GRAPE

Group for Research in APplied Economics

GRAPE Working Paper #66

Statistical gender discrimination: evidence from young workers across four decades and 56 countries

Lucas van der Velde and Joanna Tyrowicz

FAME | GRAPE, 2022

Foundation of Admirers and Mavens of Economics
Group for Research in Applied Economics

Statistical gender discrimination: evidence from young workers across four decades and 56 countries

Lucas van der Velde
Warsaw School of Economics &
FAME|GRAPE

Joanna Tyrowicz
University of Warsaw &
FAME|GRAPE

Abstract

Statistical discrimination offers a compelling narrative on gender wage gaps among younger workers. Employers could discount women's wages to adjust for probable costs linked to childbearing. Given trends towards lower and delayed fertility one should observe a lower discount in wages and a reduction in the gender wage gap among entrants. We put this conjecture to test. We provide a novel collection of adjusted gender wage gap (AGWG) estimates among young workers from 56 countries spanning four decades. We use these estimates to study the effects of postponing childbirth on AGWG. We find that postponing first parity by a year reduces AGWG by two percentage points (15%). We further benchmark the implied gender inequality with the help of time-use data

Keywords:

youth, gender wage gap, statistical discrimination

JEL Classification

J71, J31, J13

Corresponding author

Joanna Tyrowicz, j.tyrowicz@grape.org.pl

Acknowledgements

The authors express gratitude for insightful comments and suggestions by Michele Belot, Shelly Lundberg, Anna Matysiak, Michal Myck, Filip Pertold, Barbara Petrongolo, Olga Popova, Nuria Rodriguez-Planas, Andrea Weber, and participants of EEA 2019, Gender Workshop in Prague 2019, EACES 2021, LABFAM seminar as well as seminars at GRAPE, WSE and University of Warsaw. We have received great help from John Bongaarts, Janos Kollo, Mathias Nathan, and Yaer Zhuang. The assistance of LISSY data facility by LISER is gratefully acknowledged. Kamil Sekut provided wonderful research assistance. Financial support by National Science Center grant no. #2017/27/L/HS4/03219 is acknowledged. This paper is an outcome of a joint NCN-LNT DAINA innitiative. The authors do not have any conflicts of interest to disclose.

Published by: FAME | GRAPE
ISSN: 2544-2473

Foundation of Admirers and Mavens of Economics
Koszykowa 59/7
00-660 Warsaw
Poland

W | grape.org.pl
E | grape@grape.org.pl
TT | GRAPE_ORG
FB | GRAPE.ORG
PH | +48 799 012 202

1 Introduction

The risk that the workers become unavailable due to child-rearing is often cited as *the rational reason* for *statistical discrimination*, a notion put forward by Kenneth Arrow and Edmund Phelps (Schwab 1986, Norman 2003). Rationality is premised by accuracy in the perception of the risk. If the incidence of child-rearing is exaggerated, rational choice under uncertainty morphs into harmful stereotyping akin to distaste. Specifically, rational employers should accommodate lower and delayed childbearing, as observed across many countries during the past four decades, through lowering the discount on young women's wages relative to young men. We put this conjecture to empirical test, by studying the link between changes in fertility patterns and adjusted gender wage gaps among labor market entrants.

The existing literature shows that once realized and thus observable, fertility matters for the employers in wage setting as well as hiring decisions. The literature so far does not inquire if the employers adjust their statistical expectations at par with the actual costs that they could face. With the decline and delaying of fertility, an employer faces a lower probability of employing a primary care-giver when employing a young woman. This in turn implies – if statistical discrimination is indeed the underlying cause of prevailing gender wage gaps differences – that wages should become more equal.

