Essays in Labor Economics

Inaugural-Dissertation

zur Erlangung des Grades eines Doktors
der Wirtschafts- und Gesellschaftswissenschaften

durch

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

vorgelegt von

Lenard Paul Simon

aus Sinsheim

Bonn

2024

Dekan:	Prof. Dr. Jürgen von Hagen
Erstreferent:	Prof. Dr. Thomas Dohmen
Zweitreferent:	JProf. Dr. Amelie Schiprowski
Tag der mündlichen Prüfung:	02.02.2024

Acknowledgements

I want to thank my advisors Thomas Dohmen and Amelie Schiprowski for their invaluable support and guidance throughout my academic journey. From the very beginning of the dissertation phase, I always felt unconditionally supported by both of them, which made my Ph.D. experience much more enjoyable for me.

Writing this thesis would not have been possible without my great co-authors. I want to thank Tim Bayer, Hans-Martin von Gaudecker, Radost Holler, Jakob Wegmann, and Christian Zimpelmann.

Working together with Jakob was one of the greatest experiences during my Ph.D. We shared most of our academic journey, starting from the Bachelor's and Master's in Mannheim to living and working together in Bonn. I am very grateful that we have spent this journey together and hope we can continue our great cooperation in future projects!

Another highlight of my Ph.D. was meeting Radost. Sharing an office and numerous coffee breaks made us very good friends and co-authors. I benefited greatly from your support and knowledge and hope to continue doing so in the future!

Besides my coauthors, I also want to thank all the friends I made along the way, among them, Anna, Alex, Andrea, Dominik, Fabian, Florian, Gianmarco, Simon, and in particular, Miguel.

I also want to thank my parents and my uncle for always supporting me. As I progressed through the academic system, it became harder to explain what I am doing, but your support never ceased.

Last but not least, I am very grateful to Simone Jost, who made my interactions with the university administration a lot less stressful and proactively solved all my administrative problems. Further, I would like to thank the Excellence-Cluster Econtribute and the CRC Transregio 224 for always supporting me. They did not only provide me with the necessary financial support for my projects and my research stay in Barcelona but also allowed me to benefit from frequent interactions with other young researchers in Cologne and Mannheim.

Contents

List of Figures

List of Tables

Introduction

As Western societies age, their working-age populations are declining, resulting in a scarcity of labor. Consequently, these societies face reduced production opportunities while shouldering an increasing burden of elderly care. This leads to governments encountering rising expenditures alongside a decreasing tax base. One potential avenue to mitigate this problem is by increasing the labor supply among the working-age population.

In the canonical model of the labor market, an individual's labor supply is her optimal choice of working hours derived from the trade-off between consumption and leisure, given equilibrium wages and consumption prices. However, in reality, numerous institutional frictions impact labor supply as well, such as misaligned incentives, limited geographic mobility, and inadequate work-family compatibility, among others. Due to these frictions, individuals may choose to work fewer hours than what they would consider optimal when faced with a mere leisure-consumption trade-off. Hence, reducing these frictions could potentially expand the labor supply of the working-age population and, consequently, alleviate the prevailing scarcity of labor in Western societies today. This thesis consists of three independent chapters covering the causes and consequences of such frictions and how they can be eliminated with the help of policies.

Chapter 1 (joint work with Jakob Wegmann and Tim Bayer) investigates how misperceptions about the taxation of labor income affect the labor supply of married women in Germany. We study this question in the context of withholding taxes paid by married couples. In a first step, we document with the help of a survey that less than 20 % of the interviewed married individuals understand that withholding taxes are tax prepayments which are fully credited against the final income tax and, therefore, do not determine the income tax burden. Making use of a reform that decreased the withholding tax burden for some married women more than for others, while inducing no differences in income taxes, allows us to then estimate the elasticity of labor income with respect to the withholding tax. In line with our survey findings, we show with administrative tax records that women adjust their labor supply following a change in withholding taxes. Importantly, the German institutional setting allows couples to partly redistribute the withholding tax burden from one partner to the other, and the majority shifts

parts of the withholding tax burden from the husband to the wife. Our results suggest that the increased withholding tax burden of married women in Germany contributes to their low labor supply. The finding also highlights that governments should be aware that overwithholding results in an overestimation of the actual income tax and thus distorts labor supply incentives.

Chapter 2 (joint work with Hans-Martin von Gaudecker, Radost Holler, and Christian Zimpelmann) focuses on, how, the shift to remote work induced by the CoVid-19 pandemic affected the parental division of non-market and market work. We do so using representative panel survey data and population-wide administrative data from the Netherlands, spanning the years 2012–2021. We argue that we are able to isolate the effect of working from home since (1) the remote work potential was realized to a large extent only during the Covid-19 pandemic, (2) schools and daycare were available in the Netherlands throughout the pandemic, with a brief exception in spring 2020, and (3) generous support schemes kept working hours, unemployment and earnings at a similar level as before the pandemic. First, we find that parents who report a large potential to work remotely strongly increase their remote work hours and reduce their commuting hours after the onset of the Pandemic. We further show that fathers have more remote work potential than mothers mainly due to their higher working hours. As a consequence, they also asymmetrically gain more job flexibility through the shift to remote work than mothers. We then explore how these differences in flexibility gains affect the division of labor within households. First, we show that parents who can work from home increase their childcare provision in response to the shift to remote work. Due to the asymmetric distribution of the potential to work from home, this leads to a decrease in the gender gap in childcare provision as fathers provide more childcare and mothers less. Using the large-scale administrative data together with an event-study difference-in-differences design, we examine whether the shift in remote work also impacted the labor supply decisions of parents. We find that, indeed, spouses whose partners profit more from the shift to remote work, increase their labor supply in response to this shift. This study shows that residual gender differences in the division of labor are not solely driven by gender-related preferences or societal norms, but also by the challenges of combining full-time employment with childcare needs. This often results in the adoption of the male breadwinner model after the birth of the first child, where the father works full-time and the mother chooses either no or part-time employment. Government policies that increase the adoption and acceptance of remote work could therefore be an effective measure to reduce gender inequalities among parents and to increase the labor supply of mothers.

Chapter 3 analyzes how the initial allocation of refugees in Germany and subsequent restrictions on their spatial mobility affect their labor supply and social integration into society. I study these questions using data from the IAB-BAMF-SOEP Survey of Refugees, a representative survey of refugees who arrived in Germany

between 2013 and 2016. For the effects on labor supply, I analyze the likelihood of having found full- or part-time employment and for the effects on social integration, I examine self-reported measures for the frequency of social interactions with Germans. As a first step, I confirm the finding that being assigned to a county with high unemployment greatly worsens the economic and social integration of a refugee. I then use a difference-in-differences approach to analyze the impact of local residence restrictions on the integration of refugees in Germany making use of an institutional reform that limited the free choice of residence of refugees differently across federal states. My results show that stronger enforcement of residence restrictions at the level of counties and municipalities has heterogeneous effects on the integration of refugees. While I find evidence that such local residence restrictions have a positive impact on the economic and social integration of refugees being assigned to low-unemployment counties, I detect negative effects for those being assigned to counties with high unemployment. Taken together, these findings suggest that when designing refugee dispersal policies, policymakers should be aware of the strong negative impacts of unfavorable labor market conditions on the integration of a refugee. Moreover, I show that local residence restrictions, as designed in Germany, are not necessarily an effective measure to improve the integration of refugees as their success is highly influenced by local labor market conditions.

Chapter 1

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

Joint with Jakob Wegmann and Tim Bayer

1.1 Introduction

Most countries use third-party withholding to collect income taxes during the year. Typically, employers withhold monthly prepayments to income taxes which are then fully credited against the final income tax liabilities of their employees. This provides governments with a constant income stream during the year and increases tax compliance (Slemrod, 2019; Bagchi and Dušek, 2021). However, withholding tax rates do not necessarily reflect true effective income tax rates. Often, there is overwithholding as many taxpayers pay higher withholding taxes than actual income taxes (Engström, Nordblom, Ohlsson, and Persson, 2015; Rees-Jones, 2018; Gelman, Kariv, Shapiro, and Silverman, 2022; Hauck and Wallossek, 2023). In this case, a lump-sum tax refund is paid to employees by the government after the end of the tax year. Conversely, in the case of under-withholding, employees must make an additional lump-sum tax payment to the government. This interlinkage between withholding taxes and income taxes makes it more complex to understand the income tax system. As a consequence, the design of withholding taxes can distort labor supply when individuals use their monthly take-home pay to infer their income tax burden.

It is difficult to study the effects of withholding taxes, as they are typically a function of the income tax. Therefore, it is usually not possible to use reforms of the income tax system to draw conclusions regarding the role of withholding

* 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.

taxes. However, the German income tax system offers an institutional setting that allows investigating the effects of a reform of withholding taxes on labor supply. We illustrate the core feature of the institutional setting in Figure 1.1.1 which displays average withholding tax rates by gender and labor income in Germany. Conditional on labor income, married women pay, on average, higher withholding tax rates than married men. This is the consequence of the German withholding tax system that allows couples to shift parts of the withholding tax burden from one partner to the other by choosing certain withholding tax classes (*"Lohnsteuerklassen"*). As a consequence of the choice of withholding tax classes, couples with identical income structures can end up paying different withholding taxes. Importantly, the decision on withholding tax classes does not affect the final income tax rate. However, a married couple can minimize its joint withholding tax burden by shifting some part of the withholding tax burden 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 1.1.1: Married women are typically the secondary wage earner and hence face, on average, a higher withholding tax rate conditional on labor income.

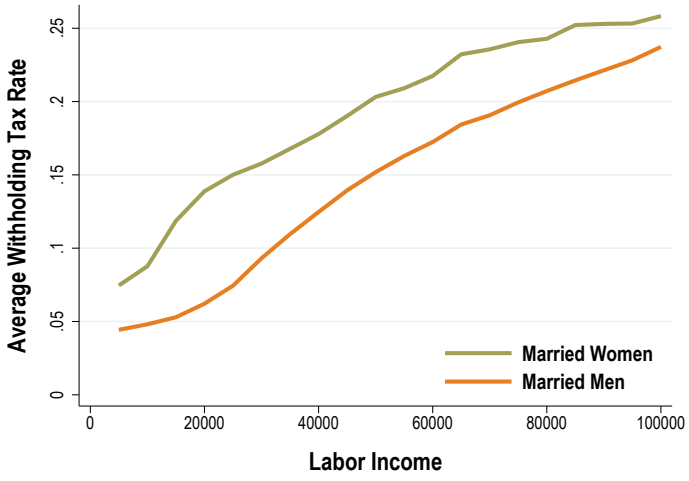


Figure 1.1.1. Average Withholding Tax Rate by Gender

Notes: The figure displays the average realized withholding tax rate by gender for married couples in Germany for annual labor income levels of up to 100,000 €. Calculations are based on a 10 % sample of income tax returns in 2010. The figure illustrates that through the choice of withholding tax classes (*"Lohnsteuerklassen"*), married couples shift a substantial share of the withholding tax burden from men to women (RDC of the Federal Statistical Office and Statistical Offices of the Federal States, 2010).

Given a fixed income tax schedule, these differences in withholding tax rates should have no real effects.[1] If individuals react strongly to withholding taxes, this suggests that withholding taxes are misunderstood and used as a proxy for income taxes. This could be due to the larger salience of withholding taxes compared to income taxes. While withholding taxes are directly observed on the monthly payslip, the actual income taxes can only be inferred after receiving the final income tax statement.[2] Withholding taxes could therefore constitute a central cornerstone in understanding how people learn about the tax rates they face.

An additional motivation to study the effect of withholding taxes on labor supply is based on joint taxation. The underlying rationale of joint taxation is based on the idea that married households act as one economic unit like in unitary household models, and are, consequently, taxed jointly. Thereby, governments set the identical economic incentives for both partners irrespective of their individual wage income and governments remain impartial about the distribution of labor supply within the household. Consequently, in joint taxation systems, the government also does not have to take a stance on how the marriage bonus induced by joint taxation should be distributed within the household, as it is paid out to the household as a whole.

However, withholding taxes are inherently individual taxes. Therefore, in countries with tax withholding and joint taxation, policymakers cannot remain impartial when deciding how the marriage bonus should be distributed between partners throughout the year. Also, governments are forced to decide on the individual marginal and average withholding tax rates. Interestingly, the implemented solutions vary substantially between countries, as we will discuss in more detail later in the paper. The gap in average tax rates presented in Figure 1.1.1 reflects the consequence of the implemented withholding tax system for married individuals in Germany. If individuals use payslips or monthly transfers to infer their net income, the gender gap in withholding taxes might contribute to a misperception about the returns to labor within couples.

To measure knowledge about withholding taxes and to better understand decision-making processes on household finances in the German population, we conducted a pre-registered online survey. We find that more than 80 % of the surveyed married and employed individuals wrongly think that the choice of withholding tax classes affects the final income tax burden. This suggests that individuals with the same income tax burden, but with differing withholding tax rates,

1. This holds in a unitary household model and in the absence of interest rates and liquidity constraints. Shapiro and Slemrod (1995) find that the financial situation of households is not correlated with the propensity to consume withholding tax savings.

2. Moreover, we find with the help of a survey that in only 37 % of interviewed married couples who file a joint tax declaration both spouses take part in preparing the tax declaration. This indicates that many individuals do not invest much time in understanding the final income tax statement.

might perceive their income tax burden differently and consequently make different labor supply decisions. Additionally, we investigate the impact of the system of withholding tax classes on the organization of household finances. As seen in Figure 1.1.1, couples often choose withholding tax classes that shift parts of the withholding tax burden from men to women. For only about 40 % of these couples, we monitor patterns that are consistent with compensating these women, i.e., the husband making a relatively larger monetary transfer to his wife, or to a shared bank account, than vice versa. If women are not compensated for the unequal distribution of the withholding tax burden, the observed pattern of assignment of withholding tax classes lowers their own disposable net income. Hence, they might overestimate their individual income tax burden, which can decrease their incentives to work and potentially also affect their bargaining power within the couple.

Motivated by these findings, we investigate empirically whether withholding taxes impact labor income. The German institutional context provides us with a unique opportunity to causally study the effects of withholding taxes. Germany offers different withholding tax schedules for couples so that households with an identical income structure and income tax burden can pay very different withholding taxes. However, the choice of withholding tax schedules is not random as households self-select into them. Accordingly, the differing levels of withholding taxes stemming from the different schedules cannot be exploited for a causal analysis. We circumvent this problem by analyzing a tax reform in 2010 that cut withholding taxes for married women differently across withholding tax schedules. Applying a Difference-in-Differences setup with continuous treatment intensity, we are able to investigate how married women react to a cut in withholding taxes while keeping income tax payments constant. The reform is the result of a technical detail in the automatic deduction of health care costs that passed the German parliament as part of a larger income tax reform. There was no discussion about the change in public sessions of the parliament and there exists not a single newspaper article about the reform. Hence, we expect no anticipation effect, and any change in labor income can be traced back to the cut in withholding taxes.

We conduct the analysis using administrative tax records from a 5 % sample of the German Taxpayer Panel (TPP) which contains extensive information on the population of taxpayers in Germany for the years 2001 to 2018 (RDC of the Federal Statistical Office and Statistical Offices of the Federal States, 2018). The dataset is very well suited for studying the effects of withholding taxes as it does not only provide information on income and the withholding tax class but also includes numerous other characteristics of the household that allow us to better understand underlying mechanisms in the heterogeneity analysis.

For married women, we estimate an elasticity of labor income with respect to the marginal net-of-withholding tax rate of about 0.1 using a static Diff-in-Diff. Estimating an event study Diff-in-Diff, we find that the treatment effect increases

monotonically over time. We attribute this change in the size of the treatment effect over time to the way taxpayers learn about their tax rates. We argue that employees use the information on withholding taxes from their monthly payslips to learn about their income taxes which takes time as they first have to realize that their monthly net wage has changed and then recognize the persistence of this change. Moreover, it might also take time to adapt one's labor supply, possibly after negotiations with one's employer or a change of employer.

The fact that individuals react to withholding taxes implies that governments should be careful when designing withholding tax schedules. As we show in this study, many countries introduce withholding tax systems for married couples to reduce overwithholding stemming from joint taxation benefits. We demonstrate that as soon as countries try to set withholding tax rates for married couples they have to decide what the individual marginal and average withholding tax rates are that each spouse faces and thereby have to make a decision on how the joint taxation benefit is divided among spouses. We show that different implementations can result in significantly different withholding tax rates for primary and secondary earners. Given our empirical results, this implies that the design of withholding tax systems cannot be incentive-neutral but always involves a decision on how to affect the incentives to work of primary and secondary earners. This is especially relevant as it shows that the withholding tax system can be used to increase the labor market participation of secondary earners.

Related Literature. In this paper, we provide the first real-world evidence on the effects of withholding taxes on labor supply.[3] Previous evidence comes from a laboratory experiment by Becker, Fooken, and Steinhoff (2019). Their paper studies the hypothesis that taxpayers have false perceptions of net labor income due to withholding taxes. Using treatments with different levels of withholding tax rates, they design their experiment in a way that these withholding tax rates and the corresponding adjustments of lump-sum payments should not influence the behavior of rational agents.[4] Contrary to standard economic theory, the authors, however, find that people describing themselves as money-motivated significantly reduce their effort when facing higher withholding tax rates.

We are not the first to study the effects of withholding taxes. Economists have generally expressed a positive view on overwithholding, as withholding taxes are associated with more savings, less consumption, liquidity for the government and

3. Buettner, Erbe, and Grimm (2019) show how the choice of withholding tax classes depends on spouses' labor income but they do not study the effect of withholding tax class choice on labor income.

4. Here, they model a world without interest rates and liquidity constraints which do not perfectly fit the real economy. Positive interest rates might give an incentive to have a low withholding tax rate because interest can be earned between paying the withholding tax and having to pay additional tax payments. Liquidity constraints might also give an incentive to have a low withholding rate to not run out of money during the year.

higher tax compliance. Hence, our results suggest that policymakers face a difficult trade-off when designing withholding taxes. We discuss the related research in Section 1.6.

Our paper contributes to the existing literature on the complexity of tax systems. Undoubtedly, the interlinkage between withholding tax and income tax and particularly the possibility to choose withholding tax classes add complexity for taxpayers. This complexity might impact their decision-making. Using an experimental setting, Abeler and Jäger (2015) find that taxpayers subject to more complex tax systems do not react to new taxes sufficiently. This shows that the complexity of tax systems can induce taxpayers to make irrational decisions. It is therefore relevant that, as well documented, an overwhelming majority of taxpayers do not understand how income taxation works. For example, many individuals do not know which tax rates apply to them personally (Enrick, 1963; Enrick, 1964; Wagstaff, 1965; Fujii and Hawley, 1988; Chetty, Friedman, and Saez, 2013; Lardeux, 2022), and they do not understand the difference between marginal and average tax rates (Liebmann and Zeckhauser, 2004; Gideon, 2017; Rees-Jones and Taubinsky, 2020). However, the literature on income taxation finds large elasticities of taxable income with respect to the income tax, which shows that people react to the amount of taxes they have to pay (Gruber and Saez, 2002; Saez, Slemrod, and Giertz, 2012; Neisser, 2021). People thus respond to income taxes even though they do not have a good understanding of them due to, e.g., mental or time constraints. This poses the question which heuristics individuals use to determine their response to income taxation. Throughout this paper, we document that withholding taxes serve as one of these heuristics in a complex system of income taxation.

Other research finds that taxpayers act on more salient parts of a tax system. Using a field experiment in a grocery store, Chetty, Looney, and Kroft (2009) find that, although consumers are aware of which tax rate they have to add, consumers' demand for goods is higher when sales taxes are not added to the price tag.[5] As linear commodity prices are relatively simple to understand and calculate, the authors take that as an indication for the hypothesis that behavioral responses of taxpayers could be very different from those predicted by standard economic theory in cases of more complex taxes such as income taxes. It has already been shown in the literature that withholding taxes substantially impact real-world decisions other than labor supply, namely saving, consumption and tax compliance. A detailed discussion of these papers can be found in Section 1.6.1.

The particularities of the German institutional setting allow us to also contribute to the literature on the determinants of the gender earnings gap which is particularly pronounced in Germany. As previously shown in Figure 1.1.1, mar-

5. Feldman and Ruffle (2015) arrive at a very similar finding.

ried women pay, conditional on labor income, higher withholding tax rates than married men. Therefore, we argue that, given our estimates, a reduction of withholding tax payments for married women in Germany might increase labor supply and thus labor income of married women. The existing system of withholding tax classes for married couples might then be an additional contributing factor to the gender gap in labor supply in Germany. We therefore contribute to a discussion of how to optimally design a tax system while creating the smallest possible detrimental incentives for labor supply of women and keeping states' budgets stable.

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 systems with joint taxation of married couples, in which marginal and average tax rates of secondary earners are increased, while those of primary earners are decreased. LaLumia (2008) studies the effects of the United States turning from an individual taxation scheme to joint taxation of married couples in 1948. She estimates that the reform decreased the employment likelihood of highly-educated married women by about two percentage points. Examining the 1971 abolishment of joint taxation of married couples in Sweden using register panel data, Selin (2014) finds that employment rose significantly more for wives of high-income earners after the reform. This is in line with expectation because this is the group that profited most from joint taxation so that joint taxation should have kept their labor supply substantially lower than it would have been without it. Bick and Fuchs-Schündeln (2017), based on Bick and Fuchs-Schündeln (2018), look at the United States and 13 European countries with joint taxation of married couples. They estimate that changing to a system of individual taxation while keeping government revenue constant would increase hours worked by women by more than 70 hours per year in ten of these countries. As an example, they calculate benefits of 113 annual hours for the United States and of even 280 annual hours for Germany.

Outline. The rest of the paper is structured as follows: Section 1.2 presents in detail the institutional setting and the results from our survey, thereafter Section 1.3 presents the data and the sample selection, and Section 1.4 explains our empirical strategy. Section 1.5 discusses the results and Section 1.6 investigates the policy implications of our findings. Section 1.7 concludes.

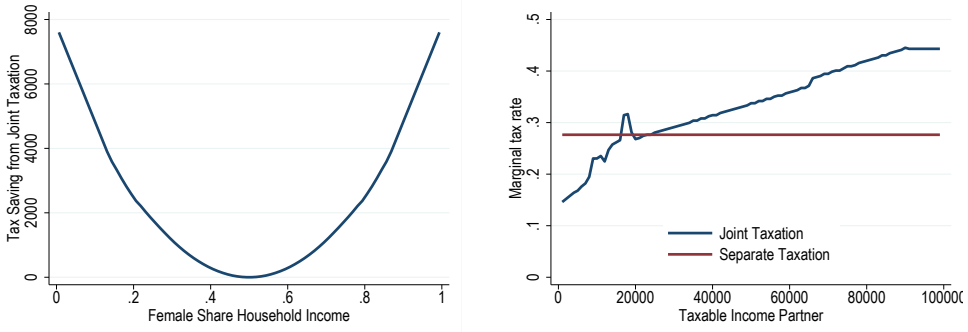
1.2 Institutional Setting

In this section, we first provide necessary 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. Finally, we shed more light on

the understanding of withholding taxes among married couples in Germany by presenting the results of our survey.

1.2.1 Income Taxation of Married Couples

Married couples in Germany have two different options when it comes to filing their income taxes. They can choose to either file their income taxes separately as if they were still single, or to file their income taxes jointly. By choosing the latter, couples can potentially realize joint taxation benefits.[6] 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 burden is then doubled. Due to the progressivity of the German income tax system, this creates joint taxation benefits for couples with differing marginal income tax rates. Put differently, for a fixed household income, a couple receives more joint taxation benefits the more unequal the intrahousehold distribution of income.



(a) Joint Tax Benefits for Joint Income 80,000 € **(b)** Marginal Income Tax for Income of 24,000 €

Figure 1.2.1. Joint Income Taxation

Notes: The figure illustrates the system of joint income taxation in Germany. Panel (a) plots the joint taxation benefits depending on the intra-household income distribution for a household with joint income of 80,000 €. Panel (b) 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. In this example, we assume that both spouses contribute to public health care, to the public pension system, and claim no further deductions.

We illustrate this feature in Figure 1.2.1a, where we plot the joint taxation benefit of a couple with an household income of 80,000 € against the female share in the household income. If both spouses contribute equally to the household

6. In fact, for the vast majority of couples choosing joint taxation is at least weakly better than choosing separate taxation. Only couples in which one partner has a significant amount of income replacement payments 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").

income, there are no benefits from joint taxation. If, however, one partner for example contributes 80 % to the household income, opting for joint taxation will save the couple around 2,000 € in yearly income taxes.

As a side effect of this joint taxation system, 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. Figure 1.2.1b shows that as soon as the partner income exceeds the own income, an individual is confronted with substantially higher marginal income tax rates. The marginal tax rate for an individual with an own income of 24,000 € is 27.5 % under separate taxation, but increases to approximately 35 % under joint taxation if their spouse has an income of 60,000 €.

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 burden.[7] Married couples can influence both the sum of their monthly withholding tax payments and the allocation of the withholding tax burden to each spouse. They can effectively choose between three different withholding tax schedules.[8] These withholding tax schedules assign each partner a certain withholding tax class, which determines the personal withholding tax payments.

Symmetric Schedule. After marriage, each couple in which both spouses receive labor income gets assigned the same “default” withholding tax schedule, which we will call the symmetric schedule. This withholding tax schedule is symmetric since it assigns each spouse the same withholding tax class “IV”. In this withholding tax

7. Germany levies withholding taxes, which are prepayments to the final income tax and which are withheld at source by employers on behalf of their employees. Usually, the withholding taxes are deducted from the monthly paycheck and then credited against the income tax liability at the end of the tax year. Such withholding taxes are part of the tax system of all developed countries with Switzerland as a special case as only employees living outside of Switzerland and/or without a permanent Swiss residence permit are subject to withholding taxes.

8. In our analysis, we leave out the fourth, least commonly chosen withholding tax schedule. This withholding tax schedule is called "IV with factor" and was introduced in 2010 with the goal to mitigate the negative effects of the men-/women-favoring withholding tax schedules, while still enabling couples to profit from the advantage of joint taxation during the year. To do so, 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 "IV with factor". Official government agencies estimate, however, that even 10 years after its introduction less than 1 % of the couples are using this schedule (*Kleine Anfrage Bundestag* 2019). We observe "IV with factor" as "IV" in the data.

class, the monthly withholding tax payments are calculated as if the individual was single, only taking into account the own individual income. Hence, for a couple without joint taxation benefits, the withholding tax would be the same as the income tax. If a couple realizes joint taxation benefits, the paid withholding tax of both spouses will exceed their final income tax liability and the couple will receive a tax refund after filing an income tax return. We illustrate this in Figure 1.2.2 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 benefits of 288 € as a lump sum tax refund after filing their income taxes.

To avoid this overpayment of withholding taxes during the year, a couple can decide to switch from the "default" symmetric schedule to a withholding tax schedule that aims at reducing the monthly withholding tax payments to account for the joint taxation benefits.[9]

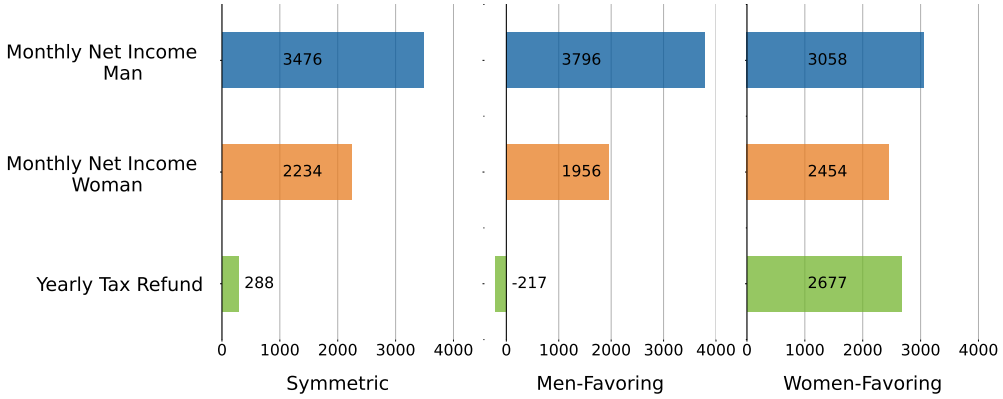


Figure 1.2.2. Example illustrating the different withholding tax schedules

Notes: The figure illustrates how the different withholding tax schedules affect the monthly net incomes of both spouses and the yearly tax refund in the year 2022. Net incomes are calculated for a household in which the husband earns 50,000 € and the wife earns 30,000 €. The assessed yearly income tax burden of the household is 11,181 € under the assumption that the couple claims no additional deductions. The figure shows how the different withholding tax schedules shift the withholding tax burden from one partner to the other and how they can affect the yearly refund from the final income tax.

Men- or Women-favoring Schedule. The most popular alternative withholding tax schedules are the men-/women-favoring withholding tax schedules. In those schedules, one spouse is assigned the favorable withholding tax class ("III"), while

9. Switching away from the symmetric schedule requires the stated consent of both spouses. For switching back, however, unilateral action suffices. The only exception are couples in which only one spouse earns labor income. Those couples are automatically assigned the men-/women-favoring withholding tax schedule.

the other spouse is assigned the unfavorable withholding tax class (“V”). The spouse in "III" is taxed as if she was the single earner, while the withholding tax in "V" 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 burden for the spouse in "III" as compared to being in withholding tax class "IV" while increasing the withholding tax burden of the spouse in "V". The second column in Figure 1.2.2 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 "III". 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, this misallocation of favorable and unfavorable withholding tax classes rarely happens.

Effect on Tax Rates. The shift of withholding tax burden from the primary to the secondary earner cannot only reduce the joint withholding tax burden but also has large effects on the withholding taxes paid by each spouse. The left-hand side of Figure 1.2.3 displays the average withholding tax rate by 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 compared to the default withholding tax class. 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 burden 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. Similarly, the marginal withholding tax rate is also affected by the different withholding tax classes. We depict the marginal withholding tax rate by withholding tax class in Figure 1.D.2.

Choice of the Different Schedules. The right-hand side of Figure 1.2.3 shows the frequency with which the different withholding tax schedules are chosen and which withholding tax class they allocate to each spouse. Approximately 50% of the couples pick the men-favoring schedule that shifts the withholding tax burden 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.

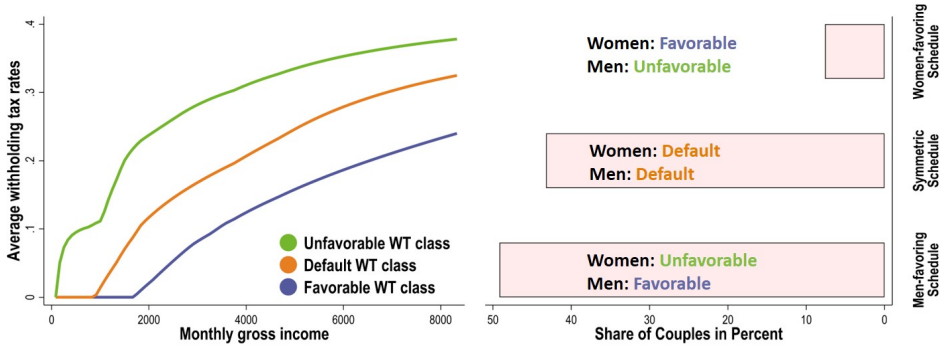


Figure 1.2.3. Illustration of different withholding tax schedules

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 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, for the year 2010 the possible withholding tax schedules and their frequency are shown 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.

While the different choices of withholding tax schedules that we have discussed here have strong effects on the amounts of withholding tax payments, they do not affect the final income tax burden of the couple. Couples cannot decrease their final income tax burden by choosing a certain withholding tax schedule, but can only change the timing of the income tax payments throughout the year.[10]

1.2.3 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 enabled taxpayers to deduct a much larger share of their contributions to health care insurance. As everyone in Germany is forced to hold health insurance, it decreased the income tax burden for all taxpayers. Conditional on income, the reform of the income tax was identical for everyone independent of the withholding tax schedule. Furthermore, as the contributions to health care insurance are automatically taken into account in the calculation of the withholding tax, the reform was equivalent to a

10. Of course, taking into account discount rates and liquidity constraints, couples can have benefits from delaying their income tax payments.

cut in withholding taxes for all taxpayers. Crucial for the identification of causal effects in our setting is that the reform, in addition, introduced that social security contributions are now taken into account for the calculation of withholding taxes for taxpayers in the unfavorable withholding tax class. Previously, they were only considered for taxpayers in the other withholding tax classes. This has the effect that the reform reduced the withholding tax - but only the withholding tax - for taxpayers in the unfavorable withholding tax class substantially more than for taxpayers in the other withholding tax classes.

Reform Effect. Figure 1.2.4 shows how annual withholding taxes changed from 2009 to 2010 by withholding tax class and annual gross labor income. For spouses in the favorable withholding tax class, the reform decreased the withholding tax burden by up to around 800 €. However, there was almost no change, and if then a slight increase, in withholding taxes for annual gross labor incomes lower than 32,000 €. For the default withholding tax class, the reform decreased the withholding tax burden by up to around 1,200 € with a substantially smaller cut for lower incomes. In contrast, women in the unfavorable withholding tax class profited from a cut by up to approximately 3,000 € with even a considerable reduction in withholding taxes for low incomes. In other years, such substantial year-to-year changes have not occurred. Figure 1.D.3 shows this for the years between 2006 and 2016 and for an annual individual income of 25,000 €, an income which is fairly common in the unfavorable tax class.[11] The described reform is the only substantial reform in withholding taxes during our sample period.

Anticipation and Salience. The reform, which was passed into law half a year before its onset, was arguably non-salient in the sense that it was unknown to taxpayers that there was a reform that changed withholding taxes depending on the withholding tax class a given taxpayer is in. There was no public debate about this part of the reform, just about the reform decreasing eventual income taxes, and there is no indication that people were made aware of the connection of the reform to withholding tax classes. This assessment is corroborated by looking at Google Trends for relevant terms. No striking movements are visible around the dates of the reform announcement and introduction. This means that couples are then not expected to have either changed their withholding tax schedules around the reform date in response or adjusted their labor supply already prior to the reform. Furthermore, the reform's non-salience means 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

11. The pattern is essentially the same for annual individual incomes of 15,000 €, 30,000 €, 50,000 €, and 70,000 €.

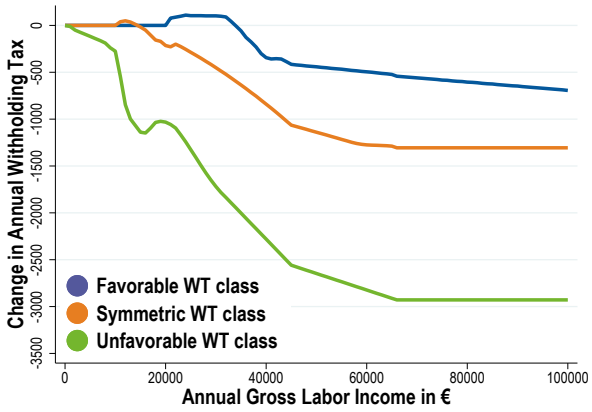


Figure 1.2.4. Effects of the 2010 Reform on Withholding Taxes by Withholding Tax Class

Notes: The figure plots the effect of the withholding tax reform 2010 on withholding tax payments depending on the withholding tax class. It illustrates the absolute change in the annual withholding tax payments caused by the reform.

taxes that was indeed salient is that they ended up getting more money after withholding taxes every month, i.e., a higher net income on their payslips.[12]

1.2.4 Survey: Exploring the Understanding of Withholding Taxes

In this study, we argue that the lack of understanding of withholding taxes can affect labor supply decisions. To underpin our argument, we conducted a pre-registered online survey to be filled out by married couples living in Germany.[13] In this survey, we ask the participants directly about their understanding of withholding taxes in Germany and try to identify channels through which a misunderstanding of withholding taxes can affect labor supply. Furthermore, we document which characteristics of couples are associated with a higher misunderstanding of the withholding tax system. By this, we want to gauge which couples could be particularly affected by distorted labor supply incentives to an extent that our administrative tax records do not allow us to. Moreover, we tentatively investigate the impact of the choice of withholding tax schedules on married couples' eventual intra-household distribution of labor income.

In this section, we focus on the core results of our final analysis sample consisting of 506 respondents (258 men, 248 women). We provide more details on

12. 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 jointly to the married couples.

13. We have pre-registered our survey at the Open Science Foundation.

our survey in Appendix 1.C. Section 1.C.1 includes information on the implementation and our sample restrictions, Section 1.C.2 displays our descriptive figures, Section 1.C.3 provides more-in-depth analysis, and Section 1.C.4 displays the original survey questionnaire in German and a translation into English.

Understanding of Withholding Taxes. The most important information we want to elicit with the help of our survey is whether married individuals understand the withholding tax system. We focus on two essential aspects by investigating (1) whether they know that withholding taxes do not affect a married couple's joint final income tax burden and (2) whether they understand that withholding taxes, however, affect their monthly payslip.

First, 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 burden. We do so by creating a realistic example of 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 burden of the couple.[14] We ask this question once at the beginning of the survey and again towards the end after the respondents have received extensive information about the withholding tax system. As discussed in Section 1.2.2, irrespective of the choice of the withholding tax schedule, the final income tax burden of the couple is the same. We find that only around 16 % of the surveyed individuals know about the irrelevance of the withholding tax schedule for the final income tax burden at the beginning of our survey. In Figure 1.C.1, we document that there exists heterogeneity in this knowledge across subgroups. Men (20 %) are better informed than women (13 %), while the knowledge is largely independent of the respondents' withholding tax classes, even though men are over-represented in the favorable withholding tax class as compared to the unfavorable one.

Second, we document whether the individuals know that and how they can influence the amount of withholding taxes they have to pay every month with the help of withholding tax schedules - so whether they know that and how they can impact the size of monthly wage transfers from their employers while keeping their gross labor income constant.[15] We document that among all respondents, we classify 61 % (63 % of men, 60 % of women) as knowledgeable. In Figure 1.C.2, we illustrate that this knowledge about the interlinkage between withholding tax classes and the monthly payslip is homogeneous with respect to gender, withholding tax classes, age except for the youngest cohort, and tax filing activity.[16]

14. See Question D7 in Appendix 1.C.4 for the exact wording of the question.

15. See Question D10 in Appendix 1.C.4 for the exact wording of the question.

16. Panel A in Table 1.2.1, however, indicates that this knowledge might be lower for respondents in the women-favoring withholding tax schedule. The same pattern can be observed for the knowledge of the irrelevance of withholding tax schedules for the final income tax burden. This

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 burden. 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). This then distorts the relative intra-household distribution of labor income as paid out by the employers.

Filing of Taxes. One way to gauge which couples are particularly affected by this distortion 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 jointly as a married couple as it is the case for our analysis sample in the administrative data.[17]

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.

As documented in Figure 1.C.1, women less often than men know that withholding taxes do not have an influence on the final income tax burden. 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 burden at the beginning of the survey. For women, knowledge increases from

might reflect that couples in the women-favoring schedule are different from others, e.g., in the importance they give to tax optimization.

17. 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.C.4 for the exact wording of the question.

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 the husband predominantly cares about taxes are also more affected by the incentive distortions arising from the shifting of the withholding tax burden 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 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 choosing the women-favoring schedule when the wife outearns the husband.[18] 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 households.[19] 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.C.3, 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.

Organization of Household Finances. The basic assumption underlying withholding tax classes is that households consisting of married couples are organized jointly and act as an economic unit. If this assumption does not hold, the choice of withholding tax schedule might have impacts on the eventual intra-household distribution of labor income and via this on the size of each spouse's budget and on their within-household bargaining power. To gain insights into such impacts we

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

19. See Question D18 in Appendix 1.C.4 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).

have to gain knowledge about potential money transfers between spouses. Sophisticated couples could make transfers from the spouse in the favorable withholding tax class to the spouse in the unfavorable withholding tax class and thereby re-establish the "default" relative intra-household earnings.

We thus asked in detail whether and how couples use shared bank accounts, onto which bank account they let their employers transfer their wage payments, whether and how they transfer (parts of) these wage payments to another bank account, and to which bank account potential tax refunds are transferred.[20] We present our main findings in the following and refer to Appendix 1.C.3 for a more in-depth derivation of these findings.

We test the basic assumption of the joint organization of household finances tentatively by interpreting the absence of a shared bank account as an indication of a lack of joint organization of household finances. 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 shifting some part of the withholding burden from one partner to the other 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. In addition, even if married couples have a shared bank account they might not use it to re-distribute labor income from one spouse to the other.

As shown in Panel B of Table 1.2.1, only 53 % of the respondents in the men-favoring withholding tax schedule state to have a shared bank account. 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 they are not commonly used to counteract the distortion of the relative intra-household distribution of labor income arising from the choice of that schedule. In particular, we consider the remaining 47 % of the couples in the men-favoring schedule unlikely to account for the distortion. We document in Panel B of Table 1.2.1 that the distortion is even aggrevated by the way couples deal with tax refunds. 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.

Turning the attention towards wage transfers between the spouses instead of tax refunds, we tentatively calculate that even among couples in the men-favoring schedule with a shared bank account 21 % do not seem to account for the distortion effects of being in the men-favoring schedule by wage transfers between the spouses so that we can monitor a counteracting strategy for only 42 % of all

20. See Questions D16a to D16f and Question D17c in Appendix 1.C.4 for the exact wording of the questions.

couples in the men-favoring schedule.[21] We thus expect relative intra-household earnings to be distorted in favor of the husband for the majority of the couples in the men-favoring schedule. Furthermore, married women's disposible net income, given constant income taxes, is lowered. This might lead them to overestimate their individual income tax burden, which can detrimentally affect their bargaining power within the household and decrease their perceived work incentives.

Table 1.2.1. Survey descriptives by withholding tax schedule

	Withholding tax schedule				
	Men-favoring	Symmetric	Women-favoring	Don't know	Overall
Panel A: Knowledge					
Irrelevance for income tax	16 %	18 %	8 %	18 %	16 %
	(0.03)	(0.03)	(0.04)	(0.05)	(0.02)
Impact on withholding taxes	66 %	63 %	48 %	54 %	61 %
	(0.04)	(0.03)	(0.07)	(0.06)	(0.02)
Panel B: Bank Accounts					
Shared bank account	53 %	58 %	58 %	51 %	55 %
	(0.04)	(0.03)	(0.07)	(0.06)	(0.02)
Tax refunds to husband's account	42 %	32 %	24 %	37 %	35 %
	(0.04)	(0.03)	(0.06)	(0.06)	(0.02)
N	177	214	50	65	506

Notes: The table summarizes our survey findings with regard to knowledge about withholding taxes and the organization of household finances. Standard errors are displayed in parentheses.

