

NBER WORKING PAPER SERIES

LONG-TERM CHANGES IN MARRIED COUPLES' LABOR SUPPLY AND TAXES:
EVIDENCE FROM THE US AND EUROPE SINCE THE 1980S

Alexander Bick
Bettina Brüggemann
Nicola Fuchs-Schündeln
Hannah Paule-Paludkiewicz

Working Paper 24995
<http://www.nber.org/papers/w24995>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2018

We thank our discussants Nezih Guner and Henry Siu, as well as Domenico Ferraro, Gustavo Ventura, and seminar and conference participants at Georgetown University, Virginia Commonwealth University, the Federal Reserve Bank of Richmond, the Carleton Macro-FinanceWorkshop, the SED Meeting 2018, and the 2018 NBER International Seminar on Macroeconomics for helpful comments and suggestions. Enida Bajgoric, Pavlin Tomov, and Mariia Bondar provided excellent research assistantship. We thankfully acknowledge financial support from NORFACE under the DIAL programme, and from the Clusters of Excellence “Formation of Normative Orders” and “Sustainable Architecture for Finance in Europe” at Goethe University Frankfurt. All errors are ours. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2018 by Alexander Bick, Bettina Brüggemann, Nicola Fuchs-Schündeln, and Hannah Paule-Paludkiewicz. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

Long-term Changes in Married Couples' Labor Supply and Taxes:

Evidence from the US and Europe Since the 1980s

Alexander Bick, Bettina Brüggemann, Nicola Fuchs-Schündeln, and Hannah Paule-Paludkiewicz

NBER Working Paper No. 24995

September 2018

JEL No. E24,H24,J22

ABSTRACT

We document the time-series of employment rates and hours worked per employed by married couples in the US and seven European countries (Belgium, France, Germany, Italy, the Netherlands, Portugal, and the UK) from the early 1980s through 2016. Relying on a model of joint household labor supply decisions, we quantitatively analyze the role of non-linear labor income taxes for explaining the evolution of hours worked of married couples over time, using as inputs the full country- and year-specific statutory labor income tax codes. We further evaluate the role of consumption taxes, gender and educational wage premia, and the educational composition. The model is quite successful in replicating the time series behavior of hours worked per employed married woman, with labor income taxes being the key driving force. It does however capture only part of the secular increase in married women's employment rates in the 1980s and early 1990s, suggesting an important role for factors not considered in this paper. We will make the non-linear tax codes used as an input into the analysis available as a user-friendly and easily integrable set of Matlab codes.

Alexander Bick
Department of Economics
W. P. Carey School of Business
Arizona State University
P.O. Box 879801
Tempe, AZ 85287-9801
alexander.bick@asu.edu

Bettina Brüggemann
Department of Economics
Kenneth Taylor Hall, Rm 432
McMaster University
1280 Main Street West
Hamilton, Ontario, L8S 4M4
Canada
brueggeb@mcmaster.ca

Nicola Fuchs-Schündeln
Goethe University Frankfurt
House of Finance
60323 Frankfurt
Germany
fuchs@wiwi.uni-frankfurt.de

Hannah Paule-Paludkiewicz
Goethe University Frankfurt
House of Finance
60323 Frankfurt
Germany
hannah.paule@hof.uni-frankfurt.de

A data appendix is available at <http://www.nber.org/data-appendix/w24995>

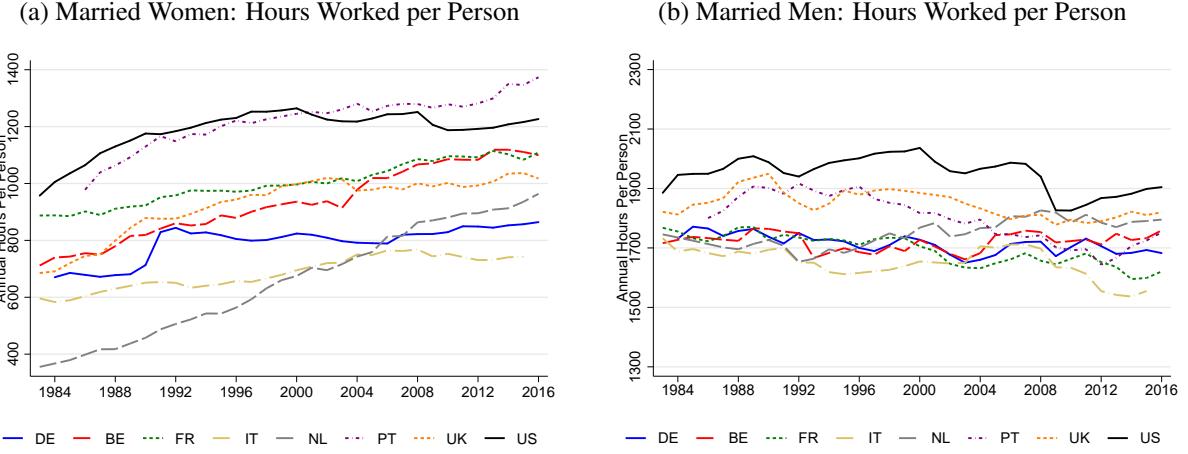
1 Introduction

Understanding the time-series changes in aggregate hours worked as well as differences across countries has long been a focus of the macroeconomic literature. Through the lens of a neoclassical growth model, [Prescott \(2004\)](#), [Ohanian et al. \(2008\)](#), and [McDaniel \(2011\)](#) show that differences in consumption and average income tax rates, combined into one linear tax rate on income, can largely explain differences in the time-series of aggregate hours worked across European countries and the US between the 1950s and today. Other papers focus on hours worked differences in the cross-section of countries, and zoom in on specific demographic subgroups. [Chakraborty et al. \(2015\)](#) and [Bick and Fuchs-Schündeln \(2018\)](#) rely on non-linear labor income taxation and the tax treatment of married couples to explain the differences in labor supply of married couples across countries. [Erosa et al. \(2012\)](#), [Wallenius \(2013\)](#), and [Alonso-Ortiz \(2014\)](#) focus on the elderly and social security systems. [Duval-Hernández et al. \(2018\)](#) and [Ragan \(2013\)](#) analyze the role of taxes and government subsidies in explaining the relatively high hours in Scandinavia, the former distinguishing by gender and education, the latter focusing on the aggregate level. Last, several papers analyze driving forces of the increase in female labor market participation in the US over the last decades (see e.g., [Greenwood et al., 2005](#); [Olivetti, 2006](#); [Attanasio et al., 2008](#); [Albanesi and Olivetti, 2009 and 2015](#); [Knowles, 2013](#); [Buera et al., 2014](#); [Jones et al., 2014](#); and [Ngai and Petrongolo, 2017](#)). Among these, [Kaygusuz \(2010\)](#) and [Rendall \(2018\)](#) focus on the role of non-linear income taxation.

Our contribution to this literature is threefold. First, we provide novel facts for the time-series of hours worked by married men and married women exploiting the data set developed in [Bick et al. \(2018\)](#). Our sample covers the US, the four largest European economies (Germany, the UK, France, and Italy), as well as Belgium, the Netherlands, and Portugal from the early 1980s to 2016. While the secular increase of (married) women's employment for these countries has been well documented, we show that, in contrast, the trends in hours worked of employed married women vary considerably across these countries. Married men's hours worked are largely flat over the sample period. Second, using a quantitative model of joint labor supply, we ask whether these trends in hours for married couples are also to a large degree driven by income taxation, just as aggregate hours. Specifically, we use the model developed in [Kaygusuz \(2010\)](#), [Guner et al. \(2012a, 2012b\)](#), and [Bick and Fuchs-Schündeln \(2018\)](#) to analyze the contribution of non-linear labor income taxes and consumption taxes, as well as of the gender and educational wage premia and the educational composition, to the evolution of hours worked of married couples over time. Third, as a crucial input into our model we use the country- and year-specific full non-linear statutory labor income tax codes, including the tax treatment of married couples. As an independent contribution, we will publish these tax codes as a user-friendly and easily integrable set of Matlab codes.

The focus of our analysis on married couples is motivated by several facts: first, among the population aged 25 to 54, married couples constitute on average around two thirds of households. Second, married women's labor supply exhibits the largest increase and cross-country variation over the last 35 years: average annual hours worked per person increased by 330 hours to on average 1060 hours for married women, but only by 50 hours (to on average 1160 hours) for unmarried women. Over the same time period, married

Figure 1: Labor Supply of Married Couples



Note: Sample consists of married couples aged 25 to 54. The jump in hours worked per person for Germany in 1991 is a consequence of the reunification of East and West Germany in 1990.

men's hours worked were largely flat or even slightly falling. The secular increase in married women's hours worked implies an increasing importance of married women's hours for aggregate hours. Figure 1 shows the stark difference in the time trends of average hours worked of married men and women.

For married women, the two margins of labor supply behave quite differently, as Figure 2 in the next section shows. On average, employment rates increased by 23 percentage points between the early eighties and 2016. In addition, employment rates display substantial convergence over time: they ranged from 35 to 60 percent in the 1980s, and converged to levels between 70 and 80 percent in 2016, with the exception of Italy. The other margin of married women's labor supply, hours worked per employed, features much more variation in levels as well as trends and shows no convergence: average annual hours worked per employed married woman vary between 1000 and 1800 hours throughout the whole sample period.

Bick and Fuchs-Schündeln (2018) demonstrate that for understanding the cross-country differences in married couples' hours worked, it is important to take non-linear labor income taxation and the tax treatment of married couples into account. This tax treatment can fundamentally differ between separate taxation, in which each spouse's taxes depend only on the own earnings, and joint taxation, in which the earnings of the other spouse matter in addition. Kaygusuz (2010) shows that the 1986 tax reform is an important driver for the US time-series of married couples' labor supply between 1980 and 1990, since it decreased the effective marginal tax rate of married women substantially. In our analysis, we take the full non-linearities of the income tax codes in each country into account. Implementing all country-specific small and large tax reforms, we analyze in how far they can explain the changes in the labor supply of married men and women. These tax reforms encompass both changes in the degree of progressivity of labor income taxes (e.g., Italy in 1998 and 2005), as well as changes in the degree of jointness of taxation (e.g. the UK in the 1990s).

Our quantitative analysis proceeds as follows. We first calibrate the preference parameters of the model

for each country to data targets from 2016, the last year in the time-series. We then separately predict the labor supply behavior for married couples in the US and the seven European countries for the years 1983 to 2016, holding preferences fixed in each country, but using the country-year-specific economic environment. Next, and focusing on the intensive margin of married women’s labor supply, we conduct a decomposition analysis in order to determine the relative roles of labor income and consumption taxes, as well as wages and the educational composition, as driving forces of the labor supply of married women. Last, we evaluate the sensitivity of our results, most importantly to assumptions about transfers, in a series of robustness checks.

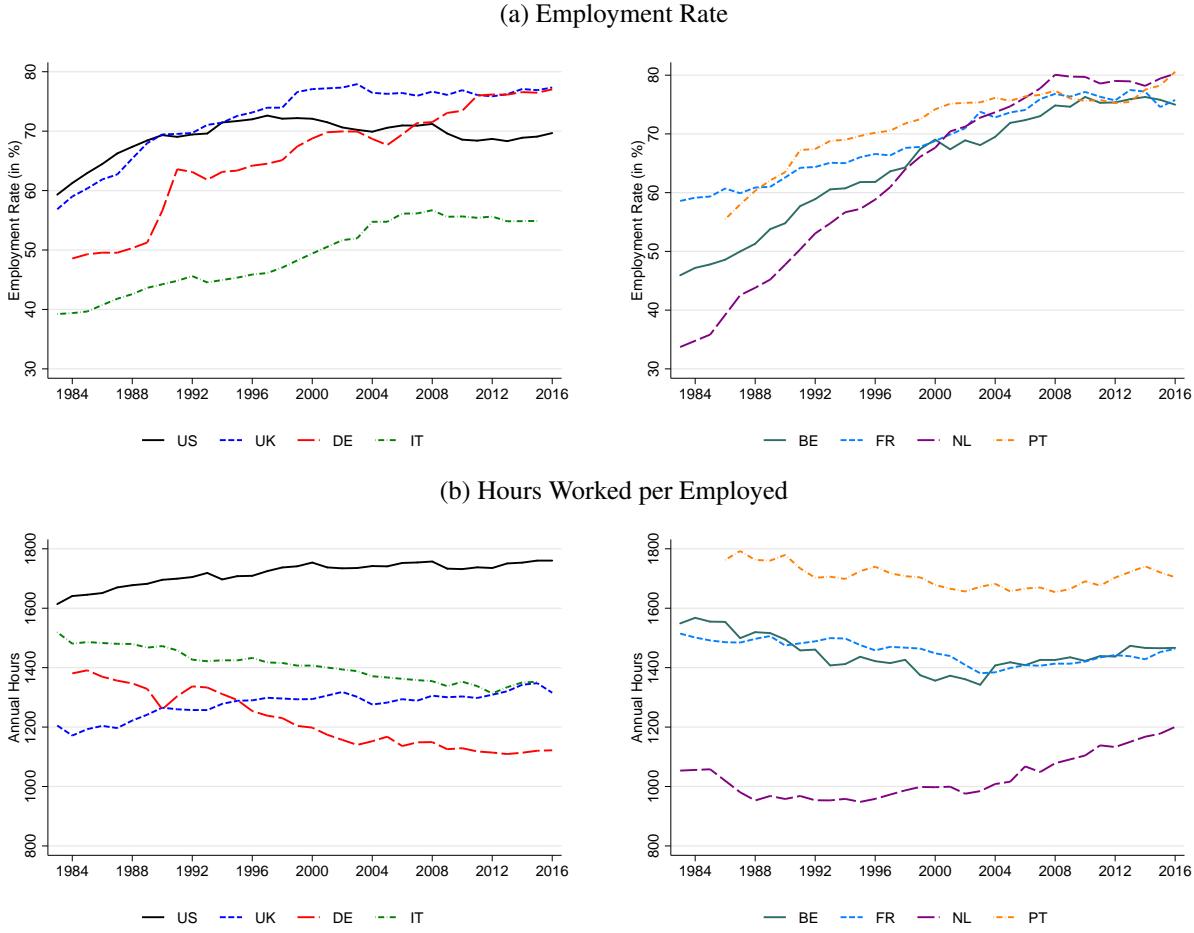
The model is quite successful in replicating the time-series behavior of hours worked per employed married woman: it explains on average 113 percent of the long-term changes in hours worked of employed married women between 1983 and 2016, and replicates the time-series variation across countries very well. Labor income taxes are the key input factor for this result; they alone explain on average 86 percent of the long-term changes in hours worked per employed married woman. The model captures however only part of the large increases in female employment rates in the 1980s and 1990s. Exceptions are the US and the UK. These two countries exhibit the smallest increases in married women’s employment, which are predicted well by the model. This implies that, perhaps not unexpectedly, other factors not considered in our model are additional important driving forces behind the pronounced increases in employment rates of married women in most European countries over the last three decades. The model correctly predicts the small changes in married men’s hours worked for the US, UK, France, Portugal, and Germany, but overpredicts the changes in married men’s labor supply in the three remaining countries (Belgium, the Netherlands, and Italy).