To test this conjecture, we collect individual level data for over fifty countries, spanning nearly four decades. Using these data, we obtain comparable estimates of adjusted gender wage gaps among labor market entrants. We then study the evolution of these adjusted gender wage gaps in conjunction with the mean age at first birth. During the analyzed period there was substantial variation in trends of mean age at first birth across countries. In some countries, a substantial delay in fertility was observed. Under statistical discrimination, employers hiring young workers observe changes in the risk of the workers becoming unavailable due to child-rearing, which should exhibit in adjusted gender wage gaps among youth. We confirm this conjecture in the data, exploring several instruments to account for endogeneity in fertility decisions and discrimination of young women. We also show that, as consistent with theory, the relationship holds for the timing of fertility rather than the number of children *per se*. Finally, we put our estimates in perspective showing that in some countries the magnitudes of gender wage gaps are consistent with those implied by our estimates, whereas in other countries the gaps appear much higher which hints that there may be additional mechanism behind adjusted gender inequality in wages.

Our inquiry builds on a growing body of studies that inspect the accuracy of statistical discrimination. The literature typically faces challenges in measuring statistical discrimination (Altonji and Pierret 2001) and evaluating the accuracy of the underlying statistical beliefs (e.g. Pager and Karafin 2009, Bohren, Haggag, Imas and Pope 2019). The obvious advantage of studying the link between fertility and adjusted gender wage gap is that the aggregate fertility decisions are perfectly observable in multiple contexts, which implies that employers can receive the correct signal about the risk of female workers' absences.

We overcome two important challenges faced by earlier studies. First, typically data limitations imply that studies cover a single country, which impedes deriving conclusions related to slow moving processes such as changes in fertility patterns. Second, the estimates available in the published papers are not directly comparable, because the authors use differentiated data, methods and pose different research questions (Weichselbaumer and Winter-Ebmer 2005). With these two constraints, it is difficult to study if *postponing* fertility has translated to lower adjusted gender wage gaps among labor market entrants. To address these challenges we develop a comprehensive collection of adjusted gender wage gaps among youth across countries and years.[1] Our estimates are comparable between countries and over time, because we obtain them from individual-level, harmonized data, and we apply the same estimation method. We obtain the adjusted gender wage gaps (AGWG) for individual characteristics using a non-parametric method (Nopo

[1]The full set of gender wage gaps estimates is available at [LINK REMOVED FOR ANONYMITY].

2008) that allows robust estimation even with a small set of controls (Goraus et al. 2017).

Overall, in our sample the AGWG *among young* workers declines with delay of fertility: increasing the mean maternal age at first birth by a year reduces AGWG by roughly two percentage points, or as much as 15% of the total gap. By contrast, there is no link between overall level of fertility and AGWG in all age groups, nor among youth. The estimates prove to be robust across estimation methods. We also show that this estimated effect may be useful to study whether the implied magnitude of (adjusted) gender wage gaps is consistent with accurate statistical discrimination when setting the wages of young workers. Our results thus corroborate that statistical discrimination constitutes a relevant component of gender wage gaps, and employers adjust to the observed changes fertility timing. Postponing fertility is clearly not the only way to reduce this risk: a more even distribution of caring responsibilities between both parents would have the same effects on employers. Our results lend support to the notion that reducing the disproportion in time devoted to caring by mothers and fathers reduces the gender wage gaps: at least in some countries the employers adjust wage policy accurately to changes in workers' time endowments.

Our paper is structured as follows. First, we discuss the relevant literature from social sciences in section 2. We use this section to discuss the state-of-the art and specify the knowledge gaps that our study aims to fill. Second, we move to data and methods in section 3. This paper involves extensive harmonization of vast collection of individual-level data, which we document in the paper and in the appendices. We also explain the methodological choices concerning the estimation of adjusted gender wage gaps. Sections 4 and 5 delve into the results and discuss their robustness as well as limitations. The final section concludes with policy implications of our study.