1.3 Data and Sample

Our study is based on a 5 % sample of extensive administrative tax records from the German Taxpayer Panel. In the first subsection, we describe this data source. In the second subsection, we describe how we construct our estimation sample and summarize basic socio-demographic characteristics of our sample.

1.3.1 German Taxpayer Panel

The German Taxpayer Panel (TPP) is an administrative dataset that contains information on the population of taxpayers in Germany for the years 2001 to 2018.[22] It

21. See Appendix 1.C.3 for a derivation of these findings.

22. RDC of the Federal Statistical Office and Statistical Offices of the Federal States (2018, DOI:10.21242/73111.2018.00.01.2.1.1)

includes information on various characteristics such as income, gender, age, number and age of children, withholding tax class and other tax-related information. The TPP consists of a total of around 63 million records for individuals for whom tax information is available for at least two years. Due to its large size, the data is primarily offered as a sample through research data centers. The waves of the TPP for the years 2001 to 2011 were created 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 pay withholding taxes. However, due to the late availability, we do not consider those taxpayers in our analysis.

1.3.2 Sample Selection and Characteristics

In our analysis, we use the administrative tax records from the years 2006 to 2016 and focus on dual-earner married couples in the two most common withholding tax schedules: the men-favoring and the symmetric schedule.[23] We do so for two reasons. First, as shown in Section 1.2.2, the vast majority of 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. Hence, these couples are very different from the couples in the other two schedules. For the men-favoring and symmetric schedules, we keep couples in which both spouses received labor income in 2009, the year before the aforementioned withholding tax reform was implemented.[24] This restriction ensures that these individuals are

23. 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.

24. We also exclude individuals earning no more than 5,400 € per year. This condition ensures that we exclude individuals in marginal employment, who can earn at most 450 € per month and are exempt from the income tax.

actually treated at the time of the reform. Moreover, we focus on couples in which both spouses are between 20 and 60 years old.[25] To ensure that labor income is the main source of income, we exclude couples in which, in the year 2009, at least one spouse received income of more than 1,000 € from self-employment.

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.

Balanced vs. Unbalanced Panel. In our main estimation sample, we only consider couples who file their income taxes for every year of the sample period and who do not violate any of the sample restrictions during the sample period. This leaves us with 11,039 couples and we refer to this sample as the balanced sample. As robustness, we also perform our analysis using an unbalanced sample in which we allow the couples to not necessarily show up in each year of the sample period. This happens if a couple does not file their income taxes in a given year or if they violate one of our sample restrictions in a given year. The unbalanced estimation sample consists of 23,233 couples. We prefer the balanced sample for two reasons. First, as mentioned before, the administrative tax records include individuals that did not file their income taxes only from 2013 onward. Since couples in the symmetric withholding tax schedule do not have the obligation to file their taxes, their labor market income could be missing in a given year in the unbalanced sample.

25. We want to abstract from early retirement decisions and thus do not consider income at older ages.

In contrast, the balanced sample makes sure that we can observe everyone's labor market outcomes for every year. Second, it allows us to abstract from extensive margin effects that could arise from married women leaving employment.[26] Our main interest is in finding out whether and to which extent married women increase their labor supply in response to a decrease in their withholding tax rate. In the unbalanced sample, other channels such as childbirth could play a role.

Descriptive Statistics. Table 1.3.1 displays descriptive statistics of basic sociodemographic characteristics for the balanced sample in the year 2009. The descriptive statistics for the unbalanced sample are similar and can be found in Table 1.D.1 in the Appendix. The results show that couples picking the men-favoring schedule have higher male income and lower female income than couples picking the symmetric schedule. Accordingly, for couples in the symmetric schedule, women earn 46 % of household income, while they earn only 29 % in households who picked the men-favoring schedule. This is not surprising as for couples with a man as the main earner, picking the men-favoring choice minimizes the withholding tax burden for the household. The table reveals that households in the two schedules are also different with respect to other observables. Specifically, couples in the men-favoring schedule are more likely to be Catholic and less likely to live in Eastern Germany.[27]

26. As we have to condition on having labor income in the pre-reform year, we cannot investigate whether they move from having no labor income to being employed.

27. Eastern Germany comprises the area of the former German Democratic Republic plus West Berlin.

Table 1.3.1. Descriptive Statistics for the Year 2009

	Men-Favoring	Symmetric
Income Wife	19651.74 (8470.72)	33321.58 (13402.3)
Income Husband	49737.3 (17046.99)	39453.28 (15233.01)
Female Income Share	0.29 (0.09)	0.46 (0.11)
Age Wife	44.63 (4.47)	44.69 (4.97)
Age Husband	46.57 (4.43)	46.39 (4.8)
Eastern Germany	0.08 (0.27)	0.36 (0.48)
Has a Child	0.67 (0.47)	0.31 (0.46)
Number of Children	1.42 (0.88)	0.76 (0.86)
Catholic Wife	0.4 (0.49)	0.23 (0.42)
Catholic Husband	0.37 (0.48)	0.2 (0.4)
Public Servant Wife	0.12 (0.32)	0.12 (0.33)
Public Servant Husband	0.2 (0.4)	0.15 (0.36)
N	5772	5267

Notes: The table displays descriptive statistics for the year 2009 for the balanced panel for couples who picked either the men-favoring or symmetric withholding tax schedule. They are calculated based on the sample restrictions outlined in Section 1.3.2. Specifically, we focus on households with dual earners in 2009, in which both partners have received no unemployment benefits and short-time work compensations in 2009, are between 20 and 60 years old in 2009, have no income from self-employment of more than 1,000 € in 2009 and whose incomes were stable between 2006 and 2009, i.e., the income for both household members fluctuated by less than 25 % from one year to the other.

All in all, the descriptives strongly suggest that the two groups are different in observable socio-demographic characteristics. However, using a Difference-in-Differences approach we do not rely on the two groups having the same observable characteristics. We discuss which assumptions we need for our identification and potential threats arising from the different sample compositions extensively in the next section.

Determinants of Schedule Choice. To further clarify which characteristics of a couple are correlated with the choice of the men-favoring schedule compared to the choice of the symmetric schedule, we regress the choice of the withholding tax schedule on various characteristics of the couple. The results in Table 1.3.2 show that a few characteristics stand out. First, living in the former East of Germany is associated with a 20 percentage points lower probability of choosing the men-favoring schedule. Since we also control for the female income share, this cannot be driven by the fact that the earning differences within couples are lower in the East due to the historically higher labor market participation of women. We suspect that more egalitarian gender norms (Campa and Serafinelli, 2019; Boelmann, Raute, and Schönberg, 2021) and lower historical institutional exposure in the East due to the take-over of West German institutions as late as 1990 lead couples to choose the men-favoring schedule less often. Second, we see that the higher the female income share, the less likely the couple chooses the men-favoring schedule. A one percentage point increase in the female income share is associated with a 1.8 percentage point decrease in the choice of the men-favoring schedule. This is intuitive since the more the man earns relative to the woman in a couple, the higher the gains in terms of withholding tax payments from choosing the men-favoring schedule. Finally, having children also significantly increases the likelihood of choosing the men-favoring schedule. The first child increases the likelihood by around 15 percentage points and every further child by another 6 percentage points. This shows that in many couples the man is likely considered the main breadwinner as soon as the couple is having children, mirroring the stylized fact that the birth of the first child is a fundamental event in explaining the persistence gender inequality in earnings (Kleven, Landais, Posch, Steinhauer, and Zweimüller, 2019).

Table 1.3.2. 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)
Income Husband (1000 Euro)	−0.0
	(0.00)
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	11039.0
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, using the balanced sample. Heteroscedasticity-robust standard errors are displayed in brackets. The regression also includes commuting days, commuting distance and a public servant dummy as regressors. As they have no explanatory power and for better readability, we do not display these regressors in this table. The full regression results including all regressors can be found in Table 1.D.2.

Taken together, this evidence illustrates that we should additionally control for some of these characteristics in our analysis. In the next section and in Appendix 1.A, we discuss how we do that by controlling for the pre-reform incomes of both spouses and for dummies indicating the parental status and the residence in East Germany using a cell fixed effects approach.

1.4 Empirical Strategy

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, since, as shown before in Table 1.3.2, the choice of withholding tax schedules is highly endogenous, 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 1.2.3. The reform disproportionally reduced the withholding tax burden of individuals in the unfavorable withholding tax class compared to individuals in the other two withholding tax classes. As argued in Section 1.3.2, 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. A naive approach would simply compare the evolution of incomes over time between these two groups using a difference-in-differences design. However, as previously shown in Figure 1.2.4, individuals' exposure to the reform is not only determined by their withholding tax class but also by their own pre-reform labor income. The latter is problematic since it implies that depending on the own pre-reform labor income there are large differences in the absolute and relative changes in withholding tax payments induced by the reform.

Treatment Intensity. 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 perform our analysis using a continuous treatment variable. The continuous treatment variable measures the percent change in the marginal net-of-withholding-tax rate of the woman induced by the reform and can therefore

be understood as a measure of treatment intensity.[28] 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.[29]

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

$$\begin{aligned}\text{Log Income}_{i,t} =& \beta \text{ Treatment Intensity}_{w,2010} \times \mathbb{1}(\text{Post Reform}_t) \\ &+ \alpha_{c,2009} \times \theta_t + \gamma X_{c,t} + \eta_i + \epsilon_{i,t},\end{aligned} \tag{1.4.1}$$

where β measures the percent change in labor income if the marginal net-of-withholding-tax rate of the woman increases by one percent. η_i controls for time-invariant individual fixed effects. Further, $X_{c,t}$ controls for time-varying characteristics of the couple c. These include the number of children, region of residence, and, for both spouses, age, age squared and a dummy for being a public sector worker. Finally, we add couple-level cell fixed effects $\alpha_{c,2009}$ interacted with year dummies θ_t. The cell fixed effects control for the strongest predictors of the withholding tax schedule, as shown in Figure 1.3.2, namely for binned own and spousal pre-reform labor income interacted with dummies for parenthood and residence in East Germany.[30] By interacting the cells with year dummies we allow for different time trends across cells. While controlling for own pre-reform income is common in the literature, additionally also controlling for partner pre-reform

28. This measure is standard in the literature on income tax elasticities. Following Saez, Slemrod, and Giertz (2012), regressing log income on this measure of treatment intensity allows us to obtain the elasticity of labor income with respect to the withholding tax.

29. The exact formula used is: $\text{Treatment Intensity}_{w,2010} = \frac{\text{MNWR}^{2010}_{w,2009} - \text{MNWR}^{2009}_{w,2009}}{\text{MNWR}^{2009}_{w,2009}}$, where $\text{MNWR}^{2010}_{w,2009}$ is the marginal net-of-withholding-tax rate woman w faces in 2009 applying the tax schedule of 2010, while $\text{MNWR}^{2009}_{w,2009}$ is the marginal net-of-withholding-tax rate woman w faces in 2009 applying the tax schedule of 2009. The subscript w denotes that we calculate the treatment intensity using the income and tax rates of the woman in each married couple.

30. We do so by dividing own and partner income into bins of 10,000 €, ranging from 0 to 100,000 €. We then interact the own and partner income bins with dummies for parenthood and residence in East Germany, leaving us with 400 couple-level cells.

income is not. In our setting, however, this is useful and necessary and we explain the underlying reason in detail in Appendix 1.A.[31]

Identifying Assumptions. The validity of our identification strategy relies on two main assumptions. First, it has to hold that there is no selection of couples into treatment. As discussed before, the reform was arguably nonsalient and therefore not anticipated by the average taxpayer.[32] However, it could be that individuals changed their withholding tax schedule as a result of the reform. This would alter the treatment intensity they are subject to and thereby bias our results. We depict the share of couples in the three different withholding tax schedules and the transitions between the different withholding tax schedules over time in Figure 1.D.1. Looking at all couples in the 5 % sample of the TPP shows that couples generally stick to the withholding tax schedule they have chosen and that there are only a few couples changing between the withholding tax schedules over time.[33] Also, there is no evidence for an increase in withholding tax schedule changes around the time of the reform. This makes us confident that there was no selection into treatment in our setting.

Second, we have to assume that the parallel trend assumption holds. It assumes that the labor market outcomes of treated and untreated individuals would have evolved the same in absence of the reform, irrespective of the treatment intensity. This implies that all observed post-reform differences in outcomes are due to differences in the treatment intensity induced by the reform.

One implication of this assumption is that we should see no economically significant effect of the treatment intensity on labor supply in the years before the withholding tax reform. We check this by also estimating a dynamic version of

31. Typically, the literature measures the elasticity of taxable income with an IV approach (see Saez, Slemrod, and Giertz, 2012). This is not needed in our setting, as the dense income cell fixed effects ensure that almost all variation in treatment intensity stems from the variation in withholding tax classes.

32. There was no public debate about the implications of the reform on withholding taxes and a search in Google Trends for relevant key words shows no signs of public discussion.

33. Typically, couples pick their withholding tax schedule at their marriage and do not adapt the withholding tax schedule thereafter.

Equation 1.4.1 in which we replace the post reform dummy with year dummies.[34] Economically insignificant estimates for the pre-reform years can make us confident that individuals with differing treatment intensity had no different trends in labor market outcomes before the reform. In Section 1.5, we will show that we indeed cannot find any economically significant estimates for the pre-reform period.

1.5 Empirical Results

In this section, we present and discuss our empirical results. We begin by presenting the static and event study Diff-in-Diff estimates. Subsequently, we explore potential heterogeneities of the treatment effect. Lastly, we investigate whether the observed treatment effects are primarily driven by changes in the marginal tax rate or average tax rate.

Static Diff-in-Diff. First, we present the results of the static diff-in-diff estimation as laid out in Equation 1.4.1. Table 1.5.1 displays regression results by gender with and without additional cell fixed effects. All regressions include individual fixed effects and control for potentially time-varying characteristics of the couple.

For women, we find significant positive estimates for the elasticity of labor income with respect to the withholding tax rate. The first column shows that without cell fixed effects the estimated elasticity for women is around 0.112. When including cell fixed effects in our preferred specification in the second column, this estimate is reduced to around 0.099. Both estimates are significant at the 1 % level. These results imply that a one percent higher marginal net-of-withholding tax rate results in an around 0.1 percent higher labor income 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

34. $\text{Log Income}_{i,t} = \sum_{t=2006}^{2016} \beta_t \left[\text{Treatment Intensity}_{w,2010} * \mathbb{1}(\text{Year}_t)\right] + \alpha_{c,2009} \times \theta_t + \gamma X_{c,t} + \eta_i + \theta_t + \epsilon_{i,t}$

tax rate increased by 10 percent from 70 % to 77 %, increased her labor income by 1 percent after the reform.

Table 1.5.1. Static Diff-in-Diff Results

	Women		Men	
	(1)	(2)	(3)	(4)
DiD Estimate	0.112***	0.099***	0.011	0.007
	(0.020)	(0.020)	(0.011)	(0.011)
Cell FE		✓		✓
N	121,429	121,429	121,429	121,429
Adj. R-Squared	0.334	0.374	0.301	0.317

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1. All regressions include individual fixed effects and control 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 East Germany, and years. Results using the unbalanced estimation sample can be found in Table 1.E.1. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

As shown in columns (3) and (4), we do not find any significant effects for men, neither with nor without cell fixed effects. The lack of a reaction of men does not come as a surprise as they are not directly affected by the withholding tax cut of their spouse. The treatment intensity, which measures the change in the marginal net-of-withholding tax rate of the female spouse, could therefore only indirectly affect the labor supply of the male spouse through possible spillover effects within the couple. We find no evidence for this.

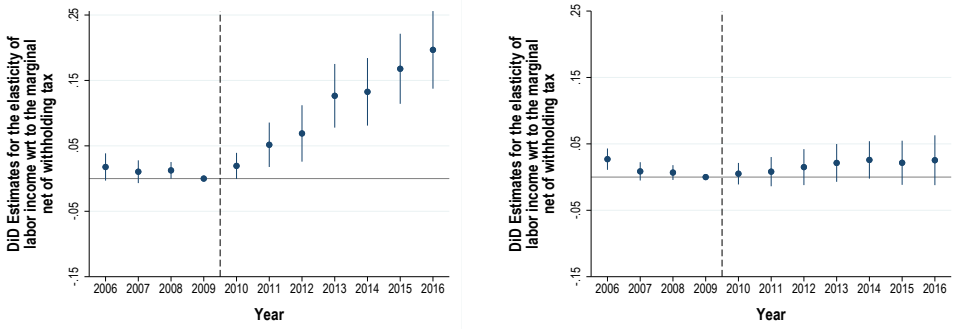
We explore the robustness of our results to using the unbalanced estimation sample in Table 1.E.1. Columns (1) and (2) show that with the unbalanced sample, the estimated elasticity for women is still statistically significant, but with around 0.05 substantially smaller than in the balanced sample. This could be caused by sample attrition. As the treatment effect estimate increases over time (see Figure 1.5.1), the static diff-in-diff estimate becomes smaller for the unbalanced sample relative to the balanced sample if observations leave the sample over time. For men, we see in column (4) that there are again no significant effects when controlling for cell fixed effects.

Event Study Diff-in-Diff. We explore the dynamics of the treatment effect in Figure 1.5.1, which graphically displays the estimates for the dynamic version of Equation 1.4.1 controlling for cell fixed effects.

For women, we see in Figure 1.5.1a that the treatment effect is increasing over time. Variation of the size of the treatment effect over time can be expected due to our assumptions about how taxpayers learn about their taxes. As we argue that employees use the information on withholding taxes from their monthly payslip to learn about their income taxes, we expect a lagged response to the reform as this learning process takes time. This lagged response is in line with Shapiro and Slemrod (1995), who 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. Further, even when taxpayers recognize a change in their withholding taxes on their monthly payslip, it might take them a few months to be sure that that change is indeed permanent. We would not expect them to react on apparently transient shocks.

Additionally, there could be a lagged response as adapting one's working hours usually takes time. Women who want to increase their working hours have to negotiate this with their employer or find a new employer that offers higher working hours. We also think that a substantial part of the treatment effect realizes 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.

The event study Diff-in-Diff results for men are displayed in Figure 1.5.1b. We find, in line with the static Diff-in-Diff results, no significant treatment effects in the post-reform period.



(a) Effect on Female Log Income **(b)** Effect on Male Log Income

Figure 1.5.1. Event Study Diff-in-Diff Estimates

Notes: The figure plots the estimates for the elasticity of labor income with respect to the withholding tax estimated based on the dynamic version of Equation 1.4.1 for women and men using the balanced sample. The dependent variable is the log income of the individual, and the independent variable is the treatment intensity. Treatment intensity is defined as the percent change in the marginal net-of-withholding-tax rate of the woman induced by the reform of the withholding tax in 2010. All regressions include individual fixed effects and control 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 East Germany, and years. Confidence intervals are plotted at the 95 % level and based on heteroscedasticity-robust standard errors. Note that the sample excludes households, where at least one member experienced a drop in income by more than 25 % from one year to the next before 2010 to ensure that no individuals directly hit by the financial crises are part of the sample. This explains the smaller standard errors before the reform. The underlying regression coefficients can be found in in columns (2) and (4) of Table 1.E.2.

As discussed in Section 1.4, one implication of the parallel trend assumption is that we should see no economically significant pre-reform effects of the treatment intensity. In fact, the pre-reform estimates for women in Figure 1.5.1a are not statistically significantly different from zero at the 5 % level and also economically insignificant and small compared to the post-reform estimates. For men, we see a similar pattern in Figure 1.5.1b. With the exception of 2006, there are no statistically significant effects before the reform. This finding of economically non-significant pre-reform effects therefore gives us additional confidence in the validity of the common trends assumption.

We, again, explore the robustness of our results to using the unbalanced estimation sample in Figure 1.E.1. As before, we see a similar pattern of lagged treatment effects. For women, the treatment effects become significant after 2012 and we even find small significant effects for men after 2013. Regarding pre-reform

treatment effects, we see for women slightly larger effects than in the balanced sample, but which are still economically insignificant compared to the post-reform period. For men, there are no statistically significant pre-reform treatment effects.

Heterogeneity of Treatment Effect. In the following, we investigate the heterogeneity of the treatment effect by observable characteristics. We start by looking at the effects for women in column (1) of Figure 1.5.2. In Panel A, we look at heterogeneities across regions. We find significant effects only for couples in former West Germany and no effects for couples in former East Germany. The reason for that could be that women in East Germany generally work more and therefore have less potential to increase their working hours. Another explanation could be that the more progressive norms towards female labor market participation in East Germany dominate incentives stemming from the withholding tax system. Moving to Panel B, we explore potential differences based on the level of pre-reform commuting days. We create a dummy variable indicating high commuting if the female spouse commuted more than 200 days before the reform. We use commuting days as a proxy for working hours to gauge the woman's potential to increase her working hours.[35] We, however, find no significant differences in the treatment effect based on pre-reform commuting days. In Panel C, we examine differences between parents and non-parents. We define couples as parents if they have at least one child below the age of 18 living in the household. We find significant effects for both parents and non-parents, with slightly larger effects observable for parents. Focusing solely on parents and distinguishing by the age of the youngest child in Panel D, we discover that the treatment effect is highly significant for parents with the youngest child older than 6 years. Conversely, the effect is non-significant for parents with the youngest child below 6 years old, although this result is primarily driven by the presence of large standard errors.

35. While there is no information on working hours in the administrative tax records, commuting days are recorded for individuals that deduct their commuting costs from the income tax.

The limited number of parents with the youngest child below 6 years old in our sample contributes to these substantial standard errors.[36]

Table 1.5.2. Heterogeneity Analysis: Static Diff-in-Diff Results

	Women (1)	Men (2)
Panel A: East vs. West Germany		
West	0.110***	0.007
	(0.021)	(0.012)
East	0.002	0.011
	(0.050)	(0.035)
Panel B: Level of Pre-Reform Commuting		
Low Commuting	0.093***	0.006
	(0.024)	(0.015)
High Commuting	0.099***	0.000
	(0.034)	(0.016)
Panel C: Parent vs. Non-Parent		
Non-Parent	0.084***	0.019
	(0.030)	(0.020)
Parent	0.104***	-0.005
	(0.026)	(0.013)
Panel D: Age of Youngest Child		
Youngest Child below 6	0.066	0.019
	(0.242)	(0.069)
Youngest Child betw 6 and 18	0.105***	-0.006
	(0.025)	(0.014)
N	121,429	121,429
Adj. R-Squared	0.362	0.311

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1, allowing for treatment heterogeneity by observable characteristics and using the balanced sample. All regressions include individual fixed effects and control for potentially time-varying characteristics of the couple. Panel A includes cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with dummies for parenthood and years. Panel B includes cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with dummies for parenthood, residence in East Germany, and years. Panel C and D include cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with years. Results using the unbalanced estimation sample can be found in Table 1.E.4. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

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

For men, we find in column (2) of Table 1.5.2 that no significant effects are detectable across subgroups. Finally, looking at the results using the unbalanced sample in Table 1.E.4, we observe very similar patterns as for our main sample.

Marginal vs. Average Tax Rate. We conclude the empirical results by examining whether our estimated elasticity of labor income with respect to the marginal net-of-withholding tax rate can be partially attributed to a reaction to changes in the *average* net-of-withholding tax rate. Typically, existing literature estimating behavioral reactions of income w.r.t. to the income tax assumes that income effects are small. In this case, households would primarily consider the marginal tax rate when reoptimizing their behavior following an income tax reform. As a result, previous studies investigating behavioral reactions to income tax reform have focused primarily on the effects of the marginal tax rate (Saez, Slemrod, and Giertz, 2012). However, even if income effects were small, households might still react to variations in the average withholding tax rate. They might consciously employ "schmeduling" techniques, such as ironing, where they assume that their average tax rate provides information about their marginal tax rate (Rees-Jones and Taubinsky, 2020). [37]

We therefore repeat our analysis and compare our estimates from the static Diff-in-Diff with and without additionally controlling for the change in the average withholding tax rate.

37. Furthermore, the design of the average withholding tax is also highly relevant for the optimal design of withholding tax schedules, as discussed in Section 1.6.

Table 1.5.3. Static Diff-in-Diff Results controlling for Average Tax Rate

	Women		Men	
	(1)	(2)	(3)	(4)
Marginal WT Rate	0.099***	0.098***	0.007	0.009
	(0.020)	(0.020)	(0.011)	(0.011)
Average WT Rate		0.006		-0.005**
		(0.004)		(0.002)
Cell FE	✓	✓	✓	✓
N	121,429	120,379	121,429	120,379
Adj. R-Squared	0.374	0.375	0.317	0.316

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1 while additionally including the change in the average net-of-withholding tax rate as an independent variable. The estimation is performed using the balanced sample. All regressions include individual fixed effects and control 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 East Germany, and years. Results using the unbalanced estimation sample can be found in Table 1.E.5. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Columns (1) and (2) of Table 1.5.3 show that, when using cell fixed effects, the estimate for the elasticity with respect to the marginal net-of-withholding tax rate for women does not change significantly when additionally controlling for the average net-of-withholding tax rate.

For men, however, we see in columns (3) and (4) that the coefficient for the average net-of-withholding tax rate is indeed significant. The higher the change in the average net-of-withholding tax rate of the female spouse, i.e., the lower her average withholding tax rate, the lower the post-reform income of the man. This could be understood as suggestive evidence that the average, rather than the marginal, withholding tax rate of wives can affect the labor supply of their husbands. Intuitively, this makes sense because a husband's (perceived) labor incentive is not directly affected by the marginal (withholding) tax rate of his wife, whereas a change in his wife's average (withholding) tax rate impacts the (perceived) joint household budget. Through income effects, a (perceived) increase of this household budget due to an increase in the average net-of-withholding tax rate of the wife can lead to a decreased male labor supply.

Looking at the results when using the unbalanced sample in Table 1.E.5, we find that the average net-of-withholding tax rate explains part of our observed effects for women. This would imply that policymakers should also partially consider the average withholding tax rate when designing withholding tax systems.

We discuss the implications of our empirical results for the design of withholding tax systems in the next section.

1.6 Policy Implications

The relationship between labor supply and withholding taxes holds important implications for welfare. Consequently, understanding the trade-offs involved in designing optimal withholding tax systems becomes essential for policymakers. Subsection 1.6.1 shortly summarizes the general considerations faced by governments when determining the appropriate level of withholding taxes. Thereafter, in Subsection 1.6.2 we investigate the trade-offs faced by countries when designing withholding taxes for couples specifically. Couples are particularly interesting because in all countries with joint taxation, the income tax rate of almost all couples is different from the income tax rate of unmarried individuals. Accordingly, in most countries with joint taxation the withholding tax schedule for married individuals is different to the withholding tax schedule of unmarried individuals. Surprisingly, however, between countries the underlying concept of the schedules for married individuals differs substantially. These differences are informative about the diverse set of objectives policy makers try to achieve by setting withholding taxes and reflect differing political objectives of policy makers.

1.6.1 Optimal Level of Withholding

Economists have generally expressed a positive view of the impact of high withholding taxes. For example, Thaler (1994, p. 191) states:

> *...most taxpayers like refunds, so raising* ***withholding taxes improves the government's cash flow*** *and* ***makes taxpayers happy****, an unusual parlay. On top of that, however, there is evidence that people* ***save more from lump-sum payments***

> *(...), so increasing the withholding rate should also increase the saving rate.* ***A free lunch!***

This view of governments being able to influence consumption and savings decisions of their citizens at no cost, no matter which of the two they want to increase, has been supported by various research contributions.[38] Shapiro and Slemrod (1995) run a survey and document that almost half of their sample planned to increase consumption as a reaction to a federal US tax reform in 1992 that decreased withholding taxes without changing the eventual tax liabilities. Feldman (2010) confirms that finding by showing that this reform decreased contributions to retirement savings accounts, likely through the channel of an increase in consumption. The behavioral reaction is particularly surprising as the US withholding tax system allows households to adapt their withholding taxes.[39] Households could have changed their withholding tax rate at any time to better reflect their income tax rate which would have allowed them to increase consumption during the year.[40]

Interestingly, despite the ability to adjust withholding taxes, most US households continue to be overwithheld so that approximately 30 % of the withheld taxes are paid back as a tax refund (see Rees-Jones, 2018; Gelman et al., 2022). This observation suggests that households see a benefit in being subject to overwithholding. Consequently, there are numerous attempts in the literature to rationalize why households are subject to overwithholding even though they could avoid it. It has been shown that active overwithholding decisions could be a tool of households to deal with limited self-control (Thaler, 1994; Neumark, 1995) and income uncertainty (Highfill, Thorson, and Weber, 1998; Gelman et al., 2022). However, another possible explanation is insufficient awareness. Indeed, Jones

38. Moreover, research has indeed shown that taxpayers like getting tax refunds and thus change tax filing behavior discontinuously at the point of exact withholding (Engström et al., 2015; Rees-Jones, 2018).

39. Adapting the withholding tax can come at a cost, as underwithheld households have to pay an interest on the underwithheld amount.

40. Messacar (2018) also finds that in Canada withholding tax rates impact saving decisions.

(2012) shows that the lack of withholding tax adjustments by taxpayers to reduce the high level of overwithholding can largely be explained by inertia.

In contrast to the existing literature, we find that withholding taxes can come at a cost with regards to labor supply. However, our results do not imply that withholding taxes should mirror income taxes as closely as possible. Instead, governments face a trade-off. On the one hand, withholding taxes are associated with more savings, less consumption, liquidity for the government and higher tax compliance. On the other hand, we document that higher levels of withholding taxes are associated with a reduction in labor supply. This implies that governments should pick the level of optimal withholding carefully, and spend effort on the optimal design of withholding taxes.

1.6.2 Implementation of Withholding Taxes for Married Couples

In countries with joint taxation, married couples typically pay lower income taxes than single households with the same income structure. Hence, when countries levy the same withholding taxes for married and single individuals, couples are typically mechanically overwithheld because the marriage bonus induced by joint taxation is not taken into account. When governments decide to reduce overwithholding of married couples, they can do so through different adaptations of the withholding tax system. However, it is important to be aware of the following dilemma: Typically, proponents of joint taxation systems view households as one unit and therefore do not want to influence labor supply decisions within the household with any tax incentives. Consequently, governments are also impartial about how the joint taxation benefits are distributed within the household. However, this changes when governments try to account for joint taxation benefits in the withholding tax system. As the withholding tax is an individual tax, they have to decide what the individual marginal and average withholding tax rates are that each spouse faces and thereby have to make a decision on how the joint taxation benefit is divided among spouses.

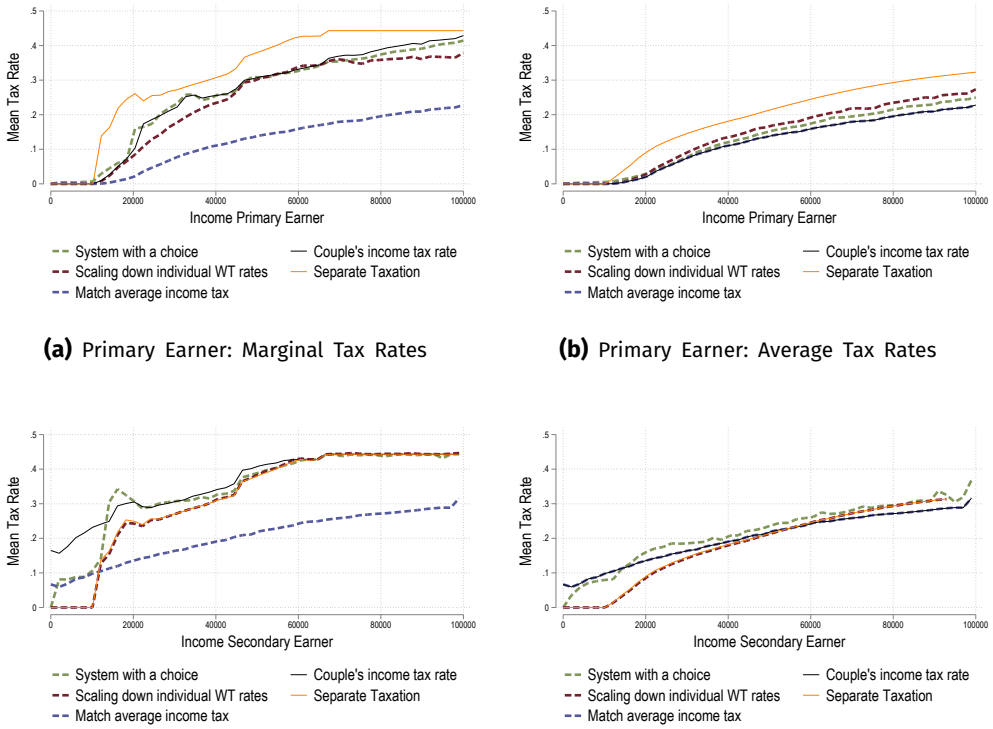
Interestingly, different countries have come up with different solutions for this problem. In the following, we evaluate some of the most common solutions by

looking at how they affect the marginal and average withholding tax rate of the primary and secondary earner. Doing so we also evaluate the degree to which the individual withholding tax rates differ from the couple's income tax rates as large differences could deceive individuals about their labor supply incentives. We will illustrate the effects of the different solutions empirically using a 10 % sample of the German tax records in 2010.

Joint taxation. Before discussing the effects of the different withholding tax systems, we first want to highlight again the effects of joint taxation on the income tax rates of both primary and secondary earners. In Figure 1.6.1, the orange line plots the individual income tax rates under separate taxation and the black line the couple's income tax rates under joint taxation. Notably, joint taxation results in higher marginal and average tax rates than under separate taxation for secondary earners, while the opposite is true for primary earners.

System with a choice. We start the discussion of the different withholding tax systems with the implications of the prevailing system in Germany, the *system with a choice* which we have already presented in detail in Section 1.2.2. It allows couples to reduce their overwithholding by choosing a different withholding tax schedule for each spouse. The system is calibrated such that there is minimal deception in marginal withholding tax rates if couples assign the secondary earner to the unfavorable withholding tax class when the gross income share of the secondary earner is around 30 % or lower. In the administrative tax records, we see that on average couples choose the withholding tax schedules such that the mean marginal withholding tax rate fits the couple's mean marginal income tax rate, i.e., there is minimal deception in marginal tax rates. We illustrate this in Figures 1.6.1a and 1.6.1c which show that the mean marginal withholding tax rate approximately fits the mean marginal income tax rate of the couple for both primary and secondary earners. Moreover, we see in Figures 1.6.1b and 1.6.1d that the *system with a choice* also leads to average withholding tax rates that approximately fit the couple's average income tax rate. This implies that both primary and secondary earners pay a share in withholding taxes roughly equivalent to their

gross-income share. However, it is important to note that we are solely examining the mean tax rates. In the appendix, Figure 1.B.1 illustrates that while the *system with the choice* fits well for the majority of the households many households experience significant differences between their withholding tax rates and the couple's income tax rates. This is the case as the available withholding tax schedules are not suitable for all income constellations. Moreover, not all households choose the withholding tax schedule that minimizes their withholding tax payments.



(a) Primary Earner: Marginal Tax Rates

(b) Primary Earner: Average Tax Rates

(c) Secondary Earner: Marginal Tax Rates

(d) Secondary Earner: Average Tax Rates

Figure 1.6.1. Average and Marginal Withholding Tax Rates in different WT regimes for married couples

Notes: The figures display the mean marginal and average 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 B and for secondary earners in Panel C and D. 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.

Alternative withholding tax systems. We contrast the *system with a choice* with another common solution to decrease overwithholding, namely the adjustment of withholding tax rates of both spouses by the expected joint taxation benefits based on past household incomes. In practice, two different implementations of this have emerged. Governments can either (1) scale down the individual withholding tax

rates of both spouses by a common factor such that the paid withholding taxes equal the expected income tax payments or they can (2) calculate the expected, effective average income tax rate paid by the couple and then set this rate as a common withholding tax rate.[41] In absence of income changes, both systems can completely eliminate the overwithholding caused by joint taxation savings. They, however, differ in how they affect the marginal and average withholding tax rates of both spouses.[42]

Scaling down withholding tax rates. For the withholding tax system that scales down the individual withholding tax rate, Figure 1.6.1a shows that for primary earners the marginal withholding tax rate is in general relatively close to the couple's marginal withholding tax.[43] Only in some income ranges the marginal withholding tax rate is lower than the couple's marginal income tax rate and thereby deceiving the primary earner about the marginal tax rate. Moreover, we see that the marginal withholding tax rate is also much lower than the marginal income tax rate under separate taxation. This is driven by the fact that many primary earners have a spouse that earns very little or no income which leads to large joint taxation benefits and a substantial down-scaling of the tax rate.

For secondary earners in Figure 1.6.1c we see that the marginal withholding tax rates are much lower than the couple's marginal withholding tax rate. As a consequence, secondary earners are largely deceived about the marginal tax rate. Interestingly, the marginal withholding tax rates are very close to the marginal income tax rates under separate taxation. This is the case as couples in which the

41. In 2010, approach (1) was introduced in Germany under the name *schedule with a factor* ("IV mit Faktor") as an additional option for married couples. The take-up rate is very low. In 2018, only around 40,000 couples (less than 0.5 % of all income-tax-paying couples) used this approach (*Kleine Anfrage Bundestag* 2019). In the German coalition agreement of 2021 (*German Coalition Agreement* 2021) the parties agreed that they plan to abolish the *system with a choice* and instead assign everybody to the *schedule with a factor*. Approach (2) was introduced in France in 2019.

42. To illustrate how withholding tax payments differ between the different systems, we calculate the withholding tax payments for a couple in which the man earns 50,000 € and the woman 20,000 € in the Appendix in Table 1.B.1.

43. We are looking only at the *short-term* marginal tax rate here. As income increases, the joint taxation benefit changes and as a consequence also the scaling factor in the following year. We illustrate the *long-term* marginal tax rates in Figure 1.B.2.

secondary earner earns income have, on average, lower joint taxation benefits and therefore no substantial down-scaling of the tax rate.

Figures 1.6.1b and 1.6.1d plot the effect on average withholding tax rates. The withholding tax system exhibits a notable shift in the withholding tax burden from secondary to primary earners, driven by the following phenomena: For secondary earners, the average income tax rate is higher than the average tax rate applicable in separate taxation due to their higher-earning partners. Consequently, when secondary earners pay a withholding tax based on separate taxation rates, their withholding tax decreases significantly. The factor further reinforces this effect, widening the gap even more compared to the average tax rate in joint taxation. This phenomenon flips only in the rare case of secondary earners earning more than around 70,000 €. On the other hand, primary earners benefit from joint taxation as it reduces their average income tax rate compared to separate taxation. However, we see that scaling down the tax rate from separate taxation still exceeds the average tax rate induced by joint taxation.[44] The consequence of these effects is that secondary earners end up paying average withholding taxes lower than the couple's average income tax, while primary earners pay average withholding taxes higher than the couple's average income tax. In other words, secondary earners contribute a lower share of withholding taxes relative to their gross income, whereas primary earners contribute a higher share than their gross income.

Matching average income taxes. For the withholding tax system which sets the marginal and average withholding tax rate to the couple's average income tax rate we see a different pattern. Figures 1.6.1a and 1.6.1c show that the marginal withholding tax rates implied by this approach are substantially lower than the marginal income tax rate of the couple for both primary and secondary earners.[45]

44. In Figure 1.6.1, the gap between "separate taxation" and "scaling down individual WT rates" is larger for primary earners than for secondary earners due to the inclusion of couples with only one wage-earner, who generally benefit the most from joint taxation.

45. We are looking only at the *short-term* marginal tax rate here. As the household income increases, the average income tax rate increases and as a consequence also the withholding tax rate in the following year. We illustrate the *long-term* marginal tax rates in Figure 1.B.2.

Both primary and secondary earners are therefore deceived in marginal tax rates. By design, the average withholding tax rate matches the average income tax rate of the couple. Both partners therefore end up paying a share of withholding taxes that matches their gross-income share.

Conclusion. Taken together, our analysis illustrates that the design of withholding tax schedules for married couples is nontrivial. We show that the *system with a choice* leads, on average, to minimal deception in marginal tax rates as the marginal withholding tax rates for both primary and secondary earners are approximately equal to the couple's marginal income tax rate. For average tax rates, we see a similar pattern as the average withholding tax rates also approximately fit the couple's average income tax rate. We then contrast this system with two other approaches. For approach (1) which scales down individual withholding tax rate by a common factor we find empirically that it leads to very low deception in marginal tax rates for primary earners, while there is a high deception for secondary earners which face marginal withholding tax rates much lower than the couple's marginal income tax rate. For average tax rates, we see that primary earners pay more withholding taxes than the couple's average income tax rate, while the secondary earners pay less than the couple's average income tax rate. In contrast, approach (2) which sets the withholding tax rates to the couple's average income tax rate leads to large deception in marginal tax rates for both primary and secondary earners. For average tax rates, we, however, see that by construction the average withholding tax rates coincide with the couple's average income tax rate.

The remaining question is now which approach should be preferred by policymakers. The answer to this depends on the preferences of the policymaker. If the policymaker does not want to deceive individuals and wants them to optimize along the couple's income tax rates, the *system with a choice* appears to be a good approach. However, given the results of our study, we expect this approach to generate negative labor supply incentives for secondary earners. This is the case because the *system with a choice* translates the high marginal income tax rates

of the secondary earner which stem from the joint taxation benefits into high marginal withholding tax rates. In contrast, policymakers interested in high labor supply incentives for secondary earners should consider deceiving them about the marginal tax rates by choosing one of the two approaches which adjusts the withholding tax rates by the joint taxation benefit.

1.7 Conclusion

In this study, we show that withholding taxes can affect labor income. For married women in Germany, we estimate an elasticity of labor income with respect to the marginal net-of-withholding tax rate of about 0.1 using a static Diff-in-Diff approach. Estimating an event study Diff-in-Diff, we find that the treatment effect increases monotonically over time.

