Long-term Changes in Married Couples' Labor Supply and Taxes: Evidence from the US and Europe Since the 1980s*

Alexander Bick
Arizona State University

Bettina Brüggemann
McMaster University

Nicola Fuchs-Schündeln
Goethe University Frankfurt, CEPR and CFS

Hannah Paule-Paludkiewicz
Goethe University Frankfurt

June 8, 2018

PRELIMINARY

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, 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, with labor income taxes being the key driving force. It also explains part of the secular increase in married women's employment rates, but the large increases among European married women in the 1980s and early 1990s are not driven by the factors considered in our study. 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.

*Email: alexander.bick@asu.edu, brueggeb@mcmaster.ca, fuchs@wiwi.uni-frankfurt.de, and hannah.paule@hof.uni-frankfurt.de.

We thank Domenico Ferraro and Gustavo Ventura, as well as seminar and conference participants at Georgetown University, Virginia Commonwealth University, the Federal Reserve Bank of Richmond, and the Carleton Macro-Finance Workshop for their helpful comments and suggestions. Enida Bajgoric, Pavlin Tomov, and Mariia Bondar provided excellent research assistantship. All errors are ours.

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 (2017a) 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 while also investigating the time-series, 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. Albanesi and Olivetti (2009), Albanesi and Olivetti (2015), Attanasio et al. (2008), Buera et al. (2014), Greenwood et al. (2005), Jones et al. (2014), Knowles (2013), Ngai and Petrongolo (2017), and Olivetti (2006)). Among those, Kaygusuz (2010) and Rendall (2015) focus on the role of non-linear income taxation.

We contribute to these literatures by investigating the driving forces behind the time-series of hours worked by married couples in the US and seven European countries from the early 1980s to 2016. The seven European countries are the four largest economies in Europe (Germany, the UK, France, and Italy), plus three other countries for which we have data on labor supply from the 1980s onwards (Belgium, the Netherlands, and Portugal).[1] Based on the data set of internationally comparable employment rates and hours worked measures for demographic subgroups introduced in Bick et al. (2018), we first document the development of employment rates and hours worked conditional on employment of married women and married men over time in the sample countries. Relying on the model developed in Kaygusuz (2010), Guner et al. (2012a, 2012b), and Bick and Fuchs-Schündeln (2017a), we then quantitatively analyze the role of non-linear labor income taxes and consumption taxes, as well as of the gender and educational wage premia, the educational distribution and the degree of assortative matching into couples, for explaining the evolution of hours worked of married couples over time. Especially, we implement the country- and year-specific full non-linear statutory labor income tax codes, including the tax treatment of married couples, for all sample countries for a much longer time period than currently available. As an independent contribution, we will publish these as a user-friendly and easily integrable set of Matlab codes. In our quantitative exercise, 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

[1]Data on labor supply from the 1980s on are also available for Denmark, Ireland, Spain, and Switzerland, but we have not coded up the tax codes for these countries.

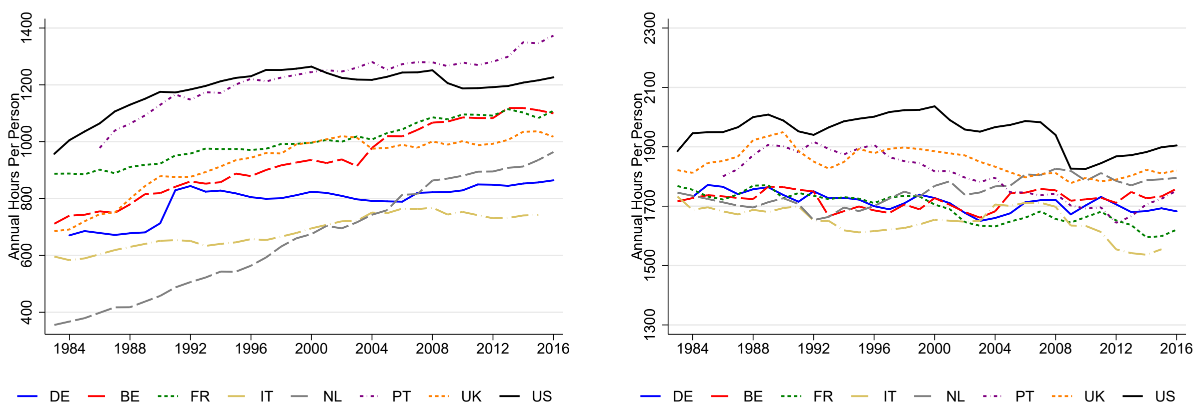
the US and the seven European countries for the years 1983 to 2016, holding preferences fixed in each country but using the country-year-specific economic environment. Last and focusing on married women in particular, we conduct a decomposition analysis in order to determine the relative roles of labor income and consumption taxes, as well as wages and the educational composition, as driving forces of labor supply of married couples.

The focus of our analysis on married couples is motivated by several facts: first, among the population aged 25 to 54, married couples are the largest group and 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 mens' 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. Bick and Fuchs-Schündeln (2017a) demonstrate that for understanding the cross-country differences in married couples' hours worked, it is important to take into account non-linear labor income taxation and the tax treatment of married couples. 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 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).

The secular increase in hours worked of married women is largely driven by the extensive margin. On average, employment rates increased by 23 percentage points between the early eighties and 2016. In addition, there was substantial convergence of employment rates, which 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, exhibits much more variation in levels as well as trends and shows no convergence: average hours vary between 1000 and 1800 hours throughout the whole sample period. The model is quite successful in predicting the time-series behavior of hours worked per employed married woman: it explains on average 115 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 86 percent of the long-term changes in female hours. The model also explains part of the secular increase in married women's employment rates, but fails to replicate the large increases in employment rates that especially European women experienced in the 1980s and early 1990s. Exceptions are the US and the UK, the two

Figure 1: Labor Supply of Married Couples

(a) Married Women: Hours Worked per Person (b) Married Men: Hours Worked per Person



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.

countries with the smallest increases, where the model performs much better with respect to the evolution of employment rates. This implies that the sharp increase in employment rates in most European countries is explained by factors outside of the model, and not due to changes in labor income and consumption taxes, or wages and the educational composition. Hours worked per employed married man vary between 1800 and 2100 hours within our sample and hardly change over time (average decrease is 33 hours). The model correctly predicts such small changes in 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 other countries (Belgium, 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, the decomposition analysis, and some robustness checks, before Section 5 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 lies in 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 latest 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 omit the younger and older work force since we do not model differences and changes in education systems or social security systems.

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 hours worked per person for married women across these countries, with average annual growth rates of 1.3 percent per year between 1983 and 2016. The Netherlands feature the largest overall increase with an average annual growth rate of 3.3 percent. At the same time, married men's hours worked were largely flat or even slightly falling: the average annual growth rate across countries amounts to -0.06 percent over the same time period. Figure A.1 in the Appendix shows that neither of the two margins of labor supply exhibits major trends in the data for married men.

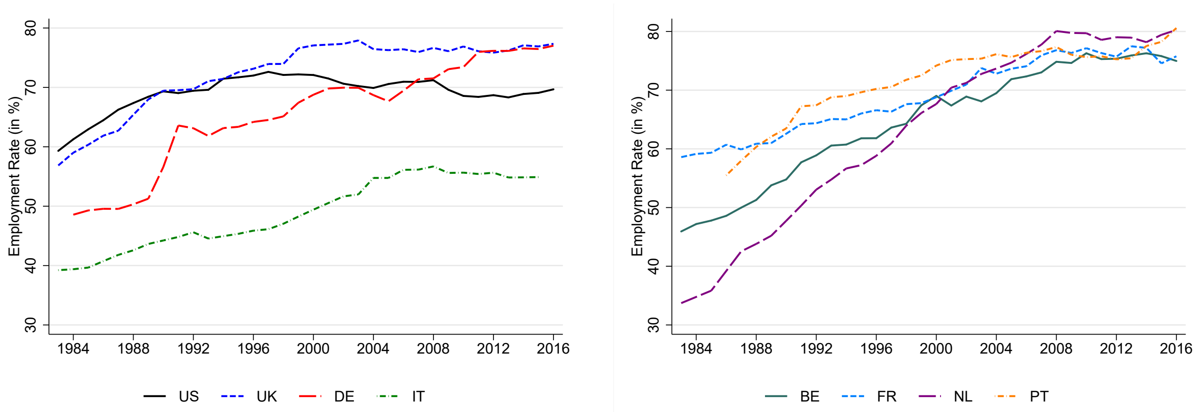
In Figure 2, we split the time series of average hours worked for married women into the extensive (Panel a) and intensive margin (Panel b). As we discuss in further detail below, we group countries according to the different patterns observed in the time trends of married women's hours worked per employed. The patterns that emerge for the two margins are quite interesting: first, it becomes clear that the overall increase in average hours per person is largely driven by a large increase in employment rates among married women. 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, suggesting that the rise in married women's employment rates started earlier in the US and was already leveling out in the early nineties, while the European countries were still catching up to (and eventually overtaking) the US. In 2016, the employment rate among married women of the US 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 differing trends across countries, as one can see in Panel (b) of Figure 2. Of the countries in the left graph of Panel (b), the US and the UK experience clear positive trends in hours worked per employed between the early 1980s and 2016, growing by on average 0.3 percent per year. The other two countries, Germany and Italy, experienced opposite trends with hours worked per employed married woman decreasing by on average -0.5 percent annually.[2] For the countries in the graph on the right side of Panel (b), Belgium, France, the Netherlands, and Portugal, a third pattern

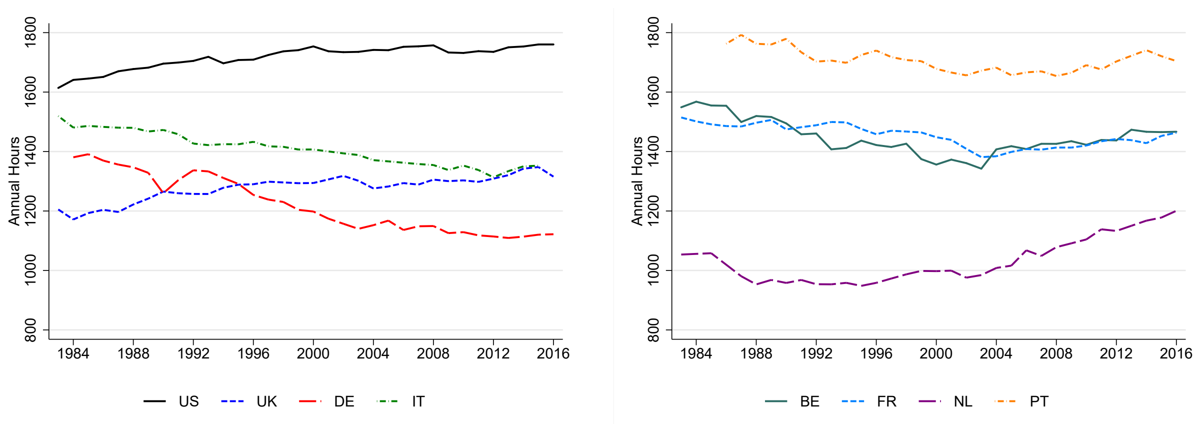
[2]The German reunification in 1990 leads to a change in the underlying population, with East Germans entering the sample. This explains the large movements in the employment rate and hours worked of married women in the German data around this time.

Figure 2: Labor Supply of Married Women

(a) Employment Rate



(b) Hours Worked per Employed



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

emerges. In these countries, hours worked per employed married woman first decreased before the trend switched sign and turned positive in later years. Unlike what we observe for employment rates, there is no convergence in married women's hours worked per employed: 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. The direction of the trends in employment rates and hours worked are robust across different age groups and family constellations as we show in Appendix Figure A.2. The goal of this paper is to shed light on the mechanisms underlying these differing trends across countries using the model we describe in the next section.

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 time. The model framework is based on Kaygusuz (2010), and Guner et al. (2012a, 2012b), and Bick and Fuchs-Schündeln (2017a), and features a maximization problem of a two person household which optimally determines male and female labor supply.

The description of the model follows closely Bick and Fuchs-Schündeln (2017a). 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

$$\sum_x \sum_z \mu(x,z) = 1. \tag{1}$$

Households draw a utility cost of joint work q from a distribution $\zeta(q|z)$ which depends on the husband's education level. 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, they always optimally choose to provide positive hours. We follow this approach because employment rates of married men aged 25 to 54 are 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 where the husband is working. The estimated hours worked and employment rates of married women are on average a little higher in this restricted sample, 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 - conditional on the husband's education level - 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 (2017a), 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, as well as the number of children in the household k through tax credits and/or child benefits. The maximization problem of a type $\{x,z\}$ household is given by

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

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

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

$$\tau_l=\tau_l(w_m(z)h_m,w_f(x)h_f,k) \tag{5}$$

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.

T represents a lump-sum transfer from the government, which redistributes a share $\lambda\in[0,1]$ of all government revenues:

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

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.

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. α_m and α_f capture the gender-specific relative weight 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.

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

The OECD documents the country-year specific statutory labor income tax codes in the so-called "Taxing Wages" reports starting in the early 1980s.[3] In these reports, the OECD describes all relevant details to cal-

[3] The "Taxing Wages" reports bear this title since 1999, before they were known as "The tax/benefit position of employees" (from 1996 to 1998) and "The tax benefit position of production workers" (from 1984 to 1995).

culate the various labor income taxes, contributions, and benefits that taxpayers face in a particular country and year, including federal, state, and local taxes, employees' social security contributions, cash benefits, and standard deductions. 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.[4] 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 the household receives (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.2.1.

For the US, we can compare the average statutory tax rates obtained from our tax codes to the estimates of actual average tax rates in Guner et al. (2014). Using a representative subsample of US taxpayers in 2000, they report effective average tax rates and average household incomes by household type, i.e. marital status and number of children, and income quantiles. The main difference is that these effective average tax rates capture all types of income taxation instead of only labor income taxation and account for itemized instead of only standard deductions. Using the reported average income by income quintile from Guner et al. (2014), we calculate average statutory tax rates using our tax codes and compare them to their estimates of effective tax rates for different household types. The two tax rates are generally close, but differ substantially for the top quintile, see Online Appendix Section A.2.2. This is not surprising since capital income and itemized deductions play a larger role for higher incomes.

It is worthwhile to point out that for the US our tax codes are a simplified version of the NBER TAXSIM program, which recreates each year's federal income tax law since 1960 and state laws since 1977 and is widely used in the literature. Specifically, the NBER TAXSIM captures all types of income taxation and accounts for itemized deductions. 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 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 that reflect key features of the labor income tax code across countries and over time for all countries in our sample, focusing on the tax burden of married couples calculated using the OECD statutory tax codes. The solid line with circles as data points 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

[4] Note that we currently exclude the years 1994, 1995, and 2001. For 1994, we simply have not yet translated the documentation into Matlab codes. For 1995 and 2001, we do not have information from the OECD.

average wage of married men prevailing in the respective year and country (we call this tax rate "average tax rate"). The dashed line with triangular data points shows the average marginal tax rate that a wife faces if she goes from not working to working the average hours worked of married women in the respective country in 2016 (called "average marginal tax rate"). We assume that she earns the average wage of married women prevailing in the respective year and 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.[5]

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 rates 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.

