only self-employment in the SOEP but still have income from dependent employment in the IEB. For these individuals, we do not know whether their SOEP wage includes only income from self-employment or whether they have also income from dependent employment. Since income from self-employment cannot be measured in the IEB, this would mechanically inflate the SOEP wage. Second, we drop individuals below the age of 20 or above the age of 65 in order to restrict our sample to the working age population. We condition the sample on individuals with SOEP and IEB wages greater than zero, thus entirely focusing on misreporting bias on the intensive margin. We acknowledge, though, that income nonresponse is a further source of misreporting (C. R. Bollinger et al., 2019).[5,6] Finally, we restrict our analysis to individuals with neither censored administrative wages nor survey wages that exceed the assessment limit for the social security contributions.[7] These restrictions leave us with 6,136 individuals for whom we observe survey and administrative wages, with about 10% based on the shorter-running SOEP IS sample. On average, we observe respondents for about ten waves, with the longest running individual linked histories going back to the very first SOEP wave in 1984 (Figure 3.B.1 in the Appendix). This leaves us with 59,118 individual-year observations. The number of observations in various steps of the subsequent analysis may very further due to item non-response in the covariates.

Definition of the Reporting Bias. We define reporting bias as the difference at the individual level between equivalent measures of gross monthly wage in both the SOEP and the IEB. In the SOEP, respondents are asked to report the gross monthly wage they have earned in the month prior to the interview. During the interview, respondents are also allowed to look at their pay slip if they do not know their wage. They are asked not to include special payments such as vacation pay or back pay, but to add overtime compensation.[8] Based on this information, we construct the gross

5. The IEB wage is reported as 0 in cases of "employment interruptions". During these periods, the employment relationship continues legally, but no remuneration is paid, e.g., illness after the end of continued pay, maternity leave and sabbaticals (Frodermann et al., 2021).

6. We also drop individuals in informal employment, i.e., who are listed with a zero wage in the administrative data, but have a positive wage from non-self-employment in the SOEP. This amounts to only about 0.35% of the spells in the merged sample before any other restrictions are imposed.

7. Following Dauth and Eppelsheimer (2020), we drop incomes with an error margin of €120 surrounding the assessment limit.

8. The exact wording in the 2019 English version of the SOEP is as follows: "*What did you earn from your work last month? Please state both: gross income, which means income before deduction of taxes and social security and net income, which means income after deduction of taxes, social security, and unemployment and health insurance. If you received extra income such as vacation pay or back pay, please do not include this. Please do include overtime pay. If you are self-employed: Please estimate your monthly income before and after taxes.*"

Table 3.1. Cross-section of individuals interviewed in 2019

		Unique Individuals by Sampling Step		
Sampling stage	Restriction	Core	IS	Total
1.	Contacted	20,525	3,000	23,525
2.	Have consented	13,166	1,846	15,012
3.	Linked	13,154	1,829	14,983
4.	Working spell within 30 days	6,903	711	7,614
5.	No pure self-employment spell (SOEP)	6,835	711	7,546
6.	Age 20-65	6,497	680	7,177
7.	SOEP and IEB wages > 0	5,997	622	6,619
8.	IEB wages < assessment limit - €120	5,592	584	6,176
9.	SOEP wages < assessment limit - €120	5,557	579	6,136
	Unique individual x year observations		59,118	

Notes: This table summarizes the number of observations in the linked SOEP - ADIAB dataset at different sampling-stages. Observation numbers are displayed separately for the SOEP Core sample and the SOEP Innovation Sample (IS). Rows 4 to 9: cross-section 2019.

monthly survey wage. In robustness checks, we construct alternative survey wage measures that include information on annual wages instead of monthly wages, as well as information on special payments (for details, see section 3.3.1 and Appendix 3.A).

The IEB data include administrative information on gross *daily* wages. These are calculated by dividing the employer's information on the total wage of the employment period by the number of calendar days in the same period. In the case of parallel spells linked to a single SOEP observation, we compute the sum of gross daily wages, as the SOEP survey question aims at income from any work and is thus not restricted to wage income from the main job. While bonus payments are included in the IEB wages, the reporting of special payments to employees is not mandatory. In order to check the extent to which such extra payments might affect our results, we perform robustness checks as described in Appendix 3.A. Two limitations of the IEB wage information have to be taken into account: First, until 1998, wages that were below the marginal employment threshold were not reported and appear as a zero wages in the administrative data. Consequently, the corresponding SOEP observation is also dropped. Second, wages are winsorized at the assessment limit for social security contributions. While there are ways to impute wages above the assessment ceiling, we refrain from doing so because this imputation procedure would artificially affect the measurement error as our key variable of interest. Instead, winsorized observations are excluded from the analysis.

To construct a *monthly* measure of wages from the reported daily wages, we follow the common practice in IEB-based studies of multiplying the daily wage by $\frac{365.25}{12} = 30.4375$, which is the average number of days in each month of a year, taking into account leap years. We report the *absolute reporting bias* as the simple difference between the wage measured in the SOEP and the wage measured in the

IEB data. We also construct a *relative reporting bias* by dividing the absolute reporting bias by the administrative gross wage at the individual level. In addition, we classify individuals according to whether they "overreport" or "underreport" their survey wage relative to the administrative wage. In this classification, we omit individuals who are close to reporting correctly, defined as deviating by less than 2.5% based on the annuals distributions of reporting bias. In robustness checks, we vary this definition from 1.25% to 5% around zero bias, and the results remain qualitatively and quantitatively unchanged.

3.3 Empirical Analysis

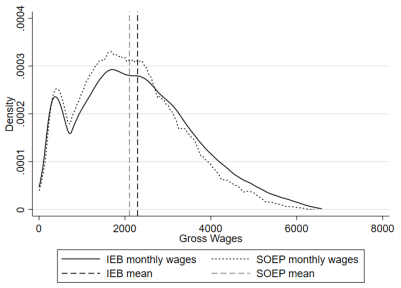
3.3.1 The Extent of the Reporting Bias

We first document the extent and heterogeneity of the reporting bias between the SOEP and the IEB. Figure 3.1a shows that the administrative and self-reported wages follow similar but not perfectly matched distributions. The means of both distributions differ only slightly between the IEB (€2,292) and the SOEP (€2,106). Both distributions show a clustering of wages at time-varying earnings thresholds for jobs exempt from social security contributions (marginal employment). Self-reported SOEP wages appear to be more centered around the mean than administrative wages, with fewer people reporting very high or very low wages. Figure 3.1b illustrates the distribution of reporting bias which is closely centered around zero. Around 71% of SOEP respondents underreport their wages, with an average underreporting of about €186 (dashed line).[9] Despite extreme cases of underreporting and overreporting by more than €5,000, such outliers are rare and do not significantly affect the average reporting bias. Restricting the analysis to biases between the 1st and 99th percentiles of each year's distribution does not significantly change the mean reporting bias, which is then €190.

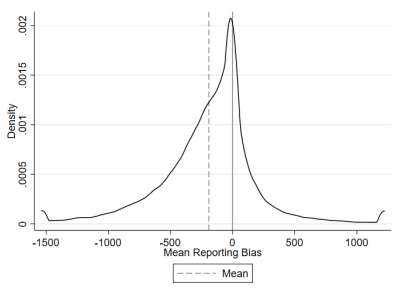
Figure 3.1c shows the reporting bias across the IEB wage distribution in both absolute and relative terms. In particular, there is significant overreporting at very low wages (below the 4th ventile). The mean bias then becomes negative and steadily larger along the income distribution. This is consistent with previous findings in the literature that find evidence for regression-to-the-mean in different settings (Bound and Krueger, 1991; Kapteyn and Ypma, 2007; Christopher R Bollinger et al., 2018, e.g.,), which is to some degree mechanical, with more scope for over-reporting among low income individuals, and more scope for under-reporting among high-income individuals.[10]

9. Only for the graphical representation of the reporting bias in Figure 3.1b biases outside the 1st and 99th percentiles are binned to the 1st and 99th percentiles for clarity.

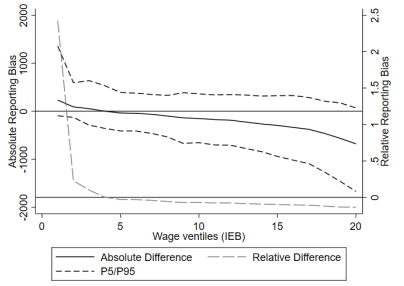
10. Figure 3.B.2 in the Appendix shows no systematic age-related trends in wage reporting bias; it remains stable across different ages.



(a) Distribution of gross monthly wages



(b) Distribution of the reporting bias



(c) Reporting bias over the income distribution

Figure 3.1. Distribution Of SOEP And IEB Wages and the Reporting Bias

Notes: Panel (a) plots the distribution of wages in SOEP and IEB (winsorized at €7,000). The mean reporting bias in IEB data is €2,292, and €2,106 Panel (b) shows the density distribution of the reporting bias, defined by the difference of SOEP and IEB wages. Reporting biases that lie below the 1st percentile or above the 99th percentile of the distribution of differences are binned into the 1st or 99th percentile respectively. Panel (c) shows the mean reporting bias and the respective 5th and 95th percentile over each IEB wage ventile in absolute (black) and relative terms (grey).

Table 3.2 provides further insight into the reporting bias by IEB wage ventile to quantify the graphical representation of Figure 3.1. Column (1) shows the mean IEB wage of each ventile, expressed in Euros. For each wage ventile, the table shows the minimum, maximum, mean and different quantiles (5th, 50th, 95th) in Euros. It also shows the proportion of over- and underreporting individuals. These results indicate a mean reversion in misreporting, with low-wage individuals overreporting and high-wage individuals underreporting their income. The mean absolute reporting bias is positive in the first four wage ventiles and then becomes negative from the fifth ventile onward. In relative terms (column 8), the reporting bias is most pronounced in the very low income group, with individuals on average overreporting their wages by more than 200% in the first ventile and by more than 20% in the second ventile. However, this pronounced mean reporting bias among very low wage individuals is largely due to outliers who overreported by more than €5,000. Across the wage distribution, the mean relative reporting bias increases only moderately, up to -13.5% for individuals in the highest wage ventile. Accordingly, the proportion of individuals who overreport their wages decreases from 57.4% in the first wage ventile to 6.7% in the last ventile. Overall, Table 3.2 corroborates the

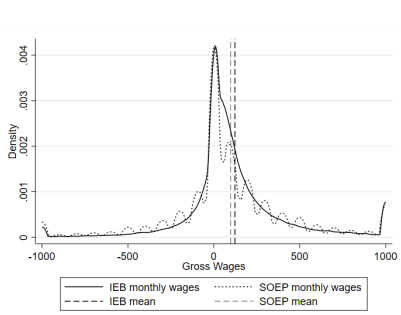
Table 3.2. Reporting Bias per Wage Ventile

		Reporting Bias										
Ventile	Mean IEB wage in €	Min in €	P5 in €	Mean in €	Median in €	P95 in €	Max in €	Avg. rel. in %	Over in %	Under in %	Zero Bias in %	N
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
1	249	-513	-94	232	23	1,356	5,127	242.5	57.4	25.4	17.2	2,985
2	439	-590	-127	91	1	594	5,151	22.9	47.0	28.8	24.2	2,989
3	682	-761	-289	53	-6	637	4,807	10.1	37.7	50.5	11.7	2,912
4	973	-1,101	-357	1	-37	532	3,114	0.3	34.6	58.7	6.7	2,952
5	1,207	-1,203	-411	-36	-59	391	4,121	-2.8	30.1	64.9	5.0	2,957
6	1,413	-1,352	-412	-44	-66	377	3,453	-3.2	30.0	64.8	5.2	2,960
7	1,593	-1,452	-463	-64	-72	349	2,784	-4.0	29.5	65.0	5.6	2,953
8	1,772	-1,670	-536	-101	-101	329	2,987	-5.7	24.7	71.3	4.0	2,959
9	1,948	-2,186	-669	-138	-124	387	3,026	-7.1	24.9	71.7	3.4	2,960
10	2,130	-2,424	-654	-152	-145	361	3,230	-7.1	25.5	72.1	2.4	2,944
11	2,310	-2,188	-702	-173	-170	341	2,458	-7.6	21.8	75.6	2.6	2,962
12	2,495	-2,342	-707	-186	-188	347	2,763	-7.5	20.0	77.8	2.1	2,950
13	2,680	-2,770	-777	-226	-221	337	2,492	-8.6	17.7	80.2	2.1	2,963
14	2,876	-2,297	-843	-265	-267	317	1,862	-9.3	15.5	83.1	1.5	2,951
15	3,073	-3,136	-942	-295	-297	325	2,105	-9.7	16.6	81.9	1.5	2,959
16	3,296	-2,674	-1,018	-336	-324	327	2,263	-10.2	15.1	83.8	1.2	2,954
17	3,557	-2,406	-1,091	-377	-369	286	1,852	-10.7	13.4	85.9	0.8	2,958
18	3,879	-3,571	-1,266	-461	-414	210	1,577	-11.9	11.6	87.7	0.7	2,954
19	4,314	-4,321	-1,462	-566	-511	173	1,318	-13.2	9.7	89.5	0.7	2,958
20	5,001	-5,121	-1,670	-679	-605	70	774	-13.5	6.7	92.9	0.4	2,938
All	2,293	-5,121	-921	-186	-138	373	5,151	7.3	24.5	70.5	5.0	59,118

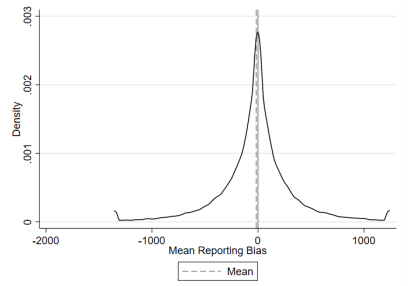
Notes: This table reports summary statistics by IEB wage ventiles. Column (1) depicts the mean IEB wage of each ventile, measured in €. Columns (2) - (11) show various information on the reporting bias, separately for each wage ventile. The minimum reporting bias in column (2), the 5th percentile value of the reporting bias in (3), the mean and median reporting bias in (4) and (5), the 95th percentile value of the reporting bias in column (6) and the maximum reporting bias in (7). Column (8) shows the average relative reporting bias as compared to the mean wage of each ventile in percent. Column (9) to (11) depict the proportion of individuals overstating their wage, understating their wage and reporting correctly within a +/- 2.5 % range around the zero bias. Column (12) gives the number of observations for each ventile.

graphical representation of Figure 3.1 and highlights an overall modest reporting bias that nevertheless varies along the income distribution. The first central result of our analysis is thus a modest but economically significant difference in mean reported wages between the SOEP and the IEB, which averages 7.3% (see column (8) in Table 3.2, last line).

