A Sources of individual level data

Structure of Earnings Survey of the European Union (EU-SES). This database is matched employee-employer database that provides administrative-quality data on earnings. The survey is conducted among firms, which report to the statistical office data directly from payroll. Consequently, neither wages nor hours worked are subject to reporting bias. In addition to high quality data, this data source is also characterized by large sample sizes, which make estimates more precise. The data are harmonized at the European level and released every four years. This data source does not have information on household such as children or residence. Marital status is reported for individual workers.

European Community Household Panel Survey (ECHP). This database is provided with annual frequency collected across the EU-15 members between 1994 and 2001. Data on wages and job characteristics are self-reported. This database provides full information on household structure and residence.

European Union Study of Income and Living Conditions (EU-SILC). This database is a follow up survey of ECHP. It has the same data coverage in terms of variables. It is more comprehensive in terms of countries, as the EU was enlarged. The data is provided with annual frequency.

American Community Survey. This is census data for the United States. We use data for 1960, 1970, 1980, annual data for 2000-2008, 2012, and 2016. This is self-reported data. It includes annual wages, annual weeks worked, hours usually worked, individual-level characteristics as well as household-level characteristics. The data is provided by IPUMS.

Census data from IPUMS-International. We use data for Mexico, Israel, Brazil and Canada. Household-level and individual level variables are comprehensively available. We utilize all the available censuses which provide data on wages and hours worked.

Living Standards Measurement Survey was a program operated jointly by the World Bank and national statistical offices around the world. Across countries, the questionnaire focuses on the characteristics of dwelling, poverty indicators, etc. The household roster provides rich data on household structure and individual-level characteristics, whereas the income modules provide data on wages and hours worked. Sample sizes in LSMS are small for some countries, though.

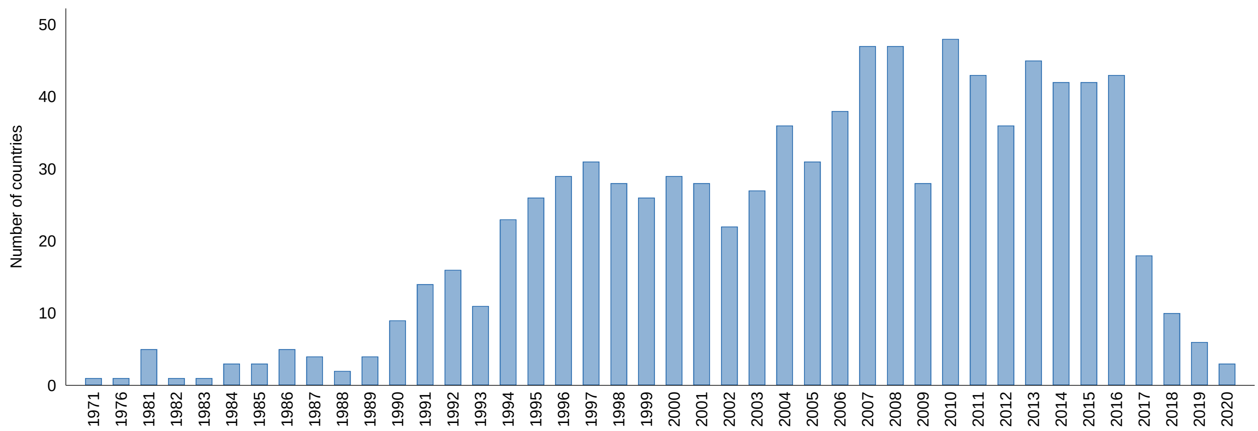
National panels. We acquire access to national longitudinal databases for Canada (Survey of Labor and Income Dynamics, SLID) Germany (Socio-Economic Panel, SOEP), Korea (Korean Labor and Income Panel Study, KLIPS), Russia (Russian Longitudinal Monitoring Survey, RLMS), Sweden (HUS), Ukraine (Ukrainian Longitudinal Monitoring Survey, ULMS) and the United States (Panel Study of Income Dynamics, PSID). All these databases provide rich information on household and individual characteristics, as well as wages and hours worked.

Labor force surveys. National statistical offices collect LFS data routinely, but only in few countries the surveys ask questions about the wages. LFS data are typically self-reported, but sample sizes are large. Unfortunately, this data is distributed at prohibitive charge in many countries. We were able to acquire data for Albania, Argentina, Croatia, France, Italy, Latvia, Poland, Serbia and the United Kingdom. All these databases provide rich information on household and individual characteristics, as well as wages and hours worked.

Household budget survey. National statistical offices often collect HBS data. This data is self-reported, but comprehensive in terms for individual-level characteristics as well as incomes earned. We acquired data for Armenia, Belarus, Georgia, Kyrgyzstan, Latvia, and Uruguay.

International Social Survey Programme (ISSP) is a rich database collected throughout the world since the 1990's. Individual-level characteristics as well as income and hours worked data are self-reported. Sample sizes in ISSP are frequently small. In addition, some databases report wages as categorical variables. Notwithstanding, ISSP is comprehensive both in terms of country coverage and periods covered.

