Estimating Labour Supply Responses: The Abolition of the German Solidarity Tax

Master Thesis [Policy Economics]

Name student:	Gregor Kloepfer
Student ID number:	564031
Supervisor:	Mareen Bastiaans
Second Assessor:	Aart Gerritsen

ERASMUS UNIVERSITY ROTTERDAM
Erasmus School of Economics

Date: August 20, 2024

Abstract

I estimate current labour supply responsiveness to income taxation in Germany and apply a Difference-in-Differences strategy, allowing for continuous treatment intensity. In this, I draw on variation caused by the abolition of the solidarity surcharge at the beginning of 2021, which changed marginal tax rates for a majority of earners. The analysis utilizes data from the Socio-Economic Panel, revealing that tax rate reductions appear to decrease annual labour hours, with a stronger effect observed among unmarried individuals. Married women show a greater responsiveness to these tax rate changes compared to married men, whereas the number of children plays no significant role in the effect. Further considering age as a driver of heterogeneous results, I found that older individuals are more responsive to tax rate changes. Overall, these findings suggest that lowering tax rates might not be the right tool for the German government to incentivise higher working hours.

Keywords: Income taxation, labour supply, Difference-in-Differences, Germany, public economics

Contents

1 Introduction

A recent report by the OECD (2023) shows the Europe-wide decrease in average hours worked per working citizen. From the perspective of many countries, it is important to prevent this: The concern of decreasing average working hours appears to force governments to take action, most often seen in tax cuts. Several European governments, such as the UK, Austria, Sweden and Lithuania, recently implemented tax reductions or announced plans to reduce income taxes (Asen, 2021; Enachen, 2024; Mengden, 2023; Reuters, 2024). This trend underlines the importance of understanding the responsiveness of individuals to changes in the tax schedule. In a 2024 article by the Financial Times (Arnold et al., 2024), similar plans from the German government were unveiled. With 1,341 annual labour hours per worker, Germany has the lowest average working hours among developed nations, showing a reduction of roughly 30 % over the last five decades (OECD, 2024). This is however, in part, driven by increased female labour force participation on a part-time basis. Facing the underlying structural problem, the German government recently proposed plans to introduce tax cuts for working overtime. Although, this plan is still up for debate, it raises questions with regards to effectiveness (and more broadly efficiency) of such measures.

Furthermore, Germany's demographic landscape is changing rapidly, as the Baby Boomer generation is starting to retire whilst the next generations are substantially smaller. Such drastic changes of the work force are expected to put a strain on tax revenues (Calahorrano et al., 2019). Germany faces economic pressure on several fronts: On the one side German workers put in fewer and fewer hours. On the other, the outlined change in demography is starting to catch up with government financing and will occupy future administrations. For well-suited governing and adaptation to the need of reforms it is crucial to accurately estimate the labour market effects of changes in the tax system. If people are less responsive to tax cuts than anticipated, such measure might pose a costly gamble for governmental budgeting. In my thesis I aim to contribute to this discussion, by analysing the responsiveness of German workers to labour income taxation.

To answer my research question, I draw on variation introduced by the abolition of the solidarity surcharge ("Solidaritätszuschlag" in German, hereafter "SOLI"), leading to decreased marginal tax rates for a majority of earners in Germany starting in 2021. In a short report, Bonin et al. (2019) estimate positive labour market responses to this reform, amounting to 70,000 full-time equivalents on the national level. To grasp the complex interplay of consequences of the abolition of the SOLI for most taxpayers, it is however crucial to understand microeconomic effects of the policy. This variation caused by the abolition of the

SOLI poses the opportunity to employ a Difference-in-Differences design, comparing affected to not-affected individuals. Previous estimates of German elasticities of labour supply (or closely related matters) have been hard to obtain for the recent period, due to little variation in the German tax schedule. Each year Germany adapts the income tax system to account for inflation and other factors. As those changes are marginal, they do not provide enough variation for an examination of labour market responses. Thus, there are no current estimates of German labour supply elasticity at the micro level. I strive to contribute to literature by extending estimates to the latest available data from 2021 and provide a baseline for comparison to other countries as well as Germany in preceding decades. Previous comparative research in the field has found similar elasticities of labor supply of men across countries, whereas women seem to show more variation (Evers et al., 2008). Hence, the estimation of differences by gender underlines the relevance of analysing the German context.

The OECD (2017) found that women are still 10 percentage points less likely to engage in the German labour market compared to men. Even though female labour force participation in Germany is above the OECD average, the gap in Germany is persistent. Furthermore, if women engage in the labour market, it is often via part-time labour, in contrast to men. This fact highlights the importance of fostering equal opportunities and removing barriers to entry and full-time employment for women. Such household-level decisions are likely to be driven partially via economic reasoning and partially via cultural norms, as a previous study demonstrated on the example of Sweden (Ichino et al., 2022). The social relevance of my paper thus shows two main aspects: understanding the intensive margin effects of changes in the German tax schedule and analyzing to what extent women are differentially affected compared to men. In the context of this analysis I go beyond literature in the field examining both men and women within the same natural experiment, and introduce a heterogeneity analysis on the role of the number of children; a potentially restraining factor negatively affecting women's labour supply. I further add to literature by analysing the whole working population and subsequently presenting results allowing for heterogeneous effects by age.

In this thesis, I found that individuals responded to a 1 percentage point decrease in tax rates with decreasing annual labour by 7.32 to 20.04 hours[1] (with varying degrees of statistical significance). Splitting the sample shows that this effect largely depends on unmarried individuals. Models excluding outliers showed that among unmarried individuals a 1 percentage point reduction in tax rates leads to a reduction of 44.82 to 85.50 annual labour hours. However, my Difference-in-Differences approach falls short of the parallel trends assumption, thus results cannot be interpreted causally. Examining heterogeneity by gender, a

[1] A minority of models considering a longer time-span showed an opposite effect sign.

larger responsiveness of women compared to men appears to be driven by differences between married individuals. Notably, the number of children plays little role in the effect, with most models showing statistically insignificant results. Lastly, I found that older individuals are more responsive to tax rate changes, underlining that differential age effects are an important aspect to consider in policy making.

The rest of this thesis is structured as follows: In Section 2 I give an overview on the current state of literature in the field, followed by an explanation of the institutional context in Section 3. After outlining the main data in Section 4, I explain the empirical design and the identifying assumption in Section 5. In Section 6 I present the main findings, and examine their robustness. To narrow down policy implications, Section 7 dives into the heterogeneity of the effect. Lastly, Section 8 outlines limitations of the thesis and concludes the work.

2 Literature Review

Economic theory lays out the contradictory roles of income and substitution effects arising as a result of income taxation and shows how taxation can lead to negative labour supply responses. In general, I focus on labour supply elasticities in this section on literature, as they are considered a measure of the responsiveness to income taxation, which in turn is considered in my thesis. Although this elasticity and the responsiveness of labour supply to changes in the tax rate examined in this thesis are not the same concepts, they can, to some extent, be used interchangeably[2].

Meghir and Phillips (2010) provide a comprehensive guide on the theory of taxation and individual utility maximization in the presence of income taxes. This book serves as a basis for understanding the distortive labour supply effects, income taxation can cause. Facing higher (lower) tax rates workers might substitute leisure for work (work for leisure), resulting in lower (higher) labour supply. This is referred to as the substitution effect and captured by the so called compensated or Hicksian elasticity. Conversely, the income effect states that given higher (lower) tax rates, the worker has to work more (less) to sustain income or more general a given standard of living. These effects work in opposing directions. From theory, it is a priori unclear which effect dominates. The overall responsiveness depends on uncom-

[2]Elasticities generally measure the percentage change in labour supply given a percentage change in the after-tax wage rate or tax rate. As I do not measure percentage changes in labour supply but hours instead in this thesis, I note the distinction between responsiveness and elasticities.

pensated elasticity of labour supply (also referred to as the Marshallian elasticity), which comprises both the income and the substitution effect. In a comprehensive literature review, Keane (2011) elaborates on different theoretical models used and surveys the literature on empirical findings across publications, separating by gender. Based on 22 studies, he found an average value of the Hicksian elasticity of 0.31 among men, which is small but economically impactful. For women, long run labour supply elasticities appear substantially larger, averaging 3.6 across five studies. The review aims at identifying key challenges in econometrically estimating labour supply functions and provides journey through different attempts of causally identifying labour supply elasticities. Key challenges identified by Keane (2011) are endogeneity concerns and measurement error in labour income.

A sizeable body of studies combine theoretical models with empirical estimations. This paragraph makes no attempt for completeness, but serves to provide a mere overview. Bingley and Lanot (2002), for instance, draw on a theoretical model to deduce an econometric evaluation based on Denmark. Their analysis is based on a theoretical life-cycle model of individuals. They found small labour supply elasticities, but detected shifts in the incidence (burden) of income taxation away from the individual, towards the employer. Cutanda and Sanchis-Llopis (2023) examined this question in the context of tax reforms in Spain, likewise building onto a theoretical model of life-cycle utility. They detected a small impact of tax reforms on hours worked, indicating a low labour-supply elasticity. However, they recognized sizeable heterogeneity, across dimensions, such as gender, age, and type of contract. Based on a theoretical model allowing for a progressive income tax system, Blomquist (1983) find a comparably small Marshallian elasticity of 0.08 on a sample of married working-aged men in Sweden.

On the purely empirical side, a large pool of studies outlines estimates for various countries. Again, this section sets no claims for completeness of these studies, but shall provide an understanding of the current state of literature in the field. Similar to Keane (2011), Evers et al. (2008) perform a qualitative study, reviewing literature and compiling quantitative estimates across countries. Generally, they found that the uncompensated (Marshallian) elasticities do not vary substantially across countries for men, whereas women show variation in international comparison. Next to the aforementioned paper by Cutanda and Sanchis-Llopis (2023), previous studies have, in general, found mostly small effects. Similarly, Müllbacher and Nagl (2017) found small estimates for the wage elasticity amongst men, with slightly larger values for women. Their analysis was based on a discrete choice model evaluated at hand of an Austrian tax reform. In a correlational study Tondani (2006) found that women's labour supply elasticity is slightly larger than men's in Italy. Ashenfelter et al. (2010), in a New York City-based study, leverage exogenous fare changes among taxi drivers, who are particularly flexible to choose their own labour supply (in terms of hours). Notably, they conclude that in their setting income effects appear to dominate substitution effects, with an elasticity of -0.2. Due to the demographic characteristics of taxi drivers, this study falls short of an analysis by gender and is unlikely to be representative of the general public.

In a more holistic study, Attanasio et al. (2015) aimed at disentangling extensive and intensive margin effects in the US and found substantial effect influences of socio-economic characteristics: On the individual level age, wealth, working hours and wage rates play an important role, next to unobserved preferences with regards to leisure. Most importantly, elasticities are larger for individuals with low working hours (part-time workers) and young women. In particular, they overall found a compensated (Hicksian) elasticity of 0.54 and an uncompensated (Marshallian) elasticity of 0.18. In a next step, the authors aggregate behavior and find that labour supply elasticities are cyclical, however depend on the demographic structure, as suggested by their heterogeneity analysis.

Gruber and Saez (2002) rework previous methodological issues in a seminal paper and provide a fundament for future empirical work in the field, on which I draw in my thesis. In a first step, they redefine the individual's marginal tax rate (hereafter MTR) as a function of the tax change preceding year's income. In that they apply the tax schedule of year t to the income of year $t-1$. This ensures to isolate tax rate changes in the identification strategy and rules out endogenous labour supply responses. Secondly, Gruber and Saez (2002) rely on multiple tax schedule changes over the 80s within the US. They draw on Goolsbee (1998) in the argument, that several reforms spread out over years, effectively ease up a causal empirical strategy: Controlling for lagged income in such a setting provides a strong identification, as this does not run into over-controlling, effectively taking all the variation from the independent variable of interest, whilst controlling indirectly for unobservables. They further state that without several tax changes, it is hardly possible to isolate income effects under controls for the base year, an important drawback for the setting discussed in this thesis. Subsequently, they employ an IV strategy (2SLS) using the synthetic MTR and after tax income share, restated in the same manner as the MTR, as instruments. Gruber and Saez (2002) find, that the overall elasticity (weighted by income) of taxable income is 0.4, with significant heterogeneity by income. For the lower-income group (which concerns a large fraction of my main sample as outlined in section 4.3) this elasticity is between 0.1 and 0.2. In line with this evidence, I formulate my first, and central, hypothesis:

$\mathbf{H_1}$**:** A reduction in marginal tax rates has a positive effect on labor supply (and vice versa).