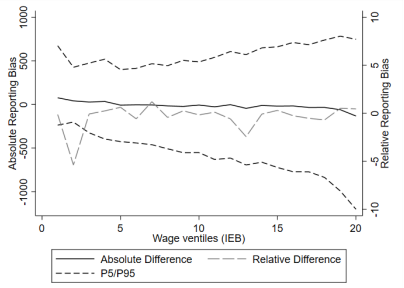
Figure 3.2 shows the reporting bias in wage changes using first differences, thereby examining the accuracy of year-to-year changes in wage income. Panel (a) compares the distributions of these first differences between SOEP and IEB and shows a high degree of similarity. The similarity leads to a narrow and symmetric distribution of deviations, as shown in Panel (b). In Panel (c), we observe that this minimal reporting bias in changes remains consistently close to zero across the entire wage distribution. Consequently, we find that the reporting bias between subjective SOEP wages and administrative IEB wages is significantly smaller when assessing changes rather than levels, suggesting time-invariant reasons for the re-



(a) Distribution of gross monthly wage changes



(b) Distribution of the difference in wage changes



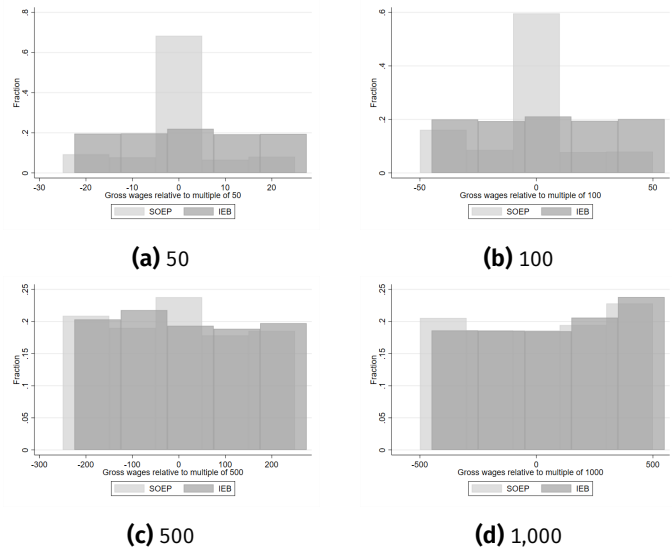
(c) Reporting bias over the distribution in wage changes

Figure 3.2. Reporting Bias in Wage Changes

Notes: Panel (a) plots the distribution of changes in SOEP and IEB (winsorized at ± €1,000) from one year to an other. Panel (b) shows the density distribution of the difference in wage changes in SOEP and IEB. Reporting biases that lie below the 1st percentile or above the 99th percentile of the distribution of differences are binned into the 1st or 99th percentile respectively. The mean (median) difference in the change is €-16.1 (€-3.7). Panel (c) shows the difference in wage changes and the respective 5th and 95th percentile over each IEB wage ventile in absolute (black) and relative terms (grey). All three figures are based on 46,196 year-to-year differences.

porting bias. This contrasts with the results from the seminal study by Bound and Krueger (1991) and Schmillen, Umkehrer, and Wachter (2024), who found that the reliability of matched CPS and administrative data declines when first differences are analyzed. As we will show later in Section 3.3.3, controlling for individual fixed effects thus turns out to be a remedy to mitigate the consequences of using misreported income.

Robustness to Other Wage Definitions. In Figure 3.A.1 in the Appendix we examine the robustness of our results to various changes in the SOEP wage definition. We find that adding one-time payments (e.g., Christmas bonuses or a 13th paycheck) or income from additional jobs, or subtracting other sources of income (e.g., self-employment or work in family businesses) does not change our conclusions quantitatively or qualitatively. We also show that using self-reported wage income in the previous calendar year, where participants are asked to report the months they



(a) 50 **(b)** 100

(c) 500 **(d)** 1,000

Figure 3.3. Rounding of Gross Monthly Income in Both Datasets

Notes: This figure plots histograms of gross monthly income in SOEP (dark grey) and IEB (light grey) surrounding full increments of €50, 100, 500, 1000€. The underlying data is generated by taking the difference between the reported wage income rounded to the next €50/100/500/1,000 and the actually reported wage income.

worked as an employee and their average gross monthly wage, leads to comparable results.

Rounding. One likely explanation for differences between self-reported and administrative wage data is rounding. Survey respondents may rely on rounded wages for convenience, which reduces cognitive load but leads to noisy information. Instead, administrative data are reported by the employer based on the official payroll. Wages based on administrative data should therefore be accurate and unbiased by rounding.

Figure 3.3 plots the fraction of SOEP and IEB wages, that surround full increments of €50 in panel (a), of €100 in panel (b) of €500 in panel (c) and of €1,000 in panel (d). The subfigures show that SOEP wages indeed show a strong excess probability for reporting "round" numbers, while IEB wages show evenly distributed probabilities. Panels (a) and (b) show that about 60% of individuals round to the nearest €50 or €100, panel (c) shows that almost 25% round to the nearest €500. Rounding to the nearest €1,000 is not more pronounced in either dataset.

To examine the extent to which rounding explains income misreporting, we plot the reporting bias after excluding "rounded" SOEP wages (Figure 3.4). Rounding does indeed partially explain misreporting, with the resulting distribution being more closely centered around 0. However, the proportion of underreporting (overre-

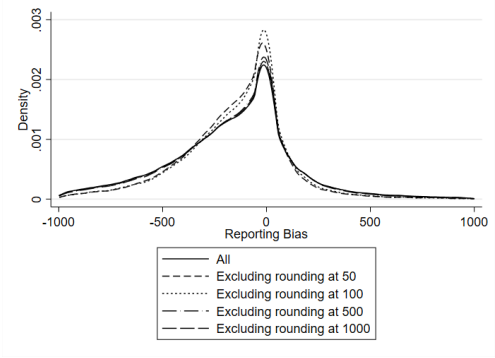


Figure 3.4. Distribution of Reporting Bias Excluding Rounding

Notes: Graph shows the density distribution of the reporting bias for all differences and for differences excluding rounding at €50, 100, 500 and 1,000. Observations with mean reporting bias larger than €1,000 are excluded for graphical reasons.

porting) individuals decreases only slightly from 24.5% to 23.1% (70.5% to 70.3%). Thus, rounding contributes only slightly to the observed total misreporting.

3.3.2 Correlates of Reporting Bias

3.3.2.1 Empirical Strategy

In the previous section, we have described moderate but economically significant deviations between SOEP and IEB wages. We now proceed to analyze factors that may be predictive of these deviations. The richness of our linked dataset allows us to simultaneously examine several groups of such potential correlates of the reporting bias: individual-level, household-level, and firm-level correlates.[11] To investigate the correlation of these groups of variables, we estimate a simple linear regression:

$$Y_{it} = \beta_0 + \beta_1 X'_{i(t)} + \xi_t + u_{it} \tag{3.1}$$

As the dependent variable Y_{it}, we use different operationalizations of the reporting bias. First, we define a binary indicator that takes the value of 1 if the reporting bias is positive, i.e., if the individual overreports her wage in the SOEP compared to her administrative wage in the IEB. The indicator takes the value of 0 in the case of correct reporting or underreporting. Correct reporting is defined as no or small reporting bias within a +/- 2.5 percentile range of the annual reporting bias.[12] Second, we use the log reporting bias, the difference in log wages between SOEP and

11. We also examine the role of survey methodology factors, as previously examined by Angel, Heuberger, and Lamei (2018), Angel et al. (2019), and Antoni, Bela, and Vicari (2019). In particular, we find little influence of panel tenure or interview mode. The results are summarized in Table 3.B.1 in the Appendix.

12. Applying alternative bounds at the 1.25th or 5th percentile does not change our conclusions (see Tables 3.B.0 and 3.B.0).

IEB. For this second outcome variable, and to facilitate interpretation of the estimated coefficients, we also run split-sample regressions separately for individuals who overreport and who underreport their wages in the SOEP.[13] Our main parameters of interest are estimated as the coefficient vector β_1 for the range of potential correlates $X'_{i(t)}$ of the reporting bias, including individual, household, and firm characteristics. We also examine time-fixed effects ξ_t for trends in the reporting behavior.

As individual-level correlates, we include socioeconomic characteristics: gender, age, migration background (none, direct, indirect), years of education, and region of residence. We also assess personality using standardized measures of the Big 5 (openness, conscientiousness, extroversion, agreeableness and neuroticism). All individual-level variables are drawn from the SOEP.

As correlates at the household level, we include indicator variables for being married, having a partner (also interacted with gender), household size, and the number of children.

Finally, we include firm-level correlates. This set of variables includes indicators for the size of the establishment (less than 10, between 10 and 50, between 50 and 250 and more than 250) and the establishment's median wage (lower, middle, or higher tercile), the industry, and indicators for the composition of the workforce (average age of employees, share of highly qualified, female, full-time, and German employees in the establishment). Establishment-level variables are based on IEB administrative data aggregated to the establishment level. In addition, we further assess individual job characteristics using indicators for working hours (working less than 35 hours, between 35 and 44 hours, and more than 44 hours per week), an indicator for being a blue-collar or white-collar worker, and an indicator for trade union membership. These indicators are derived from the SOEP data.

Shapley Decomposition for Relative Variable Importance. To determine the relative importance of different groups of predictors, we apply a Shorrocks-Shapley decomposition to a regression based on Equation (3.1) with the log reporting bias as the dependent variable. This technique decomposes the coefficient of determination R^2 into contributions from individual or groups of explanatory variables.[14] Intuitively, the Shorrocks-Shapley decomposition considers the marginal effect on the outcome of eliminating each of the contributing factors in turn, and then assigns to each factor the average of it's marginal contributions in all possible elimination sequences (Shorrocks, 2013). Unlike a simple incremental R^2, which depends

13. The interpretation of the log difference per se is not straightforward, as a negative sign of the estimated coefficient can imply both a decrease in a positive bias and an increase in a negative bias. For this reason, we estimate this difference separately for individuals with a negative and a positive bias.

14. The decomposition is implemented by using the *Shapley2* post-estimation command to compute the decomposition of the R-squared of the model based on Juarez (2012).

on the order in which the variables are added to the model, the Shorrocks-Shapley decomposition is robust to the order of the variables.

3.3.2.2 Results

In this subsection we discuss the sources of reporting bias estimated by Equation 3.1. The results are presented in Table 3.3, Appendix Table 3.B.1 additionally documents the coefficients for survey characteristics as well as year and sector dummies. In column (1) of Table 3.3, the dependent variable is an indicator of whether an individual has a positive reporting bias. The indicator is equal to one if the difference between the survey and administrative wage is larger than zero, i.e., if the individual overreports the own wage in the survey as compared to the administrative data.

In column (2), the dependent variable is defined as log reporting bias. To facilitate the interpretation of reporting bias coefficients, in columns (3) and (4) we also report the estimates separately for overreporters and underreporters, excluding correctly reporting individuals. In addition, the bottom panel of Table 3.3 summarizes the results of the Shapley decomposition, which are reported in more detail in Table 3.B.0 in the Appendix.

Individual Characteristics. The Shorrocks-Shapley decomposition (Table 3.B.0) shows that individual-level variables have a relatively small contribution to the explained variation in the log reporting bias: only about 3.9% of the R-squared can be attributed to these characteristics. The results in Table 3.3 can be read as follows: women are 2.1 percentage points less likely than men to overreport their wages (column 1), and have a smaller difference in log wages by about 1.6 percentage points (column 2). The coefficient on the female dummy is negative but insignificant in column (3), suggesting that there is no difference in the reporting bias for overreporting women and men. In column (4), however, we see that women who underreport their true wage, underreport by 0.72 percentage points more on average than men who underreport. Previous findings on gender differences in wage reporting have been inconclusive. Our results are consistent with the findings of Bound and Krueger (1991), Bollinger (1998), Angel et al. (2019), Antoni, Bela, and Vicari (2019), and Roth and Slotwinski (2021), but differ from Bricker and Engelhardt (2008), Kim and Tamborini (2014), and Angel, Heuberger, and Lamei (2018), who find no gender differences in reporting bias.