Figure A1: Number of countries across years



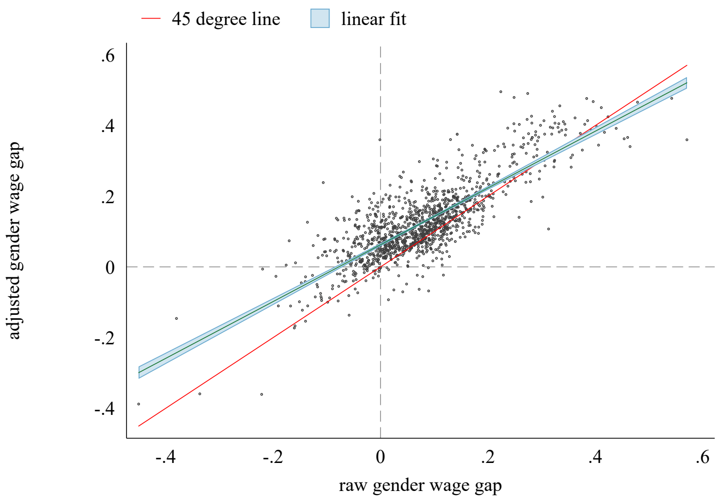
Notes: For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender. If no specification reaches 75% of individuals matched, no estimate from this country-year-data source is included in the analyses.

Table A1: Databases used in this study

Country	Census	EU	HBS/LFS	ISSP	LISSY	LSMS	longitudinal	Total
Albania						4		4
Argentina			13	1				14
Armenia			4					4
Austalia				2	9			11
Austria		18			8			26
Belarus			2					2
Belgium		21		2	19			42
Brazil	3			2	5			10
Bulgaria		8				4		12
Canada	4			1	16			37
Chile					13			13
China				3	2			5
Colombia					5			5
Cote d'Ivore					3			3
Croatia		7	12	3				22
Cyprus		11		1				12
Czechia		11		6	8			25
Denmark		15						15
DominicanRepublic					1			1
Egypt					1			1
Estonia		13			5			18
Finland		10		5	5			20
France		20	30	1	4			55
Georgia					3			3
Germany		18		9	25		32	84
Greece		15			7			22
Guatemala					3			3
Hungary		10		6	8			24
Iceland		2		1	3			6
India	1				2			3
Ireland		16			20			36
Israel	1				19			20
Italy	9	18	6	1	12			46
Japan					3			3
Kyrgyzstan			3			3		6
Latvia		11	4	7				22
Lithuania		8			9			17
Luxembourg		14			9			23
Malta		7						7
Mexico	3			3	15			21
Netherlands		21			9			30
NewZeland				1				1
Norway		5		8				13
Panama					4			4
Paraquay					6			6
Peru					5			5
Philippines				3				3
Poland		11	28	5	4			48
Portugal		20						20
Romania		8						8
Russia				13	4		24	41
Serbia		4	10		1	3		18
Slovakia		11		3	10			24
Slovenia		10		5	6			21
South Africa					5			5
South Korea							15	15
Spain		21			7			28
Sweden		8		12	3		7	30
Switzerland		7		1	13			21
Taiwan				1	11			12
Tajikistan						2		2
Turkey				2				2
Ukraine				2			3	5
United Kingdom		20		2	27		18	95
United States	19			1	29			49
Uruguay			30	2	5			37
Venezuela	2			2				4
Vietnam					2			2

B Descriptive statistics

Figure B1: Gender wage gaps among youth



Notes: For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that at least 75% of individuals of each gender find a match among the opposite gender.

Table B1: Adjusted gender wage gap in the main sample

	Mean	SD
Matched (share)		
Men	0.907	0.071
Women	0.937	0.056
Estimates adjust for		
... age	1.000	-
... education level	0.991	0.094
... location	0.669	0.471
... presence of children	0.590	0.492
... marital status	0.978	0.146
... industry	0.530	0.499
... occupation	0.443	0.497

Table B2: Construction of the main sample

	N
All youth adjusted gender wage gap estimates	1403
Excluding small sample size or out of <-0.5,0.5> range	1259
With available data on fertility timing	1168
With available instruments	1106

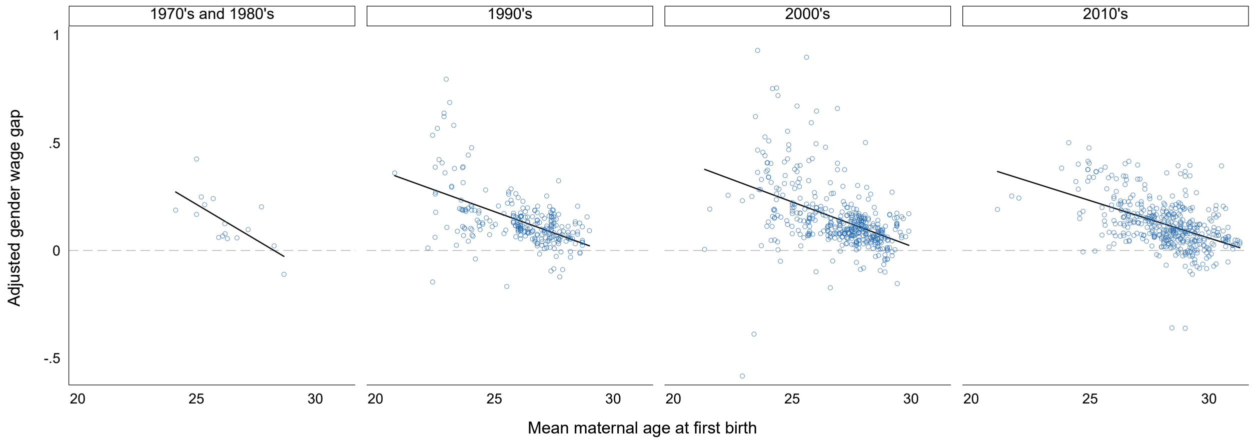
Notes: Table shows the sample size for the main estimation.