Our estimate can be compared to estimates from the literature on the elasticity of taxable income (ETI). As Neisser (2021) shows in a meta-analysis, estimates for the ETI with respect to the income tax range from about 0.2 to about 0.8. This means that our findings are in line with our expectations. On the one hand, motivated by our survey findings, we expect some effect due to individuals' lack of understanding and inattentiveness to the tax system that might make them use their withholding tax burden as proxy for their income tax burden. On the other hand, individuals' reactions should be somewhat less strong than their reactions to income taxes because, in absence of liquidity constraints, fully-informed households should not react to withholding taxes.

The fact that individuals react to withholding taxes implies that governments should be careful when designing withholding tax schedules. Typically, taxpayers receive large paybacks when filing income tax returns as the withholding tax does not take into account special deductions. For example, in the US, nearly a third of the amount of all personal income tax payments is returned as tax refunds (Gelman et al., 2022). According to the Federal Statistical Office of Germany, about 88 % of all taxpayers filing their income taxes in Germany received tax refunds for the tax year of 2018 which amounted to 1,072 € per person on average.

Our results suggest that these large paybacks go hand in hand with taxpayers overestimating their actual income tax burden, as their withholding tax is much higher than the actual income tax. Hence, governments should redesign their withholding tax systems to better reflect the actual income taxes.

A common source of overwithholding are joint taxation benefits of married couples. We investigate how different withholding tax systems that account for joint taxation benefits affect the marginal and average withholding tax rates of primary and secondary earners. We show that as soon as countries try to reduce overwithholding they have to decide what the individual marginal and average withholding tax rates are that each spouse faces and thereby have to make a decision on how the joint taxation benefit is divided among spouses. We show that different implementations result in significantly different withholding tax rates for primary and secondary earners. Given our empirical results, countries can therefore affect the labor market participation of secondary earners by choosing between different withholding tax systems.

References

Abeler, Johannes, and Simon Jäger. 2015. "Complex Tax Incentives." *American Economic Journal: Economic Policy* 7 (3): 1–28. [10]

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 (12): 104537. [5]

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

Bertrand, Marianne, Emir Kamenica, and Jessica Pan. 2015. "Gender Identity and Relative Income within Households." *Quarterly Journal of Economics* 130 (2): 571–614. [21]

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. [11]

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. [11]

Boelmann, Barbara, Anna Christina Raute, and Uta Schönberg. 2021. "Wind of Change? Cultural Determinants of Maternal Labor Supply." *CESifo Working Paper*, (9094): [28]

Buettner, Thiess, Katharina Erbe, and Veronika Grimm. 2019. "Tax Planning of Married Couples and Intra-Household Income Inequality." *Journal of Public Economics* 179 (11): 104048. [9, 21]

Campa, Pamela, and Michel Serafinelli. 2019. "Politico-Economic Regimes and Attitudes: Female Workers under State Socialism." *Review of Economics and Statistics* 101 (2): 233–48. [28]

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. [10]

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review* 99 (4): 1145–77. [10]

Engström, Per, Katarina Nordblom, Henry Ohlsson, and Annika Persson. 2015. "Tax Compliance and Loss Aversion." *American Economic Journal: Economic Policy* 7 (4): 132–64. [5, 42]

Enrick, Norbert Lloyd. 1963. "A Pilot Study of Income Tax Consciousness." *National Tax Journal* 16 (2): 169–73. [10]

Enrick, Norbert Lloyd. 1964. "A Further Study of Income Tax Consciousness." *National Tax Journal* 17 (3): 319–21. [10]

Feldman, Naomi E. 2010. "Mental Accounting Effects of Income Tax Shifting." *Review of Economics and Statistics* 92 (1): 70–86. [42]

Feldman, Naomi E., and Bradley J. Ruffle. 2015. "The Impact of Including, Adding, and Subtracting a Tax on Demand." *American Economic Journal: Economic Policy* 7 (1): 95–118. [10]

Fujii, Edwin T., and Clifford B. Hawley. 1988. "On the Accuracy of Tax Perceptions." *Review of Economics and Statistics* 70 (2): 344–47. [10]

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. [5, 42, 50]

Gideon, Michael. 2017. "Do Individuals Perceive Income Tax Rates Correctly?" *Public Finance Review* 45 (1): 97–117. [10]

Gruber, Jon, and Emmanuel Saez. 2002. "The Elasticity of Taxable Income: Evidence and Implications." *Journal of Public Economics*, 32. [10]

Hauck, Tobias, and Luisa Wallossek. 2023. "Optional (Non-)Filing and Effective Taxation." *Mimeo*, [5]

Highfill, Jannett, Douglas Thorson, and William V. Weber. 1998. "Tax Overwithholding as a Response To Uncertainty." *Public Finance Review* 26 (4): 376–91. [42]

Jones, Damon. 2012. "Inertia and Overwithholding: Explaining the Prevalence of Income Tax Refunds." *American Economic Journal: Economic Policy* 4 (1): 158–85. [42]

"Kleine Anfrage Bundestag." 2019. "Kleine Anfrage Drucksache 19/12857: Ehegattensplitting Und Gleichstellung Im Deutschen Steuersystem." [13, 47]

Kleven, Henrik, Camille Landais, Johanna Posch, Andreas Steinhauer, and Josef Zweimüller. 2019. "Child Penalties across Countries: Evidence and Explanations." *AEA Papers and Proceedings* 109 (5): 122–26. [28]

"German Coalition Agreement." 2021. "Koalitionsvertrag Zwischen SPD, Bündnis 90/Die Grünen Und FDP." [47]

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. [11]

Lardeux, Raphael. 2022. "Behavioral Cross-Influence of a Shadow Tax Bracket: Evidence from Bunching Where Income Tax Liabilities Start." *SSRN Electronic Journal*, [10]

Liebmann, Jeffrey B., and Richard J. Zeckhauser. 2004. "Schmeduling." *Working Paper*, [10]

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. [42]

Neisser, Carina. 2021. "The Elasticity of Taxable Income: A Meta-Regression Analysis." *Economic Journal* 131 (640): 3365–91. [10, 50]

Neumark, David. 1995. "Are Rising Earnings Profiles a Forced-Saving Mechanism?" *Economic Journal* 105 (428): 95–106. [42]

RDC of the Federal Statistical Office and Statistical Offices of the Federal States. 2010. "Lohn- und Einkommenssteuerstatistik 2010, Scientific Use File." [6, 46]

RDC of the Federal Statistical Office and Statistical Offices of the Federal States. 2018. "Taxpayer-Panel 2001-2018, GWAP." [8, 23]

Rees-Jones, Alex. 2018. "Quantifying Loss-Averse Tax Manipulation." *Review of Economic Studies* 85 (2): 1251–78. [5, 42]

Rees-Jones, Alex, and Dmitry Taubinsky. 2020. "Measuring "Schmeduling"." *Review of Economic Studies* 87 (5): 2399–438. [10, 39]

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. [10, 31, 32, 39]

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. [11]

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. [7, 35, 42]

Slemrod, Joel. 2019. "Tax Compliance and Enforcement." *Journal of Economic Literature* 57 (4): 904–54. [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. [15]

Thaler, Richard H. 1994. "Psychology and Savings Policies." *American Economic Review* 84 (2): 186–92. [41, 42]

Wagstaff, Joseph Van. 1965. "Income Tax Consciousness under Withholding." *Southern Economic Journal* 32 (1): 73. [10]

Appendix 1.A Empirical Strategy

In our empirical strategy, we control for binned own and spousal pre-reform labor income interacted with dummies for parenthood and residence in East Germany by adding couple-level fixed effects to our regression. We motivate the underlying reason for this in the following.

To begin, 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.A.1, which displays the percent changes in the annual withholding taxes 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 burden on the couple level but different changes in their withholding tax burden. Second, control-

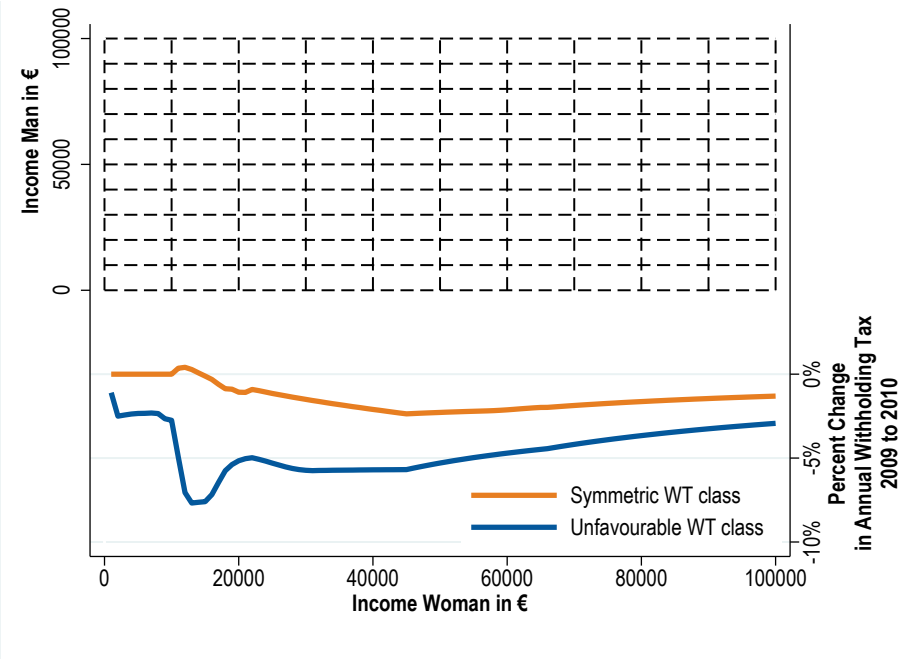


Figure 1.A.1. Illustration of Income Cell Approach

Notes: This figure illustrates the idea behind using income cell fixed effects. The lower part of the figure displays the percent change in withholding taxes induced by the reform in 2010. It is therefore a relative representation of Figure 1.2.4. 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.

ling 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.[46]

In order to address these above-outlined channels, we follow an empirical approach brought forward recently by Carbonnier, Malgouyres, Py, and Urvoy (2022) that is based on dividing observations into cells to exploit variation in treatment within each cell. In our preferred specification, we classify each individual into one of 400 cells based on own and spousal pre-reform labor income in 2009 and dummies indicating parenthood and residence in East Germany. We include the dummies to make sure that we account for the most relevant predictors of the

46. As we show in Figure 1.C.3, our survey reveals that couples in the men-favoring schedule hold more traditional gender norms than those in the symmetric schedule. Comparing only couples with a similar within-household division of labor income could mitigate this problem because, as we show in Table 1.3.1, this division is correlated with the choice of withholding tax schedule.

withholding tax class choice as shown in Figure 1.3.2. Doing so we ensure that the compared individuals are more similar in observable characteristics. The cells are created by interacting evenly spaced bins of 10,000 € of both own and partner income. Each of the 100 cells is then interacted with dummies for parenthood and residence in East 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 pre-reform income characteristics and thereby also similar couple-level gender norms. The remaining variation in treatment that we are exploiting 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.A.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 and indirectly thus also gender norms 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 and the couples' gender norms. 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.A.1.

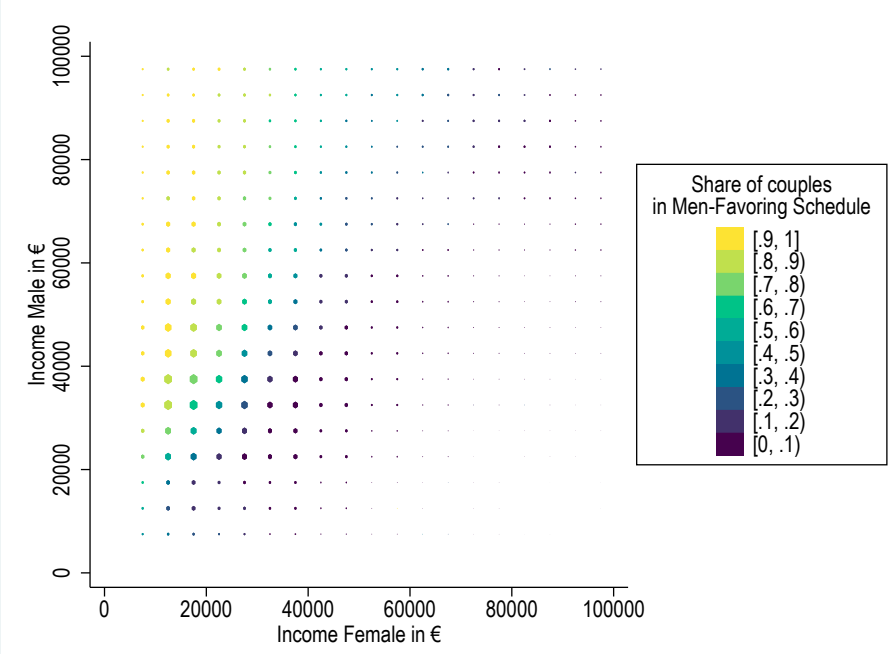


Figure 1.A.2. Heatplot: Number of Observations and Share of Couples in Men-favoring Schedule

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 € woman and man income. For example, the cell at the top right corner contains women and men with an income between 95,000 € and 100,000 €. Incomes below 5000 € are not displayed as they are not part of our analysis. The size of each bin represents the number of observations in our sample. 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 withholding tax schedule at the time of the reform. It measures how much variation between the two withholding tax classes can be exploited for each cell.

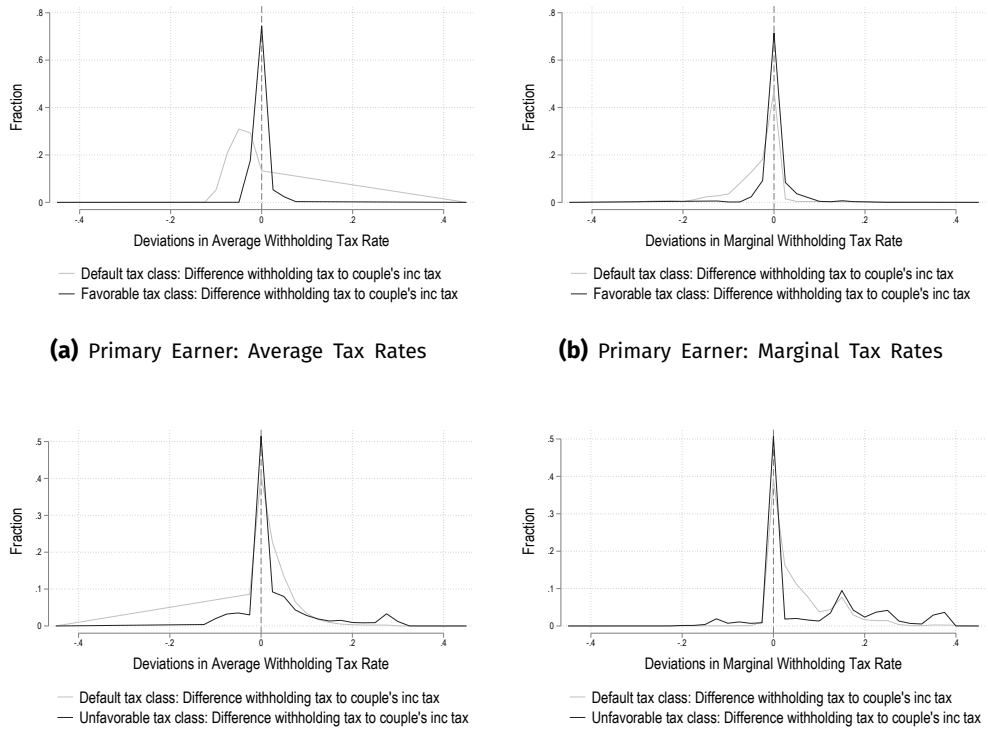
The variation that we can exploit by the bin approach is illustrated in Figure 1.A.2. It shows for each of the income cells the share of couples who are treated in a binary sense, i.e., the share of couples being in the men-favoring withholding tax schedule at the time of the reform conditional on being in the being in the men-favoring or symmetric withholding tax schedule. The size of each bin represents the number of observations, meaning that bins 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 bins there is a considerable amount of variation in the choice of withholding tax schedules.

Table 1.B.1. Different withholding tax systems

	System with a choice			
	Men-favoring	Symmetric	Matching	Scaling
Withholding Tax Women	3,100 €	800 €	2,400 €	700 €
Withholding Tax Men	4,100 €	8,500 €	6,000 €	7,700 €
Sum Withholding Taxes	7,200 €	9,300 €	8,400 €	8,400 €

Notes: The table illustrates the distribution of the withholding tax burden within a couple in the three analyzed withholding tax systems. In the example, the woman earns 20,000 €, her husband 50,000 €. All taxes are calculated based on the German tax system 2010 and the taxes are rounded to the nearest 100 €. The income tax burden of the couple sums up to 8,400 €.

Appendix 1.B Policy Implication



(a) Primary Earner: Average Tax Rates

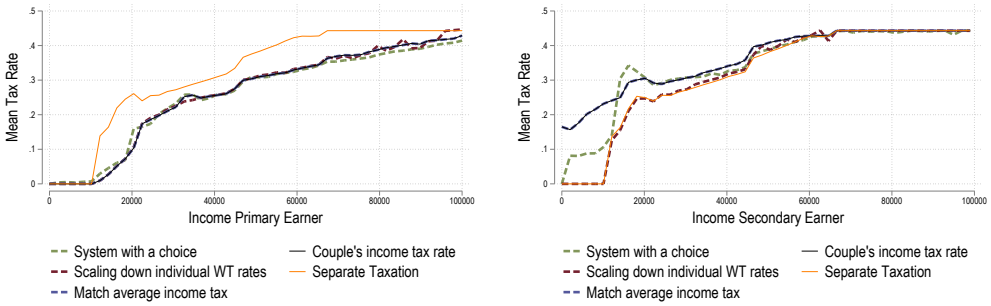
(b) Primary Earner: Marginal Tax Rates

(c) Secondary Earner: Average Tax Rates

(d) Secondary Earner: Marginal Tax Rates

Figure 1.B.1. Deviations in Average and Marginal Withholding Tax Rates

Notes: The figures display the deviations in 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. 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 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. To ease the interpretation of the figures, we ignore all non-standard deductions.



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

Figure 1.B.2. 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.C Survey

1.C.1 Implementation

We pre-registered our survey with the Open Science Foundation 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 respondents from our sample who fail at least one of two attention checks.[47] Furthermore, we restrict the sample to respondents with employed spouses. This makes sure that we can elicit information on wage transfers from and between both spouses and makes the

47. The attention checks can be found in the questions A2 and D15 in Appendix 1.C.4.

sample more comparable to the sample for our main analysis with observational data.[48] Our final sample then consists of 506 respondents (258 men, 248 women).

1.C.2 Survey Figures

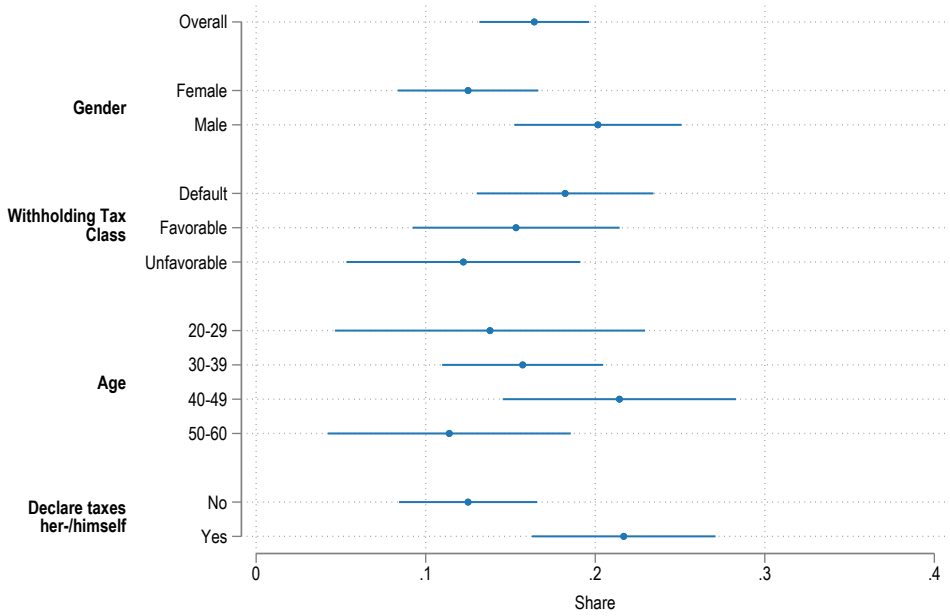


Figure 1.C.1. Knowledge of Interlinkage between Withholding Tax and Final Income Tax Burden by Subgroups

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 burden 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.C.4 for the exact wording of the question.

48. 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.

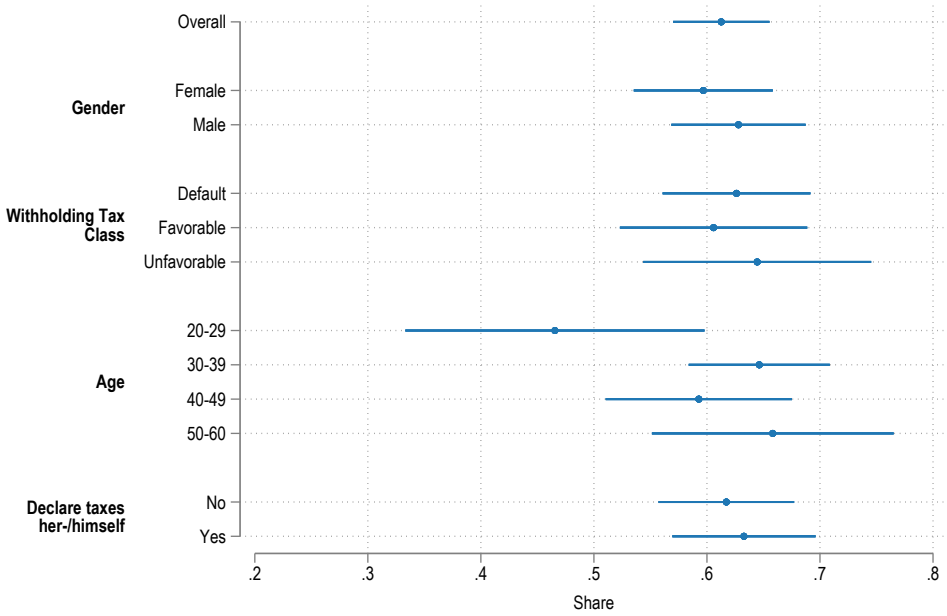


Figure 1.C.2. Knowledge of Interlinkage between Withholding Tax Classes and Monthly Payslip

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 Question D10 in Appendix 1.C.4 for the exact wording of the question.

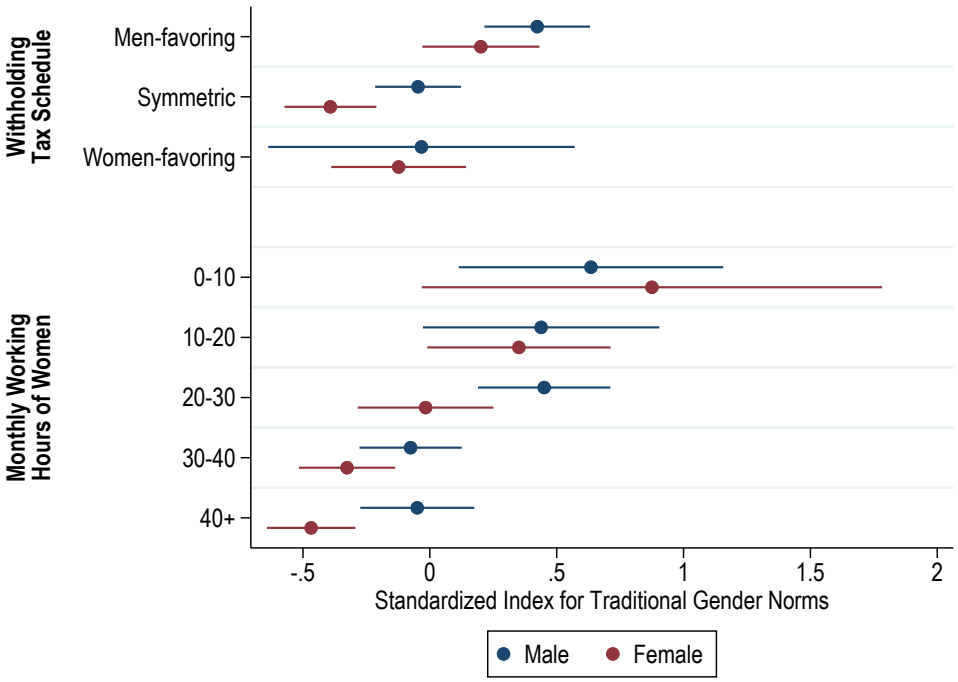


Figure 1.C.3. Gender Norms Index by Gender

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.

1.C.3 Detailed Survey Analysis

To investigate the organization of household finances, we broadly classify couples into three groups with respect to their usage of bank accounts and the destinations of the wage payments from their monthly payslips: (i) Couples without a shared bank account, (ii) couples with a shared bank account who get both their wages directly transferred to that account, and (iii) couples with a shared bank account where both spouses get their wages directly transferred to their own bank account. These categorizations are of particular relevance for couples that picked the men-favoring or women-favoring schedule because in these schedules the intra-household distribution of labor income is distorted as the withholding tax burden is partly shifted from one spouse to the other. As we focus on couples in the men-favoring schedule in our main analysis with administrative data, we

also concentrate on these here. Furthermore, we investigate onto which bank account tax refunds are transferred as they might be used to counter the distortion of the distribution of labor income.

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 remains largely unchanged as this couple is less likely to have established a compensatory sharing rule. We find that 47 % of the respondents in the men-favoring schedule do not have a shared bank account.[49] Strikingly, of those without a shared bank account, 81 % of the couples in the men-favoring withholding tax schedule let their tax refunds be transferred to the husband's bank account. This compares to 65 % of the couples in the symmetric schedule. Thus, it can be concluded that the distortions of the intra-household distribution of labor income induced by the shifting of the withholding tax burden from husbands to wives in the men-favoring withholding tax schedule are not only not diminished by the distribution of tax refunds but even aggravated.

On the other hand, 32 % of all couples in the men-favoring withholding tax schedule have a shared bank account on which both spouses get their wages directly transferred to.[50] For these households, the above-described distortion of the relative intra-household distribution of labor income appears rather unproblematic.[51] This is particularly the case because we find that almost all of these couples get their tax refund onto the shared bank account and none onto the husband's bank account. When all of a couple's labor income including any tax refund is transferred to a shared account, the choice of the men-favoring schedule likely does not directly impact the consumption opportunities of women, as they can probably use the money on the shared bank account for their private consump-

49. When considering couples irrespective of their withholding tax schedules, 45 % of the respondents state to not have a shared bank account.

50. This compares to about 21 % of the couples in the symmetric withholding tax schedule. This indicates that couples in the men-favoring withholding tax schedule use shared bank accounts as a device to compensate to a limited extent.

51. This also applies to another 3 % of the couples in the men-favoring schedule where the husband's wage income gets directly transferred to either his wife's account or the shared account and the wife's wage income gets directly transferred to her own account.

tion. However, the bargaining power within the household might still be impacted if the transfer of the withholding tax burden, induced by the men-favoring schedule, is not understood and the shifted labor income is thus mentally attributed to the husband instead of the wife.

For the 16 % of the couples in the men-favoring withholding tax schedule that have a shared bank account but receive their wage incomes to each spouse's personal bank account, it is less clear if households are compensated for the redistributive effect of the men-favoring withholding tax schedule.[52] In these cases, the money from the respective personal bank account can be seen as typically designated for the account holder's individual consumption while both partners transfer a share of their personal income to the shared bank account. We further examine in an exploratory fashion whether women are in these cases compensated for the higher withholding taxes they have to pay. Couples that take into account the redistributional consequences of the men-favoring schedule should have established a transfer rule that requires the husband to transfer a larger part of his income to the shared bank account than his wife.

We find that only 38 % of all couples in the men-favoring schedule that have a shared bank account but receive their wage incomes to each spouse's personal bank account make use of such a transfer rule. This means that even among couples in the men-favoring schedule with a shared bank account, 21 % do not seem to account for the distortion effects of being in the men-favoring schedule. Thus, we can monitor a counteracting strategy for only 42 % of all couples in the men-favoring schedule (those with a shared bank account who either already get their wages directly transferred accordingly or do compensatory payments from the husband to the wife afterward).

52. This applies even more to another 2 % of the couples in the men-favoring schedule where the husband's wage income gets directly transferred to his own account while the wife's wage income gets directly transferred to either the husband's account or the shared account.

1.C.4 Survey Questions

This section documents the survey questions. Section 1.C.4.1 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.C.4.2 provides a translation into English.

1.C.4.1 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 Datenschutz-Grundverordnung, 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.C.4.2 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 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 bear the lowest final annual tax burden (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 payroll tax depending on the withholding tax classes for a couple where both partners earn a gross monthly income of €3,500.

	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 €

You can see that the choice of withholding tax class greatly affects the payroll tax you pay. If both partners are in withholding tax class 4, both partners each pay €700 in payroll tax. If a partner is in withholding tax class 3 instead, she/he pays €350 in payroll tax. In withholding tax class 5, €1,000 in payroll tax is due. As you can see, the payroll tax you pay depends heavily on the withholding tax class you choose. But your partner's payroll 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 payroll tax, I knew that the choice of withholding tax class influences my partner's payroll 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 payroll taxes – even if both partners earn the same amount)

Page break

D11 Please take enough time to understand the following information. The final tax burden 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 burden of a married couple. Only the payroll tax is affected by the withholding tax class selection:

	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 €

The payroll tax you pay monthly as a couple is offset against the income tax at the end of the year. So if your paid payroll 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 payroll 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 burden of a married couple, but only on the payroll 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 payroll tax paid does not affect the final tax burden, I knew that the choice of withholding tax classes does not affect the final tax burden)

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 bear the lowest final annual tax burden (corresponds to income tax)? All three withholding tax class combinations mentioned exist.

(me in withholding tax class 5 and my partner in withholding tax class 3, me in withholding tax class 4 and my partner in withholding tax class 4, me in withholding tax class 3 and my partner in withholding tax class 5, whatever, don't know)

Page break

D13a Withholding Tax classes therefore have no effect on the final tax burden of a married couple, only on the payroll 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 burden of a married couple, only on the payroll 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 burden, only on the payroll 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

A27 Do you have any comments on the survey or on the subject of withholding tax classes?

Appendix 1.D Additional Descriptive Statistics

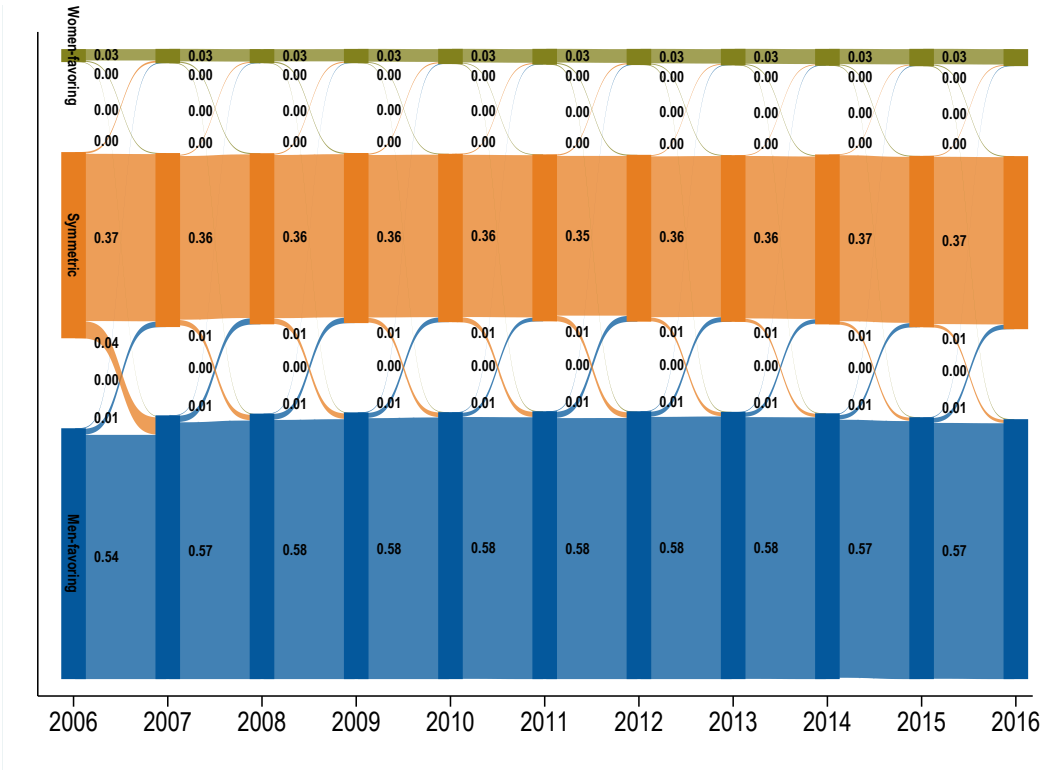


Figure 1.D.1. Changes in the Choice of Withholding Tax Classes over Time

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 graph uses information on all couples in the 5 % sample of the TPP, and no sample restrictions are applied. The figure shows that the choice of withholding tax schedules is relatively stable over time. Only a few couples change between withholding tax schedules and they typically stick with their choice of withholding tax schedules. Note that we only consider direct transitions between withholding tax schedules. 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 1.2.3 stems from the changed data composition. While this figure has no sample restrictions, the right side of Figure 1.2.3 documents the share only for couples where both spouses are working.

Table 1.D.1. Descriptive Statistics for the Year 2009

	Men-Favoring	Symmetric
Income Wife	19949.01 (8909.25)	33411.34 (13820.28)
Income Husband	49192.86 (17347.79)	39399.81 (15881.09)
Female Income Share	0.29 (0.09)	0.46 (0.11)
Age Wife	46.9 (5.83)	47.1 (6.44)
Age Husband	49.16 (5.98)	49.11 (6.41)
Eastern Germany	0.07 (0.26)	0.34 (0.47)
Has a Child	0.53 (0.5)	0.24 (0.43)
Number of Children	1.21 (0.94)	0.64 (0.82)
Catholic Wife	0.39 (0.49)	0.22 (0.42)
Catholic Husband	0.37 (0.48)	0.2 (0.4)
Public Servant Wife	0.12 (0.32)	0.14 (0.34)
Public Servant Husband	0.22 (0.42)	0.18 (0.38)
N	11366	11867

Notes: The table displays descriptive statistics for the year 2009 for the unbalanced panel for couples who picked either the men-favoring or symmetric withholding tax schedule. They are calculated based on the sample restrictions outlined in Section 1.3.2. Specifically, we focus on households with dual earners in 2009, in which both partners have received no unemployment benefits and short-time work compensations in 2009, are between 20 and 60 years old in 2009, have no income from self-employment of more than 1,000 € in 2009 and whose incomes were stable between 2006 and 2009, i.e., the income for both household members fluctuated by less than 25 % from one year to the other.

Table 1.D.2. 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)
Income Husband (1000 Euro)	−0.0
	(0.00)
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)
Public Servant Wife	0.031***
	(0.012)
Public Servant Husband	0.008
	(0.01)
Age Wife	0.003**
	(0.001)
Age Husband	0.005***
	(0.001)
Commuting Days Wife (100 days)	−0.005
	(0.004)
Commuting Days Husband (100 days)	−0.0
	(0.004)
Commuting Distance Wife (100 km)	0.009
	(0.027)
Commuting Distance Husband (100 km)	0.012
	(0.017)
Constant	0.891***
	(0.054)
N	11039.0
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, using the balanced sample. Heteroscedasticity-robust standard errors are displayed in brackets.

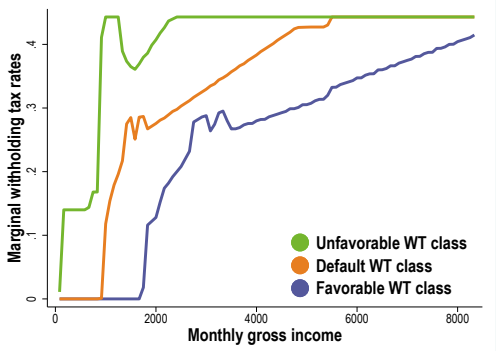


Figure 1.D.2. Marginal Withholding Tax Rates 2009

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

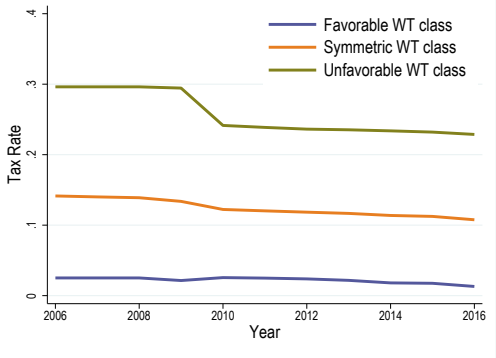


Figure 1.D.3. Development of the Average Withholding Tax Rate

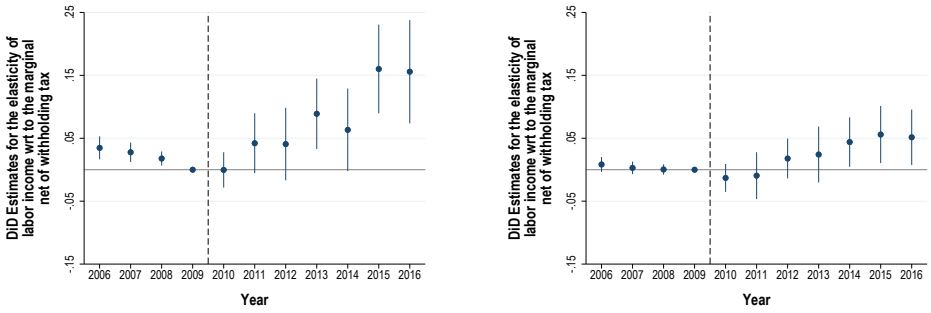
Notes: The figure plots the size of withholding tax payments depending on the withholding tax class for the period from 2006 to 2016. It illustrates for an income of 25,000 € that there were no other major reforms changing withholding tax payments except for the 2010 reform that we study in this paper. The same holds true for all other incomes.

Appendix 1.E Additional Regression Results

Table 1.E.1. Static Diff-in-Diff Results

	Women		Men	
	(1)	(2)	(3)	(4)
DiD Estimate	0.055***	0.048**	0.026**	0.016
	(0.020)	(0.020)	(0.012)	(0.013)
Cell FE		✓		✓
N	212,547	212,547	212,547	212,547
Adj. R-Squared	0.090	0.117	0.073	0.089

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1 using the unbalanced sample. All regressions include individual fixed effects and control 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 East Germany, and years. Results using the balanced estimation sample can be found in Table 1.5.1. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.



(a) Effect on Female Log Income **(b)** Effect on Male Log Income

Figure 1.E.1. Event Study Diff-in-Diff Estimates

Notes: The figure plots the estimates for the elasticity of labor income with respect to the withholding tax estimated based on the dynamic version of Equation 1.4.1 for women and men using the unbalanced sample. The dependent variable is the log income of the individual, and the independent variable is the treatment intensity. Treatment intensity is defined as the percent change in the marginal net-of-withholding-tax rate of the woman induced by the reform of the withholding tax in 2010. All regressions include individual fixed effects and control 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 East Germany, and years. Confidence intervals are plotted at the 95 % level and based on heteroscedasticity-robust standard errors. Note that the sample excludes households, where at least one member experienced a drop in income by more than 25 % from one year to the next before 2010 to ensure that no individuals directly hit by the financial crises are part of the sample. This explains the smaller standard errors before the reform. The underlying regression coefficients can be found in in columns (2) and (4) of Table 1.E.3.