The rest of the paper is organized as follows: Section 2 describes the data and some key patterns in the time-series of married women’s labor supply. In Section 3, we explain the model and its calibration, and provide details on the exogenous inputs that we feed in, especially the non-linear labor income tax codes. Section 4 discusses the results generated by the baseline model and the decomposition analysis. We perform several robustness checks in Section 5. Section 6 concludes.

2 Data and Facts

To document hours worked by married couples along the extensive and intensive margin, we rely on the data set developed in [Bick et al. \(2018\)](#). In that paper, we use two different micro data sets, namely the European Labor Force Survey (EU-LFS), and the Current Population Survey (CPS) from the US, to construct internationally comparable measures of labor supply for different demographic groups. The novelty of that data set lies in measuring not only employment but also hours worked in a consistent way across countries and over time. The main challenge for the comparability arises from the variation in reference weeks in the surveys across countries and within countries over time, which we overcome by adjusting for vacation weeks from external data sources. A detailed description of the data work and an extensive analysis of key features of the data for the recent cross-section (2013-2015) can be found in [Bick et al. \(2018\)](#). Here, we concentrate on the sample of core aged married couples, i.e. married men and women aged 25 to 54. We

Figure 2: Labor Supply of Married Women



Note: Sample consists of married couples aged 25 to 54.

omit younger and older individuals since we do not model differences and changes in education systems and incentives to retire set by the social security systems.¹

Figure 1 shows the evolution of hours worked per person for married men and women since the 1980s in the US and the seven European countries in our sample (Belgium, France, Germany, Italy, the Netherlands, Portugal, and the UK). It documents a secular increase in average annual hours worked per person for married women across these countries, with average hours increasing by 330 hours over the observation period. The Netherlands feature the largest overall increase with 570 hours. At the same time, married men's hours worked were largely flat or even slightly falling: the average change amounts to -35 hours over the same time period.

Next, we split the time series of average hours worked into the extensive and intensive margin. Neither

¹ Wallenius (2013) finds that differences in social security systems have almost no effect on labor supply behavior before retirement, i.e. in the age group we focus on.

of the two margins of labor supply exhibits major trends in the data for married men (see Figure B.1 in the Online Appendix). Figure 2 shows the extensive (Panel a) and intensive margin (Panel b) of married women's labor supply. As we discuss in further detail below, we group countries according to the different patterns observed in the time trends of married women's hours worked per employed. The patterns that emerge for the two margins are quite interesting. Employment rates of married women show a secular increase over the sample period in all countries. Starting from very different levels in the 1980s, ranging from 35 to 60 percent, employment rates increase across all countries in our sample and converge at levels between 70 and 80 percent, with Italy being a notable exception. The Netherlands experienced the largest increase in married women's employment rates from 34 percent in 1983 to 80 percent in 2016. The US stands out in starting with the highest level of the employment rate, but experiencing the smallest overall increase and a leveling off in the early nineties. At this time, the employment rates in European countries were still catching up and eventually overtook the US. In 2016, the employment rate among US married women is the second smallest in the sample countries with 70 percent, with only Italy exhibiting a significantly lower employment rate of only 55 percent, see also [Blau and Kahn \(2013\)](#).

While the trends in the employment rates of married women were uniformly positive across countries, average hours worked per employed married woman show different trends across countries, as one can see in Panel (b) of Figure 2. The left graph of Panel (b) shows a clear positive trend in hours worked per employed married woman between the early 1980s and 2016 in the US and UK, which each grow by more than 100 hours. The other two countries, Germany and Italy, experienced opposite trends with hours worked per employed married woman decreasing by 270 and 140 hours, respectively, between the first and last three years of the sample.² Belgium, France, the Netherlands, and Portugal are shown in the right graph of Panel (b). In these countries, hours worked per employed married woman first decreased and then increased in later years. Unlike what we observe for employment rates, there is no convergence in hours worked per employed married woman: the cross-country standard deviation in married women's hours worked per employed across countries increased slightly from 225 hours in 1983 to 240 hours in 2016. In Online Appendix Figure B.2, we evaluate the robustness of these trends in employment rates and hours worked of married women with respect to different age groups and family constellations. The directions of the trends are robust across three 10-year age groups, as well as across married couples with small children (not yet of school age), with school-aged children, or with neither small children nor school-aged children.

3 Model

3.1 A Model of Joint Household Labor Supply

We build a static model of married couples' hours decisions to investigate in how far changes in consumption and labor income taxes over time contribute to the changes in male and female labor supply over

²The German reunification in 1990 leads to a change in the underlying population, with East Germans entering the sample. This explains the large movements in the employment rate and hours worked of married women in the German data around this time.

time. The model framework is based on [Kaygusuz \(2010\)](#), and [Guner et al. \(2012a, 2012b\)](#), and [Bick and Fuchs-Schündeln \(2018\)](#), and features a maximization problem of a two person household which optimally determines male and female labor supply.³

The description of the model follows closely [Bick and Fuchs-Schündeln \(2018\)](#). There is a continuum of married households of mass one. Each household member exhibits one of three possible education levels, denoted by $x \in \{low, medium, high\}$ for women and by $z \in \{low, medium, high\}$ for men, which determine the offered wages $w_f(x)$ and $w_m(z)$. We denote the fraction of households of type x, z by $\mu(x, z)$ with

$$\sum_x \sum_z \mu(x, z) = 1. \quad (1)$$

Households draw a utility cost of joint work q from a distribution with the probability distribution function ζ . This cost is only incurred if the wife participates in the labor market, and thus introduces an explicit extensive margin choice for women. We abstract from modeling fixed costs of work for men. As a consequence, men always optimally choose to provide positive hours. We follow this approach because employment rates of married men aged 25 to 54 lie above 90% and display only little variation across countries and over time. In order to be consistent with the model, we restrict our sample to include only couples in which the husband is working. The estimated hours worked and employment rates are on average a little higher for this subset of married women, but overall hardly affected by this additional sample restriction. The draw q can be interpreted as a utility loss due to joint work of two household members originating from, for example, inconvenience of scheduling joint work, home production and leisure activities, or spending less family time with children, see [Kaygusuz \(2010\)](#). It captures residual heterogeneity across households regarding the participation choice. For each household x, z , there exists a threshold level $\bar{q}(x, z)$ from which onwards the utility cost of working is so high that the woman chooses not to work, i.e. $h_f = 0$.

As in [Bick and Fuchs-Schündeln \(2018\)](#), households face two types of taxes, namely a linear consumption tax at rate τ_c and a non-linear labor income tax τ_l , which depends on the gross incomes of husband and wife, taking into account tax credits and/or cash benefits.⁴

³[Greenwood et al. \(2017\)](#) present a simplified version of the model focusing on the extensive margin. Note that from a quantitative perspective, abstracting from a life-cycle might not be of first order importance. In [Bick and Fuchs-Schündeln \(2017\)](#), we find quantitatively very similar effects for the same policy reform in the US as [Guner et al. \(2012a\)](#). The latter embeds our static framework in a richer general equilibrium life-cycle model with returns to experience.

⁴The nature and amount of tax credits and/or cash benefits usually depend on the number of children in the households. We explain in Appendix Section [A.1](#) how we incorporate these features.

The maximization problem of a type $\{x, z\}$ household is given by

$$\max_{h_m, h_f} \left\{ \ln c - \alpha_m h_m^{1+\frac{1}{\phi}} - \alpha_f h_f^{1+\frac{1}{\phi}} - q \mathbf{I}_{h_f > 0} \right\} \quad (2)$$

$$\text{s.t. } c = \frac{y_{hh} - \tau_l}{(1 + \tau_c)} + T \quad (3)$$

$$y_{hh} = w_m(z)h_m + w_f(x)h_f \quad (4)$$

$$\tau_l = \tau_l(w_m(z)h_m, w_f(x)h_f) \quad (5)$$

where $\mathbf{I}_{h_f > 0}$ takes the value one if the wife is working and zero otherwise, c represents household consumption, and y_{hh} represents gross household income.

As usual in the literature explaining aggregate hours worked differences between Europe and the US, consumption and labor supply are assumed to be separable, and utility from consumption is logarithmic. Therefore, differences across countries and over time in mean wages are irrelevant, and only differences in the gender and education premia matter for labor supply decisions.⁵ α_m and α_f capture the gender-specific relative weights on the disutility of work, and ϕ determines the curvature of this disutility. This latter parameter is assumed to be the same for men and for women. T represents a lump-sum transfer from the government, which redistributes a share $\lambda \in [0, 1]$ of all government revenues:

$$T = \frac{\lambda}{1 + \tau_c} \sum_x \sum_z \mu(x, z) \left[\int_{-\infty}^{\infty} \tau_l(w_m(z)h_m^*(q), w_f(x)h_f^*(q)) \zeta(q) dq \right. \\ \left. + \tau_c \int_{-\infty}^{\infty} (w_m(z)h_m^*(q) + w_f(x)h_f^*(q)) \zeta(q) dq \right], \quad (6)$$

where $*$ denotes the optimal hours choice given the draw of q . When deciding on how much to work, households do not internalize that their choices impact the equilibrium transfer, but rather take the transfer as given.

3.2 Model Inputs

As inputs into the model, we use country-year-specific information on non-linear labor income taxes τ_l and consumption tax rates τ_c , and additionally the educational composition and matching into couples $\mu(x, z)$, male hourly wages by education $w_m(z)$, and female hourly wages by education $w_f(x)$.

3.2.1 Non-Linear Labor Income Taxes

Since the early 1980s, the OECD provides a detailed descriptions of the country-year specific statutory labor income tax codes in the so-called “Taxing Wages” reports.⁶ In these reports, the OECD describes all relevant

⁵This is explained in detail in Section C.1 of the Online Appendix to [Bick and Fuchs-Schündeln \(2018\)](#).

⁶The “Taxing Wages” reports bear this title since 1999, before that they were published as “The Tax/Benefit Position of Employees” (from 1996 to 1998) and “The Tax Benefit Position of Production Workers” (from 1984 to 1995).

elements to calculate the net tax burden of a household that receives only labor income in a given country and year. These comprise federal, state, and local taxes, employees' social security contributions, cash benefits, and standard deductions.⁷ The country-year specific tax treatment of singles vs. married couples is fully taken into account in these descriptions. From the year 2001 onwards, the OECD provides a Stata program that computes the tax-benefit position of households with precisely specifiable characteristics. Next to the sample period restriction, it is not straightforward to incorporate the output produced by this program into an empirical analysis or computational model. Relying on the descriptions in the OECD documents, we therefore implement the tax codes in Matlab for the US and seven European countries (Belgium, France, Germany, Italy, the Netherlands, Portugal, and the UK) from the early 1980s through 2016.⁸ The program allows to differentiate between single and married individuals. In the following description we directly focus on married couples. Using the number of kids as well as earnings of both spouses as inputs, the program calculates the exact amount that the household has to pay in federal, state, and local taxes, compulsory employee contributions, as well as the cash benefits received by the household (mostly relating to children). Note that our program does not include any welfare payments not directly linked to the tax codes. For more details on our tax codes see Appendix Section A.1.