2 Literature

Starting from Becker (1971), Phelps (1972) and Arrow (1973), statistical discrimination is a recognized mechanism explaining the differences in wages between a favored and a disfavored group.[2] In a nutshell, consider the following line of argumentation: as some people become parents, parenting involves a drop in productivity, for example due to absences. The employer does not know whether a young worker will become a parent during their tenure at a given business, but they expect that if that parenting occurs, then the productivity drop for women will be larger than for men. This is a form of statistical discrimination, because it involves averaging over all women on the one hand and all men on the other hand. Rational employers will discount this expected productivity drop in their offered wages. If a given candidate requests a wage that ignores this statistical discounting, no employment match is formed. Female candidates, foreseeing this discounting, may find it optimal to lower the request in order to obtain any employment.[3]

This form of statistical discrimination is different from the standard approach in the literature. Altonji and Pierret (2001) study the racial gaps in the US and estimate if employers adjust wages in reaction to observed worker performance as tenures increase. In this case, the underlying mechanism for statistical discrimination is that in the absence of knowledge about individual productivity, the employers resort to stereotyping about human capital: on average the employers initially assumed lower skill for Blacks than for Whites, even if they hold the same credential. With time, as productivity is revealed, this racial gap narrows.

Such a learning process is less applicable in the context of gender wage gaps, because child-bearing and rearing vary over time. First, not having children until given time (before a contract with a given employer), may actually imply a *higher* probability of procreation within the window of contract of that employer, *ceteris paribus*. Hence, the learning mechanism may work in the opposite direction in the case of the gender

[2] Despite immediate and excellent counter-points raised byBergmann (1973).

[3] For a thorough review of relevant social theories as well as an experimental exploration see Auspurg et al. (2017).

gaps, as compared to racial gaps. Second, even once parenthood has already occurred, the employers still experience uncertainty associated with the productivity decline due to absences, because the parenting style can change (e.g., children may require more parental involvement due to unexpected adverse health shocks). Hence, the event of childbearing does not resolve the uncertainty on the side of the employer. For both reasons, observing the gap between young men and women at the same employer along the tenure distribution cannot reveal statistical discrimination or lack thereof.

A way to circumvent this difficulty involves moving from observational studies to directly controlling expectations about productivity across genders in an experimental context. Bohren, Imas and Rosenberg (2019) explore wage-setting in repeated interactions about the compensations in digital platforms and confirm prevalence of unjustified wage inequality. In a related study, Bohren, Haggag, Imas and Pope (2019) show that systematically confronted with facts about performance, inaccurate statistical discrimination declines but it does not disappear.[4]

While in general these results could be viewed with some optimism, updating beliefs about actual workers is not enough to reduce the prevalence of inaccurate statistical discrimination. On the one hand, employers update beliefs about the *unknown* population of potential workers much slower than they update their beliefs about hired workers (Pager and Karafin 2009). On the other hand, attention discrimination effectively prevents the updating of beliefs, thus reinforcing inaccurate statistical assumptions (Bartoš et al. 2016). A wide constellation of recent papers finds sizable wage penalties for mothers relative to fathers, for example Adda et al. (2017), Fuller and Cooke (2018), Kleven, Landais, Posch, Steinhauer and Zweimüller (2019), Costa Dias et al. (2020). There is also burgeoning research on gender equality in callback rates in the correspondence studies, with ambiguous results on the effects of gender and parenthood (Petit 2007, Bygren et al. 2017, Hipp 2020, Becker et al. 2019). Audit and correspondence studies inform about the mechanisms behind hiring women and the role of parenthood. For example, Correll et al. (2007) had subjects evaluate applications of the same gender, differing by parenthood status. They find that women are penalized for being mothers, whereas men are typically not penalized and, in some cases, are even rewarded for being fathers. The penalty concerns not only the recommendation to hire a candidate, but also perceived competences and proposed wage despite the applications being exactly identical on all counts except for parenthood. Correspondence studies inquiring the call-back rates for candidates differing by gender and parenthood status find ambiguous results. Some studies point to a strong motherhood penalty (for example Petit 2007, González et al. 2019, Hipp 2020) while others do not find differences between men and women (Bygren et al. 2017, Becker et al. 2019). The overall implication of this literature is that institutional and cultural factors may reduce the bias against mothers (Baert 2018). Thus, the literature suggests that once realized and thus observable, fertility matters for the employers in wage-setting as well as in hiring decisions.