So far, the studies examined mostly concern average treatment effects. Such a broad perspective can potentially undermine important heterogeneous effects. In this thesis I aim at disaggregating average effects and consider heterogeneity across several dimensions. In that, my study touches upon models of household production functions. The decision to work within couples is taken at the household level (Pollak & Wachter, 1975). Gary Becker laid the fundament towards such thinking: In seminal work he novelly contributed towards understanding households as one unit, coordinating decision making in the so-called household production function (Heckman, 2015; Pollak, 2003). Ichino et al. (2022) consider such a household model by leveraging variation in the Swedish tax system and estimating effects on childcare. Following Gruber and Saez (2002) in reformulating tax rate changes, they find that couples originating from more conservative countries are less prone to shift childcare responsibilities to fathers when the mother's tax rate is reduced. In her influential papers, Eissa (1995) and Eissa (1996) leverages variation in tax rates caused by two US tax reforms, from 1981 and 1986. Methodologically very similar to my study, she employs a Difference-in-Differences design, comparing different income percentiles that were differentially affected by the two reforms. However, both groups in her setting experienced a treatment, hence results are based on relative changes. Further, in contrast to the context examined in this thesis, the natural experiments did not allow for the implementation of an identification strategy which allows for continuous variation in treatment intensity. Her studies find an implied elasticity of annual labour supply ranging between 1.25 and 1.60 for women. In line with previous findings, I formulate my second hypothesis:

$\mathbf{H_2}$**:** A reduction in marginal tax rates has stronger effects on women, compared to men.

My analysis is based on the individual level and only aims at considering household-level trade-offs in the heterogeneity analysis. With my study, I thus contribute to the pool of individual level analyses, providing a baseline for taxation policies with recent estimates for Germany. Previous literature has highlighted the importance of considering heterogeneity factors in the field. As one of the key drivers of differences in the individual responsiveness to income taxation or wage rate changes, gender stands out. This thesis aims at deepening the understanding of this factor, going beyond the current state of literature, by examining to what extent children mediate this effect. The number of children is to be considered a constraint in the household production function: Children can pose additional stress on the division of time. It is likely that individuals facing such constraints are less responsive to changes in the tax schedule. With this background, I state my final hypothesis:

H_3: A reduction in marginal tax rates has smaller effect on individuals with children.

3 Institutional Setting

In my study I focus on the German tax system. German labour income is taxed individually, with the option of joint taxation for married couples (and so-called registered partners[3]) (German Federal Ministry of Finance, 2022). Thus, married individuals are free to opt into joint taxation or stay in a system of individual taxation. The income tax is progressive, with marginal tax rates increasing in labour income. Tax rates are to a small extent adapted on an annual basis by the lawmakers. Such changes take place annually and due to their negligible size, I will treat them as part of the main reform examined in this analysis. In my work I will focus on the SOLI, which was abolished for most individuals from January 1st 2021. I will begin by explaining the regular income tax framework in Germany and subsequently lay out the details regarding the SOLI.

Generally § 32a EStG (Einkommensteuergesetz - income tax law) regulates the total amount of taxes to be paid. This law is summarized in Appendix Table A.1 for 2020 and 2021. It stipulates a formula from which the tax burden can be calculated. Single declarers are exempted from any income tax up to an annual income of 9.408 € in 2020 (9.744 € in 2021)[4]. After that the marginal tax rate (hereafter referred to as MTR) starts increasing from 14 % upwards to a maximum of 45 %.

The German tax system furthermore divides individuals into six different tax classes. Class 1 is for single, widowed or divorced individuals. Single parents are taxed under class 2. Continuing, classes 3 to 5 are for married individuals: These classes characterize the system as one of joint taxation. Due to this nuance, it is common that married partners share the same MTR (and thus will have to be treated differently in this analysis.). Class 4 is intended for equal earners: Here, each partner gets treated separately and effectively follows the regular tax system. Class 3 however is specified for the high-income earner. If one partner occupies class 3, the other one has to uptake class 5 (for the lower-income earner) by construction. Class 4, due to the progressive nature of the tax system, is never of monetary advantage[5], but might be preferred by couples separating finances. Lastly, class 6 is for secondary income.

[3] Registered Partnership was the predecessor of the "Marriage for All" act, enabling homosexual couples to obtain the same tax benefits as married couples.

[4] For married partners this amount is doubled.

[5] Minor advantages of class 4 arise in the reduction of additional tax payments ("Steuernachzahlungen"). However, these advantages are small enough, to be disregarded for the means of this thesis.

Besides the feature of joint taxation for married partners, the tax class does not influence MTRs but only effective MTRs taking transfers and deductibles into account.

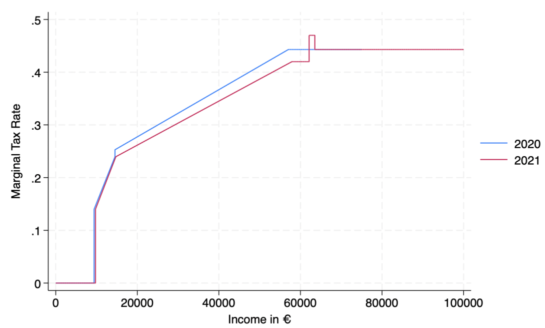
The SOLI, whose abolition forms the main source of variation studied in this thesis, was introduced in 1995 as a permanent measure, replacing an initial measure from 1991/92. The intention of this tax was to guarantee funding for the German Unity, following the reunification of West and East Germany in 1990 (German Federal Ministry of Finance, 2022). Until its phase-out began in 2021, the SOLI generated approximately €20 billion annually for the German government based on income taxes, representing a significant portion of federal tax revenue. This tax was initially introduced to boost economic development for the East after the reunification. In this setting, I analyse to what extent the abolition affected individual's labour supply, wherein the vast majority of tax payers was affected.

The SOLI is considered a special tax, levied on the total amount of income tax to be paid. In particular, it is a surcharge of 5.5 % on the income tax. To better understand the system, I provide an example based on the 2020 tax schedule: An individual that has annual labour earnings of 20,000.00 € has to pay 2,346.00 € in income taxes (this can be calculated with the formula from Panel (A) of Appendix Table A.1). Then, the SOLI is leveraged as a surchage of 5.5 % of the income tax to be paid, amounting to 129.03 € in this case. The total amount of taxes to be paid is then 2,475.03 €. The SOLI thus directly relates to the income tax system stipulated in § 32a EstG and poses an additional increase in the marginal tax rate for every individual. Note, the SOLI is stated in a single standard ("Einzelnorm"), the solidarity surcharge law[6] and is thus not directly stipulated in the main income tax law in § 32a EStG. Starting 2021 the SOLI was abolished for small and middle income earners. For earners between 62,128 € and 63,525 € in annual labour income the SOLI is phased in: 11.9 % of the income tax have to be paid in SOLI taxes, increasing MTRs for this small group. This leads up to high-income earners who, as in previous years, pay the full 5.5 % SOLI. This group is similarly small in size. Thus, the so-called phase-in range experiences an MTR increase. The resulting MTRs for 2020 and 2021 including the SOLI are displayed in Panel (A) of Figure 3.1.

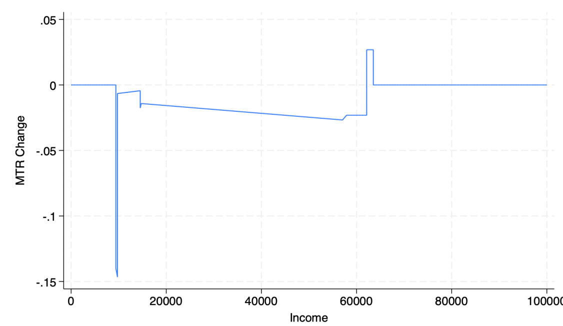
In designing an empirical strategy for the research question at hand, it appears crucial to understand the variation that arises in the MTR scheme between 2020 and 2021. Panel (B) of Figure 3.1 displays the variation in MTRs between 2020 and 2021. Firstly, individuals with income below 9,409 € and above 63,526 € are unaffected and show no change in MTRs. A small range of individuals between 9,409 and 9,744 € experiences a decrease in MTRs of

[6] In German, this is referred to as the Solidaritätszuschlaggesetz (SolZG). The applicability of the surcharge is specified under § 3 SolZG.

14.00 to 14.65 percentage points. Over the middle of the income distribution (9,745 to 62,127 €) MTRs are decreased by 0.44 to 2.31 percentage points. Thus the majority of the MTR reductions studied in this thesis are of small magnitude. Lastly, earners between 62,128 and 63,525 € experience a MTR increase (the previously outlined phase-in group) of 2.69 percentage points. They thus constitute the only group of individuals not benefiting from the reform.



Panel (A)



Panel (B)

Figure 3.1: Marginal Tax Rates including SOLI by income for 2020 and 2021 and the imputed change up to an income 100,000 €

Note: Panel (A) shows the MTRs over the considered income range for both 2020 and 2021. Panel (B) shows a line graph of the imputed change in MTRs at constant income across 2020 and 2021.

4 Data

Through the German Institute for Economic Research e.V. (Deutsches Institut für Wirtschaftsforschung e.V.), I have obtained longitudinal survey data on German households. The Socio-Economic Panel (SOEP) is generally highly regarded (Goebel et al., 2019). The survey started in 1984 and contains roughly 15,000 households amounting to over 30,000 individuals yearly up to 2021. It contains detailed information, amongst others, on employment and earnings as well as individual and household characteristics. The panel is based on multi-stage random sampling accounting for regional clustering. It is well-suited for the means of this thesis.

4.1 Dependent Variable

The central outcome variable of this thesis is labour supply, measured by "Annual Work Hours of Individual". These hours are computed by the data providers, using average values for hours worked per week, employment status and the number of weeks worked in the preceding year. Thus, work hours are a proxy for actual annual labour hours, potentially causing measurement error in my research. This problem could be addressed by using an average between two post-policy periods (2021 and 2022 in this case): However, due to the 2022 wave not having been published, such a robustness analysis is not feasible. The variable construction further entails that my analysis merely measures hours worked (in line with previous literature) and does not regard unobserved measures of labour performance, such as effort exerted[7]. An individual could work for more hours and still produce the same output, if effort decreases.

4.2 Independent Variables

4.2.1 Marginal Tax Rates

The main independent variable in this thesis is the MTR of an individual. The MTR in turn, is a direct function of income. As outlined in section 3, I differentiated the stipulated tax formulas of the German government to obtain MTRs, depending on income. Subsequently, I applied the differentiated tax schedule to individuals within the SOEP panel and obtained individual level MTRs of labour income. The determining variable in that is "Labor Earnings of Individual" as of the SOEP codebook. MTRs are computed based on §32a EStG and then adapted to account for the main source of variation, the change in the applicability of the SOLI as in § 3 SolZG.

Looking at the variation that arises in the independent variable between 2020 and 2021, a few things stand out: Firstly, I have plotted the imputed change in MTRs across income in Panel (A) of Figure 3.1. Between incomes of 9,409 € and 9744 €, MTRs are reduced by 14.00 up to 14.65 percentage points[8]. Further, it stands out, that a small range of earners is affected by an MTR increase in the range of 62,128 € to 63,525 € (with corresponding MTR changes of 2.69 percentage points). This relates to the SOLI phase range, in which individuals experience higher MTRs until the SOLI starts applying as before the policy change at 63,526 €. Hence, the variation in MTRs in this study is quite large, ranging from a reduction of 14.65 percentage points to an increase of 2.69 percentage points, though the majority of MTR reductions falls in the range of 1 to 2 percentage points. However, Figure 3.1 underlines a problem, to be further considered in setting up the empirical design: The changes in MTR

[7]Meghir and Phillips (2010) found potential displacement effects between effort exerted and labour hours, which are thus not captured by my model.

[8]Note, that the stark drop in MTRs between annual incomes of 9,409 € and 9744 € mainly arises due to an increase in income tax exemptions for low income earners, and is only marginally affected by the abolition of the SOLI.

rates are highly dependent on income, and unaffected individuals are solely concentrated on the lower and higher end of the income distribution.

