Online Appendix

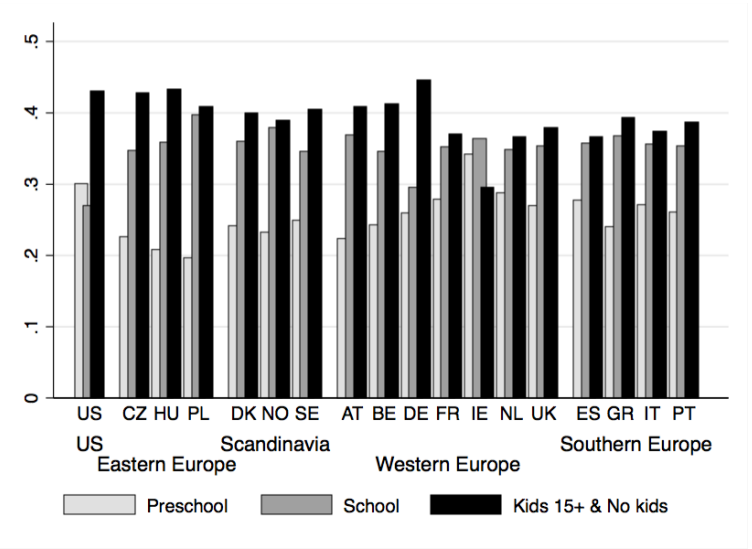
Taxation and Labor Supply of Married Women across Countries: A Macroeconomic Analysis

Alexander Bick
Arizona State University

Nicola Fuchs-Schündeln
Goethe University Frankfurt, CEPR and CFS

March 31, 2016

Figure W.1: Sample Composition by Presence of Children in Household



W.1 Married Mothers' Labor Supply

Figure W.1 shows the percentage of married women in our sample with a youngest child of preschool age (aged 0-4 in Europe, 0-5 in the US), a youngest child of school age (aged 5-14 in Europe, 6-13 in the US), and no child of either preschool or school age.[29] The group of women with preschool children is the smallest group in all countries, making up on average 25.5 percent of our sample of married women, and never exceeding 30 percent except for Ireland, where they constitute with 34 percent the second largest group.[30]

Figure W.2 shows differences in both margins of labor supply (employment rate in panels (a) and (b), and hours worked per employed in panels (c) and (d)) from the behavior of the same group in the US, contrasting women with preschool children on the y-axis and women without children younger than 15 (i.e., either with children aged 15 or above in the household, or no children in the household) on the x-axis in panels (a) and (c). In panels (b) and (d) the same comparison is drawn for women with school aged children on the y-axis. The age always refers to the youngest child. Since we are focusing on a comparison to the respective US group, we already take care of general differences between these three groups that apply in all countries. In fact, in the US women with preschool (school) children have an 18 (6) percentage point lower employment rate than women without children younger than 15, and work 14 (8) percent lower hours conditional on working. Focusing on women with school children compared to women without children younger than 15 in panels (b) and (d), differences in employment rates and hours worked per employed across countries are generally quite small. This is not the case for mothers of preschool children, where specifically the employment rate of women shows a lot of country-specific idiosyncracies, as panel (a) shows. Strikingly low employment rates in the Czech Republic and Hungary for women with preschool children, and relatively high ones in all Southern European countries, especially Portugal, and in Belgium and the Netherlands stand

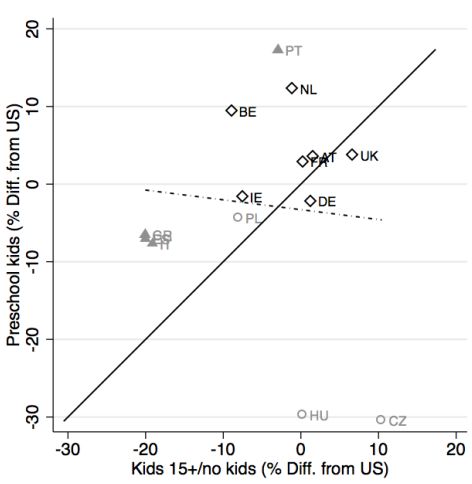
[29] Note that for Scandinavia these data come from the EU-SILC.

[30] In the US, the group of women with school children make up a slightly smaller share than the group of women with preschool children. Note, however, that the age definition of the groups is slightly different for the US, which in itself leads to a larger share of women with preschool children and a smaller share of women with school children compared to the definition applied in the European countries.

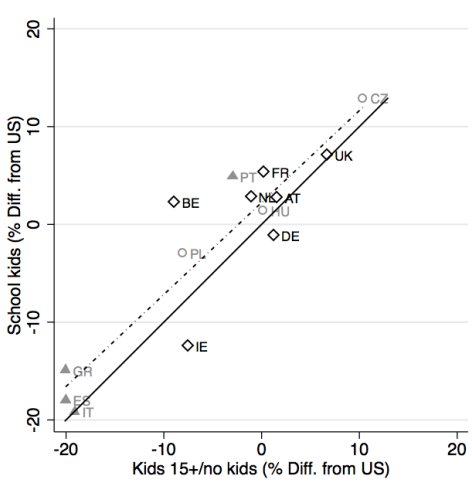
Figure W.2: Differences Relative to US by Presence of Children

Employment Rates

(a) Preschool vs. Kids 15+/No Kids

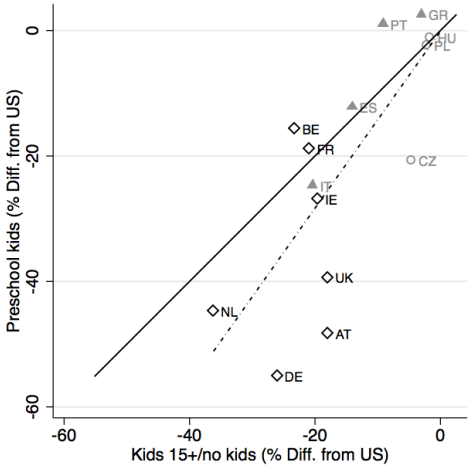


(b) School vs. Kids 15+/No Kids

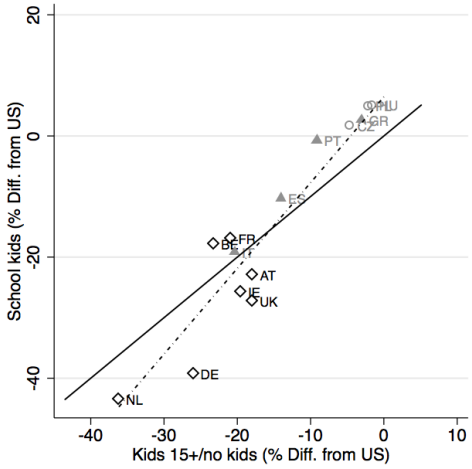


Hours Worked per Employed

(c) Preschool vs. Kids 15+/No Kids



(d) School vs. Kids 15+/No Kids



Note: Employment rate differences relative to the US are shown in percentage points and hours worked per employed differences in percent.

Table W.1: Hours Worked per Person Differences to US by Presence of Children

Region	All women	Preschool kids	School kids	Kids 15+/no kids
Eastern Europe	−2.6	−40.6	9.2	−1.8
Scandinavia	–	–	–	–
Western Europe	−26.4	−31.0	−26.6	−24.1
Southern Europe	−30.9	−22.1	−30.9	−34.7
Europe	−21.9	−31.2	−19.3	−21.4

Note: Portugal is excluded from the table.

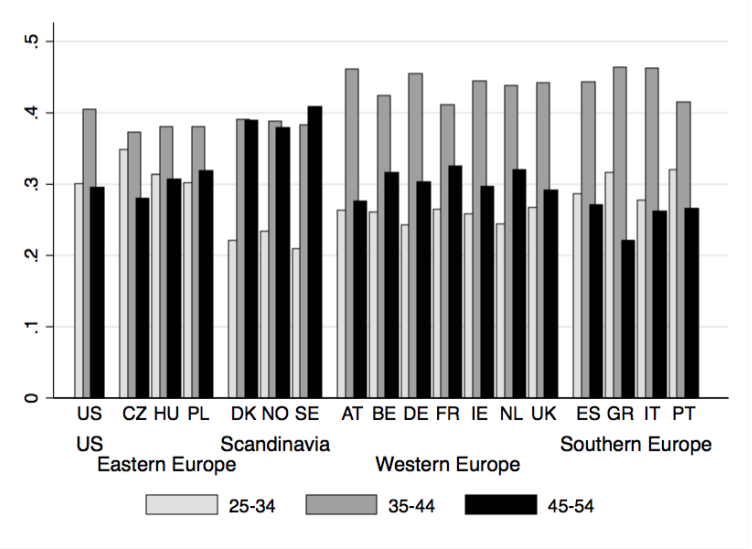
out. For hours worked per employed in panel (c), the country-ordering is similar for both groups of women, but the differences for mothers with preschool children relative to the US are substantially larger in four of the seven Western European countries and the Czech Republic than the differences for women without children younger than 15.