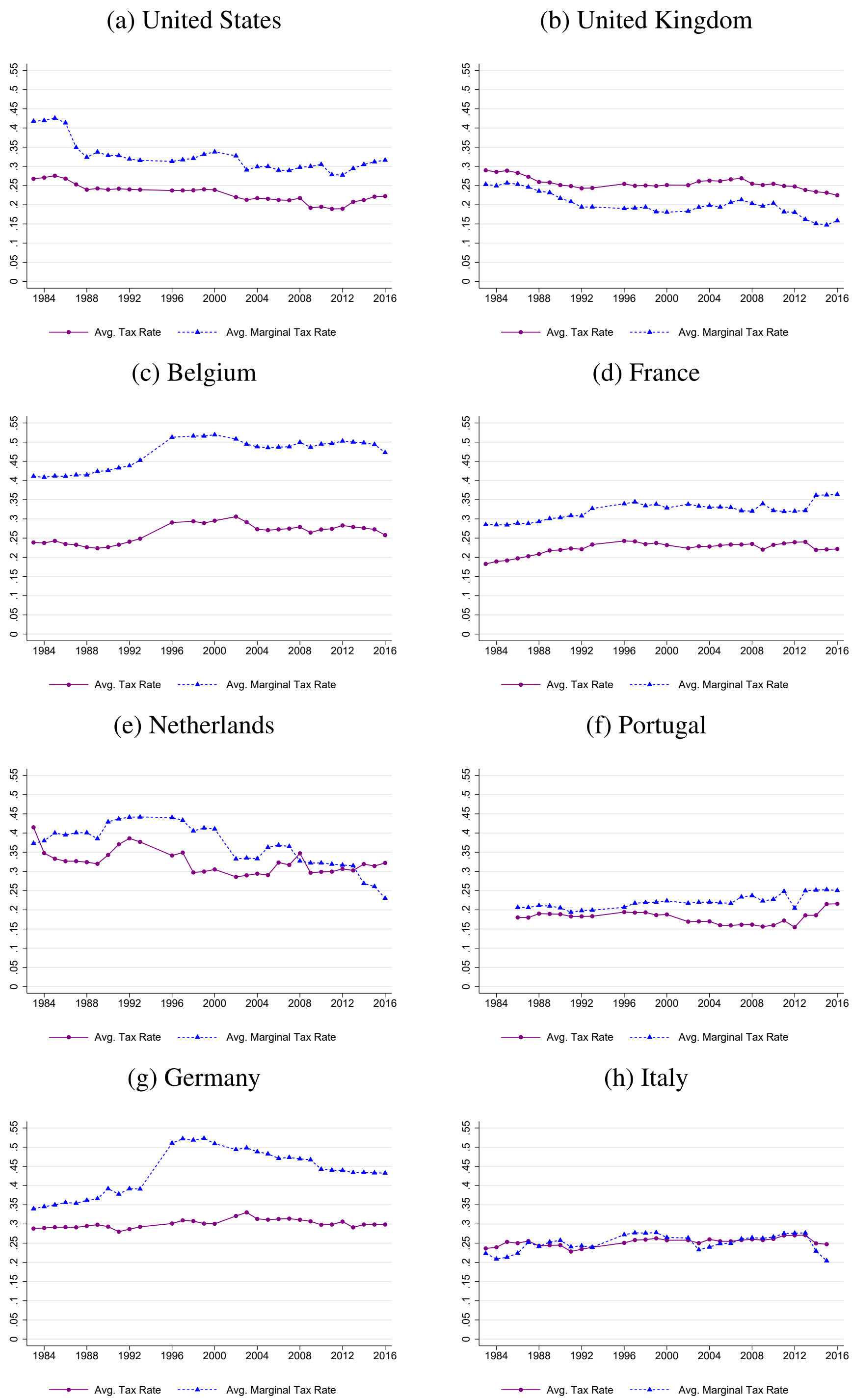
The gap between the average tax rate and the average marginal tax rate provides insights into the tax treatment of married couples used in a country. In countries with strictly separate taxation, each spouse is taxed as if he or she was single. In these cases, average marginal tax rates of second-earners 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.[6] When couples are taxed jointly, each spouse's taxes depend not only on the own earnings, but also 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, the remaining countries feature this pattern of tax rates, but to a differing degree. The gaps between the two suggestive tax rates are 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 (2017a) and Bick and Fuchs-Schündeln (2017b).

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. To name one example, the 1986 tax reform in the US, which lowered the progressivity of income taxation, is apparent for both tax rates, but has a larger effect 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, saw large increases in the mid 1990s especially in the average marginal tax rates, while the Netherlands saw several decreases in this tax rate throughout the late 1990s and early 2000s. As Figure 3 shows, sometimes tax reforms have

[5] This is true for all countries except Italy, where the latest available year in the EU-LFS, which we use to calculate average hours worked, is 2015.

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

Figure 3: Illustrative Tax Rates: $\tau_l(0)$ and $\tau_l'(h_f^{US})$



Note: $\tau_l(0)$ is the 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 year-specific mean married male wage and the wife does not work. $\tau_l'(h_f^{US})$ 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 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 1994-1995 and 2001 from the graphs because we have not yet up written up the tax code (1994) or we do not have information from the OECD (1995 and 2001).

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), who calculates consumption tax rates from NIPA data for almost our entire sample period. For all countries, the value for 2016 is missing and we extrapolate the time series using the consumption tax from 2015. For Portugal, the earliest available consumption tax rate is from 1995, and we extrapolate the series backward to 1983 using that earliest available consumption tax rate. The advantage of these tax rates over simple value added tax rates is that they also capture excise taxes, exemptions from the value added tax, etc.

Panels (a) to (h) of Figure A.3 show that consumption tax rates have been fairly stable for the US and the UK. France and Belgium experienced increases in the consumption tax rate in the 1990s, but they were subsequently lowered in the late 1990s and during the Great Recession. The decrease during the recession is also visible for the Netherlands, Portugal, and Italy, but their consumption tax rates have otherwise generally been increasing since the 1980s. This is also true for Germany, where consumption tax rates have been steadily increasing since the 1980s.

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 our data set described in Bick et al. (2018), relying on the three education groups low, medium, and high. 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 the European countries in our sample except Germany. Concretely, we use the information on education shares in the Barro-Lee Educational Attainment Data (Barro and Lee, 2013) to extrapolate the time series for the education and matching shares backwards until 1983. The Barro-Lee Educational Attainment Data is available by gender and age groups in 5-year intervals from as early as 1950. We first interpolate the data to account for missing years. Then, we regress the educational shares of married men and women aged 25 to 54 until the year 2000 on the Barro-Lee educational shares for each age group between ages 25 to 54 (25-29, 30-34, 35-39,...).[7] For Germany, we use the Socio-Economic Panel that allows us to calculate educational 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 Appendix Section A.2.3.

Figure A.4 documents the massive shifts in the educational composition among married women aged 25 to 54 in our sample countries. The share of highly educated married women increased in all countries, while the share of low-educated women fell. Only in the US exceeds the share of high-educated women the share of low-educated women for all years of our sample, whereas in Europe the number of highly-educated women started to exceed the number of low-educated ones only in the late 1990s to early 2000s. Portugal

[7] We stop in 2000 because including all years available in the Barro-Lee data (available until 2010) considerably worsened the overall explanatory power of the regression used to predict educational shares in the 80s and early 90s.

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, as Figure A.5 shows.

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 provides suitable earnings data to calculate/estimate hourly wages for married men and women, unfortunately, the EU-LFS does not provide any earnings data for the European countries. We therefore rely on a number of other datasets to get reliable estimates. For Germany, we 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. This European household data set captures income and usual hours, but features a sample size an order of magnitude smaller than the EU-LFS. From 1994 to 2001, we use the European Community Household Panel (ECHP), the EU-SILC's predecessor. We use linear interpolation to obtain gender- and education-specific wages for the missing years between the end of the ECHP and the beginning of the EU-SILC. For the remaining years prior to 1994, we have to rely entirely on estimations to impute gender- and education-specific wages. To do that, we first calculate average hourly wages on the aggregate level using a consistent earnings series that is available for our whole time series. The only aggregate earnings series that fulfills these requirements are the average annual wages of production workers that the OECD publishes along with their tax documentation described in Section 3.2.1. Using usual hours of full-time employees from Bick et al. (2018), we first transform those annual wages into a measure of average hourly wages of production workers. For each country, we then regress the gender-education-specific wages from the microdata 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. More details on our imputation strategy can be found in Appendix Section A.2.4. In a robustness check in Appendix Section A.4.3 we try to get a sense of the impact of these wage imputations by comparing the model outcomes when using actual vs. imputed wages for the US, where we have the full time series of wages available.

Figure A.6 documents the evolution of the gender wage gap for the countries in our sample after all imputations have occurred. In the majority of countries the wage gap has narrowed over time, with the largest increase 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 (the dashed line shows the value of the wage gap in 2016). But also the UK, Belgium, France, the Netherlands, and Italy have experienced a narrowing of the gender wage gap, albeit to a much smaller degree especially in the latter three countries. Notable exceptions are Portugal and Germany: In Portugal, female wages relative to male wages have slightly decreased over time, and in Germany the wage gap abruptly narrowed as a consequence of the re-unification but has since widened to ratios very similar to those 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 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 Appendix Figure A.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.

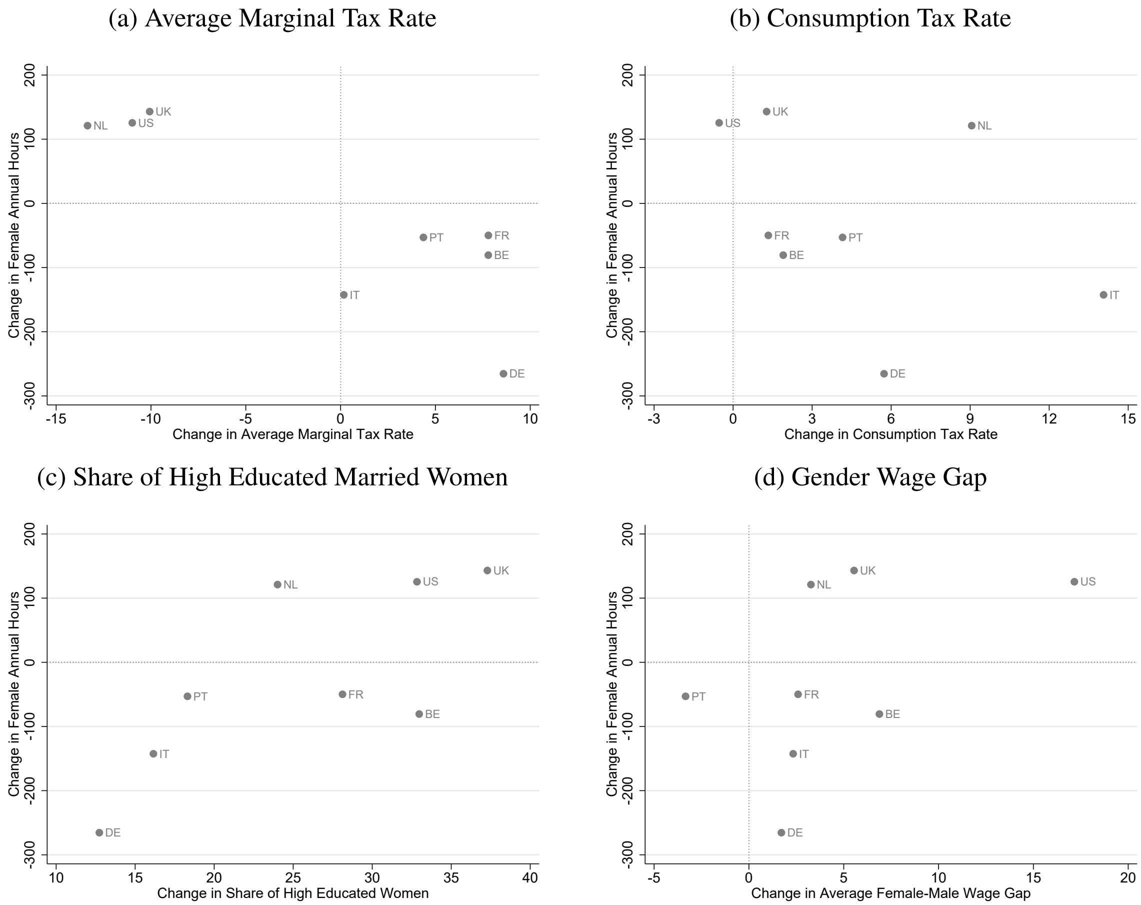
Panel (a) of Figure 4 shows the strongest empirical relationship of the four: changes in average marginal tax rates exhibit a strongly negative correlation (correlation coefficient of -0.85) with changes in female hours worked. Countries with the largest decreases in the average marginal tax rate see 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 taxes and labor supply is much weaker and the coefficient of -0.37 is not significant. As one would expect, increases in the share of highly educated married women and the female-male wage ratio are positively correlated with changes in their hours worked, but the correlation coefficients of 0.74 and 0.49 are lower and less significant than the one for marginal tax rates. 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 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 them) and hours worked per employed woman (α_f) .

Again following Kaygusuz (2010) and Guner et al. (2012a, 2012b), the utility cost parameter is distributed according to a flexible gamma distribution, with the shape parameter k_z and scale parameter θ_z

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



being conditional on the husband's type:

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

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 2016). For given preference parameters α and ϕ, and conditional on being married to a type z husband, the three different education levels x and implied wages generate three different threshold levels $\bar{q}(x,z)$ at which a woman of type x 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.[8]

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 4.3.1 we report the results when we assume that only half or none of the government revenues are redistributed.

The model fit for targeted and untargeted moments for each of our sample countries is summarized in Section A.3 of the Appendix. This section also documents the calibrated 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 by gender and the employment rate of married women with the data. In a decomposition analysis, 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 in its ability to replicate the time series behavior of hours worked per employed and employment rates from two perspectives. First, we analyze how well the model captures the full time-series variation of each variable in the data. This is especially interesting for the countries which exhibit a shift in the hours trends over time. Second, we compare the changes between 1983-85 and 2014-16 in the data and the model to assess the model's performance with respect to replicating long-term changes in married women's labor supply.

Table 1 reports the time-series correlation between data and model output 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). We split the sample of countries into three groups, following the different patterns in the time trends of married women's annual hours that we documented in Figure 2. In the US and UK trends in married women's average annual hours worked per employed are positive, they change from

[8] 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.

Table 1: Correlation between Data and Model

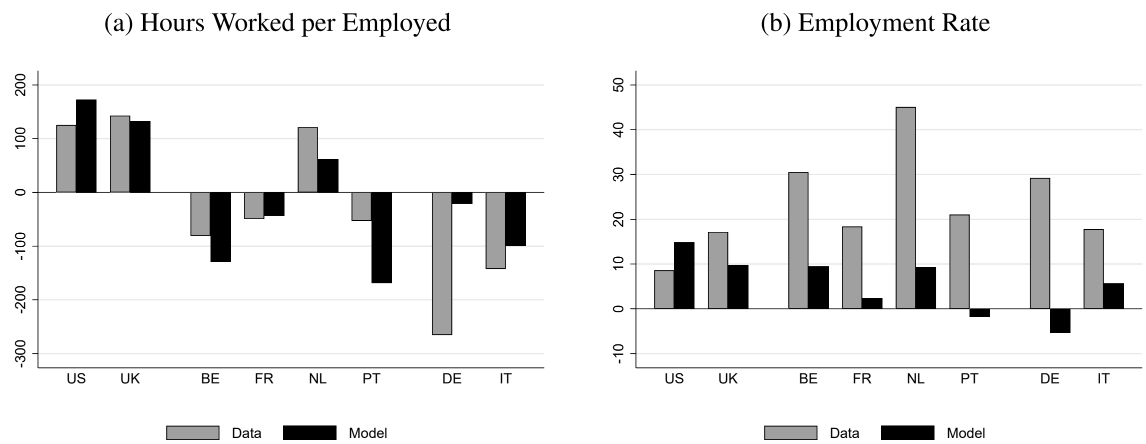
Country	Female HWE	Female ER	Male HWE
Positive Hours Trend			
United States	0.92	0.79	0.46
United Kingdom	0.91	0.81	0.88
Changing Hours Trend			
Belgium	0.66	0.64	−0.21
France	0.88	0.66	0.55
Netherlands	0.79	0.65	−0.39
Portugal	0.33	−0.77	−0.18
Negative Hours Trend			
Germany	0.64	−0.69	−0.52
Italy	0.81	0.34	−0.26

negative to positive in Belgium, France, the Netherlands, and Portugal, and are negative in Germany and Italy. The model does a good job capturing the time-series variation in female hours worked, especially in the United States, the United Kingdom, and France with correlation coefficients of or above 0.88. Also for Italy, the Netherlands, Belgium, and Germany the correlation coefficient between model and data hours is high, with values between 0.81 and 0.64. Only for Portugal the correlation between hours in data and model amounts only to 0.33. It is quite remarkable that the model can largely generate the patterns in married women's hours worked per employed in all eight sample countries, as Appendix Figure A.9 shows, despite the quite different evolution of these hours across countries.