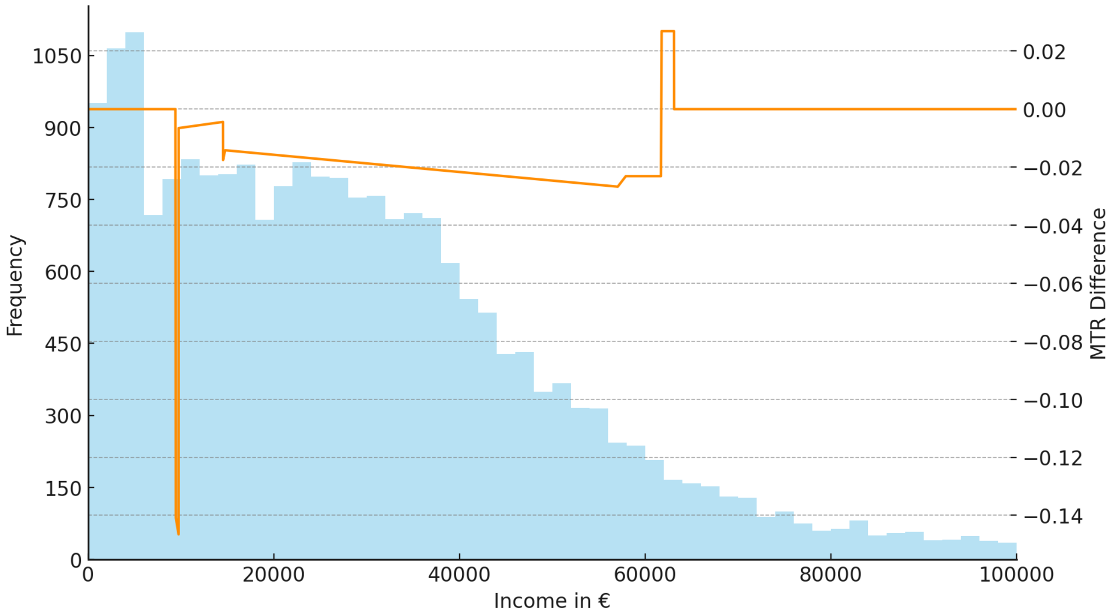


Figure 4.1: Histogram of 2020 Income and Imputed Changes in Marginal Tax Rates

Note: The Figure displays a histogram of income in 2020, the treatment (MTR) determining variable with corresponding scaling on the left y-axis. The changes on MTR are indicated in orange, based on the right y-axis. These changes are computed at constant, 2020, income. The sample is restricted as for the main analysis.

4.2.2 Socio-Demographic Controls

I control for age, number of children, and marital status, in line with previous literature. As income is a direct determinant of the MTR, it cannot be controlled for. However, I draw on 2019 lagged income, where I control for missing observations with a dummy. The age of an individual is computed by subtracting the individual's birth year from the current year and thus given in years. As the SOEP surveys on a household level, the number of children is not inferred from individual responses, but computed as the number of minors (below 18 years old) living in a given household. Lastly, marital status differentiates individuals as "married/living with a partner", "single", "widowed", "divorced" or "separated (legally married)". The first category poses a problem for the isolation of individuals who can draw on joint income taxation, as this is only available for legally married, not necessarily co-residing individuals. I will treat everyone in this category as married, combined with the legally married individuals from the "separated" category.

4.3 Sample Selection

Firstly, I restrict my sample to individuals with annual labour earnings in 2020 of less than 100,000 €. There is some variation in MTRs in the range of 270,501 € and 274,612 €, however this group is in the sample negligibly small and thus not suited for the analysis[9]. Such high income earners are furthermore not representative of the broader public and do not serve as a good control group, regardless. With the variation in MTRs following of the abolition of the SOLI it is unfortunately not possible to examine in the scope of this work estimates for high-income earners for the current German context. Secondly, the sample is restricted to individuals with positive labour earnings. This does not necessarily entail that all individuals show positive working hours, as it is possible to have earnings without working hours (for example during maternity leave). I further exclude the small minority of individuals, lacking information on gender.

Furthermore, married couples[10] need to be treated separately for the purpose of this analysis. Married within the survey are individuals who are "married/living with a partner" and those who are "separated (legally married)". I use the extensive SOEP questionnaire information to I identify the two main earners of the household: household head and the partner of the household head. Given that they are married, I can then jointly analyze their income and apply the German income tax formulas on their joint earnings. However, this method has three caveats: For one, it assumes tax bracket behavior, rather than uniquely identifying it. I assume that couples always chose the financially most beneficial option. This is due to a lack of data within the survey on tax brackets chosen (the relevant question was discontinued in 2016 by the DIW). Secondly, this approximation method is not able to recover MTRs (or for that matter, any tax-related information) on married couples not co-residing, by definition of the process. Hence, this work entirely disregards married individuals, who live (and consequently were surveyed) without their partner. Note, that this is a small subsample of the greater population and not of high relevance: The German Federal Agency for Civic Education found in a report based on 2019 microcensus data for the statistical bureau, that 5.5 % of all individuals living alone are married[11] (BPB, 2021). Lastly, as mentioned in section 4.2.2, the main marriage indicator itself is just a proxy, as it can include individuals who co-reside, but are not legally married. Such individuals are obliged to follow individual

[9] In my sample this group comprises six individuals.

[10] Married in this context refers to legally married individuals and disregards registered partners, the legal equivalent accessible for homosexual couples before the legalization of same-sex marriage in 2017. This comes as the SOEP did not adapt labelling for homosexual couples, rendering the identification of applicable tax classes impossible.

[11] However, this statistic also comprises separated individuals, which are still married but not together anymore. Thus, the sample excluded in my analysis can be expected to lie well below 5.5 % of individuals living alone.

income taxation, however I cannot separate them. Thus, I will proceed as follows: The main analysis will be set up including all marriage status categories, after which I will split the sample along my marriage proxy.

Further, I need to demarcate the sample with regards to the timeframe considered. Figure 4.2 shows attrition of the sample going back in time. Note, as the sample decreases going back in time, it is not only attrition affecting sample size, but also new entries to the labour market and/or the SOEP panel itself. Based on the treatment determining year 2020, I have a maximum number of 12,997 unique individuals, under the constraints described above. Due to the significant decrease in annual observations with every year further from the policy implementation in 2021 and consequently inhibited representatives of the pre-treatment period, I analyse three distinct samples (as shown by the red lines in Figure 4.2): The largest sample comprises 12 years, starting in 2010. A smaller sample analyses 5 years, beginning with 2017 and the smallest sample solely comprises the pre- and post-treatment years, 2020 and 2021.

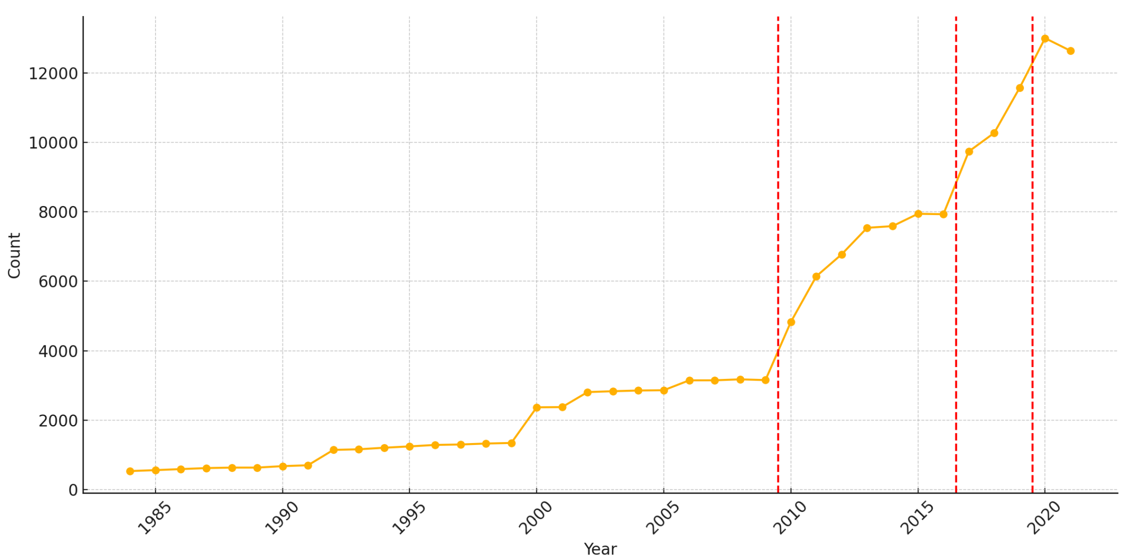


Figure 4.2: Number of Individuals in the Sample over Time with Three Sample Cut-Offs

Note: The Figure displays a count of individuals present in the filtered sample with non-missing treatment values. The three proposed sample cut-offs, correspond to years of significant sample attrition. The cut-offs are the years 2020, 2017 and 2010, respectively and indicated by red vertical lines in the Figure.

4.4 Descriptive Statistics

Table 4.1 outlines the summary statistics for the main variables of interest across different groups. Looking at Table 4.1 shows that the average individual works 1,397 hours per year as of column (1): That entails a weekly average of above 26 hours, disregarding holidays and maternity leave. This number is substantially higher for married individuals, as compared to unmarried ones. However, looking at gender in columns (4) and (5) reveals a different picture. Males work 35 % more than females. Furthermore, treated individuals appear to work more than untreated ones, a difference likely driven by the underlying income dimension.

Further, note that Table 4.1 briefly outlines average values for the MTR at constant income of 2020 - as elaborated upon in equation (1). The average individual has an MTR (including SOLI) of 26 %. Treated individuals experience substantially higher MTRs at 32 %, compared to a mere 15 % among untreated. This comes as the control group comprises a substantial fraction of tax exempted individuals. Columns (6) and (7) of Table 4.1 highlight substantial differences between treatment and control group in terms of age and number of children.

The histogram of 2020 labour earnings in Figure 4.1 shows, that income is mostly distributed around the bottom and middle of the distribution. Observations quickly get more scarce in the top range considered in this analysis. Figure 4.1 furthermore includes a line plot of the change in MTRs at constant income, as of Panel (B) in Figure 3.1. This serves to visualize, what income groups are affected to what extent by the change in the tax system.

5 Methodology

In order to obtain causal estimates on the effects of the MTR on labour supply, some creativity is needed. A simple regression of labour supply on MTRs overlooks potential simultaneity bias and endogeneity concerns. To resolve such concerns of a simple naïve comparison, I adapt elements of Ichino et al. (2022) and the seminal work by Gruber and Saez (2002) in the field of labour supply elasticity to recover causal estimates. In this section I will outline the econometric specifications that serve the purpose of my research question and provide an overview and evaluation of the corresponding assumptions.

5.1 Empirical Specifications

Applying elements of previous empirical designs in a new context, I firstly need to reformulate the MTR in the post-policy year 2021 as a function of labour income in the pre-policy year

Table 4.1: Descriptive Statistics of the Main Variables

	(1) **All**	**(2)** **Married**	**(3)** **Unmarried**	**(4)** **Female**	**(5)** **Male**	**(6)** **Treated**	**(7)** **Untreated**
Labor Earnings	28,784.38	30,666.71	26,347.47	27,352.16	30,308.30	29,206.68	27,896.03
	(20,061.75)	(19,602.84)	(20,384.36)	(19,436.03)	(20,597.90)	(15,381.21)	(27,391.67)
Annual Working Hours	1,397.48	1,475.62	1,296.31	1,194.55	1,613.39	1,511.74	1,157.10
	(1,003.60)	(984.50)	(1,018.94)	(915.60)	(1,047.13)	(967.65)	(1,034.92)
MTR (imputed)	0.26	0.30	0.24	0.26	0.27	0.32	0.15
	(0.15)	(0.14)	(0.16)	(0.15)	(0.15)	(0.07)	(0.21)
MTR change (imputed)	-0.01	-0.01	-0.01	-0.01	-0.01	-0.02	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Age	45.14	48.86	40.33	44.82	45.48	43.64	48.29
	(13.39)	(11.46)	(14.15)	(12.91)	(13.87)	(11.76)	(15.83)
Gender (Female)	0.52	0.51	0.53	1.00	0.00	0.51	0.52
	(0.50)	(0.50)	(0.50)	(0.00)	(0.00)	(0.50)	(0.50)
Number of Children	0.83	1.08	0.50	0.83	0.83	0.90	0.68
	(1.10)	(1.21)	(0.85)	(1.09)	(1.12)	(1.11)	(1.09)
Observations	105,952	59,778	46,174	54,619	51,333	71,814	34,138

Note: This Table presents the mean values by different groups, indicated in the respective columns. The sample is restricted as outlined in section 4.3 and covers all years starting from 2010. Labour Earnings are measured in Euros, Age in years, and number of children is a count. Standard errors are given in parentheses.