For more educated individuals, we observe a lower but insignificant probability of overreporting (column 1), and a lower negative reporting bias (column 4). Again, previous findings on the relationship between education and reporting bias are mixed: Bricker and Engelhardt (2008), Angel, Heuberger, and Lamei (2018), and Angel et al. (2019) also find a positive correlation between education and the reporting bias, while Bound and Krueger (1991) and Gottschalk and Huynh (2005) find no correlation and Kim and Tamborini (2014) and Antoni, Bela, and Vicari

(2019) find a negative correlation. Our results also suggest little regional variation in the reporting bias. Beyond these basic socioeconomic variables, personality traits of the worker also have little predictive power for reporting bias. One exception is individuals with higher extroversion, who are more likely to overstate their wages, in line with the findings of Antoni, Bela, and Vicari (2019).

Household characteristics. Similar to the individual predictors, household-level variables are, on average, weak predictors of reporting bias. Only 15.6% of the explained variation in the log reporting bias can be associated with these characteristics (see Table 3.B.0). The most important household characteristic is partner income. Conditional on having a partner, a difference of €1,000 in partner income increases the probability of being an overreporter by about half a percentage point (column 1). This effect may be linked to the findings of Roth and Slotwinski (2021) for Austria and Switzerland, who found that couples in which women out-earn their partners systematically underreport their income to conform to the male breadwinner model (Bertrand, Kamenica, and Pan, 2015). Instead, our results suggest a tendency to equalize reported income, instead of having a tendency to maintain income rankings in the household.

Job and Firm Characteristics. Job and firm characteristics have the strongest predictive power, accounting for by far the largest share of explained variance, about 69.2% (see Table 3.B.0). Virtually every observed factor at the job or firm level shows strong correlations with reporting bias. At the job level, white-collar versus blue-collar employment is negatively correlated with the likelihood of overreporting. Union membership reduces the probability of overreporting by 3.3 percentage points (column 1) and reduces the log reporting bias (column 2). Looking at over- and underreporters separately in columns (3) and (4), we find that union membership has no effect on the magnitude of a positive reporting bias (overreporting), but increases negative reporting bias by 0.7 percentage points, i.e., union members tend to underreport their wages more than non-union members. Working full-time reduces the probability of overreporting wages by 6.4 percentage points compared to those working less than 35 hours per week.

At the firm level, larger firm size is negatively correlated with the likelihood of wage overreporting. Workers in firms with more than 250 employees are 6.2 percentage points less likely to overreport than workers in small firms with up to nine employees, and show a 6.4% lower reporting bias, driven by both lower overreporting and higher underreporting. A similar picture emerges for the average wage level of a firm. Workers employed by firms in the 3rd tercile of the firm premium distribution are 19.3 percentage points less likely to overreport their wages than workers in the 1st tercile, mostly driven by stronger underreporting. The composition of the workforce (in terms of qualification, gender, nationality and age, as well as the share of full-time employees) are strong predictors for the reporting bias. Higher propor-

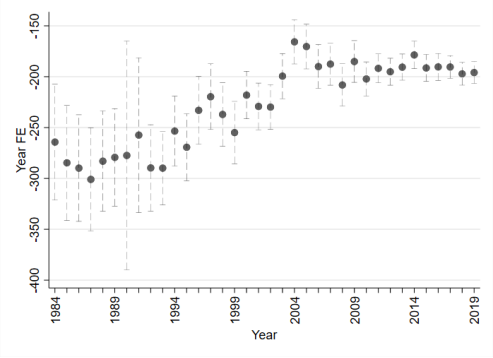


Figure 3.5. Year Fixed Effects

Notes: Graph shows the year fixed effects estimated based on regression 3.1 with the difference between log SOEP and log IEB wages as dependent variable.

tions of women, German nationals and full-time employees all predict lower levels of wage misreporting in the firm.

We can only speculate about the reasons for this strong relationship, which may be explained by heterogeneity in organizational culture and transparency in compensation practices. For example, larger and higher-paying firms may have more formalized pay structures and human resource policies, which may reduce the propensity to overreport wages by providing clearer benchmarks for employee compensation.[15]

Reporting Bias over Time. Finally, we look more closely at the year fixed effects included in Equation (3.1) to understand potential dynamics in the reporting bias. The year fixed effects account for about 10.6% of the explained variance and thus add non-negligible to the explanation of the reporting bias. Figure 3.5 shows that the average reporting bias has increased since the beginning of the SOEP in 1984, but has leveled off since about 2010. We acknowledge, however, that we are only observing a selected sample of panel survivors for the early years, as we can only link information for those respondents who have consented since 2019. Thus, the observed pattern may also represent an effect of panel tenure.

3.3.3 Consequences for Economic Analysis

So far, we have established a modest but economically significant extent of reporting bias between SOEP and IEB (Section 3.3.1). This reporting bias can be predicted

15. Table 3.B.1 in the Appendix additionally shows results for the predictive power of several survey characteristics for the reporting bias, which turn out to have little explanatory power for the reporting bias. An exception are in-person interviews, which appear to reduce the amount of misreporting.

Table 3.3. Factors of Reporting Bias

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Individual Characteristics				
Age	-0.0005	0.0000	0.0002	0.0002
	(0.0003)	(0.0002)	(0.0007)	(0.0001)
Woman	-0.0207*	-0.0160*	-0.0195	-0.0072*
	(0.0087)	(0.0067)	(0.0265)	(0.0029)
Years of education	-0.0015	0.0007	0.0001	0.0017***
	(0.0010)	(0.0009)	(0.0031)	(0.0004)
Region (ref. cat: North)				
East	0.0022	-0.0080	-0.0383*	0.0002
	(0.0104)	(0.0056)	(0.0151)	(0.0035)
South	-0.0032	0.0092	0.0217	0.0037
	(0.0104)	(0.0070)	(0.0221)	(0.0035)
West	0.0103	0.0168*	0.0236	0.0083**
	(0.0096)	(0.0065)	(0.0208)	(0.0032)
Migration background (ref. cat: none)				
Direct migration background	-0.0156	-0.0098	-0.0168	-0.0031
	(0.0124)	(0.0066)	(0.0155)	(0.0048)
Indirect migration background	-0.0105	-0.0019	-0.0017	-0.0010
	(0.0163)	(0.0091)	(0.0228)	(0.0071)
Personality Traits				
Openness (std.)	0.0015	-0.0016	-0.0037	-0.0014
	(0.0045)	(0.0027)	(0.0080)	(0.0015)
Conscientiousness (std.)	0.0033	-0.0008	-0.0089	0.0008
	(0.0046)	(0.0030)	(0.0090)	(0.0018)
Extraversion (std.)	0.0136***	0.0067**	0.0071	0.0005
	(0.0041)	(0.0026)	(0.0082)	(0.0015)
Agreeableness (std.)	-0.0084	-0.0049	-0.0007	-0.0018
	(0.0045)	(0.0031)	(0.0100)	(0.0015)
Neuroticism (std.)	0.0008	-0.0003	-0.0019	-0.0003
	(0.0042)	(0.0027)	(0.0084)	(0.0013)
Household Characteristics				
Married	0.0004	0.0022	0.0226	-0.0060*
	(0.0091)	(0.0046)	(0.0123)	(0.0029)
Household size	-0.0075*	-0.0019	0.0147	-0.0038**
	(0.0036)	(0.0027)	(0.0094)	(0.0013)
Number of kids in HH	0.0085	0.0023	-0.0170	0.0053***
	(0.0045)	(0.0035)	(0.0119)	(0.0016)
Partner	0.0048	0.0011	-0.0271	0.0094**
	(0.0103)	(0.0056)	(0.0143)	(0.0034)
Partner × Partner inc in €1000	0.0057***	0.0031**	0.0020	0.0013*
	(0.0016)	(0.0012)	(0.0026)	(0.0005)

Continued on next page

Factors of Reporting Bias - continued

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Job and Firm Characteristics				
White collar worker	-0.0493***	0.0031	0.0315	0.0104***
	(0.0087)	(0.0062)	(0.0186)	(0.0030)
Union memberhsip	-0.0325***	-0.0112*	0.0219	-0.0073**
	(0.0075)	(0.0056)	(0.0218)	(0.0028)
Working Hours (ref. cat.: <35 hours)				
Full-time (35-44)	-0.0643***	-0.0013	0.0060	0.0297***
	(0.0080)	(0.0055)	(0.0184)	(0.0031)
Over-time (45+)	-0.0183	0.0129	0.0158	0.0276***
	(0.0101)	(0.0067)	(0.0206)	(0.0035)
Firm Size (ref. cat.: <10 employees)				
10-49 employees	-0.0328**	-0.0374***	-0.0509	-0.0127***
	(0.0116)	(0.0101)	(0.0270)	(0.0036)
50-249 employees	-0.0469***	-0.0487***	-0.0615*	-0.0158***
	(0.0118)	(0.0104)	(0.0289)	(0.0038)
250+ employees	-0.0627***	-0.0641***	-0.0788**	-0.0263***
	(0.0124)	(0.0101)	(0.0275)	(0.0040)
Pay Tercile Firm (ref. cat.: 1st Tercile)				
2. Tercile paying firm	-0.1126***	-0.0598***	-0.0271	-0.0248***
	(0.0085)	(0.0056)	(0.0143)	(0.0027)
3. Tercile paying firm	-0.1932***	-0.1083***	-0.0257	-0.0543***
	(0.0101)	(0.0074)	(0.0215)	(0.0035)
Workforce composition				
Share of highly qualified employees	0.0037	0.0155	-0.0198	0.0175**
	(0.0191)	(0.0144)	(0.0532)	(0.0067)
Share of female employees	-0.0600**	-0.0629***	-0.1537***	0.0036
	(0.0195)	(0.0151)	(0.0411)	(0.0057)
Share of German employees	-0.1288***	-0.0076	0.0515	0.0213
	(0.0385)	(0.0212)	(0.0478)	(0.0134)
Share of fulltime employees	-0.1226***	-0.1113***	-0.2357***	-0.0016
	(0.0174)	(0.0165)	(0.0448)	(0.0054)
Mean age of employees in firm	-0.0015*	-0.0003	0.0002	-0.0001
	(0.0007)	(0.0005)	(0.0013)	(0.0002)
Constant	0.6368***	0.0812*	0.3645***	-0.1859***
	(0.0642)	(0.0369)	(0.1047)	(0.0213)
Firm sector controls	✓	✓	✓	✓
Survey characteristics controls	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Shapley Decomposition:				
Individual characteristics			3.93%	
Household characteristics			15.61%	
Job and firm characteristics			69.24%	
Survey characteristics			0.61%	
Year FE			10.59%	
N	42,046	42,046	9,591	31,732
R2	0.0909	0.0734	0.0557	0.0603
Mean dep. var.	0.2281	-0.0678	0.1687	-0.1408

Notes: This table summarizes coefficients of various operationalisations of the reporting error on potential correlates on the individual, household- and firm level, based on Equation 3.1. Dependent variables are a binary indicator for over-reporting (column 1), the difference in log wages (column 2), and the difference in log wages split between over-reporting (column 3) and under-reporting individuals (column 4). Shapley decomposition values indicate shares of explained variation attributed to groups of variables and are reported for the estimates in column (2). All regressions include year-fixed effects. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

to some extent by firm-level factors. Household and individual characteristics have little predictive power. In this section, we examine the implications of the reporting bias for common empirical applications in applied economic research.

3.3.3.1 Wage as Dependent Variable

We first examine the consequences of using self-reported wages rather than administrative wages as the dependent variable in the empirical analysis. We estimate simple versions of two commonly studied economic relationships: the returns to education and the gender wage gap. These topics are of high societal importance, and have received extensive attention from applied economists, partly based on the SOEP data itself.

Returns to Education. Our first case study examines how estimates of returns to education are affected by the use of self-reported versus administrative data. Labor economists have traditionally studied the relationship between education and subsequent labor market outcomes to show how investments in education or training can lead to higher future earnings despite initial opportunity costs such as forgone earnings. This concept, first proposed by Becker (1962) and extensively studied using variations of the Mincer equation (Mincer, 1974), relates wages as the dependent variable to education and experience as explanatory factors. The coefficient on education, usually measured in years, is interpreted as an estimate of the annual return to education, i.e., the financial benefit of an additional year of education. Through this empirical analysis researchers can address questions about the private returns associated with additional years of education and how these returns vary over time, across different levels of education, and for individuals with different background characteristics.

We examine the sensitivity of using either self-reported SOEP or administrative IEB wages as the dependent variable by estimating the following equation:

$$ln(Y)_{it} = \alpha + \beta_1 YED_{it} + \gamma X'_{i(t)} + \varepsilon_{it} \tag{3.2}$$

where $ln(Y)_{it}$ is the natural logarithm of the monthly wage taken from either the SOEP or the IEB, and YED_{it} is the years of education attained measured by the SOEP. The control variables $X'_{i(t)}$ consist of a minimum (age, age squared, firm tenure in years, a dummy for being full-time employment ($>$ 35 hours), year dummies) or an extended set (basic controls plus a dummy for being female, a dummy for having migration background, full-time and unemployment experience in years, industry dummies, county of residence dummies, firm size indicators, a dummy for having children under 16 in the household, marital status indicators) of individual-specific characteristics. Sample restrictions remain similar to those described in Table 3.1, i.e., we consider individuals aged 20-65 in any dependent full- or part-time employ-

ment for whom all relevant controls are non-missing. We cluster standard errors at the individual level.[16]

Figure 3.6, Panel (a), shows the estimated coefficients of β as a measure of the returns to an additional year of education, based on either SOEP or IEB. We estimate three different specifications: a "plain vanilla" OLS specification with no additional controls, a "basic controls" specification, and an "extended" specification, as described above. In all specifications, the estimated returns to education are of the same order of magnitude, and the confidence intervals of the IEB and SOEP estimates overlap. An additional year of education significantly increases wages by 7% to 7.8% when using SOEP wages and by 6.8% to 7.3% when using IEB wages. Different sets of controls have only little effect on these differences. When estimating the two regression equations jointly with GMM, a two-sample t-test does not detect significant differences in all but the extended specification.[17] We therefore conclude that the choice of data in the case of the returns to education appears to be largely inconsequential and does not affect its interpretation in an economically relevant sense.