For the US, the UK, France, Belgium, and the Netherlands the correlation between employment rates in model and data is relatively high with values between 0.64 and 0.81. For Italy, the correlation is much lower with a value of only 0.34. The model is not successful at all in replicating the behavior of married women's employment rates in Portugal and Germany: there, the correlation coefficients are negative with a large absolute value. The model does worst in capturing the time-series variation in male hours worked. Only three countries show a positive correlation coefficient between model and data, while five countries exhibit a negative correlation. However, male hours worked changes in both model and data are relatively small. This suggests that the model can provide most insight into the labor supply behavior of married women, on which we focus from here on. We provide details on our findings for married men in the Appendix.

Figure 5 compares the changes between 1983-1985 (1986-88 in Portugal) and 2014-16 (2014-15 in Italy) for the two margins of labor supply of married women. The two figures confirm the pattern that emerged in Table 1, namely that the model captures the evolution of hours worked per employed better than that of

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



employment rates. Panel (a) of Figure 5 shows that the model predicts changes in hours worked that are roughly in line with what we see in the data for most countries. Exceptions are Portugal and Germany, but ignoring the relatively poor performance of the model in explaining long-term trends for these two countries, the model accounts for 50 to 160 percent of hours worked changes for the other countries.

Panel (b) of Figure 5 shows that while the model also captures the (relatively moderate) increases in the employment rates of married women in the US and UK, it is not successful in producing the large increases in employment rates that European married women experienced in the eighties and early nineties. Germany and Portugal again stand out in that the model completely fails to predict any increase in employment rates, but even among the other European countries, it only accounts for 13 to 57 percent of the "secular" increase observed in the data. This suggests that the sharp increase in employment rates in the European countries is explained by factors outside of our model.[9] Candidates for such outside factors would be anything that affects the fixed cost q associated with women working in the model. Higher fixed costs in the 80s would imply lower employment rates. When currently calibrating the fixed costs to the level of employment in 2016, the model does not generate sufficiently large changes in employment rates over time, with the exception of the US. One can think of several explanations for such higher fixed costs in the 1980s (e.g., general attitudes towards working women, lower availability of child care, etc.), but quantifying them in detail is beyond the scope of this paper.

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, but the drastic increases in employment rates among married women are not due to the same factors, but can only be explained by

[9] The full time-series predictions for the employment rates of married women can be found in Appendix Figure A.10.

factors outside of our model.[10]

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 eighties 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. Since our model performs better in replicating the hours worked of married women than their employment rates, we largely focus on hours worked in this analysis and only briefly summarize the decomposition results for the evolution of the employment rates at the end of the section.

4.2.1 Non-Linear Labor Income Taxes

First and foremost we are interested in the role that the non-linear labor income taxes play for the evolution of married women's labor supply and especially their hours worked per employed in each of our sample countries. Figure 6 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 grey line).

The small distance between the black (full model results) and blue (only time-variation in labor income taxes) lines for all eight countries in Figure 6 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, changing taxes account for a significant part of that increase. In the US, the tax experiment alone accounts for 54 percent of the total increase between 1983 and 2016 (the full model accounts for 167 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 (remember Panel (a) in Figure 3). In the UK, changing only taxes over time generates 20 percent of the actual increase in hours worked by married women.

As noted earlier, the model generated hours per employed married woman 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 6 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 generate flat hours as in the data for a long time, and an increase towards the end of the sample period. For Portugal, where we also see changing trends in the data, panel (f) in Figure

[10] In Bick and Fuchs-Schündeln (2017a), 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 find that the model predicts differences in hours worked per employed married woman better than differences in employment rates.

Figure 6: Model output when only varying tax code

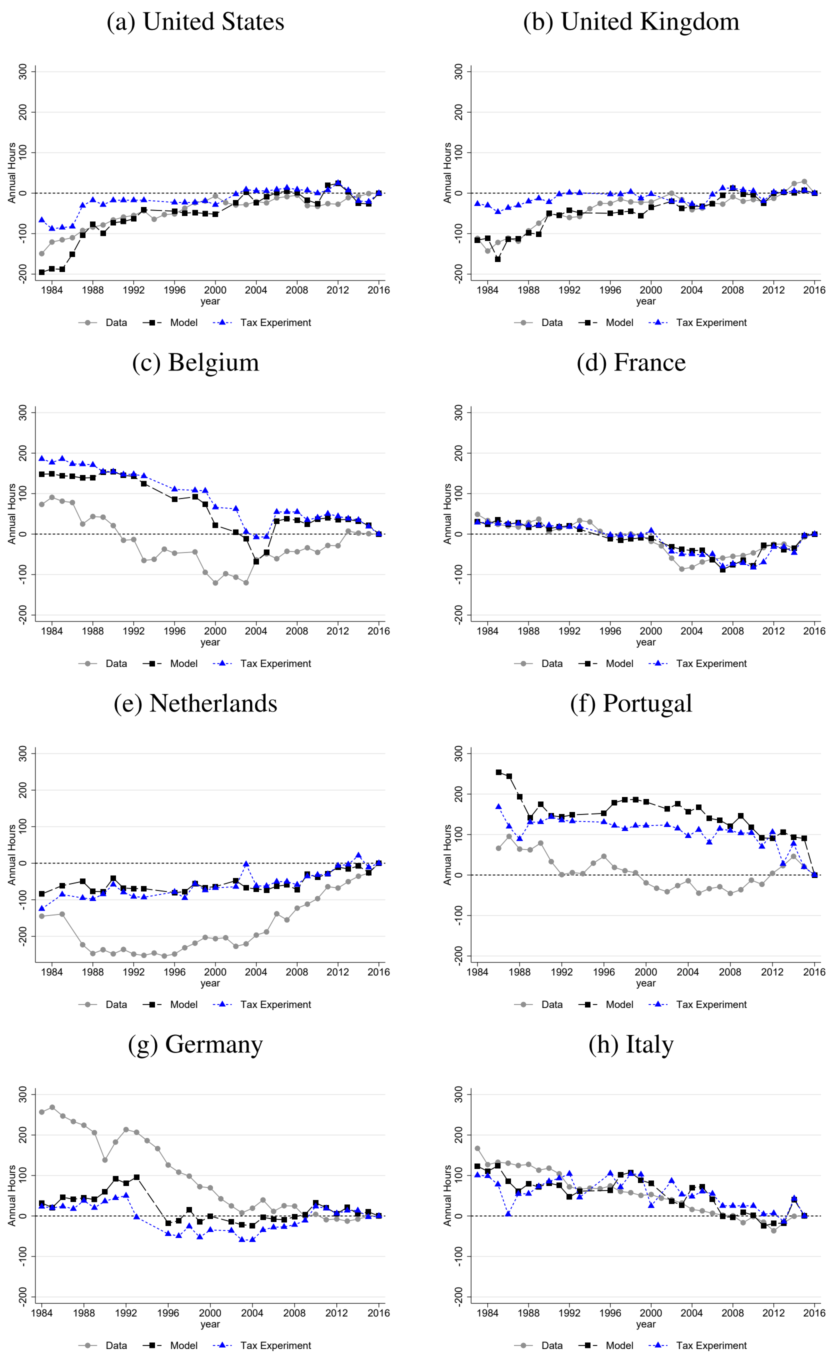


Table 2: Correlation between Data and Decomposition Output

Country	Total	Tax Code	Cons. Tax	Educ.	Wages
Positive Hours Trend					
United States	0.92	0.79	0.28	0.46	0.85
United Kingdom	0.91	0.68	−0.16	0.80	0.07
Changing Hours Trend					
Belgium	0.66	0.57	0.83	−0.42	0.46
France	0.88	0.88	0.07	−0.36	−0.67
Netherlands	0.79	0.72	−0.67	−0.87	0.32
Portugal	0.33	0.16	0.23	−0.23	0.20
Negative Hours Trend					
Germany	0.64	0.45	0.84	0.09	0.30
Italy	0.81	0.59	0.84	0.53	0.29

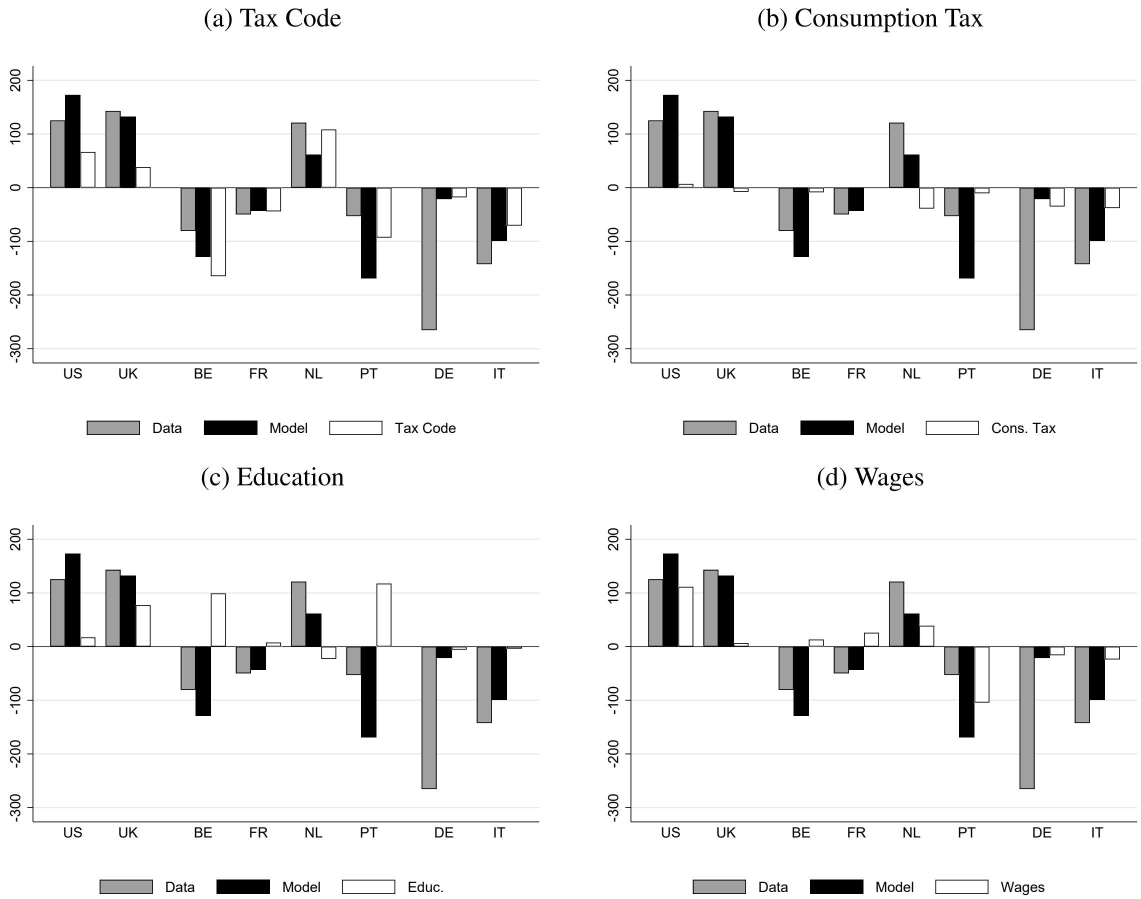
6 shows that the model is not able to predict that pattern, but instead predicts a steady decrease which is largely generated by changes in taxes.

In Germany and Italy, married women's hours worked per employed have been steadily decreasing since the early eighties. For Italy, the model is successful in generating about 69 percent of that decrease, and again, the tax experiment shows that taxes plays a big role: letting only taxes vary can explain 50 percent of the decrease observed in the data (see Panel (h) of Figure 6). The model is not able to replicate the large decrease in hours observed in Germany, and changes in the tax code especially seem to imply a much more subdued evolution of hours worked by married women than what actually happened (Panel (g) in Figure 6).

Table 2 and Figure 7 show two ways of summarizing the results across countries. 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 is replicated by the model. The second column of Table 2 reports the correlation between data and model when only taxes vary 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, all of the time-series variation captured by the full model can be generated by only letting taxes vary over time.

Figure 7 shows the decomposition of the long-term trends. Panel (a) provides further insight into the role of taxes for married women's hours worked. 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

Figure 7: Changes in Married Women's Hours per Employed between 1983-85 and 2014-16: Decomposition



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 toward explaining the long-term changes observed in the data. For France and the Netherlands the tax experiment is most successful and the tax changes can explain about 90 percent of the overall change in hours worked. The tax experiment performs worst for Germany, where changes in the tax rates can only account for 7 percent of the change observed in the data.

4.2.2 Consumption Taxes

Of the two tax inputs that we feed into the model, consumption taxes have the lesser predictive power than labor income taxes. The third column in Table 2 shows mixed results from an experiment where all inputs are kept at their 2016 level and only consumption taxes are allowed to vary over time. 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 is either hardly correlated with the data (United States, France, Portugal) or even negatively correlated (United Kingdom, Netherlands).

These mixed results are echoed by the bar graph in Panel (b) of Figure 7, which plots the changes in married women's hours worked between 1983-85 and 2014-16 in the data and full model against changes generated by the consumption tax experiment. The extent to which changes in the consumption tax rate affect changes in hours worked seems to be fairly limited, with the maximum change in hours generated by the experiment amounting to only -40 hours in the Netherlands. The small effects that changes in consumption taxes have on female hours worked is also evident when looking at the whole time series in Figure A.12.

4.2.3 Educational Composition and Matching into Couples

As shown in Figure 4 and Figure A.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 and matching into couples while holding all other inputs constant at their 2016 level, the model does not replicate the time-series behavior of hours well, as can be seen in column 4 of Table 2. Among the four countries with changing hours trends, the changes in the educational composition and matching into couples predict changes in hours worked that are opposite to the actual changes in the data. As Panel (c) of Figure 7 shows, for some countries the education experiment predicts a slight decrease in hours worked of 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 Section 4.2.5). As a consequence, government revenues and thus transfers increase, which has a countervailing effect on the intensive margin. For some countries, this effect dominates. Through the lens of the model, the only country where education and matching patterns strongly contribute toward explaining married women's hours worked is the UK. Here, the education experiment generates an increase of 77 hours compared to 143 hours in the data, which can be seen in Panel (c) of Figure 7 as well as Figure A.13. Similar to what we saw in the consumption tax experiment, changes in the distribution over educational levels and matching predict smaller hours changes than in the data.

4.2.4 Gender- and Education-Specific Wages

In the last counterfactual experiment, we let wages by gender and education vary over time while keeping the other three inputs at their 2016 level. Just like for consumption and hours worked, the time-series correlation between hours worked in the data and as generated by the wage experiment delivers mixed results, as column 5 in Table 2 shows. For the US, wages play a large role in replicating the evolution of hours worked over time, larger even 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 125 hours in the data. This large impact of the change in wages on the hours worked by married women is not entirely surprising since the US is also the country with the largest increase in the ratio of female over male wages (17 percentage points), as displayed in both Panel (d) of Figure 7 and Figure A.6.

