Men and Women at Work

Gains from Individual Taxation and the Macroeconomic Importance of Workforce Diversity

November 1, 2022

Jonas Nauerz[a]

Abstract

Joint income taxation of married couples imposes a high marginal tax rate for secondary earners. Moving to individual taxation encourages female labor supply and therefore alters workforce composition. This paper develops a general equilibrium model with heterogeneous agents and family labor supply to characterize the composition channel's quantitative importance. Calibrating the model to Germany, I quantify the consequences of abolishing joint taxation. To incorporate workforce diversity, I allow for imperfect substitutability of male and female labor inputs in output production, disciplined by my estimate of the elasticity of 0.7. The complementarities between men and women substantially amplify the positive output effects of the tax reform. Labor force participation of married women rises by 12.6%, and overall output increases by 6.8%. Assuming perfect substitution, output rises by half as much because there is no demand response for male labor, and men participate less.

Keywords: Taxation, Labor Force Participation, Diversity, Elasticity of Substitution

JEL Classification: E62, H31, J16, J21, J31

[a]University of Notre Dame, Department of Economics. Email: jnauerz@nd.edu. I am tremendously grateful to my advisor, Benjamin Pugsley, for invaluable guidance, mentorship, and advice. I would also like to thank Rüdiger Bachmann, Michael Pries, Nezih Guner, Bence Bardóczy, Georgi Kocharkov, Katharina Wrohlich, and seminar participants at the University of Notre Dame and the 2022 NBER Heterogeneous-Agent Macro Workshop for valuable comments.

1 Introduction

Many countries, including Germany, have made increasing female labor supply an explicit policy goal, and taxation is a widely discussed instrument among economists and policy-makers (Bick & Fuchs-Schündeln 2017). Secondary earners' marginal tax rate, the element of joint taxation of married couples that constitutes a disincentive for many married women, is at the center of the debates (Guner et al. 2012, Dao et al. 2019, Bach et al. 2020, Borella et al. 2021). The German government has made the issue especially relevant by showing willingness to reform its current system in their recent coalition agreement.

I contribute to the debate by highlighting that the effects of abolishing joint taxation on labor force participation and GDP are twofold. One channel is that raising female labor inputs increases output by adding more workers to the labor force. Second, the reform also changes the labor force's composition, allowing for added effects from gender diversity.

These workforce composition changes may have further effects if women complement male employment because more female labor supply would then feed back into male labor demand and wages. To test for such complementarities, I exploit a natural experiment in Germany, the introduction of a legal claim for public childcare for children under three in 2005. In regions where more childcare became available, women worked more, relative to regions with less childcare. Using this variation, I estimate the effect of greater female participation on female and male wages and find evidence of complementarities. I reject the hypothesis that men and women are perfect substitutes in output production. This result implies that production functions assuming perfect substitution, common in the macroeconomic tax reform literature, abstract from an important aspect of the data.

Motivated by my empirical findings, I develop a heterogeneous-agent model of family labor supply. I include gender diversity by allowing for imperfect male/female labor input substitution and demonstrate the quantitative importance of relaxing the assumption. I examine the responses to a reform abolishing joint taxation of married couples and find that the tax reform substantially affects household decision making. Married women's labor force partic-

ipation increases by approximately 12.6%, and married men increase participation by 5.5%. In response to the labor supply improvement, output rises by about 6.8%.

Without gender diversity, in the perfect substitutability case, output increases by 3.3%. Thus, the output gain is more than twice as large under complementarity than it is when assuming perfect substitution. Under perfect substitution, the increase in female participation crowds out male participation. In contrast, when women and men complement one another, the male labor demand shifts as female labor increases. As a result, male wages improve, and participation rises for both genders. To further examine the model's predictions, I consider various levels of complementarity estimated by others in the empirical literature and find the output response decreases with the level of substitution.

The model, calibrated for Germany under joint taxation, effectively matches the distribution of labor force participation, gender wage gap, and wealth dispersion observed in the data. The calibrated model also implies labor supply elasticity consistent with values estimated by others. Finally, the incomplete market model components help generate a reasonable joint distribution of income, wealth, and marginal propensity to consume.

Using German administrative data, I estimate the elasticity of substitution between male and female labor inputs at 0.7. My approach exploits time and county-level (i.e., Landkreise) variation in the availability of public daycare for children under three to construct an instrument for relative labor supply. The introduction of the 2005 German daycare expansion act known as Tagesbetreuungsausbaugesetz introduced variation in childcare resulting from delays in the approval process due to county-level authorities' noncompliance with federal and state regulations and property shortages inhibiting childcare facility construction. Thus, it is arguably exogenous to the labor supply decisions of families (Bauernschuster et al. 2015, Müller & Wrohlich 2020). My estimate compares to previous estimates based on exogenous gender-specific variation in labor supply in the U.S. of the 1950s of about 3 (Acemoglu et al. 2004) and a more recent number from an Italian context of around 1 (De Giorgi et al. 2013).

In the model, the government implements a progressive labor income tax schedule following

Heathcote et al. (2017). As in the German tax code, a married couple's tax burden in the model is determined by doubling the taxes on half the household's income. Given the progressive tax schedule, the marginal tax rate of the primary earner is lower under joint compared to individual taxation, while the secondary earner's rate is higher.

Abolishing joint taxation changes the German government's budget in two ways. First, labor force participation increases, causing the number of taxpayers to increase and unemployment recipients to decrease. Additionally, the reform eliminates the tax benefits for households with unequal incomes. Thus, the reform can potentially expand fiscal space; Bach et al. (2020) show that abolishing joint taxation could yield 10 billion EUR, or roughly 4.5% of total labor income tax revenue. Furthermore, I find that the tax reform increases the welfare of married households when allowing for workforce diversity effects. However, when female employment crowds out male labor, welfare increases only if the government redistributes the additional tax revenues.

My model builds on work by Guner et al. (2012), Bick & Fuchs-Schündeln (2017), Borella et al. (2021), Alon et al. (2020), Mankart & Oikonomou (2017), and Bardóczy (2020). Childcare is a central theme in the literature because of the role it plays in families' labor supply decision making. In this study's framework, families choose their labor supply subject to costs associated with work, including reduced time for home production and leisure, scheduling joint work time, and finding childcare, among other things. With the tax reform initiative's revenue yield, the German government could relieve families through increased child benefits, child allowances, or improved childcare infrastructure.

Related Literature This paper builds on three literature strands. First, it further studies the welfare implications of taxes and transfers within heterogeneous agent models. Guner et al. (2012), De Nardi et al. (2020), and Heathcote et al. (2020) focus on taxes and transfers differing by age or gender. Bick & Fuchs-Schündeln (2017, 2018) consider joint taxation of married couples. Dao et al. (2019) and Bach et al. (2020) discuss alternative joint taxation reform approaches in Germany and their fiscal ramifications. Borella et al. (2021) and

Groneck & Wallenius (2020) study how social security benefits and old age provisions may inhibit the female labor supply. This study contributes to the literature by allowing men and women to be imperfect substitutes in production and highlighting the importance of heterogeneity and gender diversity for policy analyses.

Second, this paper contributes to empirical literature documenting complementarities between men and women in the production process. Some existing evidence suggests that men and women are not perfect substitutes in production. Acemoglu et al. (2004) and De Giorgi et al. (2013) use a plausibly exogenous variation induced by a natural experiment in the U.S. and Italy, respectively and find men and women are imperfect substitutes with an estimated substitution elasticity between 1 and 4. Ghosh (2018) uses Bartik-style instruments exploiting regional variations in industry-level changes in U.S. employment and obtains long-run substitution elasticity estimates of approximately 1.7. Espinoza et al. (2019) use aggregate and sector-level cross-country and firm-level data to examine China's manufacturing sector. Their estimate is smaller than 1 in the aggregate sample, between 1 and 2 in the sector-level sample, and between 2 and 3 in the firm-level data. This paper offers a recent estimate of substitution elasticity via quasi-random variation in a policy experiment.

Why would men and women imperfectly complement one another in the production process? The microeconomic literature finds considerable evidence that men and women bring different skills, perspectives, and preferences to the workplace, and gender composition affects teams' decision-making process and performance (Azmat & Petrongolo 2014). Further evidence suggests that gender inclusiveness can increase innovation, firm value, minority productivity, and financial stability (Terjesen et al. 2009, Dezsö & Ross 2012, Hunt et al. 2018). Another explanation suggests that while men and women might be close substitutes given education level and occupation/sector, they typically work in different occupations and sectors. If substitution between sectors is relatively low, averaging substitution elasticity across sectors results in a low aggregate value (Miyagiwa & Papageorgiou 2007).

Finally, my paper is related to the literature studying the introduction of the legal claim

to a place in public childcare in Germany. For example, Müller & Wrohlich (2020) use the variation induced by the Tagesbetreuungsausbaugesetz reform and find that it increased mothers' employment. Bauernschuster et al. (2015) study fertility choices, and Barschkett (2022) provides evidence for the substitution of illness spells of communicable diseases from elementary school to the first years of daycare.

The remainder of this paper is organized as follows: The next section discusses joint taxation in Germany and the disincentive it constitutes for labor supply. Section 3 introduces the model and section 4 discusses parameterization, model fit, and discusses crucial mechanisms that are important for the central question. Section 5 presents the results, investigates the effects of abolishing joint taxation, and the quantitative importance of varying the elasticity of substitution. The last section concludes.

2 Joint Taxation in Germany

A reform of joint tax assessment of married couples has been discussed in Germany for a long time; for a summary, see Spangenberg (2016). Criticism is focused on two aspects. First, single-earner couples with higher incomes are favored, and secondly, the status quo implies negative work incentives for the (primarily female) second earners. In the case of individual taxation, the secondary earners are taxed less. However, abolishing the joint tax assessment would place a greater burden on couples with sole earners (Bach et al. 2020).

Currently, married couples and civil partners can choose between individual tax assessment and joint tax assessment (*Ehegattensplitting*)[1]. However, since there is a marriage bonus and never a marriage penalty from filing jointly (in contrast to the U.S.), filing separately is not financially beneficial.

If a couple opts for the joint tax assessment, the tax is determined by taking half of the joint household income, determining the tax for that amount, and then doubling it. This results

[1] I will treat civil partnerships as married couples.

in lower tax liability for couples than under an individual tax assessment.[2] The larger the income differences between the partners and the higher the joint taxable income, the greater the advantage of joint taxation.