Table B3: Time trends in gender wage gaps and mean maternal age at first birth

	All age groups		Youth		Mean age
	Unadjusted GWG	Adjusted GWG	Unadjusted GWG	Adjusted GWG	at first birth
	(1)	(2)	(3)	(4)	(5)
Year	−0.160	−0.0308	−0.164**	−0.158**	0.108***
	(0.101)	(0.0662)	(0.0773)	(0.0705)	(0.0118)
Observations	1,151	1,151	1,128	1,128	1,128
R-squared	0.204	0.117	0.105	0.108	0.204
Mean value	16.28	17.60	7.93	12.23	27.03

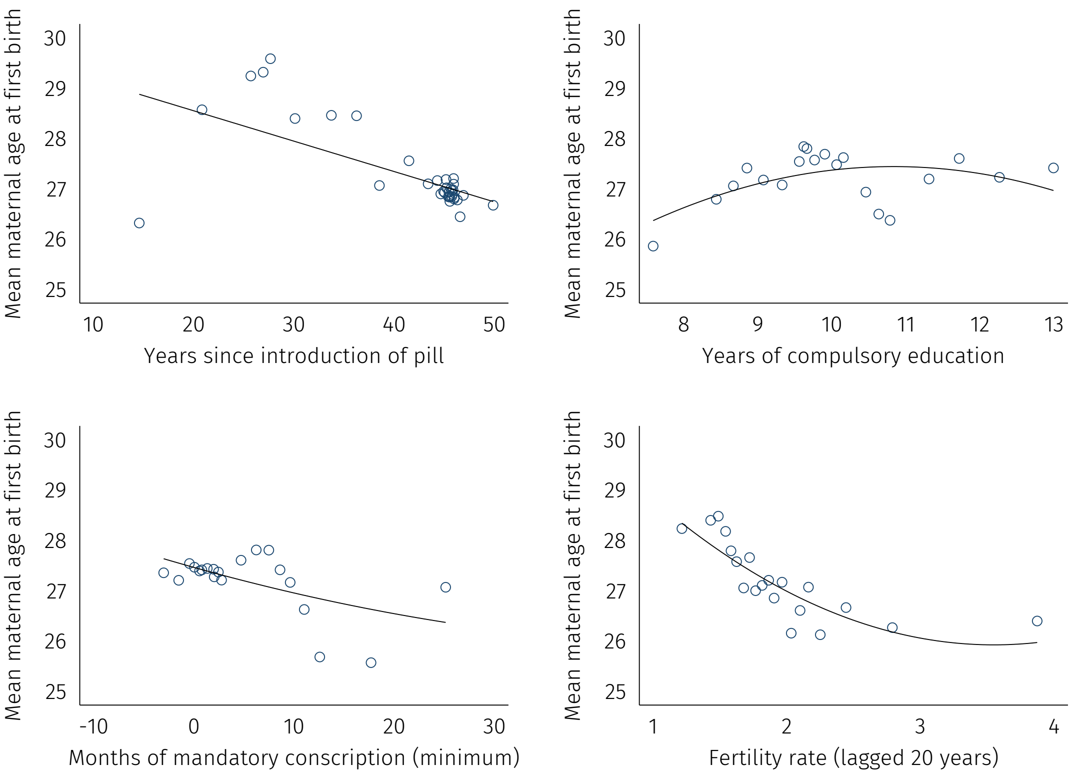
Notes: For each country, year and data source we utilize one estimate, with maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender.
All regressions include country and data source fixed effects.
Standard errors clustered at the level of country and data source. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively.

Figure B2: Adjusted gender wage gap among youth and fertility timing by decade



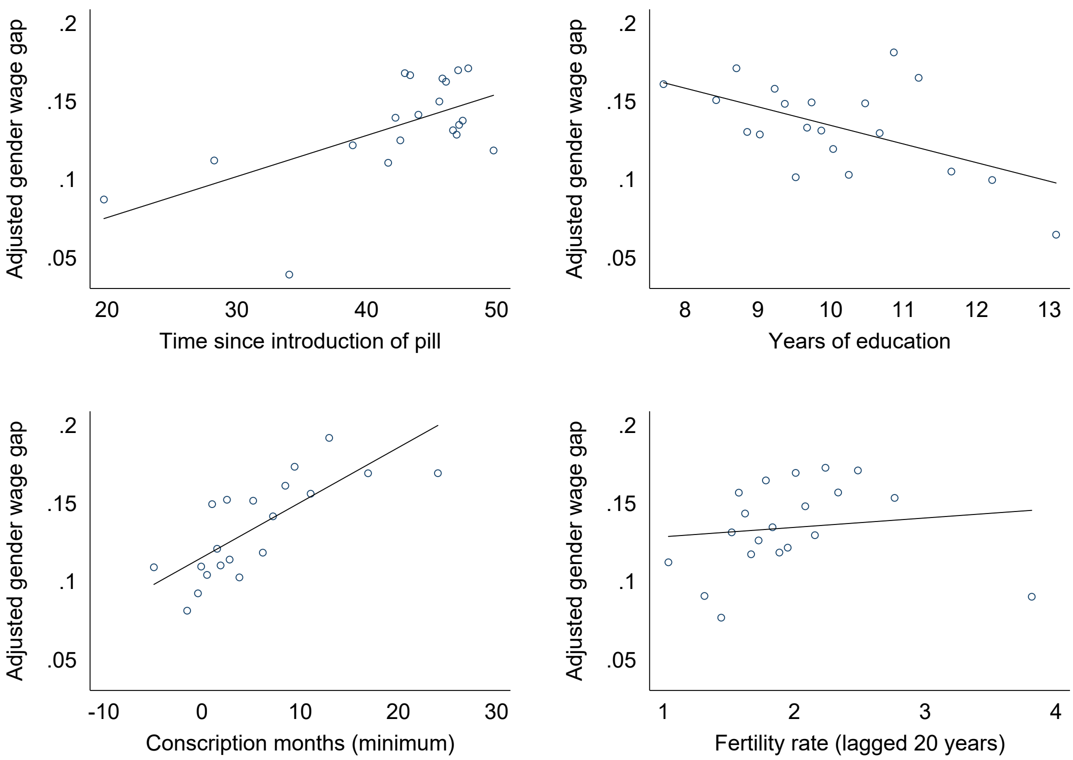
Notes: The scatter plots were obtained on samples analogous to Table 2. For each country, year and data source we utilize one estimate, selected by our standard criteria: the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender, and that estimates are within the range of -1 and 1