The remaining countries in our sample feature much more modest increases in the gender wage ratio or even decreases, and except for the US and Portugal, the predicted changes in the wage experiment are small as well. France stands out in Figure 7 (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 four out of the eight countries in the sample.

4.2.5 The Role of Inputs for Female Employment Rates

Overall, the variation in labor income and consumption taxes, educational composition, and wages explain on average 115 percent of the changes in hours worked per employed married woman between 1983 and 2016. The model is less successful in replicating the secular increase in married women's employment rates that we observe over the same time period, as shown in Figure 5: across countries, it explains on average only 37 percent of the increase. The decomposition results for the female employment rates are shown in Table A.17 and Figure A.15. Figure A.15 reveals that the small predicted changes in employment rates are due to all input factors indicating changes that are small compared to the data, rather than the input factors pointing in different directions. Moreover, as Table A.17 shows, the only input experiment that consistently positively correlates with the time-series of married women's employment rates are the educational shares, with correlation coefficients between 0.63 and 0.98 in all countries. The increase in the share of high educated women consistently predicts an increase in employment rates, as observed in the data.

4.3 Robustness Checks

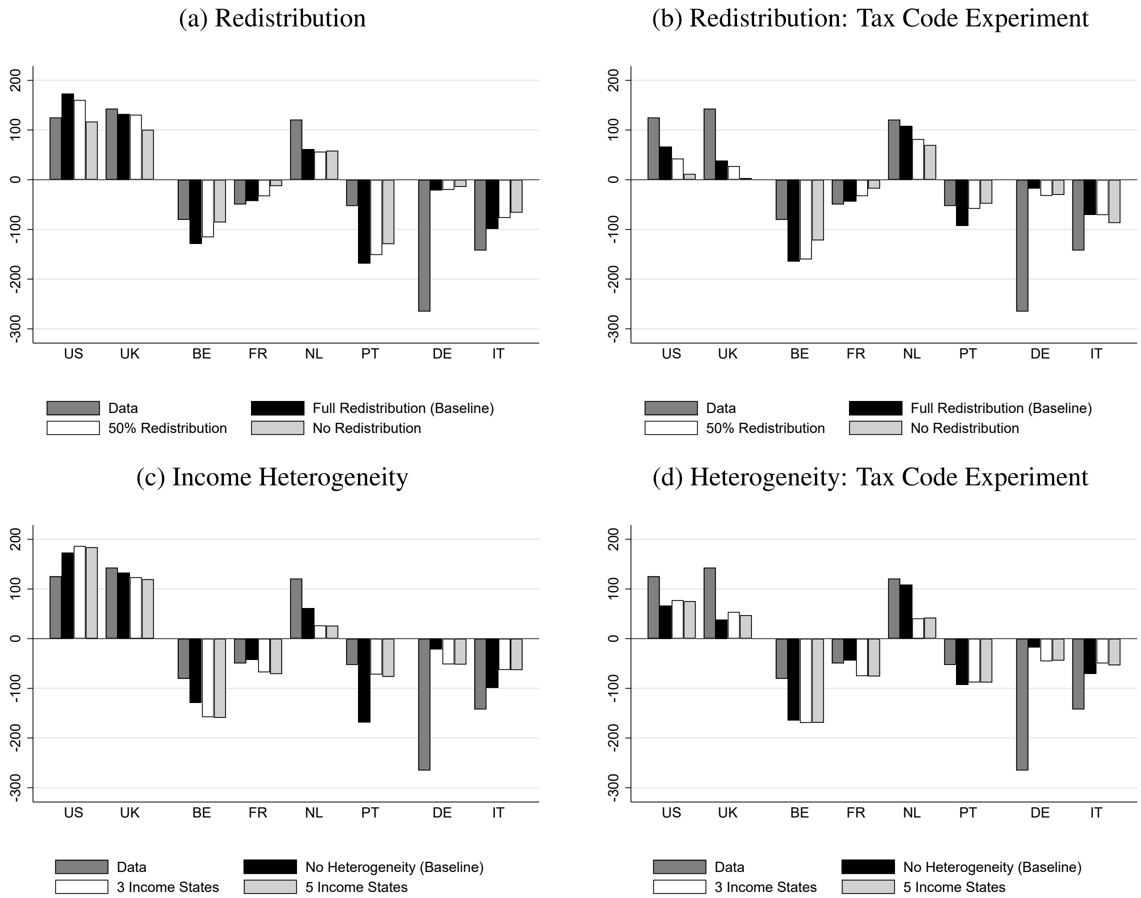
4.3.1 Redistribution of Tax Revenues

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. In this section, we explore by how much the model predictions change if we relax that assumption. Specifically, we analyze the cases of 50% redistribution ($\lambda = 0.5$) and no redistribution (that is, $\lambda = T = 0$). In each case, we recalibrate the model for each country.

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. Redistribution leads to an additional income effect of higher taxes.

Panel (a) of Figure 8 shows the model predictions for changes in hours worked per employed married

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



woman for these two alternative specifications of λ (white and light gray bars) and compares them to the baseline model with full redistribution (black bars) and the data (dark gray bars). When lowering the degree of redistribution, the model generally predicts lower changes in hours worked than with full redistribution. Via the income effect, the lump-sum transfer through which the government redistributes constitutes a disincentive to working. With less redistribution, the income effect connected to both increases and decreases in taxes is smaller. For countries where taxes decreased, like the US, the UK, or the Netherlands, this means that the model with less redistribution implies higher hours worked in the 1980s (and thus a smaller increase in hours over time) because the disincentive effect from higher taxes in the 1980s is smaller. For countries where taxes increased (that is, all other countries in the sample), the lower disincentive associated with less redistribution means that the lower tax rates in the 1980s do imply lower adjustments than with full redistribution, i.e. lower hours in the 1980s and smaller changes until 2016.

In quantitative terms, overall the results are very robust different redistribution assumptions. The higher the fraction of revenues that is redistributed, the larger the effects of the tax changes on hours worked. But redistribution does not drive all of the effects from the tax changes, as Panel (b) in Figure 8 shows. Here, we compare the results from the decomposition analysis where we only change the tax code in the full-redistribution version to the two versions with 50% (white bars) and no redistribution (light gray bars). While in some countries the effects of tax changes on female hours worked per employed are significantly reduced when decreasing the size of the lump-sum transfers, like the US, UK, or France, 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). On average, 70 percent of the long-term changes in the data are explained by tax changes if we assume 50 percent redistribution, and 53 percent are explained in the case of no redistribution (compared to 115 percent in the baseline model with full redistribution).

4.3.2 Additional Heterogeneity in Income

In the next robustness check, we 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 microdata 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. We pool individuals from all years (2004-2016) and education groups to have a sufficiently large sample. For each education group we calculate the standard deviation of the residuals from our regression as our country-education-specific measure of wage heterogeneity, which we discretize into three and five states. We assume the same wage heterogeneity for men and women. When individuals are matched into couples, we draw randomly from these states within each education type.

Panel (c) in Figure 8 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. The effects on the predicted long-term changes in hours worked go in both directions: in some countries, like the US, Belgium, France, or Germany, more heterogeneity in wages implies larger (absolute) changes in hours worked, whereas in the other countries the implied changes are smaller than in the baseline model. On average, the model with 3 income states explains 99 percent of long-term changes in married women's hours worked and the model with 5 income states explains on average 100 percent. The better fit relative to the 115 percent explained by the baseline model is mainly driven by the much better predictions obtained for Portugal. There is however still considerable cross-country variation in the fraction explained, which ranges from 20 to 200 percent. As panel (d) of Figure 8 shows, also the tax experiment is largely unaffected by the introduction of additional income heterogeneity, with the exception of the Netherlands.

5 Conclusion

We investigate the driving forces of the time-series of hours worked 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, 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 115 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 86 percent of the long-term changes in female hours.

The model also explains part of the secular increase in married women's employment rates, but fails to replicate the large increases in employment rates that especially European women experienced in the eighties and early nineties. Exceptions are the US and UK, where the model performs much better with respect to the evolution of employment rates. This implies that the sharp increase in employment rates in most European countries is explained by factors outside of the model, and not due to changes in labor income and consumption taxes, or wages and the educational composition.

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.

References

S. Albanesi and C. Olivetti. Home Production, Market Production, and the Gender Wage Gap: Incentives and Expectations. *Review of Economic Dynamics*, 12(1):80–107, 2009.

S. Albanesi and C. Olivetti. Gender Roles and Medical Progress. *Journal of Political Economy*, (forthcoming), 2015.

J. Alonso-Ortiz. Social Security and Retirement Across the OECD. *Journal of Economic Dynamics & Control*, 47:300–316, 2014.

O. Attanasio, H. Low, and V. Sanchez-Marcos. Explaining Changes in Female Labor Supply in a Life-Cycle Model. *American Economic Review*, 98(4):1517–1542, 2008.

R. Barro and J.-W. Lee. A new data set of educational attainment in the world, 1950-2010. *Journal of Development Economics*, 104:184–198, 2013.

A. Bick and N. Fuchs-Schündeln. Taxation and Labor Supply of Married Couples Across Countries: A Macroeconomic Analysis. *Review of Economic Studies*, forthcoming, 2017a.

A. Bick and N. Fuchs-Schündeln. Quantifying the Disincentive Effects of Joint Taxation on Married Women's Labor Supply. *American Economic Review Papers & Proceedings*, 107(5):100–104, May 2017b.

A. Bick, B. Brüggemann, and N. Fuchs-Schündeln. Hours Worked in Europe and the US: New Data, New Answers. *Scandinavian Journal of Economics*, forthcoming, 2018.

F. D. Blau and L. M. Kahn. Female Labor Supply: Why is the US Falling Behind? *American Economic Review: Papers & Proceedings*, 103(3):251–256, 2013.

R. Blundell and T. MaCurdy. Labor Supply: A Review of Alternative Approaches. In O. Ashenfelter and D. Card, editors, *Handbook of Labor Economics*, chapter 27, pages 1559–1695. Elsevier Science B.V., 1999.

F. J. Buera, J. P. Kaboski, and M. Q. Zhao. The Rise of Services: the Role of Skills, Scale, and Female Labor Supply. Working Paper, 2014.

I. Chakraborty, H. A. Holter, and S. Stepanchuk. Marriage Stability, Taxation and Aggregate Labor Supply in the US vs. Europe. *Journal of Monetary Economics*, 72:1–20, 2015.

D. Domeij and M. Flodén. The Labor-Supply Elasticity and Borrowing Constraints: Why Estimates are Biased. *Review of Economic Dynamics*, 9:242–262, 2006.

R. Duval-Hernández, L. Fang, and L. R. Ngai. Social Subsidies and Marketization - the Role of Gender and Skill. Working paper, 2018.

A. Erosa, L. Fuster, and G. Kambourov. Labor Supply and Government Programs: A Cross-Country Analysis. *Journal of Monetary Economics*, 59:84–107, 2012.

J. Greenwood, A. Seshadri, and M. Yorukoglu. Engines of Liberation. *Review of Economic Studies*, 72: 109–133, 2005.

N. Guner, R. Kaygusuz, and G. Ventura. Taxation and Household Labor Supply. *Review of Economic Studies*, 79:1113–1149, 2012a.

N. Guner, R. Kaygusuz, and G. Ventura. Taxing Women: A Macroeconomic Analysis. *Journal of Monetary Economics*, 59:111–128, 2012b.

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

L. E. Jones, R. E. Manuelli, and E. R. McGrattan. Why Are Married Women Working So Much? *Journal of Demographic Economics*, 81(1):75–114, 2014.

R. Kaygusuz. Taxes and Female Labor Supply. *Review of Economic Dynamics*, 13(4):725–741, 2010.

M. P. Keane. Labor Supply and Taxes: A Survey. *Journal of Economic Literature*, 49:961–1075, 2011.

J. Knowles. Why are Married Men Working so Much? An Aggregate Analysis of Intra-Household Bargaining and Labour Supply. *Review of Economic Studies*, 80(3):1055–1085, 2013.

C. McDaniel. Forces Shaping Hours Worked in the OECD, 1960-2004. *American Economic Journal: Macroeconomics*, 3:27–52, 2011.

C. McDaniel. Average Tax Rates on Consumption, Investment, Labor and Capital in the OECD 1950-2003. Working Paper, 2012.

C. B. Mulligan and Y. Rubinstein. Selection, Investment, and Women's Relative Wages over Time. *Quarterly Journal of Economics*, 123(3):1061–1110, 2008.

L. R. Ngai and B. Petrongolo. Gender Gaps and the Rise of the Service Economy. *American Economic Journal: Macroeconomics*, Forthcoming, 2017.

L. Ohanian, A. Raffo, and R. Rogerson. Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956-2004. *Journal of Monetary Economics*, 55:1353–1362, 2008.

C. Olivetti. Changes in Women's Hours of Market Work: The Role of Returns to Experience. *Review of Economic Dynamics*, 9:557–587, 2006.

C. Olivetti and B. Petrongolo. Unequal Pay or Unequal Employment? A Cross-Country Analysis of Gender Gaps. *Journal of Labor Economics*, 26:621–654, 2008.

E. C. Prescott. Why Do Americans Work So Much More Than Europeans? *Federal Reserve Bank of Minneapolis Quarterly Review*, 28:2–13, 2004.

K. Ragan. Taxes, Transfers and Time Use: Fiscal Policy in a Household Production Model. *American Economic Journal: Macroeconomics*, 5(1):168–192, 2013.

M. Rendall. Female Market Work, Tax Regimes, and the Rise of the Service Sector. Working Paper, 2015.

R. Rogerson. Structural Transformation and the Deterioration of European Labor Market Outcomes. *Journal of Political Economy*, 116:235–259, 2008.

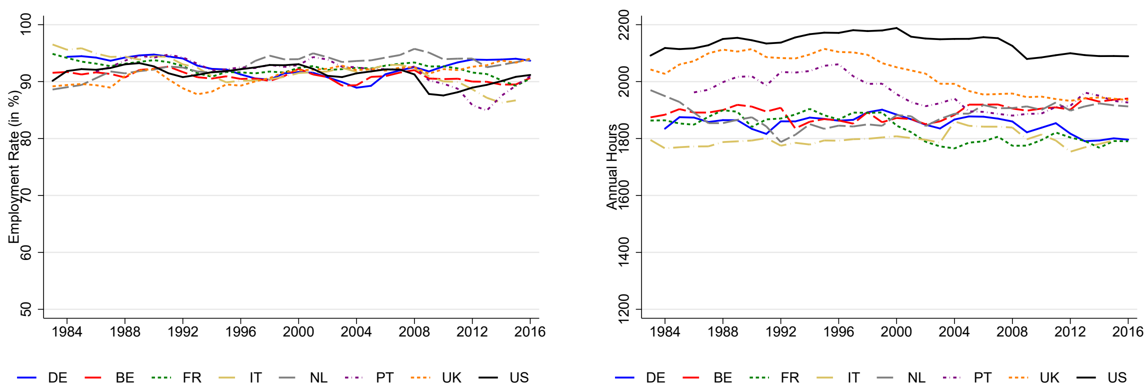
J. Wallenius. Social Security and Cross-Country Differences in Hours: A General Equilibrium Analysis. *Journal of Economic Dynamics and Control*, 37:2466–2482, 2013.