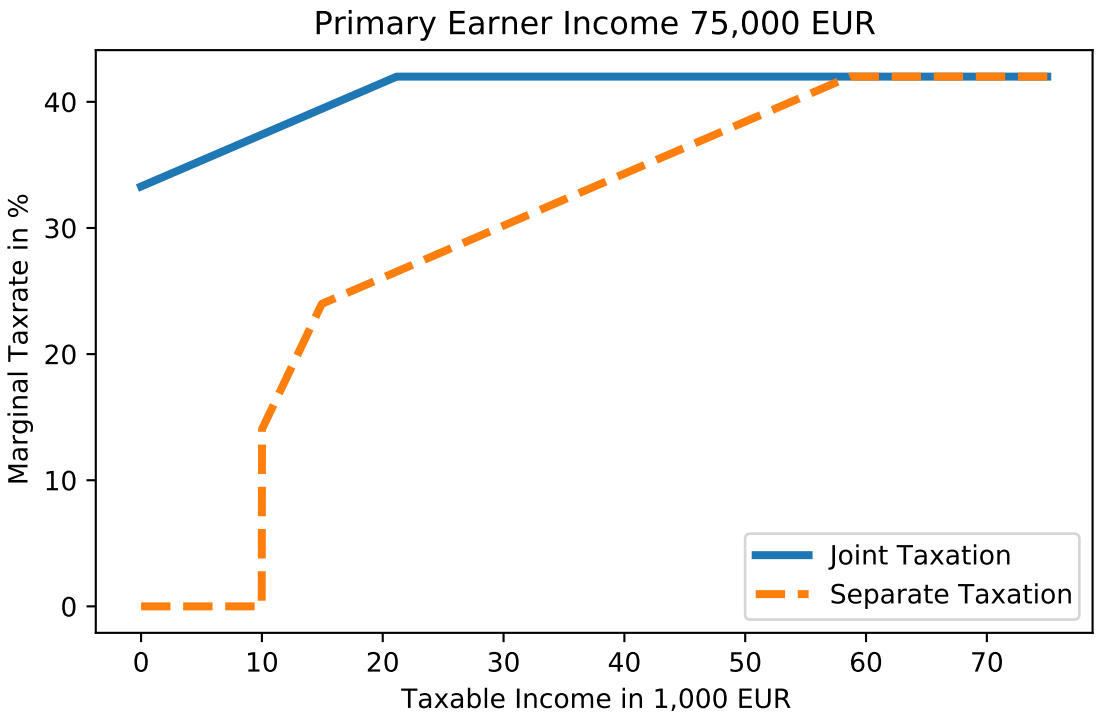

Figure 1: Marginal Tax Rate Secondary Earner

The economic implication of the differential tax treatment of jointly vs. separately filing households goes beyond the marriage bonus because it distorts the labor supply decisions of a (potential) secondary earner considering whether or not to enter the labor force. The progressivity of the German tax code implies a higher marginal tax rate on the secondary earner's income than under separate taxation. In comparison, the primary earner faces a lower marginal tax rate.

To illustrate this point, Figure 1 plots the marginal statutory tax rate of a secondary earner under both joint and separate taxation for a given income level for the primary earner. Under the current law, both partners are subject to the marginal tax rate of half their joint taxable

[2] A couple with exactly identical labor incomes is indifferent between joint and individual taxation.

income. Thus, the figure shows that the second earning partner in this example starts with a burden of 33.29% on the first Euro earned, which increases further with higher incomes. If, however, labor incomes were taxed individually, the first Euro would not be taxed. Hence, the net income from the second income is lower under joint relative to individual taxation and, with it, the motivation to take up or expand gainful employment.

In the case of individual taxation, both partners pay tax based on their own income. Accordingly, couples with single earners are taxed more heavily under individual taxation, and abolishing joint taxation changes the incentives to supply labor for the primary and secondary earners in opposite directions.

3 Model

The paper aims to examine the effects of eliminating joint taxation. To this end, I develop a heterogeneous agent model with family labor supply that builds the spousal insurance literature (Krusell et al. 2017, Mankart & Oikonomou 2017, Bardóczy 2020). In the model, individuals are subject to idiosyncratic income shocks. Men and women decide between employment and non-employment, and couples are taxed jointly. Because filing separately in Germany is not financially beneficial for married couples, I will disregard the option to choose individual taxation under the current tax system. Instead, I focus on moving from *only* joint to *only* individual taxation, which I refer to as abolishing joint taxation. The production structure allows for male and female labor inputs to be imperfect substitutes. Furthermore, the model retains the ingredients of the standard Aiyagari-type incomplete market model.

3.1 Environment and Timing

The economy consists of a unit mass of households, a representative firm, and a public sector. Households save in assets with an exogenous rate of return r, the firm uses labor to produce a final good, and the government collects labor income tax and pays unemployment benefits.

Time is discrete, and one period corresponds to a year.

I abstract from unmarried households who are unaffected by the tax reform. Thus, the model has only married households composed of females and males, indexed by $g \in \{f, m\}$. A household is characterized by a pair of labor productivities $z = (z_f, z_m)$ and shared assets $a \geq \underline{a}$.

Men and women choose between non-employment and employment, i.e., $e_f \in \{0, 1\}$ and $e_m \in \{0, 1\}$. Working is costly, and this choice enters the flow utility to avoid employment becoming a preferable state. The government pays unemployment benefits in case of non-employment. Markets are incomplete since households cannot borrow.

The within-period timing is as follows:

1. Households start with assets inherited from last period a and draw a new productivity from distribution $\pi_g(z'|z)$.
2. Individuals decide whether to participate in the labor market.
3. Households consume and save.

3.2 Heterogeneity and Uncertainty

Individuals face uncertainty in the form of idiosyncratic productivity shocks. Since markets are incomplete, households accumulate assets to partially self-insure. The individual endowment of efficiency units is denoted by z_{gt}. I allow that men and women face different stochastic processes for their idiosyncratic productivity.

Individual productivity follows an AR(1) process in logs

$$\ln(z_{g,t+1}) = \rho_{z,g} \ln(z_{g,t}) + \epsilon_{z,g,t}, \tag{3.1}$$

where $\epsilon_{z,g,t} \sim N(0, \sigma^2_{z,g})$. Since labor supply decisions are formulated at an extensive margin, if idiosyncratic productivity falls, some individuals may decide to become non-employed.

3.3 Households

Married households pool their incomes and assets to maximize joint utility. The Bellman equation can be written as

$$V(z,a) = \max_{c,a',e_f,e_m} \left\{ \frac{c^{1-\gamma}}{1-\gamma} - e_f\varphi_f - e_m\varphi_m + \boldsymbol{\xi}(e_g) + \beta\mathbb{E}\left[V(z',a',\boldsymbol{\xi}'|z',a')\right] \right\} \tag{3.2}$$

$$\text{s.t. } c + a' = y_f + y_m - T(y_f + y_m) + (1+r)a$$

$$e_f \in \{0,1\} \quad e_m \in \{0,1\} \quad a' \geq \underline{a}$$

where $T(y_f + y_m)$ are taxes on joint income. Household labor income depends on the gender-specific wages w_g (determined in general equilibrium), idiosyncratic productivity z_g, and the replacement rate b. Specifically, income of an individual with gender g is given by

$$y_g(e_g) = \begin{cases} w_g\, z_g, & \text{if } e_g = 1 \\ b\, z_g, & \text{if } e_g = 0 \end{cases}$$

All households share the same discount factor β and have the following preferences $u(c,s) = \frac{c^{1-\gamma}}{1-\gamma} - e_f\varphi_f - e_m\varphi_m$, where φ_g is the dis-utility from working.[3] The utility loss due to work can be interpreted as originating from, e.g., home production and leisure activities.

I assume that each participation decision is subject to idiosyncratic Type-I extreme value taste shocks $\boldsymbol{\xi}(e_g)$ with scale $\sigma_\xi > 0$, a standard way of smoothing discrete choice problems since Rust (1987). This assumption is convenient because it implies that the solution of the participation decision is a set of choice probabilities that have logit form and that taste shocks can be integrated out using the logsum formula (Caliendo et al. 2019).

[3] This formulation is consistent with balanced growth and wildly used in the literature on family economics. An alternative specification would allow the marginal utility of consumption to be affected by the labor force participation decision, which helps to match life cycle facts such as the hump-shaped pattern of labor supply and consumption, see, e.g., Blundell et al. (2016).

3.4 Firms

Competitive firms conduct production by using labor rented on competitive factor markets. In the production process, male and female efficiency units are imperfect substitutes. Output is given by

$$Y = \left[(B_f L_f)^{\frac{\theta-1}{\theta}} + L_m^{\frac{\theta-1}{\theta}} \right]^{\frac{\theta}{\theta-1}}$$

where the total labor input is derived from the efficiency units supplied by male and female workers is combined according to a CES aggregator. The elasticity of substitution between female and male labor inputs is equal to θ, and B_f is a factor-augmenting productivity term that influences the gender wage gap in the model.

This production structure and competitive factor markets imply that the optimal ratio of labor inputs is

$$\frac{L_m^d}{L_f^d} = \left(B_f^{\frac{1-\theta}{\theta}} \frac{w_f}{w_m} \right)^{\theta} \tag{3.3}$$

3.5 Government

Following Heathcote et al. (2017), I assume a labor income tax schedule such that the average tax rate for income level y is given by $t(y) = 1 - \eta y^{-\tau}$. The parameters η and τ govern the level and progressivity. If $\tau = 0$, average tax rates are constant as income changes, whereas $\tau > 0$ implies a progressive tax.

Thus, under joint taxation, the net income of a couple is given by $y^{net} = 2\eta \left(\frac{y_f + y_m}{2} \right)^{1-\tau}$ and tax liabilities are

$$T(y_f + y_m) = y_f + y_m - 2\eta \frac{y_f + y_m}{2}^{1-\tau}$$

I follow Heathcote et al. (2017) and denote with ω the government consumption as a fraction of aggregate output, i.e., $G = \omega Y$. I abstract from public debt, and thus the government

budget constraint holds period by period and reads as

$$b\left(\int_{e_f=0} z_{i,f}\,di + \int_{e_m=0} z_{i,m}\,di\right) + \omega \int y_i\,di = \int T(y_j)\,dj$$

where individuals are indexed by i and couples are indexed by j.

3.6 Market Clearing and Equilibirum

Labor supply by gender g is given by $L_g^s = \int_{e_g=1} z_{i,g} di$ and thus labor market clearing is given by

$$L_g^d = \int_{e_g=1} z_{i,g}\,di$$

The goods market clears when output is spent on private consumption and public consumption

$$Y = \int c_i\,di + G$$

3.6.1 Equilibrium

For a given government consumption level, unemployment benefits, and tax function, a stationary equilibrium consists of wages w_g, labor used in the production L_g^d, household decision rules, and a distribution of households over their asset and labor productivity levels such that households and firms optimize, taking wages as given, so that the government budget is balanced, the distribution is consistent with individual decisions, and markets clear.

3.7 Computation

Introducing discrete labor force participation choices in the household problem complicates the solution of the model since they can potentially lead to kinks in the value functions and discontinuities in the optimal policy rules. In order to solve the household problem

efficiently with an endogenous gridpoint method (EGM), I rely on the approach used in Bardóczy (2020).

Small taste shocks help to smooth out discontinuities in the policy functions. However, they do not fully convexify the problem, which implies that the household's first-order conditions are necessary but not sufficient (Iskhakov et al. 2017). Hence, the original EGM introduced by Carroll (2006) needs to be extended with an upper envelope step (Fella 2014, Druedahl 2021).[4]

4 Economy with Joint Taxation

This section discusses the values I have assigned to the model's parameters. I also comment on crucial mechanisms of the model that are important for the central question in this study and report on the properties of the benchmark economy while comparing them to corresponding empirical values.