Using observational data Gangl and Ziefle (2009) trace five subsequent cohorts of women in Germany, UK and the US and find substantial declines in wages after child-birth. Following an approach similar to an event study, Kleven, Landais, Posch, Steinhauer and Zweimüller (2019) proposed to explore wage gaps between mothers and fathers after child-bearing. They explore administrative data and find that whereas until child-bearing the wages of young women track closely those of young men, after the event of childbearing the gap in hourly compensation exceeds 20%. This result has proven to be robust across contexts, methods and countries Adda et al. (2017), Fuller and Cooke (2018), Kleven, Landais and Søgaard (2019), Costa Dias et al. (2020).[5] These studies do not identify the presence of bias against women *before* child-bearing, rather they

[4]The inaccuracy of statistical discrimination and information avoidance has also been studied in the context of racial and ethnic statistical discrimination and in housing as well as credit markets, Pope and Sydnor (see 2011), Ewens et al. (see 2014), Flage (see 2018), Morse and Pence (see 2020, among others). In addition, statistical discrimination appears to intensify with increasing prevalence of algorithms Kleinberg et al. (for example 2018), Raghavan et al. (for example 2020).

[5]See also Cukrowska-Torzewska and Lovasz (2020) for a descriptive comparative analysis of EU Member States.

confirm strong bias against mothers. While this issue has been much less extensively studied, previous results suggest that the wage gap at a young age actually exacerbates over career (van Staveren et al. 2018).

Studies exploiting more recent data, i.e. as of 2000s, by design ignore the sometimes remarkable progress that has occurred in terms of gender equality in employment and wages. Raw gender wage gaps are declining steadily (Blau and Kahn 2017) and access to high-aspirations occupations is rising for women (Hsieh et al. 2019). Part of that achievement was driven by improved conditions for rearing children and pursuing a professional career (Matysiak and Vignoli 2008) and a part occurred due to improved educational attainment of women, which explains why there is virtually no time trend in adjusted gender wage gaps (Weichselbaumer and Winter-Ebmer 2005). Aaronson et al. (2021) shows that the post-fertility decline in labor supply is indeed a phenomenon persistent across years, with roughly the same magnitude. The author argues that the expectations of young women concerning their future labor supply changed substantially: while in the past they might have expected to become stay-at-home moms and instead chose to work when the time came, now the reverse becomes true. Kuziemko et al. (2018) show that this reversal is substantial and exhibited in many countries.

Much less has been said about the link of fertility *timing* and gender inequality in employment and wages. Demographic literature typically studies the reverse link: the potential effects of gender wage gaps and job instability on fertility decisions (eg., Vignoli et al. 2012, Wood and Neels 2017, Vignoli et al. 2020). Okun and Raz-Yurovich (2019) study the link between holding gender-egalitarian norms and fertility. Engelhardt et al. (2004) show that in terms of time order fertility and women's employment influence one another, but this study did not explore equality in any way. Goldscheider et al. (2013) study the role of within-couple wage parity on fertility. However, within-couple differences in wages does not need to reflect gender wage inequality. It is an empirical regularity that in a high fraction of heterosexual unions the man has a higher educational attainment than the woman. Meanwhile, in many advanced economies on average women are better educated than men. Baizan et al. (2016) look at policies which were intended to foster gender equality. To the best of our knowledge, no single study looks at the effects of changes in fertility on gender inequality in wages among labor market entrants. While economists are concerned about causal identification of the effects of parenthood once it occurs, demographers and sociologists devote attention to households/couples and their fertility decisions, which left this question of paramount policy relevance somewhat orphan. In addition, typically data limitations imply that studies cover a single country, which impedes deriving general conclusions, whereas methodological differentiation makes cross-country comparisons of the estimates available in the literature a challenge.