A Appendix

A.1 Data

Figure A.1: Labor Supply of Married Men

(a) Employment Rate (b) Hours Worked per Employed

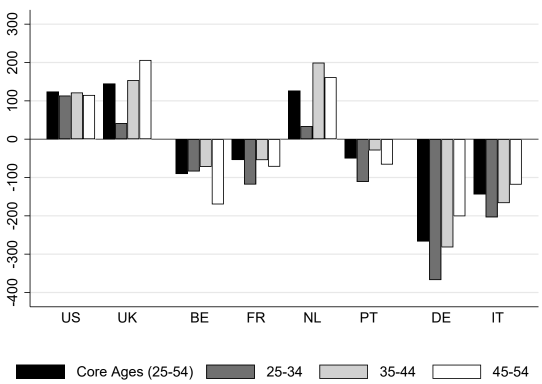


Note: Sample consists of married couples aged 25 to 54. The jump in hours worked per person for Germany in 1991 is a consequence of the reunification of East and West Germany in 1990.

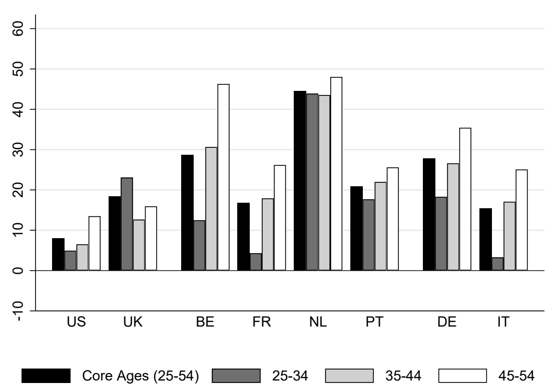
Figure A.2: Robustness of Empirical Facts: Married Women

(a) Age Groups

Hours Worked Per Employed

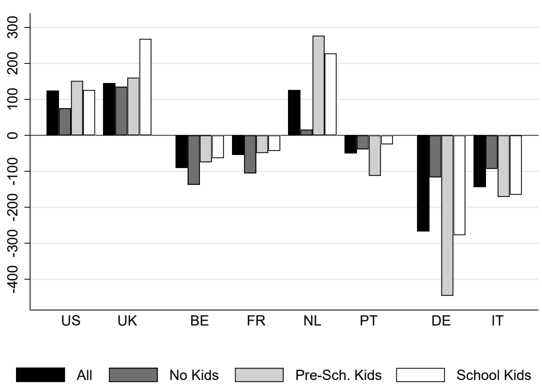


Employment Rates

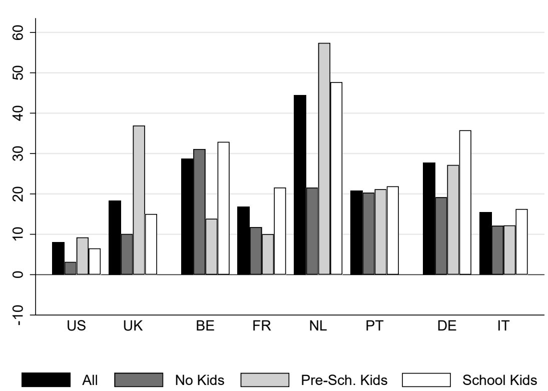


(b) Children

Hours Worked Per Employed



Employment Rates

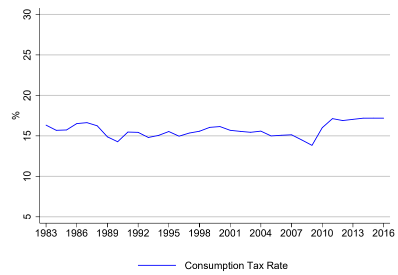


A.2 Model Inputs

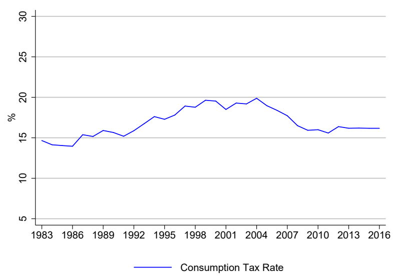
Figure A.3: Consumption Tax Rates

(a) United States

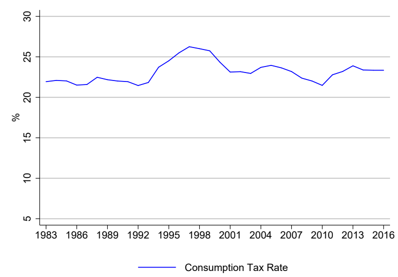
(b) United Kingdom



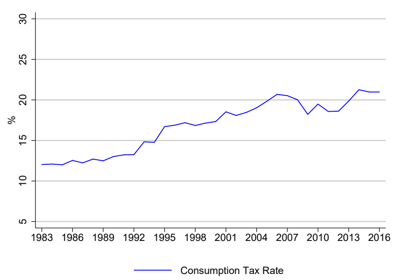
(c) Belgium



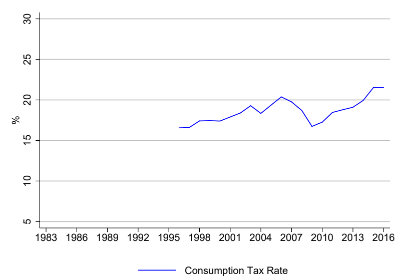
(d) France



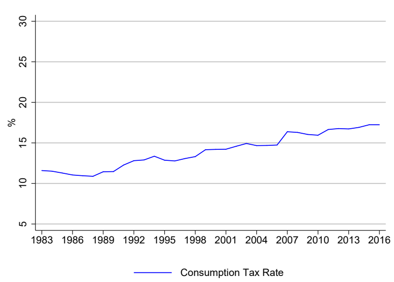
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

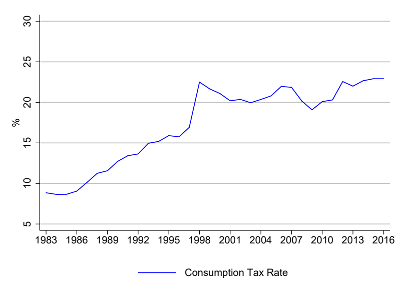


Figure A.4: Female Education Shares in %

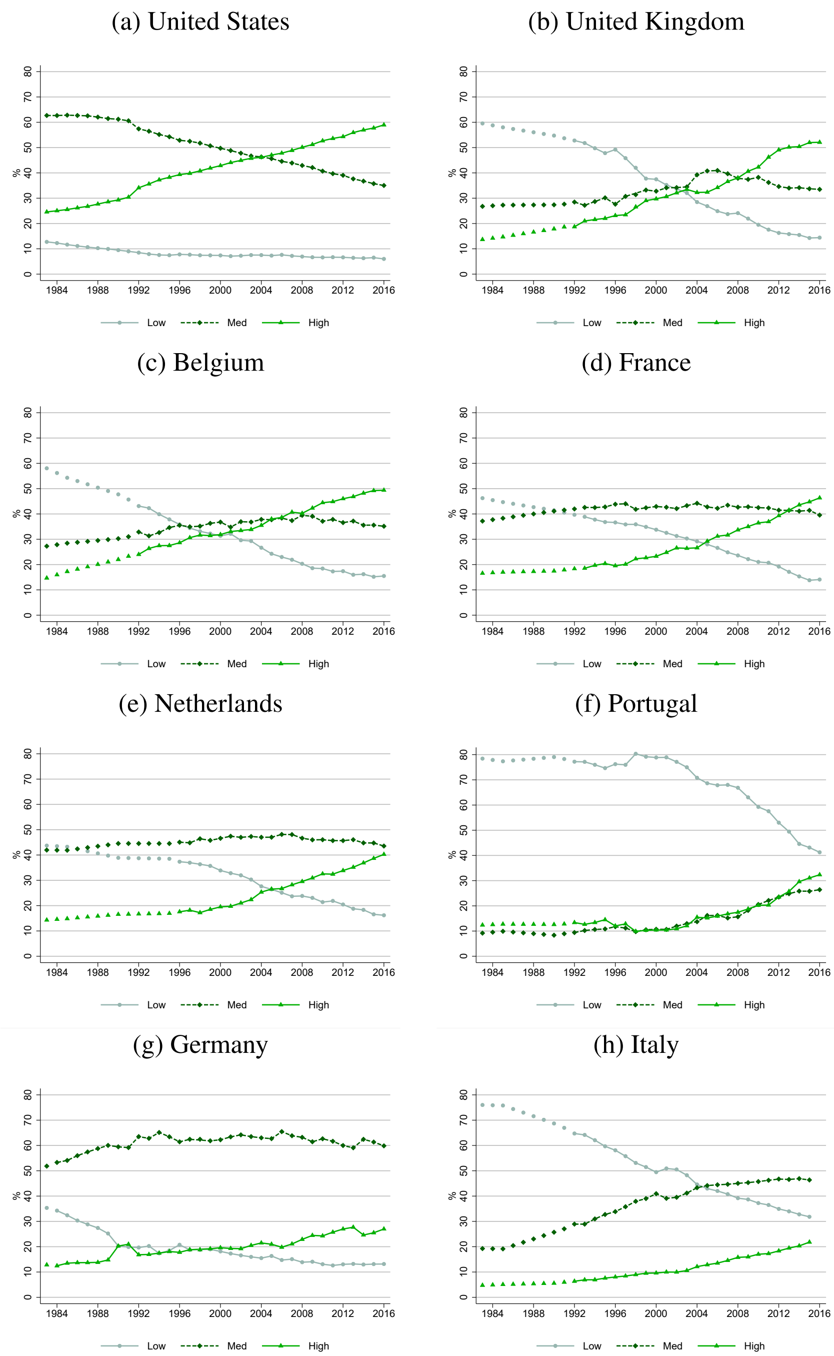
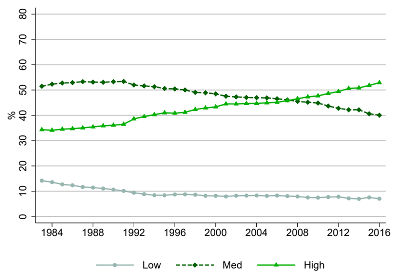
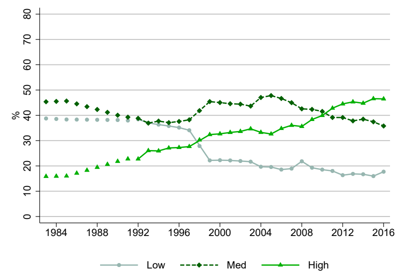


Figure A.5: Male Education Shares in %

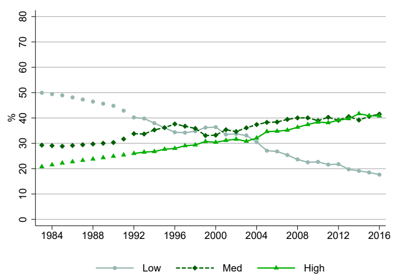
(a) United States



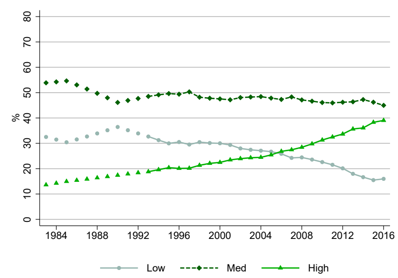
(b) United Kingdom



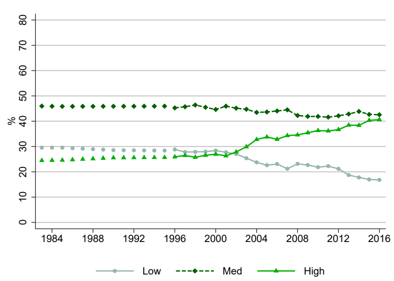
(c) Belgium



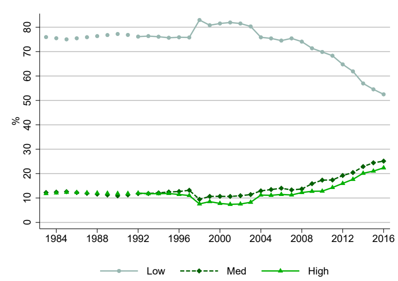
(d) France



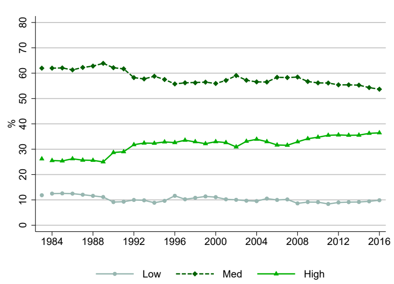
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

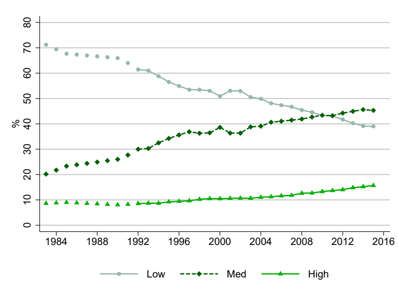
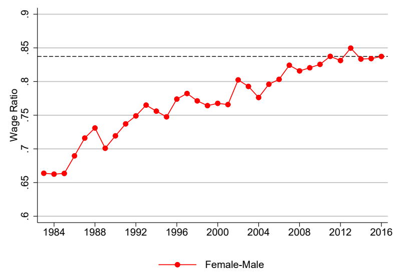
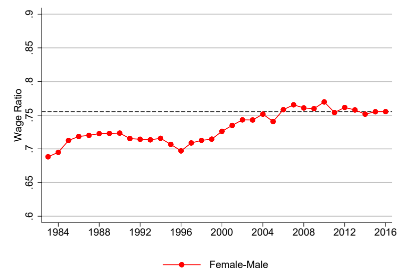


Figure A.6: Gender Wage Gap

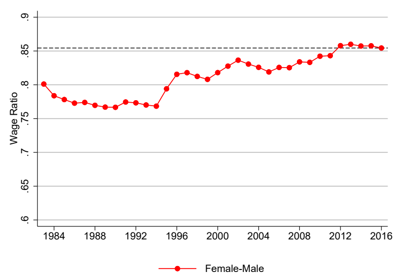
(a) United States



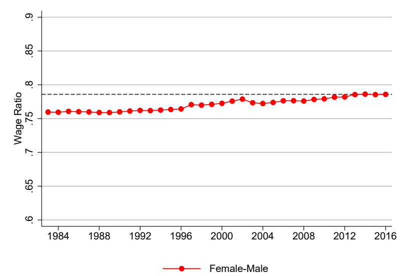
(b) United Kingdom



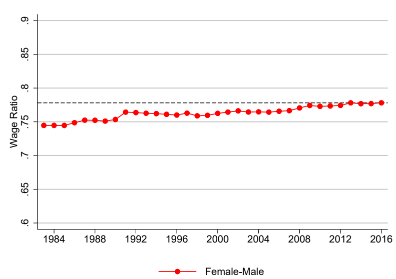
(c) Belgium



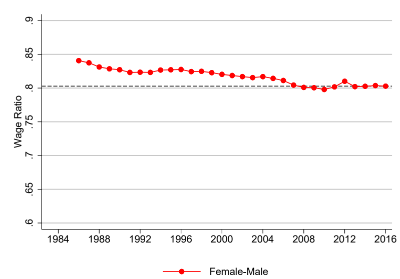
(d) France



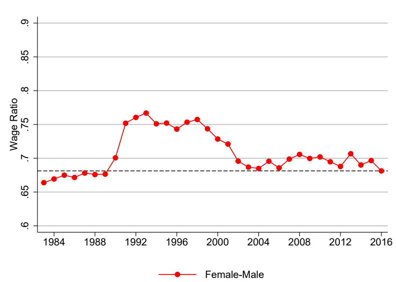
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