4.1 Calibrating the Economy

There are 14 parameters, 10 of which I calibrate externally. The remaining parameters, i.e., the dis-utility terms for labor for men and women, the replacement rate, and the factor augmenting productivity term, I calibrate internally. I calibrate the model to the German economy, Table 1 summarizes the parameter choices.

4.1.1 Externally Calibrated Parameters

Using German administrative data, I estimate the elasticity of substitution between male and female labor inputs by exploiting a natural experiment. I estimate the parameters that govern the tax function and the income processes of men and women from German survey

[4]The core upper envelope step considers every segment of the endogenous grid and finds all the exogenous gridpoints that fall into that segment. Interpolation then yields a candidate solution. However, since the endogenous grid is non-monotonic, the same point may be bracketed by another segment, and thus only the solution which yields the highest value is kept.

data and use values for the degree of relative risk aversion, the discount factor, and the borrowing constraint common in the literature.

Elasticity of Substitution I estimate the elasticity of substitution between male and female labor inputs, θ, following the approach in Acemoglu et al. (2004). Starting with the firm's first-order conditions and taking logs on both sides of Equation (3.3), we can express relative wages as a function of relative labor inputs

$$ln\left(\frac{w_f}{w_m}\right) = \frac{\theta - 1}{\theta} ln\left(B_f\right) - \frac{1}{\theta} ln\left(\frac{L_f}{L_m}\right). \tag{4.1}$$

Equation (4.1) can be directly estimated using data on average male and female wages and average male and female labor supply.

Since relative labor supply is likely endogenous to fluctuations in relative wages, I exploit a natural experiment – the Daycare Expansion Act in Germany in 2005 – to construct an instrument using spatial and temporal differences in childcare coverage in a sample of West German counties from 1994 to 2018. The variation in childcare coverage induced by the reform is arguably exogenous to relative wages, as it resulted from an extensive and complex funding process involving various levels of government, a lack of appropriate building grounds, and an approval procedure that caused holdups due to non-compliance with federal and state regulation (Bauernschuster et al. 2015, Müller & Wrohlich 2020).

My estimated coefficients reject perfect substitutability, i.e., $\theta \to \infty$. In my baseline specification, I find a negative and statistically significant effect of the relative labor supply ratio on relative wages, implying an elasticity of substitution of 0.7. Thus, my empirical results suggest that men and women in modern-day Germany may be gross complements in production. For further details on the background of the natural experiment, the identification strategy, data, and empirical results, see Appendix A.

My estimate is smaller than the one from Acemoglu et al. (2004), who find an elasticity of substitution between 2 and 3. However, their context (the U.S. during and right after World

War II) is very different from the one I study. De Giorgi et al. (2013), use the same method with Italian data from the early 2000s and find values close to 1.

A potential explanation for why I find relatively small values for θ is that, given educational level and occupation/sector, men and women are close substitutes. However, they often work in different occupations and sectors. Since the substitutability between any two factors at the aggregate level is a weighted average of the elasticity of substitution across sectors and the two factors within a sector (Miyagiwa & Papageorgiou 2007), if substitution between sectors is relatively low, this could explain why we observe low estimates in the aggregate production function.

Tax Function In order to choose the parameters of the tax function, I follow García-Miralles et al. (2019) and estimate an effective average tax function, taking into account the fact that a significant number of German taxpayers face a zero tax rate. I choose $\tau = 0.188$ and $\eta = 0.751$, based on data from the EU Statistics on Income and Living Conditions. Figure 2 plots the estimated tax schedule used in the model and a binned scatter plot of the average taxes effectively paid by households in Germany. My estimate for the effective tax-free threshold is equal to 14,819 Euro or 22% of the mean household income for married working-age couples in 2019 SOEP data. Thus, in the model, labor income taxes are calculated only on the part that exceeds this threshold of mean labor income. As a robustness exercise, I also consider values for τ and η based on the statutory tax formula. The central results and conclusions of the paper are unaffected by this parameter choice. For more details, see Appendix B.

Other Parameters To quantify idiosyncratic labor income risks, I estimate the AR(1) process for idiosyncratic productivity separately for men and women, using data from the German SOEP. To address concerns of a selection bias for women, I only use data after the year 2000. Following the methodology in Chang & Kim (2006), I find that women's income process is more volatile and less persistent; see Table 1. Finally, I calibrate the relative risk aversion, discount factor, and borrowing constraint to common values in the literature.

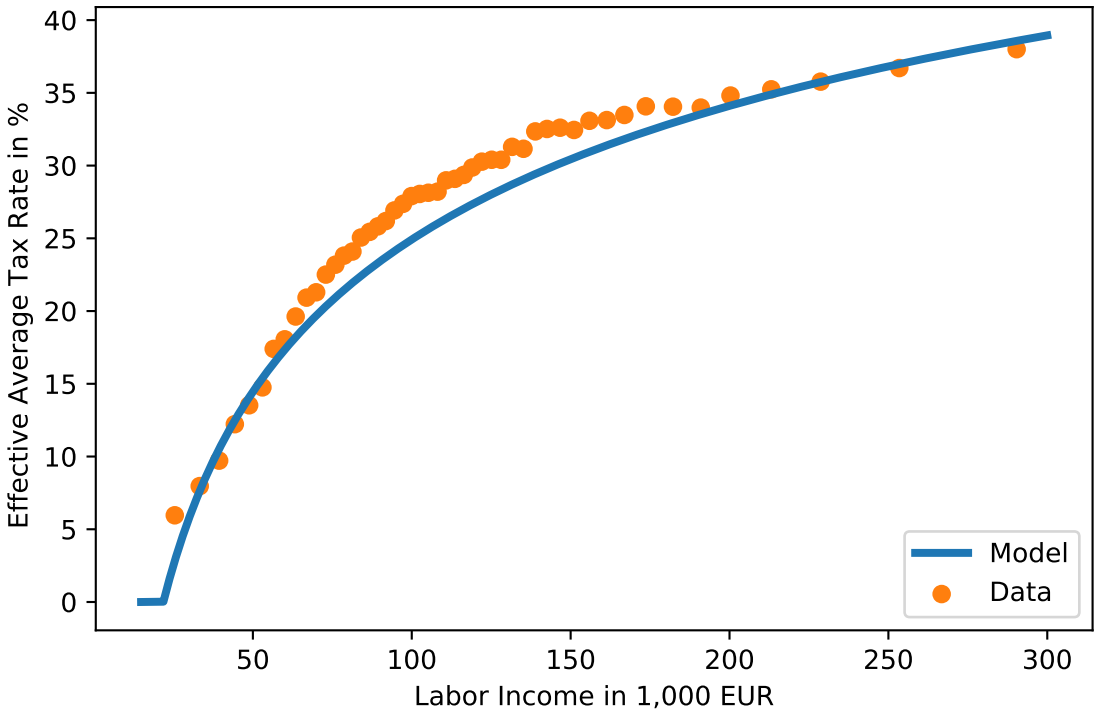

Figure 2: Tax Rates in Germany

4.1.2 Internally Calibrated Parameters

The labor force participation rate (LFPR) of married working-age men in the 2019 SOEP is 91.0%. On the other hand, the LFPR of married women is 60.7%, a difference of 30.3 percentage points or roughly 1/3. About 56% of married prime-aged households in SOEP are dual-income households, while in approximately 35%, only the husband works. I calibrate the replacement rate, b, and the disutility of labor, φ_g, such that the model matches the aggregate labor force participation rates of men and women and the fraction of households that are not participating at all in the data. Finally, I calibrate the factor-augmenting productivity term, B_f, to target a gender wage gap of 19%, as the Federal Statistical Office of Germany reported in 2019.

Table 1: Summary Table Calibration

Parameter	Description	Value
Panel A: Externally Calibrated Parameters		
γ	degree of relative risk aversion	2
β	discount factor	0.986
$\underline{a}$	borrowing constraint	0
Panel A.1: Estimated Parameters (from German Data)		
θ	elasticity of substitution male/female labor	0.7
τ	curvature parameter labor income tax	0.188
η	average tax parameter	0.751
$\rho_{z,m}$	persistence income shock	0.801
$\rho_{z,f}$	persistence income shock	0.745
$\sigma_{z,m}$	standard deviation income shock	0.295
$\sigma_{z,f}$	standard deviation income shock	0.330
Panel B: Internally Calibrated Parameters		
φ_f	dis-utility of labor	0.390
φ_m	dis-utility of labor	0.330
b	replacement rate	0.102
B_f	factor-augmenting productivity term	5.3212

4.1.3 Model and Data

Before I analyze the effects of the tax reform, I describe the properties of the benchmark economy and contrast them with data to build confidence that the model is an adequate framework to assess the research question. I focus on the following aspects of the model: how does female labor supply respond to wage rate changes, and how does it change to within-household productivity (and thus income) differences and household assets?

The calibration explicitly targets married couples' aggregate labor force participation rate and the fraction of households that do not participate. The remaining moments are the distribution of participation of married households and female labor supply elasticities with

respect to the wage rate, their husbands' productivity, and household wealth. They arise due to the mechanisms influencing labor supply within the model and are not targeted but match the corresponding moments in the data well. Furthermore, the model does reasonably well in matching the level of wealth dispersion we see in the data. The Gini coefficient is 0.477 in the model and 0.542 in the data.

As Table 2 shows, at the aggregate level, the model matches well the critical targets in the data. By construction, the relative labor force participation rate equals ⅔. The model implies that the aggregate female labor supply rises by 0.5% in response to a 1% increase in female wage rates.[5] This implied elasticity of female labor supply is a crucial benchmark used in the literature; see, e.g., Borella et al. (2021). For the elasticity of an increase in female productivity levels, I find a value of 1.1.

Table 2: Model and Data

Statistic	Model	Data
Panel A: Targeted Moments		
Labor force participation of married men	90.9%	91.0%
Labor force participation of married women	59.2%	60.7%
Labor force participation of married HH: $e_f = e_m = 0$	4.3%	4.3%
Gender wage gap	19.0%	19.0%
Panel B: Untargeted Moments		
Labor force participation of married HH: $e_f = e_f = 1$	54.4%	56.0%
Labor force participation of married HH: $e_f = 0$, $e_m = 1$	36.5%	35.0%
Labor force participation of married HH: $e_f = 1$, $e_m = 0$	4.8%	4.7%
Gini-coefficient of wealth dispersion	0.477	0.542
Implied elasticity of female labor supply	0.5	[0.36,0.75]

The literature reports various estimates for this Marshallian labor supply elasticity at the

[5] In addition to smoothing out discontinuities in the policy functions, small taste shocks allow me to utilize the algorithm developed in Auclert et al. (2021) that makes it feasible to calculate Jacobians of heterogeneous-agent problems. Using those, I can compute general-equilibrium impulse responses to approximate the elasticity of labor supply.

extensive margin, and there are multiple ways to measure it in a structural model. Guner et al. (2012), for example, find female labor force participation changes by 0.36% or 0.72% in response to a 1% change in the wage rate or female productivity levels, respectively. Blundell et al. (2016) calculate an elasticity of labor supply of 0.475 for the extensive margin, while the model in Attanasio et al. (2018) implies an elasticity of 0.75. This number is closer to the estimate of 0.7 from Haan & Wrohlich (2011). Hence, my model lies well in the range of estimated elasticities in the literature and also features a somewhat larger elasticity with respect to productivity, as in Guner et al. (2012).