Our study aims to fill the existing gaps in several ways. First, we provide a novel and comprehensive collection of adjusted gender wage gaps *among labor market entrants*. We have harmonized nearly 1,200 individual-level data sets and obtained comparable estimates of AGWG among youth. This large collection of estimates allows us to purposefully ignore time-invariant country-specificity, such as culture, legal context or social norms. We discuss the details in the next section. Second, we explore the link which has so far slipped from the radar of social scientists of many disciplines: we study whether AGWG among labor market entrants declines, as employers receive sufficiently informative signals about delayed fertility of young women. We utilize several instruments to address the issue of endogeneity and thus provide causal estimates. Our results are discussed in detail in sections 4 and 5.

3 Data and methods

The research question at hand requires estimates of the adjusted gender wage gaps among labor market entrants, across countries and for subsequent birth cohorts. These estimates constitute the explained (dependent) variable. Given that no such data set exist, we collected individual-level data sets, harmonized them, and obtained 1,199 comparable estimates of adjusted gender wage gaps among youth. In this section we describe the availability of individual-level data across countries and years; in section 3.1, we discuss the harmonizing of the acquired data and measurement of adjusted gender wage gaps in section 3.2. The key explanatory variable of interest is maternal age at first birth. We discuss in detail the sources for this variable in section 3.3.

Given a large number of countries in our study and multiple time-periods for each country, our preferred specifications account for country fixed effects. However, this approach does not warrant causal interpretation as implied by our main hypothesis. More equal labor markets could raise the opportunity cost of becoming mothers, but this is at least partially compensated by the income effect for the two-earner households. In such case, the effects on fertility timing can operate in both directions: further delay of fertility or childbearing at a younger age. To address this point, we instrument for fertility using data on duration of compulsory schooling, data on military conscription and the authorization of contraceptive pills in a given country, which we describe in section 3.4.

We conclude the description of methods and data by descriptive statistics in section 3.5.

3.1 Data on gender wage gaps among youth

We collected a large number of individual-level data bases. We introduced only two restrictions on the data sets to be included in this study. First, the data set has to comprise sufficient information to compute an hourly wage. Second, the data has to report individual level characteristics, at least gender, age and education. We relied on Eurostat, Integrated Public Use Microdata Series from the University of Michigan and LISSY service provided by Luxembourg Income Study. These data sources provide comparable samples (or permit obtaining estimates on their samples) across numerous countries based on censuses (IPUMS) or on large representative samples (Eurostat and LISSY). In addition, we also utilized data from International Social Survey Program, which is based on representative, but smaller, samples. These cross-country sources were subsequently complemented by individual-level data obtained from central statistical offices or analogous institutions around the world. We obtain panel data for Canada, Germany, Korea, Russia, Sweden, Ukraine, the UK and the US. We obtain labor force survey data or household budget survey data from Albania, Argentina, Armenia, Belarus, Chile, Croatia, France, Italy, Poland, Serbia, the UK and Uruguay. This selection of countries was driven by the availability of hourly wages rather than our arbitrary choices. Finally, the World Bank in cooperation with local statistical offices provides Living Standards Measurement Survey for several countries around the world, including Albania, Bosnia and Herzegovina, Bulgaria, Kazakhstan, Kyrgyzstan, Serbia and Tajikistan. Appendix A discusses in detail each of the data sources.

Overall, we were able to collect data for 56 countries spanning the last four decades. These databases were harmonized in order to obtain comparable estimates of adjusted gender wage gaps. The dependent variable in the decomposition is hourly wage, which is derived based on usual hours worked and total pay without bonuses. The sample is restricted to individuals aged 20 to 30 years old. Education was harmonized to three levels: primary or less, secondary, and tertiary or more. In most of our data sets, we are able to identify household structure. The harmonized measures include a dummy variable taking on the value of one if there is a child in the household and zero otherwise.[6] We are also able to recover marital status, with a dummy variable taking on the value of one if individual is in a relationship and zero otherwise. Most data sets permit identification of the size of residence (urban vs rural).