2020, as the MTR is endogenous to labour supply. With this, I can rule out simultaneity bias: MTRs are a direct function of income, whereas income is a result of labour hours. An individual, who changes their work hours (and with that income) as a result of a tax change, will inherently face different tax rates (due to the income change). I effectively rule out endogenous labour supply responses to the abolition of the SOLI and base my computations on the parameter $\widetilde{\tau}_{it}$[12]. This variable states the MTR of an individual according to the tax schedule of 2020 and 2021 ($T'_{i,2021}$), respectively, with regard to constant income of the individual in 2020 $Y_{i,2020}$:

$$\widetilde{\tau}_i = \begin{cases} T'_{i,2020}(Y_{i,2020}) & \text{for year 2020,} \\ T'_{i,2021}(Y_{i,2020}) & \text{for year 2021} \end{cases} \tag{1}$$

However, a simple ordinary least-squares (OLS) regression using $\widetilde{\tau}_{it}$ still likely suffers from endogeneity concerns. Figure 4.1 shows that changes in MTRs at constant income are highly dependent on income itself. Income on the other hand is not randomly distributed over society or my sample, but likely correlates with both observed and unobserved characteristics influencing labour supply. Even if I control for observed characteristics, such as gender and age, unobservables render an OLS estimate biased. Think of, for example, ambition: High ambitions might be correlated with higher working hours and likewise correlated to higher MTRs, via a higher income. Thus, the OLS estimator cannot be interpreted causally.

In a first step, I implement an ordinary Difference-in-Difference (DiD) specification with binary treatment, similar to models used by Eissa (1995) and Eissa (1996). In identifying an average treatment effect on the treated, this econometric method hinges on the assumption of common trends to deliver a causal estimate. In terms of potential outcomes, this assumption entails that the labour supply of the treatment group would have stayed on the same trend as in the pre-treatment period, if not for the treatment. Thus the assumption supports the idea, that the control group (comprising those individuals who experienced no change in their MTR), serves as a good counterfactual. If both the treated and the untreated follow parallel trends in the pre-treatment period, the change observed for the treated can be interpreted causally, relative to the change of the control group in the post-policy period. I can test this assumption both graphically and statistically. The results of these tests are in section 5.2.

My treatment group is formed by those individuals who experienced a change in their MTR ($\widetilde{\tau}_{i_t}$) between 2020 and 2021. Note, to guarantee comparability between models, I do not exclude the minority of individuals that experience a MTR increase[13] in the main

[12]Any mentioning of MTR hereafter refers to the synthetic MTR.

[13]Individuals with income between 62,128 and 63,525 € in 2020 experienced an increase in MTRs.

results, but present results excluding those in the Appendix. I thus define the treatment group as all individuals, that experienced a change in their reformulated MTR (for the vast majority: decrease) in this simple DiD. In contrast, the control group comprises individuals who, at constant income, showed no change in $\widetilde{\tau}_{i_t}$. I define a binary treatment dummy D_i as in equation (2). The corresponding model specification further has dummy $post_t$ taking value 1 for the post-reform year 2021 and 0 before and highlights the coefficient of interest β_3 in the estimation of labour supply L^s_{it} (given in annual hours) in the interaction term. In this model, I also account for year fixed effects (FE) γ_t. Note, that in models restricting the sample to 2 years, all year-invariant factors are captured by the coefficient of $post_t$ and I consequently drop γ_t due to perfect collinearity. Further, the panel structure of the SOEP dataset allows me to cluster standard errors at the individual level across all models in line with Abadie et al. (2023). I thus accommodate potential correlation of errors within individuals: Clustering at the individual level helps to account for heteroskedasticity by allowing the variance of errors to differ across clusters.

$$D_i = \begin{cases} 0 & \text{if } \widetilde{\tau}_{i,2020} = \widetilde{\tau}_{i,2021}, \\ 1 & \text{if } \widetilde{\tau}_{i,2020} \neq \widetilde{\tau}_{i,2021}. \end{cases} \tag{2}$$

$$L^s_{it} = \beta_0 + \beta_1 D_i + \beta_2 post_t + \beta_3 (post_t \times D_i) + \chi V_{it} + \gamma_t + \varepsilon_{it} \tag{3}$$

The SOEP offers a comprehensive set of background characteristics. I extend the DiD specification by standard demographic variables, in particular age, gender and number of children. These personal characteristics are captured by an individual control vector specified under V_{it} in equation (3). I further can control for lagged income, as in Gruber and Saez (2002). With that, I can capture unobservable determinants of previous (2019) income, which are likely correlated with the treatment determining income of 2020. Gruber and Saez (2002) also lay out the danger of over-controlling when dealing with a single tax reform, hence the controls employed in this study are a conservative bound. My selection of controls aims to capture factors, which could heterogeneously affect treated and untreated individuals.

However, the simple DiD specification (3) comes at a cost: It ignores the variation in MTRs and solely differentiates individuals who experienced a change (for the vast majority; decrease) following the reform, from those who did not. The term D_i effectively condenses the variation in $\widetilde{\tau}_{i_t}$, that is of high relevance for policy implications. I therefore introduce a continuous DiD model, allowing for continuous treatment and zoom into the effect nature. This continuous version is inspired by Callaway et al. (2024) and hinges on the same parallel trends assumption as a standard DiD. The dynamic specification in equation (4) differs from the static model in equation (3) in one aspect: Firstly, the treatment dummy D_i is replaced by the continuous treatment variable $\Delta\widetilde{\tau}_i$. The coefficient of interest is β_3.

$$L_{it}^s = \beta_0 + \beta_1 \Delta\widetilde{\tau}_i + \beta_2 post_t + \beta_3(post_t \times \Delta\widetilde{\tau}_i) + \chi V_{it} + \gamma_t + \varepsilon_{it} \tag{4}$$

where

$$\Delta\widetilde{\tau}_i = \widetilde{\tau}_{i,2021} - \widetilde{\tau}_{i,2020} \tag{5}$$

Furthermore, I examine effect heterogeneity by gender, number of children and age. Doing so by means of interaction terms effectively renders the DiD specification from equation (3) a triple Differences (hereafter referred to as DDD) strategy. The corresponding regression equation can is specified in equation (6). Variables follow the same naming as before, with exception of G_i, which is an indicator of the respective group I am isolating. The coefficient of interest is thus β_6, which isolates differential effects by group.

$$\begin{aligned} L_{it}^s = \beta_0 + \beta_1 \Delta\widetilde{\tau}_i + \beta_2 post_t + \beta_3(post_t \times \Delta\widetilde{\tau}_i) + \beta_4(post_t \times G_i) \\ + \beta_5(\Delta\widetilde{\tau}_i \times G_i) + \beta_6(post_t \times \Delta\widetilde{\tau}_i \times G_i) + \chi V_{it} + \gamma_t + \varepsilon_{it} \end{aligned} \tag{6}$$

5.2 Parallel Trends Assumption

For the control group to serve as a credible counterfactual in any DiD setting, the two groups should be similar. As the MTR change is directly based on income (as is $\Delta\widetilde{\tau}_i$ on income in 2020), this underlying logic might be violated. MTR changes are dispersed across the income range of 9,409 € and 63,525 € annual labour income[14] with varying degree of the size of MTR change. This treatment range covers most working citizens, with the exception of some part-time workers at the bottom of the distribution and top earners at the upper range. In this section I introduce evidence on the comparability between the two groups and examine if the control group serves as a suitable counterfactual in my DiD setting.

In testing the parallel trends assumption of my DiD specifications, I first introduce graphical evidence. Figure 5.1 shows the means of annual labour supply (in hours) for individuals in the treatment group (red) and control group (blue). In evaluating parallel trends, the availability of many time periods is an advantage as it allows for a more conservative view. Ignoring the labour supply fluctuations around and following the time of Germany's reunification in 1990, the trends appear somewhat parallel. The Figure effectively compares low-

[14] There is one more change for earners with income between 270,501 € and 274,612 € which by definition of the sample is excluded (see section 4.3).

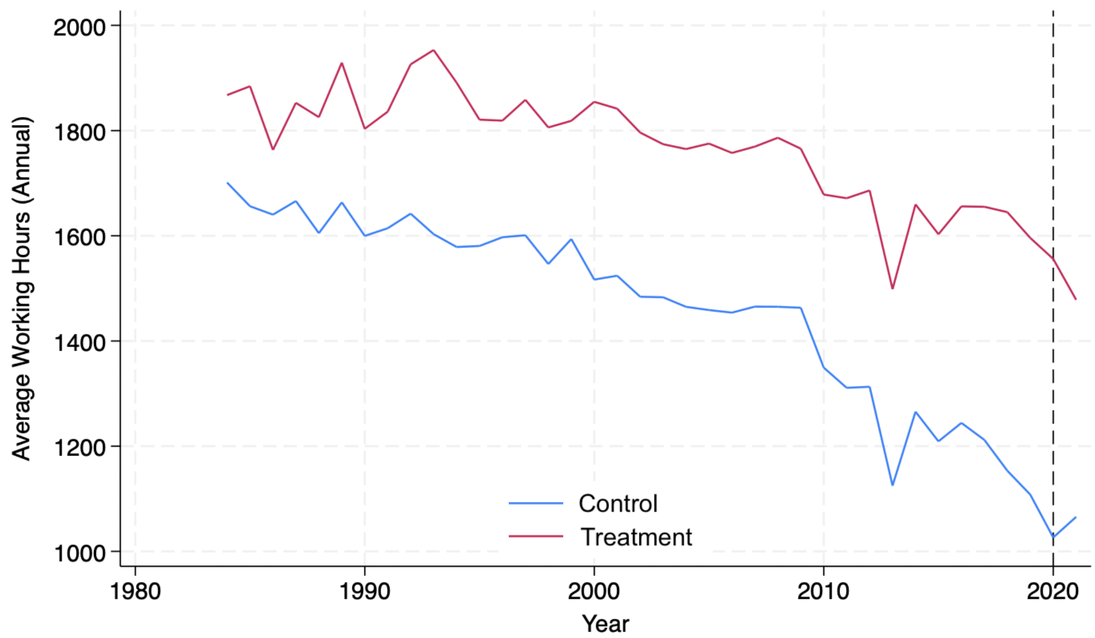


Figure 5.1: Means of Annual Working Hours for Treatment and Control Group since 1984

Note: The Figure displays average working hours by group: individuals experiencing a change in the marginal tax rate in 2021 (treatment group) and those who do not (control group). The sample is restricted as for the main analysis (see section 4.3). The dotted line separates the pre-reform and the post-reform period.

and high-income earners to the broad middle of the distribution, rendering treatment and control group different in characteristics. First graphical inspection supports the use of a "simple" DiD model as in equation (3). Trends are mostly parallel, apart from the decline observed in 2014, which affected the control group more. All in all, Figure 5.1 paints an optimistic picture of my main identifying assumption.