Gender Wage Gap. Perhaps no other measure of labor market inequality receives as much public attention as the gender wage gap. The gender wage gap describes the average difference in earnings between women and men in the labor force. It is typically expressed as a percentage of men's earnings to highlight the difference in pay for comparable work. It can be measured as a raw hourly or monthly wage gap, or as an adjusted gender wage gap, which takes into account factors such as occupation, industry, work experience, education, hours worked, and job tenure. Its extent and development is described over time and contexts (Goldin, 2014; Blau and Kahn, 2017). The gender wage gap has received renewed attention in recent years through the emerging literature on the "child penalty" as a primary reason for these persistent inequalities (e.g., Bertrand, Goldin, and Katz, 2010; Goldin, 2014; Adda, Dustmann, and Stevens, 2017; Kleven et al., 2019; Kleven, Landais, and Sogaard, 2019).

To measure the sensitivity of estimates of the gender wage gap to the reliance on self-reported versus administrative wage data, we estimate the following equation:

$$ln(Y)_{it} = \alpha + \beta_2 Female_i + \gamma X'_{i(t)} + \varepsilon_{it} \tag{3.3}$$

which relates the natural logarithm of the monthly wage $ln(Y)_{it}$, taken from either SOEP or IEB, to an indicator $Female_i$ that takes the value of 1 for women. The

16. The results remain qualitatively and quantitatively similar when we cluster at the household level instead in order to correct for a potential correlation of reporting bias within households (see Figure 3.B.3 in Appendix 3.B).

17. The respective p-values of the test for significance are 0.30 for the regression without controls, 0.82 with basic controls and 0.03 with extended controls.

coefficient β approximates of the gender wage gap. Control variables in $X'_{i(t)}$ are equivalent to those used in equation 3.2, additionally controlling for an individual's education level. Again, the sample restrictions correspond to those described in Table 3.1. In addition, we impose a restriction on full-time workers in order to identify wage differences rather than differences in working hours.

Figure 3.6, Panel (b) shows that the estimated gender wage gap is substantially larger in all specifications when based on self-reported SOEP data. Without controls, it is about 15.1% based on SOEP, and only about 13.6% based on IEB data.[18] With extended controls, it shrinks to 10.7% based on SOEP data and to 7.7% based on IEB data. However, the relative difference between the estimates based on the different data sources increases from 11% to 38% with extended controls. This difference is directly related to the results observed in Section 3.3.2: The negative coefficient of female gender on the reporting bias thus translates into the observed difference in the gender wage gap. We conclude that the choice of data source is relevant for the gender wage gap. The confidence intervals of the coefficients remain overlapping, but a when estimating the two equations jointly with GMM, the t-test confirms significant differences in all specifications.[19] Note that this difference runs counter the intuition of a mere attenuation bias by classical measurement error in survey wages, which would lead to a lower gender gap based on survey data.

3.3.3.2 Wage as Independent Variable

In a second step, we examine the consequences of using self-reported wages instead of administrative wages as the right-hand side explanatory variable. To this end, we replicate empirical exercises explaining individual-level subjective outcomes of satisfaction in different domains that have been shown to be affected by individual income (e.g., Frijters, Haisken-DeNew, and Shields, 2004; Boyce, Brown, and Moore, 2010; Cheung and Lucas, 2015).

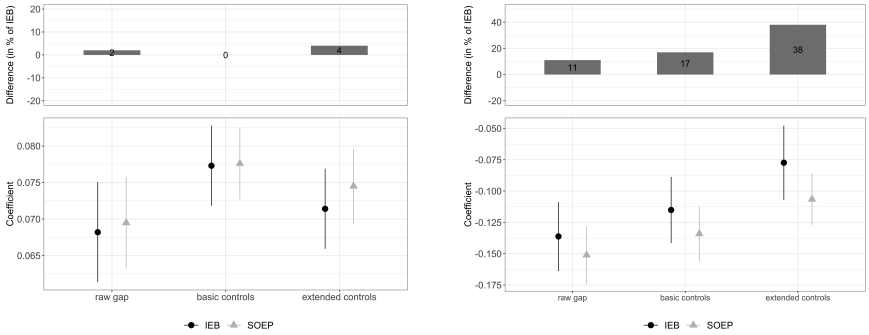
Estimation Strategy. To compare the estimates when using either the SOEP or the IEB wage as explanatory variables, we estimate the following equation:

$$SI_{it} = \alpha + \beta Wage_{it} + \gamma X'_{i(t)} + \delta_i + \varepsilon_{it} \tag{3.4}$$

where $Wage_{it}$ is the monthly wage (in €1,000), based on either IEB or SOEP data. SI_{it} is one of several indicators of subjective well-being: life satisfaction and satisfaction with job and household or personal income. Each of these items is measured

18. Both figures are substantially smaller than those reported by the German Statistical Office for the same period (around 19-24% (6-9%) for West (East) Germany in the years 2006-2022, including part-time workers) (Destatis, 2024). The relatively small gender wage gap in our sample is driven by the fact that we only consider full-time employees and that later years, in which the gender wage gap has already narrowed, are strongly overrepresented.

19. The respective p-values of the test for significance are 0.04 for the regression without controls, 0.01 with basic controls and 0.01 with extended controls.



(a) Returns to education **(b)** Gender wage gaps

Figure 3.6. Returns to Education and Gender Wage Gap Based on SOEP and IEB wages

Notes: This figure summarizes results of regressions of log wages, based on either SOEP or IEB data, on various sets of explanatory variables. Panel (a) summarizes estimates of the returns to education β_1 based on equation 3.2. Panel (b) summarizes estimates of the gender wage gap β_2 based on equation 3.3. Raw, basic and extended specifications differ in control variables as described in the text. Bar graphs illustrate the relative size of the difference by dividing the difference between the SOEP and IEB estimate by the IEB estimates. Restricted to full-time for Panel (b). The graph displays 95% confidence intervals, based on standard errors clustered at the individual level.

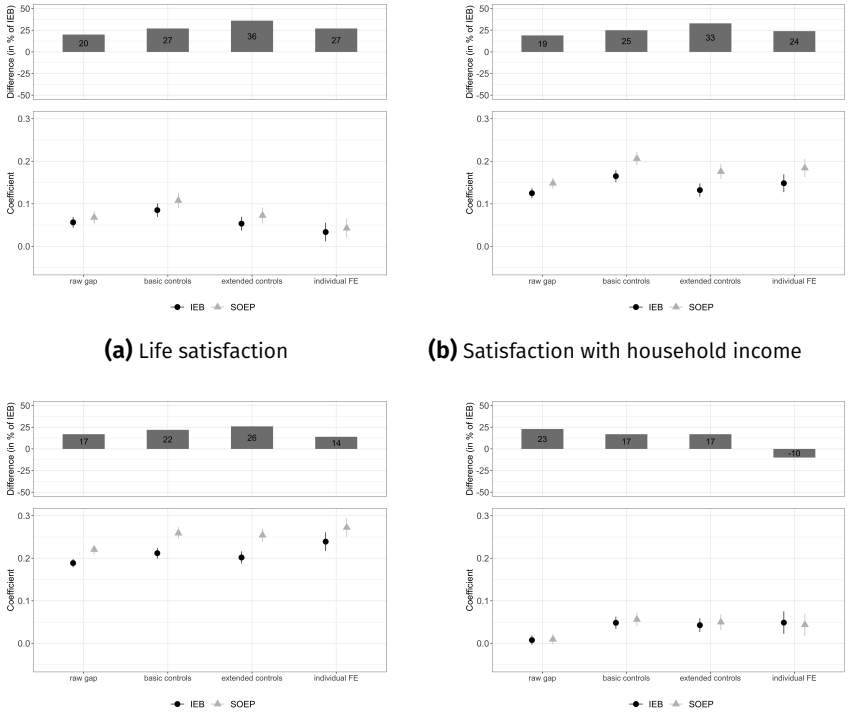
by standard questionnaire items[20] and are standardized to have a mean of zero and a standard deviation of one. Our main parameter of interest is β, which shows the association between wages and the subjective indicators. The control variables $X'_{i(t)}$ are similar to those described above. Standard errors are clustered at the individual level, sample restrictions are equivalent to those described in Table 3.1.[21]

Results. Figure 3.7 shoes the point estimates β_{IEB} in black and β_{SOEP} in gray with the corresponding 95% confidence intervals. Again, we compare specifications that differ in the number of control variables. In addition, we examine a specification that focuses on individual changes by including individual fixed effects.

The results show that the influence of misreporting depends on the type of satisfaction assessed. The type of data source appears to be unrelated to general life or job satisfaction (Panels (a) and (d)), but significantly affects the estimated income effects for satisfaction with household income or personal income, i.e., in those dimensions where it is the variable of focal interest (Panels (b) and (c)). In the latter cases, point estimates based on self-reported data consistently exceed those based

20. The exact wording in the 2019 English version of the SOEP is the following: "a) for satisfaction with job and income: *How satisfied are you today with the following areas of your life? ... (if employed) with your job?... with your household income? ... with your personal income?*; Possible answers: 0 'completely dissatisfied' to 10 'completely satisfied'; b) for life satisfaction: *In conclusion, we would like to ask you about your satisfaction with your life in general. How satisfied are you with your life, all things considered?*; possible answers: 0 'completely dissatisfied' to 10 'completely satisfied'."

21. The results remain qualitatively and quantitatively similar when we cluster at the household level instead in order to correct for a potential correlation of reporting bias within households (see Figure 3.B.4 in Appendix 3.B).



(a) Life satisfaction

(b) Satisfaction with household income

(c) Satisfaction with personal income

(d) Satisfaction with job

Figure 3.7. Relation Between (Life) Satisfaction and Wages

Notes: This figure summarizes results of regressions of subjective satisfaction indicators on log wages and control variables based on Equation 3.4, with varying sets of control variables. Outcome variables are standardized to a mean of zero and a standard deviation of one. Raw, basic and extended specifications differ in control variables as described in the text. Bar graphs illustrate the relative size of the difference by dividing the difference between the SOEP and IEB estimate by the IEB estimates. The graph displays 95% confidence intervals, based on standard errors clustered at the individual level.

on administrative data. The confidence intervals do not overlap. For example, satisfaction with personal income increases with an additional €1,000 by about 25.4% of a standard deviation when assessed on the basis of self-reported income, but by only 20.2% of a standard deviation when assessed on the basis of IEB data – a difference of 26%. We conclude that the reporting bias can lead to quantitatively different conclusions depending on the data source when examining income-related subjective outcomes. However, introducing individual fixed effects to rely on interpersonal changes mitigates the bias, reflecting the fact that misreporting appears to be less of an issue for changes than for levels. Further note that this difference cannot be explained by a mere attenuation bias in survey wages, which would hint at *smaller* coefficients when working with survey data. Rather, our results may be explained by unobservable factors that simultaneously and similarly affect both subjective satisfaction and self-reported income (Bertrand and Mullainathan, 2001). Prati (2017) argues that self-reported wages may be prone to a *hedonic recall bias*, meaning that

individuals tend to overreport their wages when they are satisfied and underreport when they are dissatisfied with their wages, leading to an upward bias in the relationship between income and satisfaction.

3.4 Conclusion

Based on a novel record linkage of the SOEP-ADIAB, which links the German Socioeconomic Panel (SOEP) to administrative wage information from the Integrated Employment Biographies (IEB), this paper examines the extent to which survey respondents misreport their wages, which observable characteristics may predict the extent of misreporting, and the consequences for applied researchers working with these misreported wages.

Our analysis yields three sets of results. First, with respect to the magnitude of the reporting bias, the IEB and SOEP wage distributions are closely aligned, with the average SOEP respondent underreporting individual wages by a modest but economically significant 7.3% of the mean. We find evidence of a mean-reverting component to the reporting bias, with respondents in the very low tails of the wage distribution overreporting their wages and a steadily increasing underreporting from there. The magnitude of misreporting appears to be smaller in terms of changes than in terms of levels. While patterns of mean reversion are comparable to previous studies, the extent of misreporting, as well as the rather good representation of wage changes seem to be a particularity of the SOEP data.