While we do not observe children for Scandinavian countries in the ELFS, we can compute employment rates of Scandinavian women by presence of children in the EU-SILC. While mothers of children 15 or older or women without children have 7 percentage points higher employment rates than the respective group in the US on average across the three Scandinavian countries, the differences for women with preschool or school children are larger with on average 18 and 14 percentage points higher employment rates. Thus, Scandinavian mothers feature relatively high employment rates.

Table W.1 shows the resulting differences in hours worked per married woman for the country group averages. In this table, we exclude Portugal from the Southern European average, as in the main results table 4 in the paper. The general pattern that hours worked differences to the US are larger for Western and Southern Europe than for Eastern Europe holds for both women with school children and for women without children younger than 15 years. For women with preschool children, however, the largest differences to the US occur in Eastern Europe, driven by the very low employment rates in the Czech Republic and Hungary. Note that the Eastern European countries are also the ones with the smallest share of women with preschool children by international comparison. In Western Europe, women with preschool children show similar labor supply patterns than the other two groups of women (always relative to the US counterparts), but in Southern Europe their hours worked are somewhat higher.

To summarize, women with preschool children often behave differently from other women in a given country in a non-systematic way (least so in Western Europe), which points to factors like child care and maternity leave policies playing a role. Since the group of women with preschool children is however relatively small, these differences are not the major determinant of the international differences of married women's labor supply.

Figure W.3: Age Group Composition



W.2 Married Womens' Labor Supply by Age

Figure W.3 shows the age composition in our sample, focusing on three different age groups, namely ages 25-34, 35-44, and 45-54. The share of the youngest group is somewhat lower in Scandinavia and Western Europe, and the share of the oldest group somewhat larger in Scandinavia. This stems from the different sample selection criteria we have to apply to Scandinavia in the ELFS because we cannot identify the age of the spouse. In the EU-SILC we know the age of the spouse also for Scandinavia. The share of 45-54 year olds in Scandinavia falls from 40 percent to 32 percent, which is in line with the other countries, if we also condition on the age of the spouse. Whether we condition on the age of the spouse in Scandinavia or not has however almost no effect on married women's employment rates in the EU-SILC, which are 84.3 percent in the former case and 83.6 percent in the latter (see Online Appendix Section W.3.2).

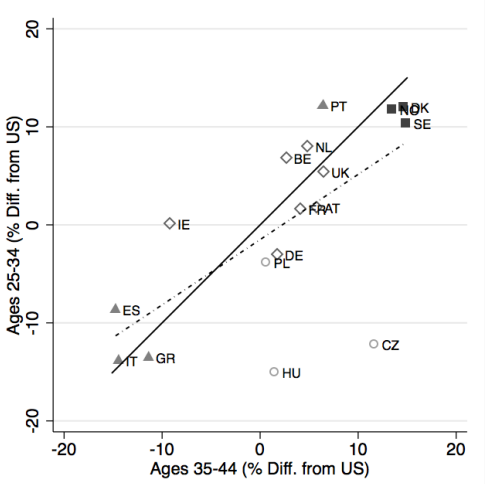
For each of the three age groups, Figure W.4 shows the deviations of both margins (employment rate in panels (a) and (b), and hours worked per employed in panels (c) and (d)) from the respective age group in the US, contrasting young and middle in panels (a) and (c), as well as old and middle in panels (b) and (d). Panel (d) shows that the hours worked per employed differences to the US are almost exactly the same for the old and the middle aged groups. The young show systematically larger hours worked per employed differences to the US than the middle aged, yet the magnitudes are small. Regarding the employment rate, the difference to the US is larger for older women than for women from the middle age group, possibly indicating cohort effects. This is especially true for Southern Europe. For the young vs. the middle aged, there are no systematic differences, but young women in Hungary and the Czech Republic feature very low employment rates. This is mostly due to the behavior of women with preschool children, as we have shown in the previous subsection (W.1).

Table W.2 summarizes what these differences add up to in terms of hours worked per person. Denmark and Portugal are excluded from the respective country group averages in this table, as in the main results table 4 in the paper. Indeed, hours worked per person differences to the US show some heterogeneity across age groups except for Western Europe. Still, the major pattern is preserved: Differences to the US are larger for Southern and Western Europe than for Scandinavia and Eastern Europe for each age group. Differences

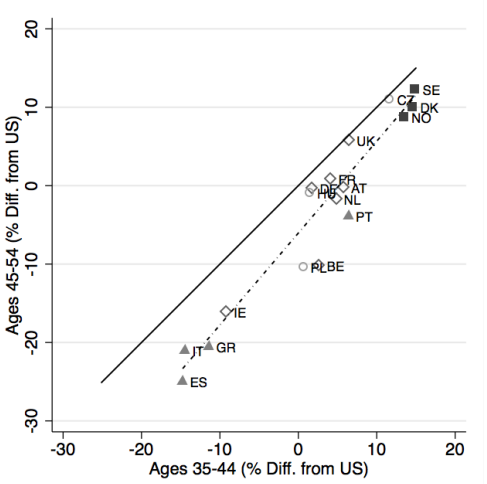
Figure W.4: Age Group Differences Relative to US

Female Employment Rate

(a) Young vs. Middle

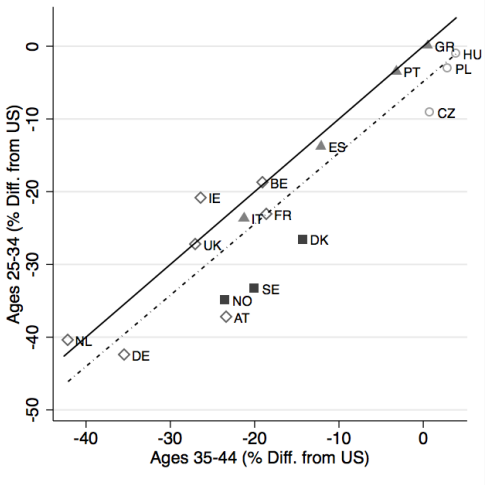


(b) Old vs. Middle

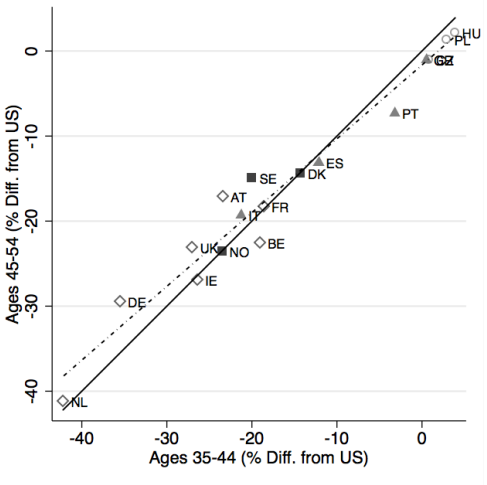


Female Hours Worked per Employed

(c) Young vs. Middle



(d) Old vs. Middle



Note: Employment rate differences relative to the US are shown in percentage points and hours worked per employed differences in percent.

Table W.2: Hours Worked per Person Differences to US by Age

Region	All women	Ages 25-34	Ages 35-44	Ages 45-54
Eastern Europe	−2.6	−19.4	9.1	0.5
Scandinavia	−8.9	−23.1	−6.0	−7.8
Western Europe	−26.4	−26.9	−24.9	−28.7
Southern Europe	−30.9	−28.5	−27.8	−37.6
Europe	−20.2	−25.2	−16.2	−21.9

Note: Denmark and Portugal are excluded from the table.

between the European country groups are generally smallest for the youngest age group. In the previous Section (W.1), we show differences by the presence of preschool children, which are relatively large in Eastern Europe and relatively small in Southern Europe, each time compared to women with school age or older children. This drives the smaller differences across European country groups for the younger age group.