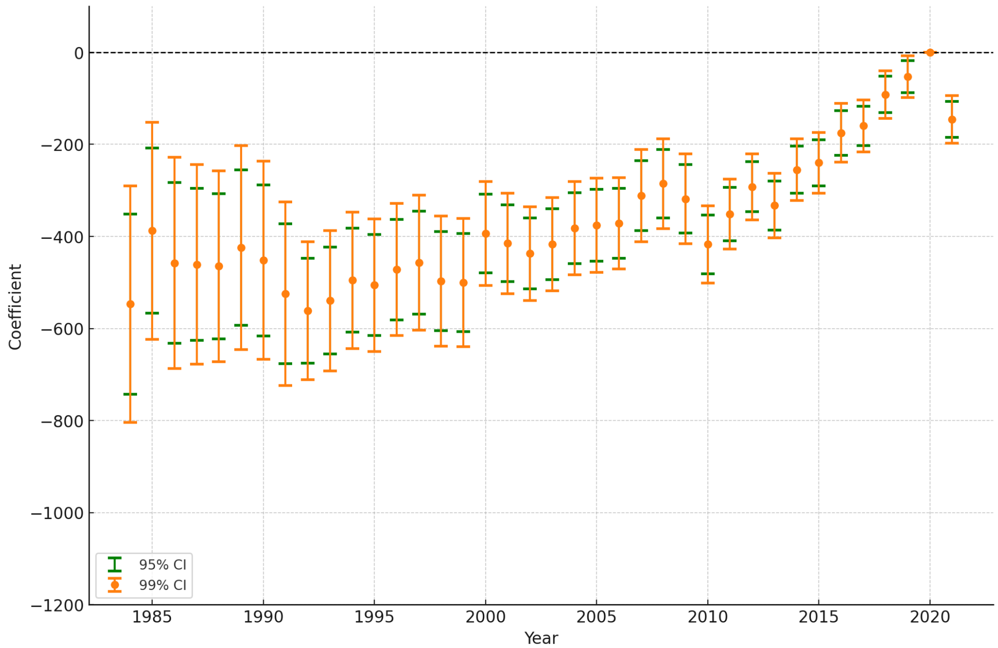
Another way to view the parallel trends assumption is by considering an event study version of the empirical specification. In that, I interact the binary treatment indicator with year-dummies. The advantage of this approach is, that it allows me to control for observed individual characteristics and effectively evaluate any divergence of the treatment group relative to the control group year by year. The corresponding regression equation (7) includes the interaction between the treatment indicator and year dummies. The coefficients of $\beta_{2,t}$ should not be significantly different from zero in the pre-policy years to support the parallel trends assumption. I thus compare treated to untreated individuals in this event study.

$$L_{it}^s = \beta_0 + \beta_1 T_i + \beta_{2,t}(T_i \times \gamma_t) + \gamma_t + \varepsilon_{it} \tag{7}$$

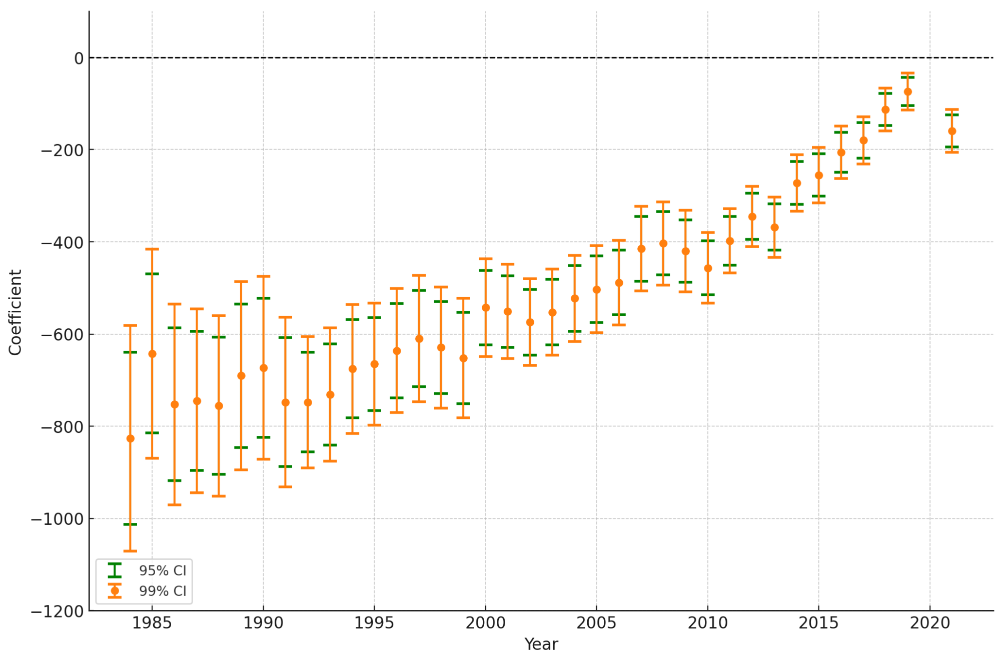
Appendix Table A.2 displays coefficients of interest $\beta_{2,t}$ of equation (7). The different columns indicate different model alterations. Coefficients and their 95 % (99 %) confidence intervals are accordingly displayed in Figure 5.2. Whilst panel (A) shows results excluding controls, panel (B) displays coefficients including controls. Treatment and control group appear to have differential patterns in labour supply over time. Coefficients displayed are at the 1 % level significantly different from 0 for most years. Unfortunately, the event study does not support the assumption of parallel trends. However, it indicates a trend, differential for treated compared to untreated individuals, which breaks in the post-policy year in 2021. With decreasing differences across treatment and control group over time up until the policy change, the negative coefficient of 2021 shows a change in trend and might still support the notion of a policy effect. Furthermore, Figure 5.2 gives first insights into the effect of the SOLI abolition. Treated individuals (those who experience a change, mostly decrease, in their restated MTRs) appear to decrease annual working hours, compared to individuals who experience no change in MTRs. However, the Figure overall highlights that DiD results of this thesis are to be interpreted with caution and can not necessarily be taken as causal.

A potential solution for this divergence between treatment and control group trends is a regression discontinuity design (RDD), to be incorporated into the DiD framework. Reducing the bandwidth around the jumps in MTR change could provide credible counterfactuals by minimizing differences in potential outcomes. However, as of Figure 4.1, the groups adjacent to the untreated individuals, are both very small and can be considered outlier cases: One group experiences substantially large reductions in MTRs (incomes between 9,409 € and 9744 €), the other undergoes an increase in MTRs (incomes between 62,128 € and 63,525 €). Due to small sample sizes in both groups an RDD approach is thus unfeasable in this context.

Lastly on the note of identifying assumptions, it is worth mentioning, that DiD models rest on another assumption which regards anticipation effects. Individuals should not change their behavior before the policy goes into place in anticipation of the treatment. In this context, anticipation effects are highly unlikely: There is no reason for individuals to adopt labour supply before a change in tax rates takes place. Figures 5.2 and 5.1 support this notion: There is no evidence of anticipation effects prior to the abolition of the SOLI.



Panel (A)



Panel (B)

Figure 5.2: Event Study Coefficients of the Year to Treatment Interaction Terms and corresponding 95 % (99 %) Confidence Intervals

Note: Panel (A) displays the coefficients of a simple regression without any controls. Corresponding results can be found in column (1) of Table A.2. Panel (B) shows coefficients of a regression including controls (age, gender, number of children in the household, marital status and 2019 income). Corresponding results can be found in column (2) of Table A.2.

6 Results

6.1 Simple Difference-in-Differences

The estimates of the static DiD model in Table 6.1 serve as a baseline of introducing DiD models in this thesis and pose a point of comparison for the subsequent continuous DiD. The Table displays the whole sample in Panel (A) and then divides individuals based on marriage status in Panels (B) and (C). Note, that this Table comprises the identical sample as the continuous DiD in Table 6.2 and thus includes individuals who experience an increase in the synthetic MTR. I repeated the analysis excluding this minority group in Appendix Table A.3, focusing on model specifications with controls. Coefficients remain virtually the same in sign, size, and significance. Given this, subsequent analyses will be based on the identical samples, to allow for comparison between the static and continuous DiD versions. Further, as columns (1) and (2) across all Panels only comprise two years, the use of year fixed-effects is inappropriate: The inclusion of $post_t$ absorbs year FE in those models.

Starting with Panel (A), the DiD interaction coefficient ranges from -157.54 in column (2) to 65.97 in column (6), whilst being statistically significant at the 1 % level across all models. This change in sign and size appears to be driven mostly by the timespan selected. Over time, the sample shows attrition, or more precisely, the sample composition changes: The larger the timespan selected, the smaller the sample in those years (as shown in Figure 4.2). This can raise issues of representativeness. However, this pattern can also reflect the violation of parallel trends assumption, examined in Figure 5.2. Given attrition concerns and new entries to the panel (effectively rendering the pre-policy period an invalid point of comparison), I prefer models with a shorter time span, such as 2 or 5 years in columns (1) to (4). Further, the inclusion of controls alters the coefficient. Column (2) shows that controlling for age, gender, marital status, number of children and 2019 income increases the absolute size of the coefficient, compared to column (1). However, this pattern is not constant across different time horizons: The coefficient decreases in absolute size (gets closer to 0) between columns (3) and (4) by the inclusion of aforementioned controls and year-FE. In the 12-year model, including controls and year-FE increases the coefficient. The same holds for Panels (B) and (C). Overall, controls appear to increase the magnitude of coefficients. Given models (2) and (4), individuals who experienced a change (reduction) in their MTRs decrease their labour supply by 47.63 to 157.54 hours annually on average, compared to individuals who do not experience a change in MTRs. This effect is statistically and economically significant: Relative to baseline means of 948 and 1,025 the effect poses a reduction in labour supply by 4.6 to 14.6 %.

However, as tax behavior can only be proxied for married individuals given the system of optional joint taxation in Germany, it is important to split the sample. Looking at Panel (B) of Table 6.1, the coefficient of interest is larger in magnitude across all models as compared to the married sample in Panel (C). For unmarried individuals, the effect size ranges between -198.36 in column (2) to 53.24 in column (5). Statistical significance varies: Estimates in the 12-year model without controls in column (5) are not statistically significant from zero. Based on my more preferred model specifications in columns (2) and (4) the coefficient is economically significant, indicating a labour supply reduction of 23.6 to 7.6 %. Among married individuals the effect size ranges from -102.47 to 81.44 in Panel (C). Here, the coefficient of the medium timeframe of 5 years in column (4) including controls is statistically insignificant and indicates that married individuals might not be significantly affected by the abolition of the SOLI. The difference between the married and unmarried sample here could be driven by decreased flexibility of married individuals in adapting their labour hours. An alternative explanation is that married partners share the same MTR in a system of joint taxation, this difference might also reflect a lower responsiveness due to the system itself.

6.2 Continuous Difference-in-Differences

Going over to a model of continuous treatment in DiD, Table 6.2 displays the main results of my thesis: The coefficients of the interaction terms in the DiD model specification from equation (4). It follows the same structure as the previously discussed Table 6.1 Panel (A) includes all observations, not differentiating between married and unmarried individuals. Without controls as of column (1), an individual increases labour supply by 20.04 hours annually on average, facing a 1 percentage point increase in MTRs. This effect is statistically significant at the 1 % level. As most MTR changes are a reduction of tax burden, I prefer the following interpretation: An individual decreases labour supply by 20.04 hours on average for a 1 percentage point decrease in MTRs, compared to unaffected individuals, all other things constant. Average MTR change values of -2 percentage points, render the coefficients smaller, compared to the static DiD model, nevertheless being economically significant: Given this average value, labour hour reductions lie between 15 and 44 hours annually. Compared to a baseline mean of 948 (1,025) annual hours, this poses a predicted decrease of around 5 % (2 %). Controlling for age, gender, marital status, number of children and 2019 income, the effect size slightly increases to 22.23 as of column (2). However, considering a model with 4 pre-policy years as in columns (3) and (4), the coefficient becomes statistically insignificant. Model (6) shows a statistically significant coefficient: Here, a 1 percentage point decrease in the MTR appears to increase annual labour supply by 14.73 hours. As with the static DiD

Table 6.1: Difference-in-Differences Regression Results of Labour Supply

	Panel (A): All Individuals					
	2 years		**5 years**		**12 years**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times D_i$	-145.57***	-157.54***	-72.03***	-47.63***	53.98***	65.97***
	(19.99)	(17.65)	(18.93)	(16.50)	(20.20)	(17.79)
Observations	25,631	25,631	57,203	57,203	105,952	105,952
R-squared	0.05	0.44	0.05	0.44	0.03	0.36
Mean (Untreated)	947.77	947.77	1,024.78	1,024.78	1,157.10	1,157.10
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes
	Panel (B): Unmarried Individuals					
	2 years		**5 years**		**12 years**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times D_i$	-176.25***	-198.36***	-104.57***	-68.89***	43.24	53.24**
	(25.67)	(22.39)	(25.24)	(20.96)	(27.59)	(22.29)
Observations	13,410	13,410	27,930	27,930	46,174	46,174
R-squared	0.05	0.59	0.05	0.60	0.03	0.52
Mean (Untreated)	839.43	839.43	911.80	911.80	1,067.90	1,067.90
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes
	Panel (C): Married Individuals					
	2 years		**5 years**		**12 years**	
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times D_i$	-102.47***	-83.54***	-58.47**	-15.90	28.97	81.44***
	(31.07)	(27.89)	(28.16)	(25.53)	(29.46)	(26.96)
Observations	12,221	12,221	29,273	29,273	59,778	59,778
R-squared	0.04	0.29	0.04	0.32	0.03	0.28
Mean (Untreated)	1,099.86	1,099.86	1,149.62	1,149.62	1,224.19	1,224.19
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes

Note: Coefficients of interest and standard errors (in parentheses) are presented. The time span varies by model: Columns (1) and (2) cover 2020-2021, (3) and (4) 2017-2021 and (5) and (6) 2010-2021. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from equation (2), indicating a change in MTRs. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * $p<0.1$, ** $p<0.05$, and *** $p<0.01$.

model, I prefer the shorter time periods of 2 to 5 years, due to sample attrition.