Table 1.E.2. Event Study Diff-in-Diff Results

	Women		Men	
	(1)	(2)	(3)	(4)
2006	0.026**	0.018	0.024***	0.027***
	(0.011)	(0.011)	(0.008)	(0.008)
2007	0.016*	0.010	0.007	0.009
	(0.009)	(0.009)	(0.007)	(0.007)
2008	0.011*	0.013*	0.007	0.007
	(0.006)	(0.007)	(0.005)	(0.006)
2009	.	.	.	.
	(.)	(.)	(.)	(.)
2010	0.025**	0.019*	0.009	0.005
	(0.010)	(0.010)	(0.008)	(0.008)
2011	0.058***	0.052***	0.012	0.008
	(0.017)	(0.017)	(0.011)	(0.011)
2012	0.085***	0.069***	0.018	0.015
	(0.022)	(0.022)	(0.013)	(0.014)
2013	0.143***	0.127***	0.025*	0.021
	(0.025)	(0.025)	(0.014)	(0.015)
2014	0.156***	0.133***	0.030**	0.026*
	(0.027)	(0.026)	(0.014)	(0.014)
2015	0.190***	0.168***	0.027*	0.022
	(0.028)	(0.027)	(0.016)	(0.017)
2016	0.223***	0.197***	0.025	0.025
	(0.030)	(0.030)	(0.018)	(0.019)
Cell FE		✓		✓
N	121,429	121,429	121,429	121,429
Adj. R-Squared	0.336	0.375	0.301	0.317

Notes: The table displays the estimates for the elasticity of labor income with respect to the withholding tax estimated based on the dynamic version of Equation 1.4.1 for women and men using the balanced sample. The dependent variable is the log income of the individual, and the independent variable is the treatment intensity. Treatment intensity is defined as the percent change in the marginal net-of-withholding-tax rate of the woman induced by the reform of the withholding tax in 2010. All regressions include individual fixed effects and control 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 East Germany, and years. Note that the sample excludes households, where at least one member experienced a drop in income by more than 25 % from one year to the next before 2010 to ensure that no individuals directly hit by the financial crises are part of the sample. This explains the smaller standard errors before the reform. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 1.E.3. Event Study Diff-in-Diff Results

	Women		Men	
	(1)	(2)	(3)	(4)
2006	0.051***	0.035***	0.011*	0.008
	(0.009)	(0.009)	(0.006)	(0.006)
2007	0.036***	0.028***	0.005	0.003
	(0.008)	(0.008)	(0.005)	(0.005)
2008	0.018***	0.018***	0.002	0.000
	(0.006)	(0.006)	(0.004)	(0.004)
2009	.	.	.	.
	(.)	(.)	(.)	(.)
2010	0.001	-0.000	-0.004	-0.013
	(0.014)	(0.014)	(0.011)	(0.011)
2011	0.041*	0.042*	0.002	-0.009
	(0.023)	(0.024)	(0.019)	(0.019)
2012	0.054*	0.041	0.026	0.018
	(0.029)	(0.029)	(0.016)	(0.016)
2013	0.108***	0.089***	0.033	0.024
	(0.028)	(0.029)	(0.021)	(0.023)
2014	0.097***	0.063*	0.056***	0.044**
	(0.034)	(0.033)	(0.019)	(0.020)
2015	0.184***	0.160***	0.076***	0.056**
	(0.036)	(0.036)	(0.021)	(0.023)
2016	0.176***	0.156***	0.069***	0.051**
	(0.041)	(0.042)	(0.022)	(0.023)
Cell FE		✓		✓
N	212,547	212,547	212,547	212,547
Adj. R-Squared	0.090	0.117	0.073	0.089

Notes: The table displays the estimates for the elasticity of labor income with respect to the withholding tax estimated based on the dynamic version of Equation 1.4.1 for women and men using the unbalanced sample. The dependent variable is the log income of the individual, and the independent variable is the treatment intensity. Treatment intensity is defined as the percent change in the marginal net-of-withholding-tax rate of the woman induced by the reform of the withholding tax in 2010. All regressions include individual fixed effects and control 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 East Germany, and years. Note that the sample excludes households, where at least one member experienced a drop in income by more than 25 % from one year to the next before 2010 to ensure that no individuals directly hit by the financial crises are part of the sample. This explains the smaller standard errors before the reform. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 1.E.4. Heterogeneity Analysis: Static Diff-in-Diff Results

	Women (1)	Men (2)
Panel A: East vs. West Germany		
West	0.058***	0.015
	(0.021)	(0.013)
East	-0.032	0.030
	(0.053)	(0.036)
Panel B: Level of Pre-Reform Commuting		
Low Commuting	0.047**	0.011
	(0.023)	(0.016)
High Commuting	0.038	0.013
	(0.036)	(0.021)
Panel C: Parent vs. Non-Parent		
Non-Parent	0.034	0.031
	(0.027)	(0.019)
Parent	0.059**	-0.003
	(0.027)	(0.017)
Panel D:Age of Youngest Child		
Youngest Child below 6	0.107	0.036
	(0.250)	(0.071)
Youngest Child betw. 6 and 18	0.058**	-0.004
	(0.027)	(0.017)
N	212,547	212,547
Adj. R-Squared	0.105	0.084

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1, allowing for treatment heterogeneity by observable characteristics and using the unbalanced sample. All regressions include individual fixed effects and control for potentially time-varying characteristics of the couple. Panel A includes cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with dummies for parenthood, and years. Panel B includes cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with dummies for parenthood, residence in East Germany, and years. Panel C and C include cell fixed effects controlling for binned own and spousal pre-reform labor income interacted with years. Results using the balanced estimation sample can be found in Table 1.5.2. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 1.E.5. Static Diff-in-Diff Results controlling for Average Tax Rate

	Women		Men	
	(1)	(2)	(3)	(4)
Marginal WT Rate	0.048**	0.038**	0.016	0.017
	(0.020)	(0.019)	(0.013)	(0.013)
Average WT Rate		0.008**		-0.009***
		(0.004)		(0.003)
Cell FE	✓	✓	✓	✓
N	212,547	210,108	212,547	210,108
Adj. R-Squared	0.117	0.151	0.089	0.091

Notes: The table displays the results of the static diff-in-diff estimation as laid out in Equation 1.4.1 while additionally including the change in the average net-of-withholding tax rate as an independent variable. The estimation is performed using the unbalanced sample. All regressions include individual fixed effects and control 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 East Germany, and years. Results using the balanced estimation sample can be found in Table 1.5.3. Standard errors are clustered on the individual level. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Chapter 2

Shift to Remote Work and Parental Division of Labor

Joint with Hans-Martin von Gaudecker, Radost Holler, Christian Zimpelmann

2.1 Introduction

Despite some progress towards gender convergence in the division of labor within households in recent decades, in many countries, mothers still tend to assume a disproportionate share of childcare and domestic responsibilities, while fathers work outside the home. This pattern is at least partially driven by the need for at least one parent's job to be compatible with childcare needs. Most parents must be able to step in at short notice when children are unable to attend school or daycare due to illness or other reasons. These responsibilities are often taken on by mothers, who may choose jobs with fewer hours or greater flexibility in order to accommodate them. Fathers, on the other hand, typically specialize in market work – potentially driven by non-linear returns to working hours (Gicheva, 2013; Bick, Blandin, and Rogerson, 2022) which diminishes the appeal for parents to share market and non-market work equally.

One promising approach to mitigating the gendered division of labor is thus to ensure that both parents' jobs are compatible with childcare responsibilities, without altering other factors such as remuneration. A potential avenue to achieve this is by increasing the amount of work that can be done from home, provided that employers do not use this as a means of selecting employees. Working from home inherently entails an increase in time spent at home, along with a typically higher degree of schedule flexibility and a reduction in commuting time and associated frictions (for the last two items, see Aksoy, Barrero, Bloom, Davis, Dolls, et al., 2023). Unfortunately, the relationship between schedule flexibility and the

division of labor within households is difficult to identify as subjects might choose jobs based on their current or expected role in the household.

In this paper, we exploit the way the COVID-19 pandemic has evolved in the Netherlands to overcome this challenge by examining the strong exogenous increase in the availability of remote work. We use representative survey data from the LISS Panel, an online survey based on a true probability sample of the Dutch population, combined with administrative labor market records from CBS Netherlands. We argue that among the multitude of effects that the pandemic had on family lives, we can isolate the effects of working from home for several reasons. First, schools and daycare were open in the Netherlands except for two (primary schools and daycare) to three (secondary schools) months in the spring of 2020. Consequently, total hours spent on childcare did not change in the months of November of 2020 or 2021 relative to 2019. Second, generous wage-support schemes were in place, which left income unchanged for most households and helped that the unemployment rate did not move much in general and actually decreased for parents. Third, we show that the potential for working from home has little explanatory power for hours worked from home just before the pandemic. This drastically changed with the onset of the pandemic and the government's advice to work from home. Put differently, the potential to work from home was there before the pandemic, but it was realized to a large extent only after March 2020.

We start out by showing that the gains in job flexibility through the shift to remote work are asymmetrically distributed among parents. On average, fathers gained more flexibility than mothers. This asymmetry is driven by two factors. First, fathers tend to work in jobs with a higher degree of remote work potential. Second, they work more hours, which is more important quantitatively.

Relying on time use data from the LISS Panel, we find that fathers, as well as mothers, use their newly gained job flexibility for childcare provision. Given the asymmetric changes in job flexibility, the gender gap in childcare provision decreased substantially. Before the pandemic, mothers provided 14 more hours of care to their children than their partners. In late 2021, this gap had shrunk to 10.7 hours. We further show that the potential hours the parents can work remotely strongly predicts childcare hours from 2020 on, but not before. Potential remote hours explain more than half of the decrease in the observed childcare gap.

To investigate the effect of the shift to remote work on labor supply, we use labor market information on the full population of Dutch parents contained in the Dutch administrative data provided by the Centraal Bureau voor de Statistiek (CBS). The larger and longer panel compared to the time use data of the LISS Panel enables us to detect more subtle changes in the labor supply as well as to implement a more sophisticated identification strategy.

Using the administrative data, we first show that a pre-existing trend of increasing full-time work among mothers accelerated during the pandemic. We then aim

to identify whether this acceleration is driven by fathers gaining more job flexibility using an identification strategy that resembles an Event-Study combined with a Difference-in-Differences approach with continuous treatment. That is, we compare the relationship of partners' remote work potential and own working hours over the 2018-2021 period with the same relationship in the years before. We find that mothers and fathers indeed increase their labor supply in response to their partners' newly gained job flexibility. For mothers, we find evidence that roughly half of this effect is driven by reduced commuting time of their partners.

Our results, thus, suggest that increased possibilities to work from home allowed couples to choose a more balanced distribution between market and nonmarket work. More generally, it highlights that policies that make it easier to combine career ambitions and childcare time can be effective in reducing gender inequality within households.

Our paper is structured as follows. We describe our data and the basic socioeconomic characteristics in the next section. Subsequently, we present the setting of our analysis: The way the pandemic evolved in the Netherlands, the background on trends in parents' labor supply and childcare division, and our measures of job flexibility. In Section 2.4, we present our results on the effects of the pandemic on parents' childcare division and labor supply. We conclude in Section 2.5.

2.2 Data

Our study is based on customized survey data from the Longitudinal Internet Studies for the Social Sciences (LISS) panel, population-wide administrative records from Statistics Netherlands, and the Dutch national working conditions survey. Both survey data sets are linked to the population registers at the individual level. We observe household members' time use only in the LISS data, consequently we will use it for all analyses regarding the division of childcare tasks. For the analysis of labor supply behavior, we can recur to the population registers, which are several orders of magnitudes larger. However, we do not directly observe the amount of work performed remotely in those data. Hence, we impute a measure for the individual remote work potential based on observed characteristics in the working conditions survey. We describe the data sets in detail in the following subsections.

Throughout our analysis, we consider heterosexual couples where both partners are between 18 and 55 years of age and who have at least one child below the age of 16 in the household. The average age of our sample members is a little more than forty years, with a difference of two years between fathers and mothers. The mean number of children is about two; the average age of the youngest child is between 6.3 and 7 years (for detailed descriptives see Table 2.A.4 in the Appendix).

2.2.1 Customized survey data from the LISS Panel

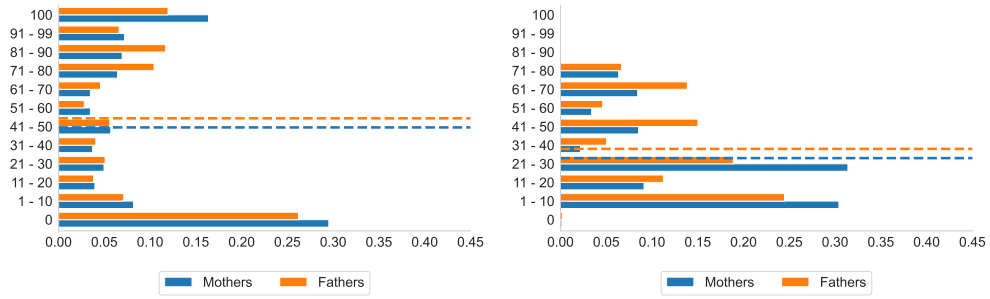
The LISS panel is based on a probability sample of individuals registered by Statistics Netherlands; it has been running since 2007 and comprises about 7000 individuals in 4000 households. The LISS panel is administered by CentERdata, a survey research institute affiliated with Tilburg University, the Netherlands. Each year, the LISS panel runs ten core surveys, which cover a wide range of topics, including health, education, work, and family. Taken together, these data are comparable in scope to popular surveys like the Panel Study of Income Dynamics (U.S.), Understanding Society (U.K.), or the Socio-Economic Panel (Germany). On top of that, the LISS panel allows researchers to run their own questionnaires. In this paper, we make use of two sets of surveys that we ran ourselves or helped design.

First, we employ time use information collected in comparable questionnaires in November 2019, April 2020, November 2020, and November 2021. In these surveys, people are asked to distribute the hours of the past week over different activities, including childcare, commuting, work at the usual workplace, and work at home. Appendix 2.A.1 reports more information on the survey; we will describe the evolution of our variables of interest below in Section 2.3.3.

Second, we make use of a series of CoViD-19 questionnaires (documented in von Gaudecker, Zimpelmann, Mendel, Siflinger, Janys, et al., 2021, see Appendix 2.A.2 for a detailed description) fielded in March–December 2020. Most importantly for this study, we elicit a measure of remote work potential. In May 2020, we ask participants "What percentage of your normal work *prior to the CoVid-19 outbreak* can you do while working from home?". We interpret the resulting answers to measure remote work potential. We repeated this question in December 2020, but inquired about the share of tasks at the current job that can be done from home instead of the pre-pandemic situation: "What percentage of your normal work can you do while working from home?". Motivated by a high correlation between the May and December measures (0.82), we take the mean of the data that is available at the individual level.[1]

We plot the distribution of measured remote work potential in Figure 2.2.1a. The distribution is very polarized. For both genders, more than a quarter of jobs require presence at the workplace for all tasks. At the same time, tasks can entirely be performed remotely in more than 10% of jobs. Average remote work potential is somewhat lower for mothers than for fathers (45 % vs. 49 %). The distribution is in line with other data. For example, a similar question asked in broad samples of

1. For more details on the correlation between the answers in May and December consult the Appendix of Zimpelmann, Gaudecker, Holler, Janys, and Siflinger (2021) which discusses the correlation between the answers in May and December as well as the marginal distributions in greater detail.



(a) Share of tasks that can be done from home

(b) Imputed remote work capability (CBS)

Figure 2.2.1. Remote working potential by gender

Notes: Figure 2.2.1a displays the distribution of the variable "share of tasks that can be done from home" by gender in the LISS sample (403 mothers and 393 fathers). Dashed lines display the mean by gender (45 % for mothers and 49 % for fathers). Figure 2.2.1b shows the distribution of the imputed remote work capability by gender in the CBS in the year 2019. For the imputation, we calculate the average remote work capability for all sector × education combinations in the NEA and impute the remote work capability in the CBS. Dashed lines indicate the mean for each gender (28 % for mothers and 33 % for fathers).

the U.S. and U.K. populations returned means of 41–43 % (Adams-Prassl, Boneva, Golin, and Rauh, 2022).

Our LISS sample consists of 403 mothers and 393 fathers for whom we have 1,190 and 1,044 observations, respectively, spread over the four waves of the time use survey. Unit non-response leads to some sampling variation across waves.[2] In order to make the descriptive statistics in Section 2.3.3 more meaningful, we re-weight the sample in terms of age of the youngest child. Doing so has no impact on our main, regression-based results.

2.2.2 Population-wide administrative data, Working Conditions Survey

We use detailed administrative microdata on the entire Dutch population from Statistics Netherlands (CBS) for our analyses of labor supply; i.e., anything that does not require time use information. We make use of gender, household composition, education, labor force status (dependent work in full time or part time, self-employment, unemployment, and being outside the labor force), sector, commuting distance, and working hours. We do not observe working hours for the self-employed. The labor market information is recorded monthly for each individual; we extract it for the months of November in the 2012–2021 period. We do

2. These numbers already account for individual linkage to the CBS data, which we will describe in the next section. We do so in order to update information on working hours and household composition, which is particularly useful when these individuals did not participate in one or more waves of the survey.

so for computational feasibility and to harmonize the timing with the LISS data. We use actual working hours from the first spell in each month. We will describe the trends in these variables in Section 2.3.2 just below.

The administrative data does not contain direct information on remote work. We thus impute remote work ability based on the National Working Condition Survey (NEA). Using survey information on *actual* remote work in the fall of 2020 from 35,000 individuals, we calculate the average share of remote work. The resulting distribution (see Appendix Figure 2.A.1) has a similar shape as in the LISS data, which is based on a completely different question.[3] We then calculate average remote work potential for all sector × education cells and use the resulting measure in the CBS data, treating it as a time-invariant characteristic. Again, the inherent assumption is that during our period of study, the potential to work remotely was exploited to much larger extent only during the pandemic. See Appendix 2.A.5 for a more detailed description of the imputation procedure and empirical evidence on the (non-) prevalence of remote work before 2020. Figure 2.2.1b shows the results for our population-wide data and reveals similar gender differences to the LISS data in Panel a. Because the imputation procedure integrates all within-cell heterogeneity, it comes as no surprise that mass is shifted from the extremes towards the middle of the unit interval.

Dingel and Neiman (2020) use an alternative approach to measure remote work potential. They classify occupations into those that can or cannot be performed from home based on typical tasks and work experiences elicited in surveys. Aggregating over higher level occupations or industries reveals the share of jobs that can be performed from home. We prefer our directly elicited measure for several reasons. First, tasks within occupations might change at the start of the pandemic in ways which cannot be captured by the descriptions of tasks elicited before the pandemic. Second, our measure is more fine-grained in the sense that it allows respondents to indicate that they can do a part but not all of their work remotely – and more than half of them do so. Finally, it captures differences in remote work potential within occupations (e.g., due to a different firm culture). Adams-Prassl et al. (2022) show their self-reported measure of remote work ability to be strongly related to occupations and sectors, but also find substantial heterogeneity within occupations. While we see this additional heterogeneity primarily as an advantage, it might raise the concern of response bias on the individual level. Our measure based on the NEA addressed this concern as it is constant within sector × education cells – while sharing the other two advantages described above (it is based on post-pandemic data and fine-grained). For the analyses based on

3. While differences across genders and the overall shape of the distribution are similar, the means in the NEA data are lower at (28 % for mothers and 33 % for fathers). We attribute this to the fact that even during the pandemic, remote work potential was not fully realized.

the LISS sample, we use the self-reported measure of remote work potential. In robustness exercises based on the NEA measure, we obtain very similar results.[4]

2.3 Setting

In this section, we describe the broader environment for our analysis along with stylized features emanating from our data. First, we illustrate the policy environment during the first two years of the CoViD-19 pandemic. We then highlight some key features of the parental division of labor regarding market and non-market work before and during the pandemic. Finally, we go through our measures of remote work—both the potential for doing so and its realizations—over the period of our analysis.

The contents of this section aim to illustrate why we deem it plausible that we can isolate the effect of remote work ability on parents' outcomes during the time period of our analysis.

2.3.1 The CoViD-19 pandemic in the Netherlands

From March 2020 until the end of the our data collection in November 2021, a set of measures were in place to slow the spread of the SARS-CoV-2 virus in the Netherlands. We will describe key features of the policy environment that are relevant for our analysis (Zimpelmann et al., 2021, features a detailed description with a focus on labor market issues). In general, measures were more lenient than in many other countries. In particular, no general curfew or stay-at-home mandate was in place at any point in time.

Figure 2.3.1 shows the timeline of relevant government policy measures at the points in time of our data collection in the LISS panel. In November 2019, the world lived in blissful ignorance of SARS-CoV-2's existence. In mid-March 2020, limits on social gatherings were imposed and many businesses involving personal contacts were closed, such as restaurants, bars, and hairdressers. Others like stores for clothes or utilities remained open if they were able to maintain the social distancing rules. Public locations were accessible and the use of public transportation was possible.

Many of these restrictions were lifted over the summer of 2020. The majority, however, were in place again during November 2020. After the winter, they were eased again and much milder measures came back in the subsequent fall/winter.

With the onset of the initial restrictions, schools and childcare facilities were closed for a period of two (daycare, primary schools) to three (secondary schools) months. From late spring and summer of 2020 on, policy makers made it very

4. **hansen2023remote** use yet another approach based on job postings to measure realized remote work across industries and sectors over time.

Figure 2.3.1. Timeline of relevant government policy measures at the points in time of our data collection.

Notes: The policy measures are obtained from the official government recommendations, which can be found on https://www.government.nl/latest/news. The unemployment rates are taken from the official statistics from CBS Netherlands.

clear that schools and childcare facilities would be the last institutions to close in case of renewed tightening of restrictions. Except for slightly prolonged vacations around Christmas 2020, this promise was kept. Actual closures were thus very limited in comparison to many other countries.

A comprehensive set of economic support measures accompanied the social distancing restrictions. The largest and most influential policy was the short-term allowance (Noodmaatregel Overbrugging voor Werkgelegenheid, NOW), which subsidized labor hoarding with a 100% wage replacements rate. Dependent employees did not see their incomes drop and there was little movement in unemployment or labor force participation (Zimpelmann et al., 2021). Figure 2.3.1 confirms that the overall unemployment rate was low throughout the 2019–2021 period. Starting in March 2020, the government strongly encouraged remote work.

2.3.2 Market work

Parents' labor force participation was high before the pandemic and increased further during 2020 and 2021. The distribution over different categories of employment (full-time employed, part-time employed, self-employed) or lack thereof (unemployment, out of the labor force) varies considerably with gender. Most notably, mothers' part-time share is with about 57 % of the population high compared to other Western countries. Moreover, their share out of the labor force is with about 20 % or more about twice as high as that of fathers.

Table 2.3.1 contains the labor market status for our sample of parents for the months of November in the 2012–2021 period. The first two columns in the upper panel show that the share of mothers who are not working decreased considerably over those years. To be precise, the fraction outside the labor force went from 25 %

to 19 % with the bulk of the decrease happening between 2016 and 2021. The unemployment share decreased from 2.5 % to 0.8 %.[5] The same trend is present for fathers, albeit at lower levels. The fraction outside the labor force went from 11 % to 9.4 %; the fraction of unemployed fathers decreased from 2.7 % to 0.6 %. In 2020, these trends were stalled but continued in 2021.

Table 2.3.1. Labor market status over time

	Mothers					Fathers				
	FT	PT	S/E	UE	OOL	FT	PT	S/E	UE	OOL
2012	8.0	58.7	6.1	2.5	24.7	65.4	9.0	12.3	2.7	10.8
2013	7.9	57.5	7.7	2.5	24.6	63.2	9.8	13.5	2.4	11.2
2014	7.8	57.3	8.2	2.1	24.8	61.1	11.8	14.0	1.7	11.5
2015	8.7	56.4	8.7	1.9	24.5	63.4	9.5	14.5	1.4	11.4
2016	9.2	56.2	9.3	1.8	23.7	63.3	9.5	15.1	1.3	10.9
2017	9.2	57.1	9.7	1.4	22.8	62.9	10.1	15.4	1.0	10.7
2018	9.6	57.6	10.2	1.1	21.6	62.7	10.3	16.0	0.8	10.3
2019	10.0	57.8	10.8	1.0	20.5	62.0	10.8	16.7	0.8	9.9
2020	10.5	56.9	11.2	1.1	20.4	61.1	10.8	17.2	0.9	10.0
2021	11.7	56.8	11.7	0.8	19.0	61.3	11.2	17.6	0.6	9.4

Notes: The table shows the labor market status of all working-age (18-55 years old) parents with a child below 16 years of age by year and gender. FT – full-time employed; PT – part-time employed; S/E – self-employed; UE – unemployed; OOL – out of the labor force. Individuals are classified as unemployed when they are receiving unemployment benefits. They are classified as out of the labor force when there are no working hours, no self-employment status, and no unemployment benefits recorded in the administrative data. Consistent with the official definition of CBS Netherlands, we classify individuals to be working part-time if they work less than 35 hours per week. Standard errors are not shown because all of them are below 0.0005 and thus rounded to zero. The data are measured in November of each year and based on administrative data of Statistics Netherlands (CBS).

In the aggregate, unemployment saw a slight increase from 4.5 % in November 2019 to 5 % in November 2020 (see Figure 2.3.1). Hence, parents with only an icrease of 0.1 percentage points towards 2020 seem to be even less affected than the general population, which may partly be explained by the type of jobs (e.g., very few parents work in the catering sector). Importantly for our purposes, there is no evidence that parents dropped out of the labor force to take care of their children. This stands in stark contrast to countries where schools and daycare facilities were closed for prolonged periods of time (e.g., Heggeness and Suri, 2021).

5. Note that unemployment is measured as receipt of unemployment benefits, so by ordinary economic definitions, we might be putting some individuals into the wrong category of inactivity.

Mothers' part time employment decreased from up from 59 % in 2012 to 57 % in 2021. In 2016, 9.2 % of mothers were employed full-time – i.e., worked 35 hours or more. The share went up by 0.7 percentage points between 2016 and 2019 and increased by another 1.8 percentage points between 2019 and 2021. Hence, there was a strong acceleration in the increase of mothers' full-time employment during the first two years of the pandemic. As a result, 11.7 % of mothers were employed full-time in 2021 as opposed to 9.9 % in 2019 and 9.2 % in 2016.

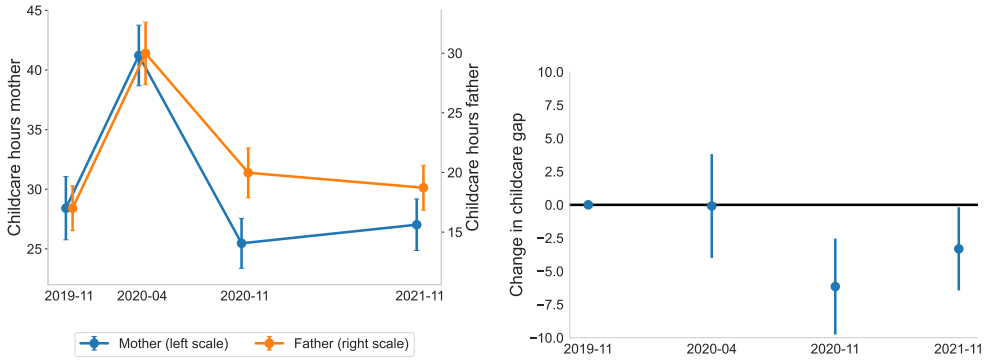
Fathers see a slight decrease in full-time employment and an increase in part-time employment over the observation period. In 2016, 63.1 % of fathers worked 35 hours or more, while 9.2 % worked less than 35 hours. The share of fathers in full-time employment decreased by 1.4 percentage points between 2016 and 2019, while the share of fathers in part-time employment increased by 1.6 percentage points to 10.8 % in 2019. During the pandemic full-time employment dropped by another 0.7 percentage points until November 2020, but recovered again by 0.3 percentage points by November 2021. Hence, decreases in fathers' full-time employment over the entire pandemic period are similar to their pre-pandemic trends. Similarly, part-time employment of fathers increased only by additional 0.3 percentage points over the two years of the pandemic.

The trends described in the previous paragraph hold up when looking at working hours of dependent employees directly instead of categories. In particular, average working hours of mothers increased from 25.2 in 2016 to 26 in 2019. This trend accelerated slightly during the pandemic and by 2021, mothers worked 26.8 hours on average (all numbers referred to in this paragraph are in Appendix Table 2.A.7). Among fathers, average working hours declined slightly from 38.4 in 2016 to 38.1 in 2019 and stayed at this level until the end of our sample period.

In July 2020, a partner leave reform came into place which increased the number of weeks of paid leave for fathers from one week to six weeks. As all six weeks have to be taken during the first six months after birth, we do not expect any effects on parents with older children in the short-term and, generally a very limited effect on the labor supply of mothers. Furthermore, the reform cannot explain the heterogeneity by ability to work from home in labor supply decisions that we are documenting below.

2.3.3 Non-market work

The flipside of the distribution of market hours is that mothers take on a much larger share of childcare work than men. Figure 2.3.2a shows that before the pandemic, mothers on average spent 29 hours per week providing care to their children. Fathers childcare hours, with units depicted on the right axis, were well below that at 17.5 hours. The location of both lines is normalized so they start at the same point, making differences in their evolution salient. During the period of



(a) Childcare hours

(b) Change in the childcare gap

Figure 2.3.2. Evolution of the childcare gap 2019–2021.

Notes: Figure 2.3.2a shows the development of childcare hours by mothers and fathers in the LISS time use data. Figure 2.3.2b shows the development of differences in childcare provision between fathers and mothers. The latter is based on a regression of childcare hours on the interaction of time dummies and gender, including as additional controls the number of children, and the standardized age of the youngest child interacted with gender. Standard errors clustered on the household level. The regression coefficients underlying the Figure are listed in Column (1) of Table 2.4.1.

closed schools and daycare facilities, combined childcare hours went up by about 25. This number is plausible given typical times spent at school/daycare and the fact that emergency childcare was available for parents working in essential occupations.[6] The large increase in April 2020 was distributed about equally among both genders. For the surveys conducted in November 2020 and November 2021, mothers' childcare hours were back to their pre-pandemic levels; in 2020 they might have been a little lower even. Fathers' hours also declined again, but at a lower rate. As a result they spent about two more hours on childcare duties than before the pandemic.

A different way to look at it is to consider the gap between genders directly. Figure 2.3.2b visualizes the result of this exercise, showing that the gender differences we described in the previous paragraphs are very robust in statistical terms. There virtually was no change in the gender gap in April 2020. Subsequently, the difference shrank by 3–6 hours. When accounting for statistical uncertainty, a range from 1 to 9 hours seems possible.

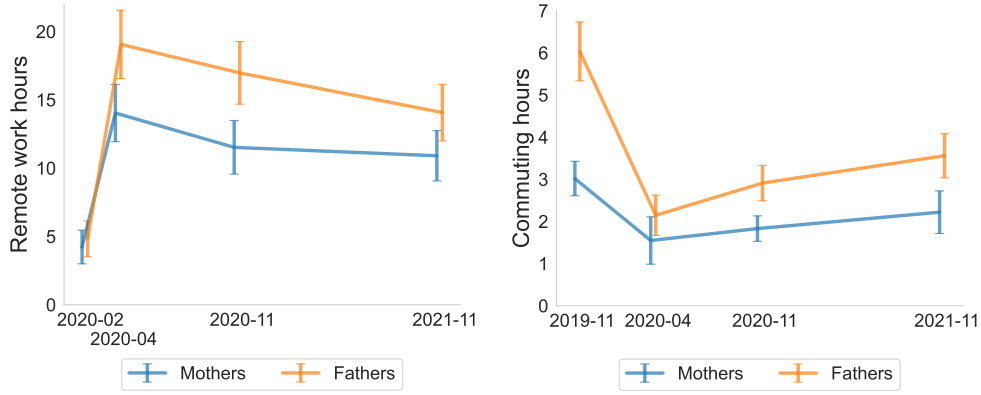
We will argue below that the change in the gender care gap can be largely explained by increased flexibility of parents when it comes to their work schedule

6. Easier access to formal childcare was the most relevant difference for essential workers in March-May 2020 and there were no relevant differences after that time; hence we do not mention them elsewhere. See Zimpelmann et al. (2021) for a more detailed analysis of essential worker status.

and location. Next, we thus describe how remote work and commuting evolved over our period of study.

2.3.4 Remote work and commuting

As early as 2016, the Netherlands introduced a law aimed at facilitating flexible work (Wet flexibel werken). This law defines processes and rights for employees to request adjustments to their working hours, their work schedules, or their work location. Before the CoViD-19 crisis, however, the effects were limited. E.g., ten Hoeve, Talman, van Mierlo, and Engelen (2021) find that 16% of employees made a request regarding flexible work along *any* of the three dimensions between 2016 and mid-March 2020. Consistent with those findings, our data shows that while 34% of individuals reported to have performed *some* work from home (Appendix Table 2.A.10), mean hours worked from home were below five for both genders (Figure 2.3.3a). To put this into perspective, this is an hour less than fathers spent on their weekly commutes on average (Figure 2.3.3b). Mothers' mean commuting time was half of that amount at three hours.



(a) Remote working hours

(b) Commuting hours

Figure 2.3.3. Realized work from home and commuting over time

Notes: Panel a displays average remote working hours in the LISS sample over time and by gender. Figure 2.3.3b displays average commuting hours in the LISS sample over time and by gender. For underlying numbers see Tables 2.A.8 and 2.A.9. Additionally, Table 2.A.10 contains the evolution over time and by gender of a variable measuring *any* remote work and Table 2.A.11 contains the evolution over time and by gender of the share of remote work. In the pre-pandemic period, remote working hours are measured in February 2020 and commuting hours in November 2019.

With the onset of the pandemic, these numbers changed dramatically for all parents. In April 2020, weekly hours worked from home increased to 14 among mothers and 19 among fathers. Put differently, about fifty percent of actual hours were done from home. Commuting time dropped to 1.5 hours for mothers and

2.1 hours for fathers. Even as the pandemic progressed, all these numbers remained closer to the values they took during the initial lockdown than to their prior levels.

Actual remote work in the LISS data is consistent with the corresponding numbers from the much-larger working conditions survey (NEA, see Section 2.2.2). The NEA data also reveal a stark increase in remote work of parents during the pandemic, from approximately 2.8 hours in late 2019 to 10 hours in late 2020. Further, investigating the remote work share by sectors (as a proxy for remote work potential), we find that in the pre-pandemic period, sectors only mildly predict the remote work share of their workers, while in late 2020 the share of hours a worker works remotely strongly depends on the sector he or she works in.[7] This supports our previous point that during the pandemic, remote work potential becomes much more important for its take-up, while take-up is more idiosyncratic before the pandemic.

The large increase in remote work implies a large flexibility gain for parents in the period after the summer of 2020, i.e., when schools and daycare were functioning normally again. In order to quantify the potential flexibility gains when it comes to providing for children, we compute potential remote working hours. We do so by multiplying the remote work share as described in Section 2.2 with working hours just before the pandemic. Figure 2.3.4 shows the distribution this measure of potential remote working hours. Compared to Figure 2.2.1, the differences between genders are substantially larger in relative terms because men work longer hours. In the LISS data, shown in Panel a, more than 30 % of fathers can work at least 30 hours from home, while only 15 % of mothers can do so. The averages are 19.4 and 13 weekly hours, respectively. In the CBS data, we can see a similar pattern but with a less polarized distribution. Mothers have, on average, a remote work capability of around 8 hours, while fathers have an average of around 12.6 hours.

Table 2.3.2 shows that before the pandemic, each potential hour of remote work translated into 12 minutes of actual remote work (this is implied by the coefficient of 0.21 in the second column). In 2020, the coefficient increases to more than 0.8, implying that individuals worked more than 45 minutes remotely for every hour they could potentially do so. In late 2021, when overall remote work was slightly lower and more individuals may have changed jobs, the coefficient drops somewhat but remains high at 0.6 (i.e., 35 minutes for every potential hour).

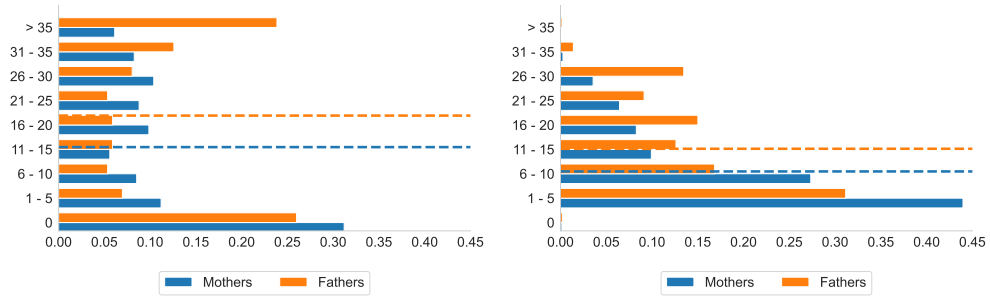
Column (4) of Table 2.3.2 reveals an even stronger pattern for realized commuting time as the dependent variable. Prior to the pandemic, a 40 hour job with the potential to do all tasks at home was associated 1.6 hours *more* time spent commuting compared to a job that would not admit any remote work. After the

7. Details are in Appendix Section 2.A.5.

Table 2.3.2. Predictive power of potential remote working hours for realized hours worked from home and commuting time

	Remote working hours		Commuting hours	
	(1)	(2)	(3)	(4)
Constant	5.68***	1.49*	4.63***	4.06***
	(1.25)	(0.79)	(0.28)	(0.38)
2020-04	12.0***	2.61***	−2.7***	−0.95*
	(0.92)	(0.86)	(0.29)	(0.50)
2020-11	9.74***	−0.16	−2.19***	−0.78**
	(0.86)	(0.68)	(0.22)	(0.34)
2021-11	7.92***	0.87	−1.65***	−0.16
	(0.82)	(0.91)	(0.27)	(0.48)
Pot. hours remote work		0.21***		0.04**
		(0.04)		(0.02)
Pot. hours remote work × 2020-04		0.63***		−0.11***
		(0.06)		(0.02)
Pot. hours remote work × 2020-11		0.61***		−0.09***
		(0.05)		(0.02)
Pot. hours remote work × 2021-11		0.42***		−0.09***
		(0.05)		(0.02)
Observations	1,876	1,876	1,876	1,876
R^2	0.081	0.471	0.069	0.11

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. Sample conditional on working pre-CoVid. Baseline commuting hours based on LISS Time Use Survey from November 2019. Baseline remote work hours obtained from LISS Covid-19 Survey and based on February 2020. Sample restricted on parents who work in November 2019. For the full table see Table 2.B.1 in the Appendix. For the the interaction by gender see Table 2.B.2 in the appendix.



(a) Potential remote working hours

(b) Potential hours of remote work (CBS)

Figure 2.3.4. Remote working potential by gender

Notes: Panel a shows the distribution of the variable potential remote work hours by gender in the LISS sample. Potential remote working hours are calculated by multiplying the share of tasks that can be done from home with the pre-covid (November 2019) working hours of an individual. Dashed lines display the gender-specific means (13 hours for mothers and 19.4 hours for fathers). Samples conditional on working before the pandemic. Panel b shows the distribution of the potential hours of remote work by gender in the CBS in the year 2019. Dashed lines indicate the mean for each gender (8 hours for mothers and 12.6 hours for fathers). For the imputation we calculate the average remote work capability by sector and education in the NEA and impute the remote work capability in the CBS with the help of those two variables. The imputed remote work capability is then multiplied with the working hours two years ago to obtain the potential hours of remote work.

pandemic's onset, the relationship was reversed and commuting time was about 4 hours *less* for a person who works full-time and can do all his tasks from home.

Summing up, we find that remote working hours have strongly increased during the pandemic years. Before the pandemic, take up of remote work was low and rather idiosyncratic. Because of the pandemic, it became intimately tied to job characteristics. The potential hours that can be worked remotely strongly vary across genders. These hours are closely related to increases in actual remote work during the pandemic and to decreases in time spent commuting.

2.4 Results

Our main results establish that the trend towards a more equal division of childcare during the pandemic was driven by the amount of flexibility parents gained as their potential to work remotely was realized. Similarly, we show in Section 2.4.2 that the same households are driving the acceleration of the trend towards mothers working longer hours.

2.4.1 Childcare

We first establish that the potential to work remotely was negatively related to hours spent on childcare before the pandemic and that this relationship flipped

rather dramatically after its onset. Beginning in early 2020, the potential to work remotely is closely associated with more time spent on childcare. We then show that remote work potential largely explains the decrease in the childcare gap between mothers and fathers, established in Section 2.3.3.

Table 2.4.1 brings together the changes in the childcare gap between mothers and fathers and the shift to remote work. Column (1) repeats the numbers underlying Figure 2.3.2b, which plotted the coefficients on the indicator variables for mother by time period during the pandemic.[8] The absolute difference in childcare provision between parents did not change in April 2020, when childcare facilities and schools were closed, because both parents increased their childcare hours in similar magnitudes. There is a sharp decline (six hours) in the gender gap in childcare in November 2020, which materializes by means of a decrease in childcare provision by mothers and an increase among fathers. The shrinking of the gap carries over to November 2021 but only at half the size.

Our key specification is column (2), which adds the potential hours of remote work. This yields a difference-in-differences design with a continuous treatment variable. The basic assumption is that in the absence of the pandemic, childcare hours would have evolved independently from remote work ability. While this assumption might be too strong, we would likely err in a direction that attenuates our coefficients of interest. In particular, before the pandemic, potential hours of remote work are negatively related with childcare hours. This is partly driven by the somewhat mechanical fact that total working hours are higher among parents with a large number of potential remote working hours. Additionally, jobs with high remote work potential tend to yield relatively high earnings, so—to the extent that income effects dominate—c.p., fathers are more likely to work longer hours and mothers are more likely to return to work early and work longer hours.

In April 2020, during the first lockdown in which childcare facilities and schools were closed, the relation between potential remote hours and childcare hours turns strongly positive. On net, one hour of potential remote work translates into 22 minutes of childcare. In November 2020, when childcare facilities and schools were open again, the relationship becomes somewhat weaker but stays significantly positive. The net effect is still slightly positive in November 2021. From 2020 on, mothers and fathers who can work more hours from home strongly increase their childcare hours compared to the time before the pandemic.

We standardize the potential hours of remote work so that we can compare the evolution of the gender care gap across specification. The coefficients on the

8. The sample includes non-working parents whose potential hours of remote work are set to zero. We report results only including parents who worked before the pandemic in Appendix Table 2.B.6, results do not change. We prefer the sample in Table 2.4.1 because when we condition on a parent working before the pandemic, we disproportionately drop mothers, leaving fathers in single-earner households in the sample. Conceptually, we prefer to avoid this imbalance.

Table 2.4.1. Evolution of the gender care gap and potential hours of remote work

	Hrs childcare		
	(1)	(2)	(3)
Constant	17.28***	18.05***	17.75***
	(1.41)	(1.43)	(1.44)
2020-04	12.8***	10.49***	10.49***
	(1.53)	(1.46)	(1.46)
2020-11	2.96**	1.20	1.69
	(1.29)	(1.29)	(1.26)
2021-11	1.55	0.31	0.36
	(1.19)	(1.20)	(1.23)
Mother	14.01***	12.44***	12.67***
	(2.13)	(2.13)	(2.14)
Mother × 2020-04	−0.09	3.85**	3.74*
	(1.99)	(1.94)	(1.94)
Mother × 2020-11	−6.15***	−3.29*	−3.36*
	(1.84)	(1.87)	(1.86)
Mother × 2021-11	−3.31**	−1.46	−1.34
	(1.60)	(1.63)	(1.62)
Pot. hours remote work (std)		−0.16***	−0.1**
		(0.05)	(0.05)
Pot. hours remote work (std) × 2020-04		0.52***	0.55***
		(0.07)	(0.09)
Pot. hours remote work (std) × 2020-11		0.36***	0.26***
		(0.07)	(0.08)
Pot. hours remote work (std) × 2021-11		0.23***	0.21***
		(0.06)	(0.07)
Pot. hours remote work (std) × Mother			−0.16
			(0.10)
Pot. hours remote work (std) × Mother × 2020-04			−0.05
			(0.14)
Pot. hours remote work (std) × Mother × 2020-11			0.24*
			(0.13)
Pot. hours remote work (std) × Mother × 2021-11			0.04
			(0.13)
Observations	2,234	2,234	2,234
R^2	0.324	0.347	0.349

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. The potential hours of remote work are demeaned to facilitate comparison of coefficients across columns. Observations are weighted to keep the composition with respect to the age of the youngest child constant over time. All specifications control for the (demeaned) age of the youngest child interacted with gender, as well as indicator variables indicating number of children, the left-out category is a single child. In column (3), we additionally interact the number of children with gender, so that the model is fully satiated. Potential remote work hours are set to zero if the individual did not work before the pandemic. The full set of coefficients can be found in Appendix Table 2.B.5. Appendix Table 2.B.6 shows results for the same specifications restricting the sample to individuals who were working before the pandemic.

mother by time period interactions measure the gender care gap, evaluated at the sample mean of potential hours of remote work.[9] Including the standardized potential hours of remote work in the regression diminishes the changes in the gender care gap in November 2020 and November 2021 by cutting coefficients in half, rendering them statistically indistinguishable from zero or marginally so. This indicates that the changes in the gender care gap can be largely explained by the shift to remote work which made, in particular, fathers more available in many families.