To be clear, our tax codes are based on the statutory tax codes for labor income including standard deductions. The implied tax burden will differ from the actual tax burden households face if they have income from other sources than labor income, and/or if they claim additional deductions. For the US, we can investigate how large the differences between statutory and actual tax rates are. [Guner et al. \(2014\)](#) report actual/effective average tax rates and average household incomes by household type, i.e. marital status and number of children, and by income quantiles for a representative subsample of US taxpayers in the year 2000. These effective tax rates capture all types of income taxation instead of only labor income taxation, and account for itemized instead of only standard deductions. We calculate the average statutory tax rates implied by our tax codes for the average income by income quintile and household type reported by [Guner et al. \(2014\)](#) and compare them to their estimates of effective tax rates for different household types. With the exception of the top quintile, our statutory average tax rates provide a close approximation of the corresponding effective tax rates, see Online Appendix Section B.2.1. The larger difference for the top quintile is not surprising, since in this income range capital income and itemized deductions become important. Assuming this insight also holds for the other countries, our tax codes are useful in applications for all but the richest households.

It is also worthwhile to point out that for the US our tax codes are a simplified version of the NBER TAXSIM program. This program recreates each year's federal income tax law since 1960 and state laws since 1977. Specifically, the all types of income taxation are captured, itemized deductions are accounted for and the tax burden can be calculated by the state of residence. Thus, while the NBER TAXSIM program offers more detail for the US, the value-added provided from our tax codes stems from the cross-country

⁷Local and state level tax rates typically vary within a country. The OECD reports the tax codes of either a typical area within a country (e.g. in the US and Italy) or the average rates across the country (e.g., in Belgium).

⁸For the years 1995 and 2001 the OECD does not provide a documentation of the tax codes.

panel dimension.

While it is impossible to summarize the complex non-linear labor income tax systems in a few numbers, Figure 3 presents two illustrative measures for married couples. Both reflect key features of the labor income tax code across countries and over time for all countries in our sample. The solid line with circles shows the average labor income tax rate of a household in which the wife does not work, assuming for all years that the husband works the average hours worked of married men in 2016 in the respective country and earns the average wage of married men prevailing in the respective year and country (we call this tax rate “average tax rate”).⁹ The dashed line with triangles shows the average marginal tax rate that the wife in this household faces if she goes from not working to working the average hours worked of married women in the respective country in 2016 (we call this tax rate “average marginal tax rate”, it is also sometimes called the “participation tax rate”). We assume that she earns the average wage of married women prevailing in the respective year and that her husband’s hours are unchanged. For both tax rates, we keep the hours fixed at the 2016 value to ensure comparability across time, and consider the case of childless couples.

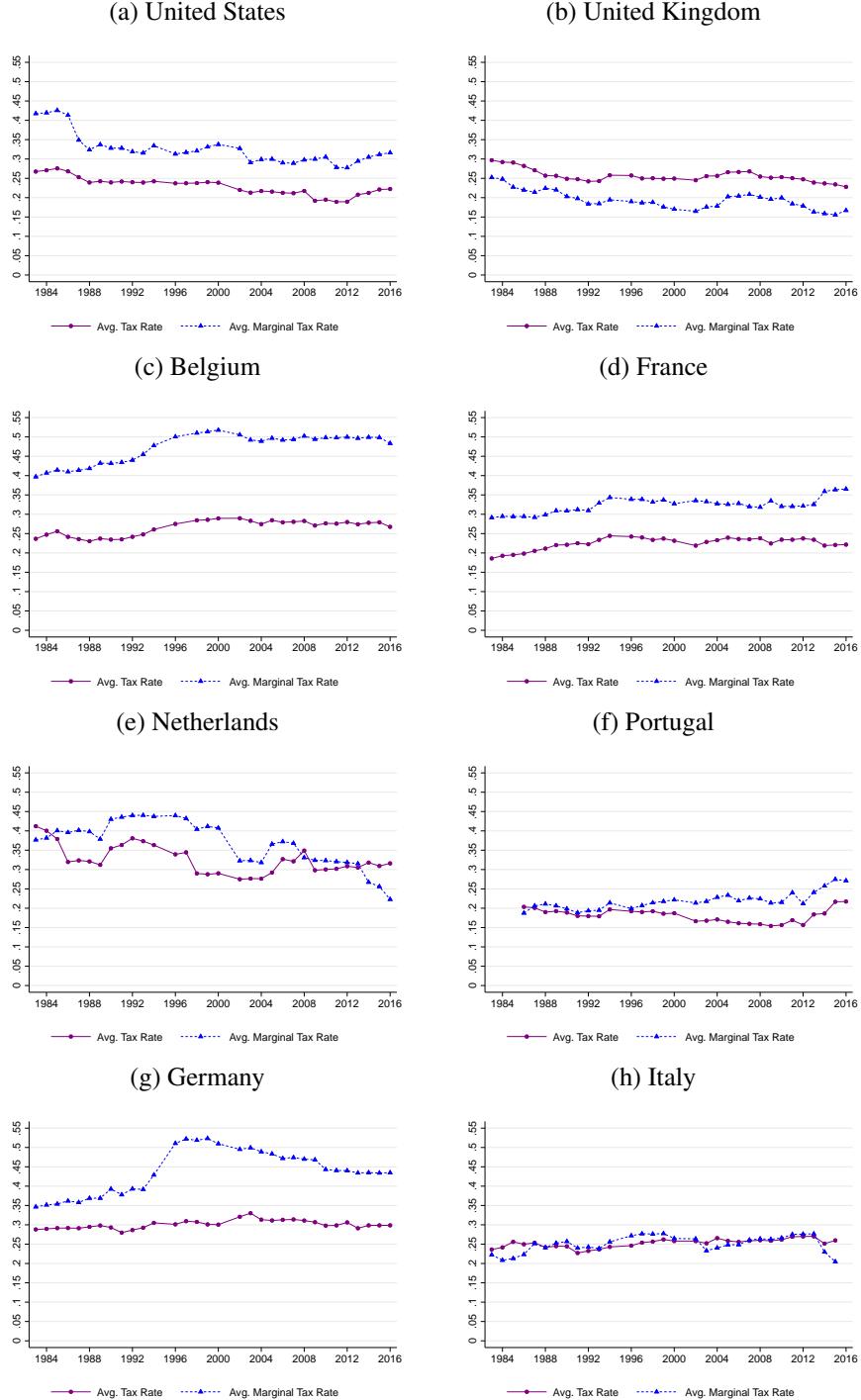
Figure 3 shows that average tax rates are more similar across countries than average marginal tax rates faced by married women. Portugal exhibits the lowest average tax rate with on average 19 percent, and the Netherlands feature the highest average tax rate with on average 33 percent. By contrast, the average marginal tax rates faced by married women range from on average 21 percent in Portugal to on average 50 percent in Belgium. The gap between the average tax rate and the average marginal tax rate provides insights into the tax treatment of married couples in a country. In countries with strictly separate taxation, each spouse is taxed as if he or she was single. In these cases, average marginal tax rates of second-earners tend to be smaller than the average tax rate due to progressivity and lower earnings of the secondary income earner. In Figure 3, the United Kingdom is the only country that exhibits this pattern of tax rates.¹⁰ When couples are taxed jointly, each spouse’s taxes depend not only on the own earnings, but also on the ones of the partner. In the most extreme case of joint taxation, implemented in the US and Germany, the wife’s earnings are added to the husband’s earnings and taxed according to a joint income tax schedule. Joint taxation thus typically implies a higher average marginal tax rate on the wife’s additional earnings than the average tax rate of a single-earner household. In our sample, all countries but the UK feature this pattern of tax rates, but to a differing degree. The gap between the two illustrative tax rates is largest in Belgium and Germany. This gap is not only influenced by the exact working of the joint taxation system, but also by the degree of progressivity in the tax code, with higher progressivity leading to larger effects. The characteristics and consequences of the different tax regimes are discussed in further detail in both [Bick and Fuchs-Schündeln \(2017, 2018\)](#).

In addition to cross-country differences, Figure 3 also shows the impact of tax reforms and changes to the tax code in our sample countries. One example constitutes the 1986 tax reform in the US: the progressivity

⁹For Italy we use the year 2015, the latest available year in the EU-LFS.

¹⁰In the UK, couples were actually taxed jointly prior to 1990. The average marginal tax rate is still lower than the average tax rate because tax brackets were so wide that in many cases the husband’s earnings had no effect on the average marginal tax rate of the wife. The difference between the two tax rates widens after the introduction of separate taxation.

Figure 3: Illustrative Tax Rates: Average Tax Rate and Average Marginal Tax Rate



Note: “Avg. Tax Rate” is the country-year-specific average tax rate evaluated at the average country-specific annual hours worked by married men in the year 2016, assuming the husband is earning the country-year-specific mean married male wage and the wife does not work. “Avg. Marginal Tax Rate” is the average marginal tax rate if the woman goes from not working to working the mean hours of married women in each country in 2016 and earns the country-year-specific mean married female wage, i.e. $[\tau_l(w_m h_m^{2016}, w_f h_f^{2016}) - \tau_l(w_m h_m^{2016}, 0)]/[w_f h_f^{2016}]$. We exclude the years 1995 and 2001 from the graphs because the OECD does not provide a documentation of the tax codes for those years.

of income taxation was lowered, as is apparent for both tax rates, but the effect was larger on the average marginal tax rate of married women. Other examples that led to reductions in both tax rates are the Bush tax cuts in the early 2000s, as well as the tax cuts in response to the 2007 Financial Crisis. Other countries also experienced large changes in their tax structure. Germany and Belgium, for example, experienced large increases in the mid 1990s especially in the average marginal tax rates, while the Netherlands experienced several decreases in this tax rate throughout the late 1990s and early 2000s. As Figure 3 shows, sometimes tax reforms have similar effects on both tax rates, while sometimes they affect one tax rate more than the other.

3.2.2 Consumption Taxes

Consumption tax rates for our sample countries are provided by [McDaniel \(2012\)](#). They are calculated based on NIPA data and capture next to basic value added taxes also all exemptions as well as excise taxes. The data series ends in 2015, and we use the 2015 rates also for the year 2016. For Portugal, the earliest available consumption tax rate is for 1995, and we use the 1995 rate for all previous years.

Panels (a) to (h) of Figure B.3 in the Online Appendix show the time-series of this measure of the consumption tax rates. They have been fairly stable for the US and the UK. France and Belgium experienced rate increases through the early 1990s, which subsequently returned to their initial levels. In the remaining countries consumption tax rates increased significantly throughout our sample period, ranging from an increase of about 5 percentage points in Germany to nearly 15 percentage points in Italy.

3.2.3 Educational Composition and Matching into Couples

We take the percentage of husbands and wives per education group, as well as their matching into couples, directly from the data. We use the three education groups of low, medium, and high education.¹¹ Since the European Labor Force Survey contains information on education only from 1992 onwards, we have to rely on other data sources to impute data for earlier years for all European countries in our sample except Germany. Concretely, we rely on the Barro-Lee Educational Attainment Data ([Barro and Lee, 2013](#)), which provide education shares by gender and age groups. The data are available from 1950 onwards, but only in 5-year intervals, so we first interpolate to account for the missing years in-between. Then we regress the matching shares of married couples aged 25 to 54 from the EU-LFS from 1992 through 2000 on the Barro-Lee educational shares for each gender and age group between ages 25 and 54 (25-29, 30-34, ..., 50-54), and use the estimated coefficients to predict matching shares for all years prior to 1992.¹² Using these predicted

¹¹Low education is defined as primary and lower secondary education (ISCED categories 0 to 2), medium education as upper secondary and non-tertiary post-secondary education (ISCED categories 3 and 4), and high education as any tertiary education (ISCED categories 5 and above). In the US, low education is defined by having completed at most 11th grade of high school; medium education by having completed the 12th grade of high school, having a high school diploma, or attended some college; and high education by having at least a college degree.

¹²Although the Barro-Lee data are available until 2010, we only use data through 2000. Including years beyond 2000 worsens the in-sample fit as measured by the R^2 .

matching shares we then calculate the distribution across educational groups for married men and women for the same years.

For Germany, we rely on the Socio-Economic Panel, which allows us to calculate educational and matching shares for all years since 1984. For the US, we can use the CPS data on educational and matching shares for the full time series. More details on the imputation strategy can be found in Online Appendix Section [B.2.2](#).

All sample countries feature massive shifts over time in the educational composition among married women aged 25 to 54 (see Figure [B.4](#) in the Online Appendix). The share of high-educated married women increased in all countries, while the share of low-educated women fell. Only in the US does the share of high-educated women exceed the share of low-educated women for all sample years. In Europe the number of high-educated married women started to exceed the number of low-educated ones only in the late 1990s to early 2000s. Portugal is the only country in our sample where low-educated married women still constitute the largest group in 2016. For men, the developments were similar, albeit less pronounced (see Figure [B.5](#)).