Figure B3: Relationship between fertility timing and instrumental variables (the first stage)



Notes: For each country, year and data source we utilize one observation. The figure presents the relationship between each instrument and mean maternal age at first birth after accounting for a linear time trend.
The outlier in the top left panel is Brazil. The country has a introduced the pill relatively late, and has a very low mean maternal age at first birth. Note also that the negative relationship holds even if the sample is restricted to EU countries, detailed results are available upon request.

Figure B4: Adjusted gender wage gap and instrumental variables (the reduced form)



Notes: The figure presents the relationship between each instrument and the adjusted gender wage gaps among the youth. All estimates account for year fixed effects and data source fixed effects.

C Robustness

Table C1: The effect of delayed fertility on AGWG - robustness to alternative IV estimators, restricted sample

	2SLS	HDFE	Quantile Regression			Heterogeneous fertility	
			Q25	Q50	Q75	Intercepts	Slopes
Fertility timing	(1)	(2)	(3)	(4)	(5)	(6a)	(6b)
FT	−0.020 ***	−0.030 ***	−0.022 ***	−0.021 ***	−0.023 ***		
	(0.01)	(0.00)	(0.00)	(0.00)	(0.01)		
FT < Q25						0.112 ***	−0.030 **
						[0.05,0.17]	[−0.05,−0.01]
FT ∈ [Q25, Q75]						0.005	−0.031 ***
						[−0.03,0.04]	[−0.05,−0.01]
FT > Q75							−0.030 ***
							[−0.05,−0.01]

Notes: Standard errors clustered at the country-data source level in parentheses in columns (1)-(5). Confidence intervals (95%) in brackets in columns (6) and (7). We report estimations analogous to column (1) from Table 2. In column (1) we use fixed effects IV estimator (2SLS IV). In column (2) we use High-Dimensional Fixed Effect IV estimator (HDFE). In columns (3)-(5) we utilize Firpo et al. (2009) recentered influence function transformation for AGWG of the model at 25th, 50th and 75th percentile, respectively. In columns (6a) and (6b) we account for the distribution of mean maternal age at first birth (intercepts and slopes).
For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender, with the additional constraint that the control factors must include industry and occupation.
All specifications include time trends, data source fixed effects, and adjust for the AGWG model specification.
Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from first and second stage regressions is available upon request.

Table C2: The effect of fertility timing on AGWG – robustness to additional controls

	(1)	(2)	(3)
Fertility timing	−0.025***	−0.016*	−0.015**
	(0.0063)	(0.0082)	(0.0075)
GDP per capita	No	Yes	Yes
Fertility rate	No	Yes	Yes
Unemployment: all and women	No	Yes	Yes
Youth unemployment: all and women	No	Yes	Yes
Tertiary enrollment: all and women	No	No	Yes
Observations	1106	1044	876
R-squared	0.29	0.34	0.41
F-statistic	25286.6	35712.2	105174.2

Notes: IV specifications using Baltagi (1981) estimator where the dependent variable is the Adjusted Gender Wage Gap (AGWG) among the youth. Column (1) above replicates Column (1) in Table 2 for convenience. The data on GDP per capita (NY.GDP.PCAP.PP.KD), unemployment (SL.UEM.TOTL.ZS and SL.UEM.TOTL.FE.ZS for women), youth unemployment (SL.UEM.1524.ZS and SL.UEM.1524.FE.ZS for women), tertiary enrollment (SE.TER.ENRR and SE.TER.ENRR.FE for women), youth NEET (SL.UEM.NEET.FE.ZS) and fertility rate (SP.DYN.TFRT.IN) were taken from The World Bank.
For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender
All specifications include time and data source fixed effects, and adjust for the AGWG model specification.
Standard errors clustered at the country-data source level in parentheses. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from the first and the second stage regressions is available upon request.

Table C3: The effect of delayed fertility on AGWG (all age groups)

AGWG estimates	IV (β^{IV})			OLS (β)
	(1)	(2)	(3)	(4)
Panel A: *Full sample (all AGWG estimates)*				
Fertility timing	−0.012***	−0.027***	−0.011**	−0.013*
	(0.0044)	(0.0075)	(0.0046)	(0.0065)
Observations	1165	1220	1173	1229
R-squared	0.22	0.27	0.22	0.85
F-statistic	20795.5	495.4	16040.3	
Panel B: *AGWG estimates **without** occupation and industry of employment*				
Fertility timing	−0.0084	−0.029***	−0.0072	−0.0019
	(0.0054)	(0.0080)	(0.0058)	(0.0074)
Observations	1163	1218	1171	1227
R-squared	0.19	0.26	0.18	0.83
F-statistic	16889.9	509.9	13283.2	
Panel C: *AGWG estimates **with** occupation and industry of employment*				
Fertility timing	−0.015***	−0.019***	−0.014***	−0.015**
	(0.0041)	(0.0062)	(0.0041)	(0.0066)
Observations	826	865	835	874
R-squared	0.35	0.36	0.35	0.86
F-statistic	7704.4	391.6	3330.3	
Clustering SE	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes

Notes: IV specifications using Baltagi (1981) estimator with time varying and time-invariant components, random effects models, include year, specification and data source fixed effects. Column (1) with all instruments jointly. Column (2) with the pill authorization as the only instrument. Column (3) with all instruments *but* the pill authorization. In all IV specifications, we include linear term and base functions up to a fourth polynomial.