Column (3) additionally includes an interaction between standardized potential hours of remote work and the mother dummy. This does not change the previous results. Further, it shows that mothers tend to be more inclined to use their potential hours of remote work for childcare. The effect is, however, only statistically significant for November 2020.

2.4.2 Labor Supply

The result that remote work induced a decrease in the childcare gap gives rise to the question whether these changes also translate to effects on labor supply. In particular, mothers whose partners are now taking over a larger share of childcare duties might be willing to increase the time spent on market work. In Section 2.3.2, we saw that the trend of increasing full-time work of mothers accelerated over the 2020–2021 period. In this section, we analyze whether partners' remote work induces individuals to work more and to which extent this effect operates through a direct effect of increased remote working hours and to which extent through reduced commuting.

The mechanisms at play are more subtle and likely to operate with some time lag. In April 2020, there was an immediate need for childcare and parental involvement had to be adjusted instantly. In contrast, changing one's (paid) hours of work requires at least some preparation and potentially negotiations with the employer as well as within the household. Hence, we would expect changes in working hours to lag behind changes in childcare hours. Because effects are rather small, we cannot expect to find much in the LISS data. Hence, we stick to the CBS data only.

To estimate the effects of remote work on labor supply, we apply a similar Diff-in-Diff empirical strategy like in the previous section. Treatment intensity, i.e., remote work potential, is measured in each year based on the job two years prior to the respective year. This allows us to include subjects who are not working in a given year, as long as they have worked two years before. This sample selection is the same for each year and will, hence, not drive our results.

9. Defined as potential hrs remote work (std) $=$ (potential hrs remote work $-\mu$) with the sample mean $\mu = 12.2$.

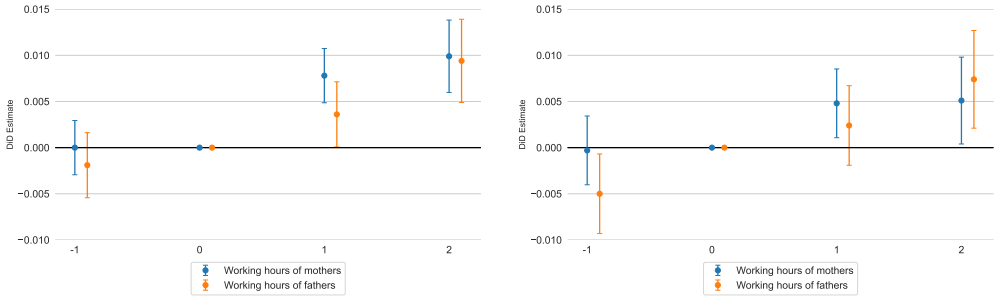
There are three important differences, however. First, we focus on the partners' flexibility gains for obvious endogeneity concerns when it comes to own remote work potential.[10] This is also the effect that is relevant for the mechanism above: When my partner gains flexibility through additional remote work, he may take over more childcare, freeing up potential for me to work more. However, all regressions still control for own remote work potential. Second, we run all regressions separately for mothers and fathers; this is just to display the results side-by-side rather than in a stacked way (the model in the previous section was nearly saturated in terms of gender interactions, too). Third and most importantly, we add an event study design on top of the specification because of the time trends in female labor force behavior seen in Section 2.3.2. That is, one might be worried that these trends are related to working from home potential, e.g., because well-educated office workers might have more progressive gender norms than workers in blue-collar occupations.

We therefore calculate the difference in working hours over time between parents with high partners' potential remote work hours, and those with lower partners' potential remote work hours, for individuals in the time period 2018-2021, when the Covid shock materialized, and for individuals in pre-Covid time periods, when no Covid shock took place.[11] We then take these separate Diff-in-Diff estimates and analyze how the difference between the estimates evolves over time. This ensures that we account for general trends in labor force participation related to the potential to work from home and only identify the effects that materialized through the actual realization of remote work during the Pandemic.

We present our results graphically in Figure 2.4.1 and display the underlying coefficients in Table 2.4.2. We plot the coefficients which show the effect of partners' remote work potential on labor supply during the pandemic period relative to prior 4-year periods. For the pandemic period, t=0 refers to November 2019. Focusing first on the results without accounting for commuting gains in Figure 2.4.1a, we see that there is a clear uptick relative to prior periods for mothers whose partner has a large potential for remote work. The coefficient of around 0.01 means that when comparing groups of women whose partners are in the lowest category of work from home potential identified in the CBS data (0 to 5 hours) and in the highest category (30 to 35 hours), three in 10 women work one more hour per week. Interestingly, we see similar effects for fathers. However, as the potential to work from home is asymmetrically distributed across genders, fathers will be less affected as their partners have relatively low gains in work from home (see Figure 2.3.4b).

10. The endogeneity concern arises as the own remote work potential is highly correlated with own working hours.

11. We calculate the effects relative to three pre-Covid time periods, namely 2014-2017, 2015-2018, 2016-2019.



(a) Effects of Partner Remote Work

(b) Effects of Partner Remote Work, controlling for Commuting Gains

Figure 2.4.1. Effects of Partner Remote Work

Notes: The figure displays the event-study DiD estimates for the effect of the potential hours of remote work of the partner on own working hours relative to the year of the Covid/Placebo shock. Results are reported separately for mothers and fathers. All regressions also control for own potential hours of remote work. The regressions results can be found in Table 2.4.2, Panel A illustrates the estimates from columns (1) and (3), while Panel B shows the effects when additionally controlling for gains in commuting in columns (2) and (4). Standard errors are obtained by clustering on the individual level.

Part of the effects are driven by gains from commuting. As shown in Figure 2.4.1b, the effect reduces by about 50 % for mothers when we additionally control for potential commuting gains, defined as potential remote work share two years prior times the commuting distance two years prior.[12] In column (3) of Table 2.4.2, we also see that the potential commuting gain of the partner is positively related to increases in working hours compared to pre-pandemic times. For fathers, the estimates do not change when additionally controlling for the partner's potential gains in commuting.

12. As we have only access to the commuting distance from the year 2014 onwards, we take the commuting distance in the year prior for observations in 2015 and the commuting distance from the year itself for the observations in the year 2014. Importantly, our results remain unchanged if we restrict ourselves to only using the periods from 2016 onwards, for which we can calculate the commuting gain measure using the information from the job two years prior.

Table 2.4.2. The effect of potential remote working hours on working hours

	Mothers		Fathers	
	(1)	(2)	(3)	(4)
Part: Pot hrs wfh × t = -1 × Pand	0.00	−0.0	−0.002	−0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Pot hrs wfh × t = 1 × Pand	0.008***	0.005**	0.004**	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Pot hrs wfh × t = 2 × Pand	0.01***	0.005**	0.009***	0.007***
	(0.002)	(0.002)	(0.002)	(0.003)
Part: Pot comm gain × t = -1 × Pand		0.00		0.004***
		(0.001)		(0.001)
Part: Pot comm gain × t = 1 × Pand		0.003***		0.00
		(0.001)		(0.001)
Part: Pot comm gain × t = 2 × Pand		0.004**		0.001
		(0.001)		(0.002)
R^2	0.175	0.193	0.015	0.018

Notes: This table reports a subset of coefficients of the event study Diff-in-Diff regression. The dependent variable is unconditional working hours, i.e., the variable is zero if the individual does not work. The event study is run on the period from 2014 to 2021 on sets of four years (i.e., 2014 to 2017, 2015 to 2018, 2016 to 2019, and 2018 to 2021). Only for the last set of years, the dummy 'Pand' is set to one. As before, we use data from November in each year. We control for the number of children, age, and age of the partner. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the individual level. Full list of coefficients in Table 2.B.9.

Importantly, our results are very robust to alternative specifications. In Table 2.B.7, we use the remote work share instead of potential remote work hours as treatment variable. Results are very similar for mothers while we find a marginally significant pre-trend for fathers. Moreover, imputing remote work share based on sector times education times gender instead of only sector times education does not alter any of our main results.

2.5 Conclusion

We have investigated how the acceleration in the shift towards remote work during the CoViD-19 pandemic has impacted the division of childcare duties and working hours. The way the pandemic has been handled in the Netherlands—the most important feature being relative short school and daycare closures—has allowed us to isolate this effect. Our analysis has relied on self-collected survey data and population-wide administrative data.

We find that the average gap between mothers' and fathers' childcare provision shrinks by 3.3 hours or 24 % in the period from November 2019 to November 2021. Most of this decline can be attributed to households where the remote work potential of the father was high. The partner's potential remote work also helps to explain the trend towards mothers working longer hours, which accelerated during the pandemic.

Our results show that remote work can help many households to find a division of labor that is more equal across genders. It is likely that more working from home will remain very common in the future, so employers will be less able to condition wages and career progression on it than they were before the pandemic. This also means that a convenient excuse for some parents, in particular fathers, for not being available for childcare duties is gone on some days.

In other institutional environments, the effects we found might take longer to materialize. The infrastructure for remote work and childcare is well-developed and reliable in the Netherlands. Mothers had a high labor force participation rate—albeit with low hours—already before the pandemic, while fathers' weekly hours were low in international comparison (Bick, Brüggemann, and Fuchs-Schündeln, 2019). Finally, of course, in many countries the pandemic had a differentially larger direct effect on the labor market outcomes of women (Alon, Coskun, Doepke, Koll, and Tertilt, 2022).

Overall, our results have shown that working from home might have a bright side in bringing about more gender equality within households.

References

Adams-Prassl, Abi, Teodora Boneva, Marta Golin, and Christopher Rauh. 2022. "Work That Can Be Done from Home: Evidence on Variation within and across Occupations and Industries." *Labour Economics* 74: 102083. [99, 100]

Aksoy, Cevat Giray, Jose Maria Barrero, Nicholas Bloom, Steven J. Davis, Mathias Dolls, and Pablo Zarate. 2023. "Time Savings When Working from Home." *AEA Papers and Proceedings* 113: 597–603. [95]

Alon, Titan, Sena Coskun, Matthias Doepke, David Koll, and Michèle Tertilt. 2022. "From Mancession to Shecession: Women's Employment in Regular and Pandemic Recessions." *NBER Macroeconomics Annual* 36 (5): 83–151. [116]

Bick, Alexander, Adam Blandin, and Richard Rogerson. 2022. "Hours and Wages." *Quarterly Journal of Economics* 137 (3): 1901–62. [95]

Bick, Alexander, Bettina Brüggemann, and Nicola Fuchs-Schündeln. 2019. "Hours Worked in Europe and the United States: New Data, New Answers." *Scandinavian Journal of Economics* 121 (4): 1381–416. [116]

Dingel, Jonathan I., and Brent Neiman. 2020. "How Many Jobs Can Be Done at Home?" *Journal of Public Economics* 189 (9): 104235. [100]

Gicheva, Dora. 2013. "Working Long Hours and Early Career Outcomes in the High-End Labor Market." *Journal of Labor Economics* 31 (4): 785–824. [95]

Heggeness, Misty, and Palak Suri. 2021. "Telework, Childcare, and Mothers' Labor Supply." Preprint. Institute Working Paper. [103]

ten Hoeve, Youri, Jildou Talman, Jorrit van Mierlo, and Mirjam Engelen. 2021. "Evaluatie Wet Flexibel Werken." Onderzoek in Opdracht van Het Ministerie van Sociale Zaken En Werkgelegenheid. [106]

von Gaudecker, Hans-Martin, Christian Zimpelmann, Moritz Mendel, Bettina Siflinger, Lena Janys, Jürgen Maurer, Egbert Jongen, Radost Holler, Renata Abikeyeva, Felipe Augusto Azuero Mutis, Annica Gehlen, and Eva Lucia Kleifgen. 2021. "CoVID-19 Impact Lab Questionnaire Documentation." [98]

Zimpelmann, Christian, Hans-Martin von Gaudecker, Radost Holler, Lena Janys, and Bettina Siflinger. 2021. "Hours and Income Dynamics during the Covid-19 Pandemic: The Case of the Netherlands." *Labour Economics* 73 (12): 102055. [98, 101, 102, 105]

Appendix 2.A Details on data sets and descriptives

2.A.1 Time-use data LISS

The time-use data is usually collected every other year in the LISS panel, but the design of the questionnaire changed substantially in November 2019, so that earlier waves cannot be compared to post-2019 waves. During the CoVid-19 pandemic, two additional waves were added. In the study, we employ time use information collected in comparable questionnaires in November 2019, April 2020, November 2020, and November 2021. In these surveys, people are asked to distribute the hours of the past week over different activities. Table 2.A.1 gives an overview of the categories that have been asked for in the different questionnaires. It shows that some categories have been split up or aggregated over time, e.g. beginning with the April 2020 wave, working hours are recorded separately by whether work was done at the usual workplace or at home (and whether the subject has been responsible for a child while working at home). We use the information on time spent working, commuting, and on childcare and aggregate the respective categories if necessary. See van Soest, Been, Pinger, von Gaudecker, and Centerdata (2019), von Gaudecker and Centerdata (2020a), von Gaudecker and Centerdata (2020b), and Been and Centerdata (2021), respectively, for the documentation of the four questionnaires.

Before the anlysis, we dropped observations containing extreme outliers which indicate that the questionnaire was not filled out properly. These are observations that allocated more than 120 hours two a single activity with the exception of resting and childcare. We did not put any restriction on hours spent resting because, individuals can, in principle, spent all their hours resting, for instance, when they are sick. We use a higher threshold (140 hours) for hours spent doing childcare as it is also, in principle, possible that an individual spends all their non-resting hours doing childcare, for instance, if they have a newborn or their child is sick. In addition to dropping extreme outliers, we winsorized the variables to avoid

Table 2.A.1. Time use categories in the LISS time use questionnaires

	Nov 2019	Apr 2020	Nov 2020	Nov 2021
Commuting	x	x	x	x
Education	x	x	x	x
Job search				x
Paid work	x			
Paid work at workplace		x	x	x
Paid work at home		x	x	x
Paid work at home while responsible for children		x	x	x
Homeschooling		x	x	x
(Other) childcare	x	x	x	x
Care for parents	x			
Care for other relatives	x			
Care for family members	x			
Care for parents or (non-)family members		x	x	x
Cooking				x
Shoping				x
(Other) chores	x	x	x	x
Leisure	x			x
Leisure with others online		x	x	
Leisure with others in person		x	x	
Leisure alone		x	x	
Resting	x	x	x	x
Other	x	x	x	x

Notes: The table displays the time use categories that are available in the LISS time use questionnaires. The questionnaires in which the respective category was part of are marked with an 'x'.

that our analysis becomes outlier driven. In particular, we winsorize hours spent working, hours spent working at the workplace, hours spent working from home, hours spent working from home while taking care of children, hours spent on job search, hours spent in education, hours spent doing chores, shopping, cooking, and searching for a job at 80 hours. Hours spent commuting is winsorized at 40 hours. Hours helping others are winsorized at 60. Leisure activities and other activities are winsorized at a 100 hours. Thresholds are defined to capture the different distributions across the variables as well as inducing stability across waves. After winsorizing, we rescaled all variables per wave per individual such that the relative share of hours of each activity after winsorizing stayed the same and the hours still sum up to 168.

2.A.2 COVID-19 modules LISS

We fielded six questionnaires on the impact that the CoViD-19 pandemic had on peoples' lives in the period between mid-March and December 2020 (the questionnaires are documented in von Gaudecker, Zimpelmann, Mendel, Siflinger, Janys, et al., 2021). From those surveys, we employ two variables. We will use the working hours at the point in time before the pandemic started affecting working lives. Most importantly, we measure remote work potential. In May 2020, we ask participants "What percentage of your normal work *prior to the coronavirus outbreak* can you do while working from home?". We repeated this question in December 2020, but inquired about the share of tasks at the current job that can be done from home instead of the pre-pandemic situation.[13]

The resulting answers measure the remote work potential, abstracting from any changes in task content that happened during the period of social distancing. The fact that we ask this when the pandemic was already in full swing allows individuals to better assess the *potential* for remote work – it would not have occurred to many people that essentially all meetings could be held in virtual formats. Put differently, we assume that the potential to work remotely was utilized to (almost) full extent only when the government strongly advised doing so as a component of social distancing. The correlation between the May and December measures is 0.82.[14] Reassured by the measures' high degree of stability, we take the mean of the data that is available at the individual level (participation is not identical across waves).

2.A.3 Description of variables

This section provides further information about the source and calculation of variables in the LISS sample (used for all time-use analyses) and in the full population sample.

LISS sample.

Age, gender Obtained from the background questionnaire.

Education Obtained from the administrative records. Based on achieved educational level. The Dutch educational levels are categorized as follows:

Lower secondary and below: primary school, vmbo

Upper secondary: mbo, havo, vwo

Tertiary: hbo, wo

13. The question in December 2020 reads: "What percentage of your normal work can you do while working from home?".

14. The Appendix of Zimpelmann, Gaudecker, Holler, Janys, and Siflinger (2021) discusses the correlation between the answers in May and December as well as the marginal distributions in great detail.

Age of the youngest child Obtained from Covid questionnaire. If possible, we take the information on the age of the youngest child from the administrative records.

Remote work share Obtained from Covid questionnaire. Variable creation is described in detail in Section 2.2.1.

Potential hours remote work Obtained from Covid questionnaire and time use questionnaire. Variable creation is described in detail in Section 2.3.4. If possible, we take the information on working hours from the administrative records.

Childcare hours Obtained from time use questionnaire. In November 2020, the time use questionnaire only contains one question that asks for hours spent doing childcare per week. In April and November 2020, childcare is spread across three questions: hours spent doing homeschooling, hours spent performing childcare while working remotely and hours spent doing other childcare. In November 2022, only the ladder two are asked. For each year, we take the sum of hours across the different childcare related activities.

Commuting hours Obtained from the time use questionnaires

Hours worked from home Values taken from the CoVid-19 questionnaire to measure pre-pandemic hours worked from home. In March 2020, we asked individuals how many hours they worked from home in February 2020. Hours worked from home in November 2020 and November 2021 are taken from the time use questionnaire by adding the hours worked from home with and without children.

Full population sample.

Age, gender Obtained from the administrative records.

Education Obtained from the administrative records. Based on achieved educational level. The Dutch educational levels are categorized as follows:

Lower secondary and below: primary school, vmbo

Upper secondary: mbo, havo, vwo

Tertiary: hbo, wo

Age of the youngest child Obtained from the administrative records.

Remote work share Obtained from administrative records. Imputed with the help of the NEA working conditions survey. Imputation is described in detail in Section 2.A.5.

Potential hours remote work Obtained from administrative records. Variable creation is described in detail in Section 2.4.2.

Commuting distance Obtained from administrative records. Measures the distance from the home to the workplace.

Working hours Obtained from administrative records. Variable creation is described in detail in Section 2.2.2.

Table 2.A.2. Socio-demographic variables by data source and gender in 2019

	LISS		CBS	
	Fathers	Mothers	Fathers	Mothers
Age	42.63	40.55	40.95	38.54
	(6.26)	(6.18)	(6.77)	(6.51)
Age youngest child	6.94	7.17	6.14	6.11
	(4.46)	(4.47)	(4.72)	(4.72)
Number of children	2.1	2.05	1.94	1.94
	(0.8)	(0.77)	(0.75)	(0.75)
Education: High	0.48	0.45	0.41	0.46
	(0.5)	(0.5)	(0.49)	(0.5)
Education: Middle	0.25	0.25	0.3	0.3
	(0.43)	(0.44)	(0.46)	(0.46)
Education: Low	0.02	0.03	0.07	0.06
	(0.15)	(0.17)	(0.25)	(0.23)
Education: Unknown	0.18	0.19	0.23	0.18
	(0.39)	(0.39)	(0.42)	(0.38)
Observations	229	260	610,577	622,972

Notes: The first column displays basic demographic characteristics of the LISS sample by gender in November 2019. The age variable is taken directly from the LISS survey. The values for the variables age of youngest child and number of children are taken from the administrative records for all linked individuals and from the LISS survey for all those who are not linked. The education variable is taken from the administrative records and therefore only available for linked individuals (note that even for linked individuals it is possible that the education is unknown). The second column displays basic demographic characteristics of all working-age (18-55 years old) parents with a child below 16 years old who were employed some time in 2018 and 2019 by gender in November 2019. The education variable is unknown if there is no available administrative record on the education for the individual. See Table 2.A.4 for the numbers over time.

2.A.4 Descriptive statistics

Table 2.A.2 displays the socio-demographic characteristics for the LISS-Sample which we use for the childcare analysis and for the CBS-Sample which we use for the labor supply analysis. For both samples we display the socio-demographic characteristics in November 2019. The table reveals that most socio-demographic statistics line up well between the LISS sample and the population data. Mothers are somewhat younger than fathers, families comprise around 2 children on average and the age of the youngest child falls just below the middle of the age interval we require.

The one exception is that respondents in the LISS panel are better educated. In particular, 3 % of parents do not have a secondary degree. This compares to 10 % in the CBS data and it is a well-known bias in surveys. The composition of our LISS sample changes somewhat over time. In particular, the average age of the youngest child is lower for mothers who respond in 2021 compared to 2019 and April 2020 (6.3 years vs 7.2 years, see Table 2.A.4). As this could affect the analysis of childcare hours, we re-weight our sample in terms of age of the youngest child. The waves in 2020 and 2021 are re-weighted such that they match the composition of the sample in 2019, which results in a stable composition over time.

2.A.5 Imputation of remote work potential in the administrative data

For the imputation of the remote work capability in the administrative records, we make use of the National Working Condition Survey (NEA). It is currently available until 2020, i.e., the wave of 2021 is not yet published. Its goal is to gather information on the topics of working conditions, occupational accidents, work content, employment relationships and employment conditions of employees. The NEA is carried out yearly since 2005 by Statistics Netherlands and TNO, in collaboration with the Ministry of Social Affairs and Employment. Its target population are all employees aged 15 to 74 who work in the Netherlands, from whom a sample is surveyed during the period of 1st of October to 31st of December of each year.[15]

Around 50,000 individuals answer the survey each year and around 30,000 to 35,000 of those respondents answered the questions on remote work, which we use for our imputation. In particular, we use the following variables for calculating a remote work share:

- Remote Work Hours (Afl_AantUurTW): “On average, how many hours a week do you work from home for your employer?”
- Remote Work Dummy (Afl_Telewerk): “Teleworker (works at least half a day a week outside the company location with access to the company’s IT system)”
- Working Hours (Afl_Uren): “Working hours in hours per week in current job”

We calculate a remote work share for each individual by dividing the remote work hours by total working hours. For individuals for whom we do not observe information on the remote work hours, but for whom we observe the remote work dummy being 0, we impute a remote work share of 0.

Figure 2.A.1 displays the distribution of the remote work share by gender in the NEA in the year 2020. Dashed vertical lines indicate the mean for each

15. The documentation of the survey and all questionnaires are available at https://www.cbs.nl/nl-nl/onze-diensten/methoden/onderzoeksomschrijvingen/korte-onderzoeksbeschrijvingen/nationale-enquete-arbeidsomstandigheden--nea--.

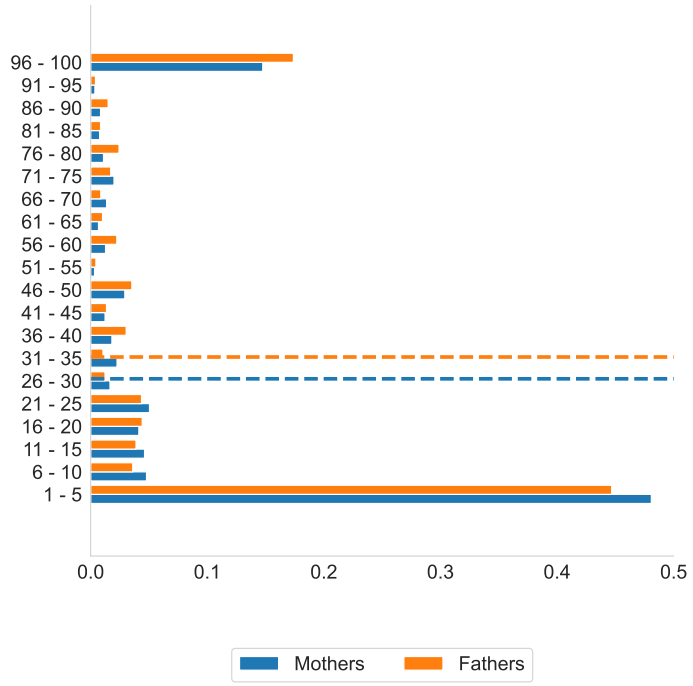


Figure 2.A.1. Remote work capability (NEA)

Notes: Figure 2.A.1 displays the distribution of remote work capability by gender in the National Working Conditions Survey (NEA) in the year 2020. Dashed vertical lines indicate the mean for each gender. Remote work capability is calculated by dividing the hours of remote work by total working hours. The figure shows that the remote work capability in the NEA exhibits a similar distribution like the share of tasks that can be done from home variable in the LISS Sample. The distribution is bi-modal and men work in jobs with an, on average, higher remote work capability than women.

gender. The figure shows that the remote work share in the NEA exhibits a similar distribution like the share of tasks that can be done from home variable in the LISS Sample (see Figure 2.2.1a). The distribution is bi-modal and men have, on average, a higher remote work share than women.

Figure 2.A.2 displays the mean remote work share aggregated by sector over time. Before the pandemic, remote work shares were on a low level and differences between sectors were much smaller than in the year 2020.

To be able to impute the remote work capability for each individual in the administrative records, we have to find predictive characteristics along which we can make the imputation. Table 2.A.3 displays the regression results from regressing the remote work share in 2020 on education, gender and sector. The table shows that education and sector are highly predictive for the remote work share, while gender is not predictive. We therefore perform the imputation with the help of education and sector.

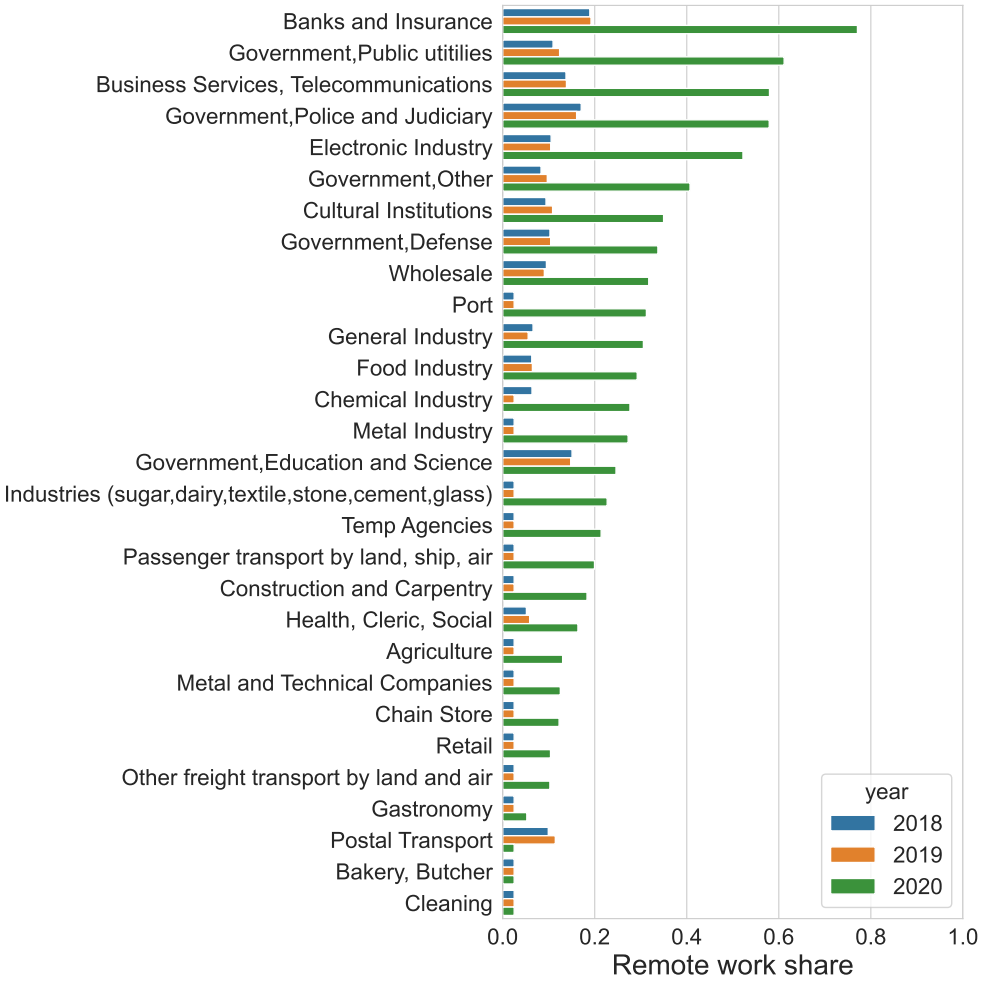


Figure 2.A.2. Share of remote work by sector over time

Notes: This figure displays the mean remote work share aggregated by sector for the years 2018, 2019, and 2020. Note that shares below 5 % are anonymised by the export procedure of CBS Netherlands. For this figure, we set those values to 2.5 %. The data is based on the National Working Conditions Survey (NEA) in the respective year.

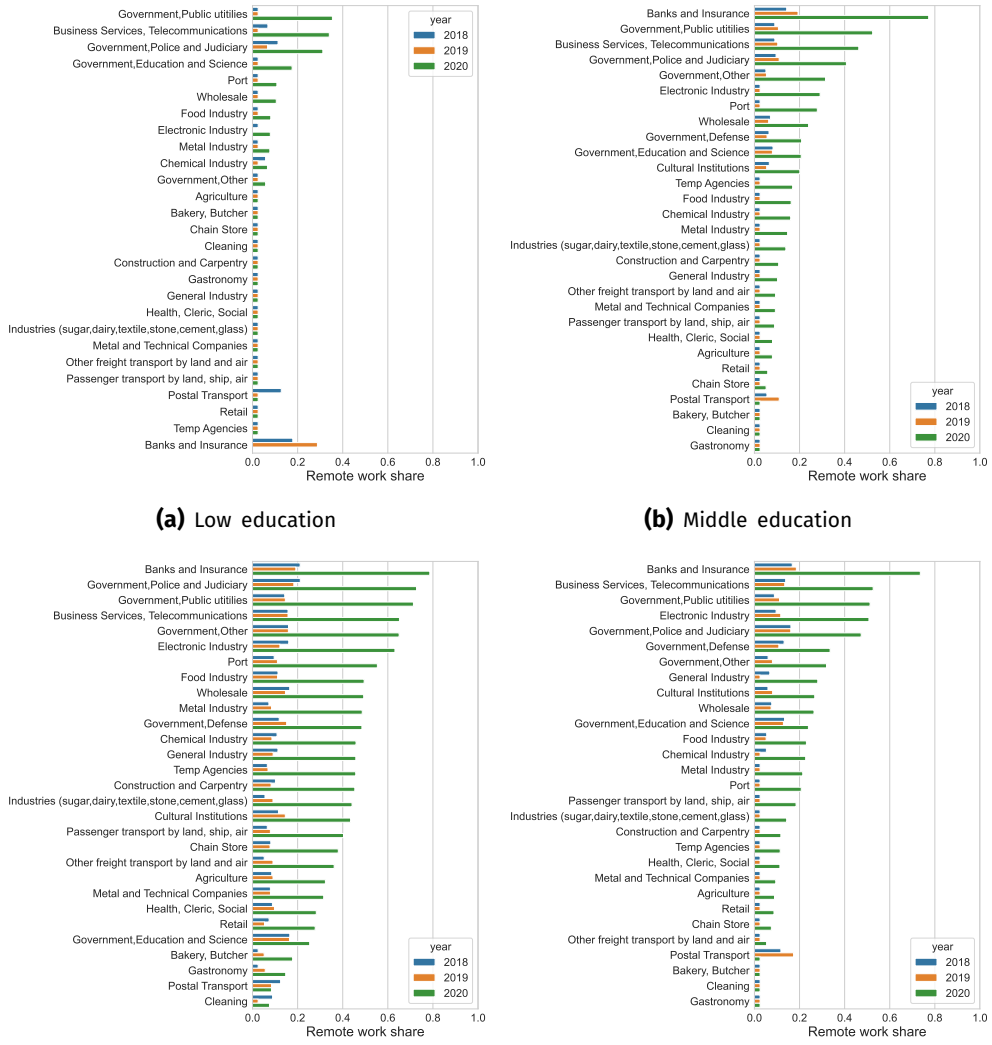


Figure 2.A.3. Share of remote work by education and sector over time

Notes: This figure displays the mean remote work share aggregated by sector for the years 2018, 2019, and 2020 for four educational categories. Note that shares below 5 % are anonymised by the export procedure of CBS Netherlands. For this figure, we set those values to 2.5 %. The data is based on the National Working Conditions Survey (NEA) in the respective year.

Table 2.A.3. Determinants remote work share

	2018	2019	2020
Constant	0.18***	0.19***	0.44***
	(0.0048)	(0.0046)	(0.0086)
Education: Low	−0.079***	−0.083***	−0.29***
	(0.0042)	(0.004)	(0.0079)
Education: Middle	−0.067***	−0.062***	−0.19***
	(0.0027)	(0.0024)	(0.0046)
Education: Unknown	−0.045***	−0.042***	−0.16***
	(0.0024)	(0.0022)	(0.0043)
Father	0.018***	0.017***	−0.0044
	(0.0021)	(0.002)	(0.0038)
Agriculture	−0.13***	−0.12***	−0.15***
	(0.0096)	(0.0089)	(0.018)
Bakery, Butcher	−0.13***	−0.14***	−0.21***
	(0.013)	(0.012)	(0.024)
Banks and Insurance	0.014*	0.0085	0.41***
	(0.0078)	(0.0075)	(0.014)
Business Services	−0.035***	−0.041***	0.23***
	(0.005)	(0.0047)	(0.0088)
Chain Store	−0.12***	−0.12***	−0.16***
	(0.0081)	(0.0077)	(0.015)
Chemical Industry	−0.097***	−0.13***	−0.029
	(0.01)	(0.0098)	(0.019)
Cleaning	−0.12***	−0.13***	−0.21***
	(0.0097)	(0.0096)	(0.019)
Construction and Carpentry	−0.12***	−0.13***	−0.1***
	(0.007)	(0.0066)	(0.013)
Cultural Institutions	−0.076***	−0.066***	−0.0071
	(0.013)	(0.013)	(0.023)
Electronic Industry	−0.064***	−0.079***	0.18***
	(0.012)	(0.01)	(0.021)
Food Industry	−0.099***	−0.1***	−0.014
	(0.011)	(0.01)	(0.02)
Gastronomy	−0.12***	−0.12***	−0.22***
	(0.0084)	(0.0077)	(0.016)
General Industry	−0.097***	−0.12***	−0.019
	(0.0084)	(0.0082)	(0.015)
Government, Defense	−0.059***	−0.068***	0.028
	(0.013)	(0.012)	(0.023)
Government, Education	−0.027***	−0.037***	−0.14***
	(0.0049)	(0.0046)	(0.0087)
Government, Other	−0.074***	−0.073***	0.099***
	(0.0075)	(0.0077)	(0.014)
Government, Police	0.0024	−0.016**	0.24***
	(0.0074)	(0.0068)	(0.013)
Government, Public utitilies	−0.058***	−0.053***	0.27***
	(0.0068)	(0.0064)	(0.012)
Health, Cleric, Social	−0.1***	−0.1***	−0.15***
	(0.0049)	(0.0047)	(0.0089)
Industries (sugar, dairy, textile, stone, cement, glass)	−0.12***	−0.12***	−0.076***
	(0.011)	(0.01)	(0.021)
Metal Industry	−0.12***	−0.12***	−0.03**
	(0.0075)	(0.0075)	(0.014)
Metal and technical companies	−0.12***	−0.13***	−0.15***
	(0.006)	(0.0058)	(0.011)
Other freight transport	−0.13***	−0.13***	−0.16***
	(0.0078)	(0.0078)	(0.015)
Passenger transport	−0.12***	−0.13***	−0.093***
	(0.0086)	(0.0083)	(0.016)
Port	−0.1***	−0.12***	0.022
	(0.01)	(0.0092)	(0.021)
Postal Transport	−0.045***	−0.04***	−0.23***
	(0.014)	(0.015)	(0.029)
Retail	−0.12***	−0.12***	−0.17***
	(0.0069)	(0.0065)	(0.013)
Temp Agencies	−0.12***	−0.12***	−0.068***
	(0.0094)	(0.0091)	(0.014)
Wholesale	−0.063***	−0.077***	0.016
	(0.0058)	(0.0055)	(0.01)
N children	0.0036***	0.0031***	−0.0043***
	(0.0009)	(0.0008)	(0.0016)
Observations	31000	38000	37000
R^2	0.12	0.11	0.27

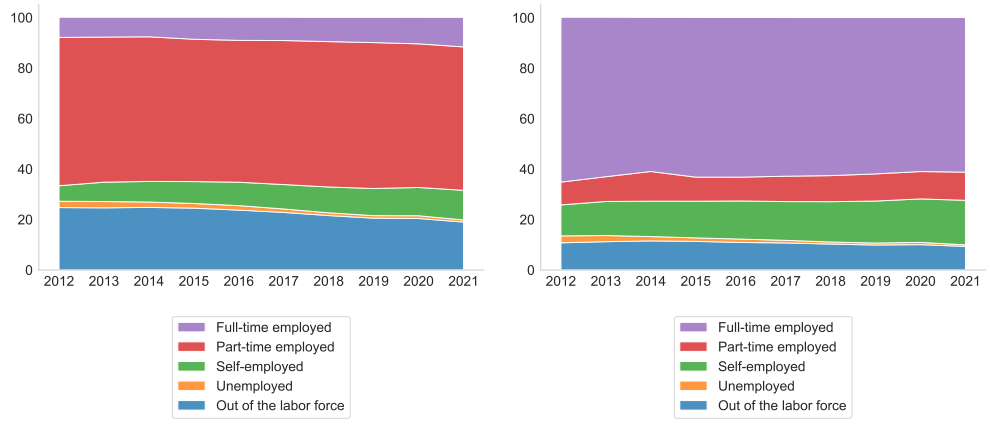
Notes: The table displays the regression results from regressing the remote work share in 2018, 2019 and 2020 on education, gender and sector. The population are all individuals in the NEA sample for whom we have information on actual remote work. The table shows that education and sector are highly predictive for the remote work share, while gender is not predictive. We therefore use education and sector for our imputation.

Table 2.A.4. Basic demographics for LISS Sample by gender over time

		Age	Age youngest child	Number of children	Education: High	Education: Middle	Education: Low	Education: Unknown	Observations
Mothers	2019-11	40.55	7.17	2.05	0.45	0.25	0.03	0.19	260
		(0.38)	(0.28)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	
	2020-04	40.89	7.12	2.07	0.44	0.26	0.04	0.17	280
		(0.36)	(0.27)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	
	2020-11	40.37	7.08	2.04	0.44	0.25	0.04	0.18	339
		(0.35)	(0.24)	(0.04)	(0.03)	(0.02)	(0.01)	(0.02)	
	2021-11	40.59	7.02	2.06	0.53	0.30	0.02	0.15	311
		(0.35)	(0.25)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	
Fathers	2019-11	42.63	6.94	2.10	0.48	0.25	0.02	0.18	229
		(0.41)	(0.29)	(0.05)	(0.03)	(0.03)	(0.01)	(0.03)	
	2020-04	43.14	6.85	2.07	0.42	0.27	0.05	0.18	257
		(0.40)	(0.28)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	
	2020-11	42.62	6.96	2.08	0.43	0.25	0.04	0.19	283
		(0.39)	(0.27)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	
	2021-11	42.62	6.86	2.13	0.50	0.28	0.04	0.18	275
		(0.38)	(0.27)	(0.05)	(0.03)	(0.03)	(0.01)	(0.02)	

Notes: The table displays means and standard errors of basic demographic characteristics of the LISS sample by month and gender. The age variable is taken directly from the LISS survey. The values for the variables age of youngest child and number of children are taken from the administrative records for all linked individuals and from the LISS survey for all those who are not linked. The education variable is taken from the administrative records and therefore only available for linked individuals (note that even for linked individuals it is possible that the education is unknown).

2.A.6 Additional descriptive statistics



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

Figure 2.A.4. Labor force participation and hours categories over time

Notes: The figure provides an illustration of in Table 2.3.1. The data source are all working-age (18-55 years old) households with a child below 16 years of age by month and gender. Individuals are classified as unemployed when they are receiving unemployment benefits and classified as out of the labor force when there are no working hours, no self-employment status, and no unemployment benefits recorded in the administrative data. Consistent with the official definition of CBS Netherlands, we classify individuals to be working part-time if they work less than 35 hours per week.