3.2.4 Gender- and Education-Specific Wages

To calculate hourly wages, we divide earnings by hours. Naturally, for women the issue of self-selection into employment arises. If high ability women of each education group are more likely to join the labor force, then observed mean wages overestimate the mean of offered wages, see e.g. [Olivetti and Petrongolo \(2008\)](#). We therefore apply a simple two-stage Heckman procedure to impute wages of non-working women. The exclusion restrictions are that the income of the husband as well as the presence of children do not directly influence the wage of a woman, see e.g. [Mulligan and Rubinstein \(2008\)](#).

While the CPS has suitable earnings data to estimate hourly wages for married men and women, the EU-LFS do not provide any information on earnings. We therefore rely on a number of other data sets. For Germany, we again use data from the Socio-Economic Panel to compute mean wages for each gender-education cell for the whole time series. For all other European countries, we use a variety of data sources: for the most recent years starting in 2004, we use the EU Statistics of Income and Living Conditions (EU-SILC) to calculate wages by gender and education. From 1994 to 2001, we use the European Community Household Panel (ECHP), the EU-SILC's predecessor.

For the years prior to 1994, we impute gender- and education-specific wages as follows. The only consistent annual earnings measure available over the entire sample period and for all countries are the “average annual wages of production workers” published by the OECD along with their tax documentation described in Section [3.2.1](#). We construct wages by dividing these values with the usual hours of full-time employees from [Bick et al. \(2018\)](#). For each country, we then regress the gender-education-specific wages from the micro data covering 1994 to 2016 on these average production worker wages in each country. Using the estimated coefficient for average production worker wages from each regression, we then predict gender-education-specific wages for all years between 1983 and 2016. We use imputed rather than actual wages throughout in order to smooth high frequency variation in the latter. More details on our imputation

strategy are in Online Appendix Section B.2.3.¹³

In the majority of our sample countries the calculated gender wage gap has narrowed over time, with the largest change happening in the US, where the ratio of wages of married women to wages of married men has increased from 0.66 to 0.84 in 2016 (see Figure B.6 in the Online Appendix). The UK, Belgium, France, the Netherlands, and Italy also experienced a narrowing of the gender wage gap, albeit to a much smaller degree. Notable exceptions are Portugal and Germany: in Portugal, female wages relative to male wages have slightly decreased over time, and in Germany the gender wage gap abruptly narrowed as a consequence of the re-unification but has since then reverted to the values prevailing in the 1980s.

3.3 Correlation of Inputs with Hours Worked by Married Women

Since we are interested in how much taxes and the other inputs contribute to the overall development of the labor supply of married women between the early 1980s and today, we briefly show some facts on the empirical relationship between the four inputs and married women's hours worked per employed. We focus on hours worked per employed instead of employment rates since there is more cross-country variation in the changes over time in hours; similar correlations for married women's employment rates can be found in Online Appendix Figure B.7. Concretely, Figure 4 plots the level change in married women's hours worked per employed between 1983-85 and 2014-16 against the percentage point change of each of our four input variables over the same time period.

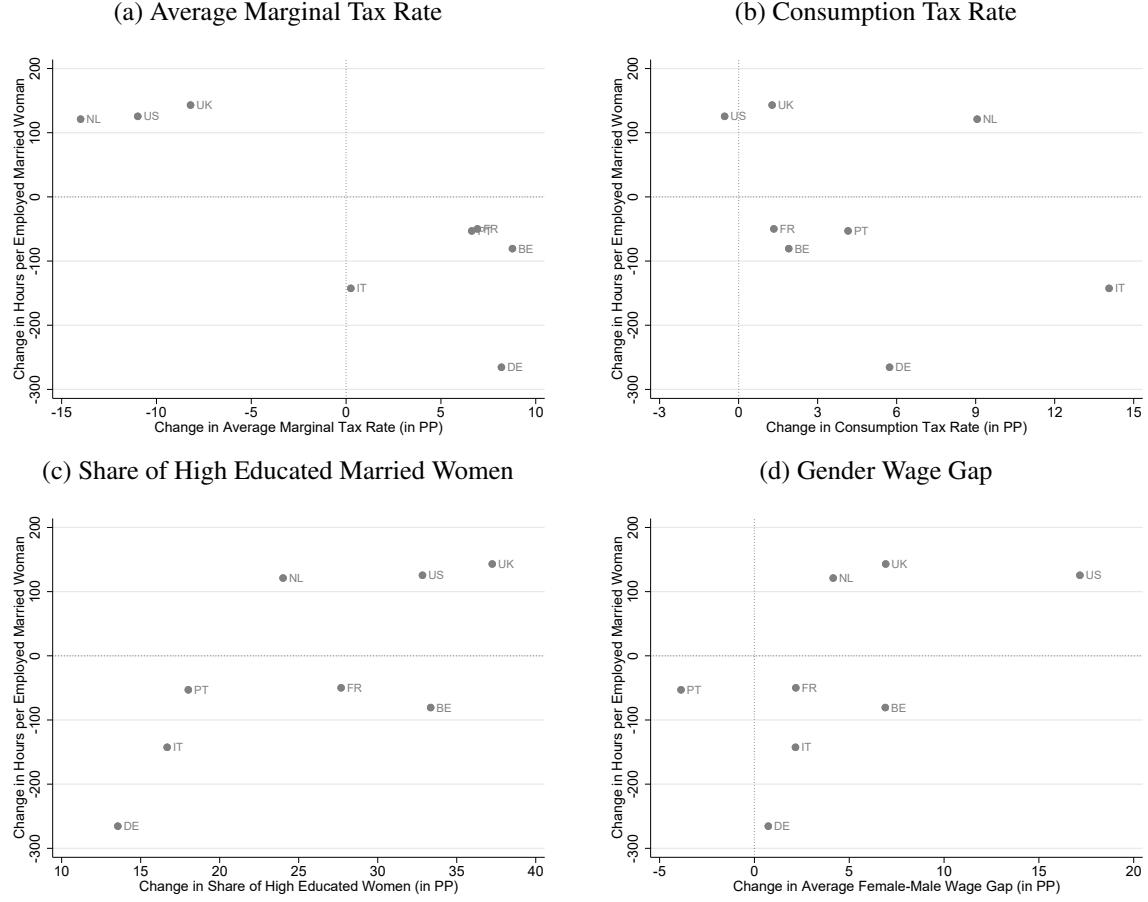
Panel (a) of Figure 4 shows the strongest empirical relationship of the four: changes in average marginal tax rates exhibit a negative correlation with changes in female hours worked of -0.85. Countries with the largest decreases in the average marginal tax rate see the largest increases in hours worked, while the opposite is true for countries experiencing increases in average marginal tax rates. While also negative, the correlation between changes in consumption tax rates and hours worked is much weaker at only -0.37. As one would expect, increases in the share of high-educated married women and in the female-male wage ratio are positively correlated with changes in hours worked per employed married woman. The correlation coefficient with the share of high-educated women is 0.74. The correlation coefficient for the gender wage ratio amounts to only 0.49. In the decomposition exercise further below, we investigate formally to what degree each of the input factors contributes to explaining the changes in labor supply.

3.4 Calibration of Preference Parameters

The preference parameters are calibrated separately for each country in our sample by matching moments from the data for the year 2016. The parameters are then kept constant within each country over time. The only parameter not calibrated but taken from the literature is the curvature in the disutility of working. As in [Kaygusuz \(2010\)](#), we set the labor supply elasticity $\phi = 0.5$, which is consistent with the estimates surveyed in [Blundell and MaCurdy \(1999\)](#), [Domeij and Flodén \(2006\)](#), and [Keane \(2011\)](#). The weights on

¹³In a robustness check in Online Appendix Section B.5.1, we compare actual and imputed wages for the US, and also investigate the impact of our imputation procedure by comparing the model outcomes when using actual vs. imputed wages.

Figure 4: Changes in Married Women's Hours Worked per Employed Against Various Inputs



the disutility of work are calibrated to match the country-specific average hours worked per person by men (α_m ; recall that we do not model an explicit intensive margin for men) and hours worked per employed married woman (α_f).

Following [Kaygusuz \(2010\)](#) and [Guner et al. \(2012a, 2012b\)](#), we assume that the fixed costs q are drawn from a flexible gamma distribution with the shape parameter k_z and scale parameter θ_z being conditional on the husband's type. The probability density function is given by

$$q \sim \zeta(q|z) \equiv q^{k_z-1} \frac{\exp(-q/\theta_z)}{\Gamma(k_z)\theta_z^{k_z}}, \quad (7)$$

where $\Gamma(\cdot)$ is the Gamma function. For each husband's education level z , we select the parameters k_z and θ_z to match as closely as possible the country-specific female labor force participation rates by their wives' own education levels $x \in \{low, medium, high\}$, again for the year 2016. For given preference parameters

α_m , α_f , and ϕ , and conditional on being married to a type z husband, for each woman of education level x there exists a threshold level $\bar{q}(x, z)$ at which she is indifferent between working and not working. Assume for simplicity that all type z husbands work the same amount of hours. Women with more education, i.e. a higher wage, will have a higher threshold q , and therefore a higher labor force participation rate, for any given distribution of q . This pattern is also prevalent in the data, i.e. conditional on the husband's education, the female labor force participation rate is increasing in the woman's own education. The parameters k_z and θ_z are then selected to ensure that the mass of women below these thresholds corresponds to the empirically observed female participation rates by female education conditional on the husband's education in each of our sample countries.¹⁴

The government redistributes a fraction $\lambda \in [0, 1]$ of all government revenues back to the households in a lump-sum fashion. In the benchmark calibration, we follow [Rogerson \(2008\)](#), [Ohanian et al. \(2008\)](#), and [Ragan \(2013\)](#), and assume full redistribution of government revenues and thus set $\lambda = 1$. In Section 5.1 we report the robustness of the results for different assumptions regarding the degree and variation of redistribution.

The model fit for targeted and untargeted moments for each of our sample countries is summarized in Section B.3 of the Online Appendix. That section also lists the calibrated preference parameters for each country.

4 Results

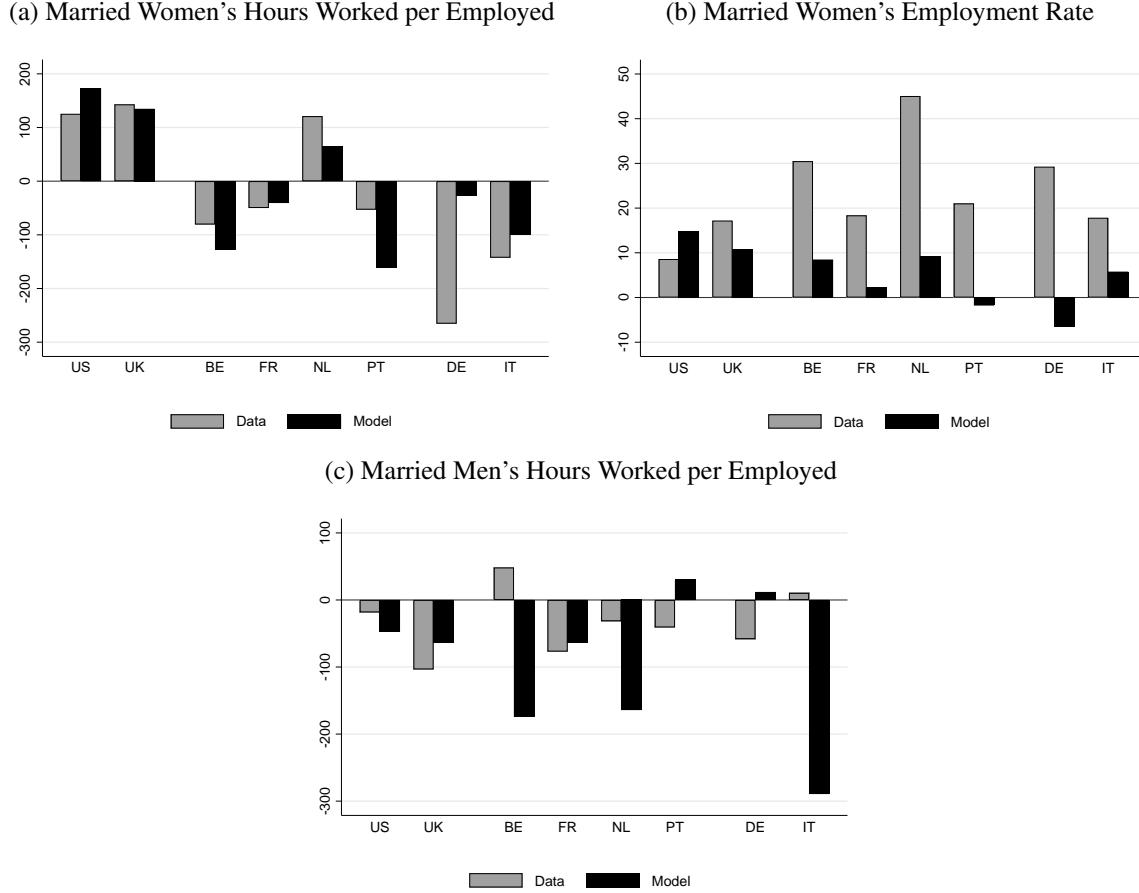
For the quantitative analysis, we proceed as follows country by country: Keeping the country-specific preference parameters fixed over time, we use the country's historical labor income tax systems and consumption tax rates, plus wages and the educational composition (i.e. the educational distribution by gender and the matching into couples), in order to obtain predicted hours worked and employment rates of married couples for each year. We contrast the time-series predictions for hours worked per employed by gender and the employment rate of married women with the data. In a decomposition analysis in the next section, we then evaluate the relative importance of labor income and consumption taxes in explaining the time-series of married women's hours worked. Last, we also analyze the relative importance of time changes in the educational compositions and wages for the results, and their interactions with the labor income tax code.