In Panel A, for each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender. In Panel B, for each country and year we apply the same restriction as in A, but only among estimates without controls for industry and occupation. In Panel C, for each country and year we impose an additional restriction that estimates should adjust for occupation and industry.

All specifications include time trends, data source fixed effects, and adjust for the AGWG model specification. The OLS specifications adjust for $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of data sources for given country in a given year. Standard errors clustered at country-data source-controls level. Asterisks ***, **, and * denote significance at 1%, 5% and 10%, respectively. Full set of estimates from the first and the second stage regressions is available upon request.

Table C4: The effect of fertility level on gender wage gap

	UGWG		AGWG	
		All	w/o ind. & occ	w/ ind.& occ
	(1)	(2)	(3)	(4)
Panel A: *GWG for the young*				
Fertility level	0.0026	−0.016	−0.012	−0.0045
	(0.036)	(0.026)	(0.033)	(0.025)
Observations	1241	1241	1238	938
R-squared	0.62	0.74	0.71	0.57
Panel B: *GWG for entire population*				
Fertility level	0.019	0.013	0.026	−0.011
	(0.033)	(0.022)	(0.023)	(0.018)
Observations	1301	1301	1299	939
R-squared	0.80	0.83	0.81	0.80
Clustering SE	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes

Notes: OLS estimates of the relation between gender wage gaps and fertility levels measured by Total Fertility Rate (TFR). Column (1) the dependent variable are gender wage gaps without adjusting (UGWG). Column (2) includes one estimate per country year, the one with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender. Column (3) applies the same restriction, but only among those estimates without controls for industry and occupation. Finally, Column (4) applies the same restriction to estimates that contain controls for industry and occupation.

In Panel A, estimates correspond to the sample of respondents aged below 30 years old. In Panel B, estimates were obtained from the entire samples.

All specifications include time trends, data source fixed effects, and adjust for the AGWG model specification. The specifications adjust for $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of data sources for given country in a given year.

Standard errors clustered at country-data source level. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively.

Table C5: The effect of fertility timing on AGWG – including one set of instruments at a time

	(1)	(2)	(3)	(4)
Fertility timing	−0.035***	−0.025***	−0.036***	−0.026***
	(0.0071)	(0.0086)	(0.0078)	(0.0061)
Observations	1161	1137	1149	1142
R-squared	0.29	0.29	0.29	0.29
F-statistic	461.1	225.2	485.6	1038.9

Notes: IV specifications using Baltagi (1981) estimator where the dependent variable is the Adjusted Gender Wage Gap (AGWG) among the youth. Each column employs a different set of instruments. Column 1 includes time since the introduction of the pill, Column 2 includes years of mandatory education, Column 3 includes duration of conscription and Column 4 Fertility in the generation of parents.

For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender.

All regressions include year and data source fixed effects, and adjust for the AGWG model specification.

Standard errors clustered at the level of country and data source in parentheses. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from the first and the second stage regressions is available upon request.

Table C6: The effect of fertility timing on AGWG – including one instrument at a time

	(1)	(2)	(3)	(4)
Panel 1: Time since the introduction of the pill				
Fertility timing	−0.034***	−0.038***	−0.044***	−0.041***
	(0.0080)	(0.0072)	(0.0068)	(0.0068)
R-squared	0.29	0.29	0.29	0.29
F-statistic	263.8	272.9	281.1	286.0
Panel 2: Years of compulsory education				
Fertility timing	−0.030***	−0.030***	−0.030***	−0.029***
	(0.0087)	(0.0089)	(0.0090)	(0.0092)
R-squared	0.30	0.30	0.30	0.30
F-statistic	36.0	34.7	33.2	32.3
Panel 3: Duration of conscription (minimum)				
Fertility timing	−0.031***	−0.035***	−0.035***	−0.034***
	(0.0089)	(0.0093)	(0.0091)	(0.0089)
R-squared	0.29	0.29	0.29	0.29
F-statistic	114.1	71.8	58.7	55.9
Panel 4: Duration of conscription (maximum)				
Fertility timing	−0.027***	−0.029***	−0.030***	−0.030***
	(0.0090)	(0.0093)	(0.0089)	(0.0086)
R-squared	0.29	0.29	0.29	0.29
F-statistic	126.7	75.7	53.7	46.8
Panel 5: Fertility rates lagged 20 years				
Fertility timing	−0.023***	−0.022***	−0.022***	−0.023***
	(0.0086)	(0.0076)	(0.0070)	(0.0069)
R-squared	0.29	0.29	0.29	0.29
F-statistic	72.9	69.7	73.5	106.5

Notes: IV specifications using Baltagi (1981) estimator where the dependent variable is the Adjusted Gender Wage Gap (AGWG) among the youth. Each panel refers to a different instrument, and each column to a different power of that instrument. For example, in Panel 1 Column 2, the instrument is the time since the introduction of the pill squared.
Sample includes one estimate per country year, the one with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender.
All regressions include year and source fixed effects and adjust for the AGWG model specification.
Standard errors clustered at the level of country and data source in parentheses. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from the first and the second stage regressions is available upon request.