Table W.3: Data Sources

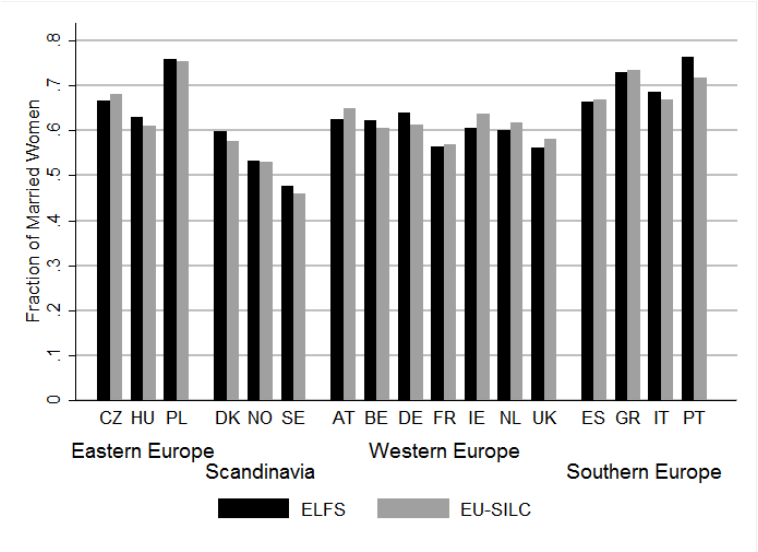
Variables	Countries	Data Source
Labor supply measures	US	CPS
	Germany	Microcensus
	Scandinavia	ELFS
	Remaining countries	ELFS
Fraction of married couples by educational match	US	CPS
	Germany	Microcensus
	Scandinavia	EU-SILC
	Remaining countries	ELFS
Children by educational match	US	CPS
	Germany	Microcensus
	Scandinavia	EU-SILC
	Remaining countries	ELFS
Wages by own education	US	CPS
	Germany	EU-SILC
	Scandinavia	EU-SILC
	Remaining countries	EU-SILC
Statutory Labor Income Tax Codes	All countries	OECD Taxing Wages modules
	US prior to 2001	NBER TaxSim module
Consumption Tax Rates	Hungary	OECD
	All remaining countries	McDaniel (2012)

W.3 EU-SILC

Individuals in our model differ by their gender and education level, which map one to one into differences in wages. Since the ELFS neither includes information on wages nor on earnings, we use the EU Statistics on Income and Living Conditions (EU-SILC) to calculate wages by gender and education. The EU-SILC is set up to provide comparative statistics on income and social inclusion in the European Union (EU), and has both a cross-sectional and longitudinal component.[31] The general sampling criteria are the same as for the ELFS. The sample size is however much smaller, ranging from 7,038 respondents aged 15-64 in the Czech Republic in 2005 to 41,032 respondents in Italy in 2004. Moreover, the survey covers only a subperiod of the years used in our analysis, with the first available year being 2004. Most importantly, the EU-SILC does not contain any information on actual hours worked, but only on usual hours worked. In Bick et al. (2015), we explain that actual hours are however crucial to construct internationally comparable annual hours worked measures. Table W.3 lists the data sources for all variables used in the paper and thus makes it clear when we rely on the EU-SILC.

[31] For details see http://ec.europa.eu/eurostat/web/income-and-living-conditions/overview.

Figure W.5: Female Marriage Rates



W.3.1 Comparison of EU-SILC and ELFS Data

Since we impute wages from the EU-SILC into the ELFS, it would be reassuring to know that both data sets look similar along important dimensions. To analyze this, we first compare marriage rates of women aged 25 to 54 between both data sets. Then, focusing on women who fit our sample selection criteria, we compare the educational composition and the presence of children in households between both data sets. Last, we analyze whether employment rates and usual hours worked are similar across both data sets. In these comparisons, we always use for the ELFS only the sample years for which the EU-SILC is available.

Figure W.5 shows marriage rates of women aged 25 to 54 in the ELFS and the EU-SILC. These are indeed quite similar, with the largest deviation of 5 percentage points coming from Portugal.

Table W.4 shows the educational composition of women aged 25 to 54 who satisfy our sample selection criteria in both data sets. Note that here and in the next two tables, we use for Scandinavia the sample selection criteria that we can apply in the ELFS, i.e. we include all married women aged 25 to 54 with non-missing education. For all other countries, the additional selection criteria of no more than two married individuals in the household, husband's age between 25 to 54, and husband's education non-missing are also applied in both data sets. Again, one can see that differences are generally small, which is also the case for the educational composition of men (not shown here). A notable exception is the UK, where the share of low educated is 15.6 percentage points higher in the ELFS than in the EU-SILC, and the share of medium educated is 12.5 percentage points lower. We investigate the reason for this difference. As described in footnote 16, the ELFS defines low education as ISCED levels 0 to 2, medium as ISCED levels 3 and 4, and high as ISCED levels 5 and above. There is however one exception to this: the subcategory 3c "upper secondary education shorter than two years" is defined as low education. While in the earlier sample years the ELFS only provides the three main education categories low, medium, and high, in the later sample years we can also observe the number of observations in each subcategory. The category 3c "upper secondary education shorter than two years" indeed plays a significant role in the UK, but in no other sample country. The EU-SILC seems to place this category into upper secondary and therefore medium education. Unfortunately the EU-SILC does not provide the finer subcategories, which would have allowed

Table W.4: Fraction of Women by Educational Level in ELFS and EU-SILC

	Low		Medium		High	
Country	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC
Czech Republic	7.6	7.7	78.9	78.1	13.6	14.3
Hungary	19.0	18.6	60.4	59.7	20.6	21.7
Poland	8.8	8.4	68.7	68.8	22.5	22.8
Mean	*11.8*	*11.5*	*69.3*	*68.8*	*18.9*	*19.6*
Denmark	20.2	20.2	43.0	44.5	36.8	35.3
Norway	15.4	14.3	44.8	46.8	39.8	38.9
Sweden	14.3	9.0	46.5	49.0	39.2	42.0
Mean	*16.7*	*14.5*	*44.8*	*46.7*	*38.6*	*38.8*
Austria	22.6	21.9	62.2	61.0	15.2	17.1
Belgium	25.9	19.1	37.5	37.6	36.6	43.3
France	27.8	24.6	42.4	45.1	29.8	30.3
Germany	17.0	13.0	62.3	63.5	20.6	23.5
Ireland	24.6	27.6	42.6	38.8	32.8	33.6
Netherlands	25.9	24.7	46.9	46.4	27.2	29.0
United Kingdom	26.6	11.0	39.4	51.9	34.0	37.1
Mean	*24.3*	*20.3*	*47.6*	*49.2*	*28.0*	*30.5*
Greece	31.5	32.3	45.6	42.9	22.9	24.8
Italy	43.6	44.4	43.3	42.3	13.1	13.3
Portugal	69.4	71.4	14.9	13.7	15.7	14.9
Spain	46.1	45.6	23.1	23.3	30.8	31.1
Mean	*47.6*	*48.4*	*31.7*	*30.6*	*20.6*	*21.0*

us to categorize individuals with the ISCED level 3c as low educated. As a consequence, we have to live with this discrepancy. As Table 2 shows, the educational shares in the UK (from the ELFS) and also the education gradient in wages (from the EU-SILC) are both in line with other Western European countries. Regarding the sample composition in terms of number of children in the household, presented in Table W.5, differences between the ELFS and the EU-SILC are all minor.

Having shown that the sample composition is similar in both data sets, we now compare employment rates of our selected samples in both data sets. Because actual hours worked per employed are not reported in the EU-SILC, we cannot compare hours worked per employed, but we can compare usual weekly hours worked per employed. The results are shown in Table W.6. Overall, labor supply measures in both data sets are very similar. Employment rate differences between both data sets are typically smaller than 3 percentage points, with the exception of the male employment rate in Ireland, which is 3.6 percentage points higher in the ELFS than in the EU-SILC, and female employment rates in the Netherlands, Austria, and Hungary, where the difference between both data sets amounts to 8.2, 7.6, and 5.5 percentage points, respectively. Usual weekly hours worked per employed differences between both data sets never exceed 3 hours for women, but do so for men in Hungary, Norway, Belgium, and Ireland. For the cases of larger differences,