4.1 Time-Series Predictions of Hours Worked and Employment Rates

We evaluate the performance of our model regarding its ability to replicate the time series behavior of hours worked per employed and employment rates from two perspectives. First, we compare the changes between 1983-85 and 2014-16 in the data and the model. This allows us to assess the model's performance with

¹⁴A priori, there is no reason why this fixed cost distribution should be modelled as being conditional on the husband's education. In fact, it might seem more intuitive to assume that the fixed cost distribution is conditional on the woman's own rather than the husband's education. In that case, in the model the female employment rate would be decreasing in the husband's education holding fixed the woman's own education. This pattern is however at odds with the data.

Figure 5: Changes in Married Couple's Labor Supply between 1983-85 and 2014-16



respect to replicating long-term changes in married couple's labor supply. Second, we report the time-series correlations between the data and the model. Although this measure captures some high-frequency movements that our model cannot speak to, the correlation also captures lower frequency variations in-between the beginning and the end of the sample. This is particularly relevant for the countries that exhibit trend changes over these three decades. In Online Appendix Section B.4.1 we compare the full time-series for all countries and all three variables in the data and the model.

As in Figure 2, we split the sample of countries into three groups: the US and the UK, where married women's average annual hours worked per employed increased throughout the sample period; Belgium, France, the Netherlands, and Portugal, where married women's average annual hours worked per employed first decreased and then increased; and Germany and Italy, where they decreased throughout the sample period.

Figure 5 compares the changes in data and model between 1983-1985 (1986-88 in Portugal) and 2014-

16 (2014-15 in Italy) for hours worked per employed and employment rates of married women as well as hours worked of married men (remember that we assume full employment for married men). The model produces the best fit for married women's hours worked per employed, see Panel (a) of Figure 5. It correctly predicts the direction of changes for all countries. The best fit arises in the UK, the country with the largest increase in married women's hours worked per employed; 94 percent of that increase are replicated. The model overpredicts the increase for the US and accounts for slightly more than half of the increase in the Netherlands. Among the countries with an overall decrease in hours, the best fit is for France with about 80 percent of the decrease being explained. The model overpredicts the decrease for Belgium by 60 percent, and accounts for 70 percent of the decrease for Italy. Hence, overall the fit is not only qualitatively but also quantitatively good. The two exceptions are Portugal, where the small decrease in hours is widely overpredicted, and Germany, where only a small fraction of the large decrease is accounted for.

Panel (b) of Figure 5 compares the long-term percentage point changes of the employment rate in the data and the model. First, just like for hours the model overpredicts the employment rate change for the US and is not able to account for the observed patterns in Portugal and Germany. For the remaining European countries, the model accounts for 13 to 57 percent of the "secular" increase in employment rates. Thus, factors outside of those considered in our model play a more important role for married women's employment rate changes than for married women's hours worked per employed. The fixed costs play a crucial role here. It seems reasonable that these fixed costs, reflecting e.g. general attitudes towards working women or availability of child care, were higher in the 1980s than in 2016.

The model is least successful in capturing the long-term changes in male hours worked (see Figure 5c). It predicts the correct direction of long-term changes for only four out of eight countries (namely the US, UK, France, and the Netherlands). One may worry that the model's failure to predict male hours worked and female employment rates is driving the model's good fit for married women's hours worked per employed. In two robustness checks in Section 5.3, we show that this concern is unwarranted.

We now turn to our second goodness of fit measure. Table 1 reports the time-series correlation between data and model for our three labor supply variables. It confirms the conclusions from Figure 5. The best fit is obtained for married women's hours worked per employed, with correlations between model and data that exceed 0.60 for all countries except Portugal, where the correlation is 0.33. These strong positive correlations are quite remarkable, given the different evolutions of hours across countries. The correlations for married women's employment rates are generally also positive and high, but negative for Portugal and Germany. For married men's hours worked, by contrast, the majority of countries feature a negative correlation between model and data.

To summarize, through the lens of the model, changes in labor income and consumption taxes, as well as changes in the educational composition and wages, have substantial explanatory power for the development of hours worked per employed married woman across our eight sample countries. In contrast, changes in male hours and the stark increases in employment rates among married women are less well accounted for

Table 1: Correlation between Data and Model

Country	Female HWE	Female ER	Male HWE
<i>Positive Hours Trend</i>			
United States	0.92	0.77	0.47
United Kingdom	0.89	0.83	0.93
<i>Changing Hours Trend</i>			
Belgium	0.63	0.59	-0.23
France	0.84	0.65	0.45
Netherlands	0.79	0.64	-0.43
Portugal	0.33	-0.78	-0.13
<i>Negative Hours Trend</i>			
Germany	0.71	-0.75	-0.53
Italy	0.82	0.37	-0.30

by these factors.¹⁵ Thus, in the remainder of the paper, we focus on the intensive margin of married women's labor supply.

4.2 Decomposition Analysis

To understand which input factor is responsible for the success of the model in replicating the evolution of hours worked per employed married woman between the early 1980s and today, we perform a decomposition exercise: one by one, we let only the non-linear labor income taxes, consumption taxes, gender-education-specific hourly wages, or educational attainment and matching vary over time, while holding all other inputs at their 2016 level. A similar decomposition exercise for married women's employment rates is conducted in Online Appendix Section B.5.

When applying the year-specific labor income tax schedule, we account for the fact that tax systems are defined relative to the wage level in a specific year: for each combination of husband-wife hours choices, we calculate the tax rate for a specific year using the 2016 gender-specific education premia and the year-specific mean wage, and then apply this tax rate to the 2016 gross earnings implied by the same husband-wife hours choices. Similarly, when varying the gender-education-specific wages over time, we leave the mean wage at the 2016 level such that the 2016 labor income tax schedule can be meaningfully applied. We describe the details of the procedure in Appendix Section A.2.

¹⁵In Bick and Fuchs-Schündeln (2018), we focus on the cross-country variation in hours worked per person for married men and women in the early 2000s, and find that these four input factors can explain a substantial share of this variation in a larger sample of 18 countries. Regarding the two margins, we also find that the model predicts differences in hours worked per employed married woman better than differences in employment rates.

For all decomposition exercises, we summarize the effects on hours worked per employed married woman for all countries in Figure 6 and Table 2. Since we are especially interested in the role that non-linear labor income taxes play for the evolution of married women's labor supply, we also show the full time-series predictions of the tax experiment for each country separately in Figure 7. Similar graphs for the other three experiments can be found in Online Appendix Section B.4.2.

4.2.1 Non-Linear Labor Income Taxes

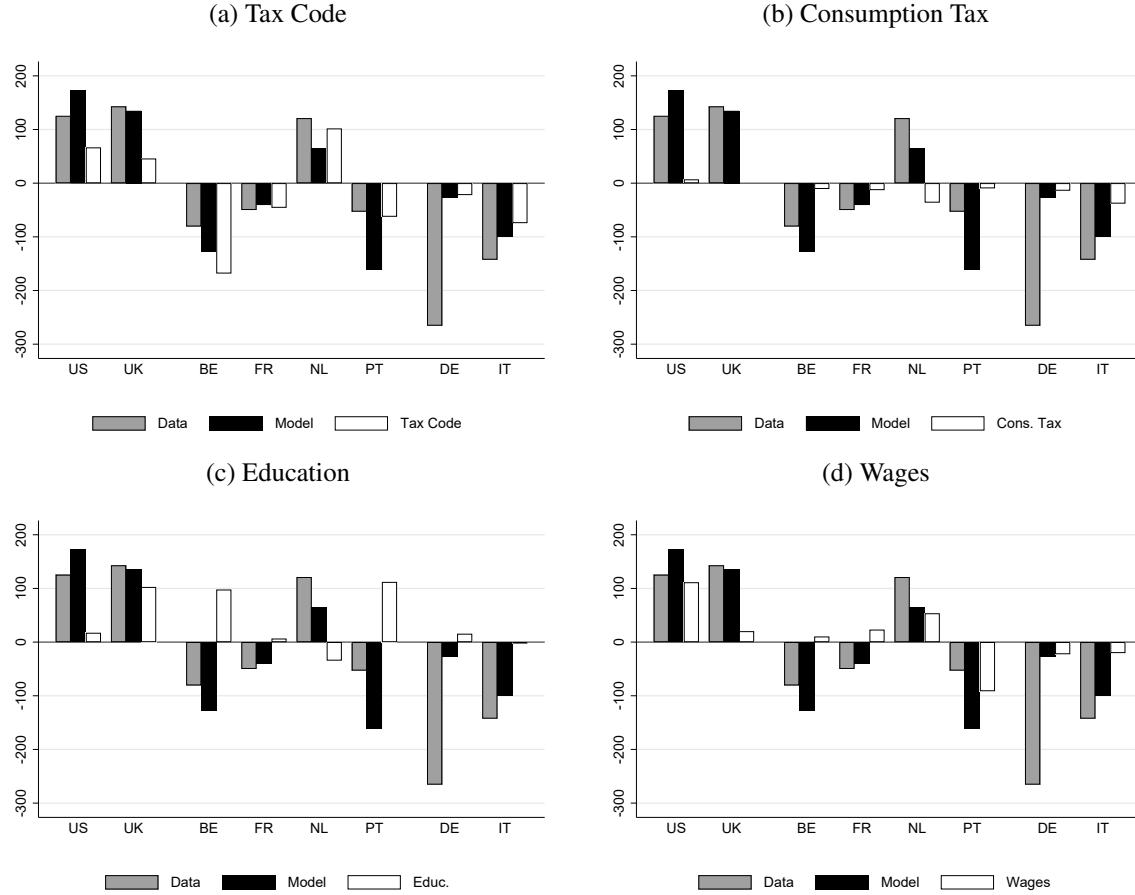
In the first decomposition exercise we let non-linear labor income taxes vary while keeping all other constant at their 2016 level. In Panel (a) of Figure 6, we extend the plot from Figure 5 comparing long-term changes in average hours worked by employed married women in data and full model by adding the implied changes when only taxes vary (white bars). This gives an idea how much of the overall change between 1983-85 and 2014-16 in the data can be explained by tax changes only. In all countries, taxes contribute positively towards explaining the long-term changes observed in the data. For France and the Netherlands, tax changes alone can explain 92 and 84 percent, respectively, of the overall changes in hours worked. The tax experiment performs worst for Germany, where changes in the tax rates can account for only 8 percent of the hours change observed in the data, but are still the most important among the considered driving factors.

The first column of Table 2 repeats the correlation between the data and the output from the full model that we already discussed in Table 1, capturing how much of the time-series variation in hours worked per employed married woman is replicated by the model. The second column of Table 2 reports the correlation between data and model when only taxes vary over time but all other inputs are fixed at their 2016 level. The reported values of the country-specific correlation coefficients confirm the picture that emerged above: taxes play a major role in generating the time-series variation that we observe in the full model. The hours worked per employed married woman generated in the counterfactual analyses are positively correlated with the data for all countries. For France, varying only taxes produces time series variation in hours worked that is more strongly correlated with the data than the hours produced by the full model.

Figure 7 presents the full time series predictions for each country and compares the total model output when all inputs vary over time (the long-dashed black line), to the counterfactual when only taxes vary and all other inputs are kept fixed at their 2016 level (the short-dashed blue line), and the data (the gray line).

The small distances between the black (full model results) and blue (only time-variation in labor income taxes) lines for all eight countries in Figure 7 clearly shows that between the four input factors non-linear labor income taxes play the most important role in shaping the evolution of hours worked per employed married woman over time. In the US and the UK, where average hours increased for employed married women, tax changes account for a significant part of that increase. In the US, the tax experiment alone accounts for 53 percent of the total increase between 1983 and 2016 (the full model accounts for 138 percent). Much of this effect is due to the 1986 tax reform which had a major impact on the average marginal tax rate faced by married women (see Panel (a) in Figure 3). In the UK, changing only taxes over time generates 32 percent of the actual increase in hours worked by employed married women.

Figure 6: Changes in Married Women's Hours per Employed between 1983-85 and 2014-16: Decomposition



As noted earlier, variations in hours per employed married woman generated by the model are highly correlated with hours in the data also for Belgium, France, and the Netherlands, which exhibit a switch in the hours trend in the data. Panels (c), (d) and (e) of Figure 7 show that much of this success is driven by changes in the non-linear labor income taxes. This is especially true for France, where tax changes alone generate the decrease in hours worked per employed married woman in the early part of the sample, and a slightly delayed trend reversal. The same holds for Belgium. For the Netherlands, taxes cannot explain the decrease in hours at the start of the sample period, but in line with the data generate flat hours for a long period right afterwards as well as an increase towards the end of the sample period. For Portugal, where we also see changing trends in the data, panel (f) in Figure 7 shows that the model is not able to predict that pattern, but instead predicts a decrease throughout, which is largely generated by changes in taxes.