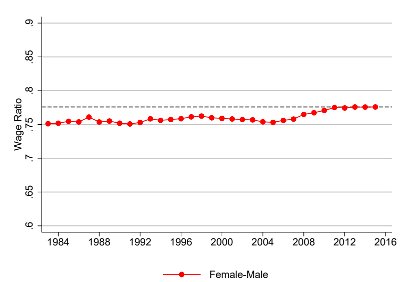
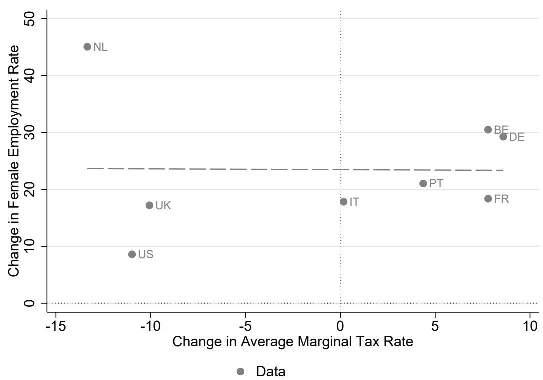
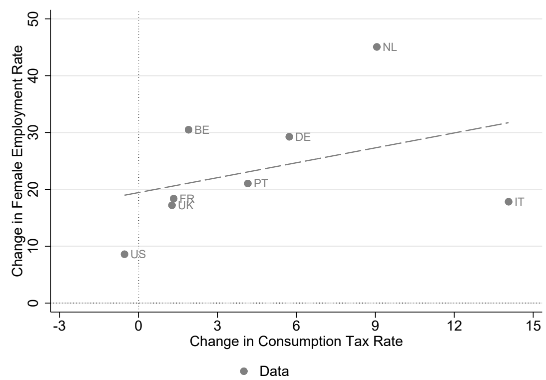


Figure A.7: Correlation of Changes in Female Employment Rates with Changes in Various Inputs

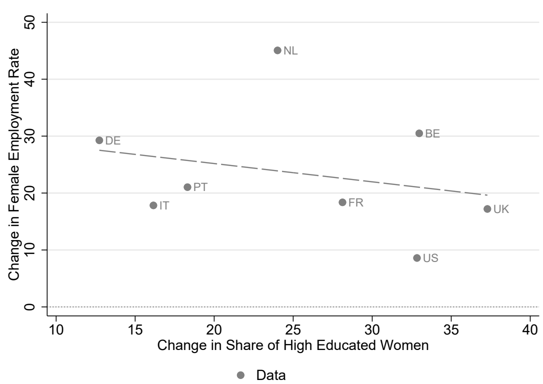
(a) Average Marginal Tax Rate



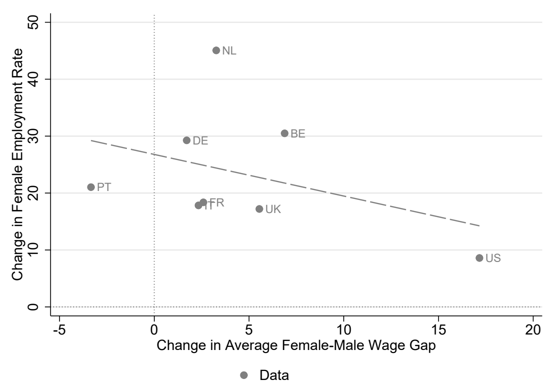
(b) Consumption Tax Rate



(c) Share of High Educated Married Women



(d) Gender Wage Gap



A.2.1 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.[11] 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[12], cash benefits, and the final net-take home pay that a household in a particular country and year faces.

These codes are based on the Taxing Wages reports published annually by the OECD.[13] 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. Based on these algorithms and descriptions we 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 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.

Following Bick and Fuchs-Schündeln (2017b), 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.

The OECD reports adopt some assumptions, which carry over to our codes. We assume that kids are between six and eleven years old (inclusive).[14] 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 consider taxes imposed on the state level, provincial level, and local level. As the local tax system often also varies within the country, we either assume 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

[11] 1995 and 2001 are missing as there is no OECD documentation for these years. 1994 still has to be coded up.

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

[13] 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".

[14] The most generous reliefs or transfers are assumed if the amounts differ within this age group.

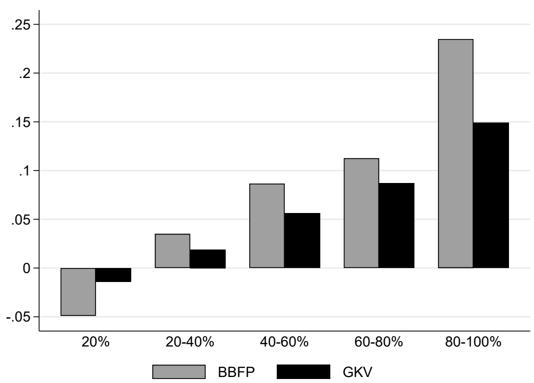
annual basis.

From the year 2001 onwards, the OECD has implemented the tax codes in Stata. This OECD program 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.[15] 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.

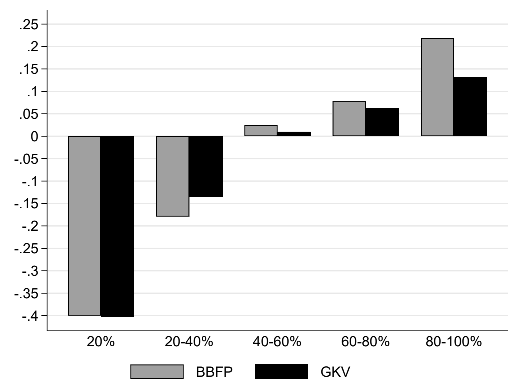
A.2.2 Comparison to Effective Average Tax Rates from Guner et al. (2014)

Figure A.8: Comparing Statutory and Effective Average Tax Rates

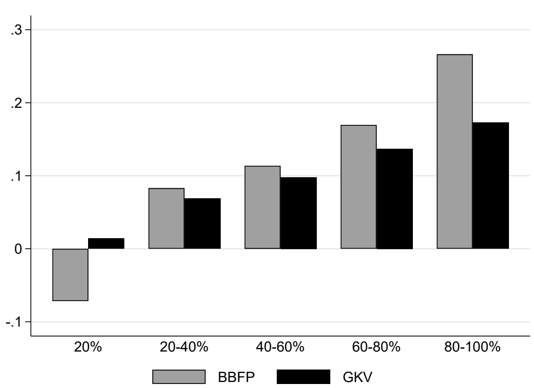
(a) Married, No Children



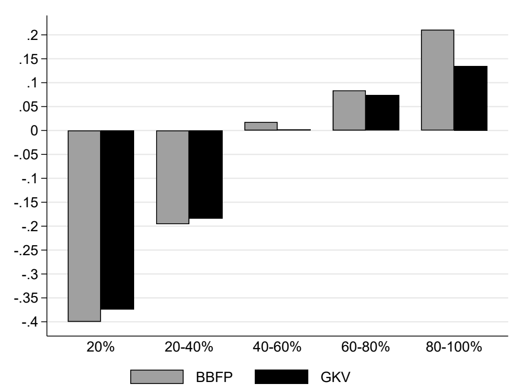
(b) Married, 2 Children



(c) Unmarried, No Children



(d) Unmarried, 2 Children



[15] Some of these benefits are just incorporated into the model for some specific countries and years.

A.2.3 Education and Matching Imputation

In this section, we provide more details on the imputation of education and matching shares for missing years. The EU-LFS includes data on education only from 1992 onwards, so we have to rely on other data sources to impute data for earlier years for the European countries in our sample except Germany. Concretely, we use the information on education shares in the Barro-Lee Educational Attainment Data (Barro and Lee, 2013) to extrapolate the time series for the education and matching shares backwards until 1983. The Barro-Lee Educational Attainment Data is available by gender and age groups in 5-year intervals from as early as 1950. We first interpolate the data to account for missing years. Then, we regress the educational shares of married men and women aged 25 to 54 until the year 2000 on the Barro-Lee educational shares for each of the six age groups between ages 25 to 54 (25-29, 30-34, 35-39,...). The exact regression equations are given by

$$\mu(x) = \alpha_x + \sum_{i=1}^{6} \beta_{x,i}\mu_i^{BL}(x) + \varepsilon_x, \text{ and} \tag{A.1}$$

$$\mu(z) = \alpha_z + \sum_{i=1}^{6} \beta_{z,i}\mu_i^{BL}(z) + \varepsilon_z, \tag{A.2}$$

where $\mu(x)$ and $\mu(z)$ denote the educational shares for women (x) and men (z), and μ_i^{BL} stands for the educational shares by age group from the Barro-Lee Educational Attainment Data, with $i \in \{25\text{-}29, 30\text{-}34, 35\text{-}39, 40\text{-}44, 45\text{-}49, 50\text{-}54\}$. We then use the estimated values for α and β as well as the available Barro-Lee data for the 80s and early 90s to predict educational shares prior to 1992.

For the matching shares, we proceed in the same way. We regress each matching share on the female educational shares from the Barro-Lee data and use the obtained estimates to predict the time series backwards. The corresponding linear regression equation is given by:

$$\mu(x,z) = \alpha_{x,z} + \sum_{i=1}^{6} \beta_{x,z,i}\mu_i^{BL}(x) + \varepsilon_{x,z} \tag{A.3}$$

A.2.4 Wage Imputation

In this section, we describe the imputation of wages for missing years in more detail. The EU-LFS does not provide any earnings data for the European countries. We therefore rely on a number of other datasets to get reliable estimates. For all European countries except Germany, 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. This European household data set captures income and usual hours, but features a sample size an order of magnitude smaller than the EU-LFS. From 1994 to 2001, we use the European Community Household Panel (ECHP), the EU-SILC's predecessor. We use linear interpolation to obtain gender- and education-specific wages for the missing years between the end of the ECHP and the beginning of the EU-SILC. For the remaining years prior to 1994, we have to rely entirely on estimations to impute gender- and education-specific wages. To do that, we first calculate average hourly wages on the aggregate level using a consistent earnings series that is available for our whole time series. The only aggregate earnings series that fulfills these requirements are the average annual wages of production workers that the OECD publishes along with their tax documentation described in Section 3.2.1. Using usual hours of full-time employees from Bick et al. (2018), we first transform those annual wages into a measure of average hourly wages of production workers, denoted by w_t^{OECD}. For each country, we then regress the gender-education-specific wages from the microdata covering 1994 to 2016 on these average production worker wages in each country using the following regression model:

$$w_{f,x,t} = \alpha^{w}_{f,x} + \beta^{w}_{f,x} w^{OECD}_{t} + \varepsilon_{f,x,t} \tag{A.4}$$

$$w_{m,z,t} = \alpha^{w}_{m,z} + \beta^{w}_{m,z} w^{OECD}_{t} + \varepsilon_{m,z,t} \tag{A.5}$$

Using the estimated constant and coefficients for average production worker wages from each regression, we then predict gender-education-specific wages for all years between 1983 and 2016. We use the education-specific predicted wages for married women $\hat{w}_{f,x,t}$ and men $\hat{w}_{m,z,t}$ for all years in our sample (instead of using raw data for the available years).

With the CPS, we have wage data for the US for all available years. This enables us to run a robustness check where we compare the model predictions when using the wage inputs obtained from the micro data to the results when using wages predicted through the procedure described above. The results can be found in Section A.4.3.

A.3 Targeted & Non-Targeted Moments (2005-2007)

Table A.1: Data Targets for U.S.

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.390$	2089	2080	-9
HWE_f	$\alpha_f = 0.414$	1757	1766	9
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 0.519$, $\theta_{low} = 0.756$			
Low educ. woman		39.4	40.7	1.3
Medium educ. woman		56.2	53.9	−2.3
High educ. woman		66.5	67.8	1.3
Medium educ. husband:	$k_{med} = 1.297$, $\theta_{med} = 0.138$			
Low educ. woman		40.0	41.5	1.5
Medium educ. woman		66.6	64.0	−2.6
High educ. woman		82.9	85.4	2.5
High educ. husband:	$k_{high} = 0.486$, $\theta_{high} = 0.327$			
Low educ. woman		51.6	51.9	0.3
Medium educ. woman		64.5	64.1	−0.4
High educ. woman		76.2	76.4	0.2

Table A.2: Untargeted Moments for U.S.

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1939.5	1896.1	−43.4
Medium education	2059.9	2072.0	12.1
High education	2131.7	2110.6	−21.1
Hours Worked per Employed Woman			
Low education	1674.5	1486.4	−188.1
Medium education	1726.2	1698.4	−27.8
High education	1776.6	1835.4	58.8

Table A.3: Data Targets for U.K.

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.464$	1935	1931	-4
HWE_f	$\alpha_f = 0.781$	1315	1313	-2
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 2.090$, $\theta_{low} = 0.076$			
Low educ. woman		57.3	64.0	6.7
Medium educ. woman		79.0	71.3	−7.7
High educ. woman		86.3	88.2	1.9
Medium educ. husband:	$k_{med} = 1.095$, $\theta_{med} = 0.119$			
Low educ. woman		67.7	69.9	2.2
Medium educ. woman		77.7	75.5	−2.2
High educ. woman		87.1	87.9	0.8
High educ. husband:	$k_{high} = 0.508$, $\theta_{high} = 0.231$			
Low educ. woman		71.7	72.1	0.4
Medium educ. woman		76.3	75.7	−0.6
High educ. woman		83.7	83.9	0.2

Table A.4: Untargeted Moments for U.K.

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1874.4	1817.4	−57.0
Medium education	1926.7	1897.0	−29.7
High education	1964.6	2000.0	35.4
Hours Worked per Employed Woman			
Low education	1276.1	1137.7	−138.4
Medium education	1258.6	1235.7	−22.9
High education	1356.1	1410.3	54.2

Table A.5: Data Targets for Belgium

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.300$	1940	1948	8
HWE_f	$\alpha_f = 0.384$	1471	1474	3
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 4.886$, $\theta_{low} = 0.031$			
Low educ. woman		48.7	58.3	9.6
Medium educ. woman		73.0	64.8	−8.2
High educ. woman		83.0	81.9	−1.1
Medium educ. husband:	$k_{med} = 2.845$, $\theta_{med} = 0.044$			
Low educ. woman		55.5	67.1	11.6
Medium educ. woman		77.8	72.1	−5.7
High educ. woman		89.0	84.3	−4.7
High educ. husband:	$k_{high} = 4.496$, $\theta_{high} = 0.027$			
Low educ. woman		44.1	59.6	15.5
Medium educ. woman		73.9	66.2	−7.7
High educ. woman		88.4	83.1	−5.3

Table A.6: Untargeted Moments for Belgium

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1798.0	1888.5	90.5
Medium education	1936.3	1948.5	12.2
High education	2006.0	1972.4	−33.6
Hours Worked per Employed Woman			
Low education	1255.1	1255.2	0.1
Medium education	1405.6	1269.0	−136.6
High education	1549.9	1688.8	138.9

Table A.7: Data Targets for France

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.564$	1791	1783	-8
HWE_f	$\alpha_f = 0.626$	1469	1466	-3
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 2.055$, $\theta_{low} = 0.073$			
Low educ. woman		51.1	57.4	6.3
Medium educ. woman		73.6	66.1	-7.5
High educ. woman		82.8	85.2	2.4
Medium educ. husband:	$k_{med} = 2.661$, $\theta_{med} = 0.049$			
Low educ. woman		60.0	64.3	4.3
Medium educ. woman		79.7	73.9	-5.8
High educ. woman		87.9	91.0	3.1
High educ. husband:	$k_{high} = 2.765$, $\theta_{high} = 0.039$			
Low educ. woman		48.3	56.0	7.7
Medium educ. woman		77.2	66.1	-11.1
High educ. woman		85.0	88.5	3.5

Table A.8: Untargeted Moments for France

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1610.1	1706.7	96.6
Medium education	1760.4	1736.8	-23.6
High education	1898.9	1868.8	-30.1
Hours Worked per Employed Woman			
Low education	1316.5	1378.2	61.7
Medium education	1452.8	1413.6	-39.2
High education	1511.7	1536.8	25.1