In addition to these basic controls, we harmonized industry and occupation, whenever it was available. Industry variable was converted into a categorical variable with six levels agriculture, construction, manufacturing, market services, non-market services and utilities. Occupation variable was recoded to

[6] Data permitting, we used the threshold for the school age.

match one-digit International Standardized Classification of Occupations. For consistency with estimates distributed by LISSY, these categories were aggregated to three levels: managers/professionals (ISCO levels 1-2), laborers and elementary workers (ISCO levels 6-9) and the residual category of occupations.

3.2 Measuring the adjusted gender wage gaps

We decompose wage differences using the approach proposed by Nopo (2008). This method is based on exact matching. Consequently, the estimates account for wages of *comparable* men and women. Moreover, given that the approach is non-parametric, the resulting estimates are less sensitive to inaccurate model specification than regression-based decompositions. This feature is particularly important given that we apply the decomposition to a large collection of highly heterogeneous countries. Prior research found that estimates of the adjusted gap obtained using Nopo (2008) decomposition prove robust to the inclusion of additional control variables (Goraus et al. 2017), which is useful given that the full set of controls is not always available for each country and source.

One major advantage of Nopo (2008) decomposition is that it simultaneously reports the share of (*un*)matched men and women. This statistic is akin to reporting the share of individuals – young men or young women – for whom a statistically equivalent individual of opposite gender is (not) available in the data. The subsequent analysis focuses on those countries with a sufficiently high match fraction for both men and women, that is the young male and female workers are sufficiently similar in terms of individual characteristics.

For each harmonized data set, we identify the availability of control variables and obtain adjusted gender wage gaps for the most comprehensive set of controls, but also for subsets controls, with and without industry and occupation. *Ex ante*, it is unknown which fraction of individuals is matched based on the given set of controls; though more comprehensive control sets result in a lower probability of finding a person with such characteristics in the male and female subsamples. In small databases, the problem is more acute. In order to strike a balance between comprehensiveness of adjusted gender wage gap measure and the comprehensiveness of sample on which it was computed, we obtain a variety of estimates in each sample for combination of control variables. Given that for each sample we obtain multiple AGWG estimates, we subsequently utilize the estimate with the highest number of controls subject to the constraint that at least 75% of men and 75% of women are matched. As a robustness check, we replicate all the estimates on a collection of restricted estimates. Specifically, we study the subsample of estimates that fulfills both the criterion of 75% men and women matched *and* adjust for occupation and industry.

3.3 Data on maternal age at first birth

Our primary interest in this study is the decline in probability of childbearing by young female workers, the fertility timing. We measure this process using data on mean maternal age at first birth. In any given year, increases in the mean age at first birth serves as an indication that women postponed childbearing, and employers would believe that it is less likely that women below the former mean age at first birth would bear children.

We combine multiple sources to collect data for maternal age at first birth for the countries and years covered by the individual-level data. The data for most European countries is provided by the Eurostat (variable `AGEMOTH1`). United Nations Economic Commission for Europe (UNECE) as well as Organization for Economic Cooperation and Development (OECD) extend this data source to include some non-EU members and reports full time series from 1960 onward.[7] In addition, Human Fertility Database and Human Fertility

[7] Distributed online as http://www.oecd.org/els/family/database.htm. The OECD database covers Canada, Israel, Japan, Korea, United Kingdom and the United States and reports specific sources. UNECE distributes data online as https://w3.unece.org/

Collection report maternal age at first birth for some developing countries around the world.[8] Bongaarts and Blanc (2015) provide data for a large collection of countries using Demographic and Health Surveys program. Data for China comes from He et al. (2019). Australian Bureau of Statistics provides full extent of first birth data by the age of the mother spanning 1975 to 2019, which we use to calculate the means for each year. The central statistical office from South Africa provides extensive birth data based on 2011 census. Last, Population Bulletin of the United Nations reports data for selected years in the case of Brazil.