Table W.5: Children Distribution in ELFS and EU-SILC

	0 Kids		1 Kids		2 Kids		3 Kids		4 Kids	
Country	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC
Czech Republic	26.5	27.4	30.0	31.6	36.1	34.7	6.0	5.2	1.4	1.0
Hungary	28.0	25.2	31.1	32.4	29.3	30.5	9.1	9.2	2.5	2.7
Poland	24.3	25.6	33.4	33.7	29.7	29.6	8.9	8.1	3.7	2.9
Mean	*26.3*	*26.1*	*31.5*	*32.6*	*31.7*	*31.6*	*8.0*	*7.5*	*2.5*	*2.2*
Denmark	–	28.7	–	22.3	–	34.5	–	11.8	–	2.7
Norway	–	26.0	–	22.0	–	33.5	–	15.0	–	3.5
Sweden	–	25.0	–	23.3	–	35.6	–	12.6	–	3.5
Mean	–	*26.6*	–	*22.5*	–	*34.5*	–	*13.1*	–	*3.2*
Austria	26.2	28.1	28.9	29.0	32.5	30.9	9.7	9.6	2.7	2.4
Belgium	26.6	27.1	26.2	26.2	31.7	29.1	11.6	13.2	3.9	4.3
France	22.9	21.1	26.7	27.2	34.2	37.4	12.8	11.5	3.5	2.8
Germany	30.4	28.9	29.6	27.6	30.1	32.6	7.9	9.0	2.1	1.9
Ireland	18.8	16.7	22.3	24.8	31.9	33.4	18.5	18.3	8.5	6.8
Netherlands	22.7	22.7	22.6	22.5	38.1	36.8	13.2	13.6	3.4	4.4
United Kingdom	26.9	28.4	24.4	25.7	33.8	33.5	11.1	9.8	3.7	2.7
Mean	*24.9*	*24.7*	*25.8*	*26.1*	*33.2*	*33.4*	*12.1*	*12.2*	*4.0*	*3.6*
Greece	26.6	23.2	29.5	29.9	35.3	43.6	7.1	2.5	1.6	0.7
Italy	24.0	23.4	35.8	35.8	33.3	34.0	6.0	6.1	0.9	0.7
Portugal	25.2	23.4	41.1	43.2	27.7	28.9	4.8	3.8	1.1	0.6
Spain	23.4	26.6	34.6	34.7	34.8	34.6	5.8	3.4	1.3	0.6
Mean	*24.8*	*24.2*	*35.3*	*35.9*	*32.8*	*35.3*	*5.9*	*4.0*	*1.2*	*0.7*

we confirm that these differences are already present when we focus on the total population aged 15 to 64, i.e. they are unrelated to our sample selection criteria.

W.3.2 Comparison of Different Scandinavian Sample Selection Criteria in EU-SILC

Since we are missing household identifiers for the Scandinavian countries in the ELFS, we have to recur to slightly different sample selection criteria than for the other countries. Specifically, while we exclude women living with more than 1 other married adult in the household, with a husband younger than 25 or older than 54, or with a husband whose education status is missing in all other countries, we cannot apply these exclusion restrictions for Scandinavia. However, in the EU-SILC we can compare the Scandinavian samples when these additional selection criteria are applied and when they are not applied. We do that in this subsection.

Table W.7a shows the age composition of the Scandinavian sample in the EU-SILC if all married women aged 25 to 54 (whose education status is non-missing) are included in the columns "No selection", and the age composition if we focus on married women aged 25 to 54 who also fulfill the additional three selection criteria (no more than 2 married individuals in household, husband aged 25-54, husband's education status non-missing) in the columns "Selection". As the table shows, the non-selected sample is on average older than the selected one: the share of 45 to 54 year old married women is on average 40 percent in the non-selected sample, and 32 in the selected sample. This difference arises because many older women are

Table W.6: Employment Rates and Usual Hours in ELFS and EU-SILC

	ER male		UHWE male		ER female		UHWE female	
Country	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC	ELFS	EU-SILC
Czech Republic	94.9	92.6	44.5	46.6	73.4	70.4	39.5	40.7
Hungary	85.1	84.6	36.5	43.8	66.4	71.9	37.6	39.8
Poland	84.7	85.5	44.4	46.4	68.8	69.3	38.8	39.9
Mean	*88.2*	*87.6*	*41.8*	*45.6*	*69.5*	*70.5*	*38.6*	*40.1*
Denmark	93.6	93.5	40.6	41.1	84.5	83.7	34.7	35.2
Norway	92.8	94.1	38.8	42.5	82.9	84.5	31.6	33.6
Sweden	91.5	91.5	40.6	41.8	84.0	82.6	35.1	35.7
Mean	*92.6*	*93.0*	*40.0*	*41.8*	*83.8*	*83.6*	*33.8*	*34.8*
Austria	92.6	92.8	44.6	44.2	74.7	67.1	31.0	31.7
Belgium	91.0	93.1	37.7	44.6	72.3	73.4	29.8	32.7
France	92.7	93.6	42.4	43.4	74.6	75.7	34.0	34.2
Germany	91.1	91.7	41.6	42.7	72.2	69.4	27.2	26.5
Ireland	92.2	88.6	35.7	44.4	63.8	61.9	28.6	29.2
Netherlands	94.6	95.6	40.1	40.8	77.2	69.0	22.9	25.3
United Kingdom	92.3	90.9	44.1	44.8	76.4	77.8	30.4	31.8
Mean	*92.4*	*92.3*	*40.9*	*43.6*	*73.0*	*70.6*	*29.1*	*30.2*
Greece	94.4	94.4	44.9	47.0	58.2	59.2	39.1	38.1
Italy	92.4	93.2	41.3	43.5	55.7	54.7	32.3	34.4
Portugal	92.3	92.1	42.1	44.3	76.5	77.4	38.1	39.6
Spain	91.3	92.5	42.4	44.5	58.9	58.5	34.6	36.1
Mean	*92.6*	*93.0*	*42.7*	*44.8*	*62.3*	*62.4*	*36.0*	*37.1*

married to men older than 54, and thus not part of the selected sample.

As Table W.7b shows, this different age composition in the two samples has however almost no effect on the educational composition, which is quite similar between both samples in all three Scandinavian countries. Moreover, and even more reassuringly, the employment rates and usual hours worked for both men and women are virtually unchanged whether the stricter sample selection criteria are applied or not, as Table W.7c shows. Therefore, we conclude that the application of slightly different sample selection criteria in the ELFS for Scandinavian countries has likely no major effect on our labor supply measures.

Table W.7: Effect of Scandinavian Sample Restrictions in EU-SILC

(a) Age Composition

	Age: 25-34		Age: 35-44		45-54	
Country	No selection	Selection	No selection	Selection	No selection	Selection
Denmark	20.7	24.2	39.8	45.3	39.5	30.5
Norway	20.5	22.4	39.4	44.4	40.1	33.1
Sweden	21.4	24.8	37.6	43.4	40.9	31.8
Mean	*20.9*	*23.8*	*39.0*	*44.4*	*40.2*	*31.8*

(b) Educational Composition

	Low		Medium		High	
Country	No selection	Selection	No selection	Selection	No selection	Selection
Denmark	20.2	17.1	44.5	46.1	35.3	36.8
Norway	14.3	13.5	46.8	46.3	38.9	40.2
Sweden	9.0	7.6	49.0	49.1	42.0	43.3
Mean	*14.5*	*12.7*	*46.7*	*47.1*	*38.8*	*40.1*

(c) Labor Supply Measures

	ER male		**UHWE male**		**ER female**		**UHWE female**	
Country	No selection	Selection	No selection	Selection	No selection	Selection	No selection	Selection
Denmark	93.5	94.2	41.1	41.1	83.7	84.7	35.2	35.3
Norway	94.1	94.8	42.5	42.4	84.5	85.2	33.6	33.6
Sweden	91.5	92.1	41.8	41.8	82.6	83.1	35.7	35.6
Mean	*93.0*	*93.7*	*41.8*	*41.8*	*83.6*	*84.3*	*34.8*	*34.8*