In Germany and Italy, married women's hours worked per employed have been steadily decreasing since the early 1980s. For Italy, the model is successful in generating about 70 percent of that decrease, and again,

Figure 7: Hours Worked per Employed Married Woman in Data, Model, and Tax Experiment

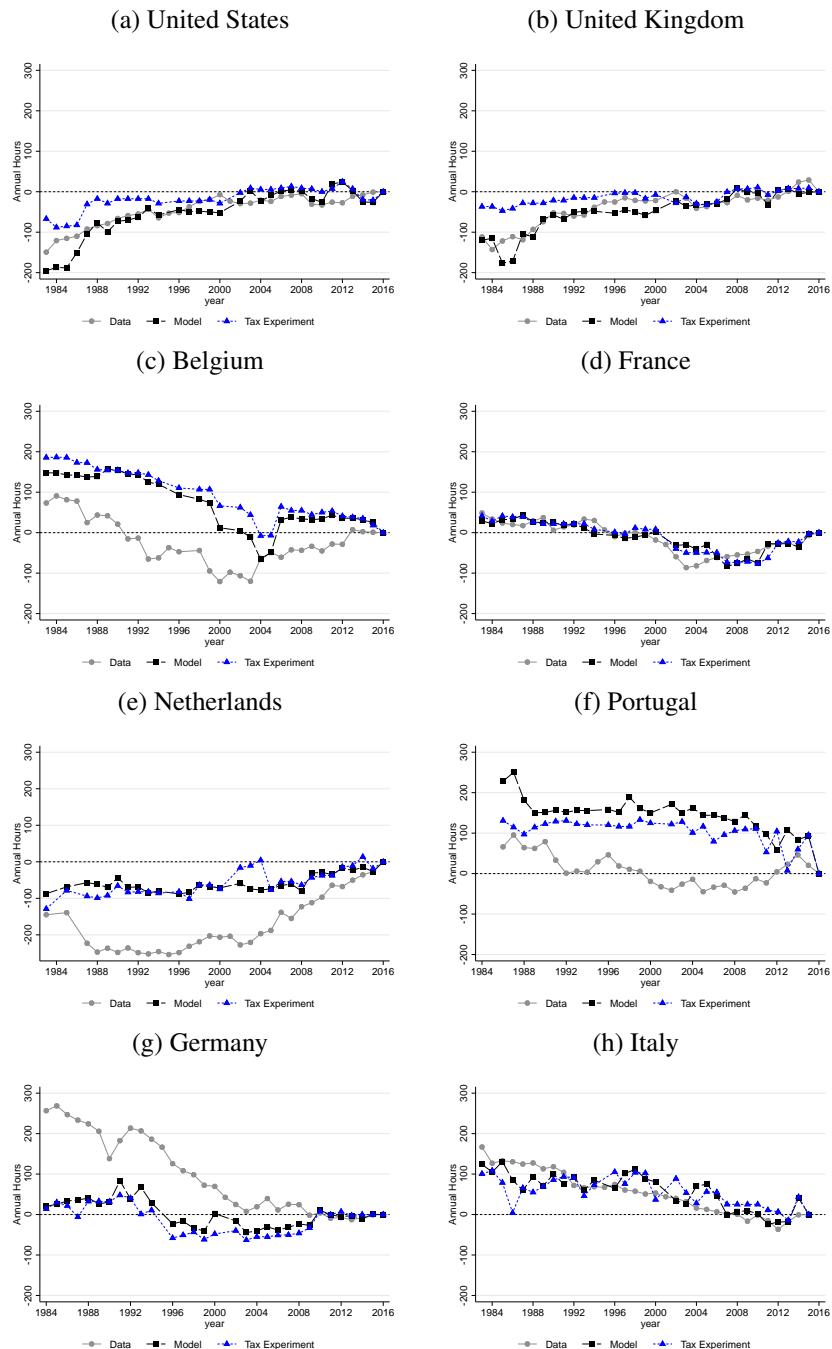


Table 2: Correlation between Data and Decomposition Output:
Hours Worked per Employed Married Woman

Country	Total	Tax Code	Cons. Tax	Educ.	Wages
<i>Positive Hours Trend</i>					
United States	0.92	0.79	0.28	0.46	0.85
United Kingdom	0.89	0.81	0.09	0.79	0.64
<i>Changing Hours Trend</i>					
Belgium	0.63	0.54	0.84	-0.42	0.51
France	0.84	0.89	0.43	-0.46	-0.63
Netherlands	0.79	0.62	-0.64	-0.49	0.39
Portugal	0.33	0.07	0.53	-0.22	0.23
<i>Negative Hours Trend</i>					
Germany	0.71	0.54	0.80	-0.89	0.66
Italy	0.82	0.62	0.84	0.51	0.24

the tax experiment shows that taxes play a big role: letting only taxes vary can explain 52 percent of the decrease observed in the data (see Panel (h) of Figure 7). The model is not able to replicate the large decrease in hours observed in Germany and changes in the tax code imply a much more subdued evolution of hours worked by employed married women than is actually observed in the data (Panel (g) in Figure 7).

4.2.2 Consumption Taxes

Panel (b) of Figure 6 plots again first the changes in married women's hours worked between 1983-85 and 2014-16 in the data and full model. In addition, it plots the changes from an experiment in which only consumption tax rates change over time, and all other inputs are kept at their 2016 level. Consumption taxes have much less predictive power than non-linear labor income taxes. The maximum change in hours generated by the experiment of letting only consumption taxes change over time amounts to only -38 hours in Italy.

The third column in Table 2 shows mixed results regarding the correlation between the time-series variation in the data and that generated by the consumption tax experiment. For Belgium, Germany, and Italy the resulting time-series variation in hours worked is highly correlated with the time-series variation observed in the data. For the other countries, the time-series variation generated by the consumption tax experiment exhibits much lower correlations with the data (United States, United Kingdom, France, Portugal) or even a negative one (Netherlands).

4.2.3 Educational Composition and Matching into Couples

As shown in Figure 4, changes in the educational composition of our sample countries were substantial between the early 1980s and today. However, when varying only the educational composition while holding all other inputs constant at their 2016 level, the model does not replicate the long-term changes in hours well, as can be seen in Panel (c) of Figure 6. Through the lens of the model, the only country where education and matching patterns strongly contribute towards explaining employed married women's hours worked is the UK. Here, the education experiment generates an increase of 103 hours compared to 143 hours in the data. As Panel (c) of Figure 6 shows, for some countries the education experiment predicts a slight decrease in hours worked of employed married women. This might seem counterintuitive given the overall increase in female educational attainment. For employment rates, the education experiment indeed predicts increases in all countries (see Online Appendix Section B.5). As a consequence, government revenues and thus transfers increase, which has a countervailing effect on the intensive margin. For some countries, this effect dominates.

Column 4 in Table 2 confirms that changes in the educational composition and matching into couples generally lack predictive power for the evolution of hours worked observed in the data. In fact, among the four countries with changing hours trends, the educational experiment predicts changes in hours worked that are opposite to the actual changes in the data.

4.2.4 Gender- and Education-Specific Wages

In the last counterfactual experiment, we let gender- and education-specific wages vary over time while keeping the other three inputs at their 2016 level. The results are shown in Panel (d) in Figure 6 and column 5 in Table 2. Figure 4 documented that the female-male wage ratio went up the most in the US. Accordingly, the effects from the wage experiment are largest for the US. Wages play an even bigger role for married women's hours worked per employed in the US than taxes: the correlation coefficient between data and the wage experiment output is 0.85 (compared to 0.79 from the tax experiment), and it produces an increase in hours worked of 111 hours, compared to an increase of 67 hours in the tax experiment and 125 hours in the data.

In Portugal, wages also play an important role, but in the opposite way. There, married women's wages decreased relative to men's, implying a reduction in their hours worked over time, as the wage experiment shows. The remaining countries in our sample feature modest increases in the gender wage ratio. As a result, the predicted changes in the wage experiment are small as well.

France stands out in Figure 6 (d) as the closing of the gender wage gap predicts a weak positive trend in hours worked per employed married woman, contradicting the negative trend we observe in the data until about 2007. Overall, the wage experiment predicts the right direction of long-term changes in hours worked for six out of the eight countries in the sample.

5 Robustness Exercises

5.1 Redistribution of Tax Revenues

In the model, we assume that the government redistributes a fraction λ of its tax revenues to the households through a lump-sum transfer T . In the baseline model we follow [Rogerson \(2008\)](#), [Ohanian et al. \(2008\)](#), and [Ragan \(2013\)](#) and assume that $\lambda = 1$, i.e. that there is full redistribution. This, especially when redistributing through a lump-sum transfer, is a stark assumption. Ideally, one would like to model transfers conditional on household characteristics. Unfortunately, there is no consistent description of transfer systems available that would allow us to do so. Estimating such functions from survey data like the ECHP or EU-SILC goes beyond the scope of this paper and would be a fruitful avenue for future research. In this section, we instead explore the role played by the degree of lump-sum redistribution in three robustness exercises. We report results for all robustness exercises only regarding married women's hours worked per employed.

First, we analyze by how much the model predictions change if we reduce the share of government revenues redistributed to households to 50% ($\lambda = 0.5$) and zero (no redistribution, $\lambda = 0$). Note that our preference specification implies that income and substitution effects of consumption taxes and (hypothetical) linear income taxes cancel each other out in the absence of redistribution. Thus, without redistribution changes in the consumption tax rate and the general level of the tax code have no effect on labor supply. However, changes in tax progressivity still matter even in the absence of redistribution.¹⁶ Redistribution through lump-sum transfers leads to an additional income effect of higher taxes, providing an additional incentive to reduce hours. For each robustness exercise ($\lambda = 0.5$ and $\lambda = 0$), we recalibrate the model for each country, again using moments from 2016. With lower transfers and thus lower income effects, this results in higher weights on the disutility from labor.

Panel i. of Figure 8a shows the model predictions for changes in hours worked per employed married woman for $\lambda = 0.5$ (white bars) and $\lambda = 0$ (light gray bars) and compares them to the baseline model with full redistribution (black bars) as well as the data (dark gray bars). When lowering the degree of redistribution, the model generally predicts lower changes in hours worked than with full redistribution. Yet, the predicted hours changes still point in the right direction, and are also quantitatively large: On average, 99 percent of the long-term changes in the data are explained by the model if we assume 50 percent redistribution, and 79 percent are explained in the case of no redistribution (compared to 113 percent in the baseline model with full redistribution). For countries where the full model over-predicts changes in married women's hours worked per employed, like the US, Belgium, or Portugal, this means that less redistribution improves the fit, for others, like the UK, France, or Italy, the fit deteriorates.

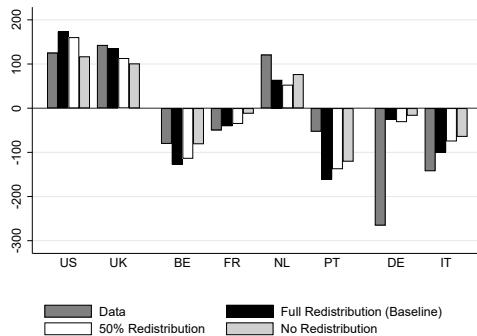
Panel ii. of Figure 8a shows the results from the tax experiment from Subsection 4.2.1 under the two different redistribution schemes. We compare the results from the decomposition analysis in which we only change the tax code for the cases of full redistribution (baseline version, black bars) to the two versions with 50% (white bars) and no redistribution (light gray bars). In some countries, the effects of tax changes

¹⁶This point is shown formally in Online Appendix Section C.1 of [Bick and Fuchs-Schündeln \(2018\)](#).