Table 2.A.5. Basic demographics, CBS sample for the analysis in Section 2.4.2

	Mothers		Fathers	
	Control: 2014-2019	Treated: 2018-2021	Control: 2014-2019	Treated: 2018-2021
Age	38.54	38.54	40.93	40.95
	(6.53)	(6.49)	(6.75)	(6.77)
Age youngest child	6.16	6.09	6.18	6.13
	(4.73)	(4.73)	(4.73)	(4.73)
Education: High	0.43	0.46	0.39	0.41
	(0.49)	(0.5)	(0.49)	(0.49)
Education: Low	0.06	0.06	0.07	0.07
	(0.24)	(0.23)	(0.25)	(0.25)
Education: Middle	0.29	0.31	0.27	0.3
	(0.45)	(0.46)	(0.44)	(0.46)
Education: Unknown	0.22	0.17	0.27	0.22
	(0.42)	(0.38)	(0.44)	(0.41)
Full-time	0.11	0.13	0.83	0.82
	(0.31)	(0.33)	(0.37)	(0.38)
Number of children	1.94	1.94	1.94	1.94
	(0.74)	(0.75)	(0.74)	(0.75)
Out of labor force	0.04	0.04	0.02	0.02
	(0.2)	(0.19)	(0.13)	(0.13)
Part-time	0.83	0.82	0.14	0.15
	(0.38)	(0.38)	(0.34)	(0.36)
Potential commuting gains	6.25	6.41	10.16	10.28
	(12.2)	(12.14)	(15.95)	(15.97)
Potential hours remote work	7.64	8.06	12.36	12.68
	(7.21)	(7.35)	(8.88)	(8.96)
Remote work capability (imputed)	27.64	28.14	32.67	33.12
	(21.87)	(21.76)	(23.18)	(23.17)
Unemployed	0.02	0.01	0.01	0.01
	(0.14)	(0.11)	(0.12)	(0.09)
Working hours	25.48	26.51	38.4	38.41
	(7.76)	(7.52)	(5.13)	(5.06)
Working hours (unconditional)	24.11	25.31	37.39	37.52
	(9.48)	(9.19)	(7.95)	(7.63)
N	7,579,149	2,498,952	7,441,669	2,452,873

Notes: The table displays means and standard deviations of the pooled event-study DiD sample for parents with a youngest child below 16. All variables are reported separately for the treatment and control group and for mothers and fathers. The difference between working hours and unconditional working hours is that the former excludes working hours of 0, while the latter is not conditional on working and therefore includes working hours of 0. Individuals are classified as unemployed when they are receiving unemployment benefits and classified as out of the labor force when there are no working hours, no self-employment status, and no unemployment benefits recorded in the administrative data. Consistent with the official definition of CBS Netherlands, we classify individuals to be working part-time if they work less than 35 hours per week. Imputed remote work capability is calculated with the procedure in Section 2.A.5. Potential hours of remote work are calculated by multiplying the imputed remote work capability two years ago with actual working hours two years ago. Potential commuting gains are calculated by multiplying the imputed remote work capability two years ago with the commuting distance two years ago, assuming that all individuals commute on five working days per week.

Table 2.A.6. Labor market status over time in the LISS data

		Out of the labor force	Unemployed	Self-employed	Part-time employed	Full-time employed
Mothers	2019-11	0.188	0.008	0.131	0.581	0.092
		(0.024)	(0.005)	(0.021)	(0.031)	(0.018)
	2020-04	0.146	0.004	0.140	0.595	0.115
		(0.021)	(0.004)	(0.021)	(0.029)	(0.019)
	2020-11	0.159	0.003	0.109	0.591	0.138
		(0.020)	(0.003)	(0.017)	(0.027)	(0.019)
	2021-11	0.134	0.008	0.122	0.632	0.103
		(0.019)	(0.005)	(0.019)	(0.027)	(0.017)
Fathers	2019-11	0.026	0.000	0.074	0.170	0.729
		(0.011)	(0.000)	(0.017)	(0.025)	(0.029)
	2020-04	0.013	0.004	0.072	0.170	0.740
		(0.007)	(0.004)	(0.016)	(0.023)	(0.027)
	2020-11	0.025	0.003	0.071	0.138	0.763
		(0.009)	(0.003)	(0.015)	(0.020)	(0.025)
	2021-11	0.020	0.003	0.076	0.155	0.746
		(0.008)	(0.004)	(0.016)	(0.022)	(0.026)

Notes: The table shows the labor market participation by month and gender for the LISS sample. For all variables means and standard errors are reported. Individuals are classified as unemployed when they are receiving unemployment benefits and classified as out of the labor force when there are no working hours, no self-employment status, and no unemployment benefits recorded in the administrative data. Consistent with the official definition of CBS Netherlands, we classify individuals to be working part-time if they work less than 35 hours per week.

Table 2.A.7. Total working hours over time

	CBS			LISS		
	All	Fathers	Mothers	All	Fathers	Mothers
2012-11	31.6	38.3	23.9			
	(0.0073)	(0.006)	(0.009)			
2013-11	31.3	37.8	24			
	(0.0072)	(0.0062)	(0.0089)			
2014-11	31	37.2	24.1			
	(0.0071)	(0.0062)	(0.009)			
2015-11	31.7	37.8	24.7			
	(0.0071)	(0.0061)	(0.009)			
2016-11	32.2	38.4	25.2			
	(0.0071)	(0.0062)	(0.009)			
2017-11	32.2	38.2	25.4			
	(0.007)	(0.0062)	(0.0089)			
2018-11	32.3	38.2	25.7			
	(0.0069)	(0.0062)	(0.0087)			
2019-11	32.3	38.1	26	28.5	37.1	20.9
	(0.0068)	(0.0062)	(0.0087)	(0.633)	(0.608)	(0.812)
2020-04				28.6	36.6	21.5
				(0.587)	(0.578)	(0.77)
2020-11	32.4	38	26.3	29.3	37	22.5
	(0.0067)	(0.0062)	(0.0086)	(0.549)	(0.539)	(0.732)
2021-11	32.6	38.1	26.8	29.6	37.4	22.7
	(0.0066)	(0.0062)	(0.0086)	(0.548)	(0.518)	(0.729)

Notes: The table displays mean and standard errors for the variable working hours by month and gender for individuals which are employed. The first column shows the average working hours of all working-age (18-55 years old) parents with a child below 16 years old. The second column shows the average working hours for the LISS sample. For all individuals in the LISS sample, which can be linked to the administrative records, we take the actual working hours from the administrative records. For the individuals which cannot be linked, we take the information on working hours from the LISS survey.

Table 2.A.8. Remote working hours over time

	LISS		
	All	Fathers	Mothers
2020-02	4.5	4.8	4.2
	(0.46)	(0.67)	(0.63)
2020-04	17	19	14
	(0.85)	(1.3)	(1.1)
2020-11	14	17	12
	(0.78)	(1.2)	(1)
2021-11	13	14	11
	(0.71)	(1.1)	(0.94)

Notes: The table shows mean and standard errors for the variable remote work hours by month and gender for the LISS sample.

Table 2.A.9. Commuting hours over time

	LISS		
	All	Fathers	Mothers
2019-11	4.6	6	3
	(0.22)	(0.36)	(0.21)
2020-04	1.9	2.1	1.5
	(0.19)	(0.24)	(0.29)
2020-11	2.4	2.9	1.8
	(0.13)	(0.21)	(0.15)
2021-11	2.9	3.6	2.2
	(0.19)	(0.27)	(0.26)

Notes: The table shows mean and standard errors for the variable remote work hours by month and gender for the LISS sample.

Table 2.A.10. Remote work dummy over time

	LISS		
	All	Fathers	Mothers
2020-02	0.34	0.36	0.33
	(0.023)	(0.032)	(0.032)
2020-04	0.61	0.63	0.59
	(0.023)	(0.031)	(0.033)
2020-11	0.55	0.6	0.5
	(0.022)	(0.031)	(0.032)
2021-11	0.54	0.58	0.49
	(0.023)	(0.031)	(0.032)

Notes: The table shows the mean and standard errors of the variable remote work dummy by month and gender for the LISS sample. We construct the remote work dummy ourselves such that it measures whether an individual did any remote work or none at all.

Table 2.A.11. Remote work share over time

	LISS		
	All	Fathers	Mothers
2020-02	0.13	0.12	0.15
	(0.014)	(0.016)	(0.023)
2020-04	0.51	0.51	0.51
	(0.025)	(0.033)	(0.037)
2020-11	0.43	0.45	0.41
	(0.025)	(0.031)	(0.039)
2021-11	0.37	0.37	0.38
	(0.022)	(0.028)	(0.034)

Notes: The table shows mean and standard errors of the the variable remote work share by month and gender for the LISS sample. The remote work share is calculated as hours worked from home divided by total working hours.

Table 2.A.12. Childcare hours over time

	LISS		
	All	Fathers	Mothers
2019-11	23.1	17	28.4
	(0.886)	(0.959)	(1.35)
2020-04	36	30	41.2
	(0.962)	(1.34)	(1.3)
2020-11	22.9	20	25.5
	(0.764)	(1.06)	(1.07)
2021-11	23.1	18.7	27
	(0.758)	(0.96)	(1.1)

Notes: The table shows mean and standard errors for the variable childcare hours by month and gender for the LISS sample.

Appendix 2.B Results

2.B.1 Remote work and Commuting

Table 2.B.1. Predictive power of potential remote working hours for realized hours worked from home and commuting time

	Remote working hours		Commuting hours	
	(1)	(2)	(3)	(4)
Constant	5.68***	1.49*	4.63***	4.06***
	(1.25)	(0.79)	(0.28)	(0.38)
2020-04	12.0***	2.61***	−2.7***	−0.95*
	(0.92)	(0.86)	(0.29)	(0.50)
2020-11	9.74***	−0.16	−2.19***	−0.78**
	(0.86)	(0.68)	(0.22)	(0.34)
2021-11	7.92***	0.87	−1.65***	−0.16
	(0.82)	(0.91)	(0.27)	(0.48)
Pot. hours remote work		0.21***		0.04**
		(0.04)		(0.02)
Pot. hours remote work × 2020-04		0.63***		−0.11***
		(0.06)		(0.02)
Pot. hours remote work × 2020-11		0.61***		−0.09***
		(0.05)		(0.02)
Pot. hours remote work × 2021-11		0.42***		−0.09***
		(0.05)		(0.02)
N children == 2	−0.56	−0.4	−0.34	−0.33
	(1.59)	(0.91)	(0.25)	(0.24)
N children == 3	−2.9	−0.93	0.07	−0.03
	(1.82)	(1.06)	(0.36)	(0.34)
N children == 4	−1.68	−0.74	1.66*	1.60*
	(2.59)	(1.96)	(1.01)	(0.96)
N children >4	−4.43	0.57	1.16	0.77
	(3.36)	(1.06)	(1.08)	(0.94)
Age youngest child (std)	−0.2	0.08	0.04*	0.02
	(0.12)	(0.08)	(0.02)	(0.02)
Observations	1,876	1,876	1,876	1,876
R^2	0.081	0.471	0.069	0.11

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. The table displays the relationship between commuting hours and remote work hours and remote work potential. All specification control for age of the youngest child standardized by subtracting the pooled sample mean (6.8) divided by the standard deviation (4.7), as well as indicator variables indicating number of children.. Sample restricted to parents working before the pandemic (Nov 2019).

Table 2.B.2. Predictive power of potential remote working hours for realized hours worked from home and commuting time by gender

	Hrs remote work (1)	Hrs commuting (2)
Constant	1.84	6.06***
	(1.12)	(0.69)
2020-04	3.46**	−2.21***
	(1.45)	(0.83)
2020-11	0.94	−1.73***
	(1.15)	(0.65)
2021-11	0.99	−0.93
	(1.52)	(0.79)
Mother	−0.97	−3.38***
	(1.62)	(0.79)
Mother × 2020-04	−1.4	2.11**
	(1.80)	(0.97)
Mother × 2020-11	−2.02	1.61**
	(1.43)	(0.75)
Mother × 2021-11	−0.14	1.31
	(1.92)	(0.97)
Pot. hours remote work	0.21***	0.00
	(0.04)	(0.02)
Pot. hours remote work × 2020-04	0.62***	−0.09***
	(0.09)	(0.03)
Pot. hours remote work × 2020-11	0.58***	−0.07***
	(0.07)	(0.02)
Pot. hours remote work × 2021-11	0.42***	−0.08***
	(0.07)	(0.03)
Pot. hours remote work × Mother	0.03	0.04
	(0.08)	(0.03)
Pot. hours remote work × Mother × 2020-04	−0.0	−0.01
	(0.12)	(0.04)
Pot. hours remote work × Mother × 2020-11	0.07	−0.01
	(0.10)	(0.03)
Pot. hours remote work × Mother × 2021-11	−0.01	−0.01
	(0.11)	(0.04)
Observations	1,876	1,876
R^2	0.166	0.473

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. The table displays the relationship between commuting hours and remote work hours and remote work potential interacted with gender. All specification control for the age of the youngest child demeaned by subtracting the pooled sample mean (6.8) interacted with gender, as well as indicator variables indicating number of children. Sample restricted to parents working before the pandemic (Nov 2019).

2.B.2 Childcare

Table 2.B.3. Hours childcare and potential hours of remote work before and during the CoVid-19 Pandemic – full table

	Childcare Hours	
	(1)	(2)
Constant	28.09***	35.52***
	(1.47)	(1.79)
2020-04	6.80***	6.49***
	(1.43)	(1.41)
2020-11	−4.99***	−5.83***
	(1.27)	(1.28)
2021-11	−3.19***	−3.6***
	(1.19)	(1.20)
Pot. hours remote work	−0.26***	−0.14***
	(0.05)	(0.05)
Pot. hours remote work × 2020-04	0.50***	0.51***
	(0.07)	(0.07)
Pot. hours remote work × 2020-11	0.39***	0.42***
	(0.07)	(0.07)
Pot. hours remote work × 2021-11	0.25***	0.26***
	(0.06)	(0.06)
N children == 2	−0.72	−0.92
	(1.25)	(1.23)
N children == 3	−1.33	−1.49
	(1.52)	(1.49)
N children == 4	−1.23	−1.49
	(2.41)	(2.59)
N children >4	2.05	−0.6
	(2.48)	(2.41)
Age youngest child (std)	−1.92***	−1.88***
	(0.10)	(0.10)
Controlling for Working Hours (2019-11)	No	Yes
Observations	2,234	2,234
R^2	0.273	0.308

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. This table displays for the LISS Sample, how the relationship between remote work ability and childcare changes over the course of the pandemic. All specification control for age of the youngest child demeaned by subtracting the pooled sample mean (6.8), as well as indicator variables indicating number of children. Potential hours of remote work is set to zero for parents who do not work before the pandemic. Column (2) restricts the sample to parents who work 35 hours or more before the pandemic, column (3) restricts the sample to parents working between 20 and 34 hours, and column (4) restricts the sample to parents working less than 20 hours.

Table 2.B.4. Hours spent on childcare and potential hours of remote work before and during the CoVid-19 Pandemic, conditional on working in November 2019

	Childcare Hours	
	(1)	(2)
Constant	25.53***	37.63***
	(1.48)	(2.17)
2020-04	5.45***	5.79***
	(1.65)	(1.62)
2020-11	−6.66***	−6.41***
	(1.34)	(1.33)
2021-11	−5.31***	−5.31***
	(1.33)	(1.32)
Pot. hours remote work	−0.19***	−0.07
	(0.05)	(0.05)
Pot. hours remote work × 2020-04	0.56***	0.55***
	(0.08)	(0.08)
Pot. hours remote work × 2020-11	0.47***	0.46***
	(0.07)	(0.07)
Pot. hours remote work × 2021-11	0.36***	0.36***
	(0.06)	(0.06)
N children == 2	−0.3	−0.78
	(1.28)	(1.25)
N children == 3	−0.04	−0.69
	(1.61)	(1.55)
N children == 4	−0.26	−0.21
	(2.79)	(2.80)
N children >4	5.07*	0.66
	(2.94)	(2.20)
Age youngest child (std)	−1.78***	−1.73***
	(0.11)	(0.11)
Controlling for Working Hours (2019-11)	No	Yes
Observations	1,876	1,876
R^2	0.294	0.332

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. This table displays for the LISS Sample, how the relationship between remote work ability and childcare changes over the course of the pandemic. All specification control for age of the youngest child demeaned by subtracting the pooled sample mean (6.8), as well as indicator variables indicating number of children. Sample restricted to parents working before the pandemic (Nov 2019). Column (2) restricts the sample to parents that work 35 hours or more, column (3) restricts the sample to parents working between 20 and 34 hours, and column (4) restricts the sample to parents working less than 20 hours.

Table 2.B.5. Evolution of the gender care gap and potential hours of remote work – full table

	Hrs childcare		
	(1)	(2)	(3)
Constant	17.28***	18.05***	17.75***
	(1.41)	(1.43)	(1.44)
2020-04	12.8***	10.49***	10.49***
	(1.53)	(1.46)	(1.46)
2020-11	2.96**	1.20	1.69
	(1.29)	(1.29)	(1.26)
2021-11	1.55	0.31	0.36
	(1.19)	(1.20)	(1.23)
N children == 2	0.72	0.73	0.66
	(1.59)	(1.55)	(1.56)
N children == 3	−0.12	0.20	0.21
	(2.07)	(2.03)	(2.04)
N children == 4	1.57	1.58	1.33
	(3.80)	(3.65)	(3.62)
N children >4	5.66	5.35	4.97
	(4.00)	(4.08)	(4.17)
Age youngest child (std)	−1.4***	−1.36***	−1.34***
	(0.14)	(0.14)	(0.14)
Age youngest child (std) × Mother	−1.07***	−1.06***	−1.1***
	(0.18)	(0.18)	(0.18)
N children == 2 × Mother	−2.07	−1.81	−1.99
	(2.13)	(2.17)	(2.17)
N children == 3 × Mother	−1.46	−1.45	−1.66
	(2.66)	(2.64)	(2.67)
N children == 4 × Mother	−4.02	−2.74	−3.02
	(5.07)	(4.99)	(4.95)
N children >4× Mother	−8.18*	−6.48	−6.72
	(4.75)	(4.96)	(5.09)
Mother	14.01***	12.44***	12.67***
	(2.13)	(2.13)	(2.14)
Mother × 2020-04	−0.09	3.85**	3.74*
	(1.99)	(1.94)	(1.94)
Mother × 2020-11	−6.15***	−3.29*	−3.36*
	(1.84)	(1.87)	(1.86)
Mother × 2021-11	−3.31**	−1.46	−1.34
	(1.60)	(1.63)	(1.62)
Pot. hours remote work (std)		−0.16***	−0.1**
		(0.05)	(0.05)
Pot. hours remote work (std) × 2020-04		0.52***	0.55***
		(0.07)	(0.09)
Pot. hours remote work (std) × 2020-11		0.36***	0.26***
		(0.07)	(0.08)
Pot. hours remote work (std) × 2021-11		0.23***	0.21***
		(0.06)	(0.07)
Pot. hours remote work (std) × Mother			−0.16
			(0.10)
Pot. hours remote work (std) × Mother × 2020-04			−0.05
			(0.14)
Pot. hours remote work (std) × Mother × 2020-11			0.24*
			(0.13)
Pot. hours remote work (std) × Mother × 2021-11			0.04
			(0.13)
Observations	2,234	2,234	2,234
R^2	0.324	0.347	0.349

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. The potential hours of remote work are demeaned to facilitate comparison of coefficients across columns. All specifications control for the (demeaned) age of the youngest child interacted with gender, as well as indicator variables indicating number of children, the left-out category is a single child. In column (3), we additionally interact the number of children with gender, so that the model is fully satiated. Potential remote work hours are set to zero if the individual did not work before the pandemic.

Table 2.B.6. Evolution of the gender care gap and potential hours of remote work – conditional on working in November 2019

	Hrs childcare		
	(1)	(2)	(3)
Constant	16.3***	16.44***	16.41***
	(1.39)	(1.34)	(1.34)
2020-04	13.38***	13.7***	13.75***
	(1.55)	(1.44)	(1.44)
2020-11	3.75***	3.46***	3.50***
	(1.32)	(1.26)	(1.25)
2021-11	2.08*	1.83	1.85
	(1.20)	(1.17)	(1.16)
Mother	12.45***	11.65***	11.71***
	(2.21)	(2.16)	(2.16)
Mother × 2020-04	1.61	2.08	1.92
	(2.11)	(2.00)	(2.02)
Mother × 2020-11	−5.61***	−4.94***	−4.93**
	(1.97)	(1.91)	(1.92)
Mother × 2021-11	−3.09*	−2.76*	−2.77*
	(1.70)	(1.67)	(1.67)
Pot. hours remote work (std)		−0.02	−0.02
		(0.02)	(0.02)
Pot. hours remote work (std) × 2020-04		0.20***	0.22***
		(0.03)	(0.04)
Pot. hours remote work (std) × 2020-11		0.15***	0.13***
		(0.02)	(0.03)
Pot. hours remote work (std) × 2021-11		0.11***	0.10***
		(0.02)	(0.03)
Pot. hours remote work (std) × Mother			−0.0
			(0.03)
Pot. hours remote work (std) × Mother × 2020-04			−0.06
			(0.05)
Pot. hours remote work (std) × Mother × 2020-11			0.04
			(0.04)
Pot. hours remote work (std) × Mother × 2021-11			0.01
			(0.04)
Observations	1,876	1,876	1,876
R^2	0.317	0.369	0.37

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the household level. The potential hours of remote work are demeaned to facilitate comparison of coefficients across columns. All specifications control for the (demeaned) age of the youngest child interacted with gender, as well as indicator variables indicating number of children, the left-out category is a single child. In column (3), we additionally interact the number of children with gender, so that the model is fully satiated. The figure displays the results for the sample restricted to parents working before the pandemic (Nov 2019).

2.B.3 Labor supply: Regression equation with pot. commuting gains

$$
\begin{aligned}
\text{Working Hours}_{i,t} = {} & \alpha + \chi \text{ Pot. hrs remote work}_i + \phi \text{ Pot. hrs remote work partner}_i \\
& + \upsilon \text{ Pot. commuting gains}_i + \nu \text{ Pot. commuting gains partner}_i \\
& + \sum_{t=-1}^{2} (\beta_t \text{ Pot. hrs remote work}_i + \delta_t \text{ Pot. hrs remote work partner}_i \\
& + \lambda_t \text{ Pot. commuting gains}_i + \phi_t \text{ Pot. commuting gains partner}_i) \\
& \times \mathbb{1}(Year = t) \\
& + \sum_{t=-1}^{2} (\gamma_t \text{ Pot. hrs remote work}_i + \omega_t \text{ Pot. hrs remote work partner}_i \\
& + \theta_t \text{ Pot. commuting gains}_i + \kappa_t \text{ Pot. commuting gains partner}_i) \\
& \times \mathbb{1}(Year = t) \times \text{Pandemic}_i \\
& + \sum_{t=-1}^{2} \mu_t \mathbb{1}(Year = t) + \sum_{t=-1}^{2} \sigma_t \mathbb{1}(Year = t) \times \text{Pandemic}_i \\
& + \pi \text{ Pandemic}_i + \rho \text{ Age youngest child}_{i,0} + \eta \text{ Number children}_{i,t} \\
& + \iota \text{ Age}_{i,t} + \xi \text{ Age partner}_{i,t} + \epsilon_{i,t}
\end{aligned}
$$

2.B.4 Labor supply

Table 2.B.9. The effect of potential remote hours on working hours

	Mothers		Fathers	
	(1)	(2)	(3)	(4)
Constant	19.823***	21.279***	40.68***	41.274***
	(0.06)	(0.057)	(0.051)	(0.046)
Part: Pot hrs wfh × t = -1 × Pand	0.00	−0.0	−0.002	−0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Pot hrs wfh × t = 1 × Pand	0.008***	0.005**	0.004**	0.002
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Pot hrs wfh × t = 2 × Pand	0.01***	0.005**	0.009***	0.007***
	(0.002)	(0.002)	(0.002)	(0.003)
Pot hrs wfh × t = -1 × Pand	−0.021***	−0.024***	−0.02***	−0.019***
	(0.002)	(0.002)	(0.002)	(0.002)
Pot hrs wfh × t = 1 × Pand	0.051***	0.05***	0.03***	0.026***
	(0.002)	(0.002)	(0.002)	(0.002)
Pot hrs wfh × t = 2 × Pand	0.073***	0.061***	0.028***	0.016***

	(0.002)	(0.003)	(0.002)	(0.002)
Part: Pot comm gain × t = -1 × Pand		0.00		0.004***
		(0.001)		(0.001)
Part: Pot comm gain × t = 1 × Pand		0.003***		0.00
		(0.001)		(0.001)
Part: Pot comm gain × t = 2 × Pand		0.004**		0.001
		(0.001)		(0.002)
Pot comm gain × t = -1 × Pand		0.004**		0.004***
		(0.002)		(0.001)
Pot comm gain × t = 1 × Pand		−0.003**		−0.001
		(0.002)		(0.001)
Pot comm gain × t = 2 × Pand		0.006***		0.004***
		(0.002)		(0.001)
t = -1	−0.711***	−0.67***	−0.736***	−0.571***
	(0.009)	(0.011)	(0.01)	(0.011)
t = 1	0.658***	0.564***	0.403***	0.279***
	(0.009)	(0.011)	(0.009)	(0.01)
t = 2	1.158***	0.954***	0.425***	0.209***
	(0.014)	(0.015)	(0.014)	(0.014)
Pand	1.511***	0.782***	0.327***	−0.291***
	(0.023)	(0.022)	(0.022)	(0.022)
t = -1 × Pand	0.461***	0.414***	0.918***	0.75***
	(0.025)	(0.026)	(0.025)	(0.026)
t = 1 × Pand	−0.858***	−0.746***	−0.931***	−0.801***
	(0.025)	(0.026)	(0.026)	(0.026)
t = 2 × Pand	−0.762***	−0.551***	−0.681***	−0.46***
	(0.033)	(0.034)	(0.033)	(0.033)
Pot hrs wfh	0.546***	0.577***	0.056***	0.051***
	(0.001)	(0.001)	(0.001)	(0.001)
Pot hrs wfh × Pand	−0.06***	−0.05***	−0.009***	0.01***
	(0.002)	(0.002)	(0.001)	(0.002)
Pot hrs wfh × t = -1	0.02***	0.024***	0.013***	0.01***
	(0.001)	(0.001)	(0.001)	(0.001)
Pot hrs wfh × t = 1	−0.029***	−0.03***	−0.01***	−0.008***
	(0.001)	(0.001)	(0.00)	(0.001)
Pot hrs wfh × t = 2	−0.053***	−0.05***	−0.014***	−0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot hrs wfh	−0.028***	−0.03***	−0.1***	−0.104***
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot hrs wfh × Pand	−0.001	0.002	−0.002	−0.0
	(0.001)	(0.002)	(0.002)	(0.002)
Part: Pot hrs wfh × t = -1	−0.001	−0.001	0.004***	0.003***

	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot hrs wfh × t = 1	0.00	−0.0	−0.001	−0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot hrs wfh × t = 2	−0.001	0.001	−0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot comm gain		0.00		0.006***
		(0.001)		(0.001)
Part: Pot comm gain × Pand		−0.001		−0.003***
		(0.001)		(0.001)
Part: Pot comm gain × t = -1		0.00		−0.0
		(0.00)		(0.001)
Part: Pot comm gain × t = 1		−0.0		0.00
		(0.00)		(0.001)
Part: Pot comm gain × t = 2		−0.001*		−0.002**
		(0.001)		(0.001)
Pot comm gain		−0.055***		−0.013***
		(0.001)		(0.00)
Pot comm gain × Pand		−0.0		−0.004***
		(0.001)		(0.001)
Pot comm gain × t = -1		−0.002***		−0.001***
		(0.001)		(0.00)
Pot comm gain × t = 1		0.006***		0.002***
		(0.001)		(0.00)
Pot comm gain × t = 2		0.004***		−0.0
		(0.001)		(0.00)
N children = 2	−1.0***	−1.106***	0.227***	0.139***
	(0.015)	(0.014)	(0.013)	(0.012)
N children = 3	−2.043***	−2.079***	0.15***	0.099***
	(0.022)	(0.021)	(0.019)	(0.017)
N children = 4	−3.589***	−3.495***	0.104**	0.115***
	(0.052)	(0.05)	(0.043)	(0.039)
N children >4	−6.149***	−6.0***	0.558***	0.552***
	(0.137)	(0.137)	(0.109)	(0.10)
Age	0.00	−0.01***	−0.067***	−0.056***
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Age	0.025***	0.026***	−0.038***	−0.046***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.175	0.193	0.015	0.018

Table 2.B.7. The effect of potential remote share on working hours

	Mothers		Fathers	
	(1)	(2)	(3)	(4)
Part: Pot share wfh × t = -1 × Pand	−0.0	−0.0	−0.001*	−0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot share wfh × t = 1 × Pand	0.004***	0.002***	0.002***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot share wfh × t = 2 × Pand	0.004***	0.002**	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot comm gain × t = -1 × Pand		0.001		0.004***
		(0.001)		(0.001)
Part: Pot comm gain × t = 1 × Pand		0.003***		0.00
		(0.001)		(0.001)
Part: Pot comm gain × t = 2 × Pand		0.004***		0.002
		(0.001)		(0.002)
R^2	0.097	0.104	0.01	0.014

Notes: $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$. Standard errors obtained by clustering on the individual level. Full results in Table 2.B.11.

Table 2.B.11. The effect of potential remote share on working hours

	Mothers		Fathers	
	(1)	(2)	(3)	(4)
Constant	19.523***	21.076***	40.806***	41.421***
	(0.064)	(0.061)	(0.051)	(0.046)
Part: Pot share wfh × t = -1 × Pand	−0.0	−0.0	−0.001*	−0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot share wfh × t = 1 × Pand	0.004***	0.002***	0.002***	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot share wfh × t = 2 × Pand	0.004***	0.002**	0.003***	0.002**
	(0.001)	(0.001)	(0.001)	(0.001)
Pot share wfh × t = -1 × Pand	−0.003***	−0.003***	−0.007***	−0.006***
	(0.001)	(0.001)	(0.001)	(0.001)
Pot share wfh × t = 1 × Pand	0.014***	0.015***	0.011***	0.01***
	(0.001)	(0.001)	(0.001)	(0.001)
Pot share wfh × t = 2 × Pand	0.02***	0.017***	0.01***	0.007***
	(0.001)	(0.001)	(0.001)	(0.001)
Part: Pot comm gain × t = -1 × Pand		0.001		0.004***
		(0.001)		(0.001)
Part: Pot comm gain × t = 1 × Pand		0.003***		0.00
		(0.001)		(0.001)
Part: Pot comm gain × t = 2 × Pand		0.004***		0.002
		(0.001)		(0.002)
Pot comm gain × t = -1 × Pand		−0.001		0.002*
		(0.002)		(0.001)
Pot comm gain × t = 1 × Pand		−0.007***		−0.002**
		(0.002)		(0.001)
Pot comm gain × t = 2 × Pand		−0.0		0.001
		(0.002)		(0.001)
t = -1	−0.724***	−0.669***	−0.772***	−0.59***
	(0.01)	(0.012)	(0.01)	(0.011)
t = 1	0.672***	0.571***	0.404***	0.273***
	(0.01)	(0.011)	(0.009)	(0.011)
t = 2	1.192***	0.98***	0.452***	0.236***
	(0.015)	(0.016)	(0.015)	(0.014)
Pand	1.569***	0.82***	0.392***	−0.213***
	(0.024)	(0.024)	(0.023)	(0.022)
t = -1 × Pand	0.438***	0.378***	0.904***	0.722***
	(0.026)	(0.027)	(0.026)	(0.026)
t = 1 × Pand	−0.899***	−0.784***	−0.933***	−0.799***

	(0.027)	(0.028)	(0.026)	(0.027)
t = 2 × Pand	−0.804***	−0.574***	−0.68***	−0.457***
	(0.035)	(0.036)	(0.034)	(0.034)
Pot share wfh	0.117***	0.112***	−0.002***	−0.009***
	(0.00)	(0.00)	(0.00)	(0.00)
Pot share wfh × Pand	−0.011***	−0.008***	−0.004***	0.001
	(0.001)	(0.001)	(0.00)	(0.00)
Pot share wfh × t = -1	0.003***	0.004***	0.005***	0.004***
	(0.00)	(0.00)	(0.00)	(0.00)
Pot share wfh × t = 1	−0.005***	−0.006***	−0.003***	−0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
Pot share wfh × t = 2	−0.01***	−0.009***	−0.005***	−0.003***
	(0.00)	(0.00)	(0.00)	(0.00)
Part: Pot share wfh	0.005***	0.007***	−0.021***	−0.022***
	(0.00)	(0.00)	(0.00)	(0.00)
Part: Pot share wfh × Pand	−0.002***	−0.001*	−0.003***	−0.001**
	(0.00)	(0.001)	(0.00)	(0.001)
Part: Pot share wfh × t = -1	0.001***	0.001**	0.002***	0.002***
	(0.00)	(0.00)	(0.00)	(0.00)
Part: Pot share wfh × t = 1	−0.001***	−0.001**	−0.001***	−0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Part: Pot share wfh × t = 2	−0.002***	−0.001***	−0.002***	−0.001***
	(0.00)	(0.00)	(0.00)	(0.00)
Part: Pot comm gain		−0.009***		−0.0
		(0.001)		(0.001)
Part: Pot comm gain × Pand		−0.0		−0.002**
		(0.001)		(0.001)
Part: Pot comm gain × t = -1		−0.0		−0.001
		(0.00)		(0.001)
Part: Pot comm gain × t = 1		−0.0		0.00
		(0.001)		(0.001)
Part: Pot comm gain × t = 2		−0.0		−0.001*
		(0.001)		(0.001)
Pot comm gain		0.004***		0.006***
		(0.001)		(0.00)
Pot comm gain × Pand		0.001		−0.001
		(0.001)		(0.001)
Pot comm gain × t = -1		−0.002**		−0.002***
		(0.001)		(0.00)
Pot comm gain × t = 1		0.005***		0.001***
		(0.001)		(0.00)
Pot comm gain × t = 2		0.007***		0.002***

		(0.001)		(0.00)
N children = 2	−1.52***	−1.639***	0.342***	0.253***
	(0.016)	(0.015)	(0.013)	(0.012)
N children = 3	−2.892***	−2.933***	0.328***	0.274***
	(0.024)	(0.023)	(0.019)	(0.017)
N children = 4	−4.813***	−4.722***	0.271***	0.285***
	(0.057)	(0.056)	(0.043)	(0.04)
N children >4	−7.841***	−7.702***	0.684***	0.694***
	(0.149)	(0.151)	(0.109)	(0.101)
Age	0.044***	0.033***	−0.071***	−0.06***
	(0.002)	(0.002)	(0.002)	(0.002)
Part: Age	0.033***	0.033***	−0.026***	−0.034***
	(0.002)	(0.002)	(0.002)	(0.002)
R^2	0.097	0.104	0.01	0.014

References

Been, J., and Centerdata. 2021. "LISS Panel - Time Use and Consumption - Part 9." [117]

van Soest, Arthur, Jim Been, Pia Pinger, Hans-Martin von Gaudecker, and Centerdata. 2019. "LISS Panel - Time Use and Consumption - Part 6." [117]

von Gaudecker, Hans-Martin, and Centerdata. 2020a. "LISS Panel - Time Use and Consumption - Part 7 (Additional Coronavirus Part)." [117]

von Gaudecker, Hans-Martin, and Centerdata. 2020b. "LISS Panel - Time Use and Consumption - Part 8." [117]

von Gaudecker, Hans-Martin, Christian Zimpelmann, Moritz Mendel, Bettina Siflinger, Lena Janys, Jürgen Maurer, Egbert Jongen, Radost Holler, Renata Abikeyeva, Felipe Augusto Azuero Mutis, Annica Gehlen, and Eva Lucia Kleifgen. 2021. "CoVID-19 Impact Lab Questionnaire Documentation." [119]

Zimpelmann, Christian, Hans-Martin von Gaudecker, Radost Holler, Lena Janys, and Bettina Siflinger. 2021. "Hours and Income Dynamics during the Covid-19 Pandemic: The Case of the Netherlands." *Labour Economics* 73 (12): 102055. [119]

Chapter 3

The Effects of Initial Labor Market Conditions and Local Residence Restrictions on the Integration of Refugees in Germany

3.1 Introduction

In the wake of large-scale refugee movements worldwide, the economic and social integration of refugees has emerged as a paramount challenge faced by host countries. Among the multifaceted challenges inherent in this process, the spatial distribution of incoming refugees has garnered considerable attention from policymakers and scholars alike. One popular policy employed by many countries is the use of refugee dispersal mechanisms, which distribute incoming refugees across the country through the use of quotas. In Germany, for instance, most incoming refugees are allocated to their initial residence based on quotas, which are determined by a combination of tax revenue and population size but which do not take into account personal characteristics of the refugee.

* 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.

Proponents of this policy argue that the resulting distribution of refugees across regions within a country ensures a better sharing of the financial burden and a better integration of refugees. However, studies that cover multiple countries, including Germany, have shown that such dispersal policies can induce large inequalities between refugees as their integration is affected by different local conditions in the initial location like labor market conditions or the size of co-ethnic networks (Edin, Fredriksson, and Aslund, 2003; Damm, 2009; Beaman, 2012; Godøy, 2017; Aksoy, Poutvaara, and Schikora, 2020; Azlor, Damm, and Schultz-Nielsen, 2020).

While refugee dispersal mechanisms can determine the initial distribution of refugees across the country, they cannot prevent subsequent sorting of refugees into different regions within the country. Some countries, therefore, employ local residence restrictions, which aim to prevent refugees from changing their initial residence for a certain period after their arrival. Policymakers often argue that such measures can enhance the integration outcomes of refugees by preventing their concentration in specific areas. However, these residence restrictions effectively increase a refugee's exposure to local economic conditions in their initial location and may consequently exacerbate the inequalities caused by the initial allocation.

In this paper, I study both the effects of labor market conditions in the initial location and the effects of local residence restrictions in the context of the integration of refugees in Germany using data from the IAB-BAMF-SOEP (IBS) Survey of Refugees, a representative survey on refugees who arrived in Germany between 2013 and 2016. Making use of the quasi-random initial allocation of refugees, I find that being allocated to a high-unemployment county negatively affects the economic and social integration of a refugee. Being allocated into a county with an above-median unemployment rate decreases the probability that a refugee has found either full or part-time employment in 2017 by around 5 percentage points.

Similarly, I find that being allocated into a county with an above-median unemployment rate leads a refugee to have around 5 to 6 fewer new German acquaintances since arrival and reduces the frequency of social contact with Germans by 0.25 of a standard deviation. This result replicates the findings of Aksoy, Poutvaara, and Schikora (2020) who also use data from the IBS Survey of Refugees to show that refugees allocated to counties with higher unemployment rates have a lower likelihood of having full or part-time employment and a lower score on a multi-dimensional integration index, which also takes into account the social integration of refugees.

I further analyze the effects of local residence restrictions in Germany, which were introduced as part of an integration law reform in 2016.[1] The integration law reform introduced the restriction that all recognized refugees are obliged to have their residence for three years after the time of recognition in the *federal state*, in which the asylum process took place. Moreover, federal states had the opportunity to instruct their local authorities to further restrict the residence more locally to a certain *municipality* or *county*. The implementation varied substantially by federal state, with certain states (enforcing states) instructing their local authorities to make use of such local residence restrictions, while other states (non-enforcing) did not.[2]

1. The draft bill of the integration law states as its official goal to "foster the integration of recognized refugees and to counteract integration-impeding tendencies of segregation" (Deutscher Bundestag, 2016) by ensuring a more equal distribution of recognized refugees across the country.

2. The residence restrictions apply to all recognized refugees who completed their asylum process after the 1st of January 2016. Local residence restrictions lie at the discretion of local authorities who can use them if they think that it will foster the integration of a refugee. Possible reasons are a better supply of housing, better possibilities to learn German or better chances to find employment. Local authorities can restrict the residence to a certain *municipality* or *county* during the first three years after the asylum decision and can name certain municipalities to which refugees are not allowed to move. Unaffected by the new rule are unaccompanied minors and individuals who themselves, whose partner or whose underage child, have a job subject to social security contributions with at least 15 weekly working hours, who are in job training or who are pursuing studies at the time of their asylum recognition. Further, a refugee can request to be freed of residential restriction if they, their partner, or an underage child can obtain the above-mentioned engagement at a different location or if their partner or underage child is living somewhere else.

The differences in the enforcement of local residence restrictions across federal states give rise to a difference-in-differences (Diff-in-Diff) design, in which refugees who were allocated into non-enforcing states serve as the control group. I detect significant effects of the stricter enforcement of local residence conditions on economic and social integration outcomes which are, however, heterogeneous by local labor market conditions. While I find evidence that refugees allocated to low-unemployment counties benefit from local residence restrictions, refugees allocated to high-unemployment counties appear to be harmed by them.

My study is closely related to two studies using the IBS Survey of Refugees to analyze the integration of refugees. As discussed before, I replicate the findings of Aksoy, Poutvaara, and Schikora (2020) who show that refugees allocated to counties with higher unemployment rates have a lower likelihood of having full or part-time employment and a lower score on a multi-dimensional integration index, which also takes into account the social integration of refugees. I extend their analysis by showing how local residence restrictions, which increase the exposure of refugees to initial local economic conditions, affect the economic and social integration of refugees. In my study, I make use of the integration law reform with its across-state differences in the enforcement of local residence restrictions. This reform has been analyzed before by Schikora (2019) using a similar Diff-in-Diff approach. She finds that the stricter enforcement of local residence restrictions had a positive effect on the completion of language courses and the level of certified German language proficiency. She, however, does not look at the effects on the economic and social integration of refugees and does not investigate potential heterogeneity of the effects by local unemployment levels. The latter is especially important as I show that local unemployment conditions do not only directly affect the integration of refugees, but can also affect the performance of policies aimed at improving their integration, like local residence restrictions.

More generally, my study complements a large strand of literature using quasi-random allocation rules of refugees to identify the effect of local economic and social conditions on their subsequent integration. Various studies have analyzed the effects of the size and composition of the co-ethnic network in the initial location on the labor market integration (Edin, Fredriksson, and Aslund, 2003; Damm, 2009; Beaman, 2012; Martén, Hainmueller, and Hangartner, 2019) and political integration (Bratsberg, Ferwerda, Finseraas, and Kotsadam, 2021). My results are in line with various papers showing that the short and long-term employment outcomes of refugees and immigrants are negatively affected by bad local labor market conditions in the initial location (Åslund and Rooth, 2007; Godøy, 2017; Aksoy, Poutvaara, and Schikora, 2020; Azlor, Damm, and Schultz-Nielsen, 2020) and more generally in line with the literature showing that economic conditions when entering the labor market affect subsequent labor market outcomes (Schwandt and Wachter, 2019; Wachter, 2020). By analyzing the effects of the integration reform, I also contribute to a growing literature evaluating policy measures targeting the integration of refugees after their arrival. These include job search assistance programs (Battisti, Giesing, and Laurentsyeva, 2019; Arendt, 2022), language training (Arendt, Bolvig, Foged, Hasager, and Peri, 2020) and employment bans (Marbach, Hainmueller, and Hangartner, 2018; Fasani, Frattini, and Minale, 2021).[3]

From a policy perspective, the results of my study highlight that while quasi-random spatial allocation mechanisms might ensure a more equal sharing of financial burden across regions they can induce large inequalities in the integration between refugees. This implies that policymakers should either adjust allocation quotas by taking local labor market conditions more into account or provide additional support measures for "unlucky" refugees that get allocated into bad local

3. Comprehensive summaries of recent research on forced migration and the labor market integration of refugees can be found in Becker and Ferrara (2019) and Brell, Dustmann, and Preston (2020).

labor markets. In this study, I also analyze one policy aimed at improving the integration of refugees by preventing their relocation from the initial location, namely the introduction of local residence restrictions in Germany. I find evidence that while this measure was successful in increasing the labor market and social integration of refugees in low-unemployment areas, it appears not to be an appropriate policy to foster the integration of refugees in high-unemployment areas. Local residence restrictions, when designed like in Germany, are therefore exacerbating existing inequalities between refugees in different regions.