Table A.9: Data Targets for Netherlands

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.350$	1913	1858	-55
HWE_f	$\alpha_f = 0.938$	1197	1189	-8
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 0.547$, $\theta_{low} = 0.237$			
Low educ. woman		65.6	68.8	3.2
Medium educ. woman		81.5	78.6	−2.9
High educ. woman		90.2	86.4	−3.8
Medium educ. husband:	$k_{med} = 1.147$, $\theta_{med} = 0.086$			
Low educ. woman		65.0	72.0	7.0
Medium educ. woman		82.6	84.0	1.4
High educ. woman		91.5	93.5	2.0
High educ. husband:	$k_{high} = 1.381$, $\theta_{high} = 0.069$			
Low educ. woman		67.4	65.6	−1.8
Medium educ. woman		79.4	80.9	1.5
High educ. woman		87.2	89.1	1.9

Table A.10: Untargeted Moments for Netherlands

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1905.9	1609.8	−296.1
Medium education	1928.8	1828.1	−100.7
High education	1899.2	1993.0	93.8
Hours Worked per Employed Woman			
Low education	1056.5	1214.5	158.0
Medium education	1117.8	1140.4	22.6
High education	1317.1	1231.8	−85.3

Table A.11: Data Targets for Portugal

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.402$	1927	1922	-5
HWE_f	$\alpha_f = 0.377$	1702	1642	-60
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 0.348$, $\theta_{low} = 0.456$			
Low educ. woman		73.4	74.3	0.9
Medium educ. woman		81.9	79.9	-2.0
High educ. woman		90.5	88.8	-1.7
Medium educ. husband:	$k_{med} = 0.065$, $\theta_{med} = 5.344$			
Low educ. woman		79.3	81.1	1.8
Medium educ. woman		84.3	82.8	-1.5
High educ. woman		91.6	85.9	-5.7
High educ. husband:	$k_{high} = 0.270$, $\theta_{high} = 0.352$			
Low educ. woman		77.9	73.7	-4.2
Medium educ. woman		75.9	79.2	3.3
High educ. woman		90.7	88.5	-2.2

Table A.12: Untargeted Moments for Portugal

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1893.8	1892.8	-1.0
Medium education	1933.7	1880.7	-53.0
High education	1995.5	2035.5	40.0
Hours Worked per Employed Woman			
Low education	1663.5	1441.2	-222.3
Medium education	1743.2	1603.2	-140.0
High education	1710.6	1930.0	219.4

Table A.13: Data Targets for Germany

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.561$	1796	1795	-1
HWE_f	$\alpha_f = 0.874$	1116	1119	3
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 0.789$, $\theta_{low} = 0.163$			
Low educ. woman		53.2	55.8	2.6
Medium educ. woman		71.9	67.1	-4.8
High educ. woman		76.7	79.6	2.9
Medium educ. husband:	$k_{med} = 1.133$, $\theta_{med} = 0.070$			
Low educ. woman		58.6	63.7	5.1
Medium educ. woman		83.1	77.7	-5.4
High educ. woman		86.4	90.3	3.9
High educ. husband:	$k_{high} = 1.189$, $\theta_{high} = 0.051$			
Low educ. woman		54.4	60.0	5.6
Medium educ. woman		81.9	73.0	-8.9
High educ. woman		82.1	87.7	5.6

Table A.14: Untargeted Moments for Germany

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1639.7	1669.6	29.9
Medium education	1753.4	1716.1	-37.3
High education	1901.7	1946.4	44.7
Hours Worked per Employed Woman			
Low education	938.1	969.6	31.5
Medium education	1088.2	1105.2	17.0
High education	1233.7	1221.9	-11.8

Table A.15: Data Targets for Italy

	Parameters	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked:				
HWP_m	$\alpha_m = 0.423$	1793	1793	0
HWE_f	$\alpha_f = 0.625$	1352	1351	-1
Female Employment Rates by Husband's and Own Education (in %)				
Low educ. husband:	$k_{low} = 2.578$, $\theta_{low} = 0.092$			
Low educ. woman		35.7	35.7	0.0
Medium educ. woman		56.8	56.6	−0.2
High educ. woman		70.3	70.4	0.1
Medium educ. husband:	$k_{med} = 2.379$, $\theta_{med} = 0.088$			
Low educ. woman		43.3	42.8	−0.5
Medium educ. woman		62.7	63.5	0.8
High educ. woman		76.8	75.7	−1.1
High educ. husband:	$k_{high} = 1.648$, $\theta_{high} = 0.108$			
Low educ. woman		53.7	51.4	−2.3
Medium educ. woman		61.7	68.4	6.7
High educ. woman		82.2	77.9	−4.3

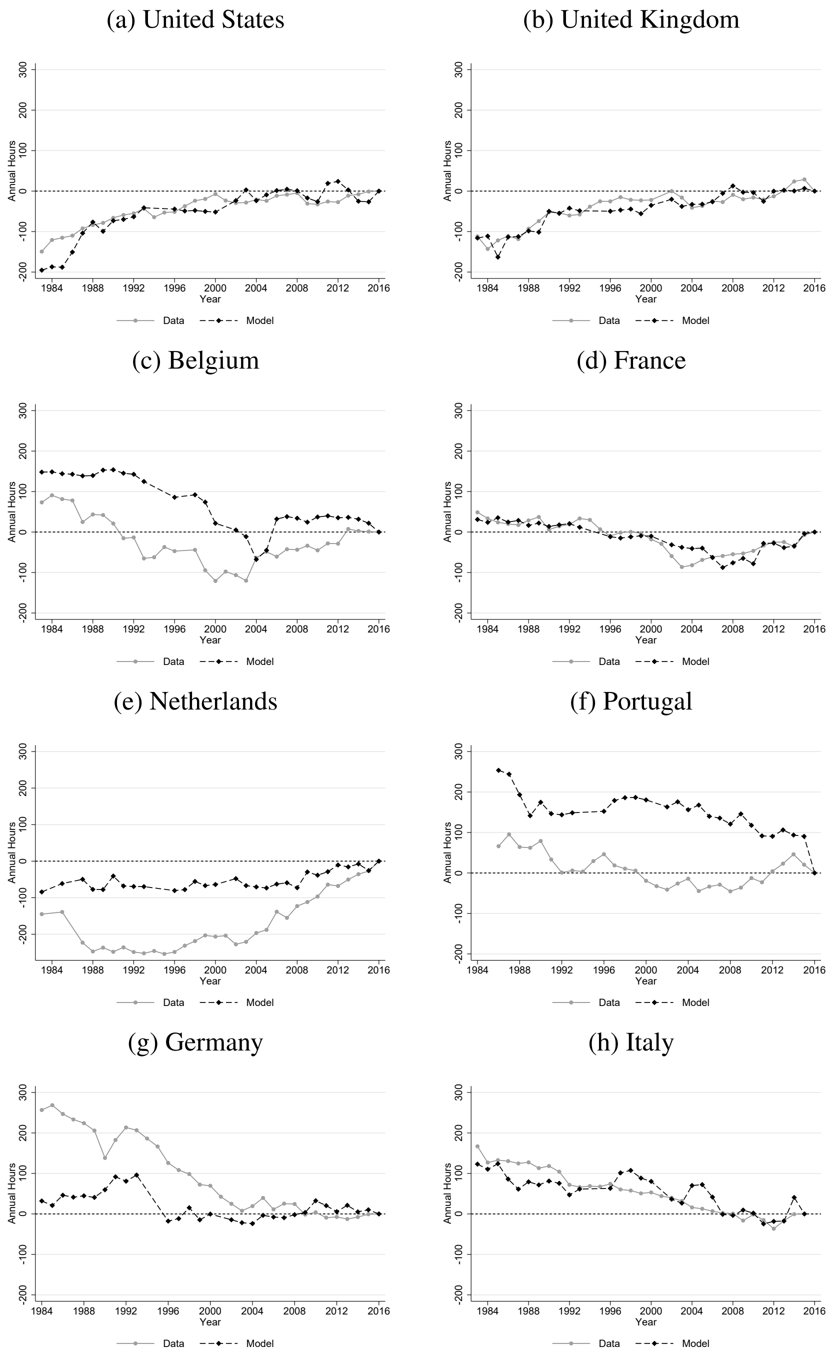
Table A.16: Untargeted Moments for Italy

	Data	Model	$\Delta_{\text{Model-Data}}$
Hours Worked per Man			
Low education	1765.9	1837.1	71.2
Medium education	1808.8	1723.2	−85.6
High education	1817.8	1883.9	66.1
Hours Worked per Employed Woman			
Low education	1362.9	1253.5	−109.4
Medium education	1361.8	1344.0	−17.8
High education	1327.0	1507.1	180.1

A.4 Results

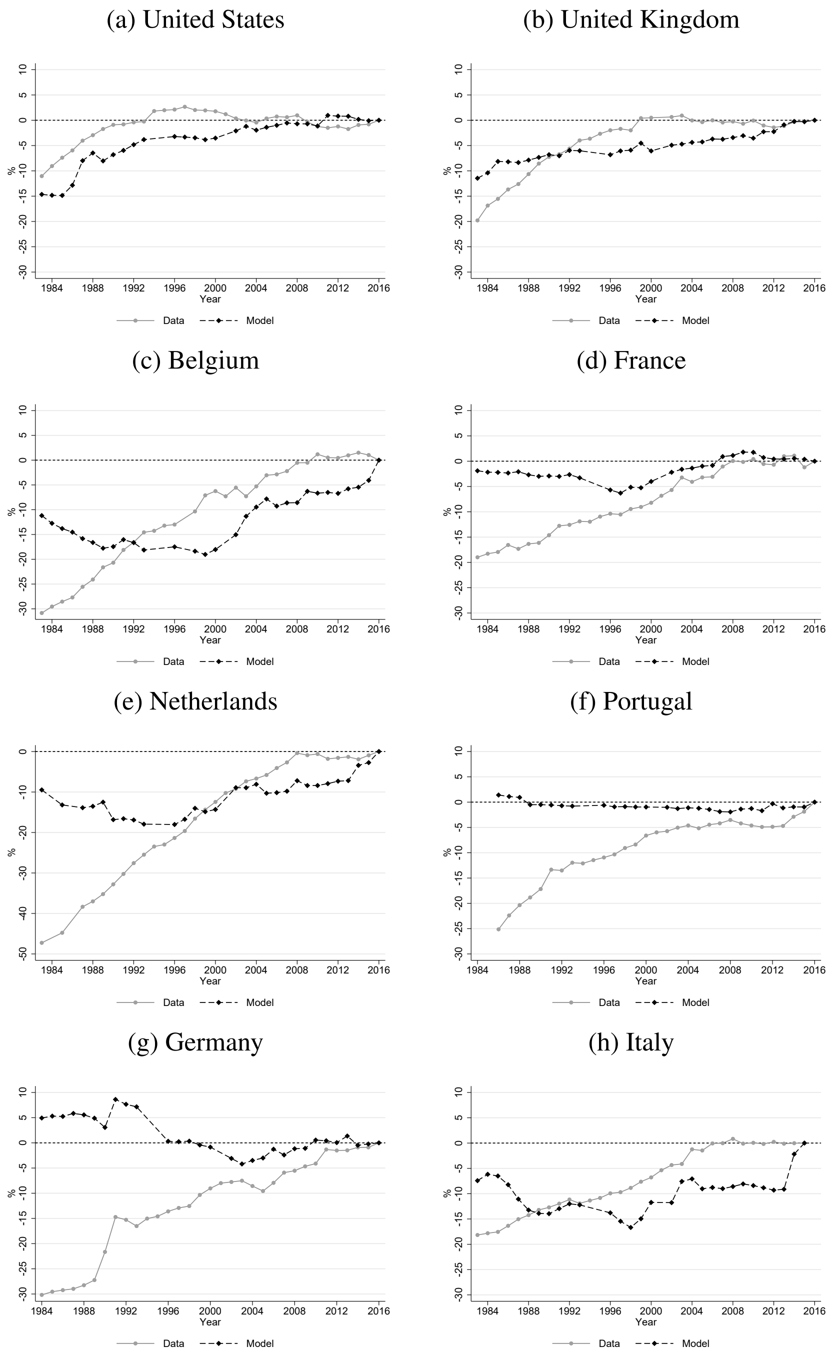
A.4.1 Time Series Predictions

Figure A.9: Time Series Predictions for Female Hours Worked per Employed



Note: We exclude the years 1994-1995 and 2001 from the graphs because the OECD does not provide tax codes for these years.

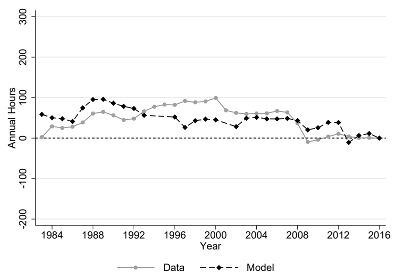
Figure A.10: Time Series Predictions for Female Employment Rates



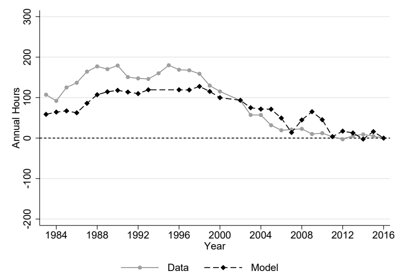
Note: We exclude the years 1994-1995 and 2001 from the graphs because the OECD does not provide tax codes for these years.

Figure A.11: Time Series Predictions for Male Hours Worked per Employed

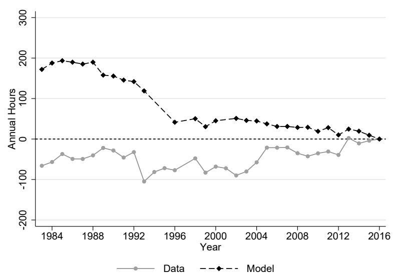
(a) United States



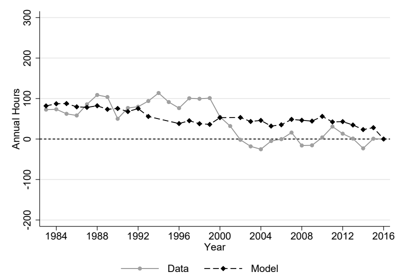
(b) United Kingdom



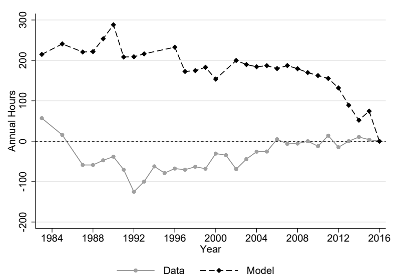
(c) Belgium



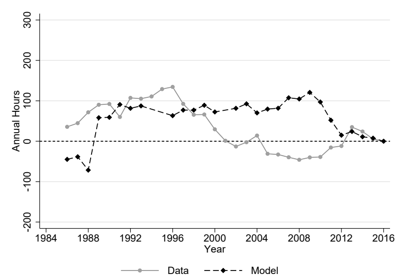
(d) France



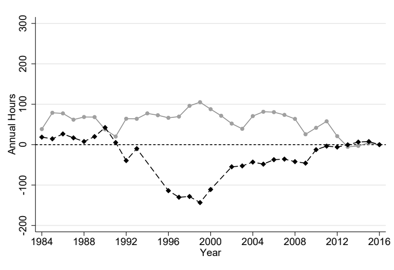
(e) Netherlands



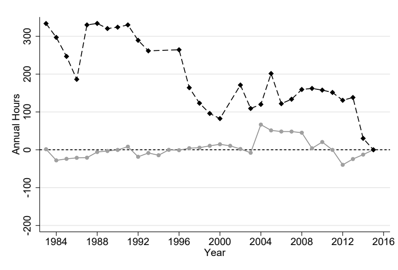
(f) Portugal



(g) Germany



(h) Italy

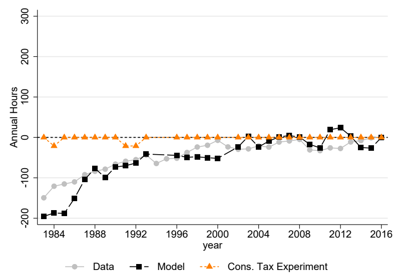


Note: We exclude the years 1994-1995 and 2001 from the graphs because the OECD does not provide tax codes for these years.