D On the relation between mean maternal age at first birth and proportion of childless women

Our main indicator of fertility timing is mean maternal age at first birth. Ideally, one would like to work with the proportion of women without children, but such measure can only be recovered in a subset of countries, usually the most developed. In this Section, we explore how these two measures are related. Similar to Section 4.2, we derive the proportion of childless women of a given age as one minus the ratio of the number of first births for each cohort up to that age and the estimated number of women.

$$\text{Proportion childless}_{c,A} = 1 - \frac{\sum_{a=16}^{A} \text{First Births}_{c,a}}{\text{N women}_{c,A}}$$

where c stands for cohort, and A for age. In this application, we selected three values of A: 20, 25 and 30 years old. This measure is an approximation to the true proportion of childless women. For example, childless women might be more likely to migrate, and this might not be fully reflected in the denominator. Moreover, this measure is only available for cohorts for whom we observe the entire history of first births up to a given age. For example, to recover the proportion of childless women under age of 20 in 1996, we need to observe first births since 1992, when these women were 16 years old. This requirement effectively restricts the sample to the most recent years and to the subset of countries that report consistent (and uninterrupted) measures.

With these caveats in mind, Table D1 shows the estimates from a linear regression of proportion of childless women by age *A* on a time trend and country fixed effects. The sample comprises only observations obtained after the year 2000, when the list of countries reporting data stabilized. For comparison, we also include a regression of the mean maternal age at first birth for this subsample.The time trend is centered around 2001, so that the constant can be interpreted as the detrended average in the reference country (US). We observe that the difference in levels is consistent with what one would expect. Around 77 percent of women are childless when they are twenty years old. The proportion of childless women falls to 47 percent at age 25, and to 27 percent at age thirty. The coefficients for time trends are all positive. These results reflect the postponement (and decline) in fertility in the last two decades. The proportion increases relatively faster in the first two groups, and somehow slower in the last one ($0.0022/(1 - 0.77) = 0.009$ for the first group, 0.011 for the second group, and 0.006 among the later group). The positive coefficient for the time trend in the mean maternal age at first birth also reflects the decision to delay fertility.

Table D1: Delaying fertility: evidence across countries

	Probability of childless			Mean age
	at age 20	at age 25	at age 30	at 1st birth
Time trend	0.0022***	0.0050***	0.0046***	0.13***
	(0.00020)	(0.00037)	(0.00057)	(0.0030)
Intercept	0.78***	0.51***	0.30***	25.1***
	(0.0038)	(0.0071)	(0.011)	(0.058)
Observations	253	247	262	262
R-squared	0.93	0.94	0.92	0.97

Notes: Table shows coefficients from regressing the proportion of childless women up to age *A* on a time trend. All estimates include country fixed effects. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively.

We further explore the relationship between the mean maternal age at first birth and the probability of being childless at different ages with the help of Figure D1. In the sample, we kept only those countries for which we had a sufficiently large number of observations. Below each plot we list the countries included in the sample. Each country has a distinct color. Lines show a linear projection of probabilities to mean maternal age at first birth.

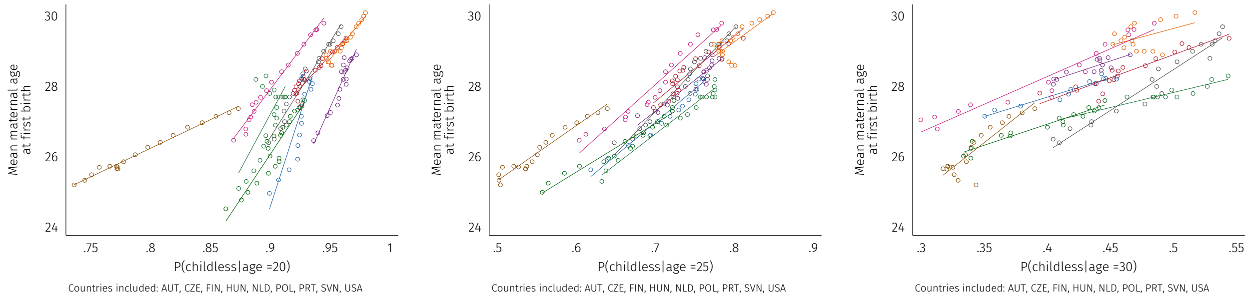


Figure D1: Mean maternal age at first birth and probability of being childless

Notes: Different colors identify different countries. Lines indicate linear projection of probabilities to mean maternal age at first birth.

These plots show a positive association between the variables. Higher probability of remaining childless is associated with a higher mean maternal age at first birth. This relationship is not mechanical. An increase in the proportion of childless people at any age can be driven by an increase in the proportion of people who forego child-bearing, in which case the mean maternal age at first birth would be constant. The positive correlation between the variables lead us to conclude that the mean maternal age at first birth is correlated with the proportion of childless women at various ages. While lines are almost parallel in the middle plot, we observe that the relationship is more heterogeneous at both extremes of the distribution. [36]

[36] In Lithuania the relationship between mean maternal age at first birth and probability of being childless flips signs in 2012 / 2013. We attribute this to an statistical artifact. In 2011, the country conducted a census, which led to a revision downwards of the number of women in previous years. If childless women aged between 25 and 30 years old were more likely to migrate, then we would observe a fall in the probability of being childless at that age. Importantly, we do not expect the same problems at earlier ages, since women might still be in education. In more recent years, where the number of women was stable, the relation between the two variables was positive, like in the remaining countries.

E ISSP data for benchmarking statistical gender discrimination

In ISSP of 2012, adult respondents report the time spent on caring. Specifically, the questionnaire asks *"On average, how many hours a week do you spend looking after family members (for example children, elderly, ill or disabled family members)?"*. This question is answered by respondent about both him/her and the partner/spouse of the respondent. Given that the question is the same across countries, this database provides comparable measurement of c.