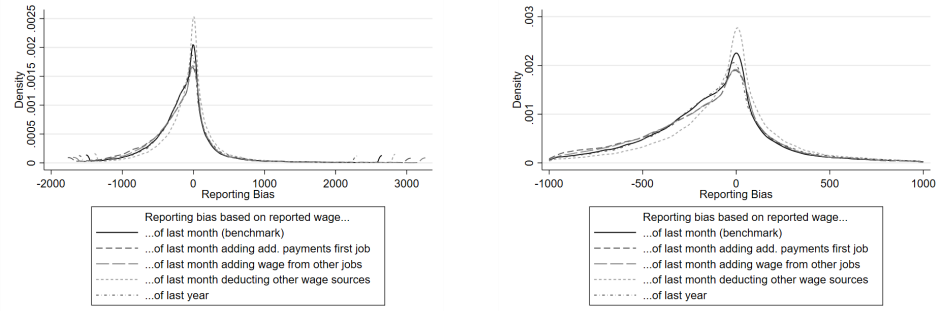
Second, with respect to the correlates of reporting bias, we find that job and firm characteristics are particularly predictive of the extent to which individuals misreport their wages. In contrast, household characteristics and especially individual characteristics contribute only weakly to explaining the variance in reporting bias.

Third, turning to the consequences of choosing survey over administrative wages, we examine the sensitivity of prominent economic relationships commonly estimated in applied economic studies. Using income as the dependent variable, we find no difference in the estimated returns to education and an economically meaningful and significant difference in the gender wage gap. Using income as an explanatory variable to estimate the relationship between satisfaction and income, the use of misreported survey data leads to a significant overestimation, especially in the dimensions of household income and personal income satisfaction.

Taken together, our findings suggest that researchers may need to be more cautious in interpreting survey-based measures of individual income in empirical research and the empirical relationships estimated on them. In our exemplary analysis, misreporting was largely inconsequential for examining the returns to education, but appeared to be more relevant for analyzing the gender wage gap – reflecting systematic variation in misreporting along the gender dimension. In addition,

we found that misreporting bias can significantly affect the results when wage is used as the independent variable, as we did in estimating the wage-satisfaction relationship. These consequences of misreported income extend beyond simple attenuation bias due to classical measurement error in survey wages. Our findings underscore that, at least in some cases, survey-based measures of individual wages can significantly bias commonly estimated empirical relationships. They also demonstrate the enormous research potential of linked administrative-survey data Bee et al. (see, e.g., also the project by 2023) which would moreover enable researchers to create a measure of "true" wage income that could be a hybrid version of the survey and the administrative information (Meijer, Rohwedder, and Wansbeek, 2012; Abowd and Stinson, 2013). Extending our analysis to different research questions will help to get a clearer picture of the importance of the misreporting problem. It may be the case that even modest misreporting biases in wages, influenced by various individual, household, job, and firm characteristics, as shown in our paper, can significantly bias (or not bias) common empirical exercises. This is an interesting avenue for future research.

Appendix 3.A Additional Wage Concepts



(a) Distribution of the reporting bias for different wage concepts

(b) Zoom in: restricted to a reporting bias of + / − €1,000

Figure 3.A.1. Reporting Bias for Different Wage Measures.

Notes: Graph shows the density distribution of the reporting bias for different wage concepts. Here, we vary the definition of SOEP wages. Income from additional sources includes reported vacation pay, Christmas pay, and the 13th and 14th monthly wage payment. For the exclusion of additional payment we deduct reported income from other employment. This variable is surveyed in SOEP only for the survey waves from 2017 to 2019, so the reporting bias is only plotted for these years. The yearly measure is based on a different set of survey items that asks about the income in the year before, specifically the monthly wage income and the number of months the individual received the wage income. Reporting biases that lie below the 1st percentile or above the 99th percentile of the distribution of differences are binned into the 1st or 99th percentile respectively.

Figure 3.A.1 investigates whether taking into account additional information from the SOEP changes the distribution of the reporting bias. Panel a) shows the whole distribution of the reporting bias. For presentational reasons, panel b) zooms in and restricts the distribution to a reporting bias of plus/ minus €1,000.

First, we will consider additional sources of wage from the primary job. The reporting bias could potentially mislead if the question is perceived as solely pertaining to wage income from the main job. Consequently, we will incorporate all declared additional income sources, including vacation pay, Christmas pay, 13th, and 14th monthly wage payments. The resulting effect is small, with a shift towards a more uniform distribution centered around 0.

Second, we supplement our wage variable from wages earned from other jobs in regular employment as reported in the SOEP. The phrasing of the primary question ("Arbeitsverdienst im letzten Monat") implies that these wages are already accounted for. We verify this. The graphs indicate that participants either incorporate or exclude wages from other jobs in the primary SOEP inquiry regarding last month's wage. Unfortunately, this survey item was only introduced in 2017, so the analysis only analyses responses for the years 2017, 2018 and 2019.

Third, we subtract reported wages from other sources, specifically self-employment, work in family businesses, and other non-regular employment jobs,

from the SOEP wage variable. Again, we do not detect a economically relevant difference.

Also, Figure 3.A.1 displays the distribution of the reporting bias based on a SOEP questionnaire. Each year, individuals are asked not only about their income from the previous month but also about their wages earned during the previous calendar year. Participants are specifically prompted to report the months they worked as an employee and their average gross monthly wage in each month. The Figure reveals that the reporting bias that is based on yearly income is very similar to the reporting bias that is based on the reported income the month before.

Appendix 3.B Supplementary Figures and Tables

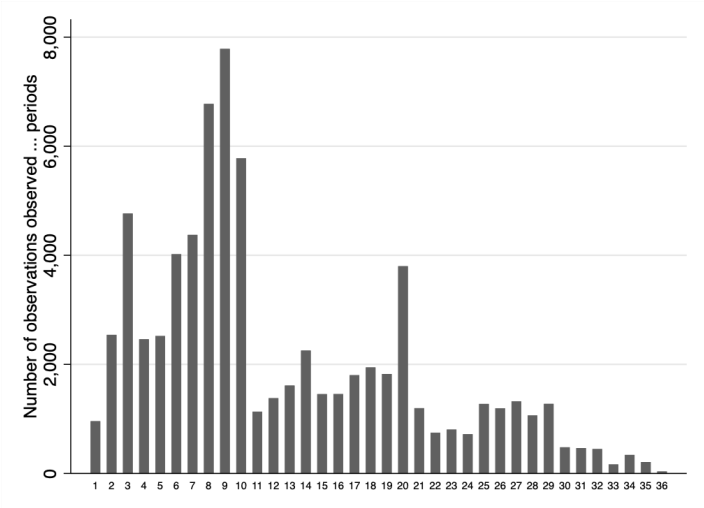


Figure 3.B.1. Number of Individuals Observed Over x Periods of Observation

Notes: This figure summarizes the panel tenure of individuals in our sample. Bars show the number of individuals that are observed for the number of years depicted on the x - axis. A large number of individuals is observed for 8-10 years.

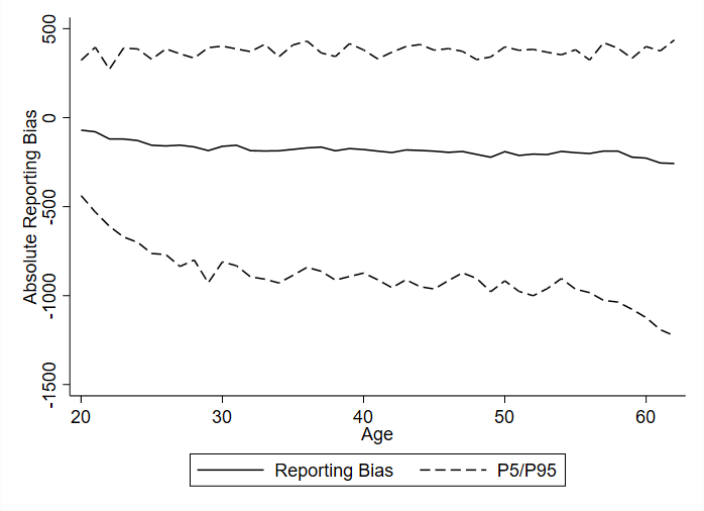
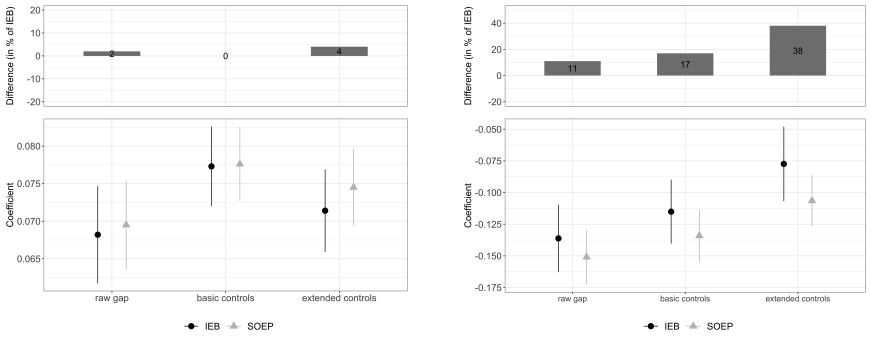


Figure 3.B.2. Distribution of Difference SOEP - IEB Over Age

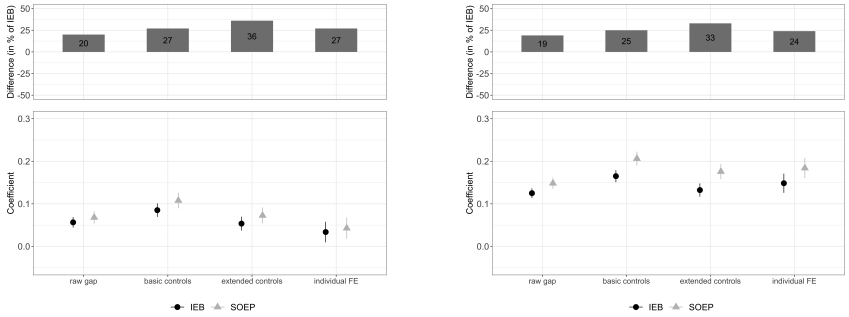
Notes: This figure shows the mean reporting bias in absolute terms (black solid line) and the respective 5th and 95th percentile (dashed lines) over age.



(a) Returns to education **(b)** Gender wage gaps

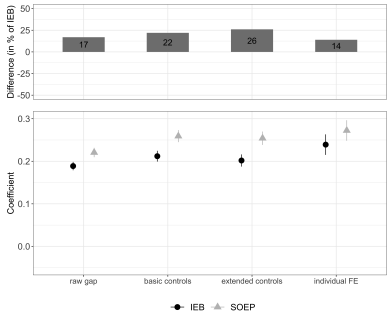
Figure 3.B.3. Returns to Education and Gender Wage Gap Based on SOEP and IEB wages

Notes: This figure summarizes results of regressions of log wages, based on either SOEP or IEB data, on various sets of explanatory variables. Panel (a) summarizes estimates of the returns to education β_1 based on equation 3.2. Panel (b) summarizes estimates of the gender wage gap β_2 based on equation 3.3. Raw, basic and extended specifications differ in control variables as described in the text. Bar graphs illustrate the relative size of the difference by dividing the difference between the SOEP and IEB estimate by the IEB estimates. Restricted to full-time for Panel (b). The graph displays 95% confidence intervals, based on standard errors clustered at the household level.

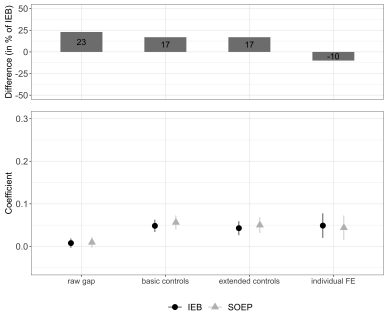


(a) Life satisfaction

(b) Satisfaction with household income



(c) Satisfaction with personal income



(d) Satisfaction with job

Figure 3.B.4. Relation Between (Life) Satisfaction and Wages

Notes: This figure summarizes results of regressions of subjective satisfaction indicators on log wages and control variables based on equation 3.4, with varying sets of control variables. Outcome variables are standardized to a mean of zero and a standard deviation of one. Raw, basic and extended specifications differ in control variables as described in the text. Bar graphs illustrate the relative size of the difference by dividing the difference between the SOEP and IEB estimate by the IEB estimates. The graph displays 95% confidence intervals, based on standard errors clustered at the household level.