The remaining paper is structured as follows. Section 2 starts by presenting the institutional framework of the integration of refugees in Germany with a focus on the changes induced by the integration reform in 2016. Section 3 presents the data sources used and shows the construction of the estimation sample. Section 4 presents the empirical strategy used, together with a discussion of the validity of the required identifying assumptions. Section 5 presents the empirical results and Section 6 concludes.

3.2 Institutional Setting

The following section sheds light on the institutional framework governing the arrival and integration of incoming refugees in Germany, while having a special focus on the regulations in the asylum law regarding labor market entry and choice of residence of refugees. Further, I discuss the implementation of and the motivation behind the integration law reform in 2016, which is being analyzed in this paper.

3.2.1 Legal Framework

Arrival and Initial Allocation. Upon arrival in Germany, all refugees[4] are obliged to immediately report to the closest government agency for registration. During the height of the refugee influx in the years 2015 and 2016, most of these first registrations took place at cities in Bavaria which are located closely to the border. In these first-arrival government agencies, a refugee gets registered and it is checked whether they already claimed asylum in another European country or whether they were already previously registered in Germany. Upon completion of this procedure, the refugee receives a proof of arrival, which entitles them to receive social benefits like food and housing, and they are located into a short-term facility for a couple of days before being relocated to a so-called reception center in a federal state (Informationsverbund Asyl und Migration, 2020).

The relocation procedure of refugees mandates that refugees are allocated across federal states according to fixed quotas. The share of refugees that each of the 16 federal state has to accept is hereby determined to 2/3 according to the tax revenue and to 1/3 according to the population count of the state two years prior.[5] The relocation procedure, therefore, takes into consideration, at least partially, the local economic conditions at the state level, and consequently assigns a greater number of refugees to economically stronger federal states. In practice, refugees are located into a federal state which has not yet fulfilled its quotas and which has a branch of the Federal Office for Migration responsible for the nationality.[6]

4. Note that I use the word "refugee" to generally describe all people who enter Germany with the intention to seek asylum. Among this group, I call refugees whose asylum application is not yet decided "asylum seekers", refugees whose asylum application was successful "recognized refugees", and refugees whose asylum application was denied but who cannot be deported to their home country for humanitarian reasons "tolerated refugees".

5. The quotas are based on the "Königsteiner Schlüssel", which is published yearly by the Joint Science Conference ("Gemeinsame Wissenschaftskonferenz (GWK)"). It was originally designed to regulate the joint funding of scientific research facilities by the federal states in Germany, but is now generally used to distribute joint financial burden among federal states. The exact quotas for the years 2010 to 2019 can be found in Figure 3.A.1.

6. Unfortunately, the Federal Office for Migration does not publish the information on which branch is responsible for which nationalities since it is considered *internal* information. In the

Importantly, a federal state cannot reject a refugee that was allocated to it and the allocated refugees themselves cannot influence the allocation decision, with the only exception being that refugees who already have close family members living at a different location in Germany can reunite with them.[7]

Following the initial allocation to a reception center in a federal state, a refugee is obliged to make an official asylum application at the closest branch of the Federal Office for Migration and thereby receives the legal status of an asylum seeker while the asylum process decision is pending. If their application is successful, the asylum seeker becomes a recognized refugee and obtains the permit to reside in Germany. However, in case of denial they are generally obliged to leave the country. Exceptions are made for refugees who have serious reasons hindering them from returning to their home country (e.g., an ongoing armed conflict). Those refugees are considered tolerated refugees and receive a limited residence permit with less rights than recognized refugees as long as the reasons hindering them from returning still persist (Bundesamt für Migration und Flüchtlinge, 2020).

While the initial allocation to the federal states is clearly governed by the fixed quotas described above, the further allocation inside the federal state on different counties is regulated by state-specific laws. With the exception of the three city states, all federal states use additional quotas to further distribute the refugees to counties and municipalities. These quotas are determined mostly by population size, however, some states also use additional decision criteria.[8]

Labor Market Entry and Choice of Residence. Entry into the German labor market is open for all recognized refugees and most asylum seekers and tolerated refugees who reside in Germany for at least three months and who are not

randomization checks in Table 3.4.1 in Section 3.4.2, I check to which extent there exist across-state differences in the composition of refugees with respect to nationalities.

7. Many of these refugees arrive under a different procedure called family reunification ("Familiennachzug") and are not part of the analysis in this paper.

8. For more details on these quotas see Table 3.A.1 in Appendix 3.A.

anymore obliged to live in a reception center. However, before the integration law was passed in August 2016, every refugee that received a job offer faced a so-called priority check ("Vorrangprüfung") before being able to accept it. The priority check consisted in the local employment agency making sure that no EU-citizen employee is available who could take the job instead of the refugee. The priority check was applied until the fifteenth month after the arrival of the refugee. Through the integration law a pause was put to this priority check for three years and it was finally abolished in 2019 (Lehrian and Mantel, 2016).

Regarding their choice of residence, asylum seekers and tolerated refugees are constrained in the first three months after arrival by a so-called residential obligation ("Residenzpflicht"). This obligation entails that an individual has to stay in the area of the respective reception center they were allocated to for the first three months. After that period, local authorities generally have the permission to continue putting residence restrictions on refugees. Prior to the integration law reform, however, applying those restrictions was not very common. This changed fundamentally when the integration law became effective in 2016.

3.2.2 Integration Law Reform

The integration law reform ("Integrationsgesetz") was implemented on the 6th of August 2016 and retroactively applied to all refugees who have received their asylum process decision after the 1st of January 2016.[9] It set new rules regarding the integration of refugees under the slogan "Promote and Demand" ("Fördern und Fordern"). In essence, it redefined the rights and obligations of asylum seekers and recognized refugees with the goal of creating better chances for their successful integration into the German society and labor market.

9. A successful asylum process decision consists in the refugee being either recognized as "Asylberechtigter" according to Art.16a GG, as "Flüchtling" according to §3 Asylgesetz, as "subsidiär Schutzberechtigter" according to §4 AsylG or if they received an "Aufenthaltserlaubnis" according to §22, 23, 25 Abs.3 AufenthG..

Content. In the following, I focus on the new rule in the integration law most relevant for this study, namely the introduction of residential restraints ("Wohnsitzauflagen"), which introduced mandatory restrictions on the choice of residence for recognized refugees.[10]

Residential restraints state that there is a compulsory, general obligation to have the residence for three years after the time of recognition in the *federal state* in which the asylum procedure took place. Further, the local authorities have the power to further restrict the residence to a certain *municipality* or *county* during these three years and to name certain municipalities to which refugees are not allowed to move.

Unaffected by the new rule are unaccompanied minors and people who themselves, whose partner or whose underage child, have a job subject to social security contributions with at least 15 weekly working hours, who are in job training or who are pursuing studies at the time of their asylum recognition. Further, a refugee can request to be freed of residential restraints if they, their partner, or an underage child can obtain the above-mentioned engagement at a different location or if their partner or underage child is living somewhere else.

Implementation. While having the residence for three years after the time of recognition in the *federal state* in which the asylum procedure took place, is compulsory for all refugees, the decision whether the refugee is further restricted to a certain *municipality* or *county*, so whether they additionally get imposed a *local* residence restriction, lies at the discretion of the local authorities. The local authorities themselves act according to how their respective federal state implemented the new law into state law. As noted in Lehrian (2018), there exist considerable differences across federal states in how the law was implemented. As shown in Table 3.A.2, there is a large group of federal states which never im-

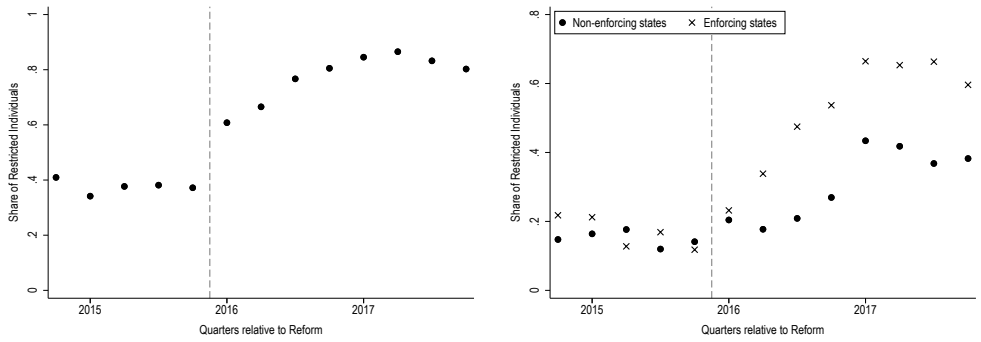
10. A detailed summary of all novelties of the integration law can be found in Lehrian and Mantel (2016).

plemented local residence restrictions into state law while other federal states either allowed or even urged the local authorities to make use of local residence restrictions. In the following, I will label those federal states which incorporated local residence restrictions quickly into state law after the reform, namely Bavaria, Baden-Wuerttemberg, Northrhine-Westphalia and Saarland, as "enforcing states" while I will label the remaining states as "non-enforcing states".

As a result of the reform, I expect to see two distinct patterns. On the one hand, I expect to see an increase in the share of individuals who report that they have been imposed a residence restriction after the reform. On the other hand, I expect that most of this increase in residence restrictions in the "enforcing" states is driven by an increase in local residence restrictions. I check this by making use of a survey question in the IBS Survey of Refugees that asks the refugees whether their location choice is restricted to a certain location, restricted to a certain federal state or not restricted at all.[11]

Figure 3.2.1 shows that, indeed, the expected patterns can be seen in the data. Figure 3.2.1a shows the share of individuals that reported at least once that they are subject to a residence restriction. For individuals who were recognized prior to the reform this share is around 40 percent and it increases quickly to around 90 percent after the reform. This is intuitive as the reform introduced state-level residence restrictions for all refugees. Additionally, Figure 3.2.1b depicts the share of individuals who report being only allowed to live in a certain location. It shows that prior to the reform, this share of individuals having a local residence restriction was approximately the same in both enforcing and non-enforcing states

11. As I expect large measurement error for this variable, I am not using it in the main analysis. Reasons for measurement error could be potential misreporting as individuals either misunderstand the question or simply do not know that they are restricted or are not aware of it since they never tried to change their location. Moreover, it is possible that individuals who are restricted but who are able to change their location since they fulfill the exemption rules might report that they have no restriction, even though technically they were subject to the residence restriction.



(a) Restricted location choice **(b)** Restricted location choice on local level

Figure 3.2.1. Prevalence of Residence Restrictions

Notes: The figure illustrates the prevalence of residence restrictions among refugees in Germany over time. Panel 3.2.1a displays the share of refugees in the sample that report having either a state or local residence restriction. It shows that after the integration law reform, the share of refugees reporting being subject to a state or local residence restriction increased substantially. Panel 3.2.1b shows the share of refugees in the sample reporting a local residence restriction. It shows that after the integration law reform, refugees in enforcing states reported more often that they are subject to a local residence restriction than refugees in non-enforcing states. Due to the small sample size, I bin observations before 2014 Q4 in the first and observations after 2017 Q4 in the last quarter.

but then increased significantly more for individuals recognized after the reform in enforcing than in non-enforcing states.

This naturally raises the question of whether the visible differences in the enforcement of local residence restrictions affect the frequency of across-county relocations. As the survey data contains information on the county of residency, I am able to construct a measure indicating whether an individual relocated from one county to another between 2016 and 2017, given that the individual was interviewed in both years. Around 54 % of my estimation sample, 1,572 individuals, appear in the data both in 2016 and 2017.[12] Of these 1,572 individuals around 6.7 % relocated from one county to another between 2016 and 2017.

First, I check whether those individuals that report a local residence restrictions relocate less frequently. Looking at individuals in both enforcing and non-enforcing states, of those that never reported a local residence restriction, around

12. I provide detailed information on the construction of the estimation sample in Section 3.3.

8.8 % changed their residence county, while only 4.1% of those that reported a local residence restriction did so. As a result, I can also see differences in the frequency of moving on the state level. Before the reform, refugees in enforcing states are slightly more likely (0.7 percentage points) to move than those in non-enforcing states. However, this flips for those recognized after the reform, as those in enforcing states become less likely (1.8 percentage points) to move than those in non-enforcing states. This highlights that, on aggregate, the stricter enforcement of local residence restrictions in enforcing states was followed by a reduction in across-county relocations. I will come back to the results on moving when discussing my empirical results in Section 3.5.

Potential Effects on Integration Outcomes. Ex-ante, it is not entirely clear how the stronger enforcement of local residence restrictions should affect the economic and social integration of refugees. The degree to which a local residence restriction indeed restricts a refugee depends on various factors. They should not be directly affected if they have either no intention to relocate from their initial residence or if they want to change their residence while having a job at the desired new location as they are then exempted from the local residence restriction.

However, refugees that want to change their location without having a job at the new location are restricted. They either end up stuck in their initial location or increase their job-finding efforts to be able to move to their desired location. If they end up stuck in their initial location they have a longer exposure to local economic conditions which could reinforce the effects of initial local economic conditions on the integration. Whether local residence restrictions have positive or negative effects on integration, therefore, depends both on the number of people that end up staying longer in the initial location and the number of individuals that increase their job search effort to find a job at their desired location. Moreover, for those individuals that stay longer in their initial location, I would expect to see different results depending on the local labor market conditions a refugee is

facing. Eventually, it is an empirical question of which of the effects dominates and I will analyze the effects of the stronger enforcement of local residence restrictions in Section 3.5.

3.3 Data and Sample

This section describes the data sources used in this study and provides information on the composition of the final estimation sample.

IBS Survey of Refugees. The main data source of this project is the IAB-BAMF-SOEP (IBS) Survey of Refugees, a representative survey of refugees and their household members, who arrived in Germany between January 2013 and June 2019 and consequently filed an asylum application. The survey is conducted yearly by the Institute for Employment Research (IAB) of the Federal Employment Agency (BA), the Federal Office for Migration and Refugees (BAMF) and the Socio-Economic Panel (SOEP) at the German Institute for Economic Research in Berlin (DIW). It aims to obtain comprehensive information on incoming refugees in order to enable policy analysis and research regarding their integration into the German society. The collected information covers a wide range of topics and includes, among others, information on the academic and professional education, current and former working experience, language skills, and also information on living conditions and social interactions of the surveyed refugees.

The IBS Survey of Refugees contains data from four survey waves, covering the years 2016 to 2019. Information is available on 4,465 adults for 2016, on 5,593 adults for 2017, on 4,344 adults for 2018, and on 3,856 adults for 2019. In total, the survey data includes 8,153 persons who have been interviewed at least once.[13] Further details on the exact sampling and interview procedures can be

13. Out of these 8,153 individuals, 2,771 persons have taken part in the survey once by 2019, 2,036 twice, 2,105 three times and 1,241 four times.

found in Brücker (2017) and Kühne, Jacobsen, and Kroh (2019). The IBS Survey of Refugees is perfectly suited for this analysis as it provides me with all necessary information to analyze the integration of refugees in Germany. Most importantly, it contains the date of the asylum process decision, information on the initial place of residence, and information on the economic and social integration of the refugee.[14]

Sample Composition. To construct the final estimation sample, I restrict the sample to individuals aged 18 to 65 in 2017 who had their positive asylum decision anytime between September 2014 and December 2017.[15] Additionally, I only consider individuals for whom I have information on the first location in Germany and non-missing values for the main employment and social integration variables.[16] I define an individual as being employed if they have found either full- or part-time work or if they are doing a vocational apprenticeship in 2017. For the social integration outcomes, I use information from two questions. The first question I use is the self-reported frequency of social contact with Germans.[17] The second question I use measures whether a refugee has found new German acquaintances since their arrival.[18] I use these two questions as those are the only

14. Following Aksoy, Poutvaara, and Schikora (2020), I use the information on the longest place of residence before the first interview as a proxy for the initial place of residence. Given that the first interview was conducted relatively shortly after the arrival in Germany and given that asylum seekers are relatively immobile before their asylum decision, I expect this procedure to correctly identify the initial place of residence for most individuals.

15. Moreover, I drop all individuals with missing birth years and unreasonable arrival dates (before Jan 2012).

16. Further, I check that all individuals have non-missing information for the observable personal characteristics summarized in Table 3.3.1.

17. The question asks “How often do you spend time with Germans?” and has six different response options ranging from "1: Daily" to “6: Never”. I recode the variable such that a higher value implies more time spent with Germans and standardize the variable by subtracting the mean and dividing by the standard deviation. The original distribution of the variable can be found in Figure 3.B.1a.

18. The variable asks about the number of new German acquaintances, due to the large range of answers I recode the variable to a binary indicator measuring whether a refugee has found new German acquaintances or not. The original distribution of the variable can be found in Figure 3.B.1b.

questions asking about social interactions with Germans and therefore constitute a good measure for the social integration of a refugee.

In principle, the IBS Survey of Refugees contains data spanning all four years from 2016 to 2019. However, the number of non-missing responses for the main outcome variables varies considerably by year. The question of whether a refugee has found new German acquaintances since their arrival is only asked once when the refugee is first surveyed. For the other two variables, most responses are recorded for the year 2017, while the other years exhibit a much lower number of responses.[19] I stick to using the responses of the 2,858 individuals that were interviewed in the year 2017 in the main specifications to avoid potential bias coming from the increasing non-response over time.[20] To check the robustness of my results, I also run my analysis on a sample in which I pool the responses of all individuals across all years.

19. The number of available non-missing observations by year is: 2016: 1,934, 2017: 2,858, 2018: 2,098, 2019: 1,804.

20. As the variable measuring whether a refugee has found new acquaintances is only asked once in the first interview, I use either the information from 2016 or 2017 while controlling for that year in the regression.

Table 3.3.1. Descriptive Statistics for Estimation Sample

	Treated	Control	Whole Sample
Age	32.60	33.0	32.81
	(0.27)	(0.26)	(0.19)
Age at Arrival	30.59	30.93	30.77
	(0.27)	(0.26)	(0.19)
Number of Children	1.91	1.88	1.90
	(0.06)	(0.05)	(0.04)
Has Children	0.60	0.61	0.60
Female	0.35	0.38	0.37
Married	0.57	0.57	0.57
Education: Primary	0.38	0.40	0.39
Education: Secondary	0.41	0.42	0.41
Education: Tertiary	0.21	0.18	0.19
Nationality: Syrian	0.61	0.54	0.58
Nationality: Afghan	0.08	0.13	0.11
Nationality: Iraqi	0.15	0.13	0.14
Nationality: Other	0.16	0.20	0.17
N	1,355	1,503	2,858

Notes: The table displays descriptive statistics for the estimation sample. I report the means of relevant personal characteristics of the refugees with standard errors in brackets. A more detailed analysis of the differences in the composition between refugees in treated and control states and how these differences change over time can be found in Table 3.4.2.

I summarize the main observable characteristics of the refugees in my sample in Table 3.3.1. The sample consists of relatively young individuals with an average age of around 33 years of which around 60 % are male. 60 % of the refugees have at least one child and the average number of children in the sample is around 2. In terms of nationalities, the majority, around 58 %, are Syrian, followed by Iraqi at 14 % and Afghan at 11 %. Finally, around 20 % have obtained a tertiary education, around 40 % secondary and 40 % primary education as their highest education level.

Table 3.3.1 also summarizes the main observable characteristics separately by treated and control states. On average, the two groups only differ in composition with regard to nationalities. I discuss differences in the composition between the two groups in great detail in Section 3.4.2 in which I also analyze differences in the composition between treated and control states before and after the reform in Table 3.4.2.

Unemployment Data. Besides survey data from the IBS Survey of Refugees, I also use administrative information on county unemployment rates. Using the information on the initial place of residence in the survey, I merge to each refugee the initial local county unemployment rate in 2016 that he or she was exposed to. The information on the county unemployment rate is obtained from the Federal Employment Agency (BA) in Germany.

3.4 Empirical Strategy

In this section, I discuss the empirical strategy used in this study. I start by presenting the empirical design and then discuss the necessary assumptions for identification and their validity in the context at hand.

3.4.1 Empirical Design

I start my analysis by making use of the quasi-random initial allocation of refugees across Germany to identify the effect of local labor market conditions on the economic and social integration of refugees. I do so by regressing the outcomes of interest on the unemployment rate in the county the refugee was initially allocated to:

$$Y_{i,c} = \alpha + \beta \text{ High-Unemployment}_c + \rho X_i + \epsilon_{i,c}, \tag{3.4.1}$$

where $Y_{i,c}$ is the outcome of interest of refugee i who was initially allocated to county c. High-Unemployment$_c$ is a dummy indicating whether the initial county

had a high, above-median, level of unemployment.[21] Finally, X_i controls for observable characteristics of the refugee, including age, gender, nationality, education, marital status and number of children.

For the coefficient β to correctly measure the effect of the labor market condition in the initial location on the integration outcome of the refugee, I need to assume that the unemployment rate in the initial location is indeed randomly assigned and not correlated with other factors that can influence the refugees integration outcome. I discuss the validity of this assumption in the next subsection.

In a second step, I examine how the stricter enforcement of local residence restrictions affects the integration outcomes of refugees. The quasi-random initial allocation of refugees across federal states in Germany, together with the across-state differences in the enforcement of local residence restrictions, gives rise to a difference-in-differences (Diff-in-Diff) setup. The idea is to compare the outcomes of individuals whose asylum decision was made after the reform and who were allocated to enforcing states, and therefore subject to stricter residence restrictions, to those that were allocated to non-enforcing states with less strict residence restrictions. Again, I will discuss the required identifying assumptions for performing the Diff-in-Diff in the next subsection.

I estimate the following regression equation:

$$\begin{aligned} Y_{i,s,c,t} &= \alpha + \beta \text{ High-Unemployment}_c + \gamma_s + \lambda_t \\ &\quad + \delta \text{ Post Reform}_t \times \text{Enforcing}_s + \rho X_i + \epsilon_{i,s,c,t}, \end{aligned} \tag{3.4.2}$$

where $Y_{i,s,c,t}$ denotes the integration outcome of individual i, being allocated into county c in state s, with an asylum decision made at time t. γ_s are binary indicators for the state of initial allocation s and λ_t are binary indicators for the asylum decision quarter t. The interaction Post Reform$_t$ × Enforcing$_s$ indicates whether

21. I check the robustness of my results by also estimating a version using the unemployment rate in the county instead of the high-unemployment dummy.

the individual was "treated", meaning whether they were allocated to an enforcing state and had their asylum process decision after the implementation of the reform. High-Unemployment$_c$ is still the indicator for whether the refugee was allocated to a county with a high level of unemployment and X_i controls for the aforementioned observable characteristics of the refugee.

The coefficient of interest in this regression is δ, which measures the intention-to-treat (ITT) effect of residence restrictions on the integration of refugees. Since not all refugees in the enforcing states after the reform also end up receiving a local residence restriction, the coefficient measures the average effect of the stricter enforcement of the reform for both restricted and unrestricted individuals in the enforcing states. This can lead me to underestimate the effect of local residence restrictions. As shown before, the survey data also contains a self-reported measure of being subject to residence restrictions. However, I refrain from using it in my analysis as I expect large measurement error for this variable. While it can give us a good overview of the general extent of local residence restrictions (see Figure 3.2.1), it cannot tell us for sure if a refugee is subject to a residence restriction. First, for some refugees, the information is simply missing in the data. Second, I expect much misreporting as individuals either misunderstand the question or simply do not know that they are restricted or are not aware of it since they never tried to change their location. Moreover, it is possible that individuals who are restricted but who are able to change their location since they fulfill the exemption rules might report that they have no restriction, even though technically they were subject to the residence restriction.

Finally, as discussed before, I expect to see heterogeneous effects of the stricter enforcement of local residence restrictions depending on local labor market conditions. I, therefore, also estimate a version of Equation 3.4.2 interacted with the high-unemployment dummy.

3.4.2 Identification

The validity of my proposed empirical design depends crucially on whether the required identifying assumptions are likely to hold. In this subsection, I discuss which assumptions are necessary for identification and provide evidence why they are likely to hold in this setting.

Initial Allocation. To identify the effect of initial local labor market conditions on the integration of the refugees, I need to assume that the initial allocation of refugees into states and counties is indeed quasi-random and not influenced by the personal characteristics of the refugee that could also directly affect their integration outcomes. As discussed in Section 3.2, incoming refugees get allocated to a federal state which has not yet fulfilled its refugee quota. Thereafter, they are further allocated to counties within the federal state with the help of local refugee allocation quotas.[22] Importantly, neither the refugee nor the federal states can veto the allocation of a refugee. For the allocation on counties within the federal state, the same principle is likely to hold. However, I cannot completely rule out the possibility that some sort of screening is happening at the county level. Nevertheless, if there is a screening of refugees on the local level, it could be only done based on observable characteristics of the refugee. This implies that I can check whether there is a different selection of refugees into counties depending on the local labor market conditions using the information of the refugees in my sample.

I perform a randomization check in Table 3.4.1 by regressing the unemployment rate in the county of the initial allocation on observable characteristics of the refugees at the time of their arrival. If the initial allocation of refugees was indeed quasi-random with respect to the local labor market conditions, I expect to

22. More information on these local quotas can be found in Table 3.A.1. The only exception to this rule are family reunions, in which the refugees are able to move to the area where previously-arrived family members are residing.

Table 3.4.1. Randomization Check

	Unemployment Rate	
	(1)	(2)
Age at Arrival	0.016***	0.011**
	(0.006)	(0.005)
Female	0.052	0.042
	(0.097)	(0.078)
Married	0.143	0.165
	(0.132)	(0.104)
Has Children	−0.706***	−0.557***
	(0.157)	(0.125)
Education: Secondary	−0.064	−0.107
	(0.107)	(0.082)
Education: Tertiary	−0.044	−0.173*
	(0.134)	(0.103)
Nationality: Syrian	−0.140	−0.052
	(0.123)	(0.094)
Nationality: Afghan	0.442**	0.341***
	(0.179)	(0.130)
Nationality: Iraqi	0.210	0.158
	(0.168)	(0.132)
State FE	No	Yes
Observations	2,858	2,858
Adj. R^2	0.0130	0.4263

Notes: The table displays the coefficients from regressing the unemployment rate in the county of the initial allocation on observable characteristics of the refugees at the time of their arrival. I perform this regression to check whether the composition of allocated refugees in a county differs by the county's unemployment rate. Heteroskedasticity robust standard errors are reported in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

see that observable characteristics of the refugee do not significantly predict the unemployment rate in the county of the initial allocation. Column (1) displays regression results *without* state fixed effects and column (2) results with state fixed effects. While most covariates do not predict the unemployment rate, the age at arrival, having children, and being Afghan has a statistically significant effect on the unemployment rate in the county of the initial allocation. These effects are, however, not economically significant. Being ten years older is associated with a

0.11 percentage point higher unemployment rate, having children with a 0.56 percentage point lower unemployment rate, and being Afghan with a 0.34 percentage point higher unemployment rate in the initial location.

Another implication of the quasi-random initial allocation is that when adding control variables to the regression equations, the coefficients of the effect of the unemployment rate in the initial location should not be significantly changed. Indeed, I find in Section 3.5 that the regression results are basically unchanged when additionally controlling for observable characteristics of the refugees. Taken together, I am confident that the initial allocation was indeed quasi-random and not significantly influenced by the observable characteristics of the refugee.

However, it is important to acknowledge that some of the observed effects in my analysis may not be directly attributable to labor market conditions alone. It is possible that local labor market conditions are associated with other variables that also influence the integration of refugees. For instance, these conditions might be correlated with local attitudes towards refugees or immigration, which can play an important role in shaping integration outcomes. Unfortunately, information regarding these attitudes is currently unavailable at the county level.[23] Obtaining information about local attitudes towards refugees and immigration at the county level would be valuable in the future when it comes to designing more targeted and effective policies as it allows policymakers to tailor their interventions better to the needs of the refugees.

Difference-in-Differences. For the causal interpretation of the Diff-in-Diff regression estimates in equation (3.4.2), two main assumptions have to hold. First, it has to be the case that the composition of refugees in the treatment and control

23. Aksoy, Poutvaara, and Schikora (2020) construct a Migrant Acceptance Index at the level of the federal state using data from the European Social Survey (ESS). They include this index as an additional explanatory variable in their analysis. Their findings show that while negative attitudes towards migrants also had a negative effect on integration, the county unemployment rate remains a significant negative predictor for successful integration.

states did not change differently over time. I check this assumption in Table 3.4.2 by displaying the means of observable characteristics before and after the reform for refugees in treated and control states. Additionally, I calculate the Diff-in-Diff of these sample means. Given the institutional framework described in Section 3.2, I expect to see no significant differences in the composition of refugees between the two groups of states. However, given that there are certain branches of the Federal Office for Migration that specialize in certain nationalities, it could be possible to see slight differences in the distribution of nationalities across the two groups of states.[24]

Table 3.4.2. Changes in Composition of Treatment and Control Group

		Before Reform			After Reform		
	Diff-in-Diff	Treat	Cont	Diff	Treat	Cont	Diff
Age	−0.75	34.48	34.31	0.18	32.06	32.64	−0.57
		(0.58)	(0.54)		(0.31)	(0.29)	
Number of Children	−0.34*	2.02	1.73	0.29*	1.88	1.93	−0.05
		(0.12)	(0.11)		(0.06)	(0.06)	
Has Children	−0.09*	0.63	0.57	0.06	0.59	0.62	−0.03
Female	0.00	0.28	0.31	−0.03	0.37	0.40	−0.03
Married	0.01	0.61	0.62	−0.01	0.55	0.56	−0.00
Education: Primary	0.02	0.29	0.33	−0.04	0.40	0.42	−0.02
Education: Secondary	−0.07	0.44	0.41	0.04	0.40	0.42	−0.03
Education: Tertiary	0.04	0.26	0.26	0.00	0.20	0.15	0.04***
Nationality: Syrian	−0.04	0.80	0.70	0.10***	0.56	0.50	0.06***
Nationality: Afghan	−0.04**	0.01	0.04	−0.02*	0.10	0.16	−0.06***
Nationality: Iraqi	0.03	0.08	0.07	0.01	0.17	0.14	0.03**
N		302	325		1,053	1,178	

Notes: The table shows how the composition of the treatment and control group changed over time. The first column displays a Diff-in-Diff estimate for each covariate, measuring whether the composition with regard to that covariate changed differently between treatment and control group over time. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

The results in Table 3.4.2 show that there indeed exist significant differences in the composition of nationalities between the two groups of states, both before

24. As mentioned before, more detailed information on the specialization of each branch of the Federal Office for Migration is not publicly available.

and after the reform. Treated states exhibit a higher share of Syrians, while control states have a higher share of Afghan nationals among the refugees. Looking at the Diff-in-Diff of the sample means, however, I find that differences in the composition between treated and control states remained relatively stable over time. Among the observed characteristics, only one changed significantly at the 5 % significance level. The difference in the share of Afghans between treated and control states became larger after the reform. However, all in all, while the overall composition of the refugees clearly changed over time, it did so in a similar way for both treatment and control states.

The second assumption that has to hold for the validity of the proposed Diff-in-Diff approach is the so-called common trend assumption which assumes that the integration outcomes of refugees in treatment and control states would have evolved similarly over time in the absence of the reform. This implies that all differences in integration outcomes over time between treatment and control states are due to the stricter enforcement of local residence restrictions in the treatment states. While this assumption is not testable, the fact that the differences in the composition of refugees stayed relatively stable over time gives me confidence that the outcomes of refugees in treated and control states would have evolved similarly in the absence of the reform. One testable implication of the common trend assumption is that treatment and control states exhibit parallel trends in outcomes prior to the integration reform. To test this, I also estimate an event study version of the simple Diff-in-Diff in Equation 3.4.2. In this dynamic version, I replace the post-reform dummy with dummies for the half-years relative to the reform. In the next section, I present the regression estimates of both the simple Diff-in-Diff and the event study Diff-in-Diff regression.

3.5 Empirical Results

In this section, I present the empirical results. I start by showing the results for the effects on employment and then present the results for the effects on the social integration of refugees. I conclude the section by investigating effects on across-county relocations of the refugees and by displaying the event study Diff-in-Diff estimates.

Effects on Employment. The quasi-random initial allocation of refugees across Germany enables me to directly identify the effects of local labor market conditions in the initial location on the employment status of refugees in Germany. In the first column of Table 3.5.1, I show that being allocated into a county with a high-unemployment rate reduces the likelihood of employment significantly. Being allocated into a county with an above-median unemployment rate reduces the likelihood of being employed either full or part-time in 2017 by around 5 percentage points.[25] As expected from the quasi-random initial allocation of the refugees, I show in column (2) that controlling for personal characteristics of the refugees does not significantly change the regression estimates.

The finding that initial labor market conditions affect a refugee's probability of finding employment is highly policy-relevant. It shows that the quasi-random initial allocation of refugees in Germany, while ensuring a more uniform distribution of refugees across the country, directly induces inequalities in employment outcomes between refugees. This result replicates the findings of Aksoy, Poutvaara, and Schikora (2020) and is in line with many papers showing that local labor market conditions affect labor market outcomes of immigrants and refugees (Åslund and Rooth, 2007; Godøy, 2017; Azlor, Damm, and Schultz-Nielsen, 2020).

25. I calculate the median unemployment rate among all 401 counties for the year 2016 and define a county to have high-unemployment if its unemployment rate is above the median unemployment rate of 5.4 %.

Making use of differences in the enforcement of local residence restrictions across federal states after the integration reform in 2016, I then estimate the effects of the stricter enforcement of local residence restrictions on the likelihood of being employed. I present the Diff-in-Diff estimates in columns (3) and (4). The estimates show that being subject to stricter enforcement of local residence restrictions has a small positive, but insignificant, effect on the likelihood of employment in 2017. However, as discussed before, local residence restrictions can increase the exposure of a refugee to the local economic conditions in the initial location. One could therefore expect to see heterogeneous effects of the stricter enforcement of local residence restrictions by initial local labor market conditions. Interacting regression equation 3.4.2 with the high-unemployment dummy, I indeed find substantial heterogeneity of the treatment effect. Columns (5) and (6) show that refugees allocated into low-unemployment counties experience a significant increase in their likelihood of finding employment of around 17 percentage points while those being allocated into high-unemployment counties see a decrease in their likelihood of being employed of around 8 percentage points. The detected effects are rather large, however, given the large standard errors, the confidence intervals are relatively wide. As a result, caution should be exercised when interpreting the precise magnitude of the effects. Nonetheless, the results clearly show that while refugees in low-unemployment counties benefited from the reform, those in high-unemployment counties were harmed.

Table 3.5.1. Full or Part-time Employment in 2017

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.053***	−0.054***	−0.035**	−0.037**	−0.132*	−0.115*
	(0.014)	(0.014)	(0.017)	(0.017)	(0.068)	(0.065)
Post Reform × Enforced			0.043	0.037		
			(0.040)	(0.038)		
Post Reform × Enforced × Low Unemployment					0.198***	0.170**
					(0.073)	(0.069)
Post Reform × Enforced × High Unemployment					−0.092*	−0.083
					(0.056)	(0.052)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.005	0.090	0.045	0.126	0.049	0.130

Notes: The table displays the regression results for the likelihood of being employed either full or part-time in 2017. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the high-unemployment dummy. Full regression results can be found in Table 3.C.4. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

I explore the robustness of these findings using the continuous unemployment rate instead of the high-unemployment dummy in Table 3.C.1 and obtain similar results. A one percentage point higher unemployment rate decreases the likelihood of having found a job by around 1 percentage point. This finding is similar to Azlor, Damm, and Schultz-Nielsen (2020), who find that a one percentage point higher employment rate increases the employment probability of refugees by 0.5 to 0.6 percentage points two to four years after their arrival in Denmark. Moreover, columns (5) and (6) of Table 3.C.1 again show that there is a positive effect of the stricter enforcement of local residence restrictions for refugees in low-unemployment counties and that this effect disappears the higher the unemployment rate in the initial county. Finally, I explore the robustness of my results to using the pooled sample, which uses information from all sample years, as my estimation sample in Table 3.C.7. Again, I obtain very similar results.

Effects on Social Integration. While being employed is arguably an important indicator of successful integration, the extent of social interaction between refugees and native Germans is also highly relevant when assessing the progress of integration. Using the rich survey data at hand, I can investigate how the initial local labor market conditions and the stricter enforcement of local residence restrictions affected the refugees' social integration using two different measures of social integration.

The first measure is the self-reported frequency of social contact with Germans.[26] The second measure is the likelihood of having found new German acquaintances. Using the same approach as for the effects on employment, I report the regression results for the frequency of social contact with Germans in Table 3.5.2 and for the likelihood of having found new German acquaintances in Table 3.5.3.

As in the case of employment, I find in columns (1) and (2) of Table 3.5.2 that bad local labor market conditions significantly decrease the frequency of social contact with Germans. Being allocated to a county with an above-median unemployment rate is associated with an around 0.25 of a standard deviation lower frequency of social interactions with Germans in 2017. Again, as expected, controlling for additional covariates in column (2) does not significantly change the coefficient. Turning to the Diff-in-Diff estimates for the effects of stronger enforcement of local residence restrictions in columns (3) and (4), I find that being subject to stricter enforcement of local residence restrictions has a significant positive effect on the frequency of social contact with Germans in 2017. However, columns (5) and (6) show suggestive evidence that this effect is again heterogeneous by local labor market conditions. Interacting regression equation 3.4.2

26. The question asks "How often do you spend time with Germans?" and has six different response options ranging from "1: Every day" to "6: Never". I recode the variable such that a higher value implies more time spent with Germans and standardize the variable by subtracting the mean and dividing by the standard deviation.

with the high-unemployment dummy indicates that refugees allocated into low-unemployment counties seem to experience a larger increase in the frequency of social contact with Germans than those being allocated into high-unemployment counties.

Table 3.5.2. Frequency of Social Contact with Germans in 2017

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.250***	−0.247***	−0.163***	−0.163***	−0.107	−0.067
	(0.037)	(0.036)	(0.046)	(0.044)	(0.128)	(0.121)
Post Reform × Enforced			0.148*	0.148*		
			(0.088)	(0.086)		
Post Reform × Enforced × Low Unemployment					0.240*	0.196
					(0.143)	(0.137)
Post Reform × Enforced × High Unemployment					0.053	0.092
					(0.136)	(0.134)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.015	0.072	0.058	0.116	0.059	0.118

Notes: The table displays the regression results for the frequency of social interactions with Germans in 2017. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the high-unemployment dummy. Full regression results can be found in Table 3.C.5. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Finally, I investigate the effects on the likelihood of having found new German acquaintances.

Column (1) and (2) of Table 3.5.3 show the same pattern as before: Being allocated to a county with bad local labor market conditions significantly decreases the likelihood of having found new German acquaintances. Being allocated to a county with an above-median unemployment rate is associated with an around 8 % lower likelihood of having found new German acquaintances. Again, as before, controlling for observable characteristics in column (2) does not significantly change the coefficient.

The estimates for the effects of stronger enforcement of local residence restrictions in columns (3) and (4) show that being subject to stricter enforcement of local residence restrictions has no significant effect on the likelihood of having found new German acquaintances. Looking at the effect separately for low and high-unemployment counties does not reveal any significant differences. Nevertheless, the results in columns (5) and (6) again provide suggestive evidence that refugees in low-unemployment counties benefit while those in high-unemployment counties are harmed, as the estimate is around 3.2 % for refugees allocated to low-unemployment counties and around -1.2 % for refugees allocated to high-unemployment counties.

Table 3.5.3. New German Acquaintances since arrival

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.080***	−0.077***	−0.078***	−0.077***	−0.049	−0.047
	(0.017)	(0.017)	(0.021)	(0.021)	(0.059)	(0.059)
Post Reform × Enforced			0.011	0.010		
			(0.039)	(0.039)		
Post Reform × Enforced × Low Unemployment					0.035	0.032
					(0.064)	(0.063)
Post Reform × Enforced × High Unemployment					−0.018	−0.012
					(0.057)	(0.058)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.021	0.045	0.033	0.055	0.034	0.055

Notes: The table displays the regression results for the likelihood of having found new German acquaintances since arrival. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the high-unemployment dummy. Full regression results can be found in Table 3.C.6. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

I explore the robustness of the estimates for the social integration outcomes using the unemployment rate instead of the high-unemployment dummy in 3.C.2 and Tables 3.C.3. As for the employment outcome, I obtain similar results using the unemployment rate. Finally, I explore the robustness of my results to using the

pooled sample, which uses information from all sample years, as my estimation sample in Table 3.C.8 and Table 3.C.9. Again, I obtain very similar results.

Effects on Across-County Relocations. Motivated by the finding that the effects of stricter enforcement of local residence restrictions are heterogeneous by local labor market conditions in the initial location, I investigate whether this heterogeneity also exists with respect to across-county relocations. As discussed in Section 3.2.2, the stricter enforcement of residence restrictions is, overall, associated with a decrease in across-county relocations.[27] Columns (1) and (2) of Table 3.5.4 show, however, that this reduction is not statistically significant, potentially due to the small remaining sample size. Looking at the effect separately for refugees allocated to low- and high-unemployment counties in columns (3) and (4), reveals that while there is some suggestive evidence that the stronger enforcement of local residence restrictions reduced the likelihood of across-county relocations in low-unemployment counties, there is no evidence for such a reduction in high-unemployment counties. However, as the coefficients for low- and high-unemployment counties are not statistically significantly different from each other, I want to refrain from over-interpreting these results.