Arguably, time-use surveys provide more accurate measurement than the ISSP, given that in the ISSP the respondents round time spent in activities to full hours. However, in those surveys, household members report the time spent on caring. Time-use surveys differ substantially in the method of collecting the data: in some data sources individuals report time spent on primary activity, in some surveys also secondary activity is reported. For example, a primary activity could be caring if an individual feeds a child without eating themselves, whereas a secondary activity could be caring if an individual feeds a child while also eating own meal. In addition to this differentiation of the time-use surveys, data collection methods evolved over time and are not the same across countries. Some countries collect data in 15-minutes intervals, in daily diaries, whereas in some countries the respondents are asked about some past time (for example the previous week) and are expected to report the start and end hours by themselves. This differentiation puts some doubt on the extent to which the data from the time-use surveys can be compared across countries, sometimes even within-countries and across periods.

The version of our benchmarking exercise relying on the ISSP data reveals that for all the countries, the ballpark implied by our model estimates is indeed close to $(c_w - c_m) \times \pi$: obtained through time-use from ISSP and age-specific fertility fall within the confidence intervals of the estimated AGWG, as predicted from equation (3). In each country, at least one of the simulated $(c_w - c_m) \times \pi$ outcomes falls within the confidence intervals. While in the case of some countries all four measures are very close to the estimated AGWG, in other countries this holds for fewer measures due to the fact that the time-use gap measures appear to be highly dispersed.[37]

Indeed, a priori, there are no arguments for or against including chores in the time-use gaps measures. This is because the employers expectations may or may not include these activities. There are stronger theoretical foundations for preferring the means-based measure over medians, notably rational expectations. However, basing expectations on medians consistently could be interpreted as an unbiased departure from rationality. For example, in the case of US and the UK, predicted AGWGs appear consistent with expectations at the median, even if they are higher than the expectations at the mean. However, measures based on means are substantially lower, which would hint that in addition to accurate statistical discrimination, differences in earnings reflect also inaccurate beliefs, stereotypes and tastes. In the case of some countries, the estimates of AGWG substantially exceed $(c_w - c_m) \times \pi$ for two or three time-use gap measures. If taken at face value, these results imply excessive statistical discrimination, which hints at biases in correctly receiving signals about the state of nature and discriminatory tastes.

Austria is an interesting case for our benchmarking exercise. Kleven et al. (2020) document convergence in unadjusted wages spanning several decades across genders in this country. They

[37]As is visible, the dispersion between mean and median measures of $(c_w - c_m)$ differs across countries: in Ireland, the Netherlands or Latvia the mean and median gaps in time-use are similar, whereas in Hungary, UK or US whereas the mean and median gaps in time-use differ substantially. Likewise, in some countries accounting for chores (squares) implies no changes to simulated $(c_w - c_m)$ relative to the mean (triangles, e.g., Belgium, Czechia, or Ireland), whereas in the case of others the chores cause large changes to obtained $(c_w - c_m)$.

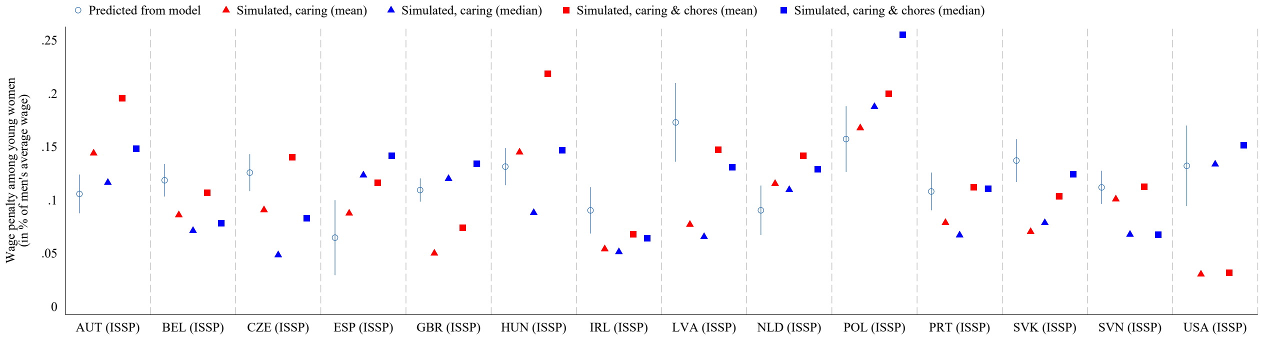


Figure E1: Benchmarking statistical gender discrimination

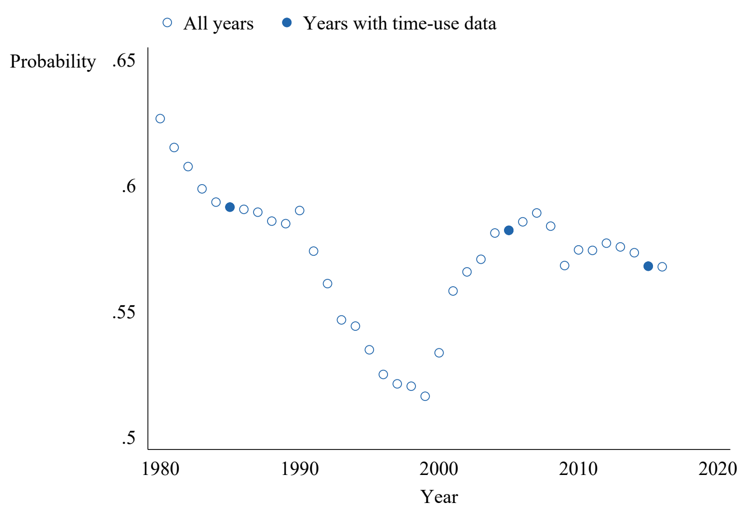
Notes: data comes from International Social Survey Program (data from 2012) Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (2), for $year = 2012$. None of the wage gaps predictions are based on ISSP data (for no country in this sample ISSP has proven to be the "best" available data for 2012). Simulations at the mean and at the median utilize Eurostat and Human Fertility Database age-specific fertility data for π. Note that equation (3) adjusts for the mean maternal age at first birth, a time trend and data source fixed effects. Hence, the only source of cross-country variation in the predictions from the model in Figure 2 is based on the mean maternal age at first birth as implied by the equation (2), because all estimates are provided for 2012. Thus, the dispersion of predicted AGWG in Figure 2 follows from the variation in fertility patterns across these countries.