Table 3.B.1. Differences between Consenters and Non-Consenters, Cross-Section 2019

	Consent Mean/SD	No Consent Mean/SD	Diff C/NC t-Test/SE
Individual Characteristics			
Age	50.70	51.86	1.16***
	(17.86)	(18.44)	(0.25)
Female (0/1)	0.54	0.54	-0.00
	(0.50)	(0.50)	(0.01)
German (0/1)	0.96	0.95	-0.02***
	(0.19)	(0.22)	(0.00)
Married (0/1)	0.57	0.58	0.00
	(0.49)	(0.49)	(0.01)
Ever partner (0/1)	1.04	1.11	0.06***
	(0.82)	(0.87)	(0.01)
Years of education	11.83	11.84	0.01
	(4.15)	(4.50)	(0.06)
Father abitur (0/1)	0.16	0.17	0.01*
	(0.37)	(0.38)	(0.01)
Mother abitur (0/1)	0.11	0.12	0.01
	(0.31)	(0.32)	(0.00)
Personality traits			
Optimistic (0/1)	0.80	0.77	-0.03***
	(0.40)	(0.42)	(0.01)
Risk lovingness (std)	0.04	-0.07	-0.11***
	(0.90)	(0.92)	(0.01)
IQ prediction (std)	0.08	0.13	0.04*
	(1.00)	(0.99)	(0.02)
Openness (std)	0.04	-0.01	-0.05***
	(0.85)	(0.85)	(0.01)
Conscientiousness (std)	-0.09	-0.07	0.01
	(0.86)	(0.87)	(0.01)
Extraversion (std)	0.03	-0.04	-0.07***
	(0.90)	(0.90)	(0.01)
Agreeableness (std)	-0.02	-0.05	-0.04**
	(0.84)	(0.83)	(0.01)
Neuroticism (std)	-0.05	0.00	0.05***
	(0.88)	(0.87)	(0.01)
Worriedness (std)	-0.02	0.03	0.05**
	(0.99)	(0.98)	(0.01)

Continued on next page

Differences between Consenters and Non-Consenters, Cross-Section 2019 - continued

	Consent Mean/SD	No Consent Mean/SD	Diff C/NC t-Test/SE
Job and Firm Characteristics			
Monthly wage SOEP	2916.50	3069.70	153.20**
	(2514.09)	(2594.24)	(46.71)
Unemployed >1 Year (0/1)	0.21	0.19	-0.02***
	(0.41)	(0.39)	(0.01)
Public sector empl (0/1)	0.28	0.27	-0.01
	(0.45)	(0.44)	(0.01)
Ever civil servant (0/1)	0.07	0.08	0.01**
	(0.25)	(0.27)	(0.00)
Max. experience fulltime | >0	19.41	20.07	0.66**
	(14.00)	(13.91)	(0.22)
Max. experience parttime | >0	7.85	8.01	0.16
	(8.08)	(8.35)	(0.16)
Max. experience unemployed | >0	3.66	3.12	-0.54***
	(5.08)	(4.45)	(0.12)
White collar worker	0.78	0.80	0.02**
	(0.42)	(0.40)	(0.01)
Union membership	0.12	0.11	-0.01**
	(0.33)	(0.31)	(0.00)
Firm size <=10	0.15	0.16	0.01*
	(0.35)	(0.37)	(0.01)
Firm size < 100	0.23	0.24	0.01
	(0.42)	(0.43)	(0.01)
Firm size < 2000	0.30	0.30	-0.00
	(0.46)	(0.46)	(0.01)
Firm sie >=2000	0.32	0.30	-0.02*
	(0.47)	(0.46)	(0.01)
N	15,014	8,514	23,528

Notes: Cross-section 2019. Column 3 shows the difference in means between column 1 and 2. Max. Experience Fulltime, Parttime and Unemployed is measured conditional on ever having a positive value for fulltime, parttime or unemployment. All individuals, both SOEP IS and SOEP Core, before any further sample restrictions are made. Compared to Table 3.1, the number of consenters deviates by 2 observations, which are missing in the IAB remote application we are using for linked dataset. Standard errors in parenthesis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.1. Factors of Reporting Bias, Year Effects & Survey Characteristics

	(1) Over Y/N	(2) Diff of Logs	(3) Diff of Logs (+)	(4) Diff of Logs (−)
Firm sector controls				
1-digit industries (ref. cat: Agriculture, hunting and forestry)				
Fishing	-0.2373	-0.0328	-0.1305**	0.0653***
	(0.2261)	(0.0194)	(0.0440)	(0.0107)
Mining and quarrying	-0.0501	-0.0096	0.1126	-0.0390
	(0.0505)	(0.0291)	(0.1585)	(0.0534)
Manufacturing	-0.0480	-0.0240	-0.0199	-0.0153
	(0.0344)	(0.0152)	(0.0264)	(0.0102)
Electricity, gas and water supply	-0.0970*	-0.0525**	-0.0134	-0.0329*
	(0.0434)	(0.0199)	(0.0419)	(0.0156)
Construction	-0.0360	-0.0345*	-0.0707*	-0.0102
	(0.0361)	(0.0157)	(0.0280)	(0.0112)
Wholesale and retail trade	-0.0911**	-0.0370*	-0.0331	-0.0112
	(0.0348)	(0.0158)	(0.0302)	(0.0103)
Hotels and restaurants	0.0298	0.0368	0.0698	-0.0362*
	(0.0423)	(0.0292)	(0.0527)	(0.0153)
Transport and communication	0.0159	0.0176	0.0179	-0.0014
	(0.0378)	(0.0204)	(0.0408)	(0.0111)
Financial intermediation	-0.0621	-0.0473**	-0.0050	-0.0448***
	(0.0381)	(0.0181)	(0.0374)	(0.0119)
Real estate, renting and business activities	-0.0323	-0.0092	-0.0025	-0.0070
	(0.0347)	(0.0159)	(0.0297)	(0.0106)
Public administration and defence	-0.1126**	-0.0384*	0.0057	-0.0158
	(0.0350)	(0.0164)	(0.0391)	(0.0107)
Education	-0.0818*	-0.0212	0.0309	-0.0077
	(0.0363)	(0.0182)	(0.0453)	(0.0114)
Health and social work	-0.0487	-0.0115	0.0133	-0.0075
	(0.0355)	(0.0184)	(0.0437)	(0.0107)
Other service activities	-0.0598	-0.0006	0.0697	-0.0063
	(0.0368)	(0.0207)	(0.0553)	(0.0114)
Private households	0.0326	0.0343	-0.0649	0.0451***
	(0.1050)	(0.0329)	(0.0707)	(0.0128)
Survey mode *(ref.cat.: in person)*				
Survey: Not In Person	-0.0352***	0.0019	0.0082	0.0123***
	(0.0079)	(0.0059)	(0.0230)	(0.0026)
Survey: Unclear	-0.0269	-0.0087	-0.0273	0.0065
	(0.0145)	(0.0083)	(0.0295)	(0.0049)
Number of SOEP interviews before	-0.0006	-0.0004	-0.0014	0.0001
	(0.0005)	(0.0003)	(0.0010)	(0.0002)

Continued on next page

Factors of Reporting Bias, Year Effects & Survey Characteristics - continued

	(1) Over Y/N	(2) Diff of Logs	(3) Diff of Logs (+)	(4) Diff of Logs (−)
Year effects *(ref.cat.: 1984)*				
1985	-0.0013	-0.0220	-0.0361	-0.0205
	(0.0322)	(0.0153)	(0.0653)	(0.0111)
1986	0.0342	-0.0288*	-0.0802	-0.0345***
	(0.0273)	(0.0134)	(0.0508)	(0.0099)
1987	-0.0169	-0.0266	-0.0783	-0.0174
	(0.0284)	(0.0141)	(0.0589)	(0.0114)
1988	0.0400	-0.0195	-0.0568	-0.0299*
	(0.0298)	(0.0154)	(0.0540)	(0.0131)
1989	0.0494	-0.0082	-0.0372	-0.0227
	(0.0313)	(0.0147)	(0.0513)	(0.0121)
1990	-0.0149	-0.0377	-0.1256*	-0.0362
	(0.0545)	(0.0309)	(0.0554)	(0.0332)
1991	0.0318	0.0043	-0.0706	-0.0121
	(0.0420)	(0.0183)	(0.0658)	(0.0148)
1992	0.0455	-0.0248	-0.0163	-0.0533***
	(0.0295)	(0.0160)	(0.0619)	(0.0119)
1993	0.0349	-0.0086	-0.0275	-0.0323**
	(0.0286)	(0.0141)	(0.0532)	(0.0105)
1994	0.0883**	0.0254	-0.0255	-0.0057
	(0.0290)	(0.0140)	(0.0530)	(0.0104)
1995	0.0576*	0.0223	-0.0046	-0.0025
	(0.0280)	(0.0137)	(0.0552)	(0.0102)
1996	0.1284***	0.0444***	-0.0555	0.0136
	(0.0297)	(0.0134)	(0.0494)	(0.0102)
1997	0.1331***	0.0564***	-0.0229	0.0194
	(0.0286)	(0.0134)	(0.0496)	(0.0102)
1998	0.1276***	0.0463**	-0.0136	0.0082
	(0.0283)	(0.0146)	(0.0553)	(0.0109)
1999	0.1296***	0.0390**	-0.0317	0.0016
	(0.0295)	(0.0142)	(0.0534)	(0.0105)
2000	0.1754***	0.0688***	0.0166	0.0172
	(0.0280)	(0.0145)	(0.0551)	(0.0099)
2001	0.1616***	0.0699***	0.0538	0.0138
	(0.0278)	(0.0155)	(0.0610)	(0.0099)
2002	0.1541***	0.0573***	0.0331	0.0059
	(0.0270)	(0.0146)	(0.0561)	(0.0101)
2003	0.2144***	0.0775***	-0.0042	0.0185
	(0.0281)	(0.0139)	(0.0507)	(0.0099)
2004	0.2728***	0.0949***	0.0009	0.0171
	(0.0287)	(0.0155)	(0.0546)	(0.0102)
2005	0.2400***	0.0937***	0.0089	0.0258*
	(0.0286)	(0.0147)	(0.0528)	(0.0101)

Continued on next page

Factors of Reporting Bias, Year Effects & Survey Characteristics - continued

	(1) Over Y/N	(2) Diff of Logs	(3) Diff of Logs (+)	(4) Diff of Logs (−)
Year effects *(ref.cat.: 1984)*				
2006	0.2430***	0.0811***	-0.0385	0.0224*
	(0.0285)	(0.0144)	(0.0522)	(0.0101)
2007	0.2389***	0.0890***	0.0102	0.0190
	(0.0285)	(0.0156)	(0.0567)	(0.0103)
2008	0.2458***	0.0748***	-0.0227	0.0063
	(0.0287)	(0.0158)	(0.0562)	(0.0108)
2009	0.2388***	0.0919***	0.0114	0.0206*
	(0.0280)	(0.0163)	(0.0587)	(0.0103)
2010	0.2277***	0.0862***	0.0127	0.0200*
	(0.0275)	(0.0157)	(0.0579)	(0.0099)
2011	0.2336***	0.0921***	-0.0058	0.0291**
	(0.0272)	(0.0156)	(0.0586)	(0.0098)
2012	0.2649***	0.0858***	-0.0391	0.0201*
	(0.0270)	(0.0154)	(0.0569)	(0.0099)
2013	0.2325***	0.0919***	0.0056	0.0265**
	(0.0270)	(0.0158)	(0.0584)	(0.0099)
2014	0.2449***	0.0967***	0.0077	0.0282**
	(0.0272)	(0.0160)	(0.0582)	(0.0100)
2015	0.2307***	0.0939***	0.0002	0.0307**
	(0.0273)	(0.0158)	(0.0579)	(0.0099)
2016	0.2523***	0.0921***	-0.0264	0.0294**
	(0.0275)	(0.0158)	(0.0578)	(0.0099)
2017	0.2342***	0.0937***	-0.0145	0.0343***
	(0.0270)	(0.0154)	(0.0573)	(0.0098)
2018	0.2358***	0.0935***	-0.0210	0.0355***
	(0.0272)	(0.0154)	(0.0572)	(0.0098)
2019	0.2519***	0.0977***	-0.0297	0.0383***
	(0.0272)	(0.0154)	(0.0572)	(0.0098)
N	42,046	42,046	9,591	31,732

Notes: This table summarizes coefficients of various operationalisations of the reporting error on potential correlates on the survey mode and year effects based on Equation 3.1. These coefficients were omitted from Table 3.3 for clarity reasons. Dependent variables are a binary indicator for overreporting (column 1), the difference in log wages (column 2), and the difference in log wages split between overreporting (column 3) and underreporting individuals (column 4). Correct reporting is defined as no or small reporting bias within a +/- 2.5 percentile range of the annual reporting bias. All regressions include year-fixed effects. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.0. Factors of Reporting Bias, Zero Bias Defined as +/ − 1.25% around 0

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Individual Characteristics				
Age	-0.0004	0.0000	0.0002	0.0002
	(0.0003)	(0.0002)	(0.0006)	(0.0001)
Woman	-0.0210*	-0.0160*	-0.0193	-0.0068*
	(0.0087)	(0.0067)	(0.0262)	(0.0029)
Years of education	-0.0017	0.0007	0.0002	0.0016***
	(0.0010)	(0.0009)	(0.0031)	(0.0004)
Region (ref. cat: North)				
East	0.0029	-0.0080	-0.0382*	0.0002
	(0.0105)	(0.0056)	(0.0149)	(0.0035)
South	-0.0030	0.0092	0.0214	0.0040
	(0.0104)	(0.0070)	(0.0219)	(0.0035)
West	0.0106	0.0168*	0.0236	0.0088**
	(0.0097)	(0.0065)	(0.0206)	(0.0032)
Migration background (ref. cat: none)				
Direct migration background	-0.0163	-0.0098	-0.0153	-0.0034
	(0.0125)	(0.0066)	(0.0153)	(0.0047)
Indirect migration background	-0.0075	-0.0019	-0.0040	-0.0007
	(0.0162)	(0.0091)	(0.0225)	(0.0070)
Personality Traits				
Openness (std.)	0.0017	-0.0016	-0.0037	-0.0013
	(0.0045)	(0.0027)	(0.0079)	(0.0015)
Conscientiousness (std.)	0.0036	-0.0008	-0.0092	0.0008
	(0.0046)	(0.0030)	(0.0089)	(0.0018)
Extraversion (std.)	0.0134**	0.0067**	0.0073	0.0006
	(0.0041)	(0.0026)	(0.0082)	(0.0014)
Agreeableness (std.)	-0.0082	-0.0049	-0.0007	-0.0021
	(0.0045)	(0.0031)	(0.0100)	(0.0015)
Neuroticism (std.)	0.0005	-0.0003	-0.0017	-0.0001
	(0.0042)	(0.0027)	(0.0083)	(0.0013)
Household Characteristics				
Married	-0.0000	0.0022	0.0234	-0.0057
	(0.0091)	(0.0046)	(0.0122)	(0.0029)
Household Size	-0.0078*	-0.0019	0.0146	-0.0037**
	(0.0036)	(0.0027)	(0.0094)	(0.0013)
Number of kids in HH	0.0090*	0.0023	-0.0172	0.0051**
	(0.0045)	(0.0035)	(0.0118)	(0.0016)
Partner	0.0047	0.0011	-0.0274	0.0089**
	(0.0103)	(0.0056)	(0.0142)	(0.0034)
Partner × Partner Inc in €1,000	0.0056***	0.0031**	0.0020	0.0014**
	(0.0016)	(0.0012)	(0.0026)	(0.0005)