Mean maternal age at first birth is especially well suited for our study. Mean age at first birth isolates first-time mothers (unlike mean age at birth, which accounts for second and subsequent births as well, thus confounding the postponement with the spacing of children at older ages). Total fertility rate measures the number of birth in a given year over the number of women aged 18 to 40 in that year, thus it confounds a decline with fertility with delay in fertility. Fraction of childless women can only be computed for those birth cohorts who completed fertility already (40 years or older), hence it does not refer to youth in a given year. Age-specific fertility rates are scarcely available across countries and years. Moreover, age-specific fertility rates includes higher parity births, for which the relation to the gender wage gap is less intuitive than for the first parity.

3.4 Instrumenting for fertility measures

We instrument for mean maternal age at first birth using the authorization of contraceptive pills. We complement this instrument with three sets of variables. We rely on drivers of family formation: (i) the length of compulsory schooling, and (ii) military conscription. In addition, we use (iii) fertility rate in the generation of mothers, that is we use total fertility rate lagged by 20 years. We describe these instruments and data sources below.

Pill authorization As the main source of identification we use the authorization of contraceptive pills. As the pill can be utilized as a medication against hormonal disorders, authorization – a purely administrative procedure – does not imply automatically access to the pill for contraceptive reasons. Whether or not an authorized medicament is available at all, via prescription or over the counter, and to all or to some selected groups of individuals closely tracks the social and gender norms, being thus endogenous to fertility timing (mean age at first birth). However, mere authorization stems from a conclusion of procedure verifying if a given product fulfills the public health criteria established by regulatory authorities, independent across countries. Indeed, Finlay et al. (2012) report wide dispersion concerning the channels of distribution and administering. For example, in some countries, once authorized, contraceptive pills were initially solely distributed as treatment for hormonal disorders, whereas in other countries it was available only to married women.

The literature on pill and women's labor supply decisions is rich.[9] To the best of our knowledge, there is no literature studying the role of pill in forming employers' beliefs. Most of our AGWG estimates correspond to cohorts joining the labor market in the mid to late 1990s rather than cohorts directly exposed to the introduction of the pill. This feature distinguishes our approach from the earlier literature: for the cohorts covered by AGWG estimates the link between the pill and fertility timing is mediated by a number of channels. First, there appears to be improvement in the quality of parents, which may strengthen the shift in

[8] Distributed online at https://www.humanfertility.org/ and https://www.fertilitydata.org/.

[9] Women postponed marital decisions and were more likely to become professionals in non-traditional sectors (Goldin and Katz 2002), women worked more hours at both extensive and intensive margins Bailey (2006). Accordingly, a significant portion of the reduction of the *raw* gender wage gap between 1980 to 1990's can be attributed to the pill (Bailey et al. 2012). Importantly, the effects likely spilled-over to the following cohorts. Women who had access to the pill postponed childbearing until finishing education, which increased quality of parenting without reduction in completed fertility (Ananat and Hungerman 2012). Fertility declined for women without tertiary education Bailey (2010).

general social norms towards gender equality in reaching professional aspirations (Ananat and Hungerman 2012). Second, inter-generational transmission of norms between mothers and daughters is a recognized phenomenon (Booth and Kee 2009, Kolk 2014, Boelmann et al. 2020), thus earlier authorization of pill in a given country raises the odds that younger generations, which we analyze in this study, had mother's generation with access to the pill. Third, prevalence of pill hints that fertility is more likely to be timed in line with professional career, which may be viewed by employers as potential for bargaining (even if only indirect).

Compulsory education The reforms in compulsory education have been previously demonstrated to causally affect fertility timing and level (Black et al. 2008, Cygan-Rehm and Maeder 2013). We use data on the number of years in compulsory education. These data are provided by UNESCO for all countries as of 1998. For the years before 1998 we infer the years in compulsory education from the available Brunello et al. (2009), Murtin and Viarengo (2011), Fenoll and Kuehn (2017). The data do not include pre-primary education. To fill the gaps for several countries missing from the sources listed above, we utilize country legislation as reported in Right to Education Initiative. In Canada, where the compulsory length of education is set at the level of provinces, we use estimates by Oreopoulos (2005).