Table W.8: Availability of Annual and Monthly Earnings in EU-SILC

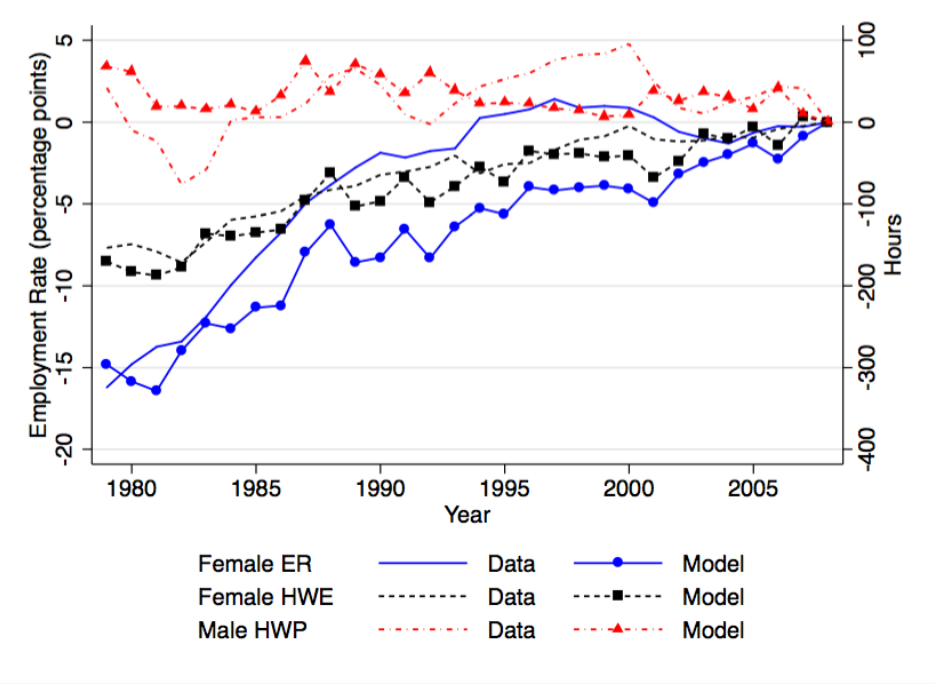
Country	Monthly Earnings	Annual Earnings
Czech Republic	—	2005 - 2008
Hungary	2006 - 2008	2005 - 2008
Poland	2005 - 2008	2005 - 2008
Denmark	—	2004 - 2008
Norway	2004 - 2006	2004 - 2008
Sweden	—	2004 - 2008
Austria	2004 - 2008	2004 - 2008
Belgium	2004 - 2005	2004 - 2008
Germany	—	2005 - 2008
France	—	2004 - 2008
Ireland	2004 - 2008	2004 - 2008
Netherlands	—	2005 - 2008
United Kingdom	2005 - 2008	2005 - 2008
Greece	2004 - 2008	2007 - 2008
Italy	2004 - 2008	2007 - 2008
Portugal	2004 - 2008	2007 - 2008
Spain	2004 - 2008	2006 - 2008
United States	2001 - 2008	—

W.3.3 Construction of Wages in the EU-SILC

In general, the EU-SILC provides two types of earnings, namely earnings in the previous month and annual earnings in the previous year. However, the actual availability varies across countries and years, as is summarized in Table W.8. Since each earnings variable is available for at most four or five years (depending from which year onwards the EU-SILC is available for a country), we pick the earnings variable per country for which more years are available. In case of a tie (i.e. in Poland, Austria, Ireland, and the UK), we choose monthly earnings because we also use monthly earnings for our benchmark country, the US, for which we only have a monthly earnings measure. A further disadvantage of the annual earnings is that they refer to the year prior to the interview, while usual hours refer to the current year. To obtain an hourly wage, we take monthly earnings or annual earnings divided by 12, and divide the respective monthly earnings by the product of weekly usual hours (which are available in all years for which we have earnings information) and average monthly weeks worked in a country (52 minus the weeks lost due to vacation days and public holidays, divided by 12).

For years prior to the first sample year, we extrapolate earnings based on country-specific OECD wage growth rates.

Figure W.6: Time-Series Predictions for the US



W.4 Time-Series Performance of the Model

While the goal of the paper is to use the model to evaluate in how far differences in taxes, wages, and the educational composition can explain cross-country differences in hours worked of married women and men, we can evaluate the predictive power of the model also by analyzing its performance in replicating the US time series of labor supply of married couples. To do that, we generate the US-specific model inputs back to the year 1979 and plug them into the model, keeping the preference parameter values fixed.[32] Figure W.6 shows employment rate differences (in percentage points, left y-axis) and hours differences (in levels, right y-axis) for each year relative to the year 2008 in the data and the model. The model correctly predicts hardly any change in hours worked of married men over the period of three decades. While hours worked per married man were 42 hours higher in 1979 than in 2008, the model predicts a difference of 68 hours. For married women, the model captures both the increase in the employment rate and in hours worked per employed very well. From 1979 to 2008 the employment rate increases by 16 percentage points in the data and by 15 percentage points in the model, with some deviations between data and model in the 1990s. Especially the increase in the 1980s, when several tax reforms favored married women (see also Kaygusuz (2010)), is captured correctly by the model. The change in hours worked per employed married woman over time of 153 hours is with 170 hours also very closely replicated by the model.

[32]For the labor income taxes in the US, we can use the NBER TaxSim module, which in contrast to the OECD Taxing Wages modules goes back to the 70s. As in the OECD modules, the state tax is taken from Michigan, which is close to the average across the US states, and the city tax is taken from Detroit.

W.5 Alternative Specifications

We carry out five robustness checks. In each of them, we recalibrate the model to match the targeted US moments.

In a first robustness check, we focus on households without children. As discussed in Section 3 and Section W.1, differences in hours worked relative to the US are not primarily driven by the different labor force behavior of women with children. In Table W.9a, we present the data and model results focusing on women without children. The data change a bit when focusing on childless married men and women, but the overall patterns remain the same. The model predicts a bit smaller differences to the US for married women in Eastern and Southern Europe when we simulate only households without children (to draw the comparison to the benchmark case with children, see Table 4). Thus, the model fit remains similar, and even slightly improves for married women in Western and Eastern Europe.

A second robustness check changes the curvature parameter ϕ on the disutility of working only for women to 1 and 1.5, respectively, thereby increasing the female labor supply elasticity. We leave the value for men at 0.5. Table W.9b shows the results. Changing the female elasticity has almost no effect on male labor supply. For married women, increasing the elasticity increases the effects of any cross-country factors, and therefore increases the difference between hours worked in the US and Europe. This leads to a slightly worsened fit for Scandinavia and Eastern Europe, but an improved fit for Western and Southern Europe.

In a third robustness check, we vary the redistribution scheme of government revenues. In the benchmark analysis, we assume full redistribution, while here we consider two alternative redistribution schemes: first, the alternative extreme of no redistribution, i.e. $\lambda^i = 0$, or secondly, a specification similar to the one used by Prescott (2004), namely setting λ^i equal to 1 minus twice the share of expenditures on military from all government expenditures. Redistribution matters in the analysis, because the redistribution of tax revenues in a lump-sum fashion to households provides important income effects that lower the incentives to work. As Table W.9c shows, going from the benchmark redistribution to the country-specific redistribution scheme used by Prescott (2004) decreases hours worked for men and women in Europe relative to the US more than in the benchmark. Due to the higher military expenses in the US than in Europe, a lower share of government revenues is redistributed in the US, which makes the income effect of high taxes via transfers relatively larger in Europe. Comparing the Prescott redistribution scenario to the benchmark scenario, the fit is always slightly better than in the benchmark scenario for married men, as well for married women in Southern and Western Europe, but is slightly worse for married women in Scandinavia and Eastern Europe. A scenario with no redistribution of government revenues, on the other hand, leads to significantly higher predicted hours worked in Europe. This lowers the model-predicted US-European difference and even turns it positive for married women in Eastern Europe and Scandinavia. Two effects lead to higher predicted hours in Europe in the model without transfers, compared to the benchmark model: first, in the benchmark scenario, the income effect from redistribution is on average larger in Europe than in the US due to higher taxes and therefore higher government revenues in Europe. This leads to larger disincentive effects on labor supply from redistribution in Europe than in the US, which are absent in the case of no redistribution. Second, given our assumption of log utility, substitution and income effects cancel out in the absence of transfers. As a consequence, cross-country differences in consumption tax rates and average income tax rates do not affect household choices anymore. Hence, disincentive effects from these higher European tax rates are no longer present in the case of no redistribution. All results are now only driven by cross-country differences in the educational composition, gender-education-wage premia, and the tax structure.

The fourth robustness check uses raw observed female wages rather than the wages computed with the Heckman correction for women as model inputs. Wages for men are unchanged. Results are in Table W.9d. Predicted hours for European married men decrease slightly when raw wages are used as model inputs,

Table W.9: Robustness Checks - Hours Worked per Married Person (% difference to the US)

(a) Childless Couples

	Men		Women	
Region	**Data**	**Model**	**Data**	**Model**
Eastern Europe	−14.4	−4.6	−4.9	−6.9
Scandinavia	–	–	–	–
Western Europe	−9.8	−6.8	−22.1	−18.1
Southern Europe	−8.9	−1.4	−32.4	−17.0
Europe	−10.7	−5.0	−20.5	−15.2

(b) Different Female Elasticities

	Men				Women			
Region	**Data**	**Benchmark**	$\phi^f = 1.0$	$\phi^f = 1.5$	**Data**	**Benchmark**	$\phi^f = 1.0$	$\phi^f = 1.5$
Eastern Europe	−11.1	−5.1	−5.3	−5.0	−2.6	−8.0	−9.9	−10.7
Scandinavia	−17.1	−10.3	−9.9	−9.6	−8.9	−8.6	−11.1	−12.7
Western Europe	−10.3	−7.4	−7.0	−6.4	−26.4	−17.9	−22.0	−24.2
Southern Europe	−9.4	−2.0	−2.0	−1.7	−30.9	−19.5	−20.6	−21.6
Europe	−11.2	−6.2	−6.1	−5.6	−20.2	−15.0	−17.9	−19.5