Following, I split the sample into unmarried and married individuals, in Panel (B) and (C), respectively. Panel (B) of Table 6.2 shows coefficients ranging from 26.30 to -4.39 for unmarried individuals, largely fluctuating with the timespan considered. The interpretation reads as before: An unmarried individual decreases labour supply between 26 and 10 hours on average, following a 1 percentage point decrease in MTRs in the preferred models of columns (1) to (4). However, not all coefficients are statistically significant. Further, note that the average married individual in the treatment group experiences a tax rate cut of about 2 percentage points. Thus, annual labour supply responses are predicted to fall in between a reduction of 53 to 20 hours for unmarried individuals.

Conversely, estimates for married individuals are smaller in size, ranging from 10.04 to -29.20 in Panel (C). Only one coefficient is statistically significant and corresponds to a less favoured regression in column (6) with a 12 year span. The overall stark differences between married and unmarried individuals underline the plausible narrative of decreased flexibility in a system of (optional) joint taxation. Further, this finding aligns with models of household production functions, such as Heckman (2015) and Pollak and Wachter (1975): Married individuals might be less flexible in adjusting their labour hours, as division on household inputs (such as employment and household labour) are often taken at the household or partner level. If this is the case, one can expect married individuals to be less responsive to changes in the tax schedule, compared to more flexible, unmarried, workers.

To see to what extent these results are driven by the minority of treated individuals, that experienced tax rate reductions of up to 14.65 percentage points, I excluded this outlier group and repeated the same analysis[15]. These results can be found in Appendix Table A.4. Overall, the exclusion of the outliers renders all coefficients substantially larger in magnitude across all three Panels and timeframes, whilst being (mostly) of high statistical significance. Coefficients in the preferred model specifications of 2 and 5 years from columns (1) and (2) indicate that a 1 percentage point reduction in MTRs leads to a reduction of 24.64 to 61.79 annual labour hours. This effect is statistically significant at the 1 % level and further economically impactful. Comparable models from the main Table 6.2 show a substantially smaller effect size between 7.32 and 22.23. The pattern, that unmarried individuals are more responsive to tax rate changes, is supported by this robustness analysis as Panel (B) and (C) show. The stark difference in magnitude following the exclusion of outliers can have two potential causes: Firstly, this income group might generally show heterogeneous effects. Secondly and on a related note, the DiD models assumes a linear effect nature. If individuals

[15]Excluded individuals are those with income in 2020 between 9,409 and 9,744 €. This group accounts for MTR reductions in the range of 14.00 to 14.65 percentage points. In that, I exclude 85 individuals from the sample.

in this group do not adjust their labour hours linearly extreme corresponding to the MTR changes they face, the coefficient can be downward biased. Excluding this group enables the model to better fit the linear relationship observed in the other data, which increases the magnitude of the coefficients. It appears that the excluded group did not adjust labour hours, as one would predict based on other observations, facing such a stark decrease in synthetic MTRs.

Generally speaking, shorter time span models accounting for controls and year FE as throughout columns (2) and (4) are my preferred specifications, both for the continuous and the static DiD model. All in all, the two models are comparable. However, estimates obtained in the continuous version are substantially smaller in magnitude, in particular when accounting for average values of treatment intensity. However this difference is mitigated, once outliers in MTR reductions are removed. I further prefer the continuous model over the static one, as it allows for a more nuanced understanding of the treatment effect. Given the evidence from the preferred models (2) and (4) in both DiD specifications, I reject Hypothesis 1. Indeed, the data shows the opposite: A decrease in MTRs causes individuals to work less hours. These findings suggest that in the context of this study income effects outweigh substitution effects. However, the interpretation of all coefficients falls short of the parallel trends assumption, which inhibits the causal nature of the identification strategy.

6.3 Robustness Analysis

Hereafter, I focus on the 2 year and 5 year models as my preferred specifications. To examine the robustness of my findings, I implement a placebo test. In that I employ bootstrap sampling to closely mimic the original income distribution: Observations for income are drawn at random from the previous sample, for the treatment determining year 2020. Note, this process samples with replacement, hence the final distribution of income is not necessarily identical but very similar. Subsequently, I repeat the process twice, once for married and once for unmarried individuals. As I found earlier these two groups are different not only in composition and treatment effect but also in treatment allocation, due to joint taxation. Thus, I obtain two new samples, each based on the distribution of income for married and unmarried individuals respectively. In general, these placebo incomes are then used to determine MTRs and my treatment variable $\Delta\widetilde{\tau}_i$. The regression specification is equivalent to the one specified for the continuous model in (4). The corresponding DiD regression results of the placebo parameter can be found in Table 6.3. As seen in Panel (A) for the whole sample, Panel (B) for unmarried and Panel (C) for married individuals, neither the short, nor the medium term models show significant results. This further does not depend on the inclusion

Table 6.2: Continous Difference-in-Differences Regression Results of Labour Supply

	2 years		5 years		12 years	
Panel (A): All Individuals	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	20.04***	22.23***	11.05	7.32	-9.47	-14.73**
	(7.01)	(5.73)	(7.15)	(5.73)	(8.01)	(6.17)
Observations	25,631	25,631	57,203	57,203	105,952	105,952
R-squared	0.03	0.42	0.03	0.42	0.02	0.35
Mean (Untreated)	947.77	947.77	1,024.78	1,024.78	1,157.10	1,157.10
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes
Panel (B): Unmarried Individuals	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	26.26***	26.30***	16.24*	10.04	1.44	-4.39
	(9.17)	(7.44)	(9.09)	(7.16)	(9.73)	(7.25)
Observations	13,410	13,410	27,930	27,930	46,174	46,174
R-squared	0.02	0.56	0.02	0.57	0.02	0.51
Mean (Untreated)	839.43	839.43	911.80	911.80	1,067.90	1,067.90
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes
Panel (C): Married Individuals	**(1)**	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	10.04	13.32	5.80	6.02	-20.02	-29.20***
	(11.01)	(8.92)	(12.11)	(10.43)	(13.15)	(10.35)
Observations	12,221	12,221	29,273	29,273	59,778	59,778
R-squared	0.03	0.28	0.03	0.20	0.02	0.27
Mean (Untreated)	1,099.86	1,099.86	1,149.62	1,149.62	1,224.19	1,224.19
Controls	No	Yes	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes	No	Yes

Note: Coefficients of interest and standard errors (in parentheses) are presented. The time span varies by model: Columns (1) and (2) cover 2020-2021, (3) and (4) 2017-2021 and (5) and (6) 2010-2021. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -14.65 and 2.69. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

of controls. The insignificant DiD estimates in the placebo study strengthen confidence in my main findings from Table 6.2.

7 Heterogeneity Analyses

Going on, this section aims at disentangling average effects discussed in section 5.1 by examining heterogeneity across several dimensions. As outlined in section 5, the heterogeneity analysis follow a DDD approach as of equation (6). Even though the continuous DiD model is my preferred specification, I perform the heterogeneity analyses using the static DiD for ease of interpreting of the coefficients. Results using the continous DDD equivalent of the DiD model in equation (4) are presented in Appendix Tables A.5, A.6 and A.7. I further only use models controlling for covariates, as they are my preferred model specifications. In general, the tables here follow the same layout as previous regression tables.

7.1 Gender

The results of the DDD model, which allows for heterogeneous treatment effects by gender, are presented in Table 7.1. Overall statistically significant positive effects from Panel (A) appear to be purely driven by married individuals in Panel (C), as opposed to unmarried individuals in Panel (B), which show insignificant differential policy effects. Married women who experience a change (again, most often decrease) in their reformulated MTRs increase their working hours by 211.65 to 284.77 hours, compared to married men. This effect corresponds to 5 to 7 full-time (40 hours) working weeks and is economically highly significant. This narrative is underlined by the evaluation in a DDD model allowing for continuous treatment intensity, as in Appendix Table A.5. Note, in contrast to the static version, only models of 5 years show statistical significance. Based on column (2) in Panel (C), married females increase labour supply by 44.05 hours on average in response to a 1 percentage point change in MTRs, compared to married males. Given average values of the main independent variable, this result is smaller in absolute size than the static effect. Given these findings, I cannot reject Hypothesis 2: Women seem indeed significantly more responsive to changes in MTRs, compared to men. However, this differential effect size seems to be solely driven by differences between married individuals.

Table 6.3: Placebo Test: Random Income Allocation

	Panel (A): All Individuals			
	2 years		**5 years**	
	(1)	**(2)**	**(3)**	**(4)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	8.26	6.45	6.26	4.60
	(5.64)	(4.51)	(5.55)	(4.34)
Observations	25,631	25,631	57,203	57,203
R-squared	0.00	0.42	0.00	0.44
Controls	No	Yes	No	Yes
Year Fixed Effects	No	No	No	No
	Panel (B): Unmarried Individuals			
	2 years		**5 years**	
	(1)	**(2)**	**(3)**	**(4)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	-6.47	4.84	-0.32	-4.71
	(7.73)	(6.50)	(7.09)	(5.69)
Observations	13,410	13,410	27,930	27,930
R-squared	0.00	0.55	0.00	0.66
Controls	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes
	Panel (C): Married Individuals			
	2 years		**5 years**	
	(1)	**(2)**	**(3)**	**(4)**
$\text{post}_t \times \Delta\widetilde{\tau}_i$	-2.76	-6.96	-2.85	-13.33
	(7.66)	(9.01)	(9.15)	(11.36)
Observations	12,221	12,221	29,273	29,273
R-squared	0.00	0.26	0.00	0.21
Controls	No	Yes	No	Yes
Year Fixed Effects	No	No	No	Yes

Note: Coefficients of interest and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -14.65 and 2.69. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table 7.1: Heterogeneity by Gender (Female): Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times F_i$	105.73***	163.17***
	(35.06)	(32.52)
Observations	25,631	57,203
R-squared	0.44	0.44
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times F_i$	-12.74	42.26
	(44.63)	(41.74)
Observations	13,410	27,930
R-squared	0.59	0.60
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times F_i$	211.65***	284.77***
	(54.87)	(50.03)
Observations	12,221	29,273
R-squared	0.29	0.32
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from equation (2), indicating a change in MTRs. Control variables are indicated as specified and include age (continuous), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

7.2 Number of Children

As of Panel (A) in Table 7.2 the number of children does not drive heterogeneous effect size: The corresponding coefficients are both positive, but statistically insignificant. However, when splitting the sample into married and unmarried individuals, it appears that married individuals might be differentially affected. Considering the 5 year model of Panel (C), column (2), married individuals with 1 child decrease labour by 59.31 hours on average following a change (most often decrease) in MTRs, relative to childless married individuals. This effect is statistically significant at the 1 % level. As the number of children is applied as a continuous variable in this model, the heterogeneous effect nature adds up. An alternative interpretation is: treated married individuals with 3 children reduce labour by 119 hours on average, compared to untreated married individuals with 1 child. Models allowing for continuous treatment intensity in Appendix Table A.6 do not emphasise this notion. All coefficients appear statistically insignificant. However, this model is not fully comparable to the static model presented in Table 7.2: In order to allow for an intuitive interpretation of the coefficient, the variable corresponding to the number of children, was restated as a binary variable in Appendix Table A.6, only differentiating between individuals with and without children. All in all, individuals with children do not appear significantly more or less responsive to tax rate changes than individuals without children. I thus reject Hypothesis 3.