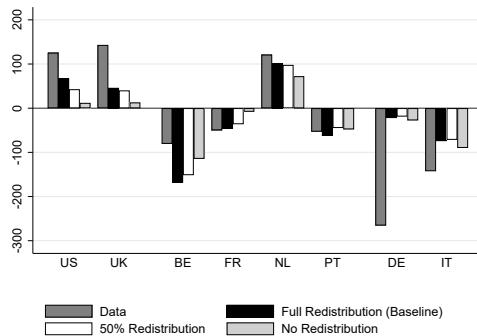
Figure 8: Robustness Checks: Redistribution
Changes in Married Women's Hours Worked per Employed between 1983-85 and 2014-16

(a) Less Redistribution

i. Full Model

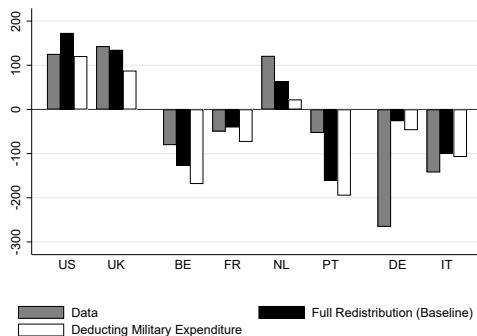


ii. Tax Experiment

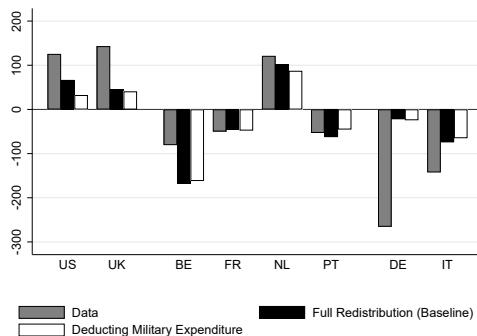


(b) Deducting Military Expenditure

i. Full Model

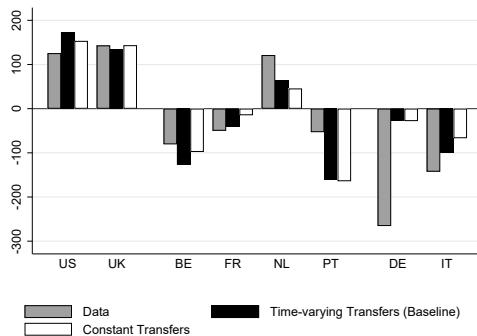


ii. Tax Experiment

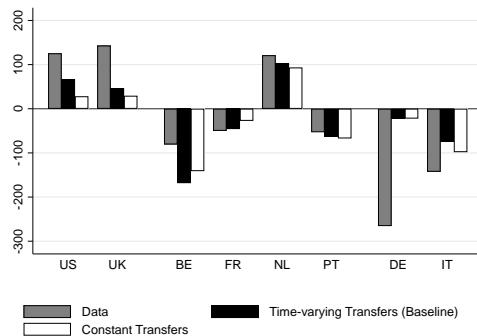


(c) No Time-Varying Redistributive Transfers

i. Full Model



ii. Tax Experiment



on female hours worked per employed are significantly reduced when decreasing the size of the lump-sum transfers, for example the US, UK, or Belgium. This implies that most of the changes in the tax code are changes in the levels of tax rates. In other countries, the tax effects are still strong and hence, are largely driven by changes in the tax structure rather than levels of tax rates (examples are Germany and Italy)

In the second robustness check concerning redistribution we vary the degree of redistribution over time and across countries, again recalibrating the model for each country. More specifically, similar to [Prescott \(2004\)](#), we set λ equal to 1 minus the share of military expenditures among total government consumption expenditures.¹⁷ The idea is that military expenditures are not a substitute for private consumption and therefore do not affect utility. A common feature of the military expenditure time series between 1983 and 2016 among our sample countries is that military spending generally decreases, meaning that the share of government expenditure used for redistribution in the model increases over time. The predicted income effect from taxes via transfers is thus lower in the early 1980s than in the 2000s, leading to a higher incentive to work and therefore higher hours in the 1980s. This is reflected in the results of this robustness check as depicted in Panel i. of Figure 8b: for countries where married women reduce their predicted hours over time in the baseline analysis (Belgium, France, Portugal, Germany, and Italy) the reduction becomes larger when deducting military spending from the amount of revenue redistributed. Countries where predicted hours increase now see smaller increases. This improves the match between long-term changes in data and model in a significant way only for the US, whereas the predictions are further away from the changes in the data for all other countries except Germany and Italy. This worse overall fit is reflected in the average fraction of the data explained by the model which increases from 113 to 124 percent.

In the last redistribution robustness exercise, we do not allow transfers to vary over time and keep them constant at the 2016 level (in relation to average wages). As a consequence, the government budget is not balanced anymore. By shutting down time-series variation in redistribution this way, we want to disentangle time-series changes caused by taxation from those caused by redistribution. The white bars in Panel i. of Figure 8c show the model-predicted changes in married women's hours if transfers are kept constant over time, while the white bars in Panel ii. show the changes in hours induced purely by changes in taxation, keeping all other variables - now also including redistributive transfers - constant. The impact of this experiment on the predicted long-term changes in married women's hours is small: the predictions change only minimally for Portugal and Germany and amount to at most 32 hours among the other countries. On average, the model with constant transfers improves the fit for three of the eight sample countries. The average fraction explained improves to 97 percent.

Summarizing, the three robustness checks regarding our assumptions about the degree of lump-sum redistributions show that in quantitative terms, the results from our baseline specification are very robust to different redistribution assumptions.

¹⁷We use data on military spending and total government consumption expenditure from the UN ([data.un.org](#)).

5.2 Additional Heterogeneity in Income

In the baseline model, wages vary only by education and gender, but there is no further heterogeneity within groups. Income heterogeneity might however be important for the role of non-linear income tax systems. In the next robustness check, we therefore allow for further wage heterogeneity within each education group instead of assuming one gender- and education-specific wage. To do that in a meaningful way, we need to make assumptions on the distribution of wages by education. We can impute the distribution of wages based on the same micro data from the EU-SILC that we also use to calculate wages. Similar to Attanasio et al. (2008), we regress male log hourly wages on a set of year and education dummies, pooling individuals from all years (2004-2016) to have a sufficiently large sample. We predict the residuals from the regression for each individual and then calculate the standard deviation of these residuals for each education group as our country-education-specific measure of wage heterogeneity, which we discretize into three or five states. For women, we assume the same wage heterogeneity as for men. When individuals are matched into couples, we draw randomly from these states within each education type.

Panel i. of Figure 9a shows the results from these robustness checks and compares them to the data and output from the baseline model. The main takeaway is the same as in the previous robustness check: the degree to which variation in the inputs can generate long-term changes in married women's hours work is fairly robust to varying the degree of heterogeneity in wage income. Changes relative to the baseline model go in either direction. On average, the model with 3 income states explains 96 percent of long-term changes in married women's hours worked and the model with 5 income states explains on average 98 percent. The better average fit relative to the 113 percent explained by the baseline model is mainly driven by the much better predictions obtained for Portugal. As Panel ii. of Figure 9a shows, the tax experiment is also largely unaffected by the introduction of additional income heterogeneity, with the exception of the Netherlands.

5.3 Exogenous Male Hours and Exogenous Female Employment Rates

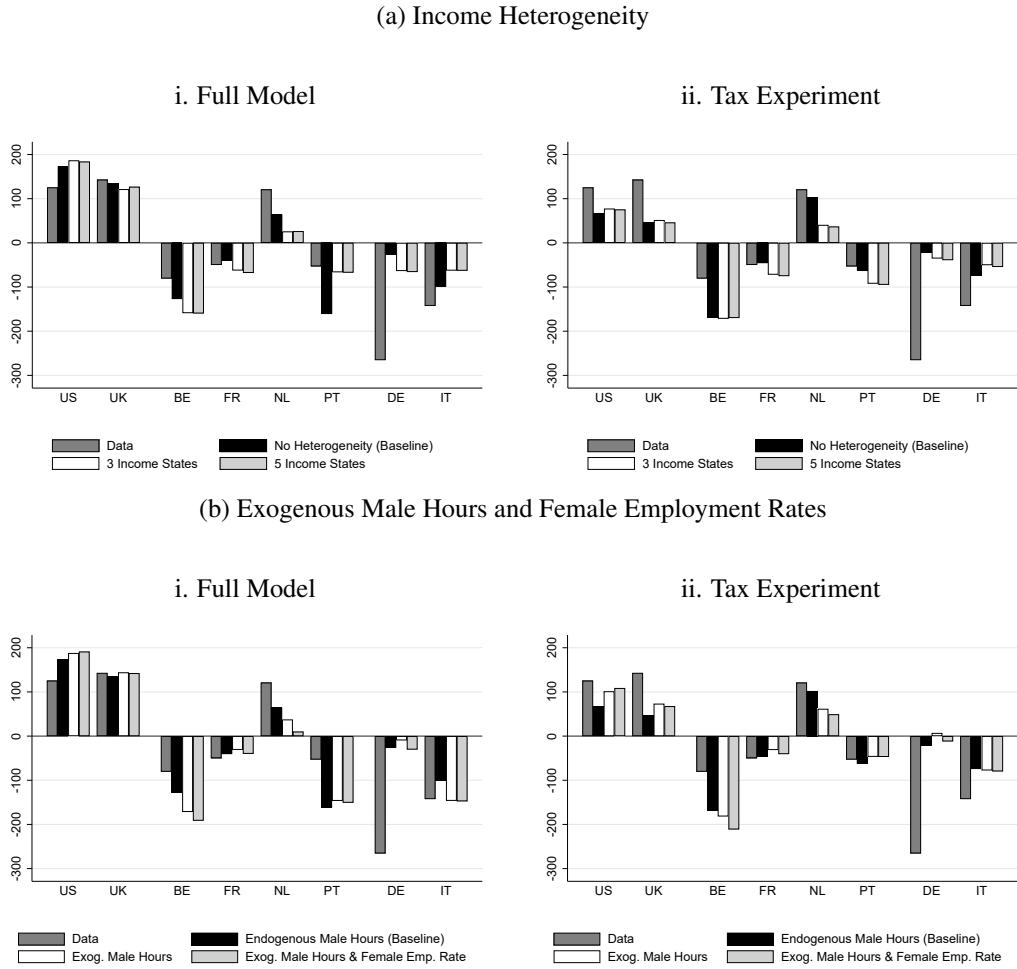
As we show above, the baseline model is not very successful in accounting for changes in hours worked by married men. One might be concerned that the imprecise reflection of married men's behavior in this model of joint decision-making also impacts the predicted changes of married women's hours. To address this concern, we run a robustness check in which we feed in average hours worked by married men exogenously instead of determining them endogenously.¹⁸ This allows us to assess whether the model predicted changes in married women's hours depend to a large degree on the changes of married men's hours. In a second step, we additionally set the average education-specific employment rates by married women exogenously, such that only hours worked per employed married woman remain as the endogenously determined variable.¹⁹

Figure 9b shows that the quantitative impact of both robustness checks on female hours worked is not

¹⁸The calibration of preferences is unchanged relative to the baseline calibration.

¹⁹Note that, since we only have information on education from 1992 onwards, we do not observe education-specific employment rates for married women prior to 1992 in the data, but instead impute them for these years. Concretely, we regress education-specific employment rates on average employment rates in the sample starting in 1992, and use the estimated coefficients to extend the time series backwards.

Figure 9: Robustness Checks: Income Heterogeneity and Fixed Male Hours and Female Employment Rates
 Changes in Married Women's Hours Worked per Employed between 1983-85 and 2014-16



large. When feeding in only male hours exogenously (white bars in Panel i), the fraction of long-term changes explained by the full model increases from 113 to 117 percent. Adding female employment rates (gray bars in Panel i) pushes that fraction further up to 122 percent. Exogenous male hours affect married women's choices in the following way: in countries in which the baseline model predicts decreases in male hours that are larger than in the data (most notably Belgium, the Netherlands, and Italy), exogenously feeding in male hours from the data implies lower male hours in the 1980s. Women partially compensate for this by providing more hours to the market in the 1980s. This leads to larger predicted negative changes (or smaller predicted positive changes in the case of the Netherlands) for married women's hours worked per employed than in the baseline case, as can be seen in Panel i. of Figure 9b. The adjustment is however only partial, and the effect thus small. The opposite holds true for countries like Germany and Portugal, where

the baseline model predicts lower hours worked by married men in the 1980s than observed in the data.

Adding exogenous employment rates for married women hardly alters these results. Conditional on working, women's hours choices depend on their husbands' hours and on the (education-specific) wages they are paid, but do not vary within a given match. The differences between the two robustness checks in Figure 9b are therefore mainly driven by changes in the composition of working women across these matches,²⁰ as well as to a lesser degree by differences in the equilibrium transfers. Varying only taxes while feeding in exogenous hours worked by married men as well as exogenous employment rates for married women as shown in Panel ii. of Figure 9b delivers very similar patterns. Taxes are thus an important driver of the adjustment of hours worked of employed married women also when forcing male hours and female employment rates to be equal to the ones in the data.

6 Conclusion

While the secular increase of (married) women's employment has been well documented, this paper provides novel facts for hours worked by married men and married women. We show that in contrast to the rather uniform increase of married women's employment rates, trends in hours worked per employed married woman vary considerably across countries. Hours for married men show only little variation over time. We use a model of joint household labor supply decisions to analyze the quantitative role of non-linear labor income taxes for explaining the evolution of hours worked of married couples over time, using as inputs the full country- and year-specific statutory labor income tax codes. We further evaluate the role of consumption taxes, gender and educational wage premia, the educational distribution, and the degree of assortative matching into couples.

The model is quite successful in predicting the time-series behavior of hours worked per employed married woman: it explains on average 113 percent of the long-term changes in hours worked of employed married women between 1983 and 2016, and replicates the time-series variation across countries very well. Labor income taxes are the key driving force: they alone explain on average 81 percent of the long-term changes in female hours. These results are robust to varying degrees and forms of redistribution, as well as the introduction of within-group income heterogeneity.

The model captures only part of the secular increase in married women's employment rates in the 1980s and early 1990s in Europe. Exceptions are the US and UK, where the model performs much better with respect to the evolution of employment rates. This suggests that, perhaps not unexpectedly, other factors not considered in our model are additional important driving forces behind the pronounced increases in employment rates of married women in Continental Europe. Deviations between the developments of employment rates of married women or their husbands' hours worked in data and model do however not affect the good performance of the model in replicating the time-series behavior of married women's hours worked per employed. These model results are robust to exogenously feeding in married women's employment rates and

²⁰In particular, we exogenously feed in education-specific employment rates, but do not let them vary by match as they are allowed to do in the baseline analysis.

married men's hours worked from the data.