(c) Different Redistribution of Government Revenues

	Men				Women			
Region	**Data**	**Benchmark**	$\lambda^{Prescott}$	$\lambda = 0$	**Data**	**Benchmark**	$\lambda^{Prescott}$	$\lambda = 0$
Eastern Europe	−11.1	−5.1	−6.2	−1.8	−2.6	−8.0	−11.2	0.2
Scandinavia	−17.1	−10.3	−10.9	−5.6	−8.9	−8.6	−11.1	11.8
Western Europe	−10.3	−7.4	−8.4	−3.3	−26.4	−17.9	−20.9	−8.9
Southern Europe	−9.4	−2.0	−3.0	−0.5	−30.9	−19.5	−22.5	−18.6
Europe	−11.2	−6.2	−7.2	−2.7	−20.2	−15.0	−17.9	−6.2

(d) Raw Wages

	Men			Women		
Region	**Data**	**Benchmark**	**Raw Wages**	**Data**	**Benchmark**	**Raw Wages**
Eastern Europe	−11.1	−5.1	−5.2	−2.6	−8.0	−6.4
Scandinavia	−17.1	−10.3	−9.8	−8.9	−8.6	−11.1
Western Europe	−10.3	−7.4	−7.7	−26.4	−17.9	−16.8
Southern Europe	−9.4	−2.0	−4.4	−30.9	−19.5	−7.3
Europe	−11.2	−6.2	−6.9	−20.2	−15.0	−12.0

(e) Wage Heterogeneity within Education Groups

	Men			Women		
Region	**Data**	**Benchmark**	**Wage Heterogeneity**	**Data**	**Benchmark**	**Wage Heterogeneity**
Eastern Europe	−11.1	−5.1	−4.2	−2.6	−8.0	0.0
Scandinavia	−17.1	−10.3	−8.6	−8.9	−8.6	−6.2
Western Europe	−10.3	−7.4	−5.8	−26.4	−17.9	−9.6
Southern Europe	−9.4	−2.0	1.2	−30.9	−19.5	−11.6
Europe	−11.2	−6.2	−4.5	−20.2	−15.0	−7.6

increasing the model fit. For women, predicted hours in Europe are slightly higher in Eastern and Western Europe when raw wages are used, but lower in Scandinavia. Quantitatively, these effects are relatively small, which is however not the case for Southern Europe, where the predicted difference in hours worked of married women to the US falls from 19.5 percent to 7.3 percent. It is not surprising that the largest differences arise for Southern Europe, which has the lowest female employment rates in the data. Therefore, selection into work is the biggest issue for countries in Southern Europe. A positive selection leads to higher observed than offered wages and thus a lower gender wage gap, which can explain the higher predicted hours of Southern European women when observed wages are used as model inputs.

In the last robustness check, we allow for further wage heterogeneity within each education group. Since we impute wages from the EU-SILC, we also impute the distribution of wages from EU-SILC. Similar to Attanasio et al. (2008), we obtain an estimate of wage heterogeneity for men only, and apply the same estimates for women. For each country, we regress male log hourly wages on a set of year and education dummies. We pool individuals from all years 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 states (using five rather than three states has virtually no impact on our results). When individuals are matched into couples, we draw randomly from these states within each education type. Table W.9e shows the main results in the model with wage heterogeneity conditional on education. Compared to the benchmark model, the fit within Europe is similarly good for both men and women: the correlation between hours worked in model and data is 0.55 for married men (compared to 0.44 in the benchmark exercise) and 0.62 for married women (compared to 0.63 in the benchmark exercise). However, with heterogeneity hours in all European regions are higher for both men and women, such that less of the US-Europe hours gap is explained. Only for Eastern European married women does the fit improve. For men, the model that allows for heterogeneity within each education group can explain on average 40 percent of the difference of European hours to the US compared to 55 percent for the model without heterogeneity, and for women on average 38 percent compared to 74 percent. All four model inputs contribute to the higher European hours with heterogeneity, and not surprisingly, wages have the largest effect.

Table W.10: Linear Taxes - Hours Worked per Married Person (% difference to the US)

	Men			Women		
Region	**Data**	**Benchmark**	**Linear Taxes**	**Data**	**Benchmark**	**Linear Taxes**
Eastern Europe	−11.1	−5.1	−6.3	−2.6	−8.0	−17.5
Scandinavia	−17.1	−10.3	−12.0	−8.9	−8.6	−29.4
Western Europe	−10.3	−7.4	−8.4	−26.4	−17.9	−27.0
Southern Europe	−9.4	−2.0	−3.6	−30.9	−19.5	−36.8
Europe	−11.2	−6.2	−7.5	−20.2	−15.0	−27.4

W.6 Linear Taxes

One major novelty of our study is that we use actual non-linear tax systems rather than average marginal tax rates as model inputs in order to predict hours worked. To understand how important this is, we compare our benchmark results to results from the same model using the same model inputs but replacing the non-linear labor income taxes with linear income taxes calculated by McDaniel (2011) as model inputs.[33]

As Table W.10 shows in the first three columns, the fit for married men improves somewhat when we use simple linear labor income taxes. On average 67 percent of the US-Europe gap is explained with linear taxes, compared to 55 percent with non-linear taxes. Since married men make up the majority of the workforce, average labor income tax rates capture their actual income tax rates quite well, and therefore the predictions are not that different whether we use the full schedule of non-linear or simple linear labor income taxes.

For women, however, results are very different whether we use the actual non-linear tax code or simple linear taxes. The model with linear taxes always predicts larger hours worked differences between Europe and the US than the model with the actual non-linear tax code. The countervailing effects of the tax structure are missing when we use simple linear tax rates. As a consequence of the larger predicted differences, the model with linear taxes performs much worse than the benchmark model in replicating female hours worked in Eastern Europe and Scandinavia, where it predicts differences of 18 and 29 percent, respectively, compared to the US. For Western Europe, on the first view it appears that the model with linear taxes performs better than the model with non-linear taxes. However, this comes from the fact that the former sometimes over- and sometimes underpredicts differences to the US, while our benchmark model always underpredicts with the exception of Belgium. Our benchmark model performs significantly better in explaining the variation within Western Europe, creating a correlation of 0.27 between model and data within Western Europe, compared to a correlation of -0.49 when using linear tax rates. Only for Southern Europe does the model with linear tax rates perform better than the benchmark model: as the countervailing effect of separate taxation is missing in the former, it creates larger female hours worked differences to the US than the benchmark model, thereby even overpredicting the differences in the data. Overall, the model with linear tax rates does a worse job in explaining cross-country differences: the correlation in female hours worked per person between data and model amounts to only 0.41 when linear taxes are applied, as compared to 0.63 in the benchmark case.

[33] Prescott (2004), McDaniel (2011) and Ragan (2013) multiply average labor income tax rates (but not social security contribution rates) by a factor of 1.6 in order to convert them into average marginal tax rates. Ohanian et al. (2008) do not multiply the labor income taxes by 1.6.

W.7 Figures

Figure W.7: Fraction of Married Women (Ages 25-54)

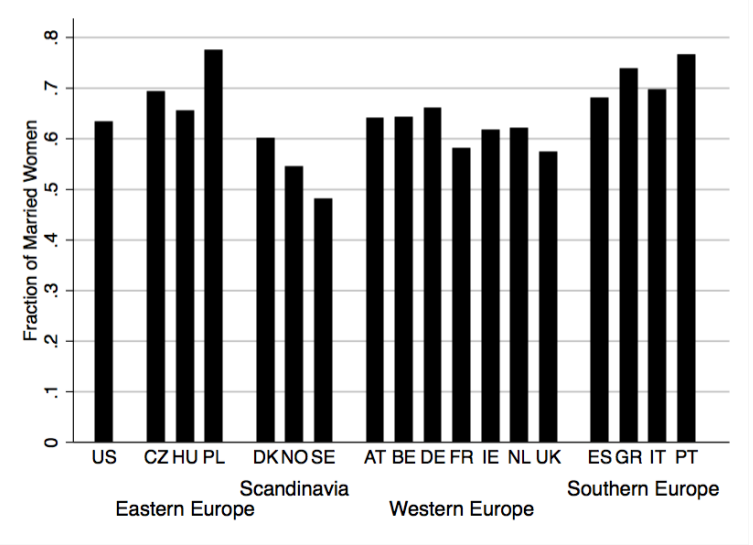


Figure W.8: Employment Rates and Hours Worked per Employed of Married Men (Ages 25-54)