Another way to better understand the workings of the model and to assess how well it captures critical moments of the data is to look at the distribution of labor force participation of households – tabulated in Panel B of Table 2. In the data, 56% of households are dual-income couples, and in only about 5% of households, the wife works, but the husband does not. The reverse, however, is more common: the fraction of couples where the sole earner is male is roughly 35%. The model matches the overall distribution well, though it somewhat over-predicts the share of male earners and underestimates the fraction of dual-earner households.

4.2 Labor Force Participation of Married Households

Next, I document the model's performance in steady-state to highlight some of the properties essential to understanding the effects of the simulated policy reform. Figure 3 illustrates how married women make labor supply choices in the model – fixing the wage rate.

The figure shows the probability that women work as a function of their household's asset position, i.e., their percentile in the steady-state asset distribution. The solid line indicates couples in which the husband has low labor productivity. In contrast, the dashed line represents husbands with medium labor productivity and the dash-dotted line with high productivity. The panels vary the labor productivity level of the wife.

The probability of participating in the labor market increases in the wife's productivity level.

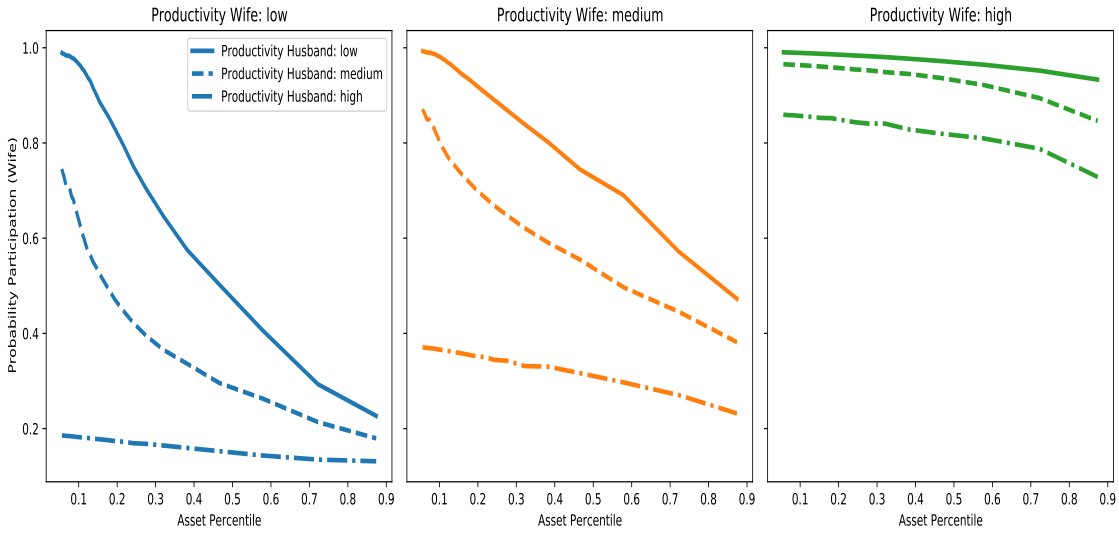

Figure 3: Labor Supply Choices of Married Women

This is because working has a fixed utility cost while income is proportional to productivity. As the rightmost panel shows, highly productive women almost always work independently of their husbands' productivity level or wealth. For all women, the probability of participation is decreasing in assets, i.e., in families with high levels of assets, women are less likely to work. This represents a wealth effect on labor supply in the model.

An interesting pattern emerges in the left panel where women have the lowest productivity level. On average, low-productive women are the least likely to participate in the labor force. We also do not observe a strong wealth effect if their husband has a high labor income. However, for women with a low productive husband, assets matter. They drastically change their decision to participate as a function of their wealth. In poorer households with an unproductive husband, even the least productive women work with a high probability since otherwise, their overall household income (and thus consumption) would be too small.

4.3 Equilibrium Effects of a Labor Supply Shift

Production functions that impose perfect substitutability abstract from a critical feature of the data and potentially lead to much smaller estimates of the benefits (for both men

and women) of increasing female labor force participation. This argument relies on the assumption that men are initially in excess supply (Espinoza et al. 2019). Then, lowering barriers to female labor supply can grow the labor force and correct the imbalance in the production process, which would lead to improvements in the productivity of men. Hence, in contrast to the perfect substitutability case, women may not necessarily displace but complement men in production, implying added benefits for both genders (and the aggregate economy) from more diversity.

In this section, before analyzing the specific reform counterfactual of moving to separate taxation, I characterize the importance of relaxing the assumption of perfect substitution. Specifically, to isolate the effect of varying θ, I study a generic labor supply shock in my model, keeping the tax system fixed. I compare the results between my baseline calibration of the model where the elasticity of substitution between male and female labor inputs equals 0.7 and one where I assume perfect substitutability.

I compute the model's response to a permanent labor supply shock by calculating the elasticity of male and female labor force participation to a shock to the dis-utility of labor φ_g. For ease of comparison, I calibrate the size of the shock in both cases such that the labor force participation response is 1% for the gender for which dis-utility is shocked. That is, I normalize the shock to φ_f such that the general equilibrium effect on female labor force participation is 1% and vice versa for φ_m.

Table 3: Equilibrium Effects of Generic Labor Supply Shift

	Complementarity		Perfect Substitution	
Shock	Female LFP	Male LFP	Female LFP	Male LFP
Female Labor supply (φ_f)	+1.0%	+0.45%	+1.0%	-0.05%
Male Labor supply (φ_m)	+0.12%	+1.0%	-0.01%	+1.0%

Table 3 tabulates the general equilibrium effects of a shift in both male and female labor supply in the model. In the baseline specification, where men and women are gross complements, a supply shock to female labor inputs, which shifts out female labor force participation, also

increases male labor force participation. This effect is sizable since the increase in male labor force participation is almost half as large as in female participation. Conversely, exogenously raising the male labor supply also shifts out female participation.

To illustrate the economic intuition behind this result, recall from Section 3.4 that the wage ratio is

$$\frac{w_f}{w_m} = B_f^{\frac{\theta-1}{\theta}} \left(\frac{L_f^d}{L_m^d} \right)^{-\frac{1}{\theta}}$$

implying that optimal demand for male labor is a function of female labor.

Hence, an increase in the quantity of female labor increases the demand for male labor. Since the demand schedule for male labor is downward sloping, an outward shift of the labor demand curve implies that male wages rise.

In the case of perfect substitution, on the other hand, the wage ratio is equal to B_f, and the shift in female labor supply does not affect demand for male labor as under gross complementarity. This explains why the response in labor force participation to a labor supply shock for the opposite gender is close to zero and, in fact, negative.

The following section describes the effects of abolishing joint taxation, a combination of a positive labor supply shock for secondary earners and a negative supply shock for primary earners.

5 Abolishing Joint Taxation

The counterfactual analysis studies a policy that aims to increase female labor force participation: abolishing joint taxation of married couples. Such a change in taxation has been studied in previous research interested in female labor supply, see, e.g., Guner et al. (2012), Bick & Fuchs-Schündeln (2017), Borella et al. (2021) who find that such a reform generates significant increases in the propensity of married women to work. My main contribution to this literature is the focus on the imperfect substitutability between men and women in the production process, which is motivated by the empirical estimates that reject this assump-

tion for modern-day Germany. As discussed in Section 4.3, under imperfect substitutability, women may not necessarily displace men in production, resulting in additional gains in labor force participation and output.

What is the intuition behind the claim that removing the joint taxation of married couples will increase female labor force participation? Consider the model economy before the reform: As in Germany, a married couple's tax burden is calculated as twice the tax burden of the average household income. Net household income is thus given by

$$y = 2\eta \left(\frac{w_f z_f e_f + w_m z_m e_m}{2} \right)^{1-\tau}$$

After the reform is implemented, hence, under separate taxation, we have

$$y = \eta \left[e_f (w_f z_f)^{1-\tau} + e_m (w_m z_m)^{1-\tau} \right]$$

Assuming for a moment that both husband and wife make their labor supply decision on the intensive margin allows me to highlight intuition about some of the differences between joint and individual taxation in closed form:[6] The marginal return for a women to work an extra hour is then given by

$$\frac{\partial y}{\partial h_f} = \eta(1-\tau) \left(\frac{w_f z_f h_f + w_m z_m h_m}{2} \right)^{-\tau} w_f z_f$$

When we compare this expression to the marginal return under individual taxation,

$$\frac{\partial y}{\partial h_f} = \eta(1-\tau) \left(w_f z_f h_f \right)^{-\tau} w_f z_f$$

we notice that the optimal decision for married women under joint taxation depends on her husband's income, while under separate taxation it does not.

[6] Bick & Fuchs-Schündeln (2018) provide these theoretical insights for the same reason. A part of their analysis is reproduced here for convenience.

Defining the marginal tax rate (MTR) as $1-\frac{\partial y}{\partial h_f}/w$, we can see that under progressive taxation, the MTR (and thus the dis-incentive to increase hours) of the *primary earner is lower*, and the MTR of the *secondary earner is higher* under joint taxation compared to the one they would face under separate taxation. However, notice that Jensen's inequality implies a marriage bonus under joint taxation since the sum of the tax paid on two different incomes is larger than twice the tax paid on the average of those two incomes. Thus, joint taxation is essentially a transfer to married households with a built-in distortion that suppresses the labor supply of the secondary earner, who, empirically, is typically the wife in Germany.[7]

5.1 Labor Force Participation under Individual Taxation

This section reports findings from steady-state comparisons of the pre-and counterfactual post-reform economy. Simulating the reform in the model, I find that moving from joint to individual taxation increases aggregate female labor force participation by 12.6%. Male labor force participation increases by 5.5%, and as a result, output increases by 6.8%. The results for female labor force participation are consistent with what others in the literature find, see, e.g., Guner et al. (2012) and Borella et al. (2021).

I want to highlight two aspects of this result. First, the reform's effect is increasing in the income difference between spouses. Figure 4 shows the labor supply choices of women as a function of their asset position. The left panel shows a household where both husband and wife have median productivity. In contrast, the right panel shows a husband with high productivity and, thus, a larger income difference between husband and wife.

The main reason the responses are so different lies in the tax structure. The marginal tax rate the secondary earner has to pay under joint taxation equals what they would have to pay if they earned the average of the two incomes. Hence, given a progressive tax schedule, the disincentive for the secondary earner to increase their labor supply is increasing in the spousal income differences. Thus, the reform reduces labor supply barriers for women in the right panel of Figure 4 by more due to their income differences within-household.

[7]The share of households in which women out-earn their husbands is 17.2% in the 2019 SOEP.