7.3 Age

This section aims to better understand what drives differences between married and unmarried individuals observed in both the static and the continuous DiD models in sections 6.1 and 6.2. To achieve this, I go beyond my hypotheses and examine the role of age in the effect. Table 7.3 displays results allowing for heterogeneous effects by age, where age is treated as a continuous variable. As in previous heterogeneity analyses, I draw on the simple DiD with binary treatment for the core analysis. Generally, higher age is associated with a more pronounced effect of the change in MTRs, though statistical significance is mostly concentrated within the 5-year models in column (2). Panel (A) shows that treated individuals decreased labour hours by 13.3 to 67.4 hours annually, compared to individuals 10 years younger. The difference between Panels (B) and (C) highlights that this effect is more pronounced among married individuals, rather than unmarried ones. This can also indicate that the heterogeneous effect of age might be non-linear: Married individuals are on average substantially older than unmarried ones (see Table 4.1). If the effect is non-linear, the results in Table 7.3 suggest that older age groups are disproportionally more sensitive to tax rate changes. In

Table 7.2: Heterogeneity by Number of Children: Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times C_i$	14.14	23.47
	(16.65)	(15.66)
Observations	25,631	57,203
R-squared	0.44	0.44
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times C_i$	-18.30	-37.25
	(25.17)	(23.99)
Observations	13,410	27,930
R-squared	0.59	0.60
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times C_i$	12.20	59.31***
	(21.62)	(20.50)
Observations	12,221	29,273
R-squared	0.29	0.32
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from equation (2), indicating a change in MTRs. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

summary, these findings suggest that older individuals are more responsive to tax changes. Such a pattern might reflect the effect of wealth, which concentrates over the life-cycle in tax responsiveness: More wealthy individuals are likely more flexible in adjusting their work input, due to being lesser constraint on sustaining labour income.

I further look at the heterogeneous effect nature within the continuous DiD model. Similar to the analysis regarding number of children in section 7.2, I have to reformulate age to be binary. Appendix Table A.7 displays corresponding results, where O_i is a binary variable, taking 1 for old individuals (aged 40 and above) and value 0 for individuals below the age of 40. Firstly, it is notable that all (despite one) coefficients of this model are statistically insignificant at the 10 % level, regardless of the model chosen, or sample selected (marriage status). Most coefficients are positive, indicating that a reduction in MTRs leads to a more pronounced decrease in labour hours among old individuals. Despite a lack of statistical significance, these results are in line with the static version in Table 7.3. Notably, among unmarried individuals, the triple interaction coefficient of the 5-year model in column (2) of Panel (B) is statistically significant at the 10 % level. Here, a 1 percentage point decrease in MTRs causes older unmarried individuals to work 22.80 hours less annually compared to younger unmarried individuals. This entails that responsiveness is more concentrated among older unmarried individuals, in contrast to older married individuals. This dynamic could reflect, that unmarried individuals have less constraining factors in their decision on working hours, such as more flexible finance.

Table 7.3: Heterogeneity by Age: Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times A_{it}$	-1.33	-6.74***
	(1.10)	(1.06)
Observations	25,631	57,203
R-squared	0.44	0.44
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times A_{it}$	0.05	-4.44***
	(1.46)	(1.38)
Observations	13,410	27,930
R-squared	0.60	0.60
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times D_i \times A_{it}$	-4.06*	-10.69***
	(2.20)	(2.05)
Observations	12,221	29,273
R-squared	0.29	0.32
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from equation (2), indicating a change in MTRs. Control variables are indicated as specified and include gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

8 Conclusion

Firstly, my main findings suggest that in the context of the abolition of the SOLI in 2021, the income effect outweighs the substitution effect: German workers appear to work less hours facing lower tax rates and thus tend to sustain a standard in living, rather than substituting work for leisure. This finding rejects the first hypothesis: A decrease in MTRs appears to decrease labour supply. This result contradicts most of the literature[16], which previously found larger substitution than income effects and goes against first computations by Bonin et al. (2019) on the effects of the SOLI abolition. However, this effect is concentrated among unmarried individuals. This might also reflect a change in attitude: individuals may prefer to work less if they can sustain a given standard of living, rather than increasing their standard of living. A recent example of such a mentality are the tedious negotiations between the German Train Drivers' Union (GDL), representing over 40,000 workers, and the "Deutsche Bahn" (German Railway): The two parties agreed on a reduction of the regular work week from 38 to 35 weekly hours by 2029 with full wage compensation (GDL, 2024). Continuing with the examination of heterogeneity by gender, married women appear to be vastly more responsive to changes in the tax schedule, compared to married men. I thus cannot reject Hypothesis 2. Looking at the number of children in a household, I found mostly insignificant results and consequently rejected the third and final hypothesis. Going further, I lastly found that older individuals are more responsive to tax changes, showing starker reductions in labour hours than younger individuals.

Notably, my thesis has several limitations. Most obviously, the identifying DiD assumption does not appear to hold, despite promising initial graphical evidence. This limits a causal interpretation of my results. I therefore propose to consider coefficients with caution and reformulate their interpretation: A 1 percentage point decrease in MTRs is associated with an average reduction of 7.32 to 20.04 (excluding outliers this effect amounts to 24.64 to 61.79) hours annually in labour supply, as compared to untreated individuals, ceteris paribus. Furthermore, the data might suffer from measurement error across several dimensions: Labour hours are partially based on the hours of the previous year. My outcome variable might thus not be fully accurate. Additionally, the tax rates for married individuals could only be proxied[17], due to a lack of data on tax brackets chosen for married individuals. Moreover, the panel does not allow for the analysis of long-run effects: Only one post-policy year (2021) is published as of now. The effect might develop over time, particularly if labour hours are

[16] As described in section 2, Ashenfelter et al. (2010) found similar results on a sample of male, New York City taxi drivers.

[17] I assumed all married individuals to take the most financially advantageous tax bracket.

rigid and take time to adjust.

Further, my sample selection has some important implications. The exclusion of high-income earners restricts potential policy implications of my findings: The responsiveness of high-income earners is of particular relevance in designing and restructuring redistributive measures. Previous literature, as highlighted in section 2, has generally found higher responsiveness of high-income earners. However, as this reform did not pose substantial changes to high-income earners, future research has to find other identification channels for this demographic group.

Another issue of the context studied is potentially low salience. The reform studied, the abolition of the SOLI, is a proportional surcharge levied on top of the regular income tax, as outlined in section 3. It is thus likely, that individuals do not fully grasp the financial consequences of this tax and fail to incorporate the tax change fully rationally into their decision making. If this is the case, the coefficients form a lower bound. Moreover and relatedly, the results are specific to the abolition of the SOLI within the German tax system.

To conclude, I suggest that future research incorporate my work into long-run data once it becomes available. Understanding how the effect develops over time is crucial for policy making. For well-structured governance, it could also be interesting to explore the effects on the age of first entry into the labour market, which could change under different tax systems. In times of globalization and digitization lower taxes might also set a lower incentive to work remotely abroad. Such unintended potential side effects are interesting to quantify in future research. Another important outcome to consider is the extensive margin effect: Even if a reduction in MTRs overall appears to decrease labour supply, there might be positive effects incentivising individuals, particularly women, to enter the labour market.

Based on my findings, I suggest care in redesigning the German tax system. My analysis suggests, that lowering tax rates for the broad middle of the income distribution, has negative labour supply effects. This could indicate that tax rate cuts are an inadequate tool to combat the national decrease in average working hours. Further, given the overall insignificant effect among married individuals, but significant heterogeneity by gender within this group, it appears that labour within a couple is shifted to the man following a tax rate decrease. This could partially be attributed to a system of joint taxation, as both partners face the same MTR. Such an internal reallocation can be financially beneficial, if the partner has a higher wage rate. If men are more likely to have a higher wage rate, this dynamic comes at a cost: Women become more financially dependent on men, both in the short-term for household finance, and more concerningly, in the long-term when considering retirement plans. Women might be at a disadvantage in such a system.

References

Abadie, A., Athey, S., Imbens, G. W., & Wooldridge, J. M. (2023). When should you adjust standard errors for clustering? *The Quarterly Journal of Economics*, *138*(1), 1–35.

Arnold, M., Strauss, D., & Fleming, S. (2024, May). Berlin explores tax breaks to get germans working longer hours [Accessed: 2024-06-18]. https://www.ft.com/content/deaba04b-89cf-447f-970a-6d732c751b8f

Asen, E. (2021). Recent changes in top personal income tax rates in europe, 2021 [Accessed: 2024-06-18]. *Tax Foundation*. https://taxfoundation.org/data/all/eu/recent-changes-top-personal-income-tax-rates-europe-2021/

Ashenfelter, O., Doran, K., & Schaller, B. (2010). A shred of credible evidence on the long-run elasticity of labour supply. *Economica*, *77*(308), 637–650.

Attanasio, O., Levell, P., Low, H., & Sánchez-Marcos, V. (2015). *Aggregating elasticities: Intensive and extensive margins of female labour supply* (tech. rep.). National Bureau of Economic Research.

Bingley, P., & Lanot, G. (2002). The incidence of income tax on wages and labour supply. *Journal of Public Economics*, *83*(2), 173–194.

Blomquist, N. S. (1983). The effect of income taxation on the labor supply of married men in sweden. *Journal of Public Economics*, *22*(2), 169–197.

Bonin, H., Buhlmann, F., Siegloch, S., & Stichnoth, H. (2019). Aufkommens-, verteilungs-und arbeitsangebotswirkungen einer stufenweisen abschaffung des solidaritätszuschlags. *Kurzgutachten für das Bundesministerium für Wirtschaft und Energie, Mannheim und Bonn*.

BPB. (2021). Alleinlebende nach geschlecht und familienstand [Available at: https://www.bpb.de/kurz-knapp/zahlen-und-fakten/soziale-situation-in-deutschland/61572/alleinlebende-nach-geschlecht-und-familienstand/].

Calahorrano, L., Rebeggiani, L., Stöwhase, S., & Teuber, M. (2019). Demographic change and income tax revenues–results from a large microsimulation model for germany. *Journal of Economic Policy Reform*, *22*(4), 399–419.

Callaway, B., Goodman-Bacon, A., & Sant'Anna, P. H. (2024). *Difference-in-differences with a continuous treatment* (tech. rep.). National Bureau of Economic Research.

Cutanda, A., & Sanchis-Llopis, J. A. (2023). Labour supply responses to income tax changes in spain. *Hacienda Publica Espanola*, (245), 71–94.

Eissa, N. (1995). Taxation and labor supply of married women: The tax reform act of 1986 as a natural experiment.

Eissa, N. (1996). Labor supply and the economic recovery tax act of 1981. In *Empirical foundations of household taxation* (pp. 5–38). University of Chicago Press.

Enachen, C. (2024). Tax burden on labor in europe, 2024 [Accessed: 2024-06-18]. *Tax Foundation*. https://taxfoundation.org/data/all/eu/tax-burden-labor-europe-2024/

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

GDL. (2024, March). GDL gelingt beispielhafter Tarifabschluss [Press release]. https://www.gdl.de/fileadmin/user_upload/www_gdl_de/pressemitteilungen/2023-03-26_GDL-Pressemitteilung_Einigung.pdf

German Federal Ministry of Finance. (2022). Einkommensteuer-handbuch 2022 [Accessed: 2024-04-10].

Goebel, J., Grabka, M. M., Liebig, S., Kroh, M., Richter, D., Schröder, C., & Schupp, J. (2019). The german socio-economic panel (soep). *Jahrbücher für Nationalökonomie und Statistik*, *239*(2), 345–360.

Goolsbee, A. (1998). It's not about the money: Why natural experiments don't work on the rich.

Gruber, J., & Saez, E. (2002). The elasticity of taxable income: Evidence and implications. *Journal of public Economics*, *84*(1), 1–32.

Heckman, J. J. (2015). Introduction to a theory of the allocation of time by gary becker. *The Economic Journal*, *125*(583), 403–409.

Ichino, A., Olsson, M., Petrongolo, B., & Thoursie, P. S. (2022). *Taxes, childcare, and gender identity norms* (tech. rep.). Working Paper.