Continued on next page

Factors of Reporting Bias, Zero Bias Defined as +/ − 1.25% around 0 - continued

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Job and Firm Characteristics				
White Collar Worker	-0.0494***	0.0031	0.0310	0.0104***
	(0.0087)	(0.0062)	(0.0184)	(0.0030)
Union memberhsip	-0.0322***	-0.0112*	0.0210	-0.0075**
	(0.0075)	(0.0056)	(0.0216)	(0.0028)
Working hours (ref. cat.: <35 hours)				
Full-time (35-44)	-0.0688***	-0.0013	0.0088	0.0262***
	(0.0080)	(0.0055)	(0.0183)	(0.0031)
Over-time (45+)	-0.0231*	0.0129	0.0191	0.0242***
	(0.0102)	(0.0067)	(0.0204)	(0.0035)
Firm size (ref. cat.: <10 employees)				
10-49 employees	-0.0346**	-0.0374***	-0.0497	-0.0155***
	(0.0116)	(0.0101)	(0.0266)	(0.0036)
50-249 employees	-0.0490***	-0.0487***	-0.0603*	-0.0193***
	(0.0118)	(0.0104)	(0.0285)	(0.0038)
250+ employees	-0.0654***	-0.0641***	-0.0766**	-0.0299***
	(0.0124)	(0.0101)	(0.0270)	(0.0040)
Pay tercile firm (ref. cat.: 1st Tercile)				
2. Tertile paying firm	-0.1140***	-0.0598***	-0.0264	-0.0261***
	(0.0085)	(0.0056)	(0.0141)	(0.0027)
3. Tertile paying firm	-0.1949***	-0.1083***	-0.0254	-0.0560***
	(0.0101)	(0.0074)	(0.0213)	(0.0035)
Workforce composition				
Share of highly qualified employees	0.0031	0.0155	-0.0198	0.0169*
	(0.0192)	(0.0144)	(0.0525)	(0.0067)
Share of female employees	-0.0624**	-0.0629***	-0.1512***	0.0019
	(0.0195)	(0.0151)	(0.0407)	(0.0057)
Share of German employees	-0.1309***	-0.0076	0.0539	0.0239
	(0.0385)	(0.0212)	(0.0475)	(0.0133)
Share of fulltime employees	-0.1250***	-0.1113***	-0.2329***	-0.0027
	(0.0175)	(0.0165)	(0.0444)	(0.0053)
Mean age of employees in firm	-0.0014*	-0.0003	0.0001	-0.0000
	(0.0007)	(0.0005)	(0.0013)	(0.0002)
Constant	0.6478***	0.0812*	0.3586***	-0.1795***
	(0.0642)	(0.0369)	(0.1038)	(0.0211)
Firm sector controls	✓	✓	✓	✓
Survey characteristics controls	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Shapley decomposition:				
Individual characteristics			3.93%	
Household characteristics			15.61%	
Job and Firm characteristics			69.24%	
Survey characteristics			0.61%	
Year FE			10.59%	
N	42,046	42,046	9,704	32,110
R2	0.0925	0.0734	0.0545	0.0613
Mean dep. var.	0.2308	-0.0678	0.1668	-0.1391

Notes: This table summarizes coefficients of various operationalisations of the reporting error on potential correlates on the individual, household- and firm level, based on Equation 3.1. Dependent variables are a binary indicator for overreporting (column 1), the difference in log wages (column 2), and the difference in log wages split between overreporting (column 3) and underreporting individuals (column 4). Correct reporting is defined as no or small reporting bias within a +/- 1.25 percentile range of the annual reporting bias. All regressions include year-fixed effects. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.0. Factors of Reporting Bias, Zero Bias Defined as +/ − 5% around 0

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Individual Characteristics				
Age	-0.0004	0.0000	0.0002	0.0002
	(0.0003)	(0.0002)	(0.0007)	(0.0001)
Woman	-0.0207*	-0.0160*	-0.0196	-0.0069*
	(0.0086)	(0.0067)	(0.0270)	(0.0029)
Years of education	-0.0012	0.0007	-0.0001	0.0017***
	(0.0010)	(0.0009)	(0.0032)	(0.0004)
Region (ref. cat: North)				
East	-0.0005	-0.0080	-0.0366*	-0.0001
	(0.0104)	(0.0056)	(0.0153)	(0.0036)
South	-0.0052	0.0092	0.0241	0.0033
	(0.0103)	(0.0070)	(0.0225)	(0.0036)
West	0.0085	0.0168*	0.0257	0.0077*
	(0.0095)	(0.0065)	(0.0212)	(0.0033)
Migration background (ref. cat: none)				
Direct migration background	-0.0156	-0.0098	-0.0177	-0.0041
	(0.0124)	(0.0066)	(0.0158)	(0.0048)
Indirect migration background	-0.0084	-0.0019	-0.0033	-0.0013
	(0.0162)	(0.0091)	(0.0230)	(0.0072)
Personality Traits				
Openness (std.)	0.0020	-0.0016	-0.0042	-0.0015
	(0.0045)	(0.0027)	(0.0082)	(0.0015)
Conscientiousness (std.)	0.0030	-0.0008	-0.0089	0.0012
	(0.0046)	(0.0030)	(0.0092)	(0.0019)
Extraversion (std.)	0.0132**	0.0067**	0.0074	0.0003
	(0.0040)	(0.0026)	(0.0084)	(0.0015)
Agreeableness (std.)	-0.0084	-0.0049	-0.0004	-0.0021
	(0.0044)	(0.0031)	(0.0103)	(0.0015)
Neuroticism (std.)	0.0008	-0.0003	-0.0016	-0.0004
	(0.0042)	(0.0027)	(0.0085)	(0.0013)
Household characteristics				
Married	0.0000	0.0022	0.0240	-0.0059*
	(0.0090)	(0.0046)	(0.0125)	(0.0029)
Household Size	-0.0076*	-0.0019	0.0155	-0.0040**
	(0.0035)	(0.0027)	(0.0097)	(0.0013)
Number of kids in HH	0.0087	0.0023	-0.0179	0.0055***
	(0.0044)	(0.0035)	(0.0121)	(0.0016)
Partner	0.0043	0.0011	-0.0277	0.0104**
	(0.0102)	(0.0056)	(0.0145)	(0.0035)
Partner × Partner Inc in 1,000€	0.0058***	0.0031**	0.0019	0.0013*
	(0.0016)	(0.0012)	(0.0027)	(0.0005)

Continued on next page

Factors of Reporting Bias, Zero Bias Defined as +/ − 5% around 0 - continued

	(1)	(2)	(3)	(4)
	Over Y/N	Diff of Logs	Diff of Logs (+)	Diff of Logs (−)
Job and firm characteristics				
White collar worker	-0.0504***	0.0031	0.0334	0.0108***
	(0.0086)	(0.0062)	(0.0188)	(0.0030)
Union memberhsip	-0.0325***	-0.0112*	0.0232	-0.0073**
	(0.0075)	(0.0056)	(0.0221)	(0.0028)
Working Hours (ref. cat.: <35 hours)				
Full-time (35-44)	-0.0583***	-0.0013	0.0019	0.0327***
	(0.0078)	(0.0055)	(0.0186)	(0.0032)
Over-time (45+)	-0.0126	0.0129	0.0116	0.0306***
	(0.0101)	(0.0067)	(0.0209)	(0.0036)
Firm Size (ref. cat.: <10 employees)				
10-49 employees	-0.0322**	-0.0374***	-0.0511	-0.0097**
	(0.0115)	(0.0101)	(0.0279)	(0.0037)
50-249 employees	-0.0441***	-0.0487***	-0.0634*	-0.0111**
	(0.0117)	(0.0104)	(0.0298)	(0.0039)
250+ employees	-0.0582***	-0.0641***	-0.0828**	-0.0215***
	(0.0123)	(0.0101)	(0.0283)	(0.0041)
Pay tercile firm (ref. cat.: 1st Tercile)				
2. Tercile paying firm	-0.1091***	-0.0598***	-0.0284	-0.0220***
	(0.0084)	(0.0056)	(0.0146)	(0.0027)
3. Tercile paying firm	-0.1895***	-0.1083***	-0.0260	-0.0506***
	(0.0100)	(0.0074)	(0.0221)	(0.0036)
Workforce composition				
Share of highly qualified employees	0.0104	0.0155	-0.0261	0.0157*
	(0.0189)	(0.0144)	(0.0542)	(0.0068)
Share of female employees	-0.0629**	-0.0629***	-0.1547***	0.0056
	(0.0194)	(0.0151)	(0.0421)	(0.0058)
Share of German employees	-0.1319***	-0.0076	0.0549	0.0249
	(0.0383)	(0.0212)	(0.0484)	(0.0139)
Share of fulltime employees	-0.1217***	-0.1113***	-0.2397***	0.0000
	(0.0174)	(0.0165)	(0.0459)	(0.0055)
Mean age of employees in firm	-0.0013	-0.0003	0.0001	-0.0001
	(0.0007)	(0.0005)	(0.0014)	(0.0002)
Constant	0.6160***	0.0812*	0.3832***	-0.2027***
	(0.0623)	(0.0369)	(0.1096)	(0.0220)
Firm sector controls	✓	✓	✓	✓
Survey characteristics controls	✓	✓	✓	✓
Year FE	✓	✓	✓	✓
Shapley decomposition:				
Individual characteristics			3.93%	
Household characteristics			15.61%	
Job and Firm characteristics			69.24%	
Survey characteristics			0.61%	
Year FE			10.59%	
N	42,046	42,046	9,375	30,979
R2	0.0879	0.0734	0.0582	0.0571
Mean dep. var.	0.2230	-0.0678	0.1725	-0.1441

Notes: This table summarizes coefficients of various operationalisations of the reporting error on potential correlates on the individual, household- and firm level, based on Equation 3.1. Dependent variables are a binary indicator for overreporting (column 1), the difference in log wages (column 2), and the difference in log wages split between overreporting (column 3) and underreporting individuals (column 4). Correct reporting is defined as no or small reporting bias within a +/- 5 percentile range of the annual reporting bias. All regressions include year-fixed effects. Standard errors in parentheses. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 3.B.0. Detailed Shapley Decomposition

Factor	All
Individual characteristics	
Socio-demographic variables	2.51%
Personality traits	1.42%
Household characteristics	
Household characteristics	15.61%
Job and Firm characteristics	
Job characteristics	1.31%
Number of employees	29.50%
Wage tercile firm	15.48%
Sector	17.66%
Workforce composition	5.29%
Survey characteristics	
Survey characteristics	0.61%
Year FE	
Year FE	10.59%
N	42,046
R2	0.07

Notes: Table shows the results of the Shapley decomposition based on an OLS regression with the log of reporting bias (log) as dependent variable and all individual, household, and firm variables included in Table 3.3 as explanatory variables. Shapley values are computed for the groups of variables.