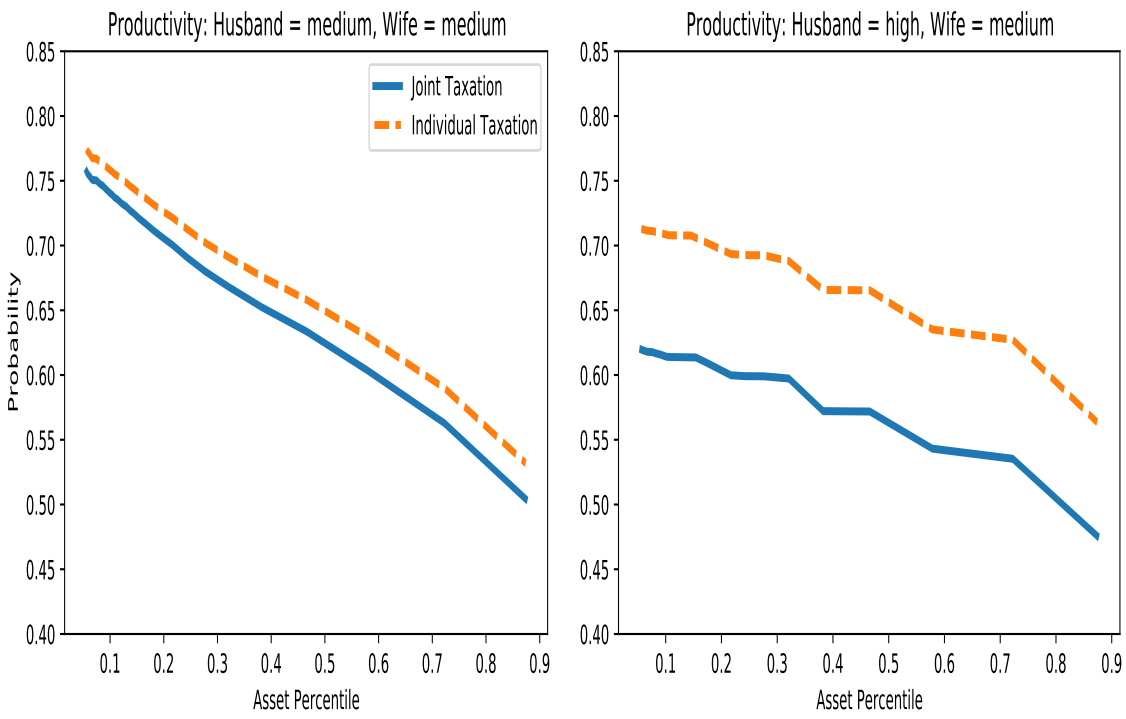

Figure 4: Labor Supply Choices before and after the Reform

In the aggregate, the significant increase in labor force participation is predominantly driven by households that previously had a single male earner. In almost one out of five of these households, the wife decides to participate in response to the reform, making them double-income households. The model also predicts a significant reduction in households with a sole female earner, indicating that many of the (relatively few) men with a working wife who previously decided not to participate are incentivized to supply labor after the reform is implemented.

5.2 Importance of Workforce Diversity

To understand the added benefits of gender diversity in production, I use the model to study how my results vary with the elasticity of substitution. Notice that the gender wage gap in the model, which I use to discipline the relative productivity parameter in the CES production function B_f, is also a function of the elasticity of substitution θ. Thus, I accommodate the

change in the elasticity of substitution by re-calibrating B_f to ensure that the model matches the pre-reform data given a specific level of θ.

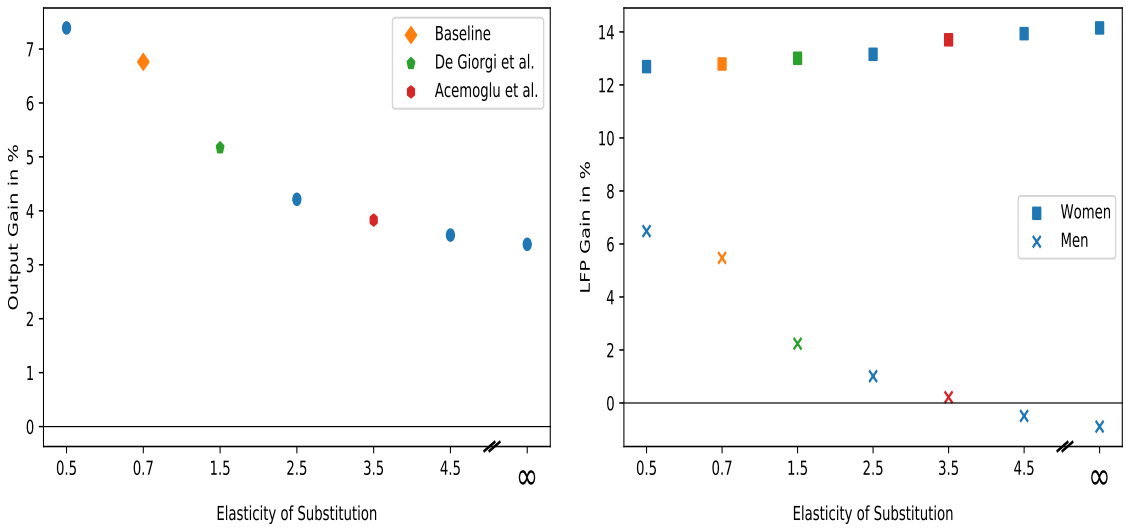

Figure 5: Benefits of Gender Diversity in Production

Figure 5 graphs the results of this exercise, the dashed line indicates the perfect substitution case. The relative output gains from the reform are non-linearly decreasing in the elasticity of substitution. At baseline, output increases by 6.8% in response to the reform, while it only increases by 3.3% in the perfect substitution case. The difference between the results under perfect vs. imperfect substitution is driven by differences in changes in male labor force participation.

While the tax reform has a strong positive effect on female labor force participation for all levels of θ, male labor force participation only increases for smaller values of the elasticity of substitution. At baseline, with $\theta = 0.7$, I find that labor force participation increases by 12.6% for women and 5.5% for men. However, for the limiting case of perfect substitution, male labor force participation decreases slightly in response to the reform – a common finding in the literature, see Guner et al. (2012).

Moving from joint to individual taxation shifts the female labor supply outward and is likely to shift the male labor supply inward, given the large fraction of men who out-earn their

wives. In the baseline scenario where men and women are gross complements, more female labor supply increases the demand for male labor. Hence, a priori, the reform's effect on the equilibrium quantity of male labor is ambiguous, but we expect the quantity of female labor to rise. Simulating the counterfactual in the baseline model, I find that male labor force participation increases, indicating that the outward shift in demand for male labor dominates the inward shift of male labor supply.

What happens to the gender wage gap? Since the demand schedule for male labor is downward sloping, an inward shift of the labor supply curve and an outward shift of the labor demand curve imply that male wages rise. An outward shift of the female labor supply curve suggests a decline in the female wage rate. However, there is a potential second-round effect on female labor demand due to the increase in male labor, implying higher wages for women, ceteris paribus. Implementing the counterfactual in the model, however, I find that the female wage rate declines after the reform, implying a wider gender wage gap.

In the case of perfect substitution, on the other hand, the shift in female labor supply does not affect demand for male labor as under gross complementarity. Hence, the gender wage gap is unaffected. Simulating the counterfactual in the model, female labor force participation increases slightly more than at baseline, but male labor force participation declines. This is the main difference between the two cases and explains the significantly lower output gain under perfect substitutability, see Figure 5.

The predictions of my model under the assumption of perfect substitution align well with values from the literature. In particular, I find that female labor force participation increases by 8.7 p.p. or 14.4%, which compares to a 7.7 p.p or 11.1% increase in Guner et al. (2012) and a 10 percentage points increase for mid-aged, married women in Borella et al. (2021).

5.3 Government Revenues and Welfare

Since the reform increases both output and consumption but also labor supply, the welfare implications are not clear, a priori. Moving to separate taxation will increase the tax burden

of most married households.[8] Thus, the aggregate tax revenue collected by the government will likely increase. Borella et al. (2021) argue that the resulting government surplus could be used to lower income taxation, leading to significant welfare gains for the vast majority of the population. Bach et al. (2020) estimate that the German government could achieve an additional tax revenue of close to 10 billion EUR, which corresponds to roughly 4.5% of labor income tax revenue in 2021, by implementing their preferred specification. They argue that with the additional tax revenue, the government could, e.g., give families higher child benefits and child allowances or finance a better childcare infrastructure which would help families with smaller children more than the savings from joint income taxation.

In this paper, I will focus on the use of additional revenue that has the potential to affect a large faction of the population. Following Borella et al. (2021), I calculate two sets of results to analyze how much the results are influenced by how the government uses the additional revenues. At baseline, I assume the government uses the additional revenue for consumption households do not value. In the second one, I balance the budget constraint by adjusting the proportional component of the income tax.

Table 4: Government Revenues and Welfare

	Complementarity				Perfect Substitution			
Gov. Budget	FLFP	MLFP	Output	Welfare	FLFP	MLFP	Output	Welfare
consumption	12.6%	5.5%	6.8%	0.57%	14.4%	-0.9%	3.3%	-0.09%
lower taxes	11.5%	5.1%	6.7%	0.98%	12.9%	-1.0%	3.1%	0.36%

Table 4 tabulates the results from this exercise. Adjusting the proportional component of the income tax lowers the average tax rate on labor income, which introduces two opposing effects on household labor supply. First, the lower tax rate incentivizes market work. However, adjusting the proportional component also has an income effect. Balancing the government's budget attenuates the labor supply effects of abolishing joint taxation for women, indicating

[8]The tax burden of a household in which husband and wife earn precisely the same is unchanged.

that the income effect dominates. However, the changes are not large enough to overturn previous results or conclusions.

Overall, these policy experiments indicate that abolishing joint taxation would increase married households' welfare when considering complementarities.[9] Assuming that men and women are perfect substitutes, welfare increases only if the government redistributes the additional tax revenues since this counteracts the adverse welfare effects stemming from the increased tax burden for sole-income households.

Note that welfare calculations are sensitive to all aspects of the model. In my case, they could be especially sensitive to modeling the population as single and married households and to modeling home production. I leave these extensions for future work.

6 Conclusion

Many policy proposals to reform income taxation of married couples have been discussed by policymakers and economists alike because the system of joint taxation is perceived by some to be unfair and an impediment to female labor force participation.

In this paper, I develop an equilibrium model with heterogeneous agents, family labor supply, and imperfect sustainability of male and female labor input in the production process to assess the effects of a policy reform that moves from a tax system in which married households are taxed to one in which individuals are the unit of taxation. I utilize German administrative data to leverage a natural experiment that helps me identify the model's central parameter: the substitution elasticity between male and female labor inputs in production.

Along with others in the macroeconomic literature, I find that such tax reform has the potential to generate significant effects on aggregate labor supply and output without changing the progressivity embedded in the tax code. My paper contributes to the literature by highlighting the added benefits of more gender diversity in the labor force.

[9] To evaluate welfare changes, I follow Bhandari et al. (2021) and calculate welfare as an integral over agents' expected utilities: $\mathcal{W} = \int_i z_i u(c_i, s_i) f(i) di$, where f characterizes the invariant distribution.