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

Meghir, C., & Phillips, D. (2010). Labour supply and taxes. *Dimensions of tax design: The Mirrlees review*, 202–74.

Mengden, A. (2023). Comparing income tax systems in europe, 2023 [Accessed: 2024-06-18]. *Tax Foundation*. https://taxfoundation.org/data/all/eu/comparing-income-tax-systems-europe-2023/

Müllbacher, S., & Nagl, W. (2017). Labour supply in austria: An assessment of recent developments and the effects of a tax reform. *Empirica*, *44*, 465–486.

OECD. (2017). The pursuit of gender equality: An uphill battle [Accessed: 2023-04-10]. https://www.oecd.org/germany/Gender2017-DEU-en.pdf

OECD. (2023). *Oecd labour force statistics 2022*. https://doi.org/https://doi.org/https://doi.org/10.1787/dc0c92f0-en

OECD. (2024). Hours worked (indicator) [Accessed: 2024-06-18]. https://doi.org/10.1787/47be1c78-en

Pollak, R. A. (2003). Gary becker's contributions to family and household economics. *Review of Economics of the Household*, *1*, 111–141.

Pollak, R. A., & Wachter, M. L. (1975). The relevance of the household production function and its implications for the allocation of time. *Journal of Political Economy*, *83*(2), 255–278.

Reuters. (2024, March). Uk's hunt announces new cuts to national insurance [Accessed: 2024-06-18].

Tondani, D. (2006). *Estimating the effects of personal income tax on labour supply in italy* (tech. rep.). Department of Economics, Parma University (Italy).

A Appendix

Table A.1: German Income Tax Rates without SOLI (§ 32a EStG)

Panel (A): 2020 **Taxable Income**	**Income Tax (in Euro)**
$\leq 9,408$ Euro	0
$9,409$ Euro to $14,532$ Euro	$(972.87 \times (z - 9,408) + 1,400) \times (z - 9,408)$
$14,533$ Euro to $57,051$ Euro	$(212.02 \times (z - 14,532) + 2,397) \times (z - 14,532) + 972.79$
$57,052$ Euro to $270,500$ Euro	$0.42 \times z - 8,963.74$
$> 270,501$ Euro	$0.45 \times z - 17,078.74$

Panel (B): 2021 **Taxable Income**	**Income Tax (in Euro)**
$\leq 9,744$ Euro	0
$9,745$ Euro to $14,753$ Euro	$(995.21 \times (z - 9,744) + 1,400) \times (z - 9,744)$
$14,754$ Euro to $57,918$ Euro	$(208.85 \times (z - 14,753) + 2,397) \times (z - 14,753) + 950.96$
$57,919$ Euro to $274,612$ Euro	$0.42 \times z - 9,136.63$
$> 270,501$ Euro	$0.45 \times z - 17,364.99$

Notes: z = Taxable income rounded down to the nearest Euro. The income tax is to be rounded down to the nearest Euro.

Table A.2: Event Study Interaction Coefficients with Labour Supply as Dependent.

	(1)	**(2)**
Controls	No	Yes
Year FE	No	No
1984 $\times D_i$	-547.12***	-826.32***
	(99.87)	(95.08)
1985 $\times D_i$	-387.43***	-642.31***
	(91.53)	(88.10)
1986 $\times D_i$	-457.62***	-752.57***
	(89.16)	(84.60)
1987 $\times D_i$	-460.99***	-745.08***
	(84.17)	(77.23)
1988 $\times D_i$	-464.65***	-755.97***
	(80.46)	(75.96)
1989 $\times D_i$	-423.94***	-690.45***
	(86.09)	(79.31)
1990 $\times D_i$	-451.91***	-673.35***
	(83.63)	(77.04)
1991 $\times D_i$	-524.31***	-747.80***
	(77.45)	(71.44)
1992 $\times D_i$	-561.78***	-747.87***
	(58.20)	(55.24)
1993 $\times D_i$	-539.40***	-731.21***
	(59.20)	(56.09)
1994 $\times D_i$	-495.29***	-675.74***
	(57.61)	(54.33)
1995 $\times D_i$	-505.79***	-664.90***
	(55.96)	(51.40)
1996 $\times D_i$	-472.06***	-636.19***
	(55.73)	(52.21)
1997 $\times D_i$	-456.92***	-610.19***
	(56.92)	(53.20)
1998 $\times D_i$	-496.96***	-629.48***
	(54.96)	(51.13)
1999 $\times D_i$	-500.39***	-652.20***

Continued on next page

	(1)	**(2)**
Controls	No	Yes
Year FE	Yes	Yes
	(54.20)	(50.42)
2000 $\times D_i$	-393.83***	-542.89***
	(43.71)	(41.09)
2001 $\times D_i$	-415.24***	-551.23***
	(42.52)	(39.69)
2002 $\times D_i$	-437.10***	-574.33***
	(39.46)	(36.46)
2003 $\times D_i$	-416.40***	-552.72***
	(39.32)	(36.23)
2004 $\times D_i$	-382.14***	-522.57***
	(39.28)	(36.29)
2005 $\times D_i$	-375.68***	-502.91***
	(39.91)	(36.83)
2006 $\times D_i$	-371.59***	-488.31***
	(38.67)	(35.71)
2007 $\times D_i$	-311.31***	-414.86***
	(38.76)	(35.79)
2008 $\times D_i$	-285.42***	-403.40***
	(37.98)	(34.98)
2009 $\times D_i$	-318.62***	-419.82***
	(37.91)	(34.52)
2010 $\times D_i$	-417.28***	-456.64***
	(32.64)	(29.82)
2011 $\times D_i$	-351.69***	-397.92***
	(29.58)	(27.09)
2012 $\times D_i$	-292.08***	-344.88***
	(27.83)	(25.48)
2013 $\times D_i$	-332.97***	-368.13***
	(27.35)	(25.50)
2014 $\times D_i$	-255.04***	-272.35***
	(26.06)	(23.68)
2015 $\times D_i$	-239.86***	-255.23***
	(25.57)	(23.40)

Continued on next page

	(1)	**(2)**
Controls	No	Yes
Year FE	Yes	Yes
2016 $\times D_i$	-175.18***	-205.78***
	(24.80)	(22.23)
2017 $\times D_i$	-159.91***	-179.97***
	(21.94)	(19.85)
2018 $\times D_i$	-91.76***	-113.04***
	(20.17)	(17.98)
2019 $\times D_i$	-52.82***	-73.98***
	(17.67)	(15.64)
2021 $\times D_i$	-145.57***	-159.38***
	(20.00)	(17.90)

Notes: Year-specific coefficients of interest and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from 2, indicating a change in MTRs. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table A.3: Static Difference-in-Differences of Labour Supply with Treatment as Exclusively Negative MTR changes

	Panel (A): All Individuals		
	(1)	**(2)**	**(3)**
	2 years	5 years	12 years
$\text{post}_t \times D_i$	-156.94***	-47.82***	65.93***
	(17.66)	(16.51)	(17.81)
Observations	25,491	56,896	105,363
R-squared	0.44	0.44	0.36
Mean (Untreated)	947.77	1024.78	1157.10
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes
	Panel (B): Unmarried Individuals		
	(1)	**(2)**	**(3)**
	2 years	5 years	12 years
$\text{post}_t \times D_i$	-197.51***	-69.05***	53.40**
	(22.41)	(20.98)	(22.31)
Observations	13,349	27,798	45,932
R-squared	0.59	0.60	0.52
Mean (Untreated)	839.43	911.80	1067.90
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes
	Panel (C): Married Individuals		
	(1)	**(2)**	**(3)**
	2 years	5 years	12 years
$\text{post}_t \times D_i$	-83.37***	-16.57	80.52***
	(27.89)	(25.54)	(26.96)
Observations	12,142	29,098	59,431
R-squared	0.29	0.32	0.28
Mean (Untreated)	1099.86	1149.62	1224.19
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes

Note: Coefficients of interest and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The treatment variable is binary and follows from equation (2), indicating a decrease in MTRs. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table A.4: Continuous Difference-in-Differences Regression Results of Labour Supply excluding Outliers

	Panel (A): All Individuals		
	(1) 2 years	**(2)** 5 years	**(3)** 12 years
$\text{post}_t \times \Delta\widetilde{\tau}_i$	61.79***	24.64***	-26.93***
	(7.84)	(7.25)	(7.91)
Observations	25,461	56,845	105,377
R-squared	0.43	0.44	0.36
Mean (Untreated)	947.77	1024.78	1157.10
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes
	Panel (B): Unmarried Individuals		
	(1) 2 years	**(2)** 5 years	**(3)** 12 years
$\text{post}_t \times \Delta\widetilde{\tau}_i$	85.50***	44.82***	-8.20***
	(10.34)	(9.37)	(10.08)
Observations	13,295	27,702	45,868
R-squared	0.58	0.59	0.52
Mean (Untreated)	839.43	911.80	1067.90
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes
	Panel (C): Married Individuals		
	(1) 2 years	**(2)** 5 years	**(3)** 12 years
$\text{post}_t \times \Delta\widetilde{\tau}_i$	29.13**	11.70	-46.40***
	(12.03)	(11.61)	(11.91)
Observations	12,166	29,143	59,509
R-squared	0.29	0.22	0.28
Mean (Untreated)	1099.86	1149.62	1224.19
Controls	Yes	Yes	Yes
Year Fixed Effects	No	Yes	Yes

Note: Coefficients of interest and standard errors (in parentheses) are presented. The time span varies by model. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -3.17 and 2.69 (as outliers were excluded for this analysis). Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table A.5: Heterogeneity by Gender (Female): Continuous Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times F_i$	-18.30	-26.04**
	(11.70)	(12.02)
Observations	25,631	57,203
R-squared	0.42	0.42
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times F_i$	-11.25	-14.53
	(15.11)	(14.80)
Observations	13,410	27,930
R-squared	0.56	0.57
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times F_i$	-23.41	-44.05**
	(17.86)	(19.51)
Observations	12,221	29,273
R-squared	0.28	0.31
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -14.65 and 2.69. Control variables are indicated as specified and include age (continuous), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table A.6: Heterogeneity by Number of Children: Continuous Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1)	**(2)**
	2 years	5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times C_i$	1.03	2.28
	(11.72)	(12.21)
Observations	25,631	57,203
R-squared	0.41	0.42
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1)	**(2)**
	2 years	5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times C_i$	2.53	7.93
	(15.27)	(15.36)
Observations	13,410	27,930
R-squared	0.56	0.57
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1)	**(2)**
	2 years	5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times C_i$	8.35	2.90
	(18.09)	(18.63)
Observations	12,221	29,273
R-squared	0.27	0.30
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -14.65 and 2.69. C_i is a dummy indicating households with children, displaying value 0 for households without children. Control variables are indicated as specified and include age (continuous), gender (categorical), marital status (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.

Table A.7: Heterogeneity by Age: Continuous Triple Differences (DDD) Regression Results of Labour Supply

	Panel (A): All Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times O_i$	5.48	15.37
	(11.36)	(11.39)
Observations	25,631	57,203
R-squared	0.43	0.43
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (B): Unmarried Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times O_i$	2.07	22.80*
	(14.16)	(13.63)
Observations	13,410	27,930
R-squared	0.56	0.57
Controls	Yes	Yes
Year Fixed Effects	No	Yes
	Panel (C): Married Individuals	
	(1) 2 years	**(2)** 5 years
$\text{post}_t \times \Delta\widetilde{\tau}_i \times O_i$	5.20	-1.97
	(18.55)	(21.29)
Observations	12,221	29,273
R-squared	0.31	0.33
Controls	Yes	Yes
Year Fixed Effects	No	Yes

Note: Triple difference coefficients and standard errors (in parentheses) are presented. The outcome variable labour supply is the same across all models and measured in annual working hours. The change in marginal tax rate was reformulated to avoid simultaneity bias and follows from equation (5): It takes continuous values between -14.65 and 2.69. O_i is a dummy indicating old individuals, displaying value 1 for individuals 40 and older and 0 for individuals below the age of 40. Control variables are indicated as specified and include gender (categorical), marital status (categorical), number of children (categorical) and 2019 income (continuous). Significance levels are indicated as follows: * p<0.1, ** p<0.05, and *** p<0.01.