We will make the non-linear labor income tax codes used as an input into the analysis available as a user-friendly and easily integrable set of Matlab codes.

References

- ALBANESI, S. AND C. OLIVETTI (2009): "Home Production, Market Production, and the Gender Wage Gap: Incentives and Expectations," *Review of Economic Dynamics*, 12, 80–107.
- (2015): "Gender Roles and Medical Progress," *Journal of Political Economy*.
- ALONSO-ORTIZ, J. (2014): "Social Security and Retirement Across the OECD," *Journal of Economic Dynamics & Control*, 47, 300–316.
- ATTANASIO, O., H. LOW, AND V. SANCHEZ-MARCOS (2008): "Explaining Changes in Female Labor Supply in a Life-Cycle Model," *American Economic Review*, 98, 1517–1542.
- BARRO, R. AND J.-W. LEE (2013): "A New Data Set of Educational Attainment in the World, 1950-2010," *Journal of Development Economics*, 104, 184–198.
- BICK, A., B. BRÜGGEDE, AND N. FUCHS-SCHÜNDELN (2018): "Hours Worked in Europe and the US: New Data, New Answers," *Scandinavian Journal of Economics*, forthcoming.
- BICK, A. AND N. FUCHS-SCHÜNDELN (2017): "Quantifying the Disincentive Effects of Joint Taxation on Married Women's Labor Supply," *American Economic Review Papers & Proceedings*, 107, 100–104.
- (2018): "Taxation and Labor Supply of Married Couples Across Countries: A Macroeconomic Analysis," *Review of Economic Studies*, 85, 1543–1576.
- BLAU, F. D. AND L. M. KAHN (2013): "Female Labor Supply: Why is the US Falling Behind?" *American Economic Review: Papers & Proceedings*, 103, 251–256.
- BLUNDELL, R. AND T. MACURDY (1999): "Labor Supply: A Review of Alternative Approaches," in *Handbook of Labor Economics*, ed. by O. Ashenfelter and D. Card, Elsevier Science B.V., chap. 27, 1559–1695.
- BUERA, F. J., J. P. KABOSKI, AND M. Q. ZHAO (2014): "The Rise of Services: the Role of Skills, Scale, and Female Labor Supply," Working Paper.
- CHAKRABORTY, I., H. A. HOLTER, AND S. STEPANCHUK (2015): "Marriage Stability, Taxation and Aggregate Labor Supply in the US vs. Europe," *Journal of Monetary Economics*, 72, 1–20.
- DOMEIJ, D. AND M. FLODÉN (2006): "The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased," *Review of Economic Dynamics*, 9, 242–262.

- DUVAL-HERNÁNDEZ, R., L. FANG, AND L. R. NGAI (2018): “Social Subsidies and Marketization - the Role of Gender and Skill,” Working paper.
- EROSA, A., L. FUSTER, AND G. KAMBOUROV (2012): “Labor Supply and Government Programs: A Cross-Country Analysis,” *Journal of Monetary Economics*, 59, 84–107.
- GREENWOOD, J., N. GUNER, AND G. VANDENBROUCKE (2017): “Family Economics Writ Large,” *Journal of Economic Literature*, 55, 1346–1434.
- GREENWOOD, J., A. SESHADEVI, AND M. YORUKOGLU (2005): “Engines of Liberation,” *Review of Economic Studies*, 72, 109–133.
- GUNER, N., R. KAYGUSUZ, AND G. VENTURA (2012a): “Taxation and Household Labor Supply,” *Review of Economic Studies*, 79, 1113–1149.
- (2012b): “Taxing Women: A Macroeconomic Analysis,” *Journal of Monetary Economics*, 59, 111–128.
- (2014): “Income Taxation of U.S. Households: Facts and Parametric Estimates,” *Review of Economic Dynamics*, 17, 559–581.
- JONES, L. E., R. E. MANUELLI, AND E. R. MCGRATTAN (2014): “Why Are Married Women Working So Much?” *Journal of Demographic Economics*, 81, 75–114.
- KAYGUSUZ, R. (2010): “Taxes and Female Labor Supply,” *Review of Economic Dynamics*, 13, 725–741.
- KEANE, M. P. (2011): “Labor Supply and Taxes: A Survey,” *Journal of Economic Literature*, 49, 961–1075.
- KNOWLES, J. (2013): “Why are Married Men Working so Much? An Aggregate Analysis of Intra-Household Bargaining and Labour Supply,” *Review of Economic Studies*, 80, 1055–1085.
- MCDANIEL, C. (2011): “Forces Shaping Hours Worked in the OECD, 1960-2004,” *American Economic Journal: Macroeconomics*, 3, 27–52.
- (2012): “Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD 1950-2003,” Working Paper.
- MULLIGAN, C. B. AND Y. RUBINSTEIN (2008): “Selection, Investment, and Women’s Relative Wages over Time,” *Quarterly Journal of Economics*, 123, 1061–1110.
- NGAI, L. R. AND B. PETRONGOLO (2017): “Gender Gaps and the Rise of the Service Economy,” *American Economic Journal: Macroeconomics*, Forthcoming.

- OHANIAN, L., A. RAFFO, AND R. ROGERSON (2008): “Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956-2004,” *Journal of Monetary Economics*, 55, 1353–1362.
- OLIVETTI, C. (2006): “Changes in Women’s Hours of Market Work: The Role of Returns to Experience,” *Review of Economic Dynamics*, 9, 557–587.
- OLIVETTI, C. AND B. PETRONGOLO (2008): “Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps,” *Journal of Labor Economics*, 26, 621–654.
- PREScott, E. C. (2004): “Why Do Americans Work So Much More Than Europeans?” *Federal Reserve Bank of Minneapolis Quarterly Review*, 28, 2–13.
- RAGAN, K. (2013): “Taxes, Transfers and Time Use: Fiscal Policy in a Household Production Model,” *American Economic Journal: Macroeconomics*, 5, 168–192.
- RENDALL, M. (2018): “Female Market Work, Tax Regimes, and the Rise of the Service Sector,” *Review of Economic Dynamics*, 28, 269–289.
- ROGERSON, R. (2008): “Structural Transformation and the Deterioration of European Labor Market Outcomes,” *Journal of Political Economy*, 116, 235–259.
- WALLENIUS, J. (2013): “Social Security and Cross-Country Differences in Hours: A General Equilibrium Analysis,” *Journal of Economic Dynamics and Control*, 37, 2466–2482.

A Appendix

A.1 Non-Linear Labor Income Tax-Codes

We implement the statutory labor income tax codes for the US, Germany, France, the UK, Belgium, Italy, the Netherlands, and Portugal for 1983-2016 in Matlab.²¹ Given the marital status, the number of children in the household, and the combination of labor income of both spouses (if they are married), the codes compute as outputs the amounts of federal and local labor income taxes, employee and employer social security contributions, cash benefits, and the final net-take home pay that a household in a particular country and year faces.²²

These codes are based on the Taxing Wages reports published annually by the OECD.²³ The publications contain country chapters that describe the legal provisions of the tax systems of each country and year in detail. Starting from 1996, the reports also contain a methodological part that presents the parameter values and equations characterizing the statutory labor income tax codes in the respective countries. We use the descriptions (until 1994) and the algorithms (starting from 1996) to code up the tax codes in Matlab. The equations follow a standardized structure: First, they define standard tax allowances. Depending on the country, these can e.g. be basic allowances, allowances related to the family structure (marriage, children etc.) or work-related allowances. Also, social security contributions are deductible in some countries. We subtract the allowances from the gross wage earnings to compute the taxable income. Based on taxable income, the equations specify the amount of the federal government tax liability based on the prevailing tax schedule. We add potential state and local income taxes to the taxes levied by the central government. Next, the algorithm considers tax credits, which may reduce the tax payments depending on the marital status and the number of children or in form of a general basic credit. Finally, the tax equations determine compulsory employee social security contributions that the household pays and cash transfers it receives. Subtracting the tax liability and the employee social security contributions from gross wage earnings and adding cash transfers yields net earnings. In a separate step, the system of equations defines the employers' compulsory social security contributions. The equations consider for every step of the calculations whether it is based on the individual income or the joint income of couples.

The OECD reports adopt some assumptions, which carry over to our codes. We assume that kids are between six and eleven years old (inclusive).²⁴ As income sources of the household we include employment income and cash benefits only. Only standard tax reliefs are taken into account, i.e. only those which accrue independent of the actual expenditure of the household. Examples of standard tax reliefs are reliefs granted based on the marital status or the number of kids in the household. As for the local income taxes, the level of sub-central governments differs between countries. We follow the OECD and consider taxes imposed on the state level, provincial level, and local level. As the local tax system often also varies within the country, the OECD either assumes the tax codes of a typical area (e.g. in the US and Italy) or the average local rates of the whole country (e.g. in Belgium). Only social security contributions to the federal government are considered. The cash benefits consist of payments to families by the federal government (usually for children). All numbers are on an annual basis.

From the year 2001 onwards, the OECD has implemented the tax codes in Stata. This OECD program

²¹ 1995 and 2001 are missing as there is no OECD documentation for those years.

²² So far, employer social security contributions are coded up for the years 1996 to 2016.

²³ The reports are called "Taxing Wages" since 1999. In 1996 to 1998 they were called "The Tax-Benefit Position of Employees" and from 1984 until 1995 they were called "The Tax Benefit Position of Production Workers".

²⁴ If the reliefs or transfers relating to children differ within this age group in a particular year and country, the OECD considers the most generous amounts.

also computes the tax-benefit position of households with precisely specifiable characteristics. In comparison to our codes it allows to vary the age of the kids, considers child care costs, and includes additional cash benefits like unemployment insurance, unemployment assistance, social assistance, and housing benefits. However, not all of these benefits are implemented in the Stata codes in all years and all countries. Although the program is richer in options than ours, the sample period covered is much more restrictive and our program is a step forward in terms of user-friendliness when it comes to incorporating the output produced into an empirical analysis or computation model.

We incorporate the output of our Matlab codes into our quantitative model following [Bick and Fuchs-Schündeln \(2018\)](#). We compute net household earnings for a grid of wives' annual earnings with 201 grid points, ranging from 0 earnings to three times the average annual earnings in the country, and for an earnings grid with 101 grid points for husbands, ranging from 0 earnings to four times the average annual earnings in the country. We then linearly interpolate in two dimensions to assign a net annual household income to each possible annual hours choice of husband and wife. While we do not model children explicitly, we want to take into account that the presence and number of children affect the tax burden of households and that this might have changed over time. We do so in the following way. From the micro data, we calculate the percentage of married couples with 0, 1, 2, 3, or 4+ children conditional on the educational match, and then take the weighted average over these tax burdens for any pair of hours choices.

A.2 Decomposition: Applying Tax Codes from Different Years

The second column of Table 2 shows the correlation between data and model when taxes are allowed to vary over time while gross household income y_{hh} in Equation (4) remains at the 2016 level, i.e.

$$y_{hh} = w_m^{2016}(z)h_m + w_f^{2016}(x)h_f. \quad (\text{A.1})$$

Progressive tax systems are in some way defined relative to the income level at a certain point in time. Simply applying the tax system of 1983 to 2016 would imply that the average household would end up in a range of the tax code featuring a much higher tax rate than the average household in 1983. We account for this in the following way. First, for each combination of husband-wife hours choices, we calculate the tax rate in year t using the gender-specific education premia of 2016 and the year-specific mean wage (\bar{w}^t). Second, we apply this tax rate to the gross earnings of 2016 implied by the same husband-wife hours choices to obtain the household's income tax liability τ_l , and set Equation (5) equal to:

$$\tau_l = y_{hh} \frac{\tau_l^t \left(\frac{w_m^{2016}(z)}{\bar{w}^{2016}} \bar{w}^t h_m, \frac{w_f^{2016}(x)}{\bar{w}^{2016}} \bar{w}^t h_f \right)}{\frac{w_m^{2016}(z)}{\bar{w}^{2016}} \bar{w}^t h_m + \frac{w_f^{2016}(x)}{\bar{w}^{2016}} \bar{w}^t h_f} \quad (\text{A.2})$$

in the household optimization problem. We proceed in a similar fashion when we analyze the effects of year-specific gender-education premia in the fifth column of Table 2. Household income in Equation (4) in this case is replaced by

$$y_{hh} = \frac{w_m^t(z)}{\bar{w}^t} \bar{w}^{2016} h_m + \frac{w_f^t(x)}{\bar{w}^t} \bar{w}^{2016} h_f, \quad (\text{A.3})$$

and the household's income tax liability in Equation (5) by

$$\tau_l = \tau_l^{US} \left(\frac{w_m^t(z)}{\bar{w}^t} \bar{w}^{2016} h_m, \frac{w_f^t(x)}{\bar{w}^t} \bar{w}^{2016} h_f \right). \quad (\text{A.4})$$

Thus, the mean wage remains unchanged, but only the gender-education premia are set year-specific.