subsequently show that changes in the duration of the maternity leave during the same period have no systematic power to explain this trend. In our benchmarking exercise in ISSP 2021, the estimates of AGWG are generally higher than the values implied by $(c_w - c_m) \times \pi$ (except median, including chores, blue square). Hence, the adjusted gender wage gaps among youth in this country is *smaller* than a rational employer would impose. Given low overall labor force participation of women in Austria, relative to other EU countries, one potential explanation of our findings for this country may be that it is actually optimal to pay a premium to the disfavored group if labor market participation is an informative signal of productivity (Blair and Chung, 2021).

F Evolution of fertility and time use over time in the US

Time-use surveys for the United States are available since 1986. Availability of the individual-level data for obtaining the AGWG estimates allows tracing the time trends in $(c_w - c_m)$, π and adjusted gender wage gaps. Figure F1 depicts the evolution of π in the US. The years marked with a full circle denote the availability of the time-use surveys. There are three distinct periods in the evolution of π: a steep decline in π between 1980 and 2000, a rise between 2000 and 2010 and relatively flat behavior of π thereafter. This complexity of changes in maternal age at first birth implies that employers were forced to frequently update their beliefs about the risk of pregnancy and child-related absences of workers. Moreover, inferring the past patterns could be misleading for the future.

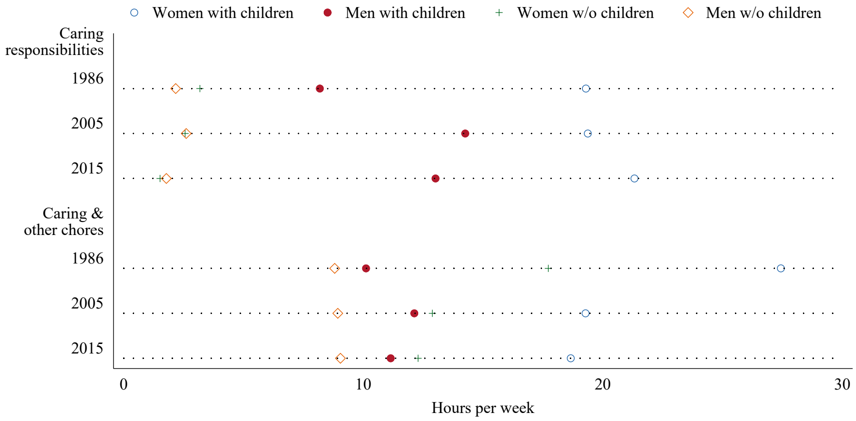
Figure F1: Evolution of fertility patterns in the United States



Notes: Data on age-specific fertility rates comes from Human Fertility Database. Full markers denote years for which time-use data are available. Data between 2005 and 2015 are available on an annual basis, but reveal similar picture. Hence, for clarity, we portray these two data points: 2005 and 2015.

Figure F2 portrays the estimates of caring time for respective years using time-use data for the United States. The weekly hours for women with children were similar across the years, it is the caring time of men with children that changed substantially. In other words, it is not that the time allocated by women declined – rather the time allocated by men increased. This implies that on the one hand the differential effect between young men and women declines, specifically $(c_w - c_m)$ declines. However, the time endowment of women is just as taxed as it was before, so the rational employer has no reasons to expect a mother to have a higher time endowment, rather the employer ought to expect a father to have a lower time endowment. If we account for household chores, there is a decline for women between 1980s and 2000s, but the evolution as of 2000 is relatively flat.

Figure F2: Evolution of caring time in the United States



Notes: Data from American Time Use Survey (ATUS).

References

Adda, J., Dustmann, C., and Stevens, K. (2017). The career costs of children. *Journal of Political Economy*, 125(2):293–337.

Altonji, J. G. and Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(1):313–350.

Amuedo-Dorantes, C. and Kimmel, J. (2005). The Motherhood Wage Gap for Women in the United States: The Importance of College and Fertility Delay. *Review of Economics of the Household*, 3(1):17–48.

Ananat, E. O. and Hungerman, D. M. (2012). The power of the pill for the next generation: oral contraception's effects on fertility, abortion, and maternal and child characteristics. *Review of Economics and Statistics*, 94(1):37–51.

Arrow, K. (1973). The theory of discrimination. pages 3–33. Princeton University Press.

Auspurg, K., Hinz, T., and Sauer, C. (2017). Why should women get less? evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, 82(1):179–210.

Baert, S. (2014). Career lesbians. getting hired for not having kids? *Industrial Relations Journal*, 45(6):543–561.

Bailey, M. J. (2006). More power to the pill: The impact of contraceptive freedom on women's life cycle labor supply. *Quarterly Journal of Economics*, 121(1):289–320.

Bailey, M. J. (2010). "Momma's got the pill": how Anthony Comstock and Griswold v. Connecticut shaped US childbearing. *American Economic Review*, 100(1):98–129.

Bailey, M. J., Hershbein, B., and Miller, A. R. (2012). The opt-in revolution? contraception and the gender gap in wages. *American Economic Journal: Applied Economics*, 4(3):225–54.

Baizan, P., Arpino, B., and Delclós, C. E. (2016). The effect of gender policies on fertility: The moderating role of education and normative context. *European Journal of Population*, 32(1):1–30.