References

Abowd, J.M., and M.H. Stinson. 2013. "Estimating measurement error in annual job earnings: a comparison of survey and administrative data." *Review of Economics and Statistics* 95 (5): 1451–67. [148]

Acemoglu, D. 1997. "Training and innovation in an imperfect labour market." *Review of Economic Studies* 64 (3): 445–64. [121]

Adda, J., C. Dustmann, and K. Stevens. 2017. "The career costs of children." *Journal of Political Economy* 125 (2): 293–337. [143]

Angel, Stefan, Franziska Disslbacher, Stefan Humer, and Matthias Schnetzer. 2019. "What did you really earn last year?: Explaining measurement error in survey income data." *Journal of the Royal Statistical Society Series A* 182 (4): 1411–37. https://doi.org/10.1111/rssa.12463. [122, 124, 125, 135, 137]

Angel, Stefan, Richard Heuberger, and Nadja Lamei. 2018. "Differences between household income from surveys and registers and how these affect the poverty headcount: Evidence from the Austrian SILC." *Social Indicators Research* 138: 575–603. [135, 137]

Antoni, Manfred, Mattis Beckmannshagen, Markus M. Grabka, Sekou Keita, and Parvati Trübswetter. 2023. "Befragungsdaten der SOEP-Core-, IAB-SOEP Migrationsstichprobe, IAB-BAMF-SOEP Befragung von Geflüchteten und SOEP-Innovationssample verknüpft mit administrativen Daten des IAB (SOEP-CMI-ADIAB) 1975-2020." *FDZ-Datenreport, 03/2023(de)* (Nuremberg), https://doi.org/10.5164/IAB.FDZD.2303.de.v1. [126]

Antoni, Manfred, Daniel Bela, and Basha Vicari. 2019. "Validating earnings in the German National Educational Panel Study. Determinants of measurement accuracy of survey questions on earnings." *methods, data, analyses* 13 (1): 32. https://doi.org/10.12758/MDA.2018.08. [122, 124, 125, 135, 137, 138]

Arni, P., M. Caliendo, S. Künn, and K.F. Zimmermann. 2014. "The IZA evaluation dataset survey: A scientific use file." *IZA Journal of European Labor Studies* 3 (6): 1–20. [122]

Becker, G. S. 1962. "Investment in human capital: A theoretical analysis." *Journal of Political Economy* 70 (5): 9–49. [142]

Becker, G. S., and Barry R. Chiswick. 1966. "Education and the Distribution of Earnings." *American Economic Review* 56 (1/2): 358–69. [121]

Bee, Adam, Joshua Mitchell, Nikolas Mittag, Jonathan Rothbaum, Carl Sanders, Lawrence Schmidt, and Matthew Unrath. 2023. "National Experimental Wellbeing Statistics." *SEHSD Working Paper, 2023-02*. [148]

Bertrand, M., C. Goldin, and L. F. Katz. 2010. "Dynamics of the gender gap for young professionals in the financial and corporate sectors." *American Economic Journal: Applied Economics* 2 (3): 228–55. [143]

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender identity and relative income within households." *Quarterly Journal of Economics* 130 (2): 571–614. [138]

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Do people mean what they say? Implications for subjective survey data." *American Economic Review* 91 (2): 67–72. [123, 124, 146]

Blau, Francine D, and Lawrence M Kahn. 2017. "The gender wage gap: Extent, trends, and explanations." *Journal of Economic Literature* 55 (3): 789–865. [143]

Bollinger, C. R., B. T. Hirsch, C.M. Hokayem, and J. P. Ziliak. 2019. "Trouble in the Tails? What We Know about Earnings Nonresponse Thirty Years after Lillard, Smith, and Welch." *Journal of Political Economy* 127 (5): 2143–85. [128]

Bollinger, Christopher R. 1998. "Measurement error in the current population survey: A nonparametric look." *Journal of Labor Economics* 16 (3): 576–94. [137]

Bollinger, Christopher R, Barry T Hirsch, Charles M Hokayem, and James P Ziliak. 2018. "The good, the bad and the ugly: measurement error, non-response and administrative mismatch in the CPS." *Working Paper, Gatton College of Business, University of Kentucky.* [122, 130]

Bollinger, Christopher R, and Iva V Tasseva. 2023. "Income source confusion using the SILC." *Public Opinion Quaterly* 87 (S1): 542–74. [124]

Borghans, L., A.L. Duckworth, J.J. Heckman, and B. Ter Weel. 2008. "The economics and psychology of personality traits." *Journal of Human Resources* 43 (4): 972–1059. [122]

Bossler, Mario, and Christian Westermeier. 2020. "Measurement error in minimum wage evaluations using survey data." *IAB-Discussion Paper, 11/2020.* [125]

Bound, John, Charles Brown, Greg Duncan, and Willard L. Rodgers. 1994. "Evidence on the validity of cross-sectional and longitudinal labor market data." *Journal of Labor Economics* 12 (3): 345–68. https://doi.org/10.1086/298348. [124]

Bound, John, Charles Brown, and Nancy Mathiowetz. 2001. "Chapter 59 - Measurement Error in Survey Data," edited by James J. Heckman and Edward Leamer, 5: 3705–843. Handbook of Econometrics. Elsevier. https://doi.org/https://doi.org/10.1016/S1573-4412(01)05012-7. [122, 124]

Bound, John, and Alan Krueger. 1991. "The extent of measurement error in longitudinal earnings data: Do two wrongs make a right?" *Journal of Labor Economics* 9 (1): 1–24. https://doi.org/10.1086/298256. [122, 124, 130, 133, 137]

Boyce, C.J., G.D. Brown, and S.C. Moore. 2010. "Money and happiness: Rank of income, not income, affects life satisfaction." *Psychological Science* 21 (4): 471–75. [144]

Bricker, Jesse, and Gary V Engelhardt. 2008. "Measurement error in earnings data in the health and retirement study." *Journal of Economic and Social Measurement* 33 (1): 39–61. [137]

Cheung, F., and R.E. Lucas. 2015. " When does money matter most? Examining the association between income and life satisfaction over the life course." *Psychological Aging* 30 (1): 120–35. [144]

Dauth, W., and J. Eppelsheimer. 2020. "Preparing the sample of integrated labour market biographies (SIAB) for scientific analysis: A guide." *Journal for Labour Market Research* 54 (10). [127, 128]

Deaton, Angus. 2003. "Health, inequality, and economic development." *Journal of Economic Literature* 41 (1): 113–58. [121]

Deaton, Angus. 2008. "Income, health, and well-being around the world: Evidence from the Gallup World Poll." *Journal of Economic Perspectives* 22 (2): 53–72. [121]

Destatis. 2024. "Qualität der Arbeit - Gender Pay Gap." https://www.destatis.de/DE/Themen/Arbeit/Arbeitsmarkt/Qualitaet-Arbeit/Dimension-1/gender-pay-gap-1.html. [144]

Duncan, Greg, and Daniel Hill. 1985. "An investigation of the extent and consequences of measurement error in labor-economic survey data." *Journal of Labor Economics* 3 (4): 508–32. https://doi.org/10.1086/298067. [124, 125]

Frijters, P., J. P. Haisken-DeNew, and M. A. Shields. 2004. "Money does matter! Evidence from increasing real income and life satisfaction in East Germany following reunification." *American Economic Review* 94 (3): 730–40. [144]

Frodermann, C., A. Schmucker, S. Seth, and P. vom Berge. 2021. "Stichprobe der Integrierten Arbeitsmarktbiografien (SIAB) 1975-2019." *FDZ Datenreport, 01/2021 (de).* [128]

Fuchs, Victor R, Alan B Krueger, and James Poterba. 1997. "Why do economists disagree about policy?" *NBER Working Paper, 6151.* [121]

Goebel, Jan, Markus M Grabka, Stefan Liebig, Martin Kroh, David Richter, Carsten Schröder, and Jürgen Schupp. 2019. "The German Socio-Economic Panel (SOEP)." *Jahrbücher für Nationalökonomie und Statistik* 239 (2): 345–60. [122]

Goldin, C. 2014. "A grand gender convergence: Its last chapter." *American Economic Review* 104 (4): 1091–119. [143]

Gottschalk, Peter. 2005. "Downward nominal-wage flexibility: Real or measurement error?" *Review of Economics and Statistics* 87 (3): 556–68. https://doi.org/10.1162/0034653054638328. [125]

Gottschalk, Peter, and Minh Huynh. 2005. "Validation Study of Earnings Data in the SIPP - Do Older Workers Have Larger Measurement Error?" *Center for Retirement Research Working Papers, 2005-07.* [137]

Gottschalk, Peter, and Minh Huynh. 2010. "Are earnings inequality and mobility overstated? The impact of nonclassical measurement error." *Review of Economics and Statistics* 92 (2): 302–15. https://doi.org/10.1162/rest.2010.11232. [125]

Hurst, E., G. Li, and B. Pugsley. 2014. "Are household surveys like tax forms? Evidence from income underreporting of self-employed." *Review of Economics and Statistics* 96 (1): 19–33. [125]

Isaoglu, Aysen. 2010. "Occupational affiliation data and measurement errors in the German Socio-Economic Panel." *SOEP Papers on Multidisciplinary Panel Data Research, 318.* [124]

Juarez, F.C. 2012. "SHAPLEY2: Stata module to compute additive decomposition of estimation statistics by regressors or groups of regressors." *Statistical Software Components S457543, Boston College Department of Economics.* [136]

Kapteyn, Arie, and Jelmer Ypma. 2007. "Measurement error and misclassification: A comparison of survey and administrative data." *Journal of Labor Economics* 25 (3): 513–51. https://doi.org/10.1086/513298. [130]

Kearl, James R, Clayne L Pope, Gordon C Whiting, and Larry T Wimmer. 1979. "A confusion of economists?" *American Economic Review* 69 (2): 28–37. [121]

Kim and Gary Solon. 2005. "Implications of mean-reverting measurement error for longitudinal studies of wages and employment." *Review of Economics and Statistics* 87 (1): 193–96. https://doi.org/10.1162/0034653053327685. [125]

Kim and C.R. Tamborini. 2014. "Response error in earnings: An analysis of the survey of income and program participation matched with administrative data." *Sociological Methods & Research* 43 (1): 39–72. https://doi.org/10.1177/0049124112460371. [122, 125, 137]

Kleven, H., C. Landais, J. Posch, A. Steinhauer, and J. Zweimüller. 2019. "Child penalties across countries: Evidence and explanations." *AEA Papers and Proceedings* 109: 122–26. [143]

Kleven, H., C. Landais, and J. E. Sogaard. 2019. "Children and gender inequality: Evidence from Denmark." *American Economic Journal: Applied Economics* 11 (4): 181–209. [143]

Künn, Steffen. 2015. "The challenges of linking survey and administrative data." *IZA World of Labor, 214.* [121, 122]

Madeira, Carlos, Paula Margaretic, Felipe Martínez, and Pedro Roje. 2022. "Assessing the quality of self-reported financial information." *Journal of Survey Statistics and Methodology* 10 (5): 1183–210. https://doi.org/10.1093/jssam/smab059. [124]

Meijer, E., S. Rohwedder, and T. Wansbeek. 2012. "Measurement Error in Earnings Data: Using a Mixture Model Approach to Combine Survey and Register Data." *Journal of Business & Economic Statistics* 30 (2): 191–201. [148]

Meyer, Bruce D, and Nikolas Mittag. 2019. "Using linked survey and administrative data to better measure income: Implications for poverty, program effectiveness, and holes in the safety net." *American Economic Journal: Applied Economics* 11 (2): 176–204. [125]

Meyer, Bruce D., Nikolas Mittag, and Derek Wu. 2024. "Race, Ethnicity, and Measurement Error." In *Race, Ethnicity, and Economic Statistics for the 21st Century*. University of Chicago Press. [125]

Mincer, J. 1974. "Age and Experience Profiles of Earnings." In *Schooling, Experience, and Earnings*, edited by J. Mincer, 64–82. National Bureau of Economic Research. [142]

Moore, J.C., L.L. Stinson, and E.J. Welniak. 2000. "Income measurement error in surveys: a review." *Journal of Official Statistics* 16 (4): 331–61. [124]

Mulligan, C.B., and Y. Rubinstein. 2008. "Selection, investment, and women's relative wages over time." *Quarterly Journal of Economics* 123 (3): 1061–110. [121]

Pencavel, John. 1986. "Chapter 1: Labor supply of men: A survey," 1: 3–102. Handbook of Labor Economics. Elsevier. https://doi.org/https://doi.org/10.1016/S1573-4463(86)01004-0. [121]

Pischke, Jörn-Steffen. 1995. "Measurement error and earnings dynamics: Some estimates from the PSID validation study." *Journal of Business & Economic Statistics* 13 (3): 305–14. https://doi.org/10.1080/07350015.1995.10524604. [122, 124]

Prati, Alberto. 2017. "Hedonic recall bias. Why you should not ask people how much they earn." *Journal of Economic Behavior & Organization* 143: 78–97. [124, 146]

Richter, David, and Jürgen Schupp. 2015. "The SOEP Innovation Sample (SOEP IS)." *Schmollers Jahrbuch* 135 (3): 389–99. https://doi.org/10.3790/schm.135.3.389. [127]

Roth, Anja, and Michaela Slotwinski. 2021. "Gender norms and income misreporting within households." *CESifo Working Paper, 7298*. [122, 123, 125, 137, 138]

Sakshaug, J.W., and F. Kreuter. 2012. "Assessing the Magnitude of Non-Consent Biases in Linked Survey and Administrative Data." *Survey Research Methods* 6 (2): 113–22. [127]

Sandmo, A. 1970. "The effect of uncertainty on saving decisions." *Review of Economic Studies* 37 (3): 353–60. [121]

Schmillen, Achim, Matthias Umkehrer, and Till von Wachter. 2024. "Measurement Error in Longitudinal Earnings Data: Evidence from Germany." *Journal for Labour Market Research,* Forthcoming. [124, 133]

Shorrocks, A.F. 2013. "Decomposition procedures for distributional analysis: A unified framework based on the Shapley value." *Journal of Economic Inequality* 11: 99–126. [136]

SOEP. 2022. "Sozio-oekonomisches Panel (SOEP), Version 37, data from 1984-2020 (SOEP-Core v37, EU-Edition)," https://doi.org/10.5684/soep.core.v37eu. [126]

SOEP-IS. 2020. "SOEP Innovation Sample (SOEP-IS), data from 1998-2019," https://doi.org/10.5684/soep.is.2019. [126]

Solon, Gary. 1992. "Intergenerational income mobility in the United States." *American Economic Review* 32 (3): 393–408. [125]

Stantcheva, Stefanie. 2022. "How to run surveys: A guide to creating your own identifying variation and revealing the invisible." *NBER Working Paper, No. 30527,* https://doi.org/10.3386/w30527. [122]

Stüber, H., M. Grabka, and D. Schnitzlein. 2023. "A tale of two data sets: Comparing German administrative and survey data using wage inequality as an example." *Journal for Labour Market Research* 57 (8). [125]

Valet, Peter, Jule Adriaans, and Stefan Liebig. 2019. "Comparing survey data and administrative records on gross earnings: nonreporting, misreporting, interviewer presence and earnings inequality." *Quality & Quantity* 53: 471–91. https://doi.org/10.1007/s11135-018-0764-z. [124]