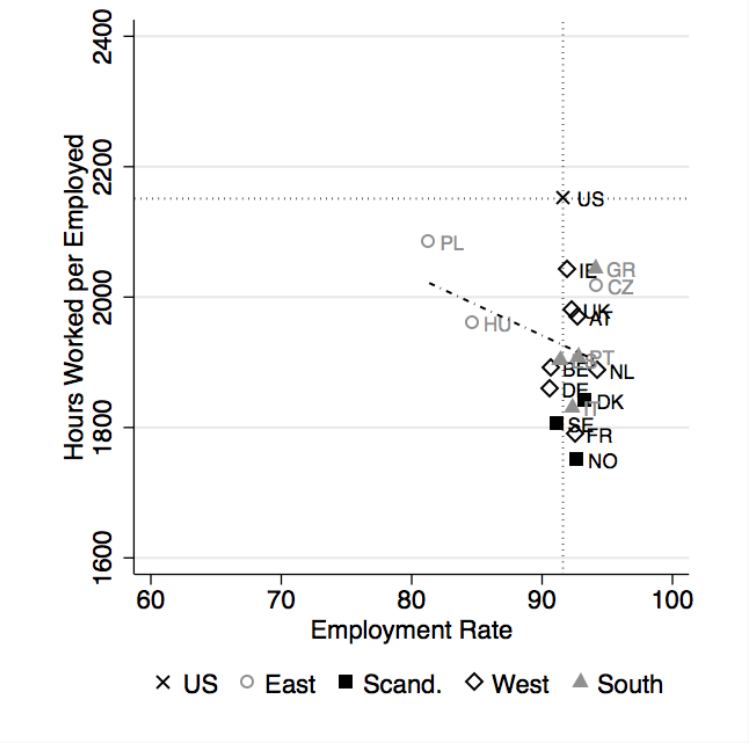
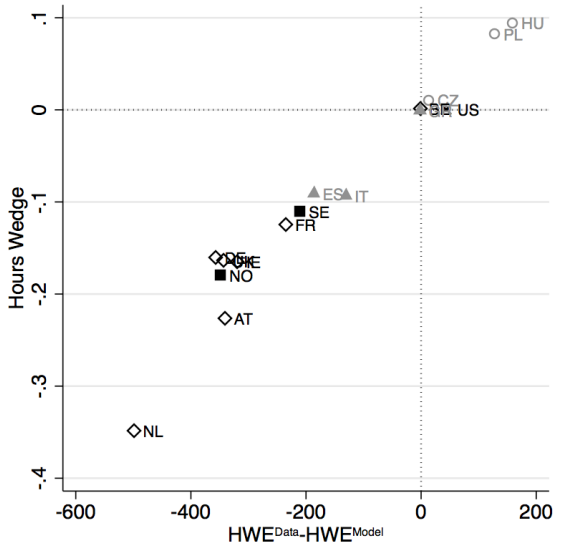
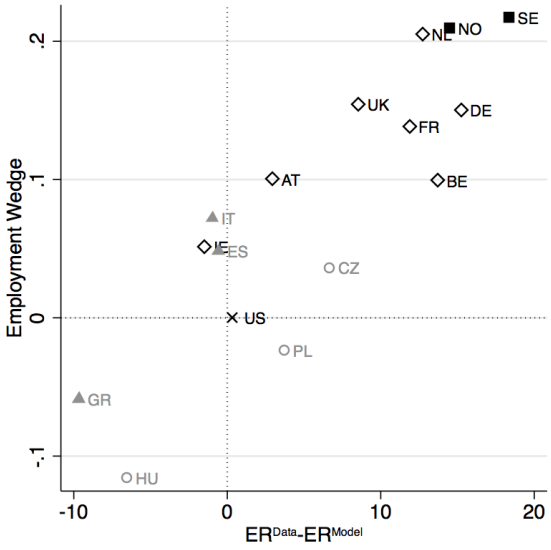


Figure W.9: Country-Specific Wedges and Data-Model Discrepancies

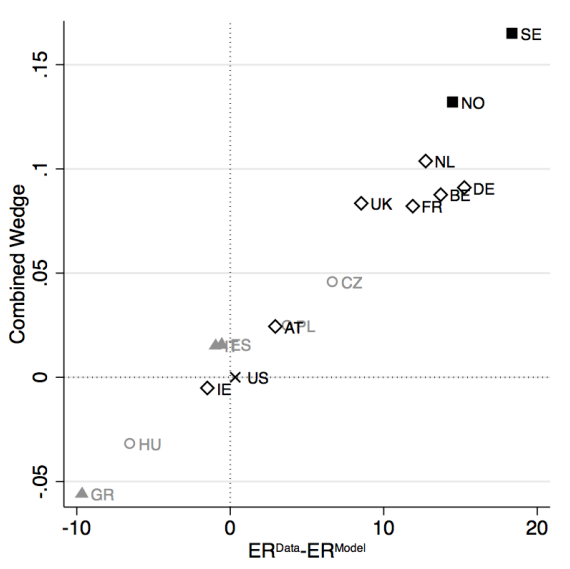
(a) $\hat{\tau}_h$ vs. Female Hours Worked per Employed



(b) $\hat{\tau}_e$ vs. Female Employment Rate



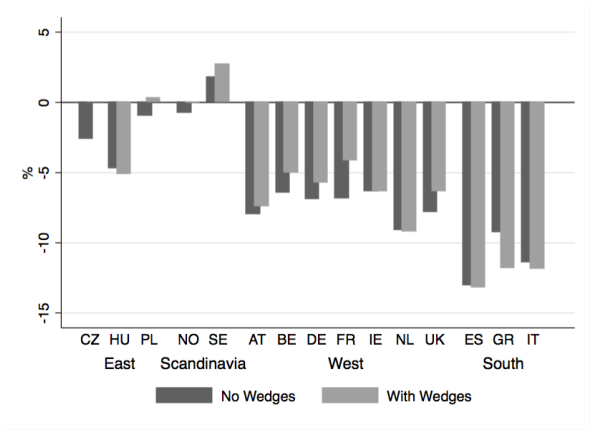
(c) $\hat{\tau}_h \cdot \bar{h}_f + \hat{\tau}_e$ vs. Female Employment Rate



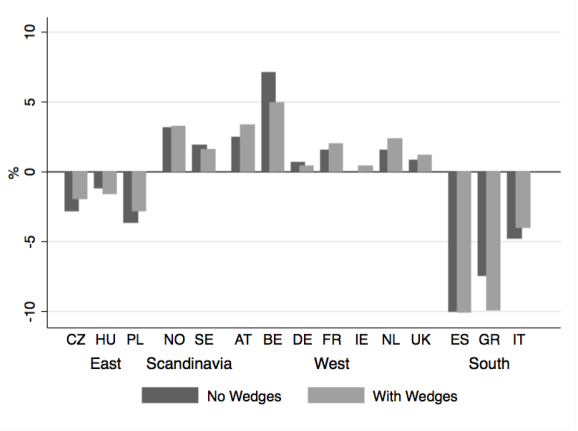
Panel (c) plots for each country the combined wedge on the y-axis, i.e. the employment wedge ($\hat{\tau}_e$) plus the hours wedge ($\hat{\tau}_h$) times the mean female hours worked per employed in the data ($\bar{h}_f$), against the employment rate difference between the predictions of the model without wedges and the data on the x-axis. The combined wedge lines up much more closely with the employment rate difference between the model and the data than the employment wedge alone shown in Panel (b). This is because the hours wedge has an indirect effect on the extensive margin choice, by changing the utility of working a certain amount of hours vis-à-vis the utility of not working.

Figure W.10: Decomposition With and Without Wedges

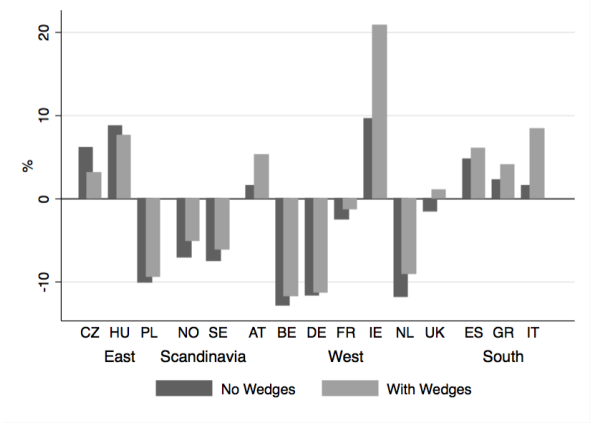
(a) Educational Composition



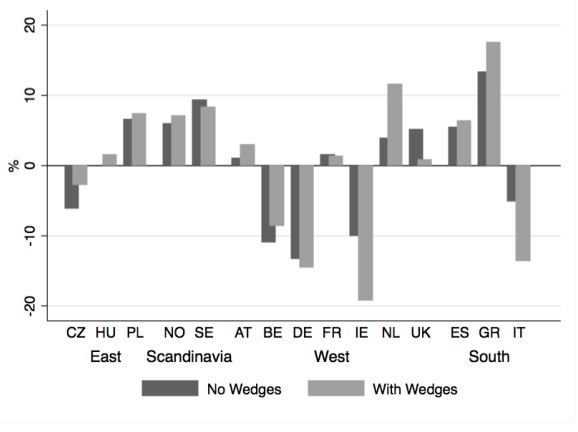
(b) Wages



(c) Average Tax Rate



(d) Tax Structure



Note: This figure shows predicted hours worked per person differences of married women to the US if only one model component is set country-specific in model with and without wedges.