These benefits are sizable and economically meaningful since the output gains are more than twice as large when imperfect substitutability is considered compared to the perfect substitutability case. The reason is that higher female labor force participation positively affects the demand for male labor. Thus, women do not necessarily replace but complement men in the production process, leading to significant gains for both genders and the overall economy.

References

Acemoglu, D., Autor, D. H. & Lyle, D. (2004), 'Women, War, and Wages: The Effect of Female Labor Supply on the Wage Structure at Midcentury', *Journal of Political Economy* **112**(3), 497–551.

Alon, T., Doepke, M., Olmstead-Rumsey, J. & Tertilt, M. (2020), This Time It's Different: The Role of Women's Employment in a Pandemic Recession, Technical report.

Attanasio, O., Levell, P., Low, H. & Sánchez-Marcos, V. (2018), 'Aggregating Elasticities: Intensive and Extensive Margins of Women's Labor Supply', *Econometrica* **86**(6), 2049–2082.

Auclert, A., Bardóczy, B., Rognlie, M. & Straub, L. (2021), 'Using the Sequence-Space Jacobian to Solve and Estimate Heterogeneous-Agent Models', *Econometrica* **89**(5), 2375–2408.

Azmat, G. & Petrongolo, B. (2014), 'Gender and the labor market: What have we learned from field and lab experiments?', *Labour Economics* **30**(C), 32–40.

Bach, S., Fischer, B. & Haan, Peter und Wrohlich, K. (2020), Reform des Ehegattensplittings: Realsplitting mit niedrigem Übertragungsbetrag ist ein guter Kompromiss, Wochenbericht 41, DIW.

Bardóczy, B. (2020), Spousal Insurance and the Amplification of Business Cycles, PhD thesis, Northwestern University.

Bauernschuster, S., Hener, T. & Rainer, H. (2015), 'Children of a (Policy) Revolution: The Introduction of Universal Child Care and Its Effect on Fertility', *Journal of the European Economic Association* **14**(4), 975–1005.

Bhandari, A., Evans, D., Golosov, M. & Sargent, T. (2021), Efficiency, Insurance, and Redistribution Effects of Government Policies, Technical report.

Bick, A. & Fuchs-Schündeln, N. (2017), 'Quantifying the disincentive effects of joint taxation on married women's labor supply', *American Economic Review* **107**(5), 100–104.

Bick, A. & Fuchs-Schündeln, N. (2018), 'Taxation and Labour Supply of Married Couples across Countries: A Macroeconomic Analysis', *Review of Economic Studies* **85**(3), 1543–1576.

Bien, W., Rauschenbach, T. & Riedel, B. (2006), Wer betreut Deutschlands Kinder? DJI-Kinderbetreuungsstudie, Technical report, Deutsches Jugendinstitut.

Blundell, R., Dias, M. C., Meghir, C. & Shaw, J. (2016), 'Female Labor Supply, Human Capital, and Welfare Reform', *Econometrica* **84**, 1705–1753.

Borella, M., De Nardi, M. & Yang, F. (2021), 'Are Marriage-Related Taxes and Social Security Benefits Holding Back Female Labor Supply?', *Review of Economic Studies* **0**, 1–0.

Bundesministerium für Familie, Senioren, F. u. J. (2013), 'Vierter Zwischenbericht zur Evaluation des Kinderförderungsgesetzes. Bundesministerium für Familie, Senioren, Frauen und Jugend (BMFSFJ)'.
URL: *https://www.bmfsfj.de/bmfsfj/service/publikationen/vierter-zwischenbericht-zur-evaluation-des-kinderfoerderungsgesetzes-kifoeg-bericht-2013–86368*

Bundesministerium für Familie, Senioren, F. u. J. (2018), 'Gesetz zur Förderung von Kindern unter drei Jahren in Tageseinrichtungen und in Kindertagespflege (Kinderförderungsgesetz)'.
URL: *https://www.bmfsfj.de/bmfsfj/aktuelles/alle-meldungen/gesetz-zur-foerderung-von-kindern-unter-drei-jahren-in-tageseinrichtungen-und-in-kindertagespflege-kinderfoerderungsgesetz–86390*

Bundesministerium für Familie, Senioren, F. u. J. (2021), 'Kita-Ausbau: Gesetze und Investitionsprogramme'.

URL: *https://www.bmfsfj.de/bmfsfj/themen/familie/kinderbetreuung/kita-ausbau-gesetze-und-investitionsprogramme-86394*

Caliendo, L., Dvorkin, M. & Parro, F. (2019), 'Trade and labor market dynamics: General equilibrium analysis of the china trade shock', *Econometrica* **87**(3), 741–835.

Carroll, C. D. (2006), 'The method of endogenous gridpoints for solving dynamic stochastic optimization problems', *Economics Letters* **91**(3), 312–320.

Chang, Y. & Kim, S.-B. (2006), 'From Individual To Aggregate Labor Supply: A Quantitative Analysis Based On A Heterogeneous Agent Macroeconomy ', *International Economic Review* **47**(1), 1–27.

Dao, M. C., Hebous, S., Klemm, A. & Perry, V. (2019), Germany: Selected Issues. Tax Pressures and Reform Options, Country Report 19/214, International Monetary Fund.

De Giorgi, G., Paccagnella, M. & Pellizzari, M. (2013), Gender complementarities in the labor market, Questioni di Economia e Finanza (Occasional Papers) 183, Bank of Italy, Economic Research and International Relations Area.

De Nardi, M., Fella, G. & Paz-Pardo, G. (2020), Wage risk and government and spousal insurance, Working Paper 28294, National Bureau of Economic Research.

Dezsö, C. L. & Ross, D. G. (2012), 'Does female representation in top management improve firm performance? a panel data investigation', *Strategic Management Journal* **33**(9), 1072–1089.

Druedahl, J. (2021), 'A Guide on Solving Non-convex Consumption-Saving Models', *Computational Economics* **58**(3), 747–775.

Espinoza, R., Ostry, J. D. & Papageorgiou, C. (2019), The Armistice of the Sexes: Gender Complementarities in the Production Function, CEPR Discussion Papers 13792, C.E.P.R. Discussion Papers.

Felfe, C. & Lalive, R. (2012), Early Child Care and Child Development: For Whom it Works and Why, CESifo Working Paper Series 4043, CESifo.

Fella, G. (2014), 'A generalized endogenous grid method for non-smooth and non-concave problems', *Review of Economic Dynamics* **17**(2), 329–344.

García-Miralles, E., Guner, N. & Ramos, R. (2019), The Spanish personal income tax: facts and parametric estimates, Working Papers 1907, Banco de España.

Ghosh, P. (2018), A study of the long-run substitutability between men and women, Technical report, University of Oklahoma.

Groneck, M. & Wallenius, J. (2020), 'It Sucks to Be Single! Marital Status and Redistribution of Social Security', *Economic Journal* **131**(633), 327–371.

Guner, N., Kaygusuz, R. & Ventura, G. (2012), 'Taxation and household labour supply', *The Review of Economic Studies* **79**(3), 1113–1149.

Guner, N., Kaygusuz, R. & Ventura, G. (2014), 'Income Taxation of U.S. Households: Facts and Parametric Estimates', *Review of Economic Dynamics* **17**(4), 559–581.

Haan, P. & Wrohlich, K. (2011), 'Can child care policy encourage employment and fertility?: Evidence from a structural model', *Labour Economics* **18**(4), 498–512.

Havnes, T. & Mogstad, M. (2011), 'Money for nothing? Universal child care and maternal employment', *Journal of Public Economics* **95**(11), 1455–1465.

Heathcote, J., Storesletten, K. & Violante, G. L. (2017), 'Optimal Tax Progressivity: An Analytical Framework', *The Quarterly Journal of Economics* **132**(4), 1693–1754.

Heathcote, J., Storesletten, K. & Violante, G. L. (2020), How should tax progressivity respond to rising income inequality?, Working Paper 28006, National Bureau of Economic Research.

Holter, H. A., Krueger, D. & Stepanchuk, S. (2019), 'How do tax progressivity and household heterogeneity affect Laffer curves?', *Quantitative Economics* **10**(4), 1317–1356.

Hunt, V., Prince, S., Dixon-Fyle, S. & Yee, L. (2018), Delivering through diversity, Technical report, McKinsey & Company.

Hüsken, K. (2011), Kita vor Ort. Betreuungsatlas auf Ebene der Jugendamtsbezirke 2010, Technical report, Deutsches Jugendinstitut.

Iskhakov, F., Jørgensen, T. H., Rust, J. & Schjerning, B. (2017), 'The endogenous grid method for discrete-continuous dynamic choice models with (or without) taste shocks', *Quantitative Economics* **8**(2), 317–365.

Krusell, P., Mukoyama, T., Rogerson, R. & Şahin, A. (2017), 'Gross worker flows over the business cycle', *American Economic Review* **107**(11), 3447–76.

Mankart, J. & Oikonomou, R. (2017), 'Household Search and the Aggregate Labour Market', *Review of Economic Studies* **84**(4), 1735–1788.

Miyagiwa, K. & Papageorgiou, C. (2007), 'Endogenous aggregate elasticity of substitution', *Journal of Economic Dynamics and Control* **31**(9), 2899–2919.

Mühler, G. (2008), Consequences of Mixed Provision of Child Care – An Overview on the German Market, Discussion Paper 8, ZEW - Centre for European Economic Research .

Müller, K.-U. & Wrohlich, K. (2016), 'Two Steps Forward—One Step Back? Evaluating Contradicting Child Care Policies in Germany', *CESifo Economic Studies* **62**(4), 672–698.

Müller, K.-U. & Wrohlich, K. (2020), 'Does subsidized care for toddlers increase maternal labor supply?', *Labour Economics* **62**, 101776.

Nollenberger, N. & Rodríguez-Planas, N. (2011), Child care, maternal employment and persistence: a natural experiment from Spain, IZA discussion paper 5888, Institute for the Study of Labor (Forschungsinstitut zur Zukunft der Arbeit, IZA).

Rust, J. (1987), 'Optimal replacement of gmc bus engines: An empirical model of harold zurcher', *Econometrica* **55**(5), 999–1033.

Spangenberg, U. (2016), 'Der lange weg zur individualbesteuerung: Gleichstellungsrechtliche perspektiven', *Steuer und Wirtschaft - StuW* **93**(4), 343–353.

Tagesbetreuungsausbaugesetz (TAG) (2004), 'Gesetz zum qualitätsorientierten und bedarfsgerechten Ausbau der Tagesbetreuung für Kinder', Bundesgesetzblatt Jahrgang 2004, Teil I, Nr 76, Seite 3852, ausgegeben zu Bonn am 31. Dezember 2004.

Terjesen, S., Sealy, R. & Singh, V. (2009), 'Women directors on corporate boards: A review and research agenda', *Corporate Governance: An International Review* **17**(3), 320–337.

Wrohlich, K. (2008), 'The excess demand for subsidized child care in germany', *Applied Economics* **40**(10), 1217–1228.

Appendix

A Estimating the Elasticity of Substitution

This section describes the background of the natural experiment and the data used to identify the elasticity of substitution between male and female labor inputs.