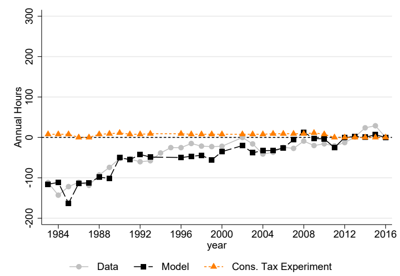
A.4.2 Decomposition Results

Figure A.12: Model output when only varying consumption tax

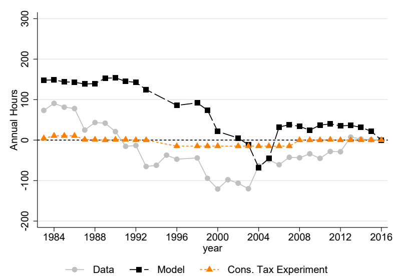
(a) United States



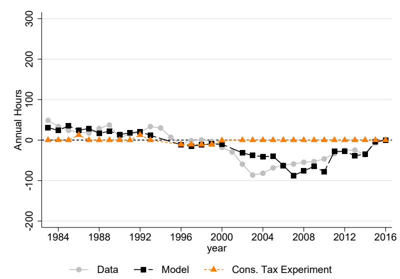
(b) United Kingdom



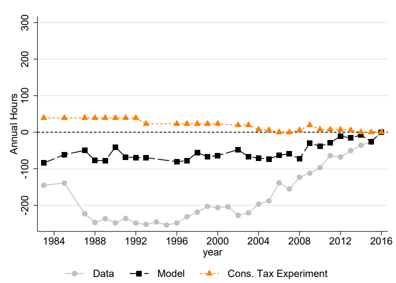
(c) Belgium



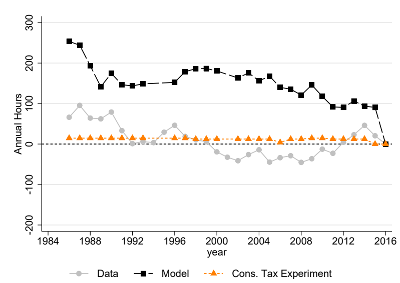
(d) France



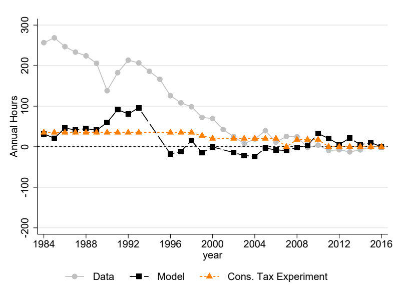
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

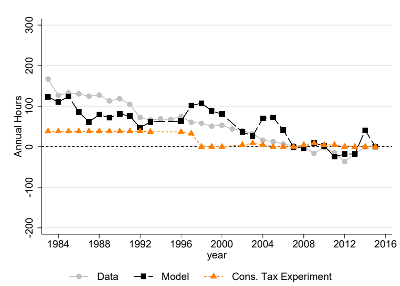
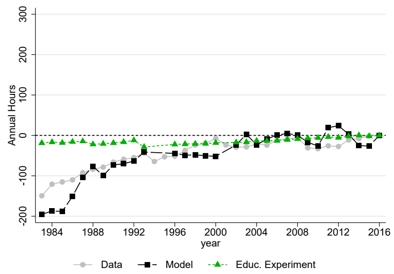
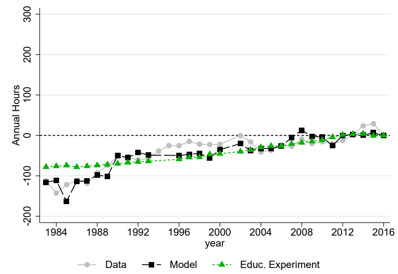


Figure A.13: Model output when only varying educational composition and matching

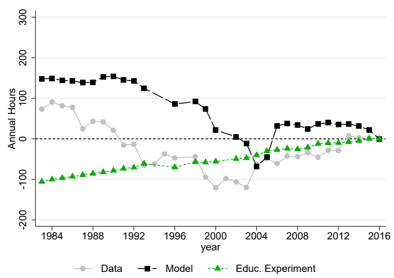
(a) United States



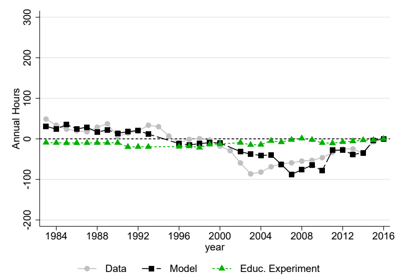
(b) United Kingdom



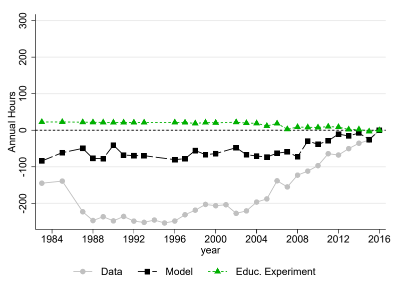
(c) Belgium



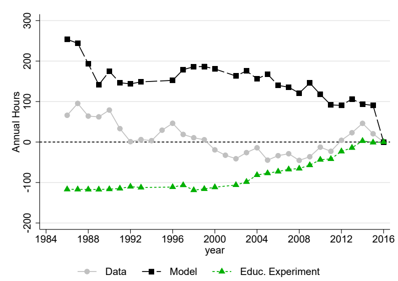
(d) France



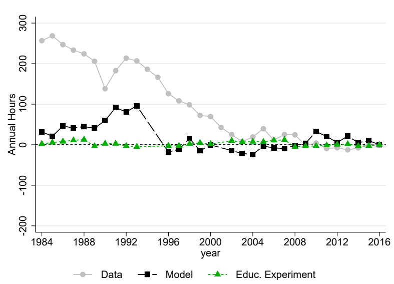
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

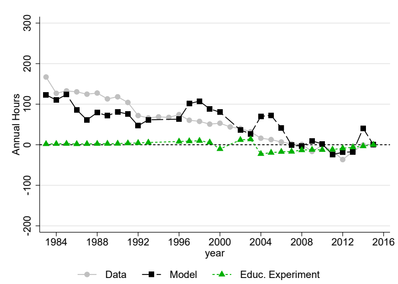
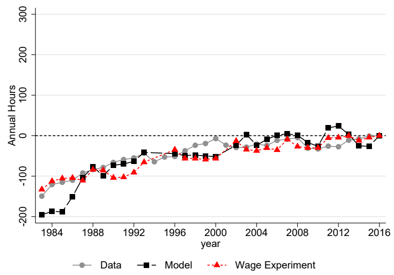
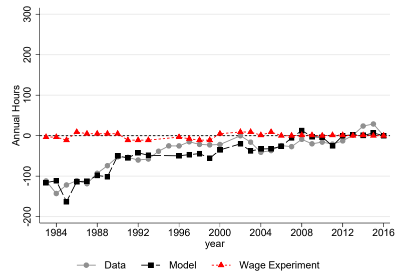


Figure A.14: Model output when only varying wages

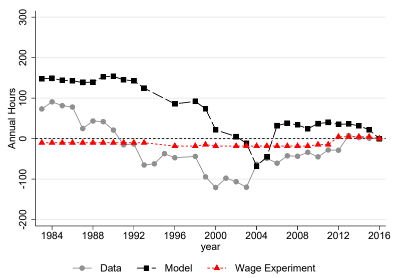
(a) United States



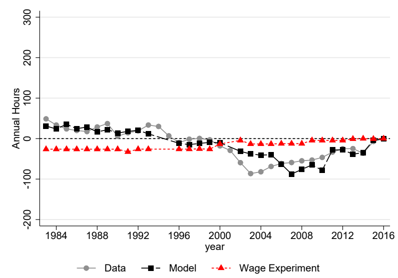
(b) United Kingdom



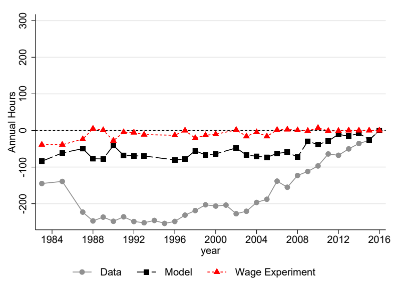
(c) Belgium



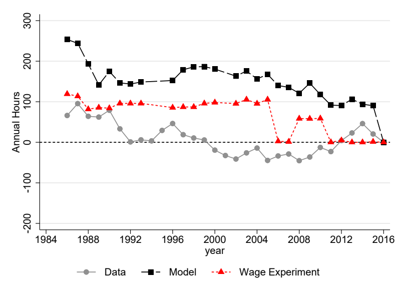
(d) France



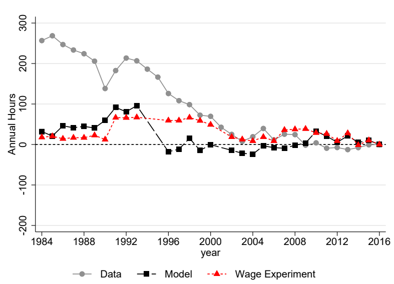
(e) Netherlands



(f) Portugal



(g) Germany



(h) Italy

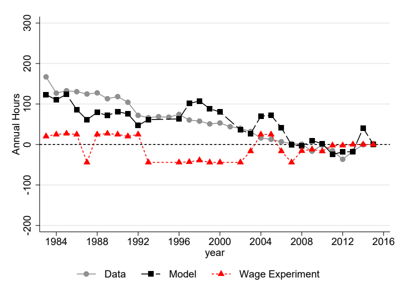
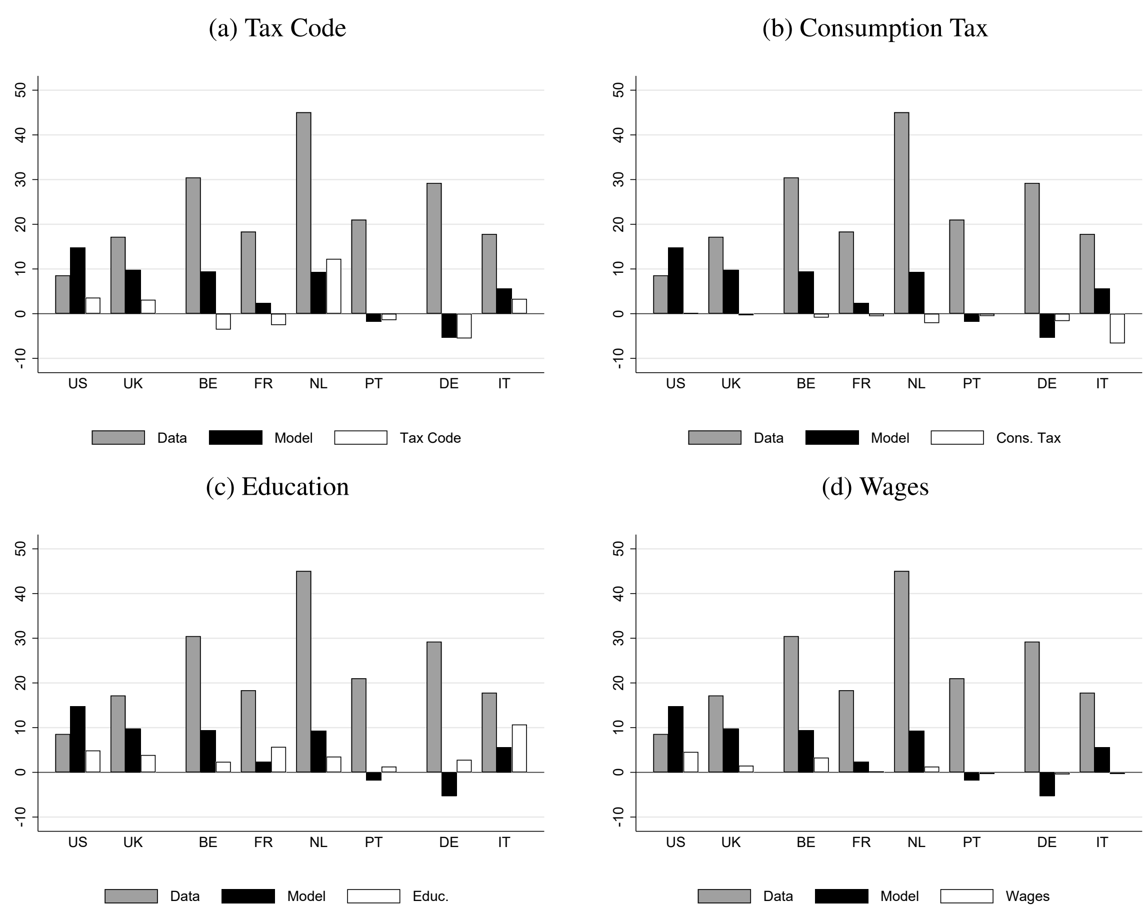


Table A.17: Correlation between Data and Decomposition Output for Female Employment Rates

Country	**Total**	**Tax Code**	**Cons. Tax**	**Educ.**	**Wages**
Positive Hours Trend					
United States	0.79	0.77	0.27	0.63	0.71
United Kingdom	0.81	0.37	0.06	0.72	0.55
Changing Hours Trend					
Belgium	0.64	−0.55	−0.53	0.97	0.87
France	0.66	−0.04	−0.23	0.94	0.56
Netherlands	0.65	0.76	−0.96	0.89	0.85
Portugal	−0.77	−0.85	−0.69	0.82	−0.66
Negative Hours Trend					
Germany	−0.69	−0.73	−0.96	0.93	−0.09
Italy	0.34	0.13	−0.91	0.98	−0.35

Figure A.15: Changes in Married Women's Employment Rates between 1983-85 and 2014-16: Decomposition



A.4.3 Imputed vs. Actual Wages

Prior to 1994, we do not have microdata on earnings that would enable to estimate gender- and education-specific wages as we do in later years. In Appendix Section A.2.4, we describe how we impute wages for those missing years.

In order to assess the consequences of these imputations we run a robustness check, where we use the fact that for the US we have access to micro data to calculate the full time-series of gender- and education-specific wages. In the baseline model, we use the micro data provided by the US CPS to directly calculate or estimate (in the case of Heckman corrected wages for married women) wages for the full time-series and use these as our input into the model. In this robustness check, we instead estimate wages as if we did not have the full time series available but instead the same number of years as for the other countries. We then use those wages as inputs into the model, and compare the results to the baseline results.

In principle we could run the same robustness check for Germany, for which we also have micro data on wages for the whole time series. But the German reunification and the concurrent abrupt changes in wages, which the predictions would not capture, make the results less meaningful, so we abstain from looking at Germany here.

The results based on imputed wages for the US are depicted as the dotted line (with triangular markers) in Figure A.16. For the years 1994 and following, using imputed or actual wages has only minor effects on the results. For the years prior to 1994, the model results based on imputed wages lie however always above the model results based on actual wages. This is because the ratio of female to male wages implied by the imputation is too high compared to the actual data as Figure A.17 shows, thus implying higher hours and employment rates. Overall, however, the differences that we find for the US are relatively minor.

Figure A.16: Time-Series Predictions for Married Women's Labor Supply in the US: Actual vs. Imputed Wages

Figure A.17: Gender Wage Gap in the US: Actual vs. Imputed Wages

(a) United States