Nonetheless, if these differences indeed exist, there could be several possible explanations for them. Firstly, considering the lack of positive employment effects for individuals in high-unemployment counties, as shown in Table 3.5.1, it seems unlikely that those refugees circumvented residence restrictions by moving while having a job in the desired location. However, it is possible that refugees in high-unemployment counties in enforcing states who were not subject to a local residence restriction after the reform moved more frequently.[28] It would be

27. As noted before, I measure across-county relocations by constructing a measure indicating whether an individual relocated from one county to another between 2016 and 2017, given that the individual was interviewed in both years. I can calculate the measure for around 54 % of my estimation sample, 1,572 individuals, who appear in the data both in 2016 and 2017.

28. Keep in mind that the reform only increased the likelihood of being treated and therefore not all individuals in enforcing states were treated after the reform.

Table 3.5.4. Across County Relocations between 2016 and 2017

	(1)	(2)	(3)	(4)
High Unemployment	0.004	0.003	−0.005	−0.009
	(0.015)	(0.015)	(0.033)	(0.033)
Post Reform × Enforced	−0.026	−0.029		
	(0.026)	(0.026)		
Post Reform × Enforced × Low Unemployment			−0.042	−0.039
			(0.044)	(0.044)
Post Reform × Enforced × High Unemployment			0.016	0.011
			(0.038)	(0.038)
Controls	No	Yes	No	Yes
N	1,572	1,572	1,572	1,572
Adj. R^2	-0.001	0.005	-0.001	0.132

Notes: The table displays the regression results for the likelihood of across county relocations between 2016 and 2017. Columns (1) and (2) show the regression results from equation 3.4.2 interacted with the high-unemployment dummy. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

intriguing to analyze whether these individuals relocated to better or worse locations compared to those not subjected to stricter enforcement. Unfortunately, due to the limited sample size, I am unable to conduct this analysis.

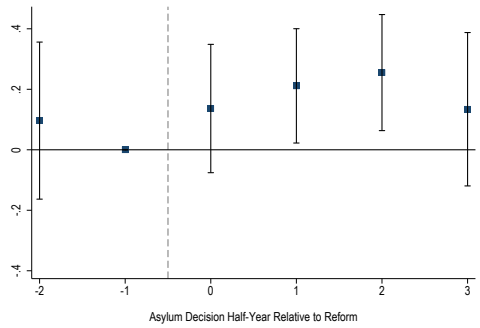
Moreover, the absence of evidence for a reduction in relocations in enforcing states could be attributed to the fact that the penalty for not complying with local residence restrictions consists "only" of the withdrawal of social benefits. It could therefore be the case that individuals in high-unemployment counties simply accept the withdrawal of social benefits and still move to their desired location. Additionally, since the withdrawal of social benefits lies ultimately at the discretion of local authorities, I cannot rule out that they may be more lenient towards refugees relocating from high-unemployment counties, resulting in fewer sanctions for breaching local residence restrictions.

Event Study Diff-in-Diff. Finally, I also investigate the dynamics of the treatment effects by estimating a dynamic version of equation 3.4.2 interacted with the high-unemployment dummy. I do so by replacing the post reform dummy with dummies for each half-year relative to the reform.[29] This allows me to evaluate possible time dynamics of the treatment effect. Further, it also enables me to provide more evidence for the plausibility of the common trend assumption. One implication of the common trend assumption is that I should see no different evolution in outcomes between treated and control states before the reform, meaning that the pre-reform event study Diff-in-Diff estimates should be non-significant.

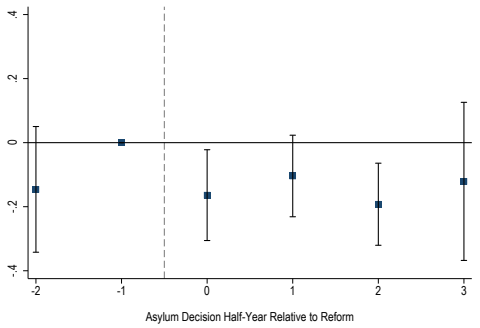
I graphically display the event study Diff-in-Diff estimates in Figure 3.5.1 using the half-year before the reform (July 2015 - Dec 2015) as the baseline period. The results for the effect on having a full- or part-time employment in 2017 are displayed in Figures 3.5.1a and 3.5.1b. Both figures show that I find significant effects for the half-years 0, 1 and 2, so for all refugees with an asylum decision between January 2016 and June 2017. For refugees with an asylum decision in the last half-year of the sample period (July 2017- Dec 2017), I cannot detect any significant effects. As for these refugees the measurement of the outcome variable overlaps with their asylum decision, it might not come as a surprise that the treatment has not produced any noticeable effects thus far.

Turning to the event study Diff-in-Diff estimates for the effects on social contact with Germans in 2017 in Figures 3.5.1c and 3.5.1d, I cannot detect any significant effects either before or after the reform. This finding is in line with the insignificant static Diff-in-Diff results in Table 3.5.2. Similarly, for the likelihood of having found new German acquaintances, I also cannot detect any significant effects either before or after the reform in Figures 3.5.1e and 3.5.1f.

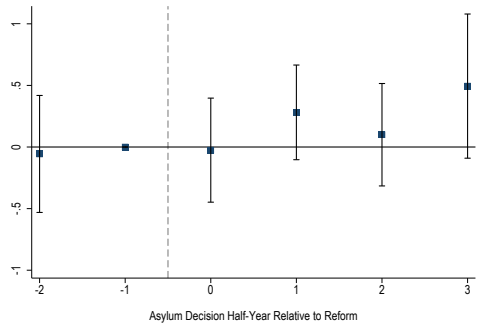
29. To enhance the reliability of my analysis given the limited sample size, I choose to aggregate the observations on the half-year level instead of the quarterly level. This approach allows me to maintain a sufficient level of statistical power for the analysis at the more granular level.



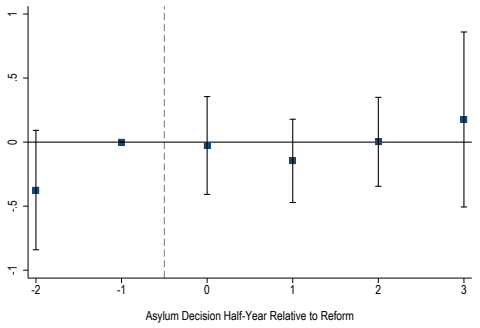
(a) Employment: Low-Unemployment



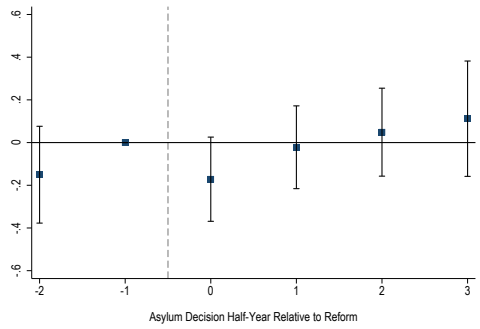
(b) Employment: High-Unemployment



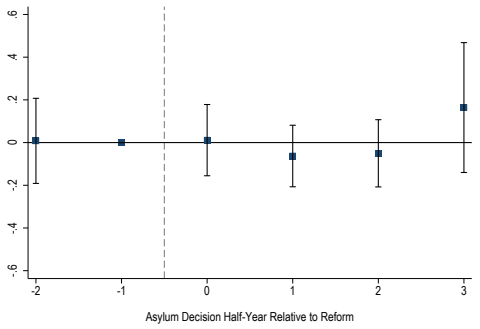
(c) Social Contact: Low-Unemployment



(d) Social Contact: High-Unemployment



(e) New Acquaintances: Low-Unemployment



(f) New Acquaintances: High-Unemployment

Figure 3.5.1. Event Study Diff-in-Diff Estimates

Notes: The figures display the event study Diff-in-Diff estimates for the three main integration outcomes separately for individuals in low and high-unemployment counties. All estimates are calculated relative to the effect in the half-year before the reform. Due to sample size restrictions, the half-year -2 also includes the observations from the one preceding quarter in my sample. The estimates are presented with 95 % confidence bands, which are calculated using 2,000 bootstrap iterations with clustering at the state level.

Taken together, I cannot detect any significant effects of the treatment pre-reform across all specifications. This finding is in line with the common trend assumption and provides me with additional confidence in the validity of the common trend assumption in this setting. Moreover, my results show that the treatment effects are most pronounced for refugees with an asylum decision between January 2016 and June 2017.

3.6 Conclusion

In this study, I show that the labor market conditions in the initial location of a refugee have a significant impact on their economic and social integration. Being allocated into a county with an above-median unemployment rate, decreases the probability that a refugee has found either full or part-time employment in 2017 by around 5 percentage points. Similarly, I find that being allocated into a county with an above-median unemployment rate reduces the likelihood of having found new German acquaintances by around 8 percentage points and reduces the frequency of social contact with Germans by 0.25 of a standard deviation. This result replicates findings of Aksoy, Poutvaara, and Schikora (2020) and is in line with many papers showing that local labor market conditions affect labor market outcomes of immigrants and refugees (Åslund and Rooth, 2007; Godøy, 2017; Azlor, Damm, and Schultz-Nielsen, 2020).

For policymakers the results imply that quasi-random spatial allocation mechanisms can induce large inequalities in integration outcomes between refugees. As a result, policymakers should either take local labor market conditions more into account when designing spatial allocation mechanisms or provide additional support measures for “unlucky” refugees that get allocated into bad local labor markets.

I extend the literature by also studying the effects of a policy aimed at increasing the stability of the initial distribution of refugees across regions, namely

local residence restrictions. I study the introduction of residence restrictions in Germany, which were argued to prevent spatial segregation of refugees and thereby improve the integration of refugees. Making use of their differential implementation across federal states, I find that being subject to stricter enforcement of local residence restrictions has heterogeneous effects on the economic and social integration depending on local labor market conditions in the initial residence. While I find evidence that refugees allocated to low-unemployment counties benefit from the stricter enforcement of local residence restrictions, refugees allocated to high-unemployment counties appear to be harmed by the stricter enforcement. Due to the small sample size, I cannot draw definite conclusions whether this result is driven by differences in compliance with the residence restrictions or differences in moving behavior. Taken together, my results, however, clearly highlight that local residence restrictions, as designed in Germany, are not necessarily beneficial for the economic and social integration of a refugee. Restricting the choice of residence increases existing inequalities among refugees introduced by the initial allocation as individuals allocated to favorable labor markets benefit while those allocated to unfavorable labor markets are further harmed.

References

Aksoy, Cevat Giray, Panu Poutvaara, and Felicitas Schikora. 2020. "First Time around: Local Conditions and Multi-Dimensional Integration of Refugees." *IZA Discussion Paper* 13914: [150–153, 163, 171, 174, 184]

Arendt, Jacob Nielsen. 2022. "Labor Market Effects of a Work-First Policy for Refugees." *Journal of Population Economics* 35: 169–96. [153]

Arendt, Jacob Nielsen, Iben Bolvig, Mette Foged, Linea Hasager, and Giovanni Peri. 2020. "Language Training and Refugees' Integration." *NBER Working Paper* 26834: [153]

Åslund, Olof, and Dan-Olof Rooth. 2007. "Do When and Where Matter? Initial Labour Market Conditions and Immigrant Earnings." *Economic Journal* 117 (518): 422–48. [153, 174, 184]

Azlor, Luz, Anna Piil Damm, and Marie Louise Schultz-Nielsen. 2020. "Local Labour Demand and Immigrant Employment." *Labour Economics* 63: 101808. [150, 153, 174, 176, 184]

Battisti, Michele, Yvonne Giesing, and Nadzeya Laurentsyeva. 2019. "Can Job Search Assistance Improve the Labour Market Integration of Refugees? Evidence from a Field Experiment." *Labour Economics* 61: 101745. [153]

Beaman, Lori A. 2012. "Social Networks and the Dynamics of Labour Market Outcomes: Evidence from Refugees Resettled in the U.S." *Review of Economic Studies* 79(1): 128–61. [150, 153]

Becker, Sascha O., and Andreas Ferrara. 2019. "Consequences of Forced Migration: A Survey of Recent Findings." *Labour Economics* 59: 1–16. [153]

Bratsberg, Bernt, Jeremy Ferwerda, Henning Finseraas, and Andreas Kotsadam. 2021. "How Settlement Locations and Local Networks Influence Immigrant Political Integration." *American Journal of Political Science* 65(3): 551–65. [153]

Brell, Courtney, Christian Dustmann, and Ian Preston. 2020. "The Labor Market Integration of Refugee Migrants in High-Income Countries." *Journal of Economic Perspectives* 34(1): 94–121. [153]

Brücker, Herbert. 2017. "IAB-BAMF-SOEP-Befragung von Geflüchteten 2016." *IAB Forschungsbericht* 13/2017: [163]

Bundesamt für Migration und Flüchtlinge. 2020. "Ablauf des deutschen Asylverfahrens." *Official Information Brochure*, [156]

Damm, Anna Piil. 2009. "Ethnic Enclaves and Immigrant Labor Market Outcomes: Quasi-Experimental Evidence." *Journal of Labor Economics* 27(2): 281–314. [150, 153]

Deutscher Bundestag. 2016. "Entwurf Eines Integrationsgesetzes Der Fraktionen Der CDU/CSU Und SPD." *Parliamentary Draft*, [151]

Edin, P.-A., P. Fredriksson, and O. Aslund. 2003. "Ethnic Enclaves and the Economic Success of Immigrants–Evidence from a Natural Experiment." *Quarterly Journal of Economics* 118(1): 329–57. [150, 153]

Fasani, Francesco, Tommaso Frattini, and Luigi Minale. 2021. "Lift the Ban? Initial Employment Restrictions and Refugee Labour Market Outcomes." *Journal of the European Economic Association* 19(5): 2803–54. [153]

Godøy, Anna. 2017. "Local Labor Markets and Earnings of Refugee Immigrants." *Empirical Economics* 52(1): 31–58. [150, 153, 174, 184]

Informationsverbund Asyl und Migration. 2020. "Das Asylverfahren in Deutschland." *Basisinformationen für die Beratungspraxis* Nr. 1: [155]

Kühne, Simon, Jannes Jacobsen, and Martin Kroh. 2019. "Sampling in Times of High Immigration: The Survey Process of the IAB-BAMF-SOEP Survey of Refugees." *Survey Methods: Insights from the Field (SMIF)*, [163]

Lehrian, Melina. 2018. "Zwei Jahre Wohnsitzregelung Nach § 12a AufenthG." *Asylmagazin* 12/2018: 416–23. [158]

Lehrian, Melina, and Johanna Mantel. 2016. "Neuerungen durch das Integrationsgesetz." *Asylmagazin* 9/2016: 290–94. [157, 158]

Marbach, Moritz, Jens Hainmueller, and Dominik Hangartner. 2018. "The Long-Term Impact of Employment Bans on the Economic Integration of Refugees." *Science Advances* 4(9): [153]

Martén, Linna, Jens Hainmueller, and Dominik Hangartner. 2019. "Ethnic Networks Can Foster the Economic Integration of Refugees." *Proceedings of the National Academy of Sciences* 116(33): 16280–85. [153]

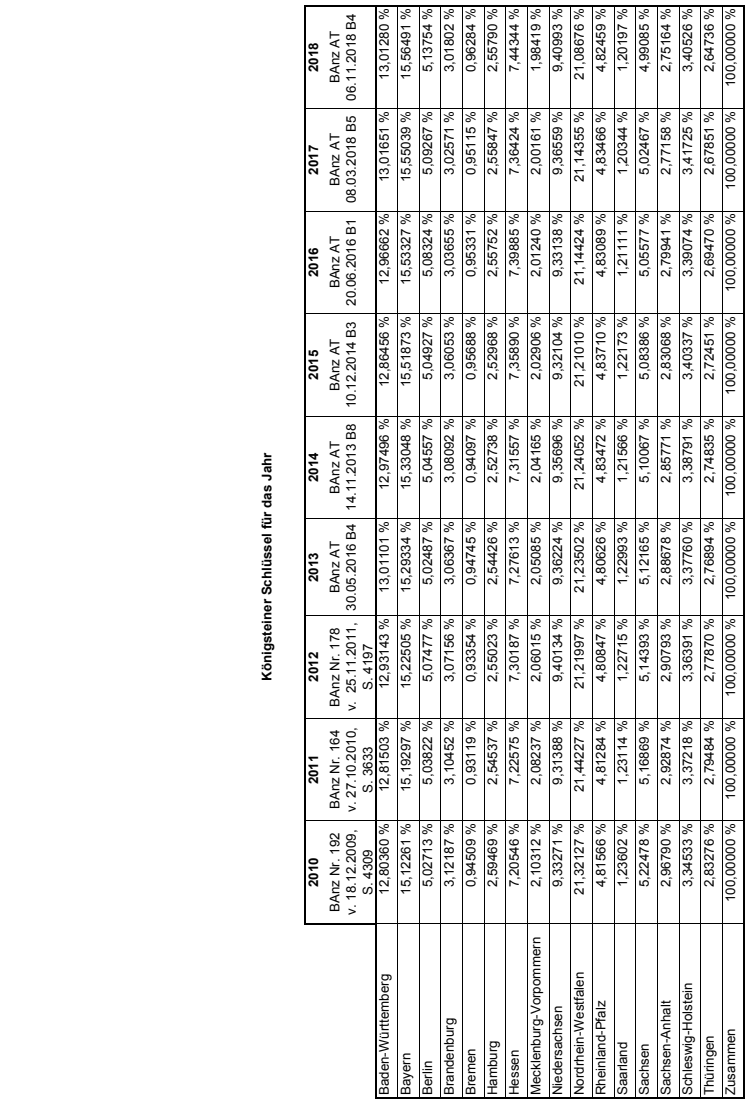
Schikora, Felicitas. 2019. "The Effect of Initial Placement Restrictions on Refugees' Language Acquisition in Germany." SOEPpapers on Multidisciplinary Panel Data Research 1035. Deutsches Institut für Wirtschaftsforschung (DIW). [152]

Schwandt, Hannes, and Till Von Wachter. 2019. "Unlucky Cohorts: Estimating the Long-Term Effects of Entering the Labor Market in a Recession in Large Cross-Sectional Data Sets." *Journal of Labor Economics* 37(1): S161–S198. [153]

Wachter, Till Von. 2020. "The Persistent Effects of Initial Labor Market Conditions for Young Adults and Their Sources." *Journal of Economic Perspectives* 34(4): 168–94. [153]

Appendix 3.A Institutional Setting

Figure 3.A.1. Federal Refugee Allocation Quotas



Königsteiner Schlüssel für das Jahr

	2010 BAnz Nr. 192 v. 18.12.2009, S. 4309	2011 BAnz Nr. 164 v. 27.10.2010, S. 3633	2012 BAnz Nr. 178 v. 25.11.2011, S. 4197	2013 BAnz AT 30.05.2016 B4	2014 BAnz AT 14.11.2013 B8	2015 BAnz AT 10.12.2014 B3	2016 BAnz AT 20.06.2016 B1	2017 BAnz AT 08.03.2018 B5	2018 BAnz AT 06.11.2018 B4
Baden-Württemberg	12,80360 %	12,81503 %	12,93143 %	13,01101 %	12,97496 %	12,86456 %	12,96662 %	13,01651 %	13,01280 %
Bayern	15,12261 %	15,19297 %	15,22505 %	15,29334 %	15,33048 %	15,51873 %	15,53327 %	15,55039 %	15,56491 %
Berlin	5,02713 %	5,03822 %	5,07477 %	5,02487 %	5,04557 %	5,04927 %	5,08324 %	5,09267 %	5,13754 %
Brandenburg	3,12187 %	3,10452 %	3,07156 %	3,06367 %	3,08092 %	3,06053 %	3,03655 %	3,02571 %	3,01802 %
Bremen	0,94509 %	0,93119 %	0,93354 %	0,94745 %	0,94087 %	0,95688 %	0,95331 %	0,95115 %	0,96284 %
Hamburg	2,59469 %	2,54537 %	2,55023 %	2,54426 %	2,52738 %	2,52968 %	2,55752 %	2,55847 %	2,55790 %
Hessen	7,20546 %	7,22575 %	7,30187 %	7,27613 %	7,31557 %	7,35890 %	7,39885 %	7,36424 %	7,44344 %
Mecklenburg-Vorpommern	2,10312 %	2,08237 %	2,06015 %	2,05085 %	2,04165 %	2,02906 %	2,01240 %	2,00161 %	1,98419 %
Niedersachsen	9,33271 %	9,31388 %	9,40134 %	9,36224 %	9,35696 %	9,32104 %	9,33138 %	9,36559 %	9,40993 %
Nordrhein-Westfalen	21,32127 %	21,44227 %	21,21997 %	21,23502 %	21,24052 %	21,21010 %	21,14424 %	21,14355 %	21,08676 %
Rheinland-Pfalz	4,81566 %	4,81284 %	4,80847 %	4,80626 %	4,83472 %	4,83710 %	4,83089 %	4,83466 %	4,82459 %
Saarland	1,23602 %	1,23114 %	1,22715 %	1,22993 %	1,21566 %	1,22173 %	1,21111 %	1,20344 %	1,20197 %
Sachsen	5,22478 %	5,16869 %	5,14393 %	5,12165 %	5,10067 %	5,08386 %	5,05577 %	5,02467 %	4,99085 %
Sachsen-Anhalt	2,96790 %	2,92874 %	2,90793 %	2,88678 %	2,85771 %	2,83068 %	2,79941 %	2,77158 %	2,75164 %
Schleswig-Holstein	3,34533 %	3,37218 %	3,36391 %	3,37760 %	3,38791 %	3,40337 %	3,39074 %	3,41725 %	3,40526 %
Thüringen	2,83276 %	2,79484 %	2,77870 %	2,76894 %	2,74835 %	2,72451 %	2,69470 %	2,67851 %	2,64736 %
Zusammen	100,00000 %	100,00000 %	100,00000 %	100,00000 %	100,00000 %	100,00000 %	100,00000 %	100,00000 %	100,00000 %

Notes: The table depicts the federal refugee allocation quotas in Germany for the years 2010 to 2018. The quotas are based on the "Königsteiner Schlüssel", which is published yearly by the Joint Science Conference ("Gemeinsame Wissenschaftskonferenz (GWK)").

Table 3.A.1. Local Refugee Allocation Quotas

No Quotas	Fixed Quotas	Fixed Quotas based on Population	Other Fixed Quotas
Berlin	Bavaria	Baden-Wuerttemberg	North-Rhine-Westfalia
Bremen	Saarland	Brandenburg	Saxony-Anhalt
Hamburg	Hesse	Mecklenburg Western Pomerania	
	Schleswig-Holstein	Lower Saxony	
	Thuringia	Saxony	
	Rhineland-Palatia		

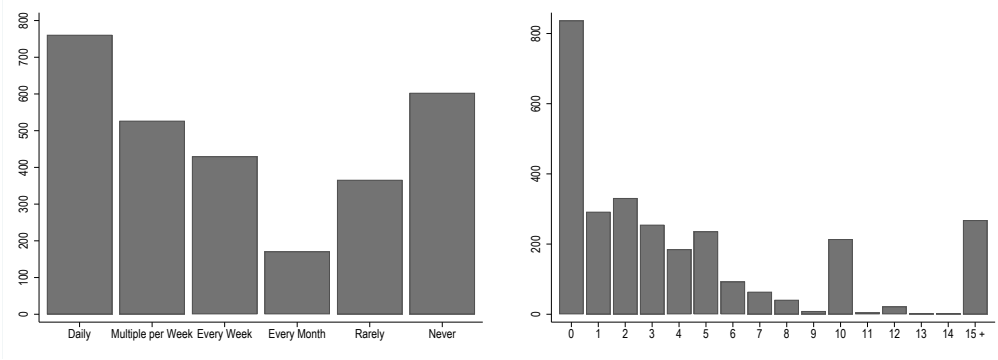
Notes: The table shows how the federal states design quotas to distribute refugees to different counties within the state. With the exception of the city-states Berlin, Bremen, and Hamburg, all federal states have implemented quotas to distribute refugees to different counties within the state. Most common are quotas according to the population size of the county (Baden-Wuerttemberg, Brandenburg, Mecklenburg Western Pomerania, Lower Saxony, and Saxony) and quotas that do not explicitly state which characteristics of the county were taken into account (Bavaria, Saarland, Hesse, Schleswig-Holstein, Rhineland-Palatia, and Thuringia). In two states, the quotas are determined using a combination of different factors. North-Rhine-Westfalia determines the quota to 80 % according to population size, to 10 % according to the area size, and to 10 % according to the level of unemployment and Saxony-Anhalt uses a not nearer specified combination of population size, unemployment, and vocational training places.

Table 3.A.2. Differences in Enforcement of Local Residence Restrictions

Early Enforcement	Late Enforcement	No Enforcement
Baden-Wuerttemberg	Hesse	Rhineland-Palatia
Bavaria	Saxony	Brandenburg
North-Rhine-Westfalia		Mecklenburg Western Pomerania
Saarland		Lower Saxony
		Thuringia
		Saxony-Anhalt
		Berlin, Bremen, Hamburg

Notes: The table shows the different enforcement of local residence restrictions across federal states. All information is taken from Lehrian (2018). The states of Bavaria, Baden-Wuerttemberg, Northrhine-Westphalia, and Saarland enforced local residence restrictions immediately following the integration law reform. The states Hesse (September 2017) and Saxony (April 2018) enforced local residence restrictions relatively late such that the refugees in my sample were not yet affected. The remaining states never enforced local residence restrictions.

Appendix 3.B Additional Summary Statistics



(a) Frequency of Social Contact with Germans **(b)** Number of New German Acquaintances

Figure 3.B.1. Distribution of Social Integration Measures

Notes: The figures display the empirical distribution of the two measures for social integration.

Appendix 3.C Additional Regression Results

Table 3.C.1. Full- or Part-time Employment in 2017 (Unemployment Rate)

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment Rate	−0.010***	−0.010***	−0.010***	−0.011***	−0.028**	−0.024**
	(0.003)	(0.003)	(0.003)	(0.003)	(0.011)	(0.011)
Post Reform × Enforced			0.042	0.036	0.261**	0.206**
			(0.038)	(0.036)	(0.110)	(0.106)
Post Reform × Enforced × Unemployment Rate					−0.034**	−0.026*
					(0.015)	(0.015)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.004	0.090	0.046	0.128	0.047	0.128

Notes: The table displays the regression results for the likelihood of being employed either full or part-time in 2017 using the unemployment rate instead of the high-unemployment dummy. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.2. Frequency of Social Contact with Germans in 2017 (Unemployment Rate)

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment Rate	−0.047***	−0.047***	−0.038***	−0.037***	−0.047*	−0.035
	(0.007)	(0.007)	(0.010)	(0.010)	(0.025)	(0.024)
Post Reform × Enforced			0.142	0.142*	0.348	0.212
			(0.087)	(0.085)	(0.245)	(0.243)
Post Reform × Enforced × Unemployment Rate					−0.035	−0.014
					(0.037)	(0.037)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.013	0.070	0.059	0.117	0.059	0.116

Notes: The table displays the regression results for the frequency of social interactions with Germans in 2017 using the unemployment rate instead of the high-unemployment dummy. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.3. New German Acquaintances since arrival (Unemployment Rate)

	(1)	(2)	(3)	(4)	(5)	(6)
Unemployment Rate	−0.019***	−0.018***	−0.024***	−0.023***	−0.024**	−0.023**
	(0.003)	(0.003)	(0.005)	(0.005)	(0.012)	(0.012)
Post Reform × Enforced			0.007	0.006	0.086	0.062
			(0.040)	(0.040)	(0.111)	(0.112)
Post Reform × Enforced × Unemployment Rate					−0.013	−0.009
					(0.017)	(0.017)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.01	0.036	0.035	0.055	0.034	0.054

Notes: The table displays the regression results for the likelihood of having found new German acquaintances since arrival using the unemployment rate instead of the high-unemployment dummy. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.4. Full- or Part-time Employment in 2017

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.053***	−0.054***	−0.035**	−0.037**	−0.132*	−0.115*
	(0.014)	(0.014)	(0.017)	(0.017)	(0.068)	(0.065)
Post Reform × Enforced			0.043	0.037		
			(0.040)	(0.038)		
Enforced × Low Unemployment					−0.238***	−0.200**
					(0.088)	(0.083)
Post Reform × Low Unemployment					−0.114	−0.095
					(0.071)	(0.067)
Post Reform × Enforced × Low Unemployment					0.198***	0.170**
					(0.073)	(0.069)
Post Reform × Enforced × High Unemployment					−0.092*	−0.083
					(0.056)	(0.052)
Female		−0.167***		−0.156***		−0.156***
		(0.012)		(0.012)		(0.012)
Age at Arrival		−0.001**		−0.002**		−0.001*
		(0.001)		(0.001)		(0.001)
Has Children		0.001		0.015		0.015
		(0.023)		(0.023)		(0.023)
Number of Children		−0.004		−0.006		−0.007
		(0.004)		(0.004)		(0.004)
Nationality: Syrian		−0.068***		−0.095***		−0.091***
		(0.020)		(0.020)		(0.020)
Nationality: Afghan		−0.076***		−0.047*		−0.046*
		(0.027)		(0.027)		(0.027)
Nationality: Iraqi		−0.101***		−0.093***		−0.092***
		(0.024)		(0.024)		(0.024)
Married		−0.053***		−0.070***		−0.069***
		(0.017)		(0.017)		(0.017)
Education: Secondary		0.057***		0.054***		0.051***
		(0.015)		(0.014)		(0.014)
Education: Tertiary		0.069***		0.055***		0.053***
		(0.019)		(0.019)		(0.019)
Constant	0.194***	0.363***	0.372***	0.568***	0.447***	0.625***
	(0.011)	(0.029)	(0.071)	(0.074)	(0.087)	(0.088)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.005	0.090	0.045	0.126	0.049	0.130

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the likelihood of being employed either full or part-time in 2017. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.5. Frequency of Social Contact with Germans in 2017

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.250***	−0.247***	−0.163***	−0.163***	−0.107	−0.067
	(0.037)	(0.036)	(0.046)	(0.044)	(0.128)	(0.121)
Post Reform × Enforced			0.148*	0.148*		
			(0.088)	(0.086)		
Enforced × Low Unemployment					0.069	0.165
					(0.184)	(0.179)
Post Reform × Low Unemployment					−0.057	−0.025
					(0.139)	(0.132)
Post Reform × Enforced × Low Unemployment					0.240*	0.196
					(0.143)	(0.137)
Post Reform × Enforced × High Unemployment					0.053	0.092
					(0.136)	(0.134)
Female		−0.285***		−0.271***		−0.273***
		(0.039)		(0.039)		(0.038)
Age at Arrival		−0.011***		−0.012***		−0.012***
		(0.002)		(0.002)		(0.002)
Has Children		−0.044		−0.002		−0.002
		(0.064)		(0.063)		(0.062)
Number of Children		0.020		0.015		0.015
		(0.015)		(0.014)		(0.014)
Nationality: Syrian		−0.012		−0.085		−0.076
		(0.051)		(0.052)		(0.052)
Nationality: Afghan		0.086		0.124*		−0.133*
		(0.072)		(0.072)		(0.072)
Nationality: Iraqi		−0.123**		−0.151**		−0.138**
		(0.067)		(0.067)		(0.067)
Married		−0.049		−0.092*		−0.095**
		(0.048)		(0.048)		(0.048)
Education: Secondary		0.291***		0.284***		0.284***
		(0.043)		(0.042)		(0.042)
Education: Tertiary		0.338***		0.322***		0.322***
		(0.052)		(0.0512)		(0.052)
Constant	0.187***	0.478***	0.467***	0.846***	0.426**	0.770***
	(0.027)	(0.084)	(0.153)	(0.172)	(0.176)	(0.186)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.015	0.072	0.058	0.116	0.059	0.118

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the frequency of social interactions with Germans in 2017. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.6. New German Acquaintances since arrival

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.080***	−0.077***	−0.078***	−0.077***	−0.049	−0.047
	(0.017)	(0.017)	(0.021)	(0.021)	(0.059)	(0.059)
Post Reform × Enforced			0.011	0.010		
			(0.039)	(0.039)		
Enforced × Low Unemployment					0.040	0.053
					(0.078)	(0.079)
Post Reform × Low Unemployment					−0.011	−0.014
					(0.064)	(0.063)
Post Reform × Enforced × Low Unemployment					0.035	0.032
					(0.064)	(0.063)
Post Reform × Enforced × High Unemployment					−0.018	−0.012
					(0.057)	(0.058)
Female		−0.021		−0.022		−0.023
		(0.018)		(0.018)		(0.018)
Age at Arrival		−0.003**		−0.002**		−0.002**
		(0.001)		(0.001)		(0.001)
Has Children		−0.019		−0.016		−0.016
		(0.030)		(0.031)		(0.031)
Number of Children		0.002		0.001		0.001
		(0.007)		(0.007)		(0.007)
Nationality: Syrian		0.054**		0.034		0.037
		(0.024)		(0.025)		(0.025)
Nationality: Afghan		0.045		0.047		0.050
		(0.034)		(0.034)		(0.035)
Nationality: Iraqi		0.012		0.001		0.006
		(0.031)		(0.031)		(0.031)
Married		0.021		0.014		0.013
		(0.024)		(0.024)		(0.024)
Education: Secondary		0.115***		0.114***		0.114***
		(0.020)		(0.020)		(0.020)
Education: Tertiary		0.157***		0.155***		0.154***
		(0.024)		(0.024)		(0.024)
Year of Interview	−0.103***	−0.097***	−0.065***	−0.068***	−0.063***	−0.067***
	(0.017)	(0.017)	(0.019)	(0.020)	(0.019)	(0.020)
Constant	0.807***	0.769***	0.906***	0.875***	0.884***	0.851***
	(0.014)	(0.039)	(0.060)	(0.075)	(0.073)	(0.087)
Controls	No	Yes	No	Yes	No	Yes
N	2,858	2,858	2,858	2,858	2,858	2,858
Adj. R^2	0.021	0.045	0.033	0.055	0.034	0.055

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the likelihood of having found new German acquaintances since arrival. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.7. Full- or Part-time Employment (Pooled)

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.052***	−0.054***	−0.033***	−0.036***	−0.068**	−0.052*
	(0.009)	(0.008)	(0.011)	(0.010)	(0.033)	(0.030)
Post Reform × Enforced			0.033	0.022		
			(0.020)	(0.019)		
Enforced × Low Unemployment					−0.095**	−0.059
					(0.043)	(0.039)
Post Reform × Low Unemployment					−0.043	−0.029
					(0.035)	(0.032)
Post Reform × Enforced × Low Unemployment					0.097***	0.070***
					(0.036)	(0.033)
Post Reform × Enforced × High Unemployment					−0.028	−0.029
					(0.029)	(0.027)
Year 2017	0.066***	0.075***	0.103***	0.106***	0.104***	0.107***
	(0.010)	(0.009)	(0.010)	(0.010)	(0.010)	(0.010)
Year 2018	0.193***	0.201***	0.226***	0.229***	0.227***	0.230***
	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)
Year 2019	0.272***	0.285***	0.301***	0.311***	0.302***	0.312***
	(0.013)	(0.012)	(0.013)	(0.012)	(0.013)	(0.012)
Female		−0.219***		−0.211***		−0.211***
		(0.008)		(0.008)		(0.008)
Age at Arrival		−0.003***		−0.003***		−0.003***
		(0.000)		(0.000)		(0.000)
Has Children		−0.008		0.001		0.001
		(0.014)		(0.014)		(0.014)
Number of Children		−0.003		−0.004		−0.004
		(0.003)		(0.003)		(0.003)
Nationality: Syrian		−0.065***		−0.087***		−0.084***
		(0.012)		(0.012)		(0.012)
Nationality: Afghan		−0.093***		−0.069***		−0.068***
		(0.016)		(0.017)		(0.017)
Nationality: Iraqi		−0.098***		−0.092***		−0.090***
		(0.015)		(0.015)		(0.016)
Married		−0.047***		−0.056***		−0.056***
		(0.010)		(0.010)		(0.010)
Education: Secondary		0.071***		0.070***		0.069***
		(0.009)		(0.009)		(0.009)
Education: Tertiary		0.093***		0.084***		0.084***
		(0.012)		(0.012)		(0.012)
Constant	0.127***	0.340***	0.177***	0.395***	0.205***	0.407***
	(0.008)	(0.019)	(0.033)	(0.036)	(0.041)	(0.042)
Controls	No	Yes	No	Yes	No	Yes
N	8,817	8,817	8,817	8,817	8,817	8,817
Adj. R^2	0.403	0.379	0.398	0.375	0.397	0.375

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the likelihood of being employed either full or part-time using information from all survey years pooled. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.8. Frequency of Social Contact with Germans in 2017 (Pooled)

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.174***	−0.176***	−0.127***	−0.125**	−0.015	0.019
	(0.021)	(0.020)	(0.026)	(0.025)	(0.067)	(0.066)
Post Reform × Enforced			0.066	0.051		
			(0.047)	(0.046)		
Enforced × Low Unemployment					0.186**	0.248***
					(0.092)	(0.090)
Post Reform × Low Unemployment					0.034	0.060
					(0.075)	(0.074)
Post Reform × Enforced × Low Unemployment					0.045	0.001
					(0.077)	(0.075)
Post Reform × Enforced × High Unemployment					0.063	0.070
					(0.068)	(0.068)
Year 2017	−0.017	−0.009	0.020	0.018	0.020	0.018
	(0.029)	(0.028)	(0.029)	(0.028)	(0.029)	(0.028)
Year 2018	−0.037	−0.030	−0.007	−0.007	−0.006	−0.007
	(0.030)	(0.030)	(0.031)	(0.030)	(0.031)	(0.030)
Year 2019	0.024	0.044	0.048	0.064	0.048	0.063
	(0.032)	(0.032)	(0.033)	(0.032)	(0.033)	(0.032)
Female		−0.279***		−0.276***		−0.279***
		(0.022)		(0.022)		(0.022)
Age at Arrival		−0.011***		−0.011***		−0.011***
		(0.001)		(0.001)		(0.001)
Has Children		−0.063*		−0.050		−0.050
		(0.036)		(0.036)		(0.036)
Number of Children		0.019**		0.014*		0.014*
		(0.008)		(0.008)		(0.008)
Nationality: Syrian		−0.069**		−0.107***		−0.102***
		(0.028)		(0.029)		(0.029)
Nationality: Afghan		−0.005		0.017		0.020
		(0.043)		(0.044)		(0.044)
Nationality: Iraqi		−0.091**		−0.133***		−0.127***
		(0.039)		(0.039)		(0.039)
Married		−0.008		−0.017		−0.020
		(0.027)		(0.027)		(0.027)
Education: Secondary		0.280***		0.280***		0.281***
		(0.024)		(0.024)		(0.024)
Education: Tertiary		0.333***		0.325***		0.325***
		(0.030)		(0.030)		(0.030)
Constant	0.151***	0.458***	0.260***	0.579***	0.174**	0.467***
	(0.024)	(0.050)	(0.078)	(0.090)	(0.091)	(0.100)
Controls	No	Yes	No	Yes	No	Yes
N	8,817	8,817	8,817	8,817	8,817	8,817
Adj. R^2	0.008	0.061	0.022	0.077	0.027	0.078

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the frequency of social interactions with Germans in 2017 using information from all survey years pooled. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.

Table 3.C.9. New German Acquaintances since Arrival (Pooled)

	(1)	(2)	(3)	(4)	(5)	(6)
High Unemployment	−0.071***	−0.069***	−0.064***	−0.061***	−0.057	−0.050
	(0.015)	(0.015)	(0.018)	(0.018)	(0.049)	(0.049)
Post Reform × Enforced			−0.020	−0.029		
			(0.033)	(0.033)		
Enforced × Low Unemployment					−0.048	−0.033
					(0.065)	(0.065)
Post Reform × Low Unemployment					−0.021	−0.020
					(0.054)	(0.054)
Post Reform × Enforced × Low Unemployment					0.037	0.024
					(0.054)	(0.054)
Post Reform × Enforced × High Unemployment					−0.093*	−0.094*
					(0.048)	(0.048)
Year 2017	−0.102***	−0.096***	−0.071***	−0.076***	−0.068***	−0.073***
	(0.016)	(0.016)	(0.018)	(0.018)	(0.018)	(0.018)
Year 2018	−0.021	−0.009	−0.005	0.002	0.000	0.007
	(0.039)	(0.039)	(0.039)	(0.040)	(0.039)	(0.040)
Female		−0.032**		−0.034**		−0.034**
		(0.016)		(0.016)		(0.016)
Age at Arrival		−0.002**		−0.002**		−0.002**
		(0.001)		(0.001)		(0.001)
Has Children		−0.020		−0.019		−0.019
		(0.027)		(0.027)		(0.027)
Number of Children		0.003		0.001		0.001
		(0.006)		(0.006)		(0.006)
Nationality: Syrian		0.030		0.015		0.019
		(0.020)		(0.021)		(0.021)
Nationality: Afghan		0.008		0.010		0.013
		(0.030)		(0.031)		(0.031)
Nationality: Iraqi		−0.002		−0.017		−0.013
		(0.027)		(0.028)		(0.028)
Married		0.034		0.030		0.029
		(0.021)		(0.021)		(0.021)
Education: Secondary		0.113***		0.111***		0.110***
		(0.018)		(0.018)		(0.018)
Education: Tertiary		0.157***		0.155***		0.154***
		(0.021)		(0.022)		(0.022)
Constant	0.801***	0.776***	0.854***	0.833***	0.852***	0.825***
	(0.012)	(0.034)	(0.050)	(0.060)	(0.061)	(0.070)
Controls	No	Yes	No	Yes	No	Yes
N	3,564	3,564	3,564	3,564	3,564	3,564
Adj. R^2	0.019	0.042	0.029	0.052	0.030	0.053

Notes: The table displays full regression results (except for asylum decision time and state dummies) for the number of new German acquaintances since arrival in 2017 using information from all survey years pooled. There are less observations than for the other variables as the question was only asked once in the first interview. Columns (1) and (2) show the regression results from equation 3.4.1, columns (3) and (4) the regression results from equation 3.4.2 and columns (5) and (6) the regression results from equation 3.4.2 interacted with the unemployment rate. Controls include age, gender, nationality, education, marital status and number of children of the refugee. Standard errors are clustered on the state level and bootstrapped with 2,000 replications. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$.