A.1 Background Childcare Reform

In 2005, the German federal government introduced the Daycare Expansion Act for a quality-oriented and needs-based daycare expansion, particularly for children under three. The law states that childcare places are to be made available as needed for children under the age of three whose parents are employed, unemployed, in vocational training, school, or university (Tagesbetreuungsausbaugesetz (TAG) 2004). The Child Promotion Act (Kinderförderungsgesetz), which defines a binding deadline for regional authorities, was introduced three years later. Since August 1, 2013, all children between the ages of one and three have a legal right to a place in childcare (Bundesministerium für Familie 2018). Therefore, parents can take legal actions against municipalities if they are unable to find childcare and sue for the costs of private arrangements and or foregone earnings (Bauernschuster et al. 2015).

Since 2008, the federal government has contributed 3.28 billion euros to expanding 560,000 additional childcare places. As a result, the national average childcare rate has almost doubled from about 18 percent to 35 percent in 2020. Nevertheless, the demand is far more sizeable: more than 49 percent of parents with children under the age of three would like a childcare place for their children (Bundesministerium für Familie 2021).

The Daycare Expansion Act was introduced during a time when demand was high – 35% of surveyed mothers with young children stated they wanted a slot[10] – but almost no public childcare was available in West Germany as coverage was below 5% (Bien et al. 2006). As

[10] Wrohlich (2008) estimates that in the mid-2000s, on average, more than half of all mothers in West Germany waited for a spot, indicating an even more drastic excess demand.

such, the Daycare Expansion Act was a declaration of intent. In line with the statement from the Federal Ministry for Family Affairs, newer data shows that excess demand and rationing are still prevalent today, though to a lesser extent (Müller & Wrohlich 2016). Interestingly, despite the lack of public childcare, practically no private market existed when the law was introduced. In the early 2000s, only about 2% of childcare slots were operated by for-profit providers (Mühler 2008). One potential explanation for the lack of private options is strict regulation (Felfe & Lalive 2012). The situation was utterly different in East Germany. The government of the German Democratic Republic supported public daycare for children (even at very young ages), and thus, for mostly historical reasons, there is public childcare available for more than half of all East German children under the age of three.

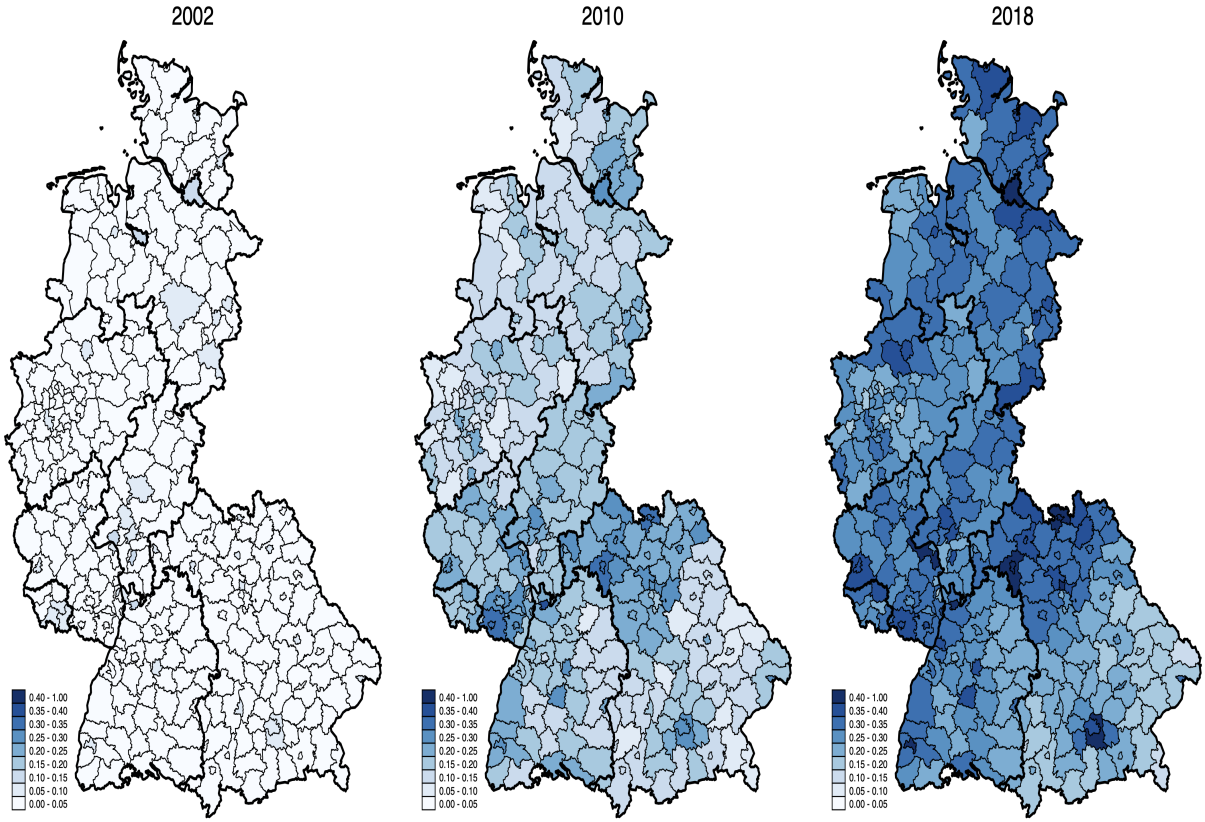

Figure 6: Public Childcare Coverage over Time

The growth in the national average childcare rate was, however, not evenly distributed across regions, federal states, and counties (Landkreise), even within federal states. In fact,

the majority of the variance in childcare coverage comes from variation within federal states. Müller & Wrohlich (2020) summarize the main contributing factor to why some counties were faster to increase their capacity than others: problems and delays in the administrative funding process, shortages of building grounds. Lengthy and complex administrative processes contributed to the delayed implementation of the reform, as Felfe & Lalive (2012) point out.

The federal government secured the financing, but local authorities and non-profit providers of childcare facilities had to estimate local demand and develop a proposal for an expansion strategy, which had to be approved at the state level. These demand projections were subject to forecast errors. Furthermore, as the application process for federal funding was not standardized by federal or state law, there were severe delays in permitting procedures or outright rejection of applications for non-compliance with state regulation (Hüsken 2011). In addition, a majority of counties reported financial as well as spatial constraints and difficulties with various regulations for building childcare centers when attempting to meet local needs (Bundesministerium für Familie 2013).

To illustrate the differences in the availability of slots across regions and time, Figure 6 shows three maps showing the childcare rate for the 324 West German counties in 2002, 2010, and 2018. Before the reform, the childcare rate was consistently below 5% in practically all western German areas. In the years after the reform, coverage then increased almost tenfold and averaged 28.3% in 2018. The maps also show that counties differ significantly in the speed of expansion. In 2010, public childcare rates fluctuated between 6.9% and 36.4%, while in 2018, these numbers were 13.6% and 44.5%. The growth from 2002 to 2018 ranges from a minimum of 11.7 percentage points to a maximum of 40.5 percentage points. However, even in 2018, no West German county has a coverage rate of over 50%.

A.2 The Elasticity of Substitution

To estimate the elasticity of substitution between male and female labor inputs in the aggregate production function, I exploit this considerable variation in childcare coverage, both across time and between West Germany's 324[11] counties generated by the reform.

As described in Section 4, the starting point of my analysis is a standard production function, featuring heterogeneus labor inputs, as in Acemoglu et al. (2004) and De Giorgi et al. (2013):

$$Y = \left[(B_f L_f)^{\frac{\theta-1}{\theta}} + L_m^{\frac{\theta-1}{\theta}} \right]^{\frac{\theta}{\theta-1}}$$

where the elasticity of substitution between between female labor inputs and male labor inputs is equal to θ.

Assuming competitive markets, the ratio of female to male wages can then be expressed in terms of the relative labor inputs

$$ln\left(\frac{w_f}{w_m}\right) = \frac{\theta-1}{\theta} ln\left(B_f\right) - \frac{1}{\theta} ln\left(\frac{L_f}{L_m}\right)$$

This equation can be directly estimated using data on average male and female wages and average male and female labor supply. Year and county fixed effects control for unobserved differences in productivity.

A.3 Data and Sample

For my empirical analysis, I use administrative data on the availability of childcare, labor force participation, and median wages at the county (Landkreise) level. These data are provided by the German Statistical Office and the Federal Employment Agency. I restrict the analysis to West Germany for two reasons: First, childcare coverage in East Germany was already very high, and, thus, the reform had virtually no effect there (Müller & Wrohlich

[11] There are 325 West German counties, but I exclude Osterode am Harz from my analysis based on missing wage and employment data for part of the sample.

2020). Second, due to district reorganizations, there exists no consistent county panel for East German counties during my sample.

Information on the supply of childcare, defined as public childcare slots divided by the population of children three years and below, is available for 1994, 1998, 2002, 2006-2018. For the years 1994, 1998, and 2002, the German Youth Institute gathered these statistics based on data from the Statistical Offices of federal states. The data before 2006 reflects the number of slots, whereas the German Federal Statistical Office reports the number of children attending childcare for the later periods. The severe excess demand for childcare implies that the number of slots should closely track the number of children in childcare.

Table 5: Child care coverage over time

	count	mean	sd	p50	min	max
1998	324	.017	.020	.009	0	.117
2002	324	.022	.023	.015	0	.131
2006	324	.074	.039	.068	.010	.233
2007	324	.093	.044	.086	.022	.286
2008	324	.117	.047	.108	.033	.351
2009	324	.141	.050	.133	.036	.358
2010	324	.170	.053	.161	.069	.364
2011	324	.196	.057	.189	.092	.376
2012	324	.221	.060	.214	.105	.404
2013	324	.240	.060	.235	.113	.448
2014	324	.269	.058	.266	.139	.469
2015	324	.273	.058	.271	.130	.470
2016	324	.271	.059	.270	.143	.479
2017	324	.278	.058	.278	.137	.454
2018	324	.283	.060	.288	.136	.445

Source: Regional Statistics Database (Federal and State Statistical Offices)

Table 5 shows that the average childcare rate in my sample is 17.8% and that coverage varies significantly from 0 to 47.9%. Before the reform, there was essentially no public childcare

available, on average, but there are some counties where roughly one in ten children under the age of three had a slot. After the reform was implemented in 2005, average coverage rates rose. In 2011, on average, there were slots available for one in five children. Even in counties with the lowest coverage rates, one in ten children were provided a spot. At the end of the sample period, the childcare availability averaged around 28%. As Bauernschuster et al. (2015) already noticed, while distribution shifted to the right, there is no convergence between counties – the standard deviation actually increases over time. This points to significant heterogeneity in the speed of the expansion across counties.

My measure of county-level employment is the ratio of employees subject to social insurance contributions by gender (at the place of residence) to the population aged 20-65 by gender. These data are publicly available and provided by the German Statistical Office. For wages, I use the median wages of full-time employees subject to social security contributions by gender on a county level, provided by the German Federal Employment Agency. Table 6 provides descriptive statistics (averages by year) for the employment and earnings data.