W.8 Tables

Table W.11: Percent of Observations Dropped from the Sample

Country	Men	Women
Czech Republic	8.7	17.2
Hungary	10.0	18.2
Poland	17.4	23.7
Mean	*12.0*	*19.7*
Denmark	0.9	1.0
Norway	2.3	3.1
Sweden	0.6	0.7
Mean	*1.2*	*1.6*
Austria	12.7	20.3
Belgium	11.9	21.4
France	7.9	15.8
Germany	12.0	20.7
Ireland	9.5	15.9
Netherlands	6.4	14.8
United Kingdom	13.0	21.6
Mean	*10.5*	*18.7*
Greece	8.1	22.8
Italy	8.1	19.0
Portugal	10.7	17.7
Spain	9.0	16.7
Mean	*9.0*	*19.1*
United States	11.4	16.1

Table W.12: Labor Supply of Married Men and Women (difference to the US) – All (Raw) and Excluding Unemployed Individuals

	$\mathbf{HWP}_m$		$\mathbf{ER}_f$		$\mathbf{HWP}_f$	
Country	Raw	w/o UE	Raw	w/o UE	Raw	w/o UE
Czech Republic	−3.6	−3.5	2.9	6.0	1.8	5.9
Hungary	−15.8	−14.8	−4.5	−2.9	−4.2	−1.7
Poland	−14.0	−8.4	−4.2	2.7	−5.2	4.5
Mean	*-11.1*	*-8.9*	*-1.9*	*1.9*	*-2.6*	*2.9*
Denmark	−12.8	−13.5	12.9	13.7	−1.3	−0.7
Norway	−17.6	−18.6	11.9	11.6	−13.3	−13.8
Sweden	−16.5	−16.1	13.7	15.0	−4.5	−3.4
Mean	*-15.6*	*-16.1*	*12.8*	*13.4*	*-6.4*	*-6.0*
Austria	−7.3	−7.2	3.3	4.2	−21.5	−20.6
Belgium	−12.9	−12.1	0.0	1.9	−20.0	−17.8
France	−16.0	−14.9	2.7	5.9	−16.5	−12.8
Germany	−14.5	−11.0	0.3	3.9	−34.8	−31.9
Ireland	−4.8	−4.9	−8.6	−8.8	−34.3	−34.1
Netherlands	−9.6	−10.7	3.9	4.1	−38.2	−38.1
United Kingdom	−7.2	−7.8	6.1	6.3	−19.4	−19.3
Mean	*-10.3*	*-9.8*	*1.1*	*2.5*	*-26.4*	*-24.9*
Greece	−2.4	−2.7	−14.4	−11.6	−20.7	−16.3
Italy	−14.1	−14.0	−16.0	−14.6	−39.3	−37.4
Portugal	−10.2	−9.5	5.3	8.3	2.6	6.5
Spain	−11.7	−10.3	−15.8	−12.7	−32.6	−28.3
Mean	*-9.6*	*-9.1*	*-10.2*	*-7.7*	*-22.5*	*-18.9*

Note: For the employment rate we show percentage point differences, and for hours worked per person percent differences.

Table W.13: Male Education Wage Premium, Male Educational Shares, and Correlation Between Education Levels of Husband and Wife

Country	$\frac{w_m^{high}}{w_m^{low}}$	μ_m^{low}	μ_m^{high}	Assortative Matching
Czech Republic	1.99	4.6	15.6	0.46
Hungary	2.43	14.1	16.5	0.55
Poland	2.32	9.8	14.9	0.52
Mean	*2.25*	*9.5*	*15.7*	*0.51*
Denmark	1.36	15.9	31.7	0.37
Norway	1.40	12.7	35.2	0.40
Sweden	1.42	16.8	27.9	0.39
Mean	*1.39*	*15.1*	*31.6*	*0.38*
Austria	1.61	12.7	21.1	0.39
Belgium	1.39	32.0	31.4	0.55
France	1.44	28.1	24.8	0.48
Germany	1.67	11.0	31.3	0.50
Ireland	1.80	35.6	28.7	0.52
Netherlands	1.59	25.5	30.7	0.44
United Kingdom	1.84	22.1	33.0	0.40
Mean	*1.62*	*23.9*	*28.7*	*0.47*
Greece	1.58	38.4	23.3	0.63
Italy	1.71	50.7	10.7	0.54
Portugal	2.67	78.1	9.6	0.62
Spain	1.61	52.0	27.0	0.53
Mean	*1.89*	*54.8*	*17.6*	*0.58*
United States	1.96	9.3	43.3	0.54

Table W.14: Untargeted Moments

	Data	**Model**	$\Delta_{\textbf{Model-Data}}$
Hours Worked per Man			
Low education	1648.6	1949.0	300.4
Medium education	1922.0	1976.1	54.1
High education	2092.2	2043.9	−48.3
Hours Worked per Employed Woman			
Low education	1682.9	1529.9	−153.0
Medium education	1739.8	1660.6	−79.2
High education	1757.6	1770.3	12.7

Table W.15: Wedges and Institutional Factors

Country	$\hat{\tau}_h$	$\hat{\tau}_e$	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Czech Republic	0.02	0.04	11	110	50.8	11.0	0.4	43.9	9.1	3.2
Hungary	0.14	−0.12	13	160	43.9	6.0	0.6	30.7	10.2	2.5
Poland	0.08	−0.02	10	26	100.0	7.0	0.3	31.4	25.4	1.5
Mean	*0.08*	*-0.03*	*11*	*99*	*64.9*	*8.0*	*0.4*	*35.3*	*14.9*	*2.4*
Norway	−0.18	0.21	2	87	42.0	17.0	1.0	5.4	11.3	2.3
Sweden	−0.11	0.22	1	60	63.4	7.0	1.3	1.1	10.8	2.3
Mean	*-0.15*	*0.21*	*2*	*74*	*52.7*	*12.0*	*1.1*	*3.3*	*11.0*	*2.3*
Austria	−0.17	0.10	12	60	68.0	17.0	0.3	42.3	9.7	2.5
Belgium	−0.03	0.10	9	32	40.1	6.0	0.6	22.5	14.4	3.0
Germany	−0.16	0.15	2	58	59.8	14.0	0.4	–	–	2.4
France	−0.12	0.14	4	42	49.5	17.0	1.2	17.7	15.3	2.1
Ireland	−0.13	0.05	8	26	34.7	45.0	0.3	14.6	15.1	0.8
Netherlands	−0.30	0.20	5	42	49.3	13.0	0.6	49.3	8.4	2.0
United Kingdom	−0.16	0.15	14	39	30.9	41.0	0.9	37.3	15.0	2.6
Mean	*-0.15*	*0.13*	*8*	*43*	*47.5*	*21.9*	*0.6*	*30.6*	*13.0*	*2.2*
Spain	−0.06	0.05	15	16	100.0	8.0	0.5	7.7	23.2	1.8
Greece	0.00	−0.06	7	43	54.4	5.0	0.1	46.0	25.4	1.1
Italy	−0.12	0.07	6	48	52.7	–	0.6	41.4	16.0	0.8
Mean	*-0.06*	*0.02*	*9*	*36*	*69.1*	*6.5*	*0.4*	*31.7*	*21.5*	*1.2*
United States	0.00	0.00	16	0	0.0	38.0	0.4	35.0	–	3.4

$\hat{\tau}_h$: Hours Wedge
$\hat{\tau}_e$: Employment Wedge
(1): Part-time generosity rank
(2): Maternity leave: paid weeks
(3): Avg. pay during maternity leave
(4): Net child care costs (% of avg. earnings)
(5): Public child care expenditure (% of GDP)
(6): % of preschool child. in informal care
(7): Avg. hours of informal care for preschool child.
(8): Divorce rate (per 1000 persons per year)