A.4 Identification

Relative labor supply in Equation (4.1) is likely endogenous to fluctuations in relative wages. To overcome the problem of endogeneity, I construct an instrument for relative labor supply exploiting the natural experiment of introducing the Daycare Expansion Act in Germany. My instrument is based on the idea that the variation of childcare availability across counties and over time (see Figure 6) has generated exogenous variation in (relative) labor supply: as more and more childcare slots were opened, more and more mothers could enter the labor market.

In my baseline specification, I include a squared childcare term to account for non-linear effects of the reform. One reason to expect that the reform had non-linear effects on female labor supply is that more coverage makes it more likely to get a spot in childcare but it also reduces the distance between home, childcare, and work. Thus the probability that

Table 6: Descriptive Statistics Labor Supply and Earnings

	Female Emp.	Female Earnings	Male Emp.	Male Earnings
1998	.456		.590	
2002	.482	2026	.602	2705
2006	.478	2082	.593	2807
2007	.485	2102	.608	2850
2008	.457	2142	.570	2914
2009	.463	2178	.562	2917
2010	.471	2212	.566	2957
2011	.487	2370	.591	3023
2012	.499	2439	.600	3097
2013	.506	2521	.602	3169
2014	.516	2589	.608	3237
2015	.527	2651	.608	3291
2016	.539	2711	.619	3332
2017	.551	2790	.633	3401
2018	.563	2883	.648	3498
2019	.574	2976	.660	3583
Total	.503	2444	.604	3118

Source: Federal and State Statistical Offices. Note that data on employment is publicly available at the Regional Statistics Database. Earnings data (by gender) is only available upon request.

the childcare spot enables a mother to work is higher the more childcare is available along her commute. To assess the robustness of my findings, I estimate a linear specification and include controls (at baseline, interacted with a time trend) for the population density in a county.

Specifically, I estimate Equation (4.1) using the coverage of public child care, d_{ct} defined as the share of children using childcare in county c in year t, across counties as an instrument

for relative labor supply:

$$ln\left(\frac{F_{ct}}{M_{ct}}\right) = \eta_c + \mu_t + \rho d_{ct} + \xi d_{ct}^2 + \zeta_{ct}, \tag{A.1}$$

$$ln\left(\frac{w_{ct,f}}{w_{ct,m}}\right) = \alpha + \beta ln\widehat{\left(\frac{F_{ct}}{M_{ct}}\right)} + \delta_c + \tau_t + \epsilon_{ct}. \tag{A.2}$$

The ratio of female to male labor supply, $\frac{F_{ct}}{M_{ct}}$, is measured by the employment-to-working population ratio in a county and $w_{ct,g}$ are median earnings for gender g. The coefficient of interest is $\beta = -\frac{1}{\theta}$. Note in this specification that the marginal effects of the expansion in childcare on relative labor supply are restricted to be constant.[12] I include county fixed effects to control for unobserved factors potentially correlated with the availability of childcare in the county but time-invariant. Identification comes from within-country differences over time. I report heteroskedasticity-consistent standard errors clustered at the level of counties to correct for potential serial correlation of the error terms.

For identification, I rely on the argument that the variation in childcare availability is exogenous to changes in the relative wages of men and women and not correlated with time-varying unobserved characteristics of a county that are related to relative labor supply.

I argue that the expansion process in response to the Daycare Expansion Act is exogenous to relative wages as it resulted from a long-winded and complicated funding process involving multiple administrative levels, insufficient space for building day-care centers, and an approval procedure that caused delays and rejection of applications for non-compliance with state regulation.[13] However, there remain concerns that, e.g., macro-level shocks potentially affect counties differently or that social norms about childcare/working and voting behavior vary systematically (Havnes & Mogstad 2011). Furthermore, excess demand and the level of the initial shortages in childcare could be systematically related to economic conditions

[12] Since I do not observe the individual choice of a childcare arrangement, this estimate reflects the intent-to-treat effect on relative labor supply, see also Müller & Wrohlich (2020).

[13] Bauernschuster et al. (2015) and Müller & Wrohlich (2020) study the effect of the reform on fertility and maternal labor supply, respectively, and make a similar argument.

and the financial conditions of local governments – and thus the speed of the expansion in childcare in a region (Nollenberger & Rodríguez-Planas 2011).

However, as long as social norms and voting behavior vary systematically across regions and counties but not also over time, fixed effects should control these factors.

A.5 Empirical Results

The results are based on the sample of West German counties from 1994 to 2018. The estimated coefficients for Equation (4.1) imply small values for θ. In my baseline specification, I find a negative and statistically significant effect of the relative labor supply ratio on relative wages. The coefficient of -1.401 means that the estimated elasticity of substitution equals 0.7, implying that men and women may actually be gross complements in production.

My estimate is smaller than the one from Acemoglu et al. (2004), who find an elasticity of substitution between 2 and 3. However, their context (the U.S. during and right after World War II) is very different from the one I study. De Giorgi et al. (2013) on the other hand, use Italian data from the early 2000s and find values between 1 and 1.5.

Table 7: Estimates of the Elasticity of Substitution

	baseline	control	linear	linear+control
	$\ln {}^{w_f}/_{w_m}$	$\ln {}^{w_f}/_{w_m}$	$\ln {}^{w_f}/_{w_m}$	$\ln {}^{w_f}/_{w_m}$
$\ln {}^{F}/_{M}$	-1.401^{***}	-1.396^{***}	-1.807^{*}	-1.698^{**}
	(0.356)	(0.435)	(0.972)	(.826)
implied θ	0.71	0.72	0.55	0.59
	$[0.37, 1.06]$	$[0.28, 1.15]$	$[-0.03, 1.14]$	$[0.027, 1.15]$
Observations	4,841	4,838	4,841	4838
KP Wald F (FS)	31.61	23.88	3.63	4.75
Clustered SE	✓	✓	✓	✓
County and Year FE	✓	✓	✓	✓

As column 1 in Table 7 shows, the instrument is fairly strong: the first-stage Kleibergen-Paap F-statistics is 31.61, resulting from an significant estimate for the effect of local childcare coverage on the relative employment, see column 1 in Table 8. My baseline estimates imply that an increase in childcare coverage rate by one percentage point, at the current mean of around 0.3, increases the relative employment rate by roughly 3 percent, which is comparable to but somewhat smaller than the effect size found in Müller & Wrohlich (2020).

The estimate for the elasticity of substitution is with 0.55 slightly smaller in the linear compared to the baseline specification. However, while the coefficient of the relative employment ratio of -1.807 is still marginally significant, not accounting for non-linear effects substantially weakens the first stage, see column 3 in Table 7.

One potential concern is that the expansion of childcare was clustered in certain geographical areas. As Figure 6 shows, some of the counties that expanded quickly are close to major cities. Thus, as a robustness exercise, I estimate all regressions also controlling for differences in population density before the reform was implemented. I find that controlling for county population does not statistically significantly change my estimates.

Table 8: First-stage Regressions

	baseline	control	linear	linear+control
	$\ln {}^{F}/_{M}$	$\ln {}^{F}/_{M}$	$\ln {}^{F}/_{M}$	$\ln {}^{F}/_{M}$
d_t	0.493^{***}	0.417^{***}	0.0294	.0474
	(0.057)	(0.055)	(0.032)	(.0303)
d_t^2	-0.961^{***}	-0.773^{***}		
	(0.117)	(0.116)		
${}^{pop_{05}}/_{area}$		$-.0002^{***}$		$-.0002^{***}$
		(.00005)		(.00005)
Observations	5,179	5,174	5,179	5,174
clustered SE	✓	✓	✓	✓
County and Year FE	✓	✓	✓	✓

B Estimating Tax Functions

The German statutory personal income tax schedule is formula based and progressive in income. There are five brackets with different marginal taxes. Income below EUR 9,984 per year is tax exempt, while a taxable income above EUR 277,826 is subject to the highest marginal tax rate of 45%. The tax calculations are based on the amount of annual taxable income X, rounded to the next full Euro, such that T, the income tax liability, is given by[14]

$$T = \begin{cases} 0 & \text{if} \quad X \leq 9,984, \\ (1,008.70\, Y + 1,400)\, Y & \text{if} \quad 9,985 \leq X \leq 14,926, \\ (206.43\, Z + 2,397)\, Z + 938.24 & \text{if} \quad 14,927 \leq X \leq 58,596, \\ 0.42\, X - 9,267.53 & \text{if} \quad 58,597 \leq X \leq 277,825, \\ 0.45\, X - 17,602.28 & \text{if} \quad X > 277,826, \end{cases} \tag{A.1}$$

where $Y = \frac{X-9,984}{10,000}$, $Z = \frac{X-14,926}{10,000}$ are parameters set by the government. As described in Section 2, in the case of spouses who are jointly assessed for income tax, the standard income tax is twice the tax amount that applies to half of their jointly taxable income according to Equation (A.1) (splitting procedure).

Notice however, that the tax formula in A.1 takes as input taxable income. In practice, deductions and tax credits may introduce significant differences between gross income and taxable income. Thus, several papers, e.g., Guner et al. (2014), Heathcote et al. (2017), and Holter et al. (2019), estimate simple parametric functions for the effective average tax rate paid by households using different data sets. I follow their approach and in order to account for the fact that a significant number of German taxpayers face a zero tax rate, I estimate

[14]Source: Einkommensteuergesetz (EStG) §32a Einkommensteuertarif

the same functional form as García-Miralles et al. (2019):

$$t(y) = \begin{cases} 0 & \text{if } y < \tilde{y} \\ 1 - \eta \left(y/\bar{y}\right)^{-\tau} & \text{otherwise} \end{cases} \tag{A.2}$$

where $t(y)$ is the average tax rate for income level y, $\bar{y}$ is the mean income, and $\tilde{y}$ is an income threshold.

To estimate Equation (A.2) for Germany I use the EU Statistics on Income and Living Conditions (EU-SILC). I use data for the years 2014-2018 and restrict the sample to those married households between ages 25-64. For each household, I observe have $I_{e;1}$ (earnings of the primary earner, one with higher earner), $I_{e;2}$ (earner of the secondary earner, which can be zero), and $I_{o;h}$ (other, asset plus transfers) income at the household level. Thus the total household income is given by $I_h = I_{e;1} + I_{e;2} + I_{o;h}$. For each household I observe the taxes paid, T_h. I restrict the sample such that the labor income for each partner and taxes paid is non-negative. Thus, I can calculate the tax rate as $t = \frac{T_h}{I_h}$

The mean total gross household income in the sample is 67,360.37 EUR, which I use as a numéraire to allow for a straightforward comparison with the parameters used in the model. The estimate for the effective income threshold 22% or 14,819 EUR. The estimate for η, the parameter governing the average tax rate is 0.7508 and $\hat{\tau} = 0.1883$.

Comparing these values to those implied by the statutory tax formula, we notice that the effective income threshold is higher. However, the German statutory personal income tax schedule formula also implies lower values for the tax progressivity parameter and the parameter governing the average tax rate, i.e., $\tau = 0.14$ and $\eta = 0.743$. Using these values in the model instead, I find that the effects of the reform on female labor force participation are very similar but slightly larger compared to the results at baseline. Importantly, none of the central results or conclusions in this paper are sensitive to the choice of tax function.