The use of this instrument is not without weaknesses. Across countries there are substantial differences in the meaning of compulsory education. In some cases, the relevant metric would have been the legal school leaving age, whereas in others length of compulsory education is just as informative. For example, compulsory education in Mexico formally lasts 14 years, whereas it lasts 9 years in Czech Republic, with the school entry age at 6. However, in the latter case, the parents are legally bound to provide for child's education until the age of maturity, that is 18th birthday, which makes education *de facto* mandatory for 12 years and high school dropout rates are much lower in Czech Republic than in Mexico. The imperfection of this measure introduces noise to our first-stage estimates.

Military conscription The third instrument corresponds to the length of military conscription in months. The length of military conscription can drive mean maternal age at first birth through several channels. On the one hand, longer military conscription could lead to postponing fertility. First, longer periods of conscription could lead men to postpone family formation until the conscription is over. Second, even if men become married before enlisting, they are could be deployed or relocated, thus facing obvious obstacles to conceiving children. Accordingly, compulsory military service may rise mean age at first birth among their partners. On the other hand, military service provides stable, guaranteed income, which may reduce earnings uncertainty and thus encourage child-bearing. Similarly, military service can provide skills relevant for future employers, thus raising earnings potential among men and encouraging child-bearing. Given that military conscription can work in either way, we do not hypothesize on the sign of the coefficient in the first stage regression. We use data on the number of months as provided by Mulligan and Shleifer (2005) and extend it for time and countries using the same source, that is the *Military Balance* which is published annually. We supplement this database with the records of War Resisters' International and the World Factbook.[10] This variable exhibits variation over time and countries. This variable ranges from zero, i.e. there is no conscription, to 48 months in Israel. Given that in some countries the duration is established as a range, models include two instruments: one for the lower bound and one for the upper bound.

Fertility among mothers' generation The fourth instrument utilizes data on total fertility rate in the generation of mothers of the individuals in our sample. For example, if a sample for a given country comes from 2000, and we restrict the individuals used in the estimation to between 18 and 30 years of

[10] https://www.wri-irg.org/ and https://www.cia.gov/the-world-factbook/

age, we take the data for 1980 (= 2000 - 20) for that country. By the year of our sample the birth cohort of mothers has completed fertility. There is broad evidence for the inter-generational transmission of fertility norms covering both the demographic transition of the 19th Century (Pearson and Lee 1899) and the current demographic changes (Steenhof and Liefbroer 2008, Kolk 2015).which makes this instrument plausibly correlated with mean age at first birth measured contemporaneously. Meanwhile, since the present cannot affect the past, the adjusted gender wage gaps in the generations of the children could not have affected the realized fertility of the mothers two decades earlier.

Among the four instruments in our study, military conscription, compulsory schooling and lagged fertility have country-by-year variation, whereas the pill authorization is essentially one year for each country. For each country-year sample in our data we construct an indicator measuring how many years have lapsed since the introduction of the pill. In our preferred specifications, we use Baltagi (1981) estimator, which is efficient in a setup with both time-invariant and time-varying instruments.

3.5 Descriptive statistics

In total, we obtained 1,233 unique estimates of adjusted gender wage gaps computed for individuals aged 18 to 30 years old. This collection of estimates covers 51 countries for 38 years. The specific number of estimates for each country and source is reported in Table A1. This number of estimates reflects the unique combinations of country, source and year. For each sample, we estimated decompositions with an increasing number of controls. For each country source and year, we kept the specification which maximizes the number of control variables conditional on matching at least 75% of individuals of each gender. Given that some samples are relatively small, we restrict the study to those databases that contained at least 100 valid observations for each gender. This restriction affects chiefly estimates obtained from ISSP data, and is virtually inconsequential for the remaining sources. The final sample includes estimates for 1,161 countries, years and data sources for estimates of AGWG among youth and 1,301 for estimates of AGWG for the working age population.

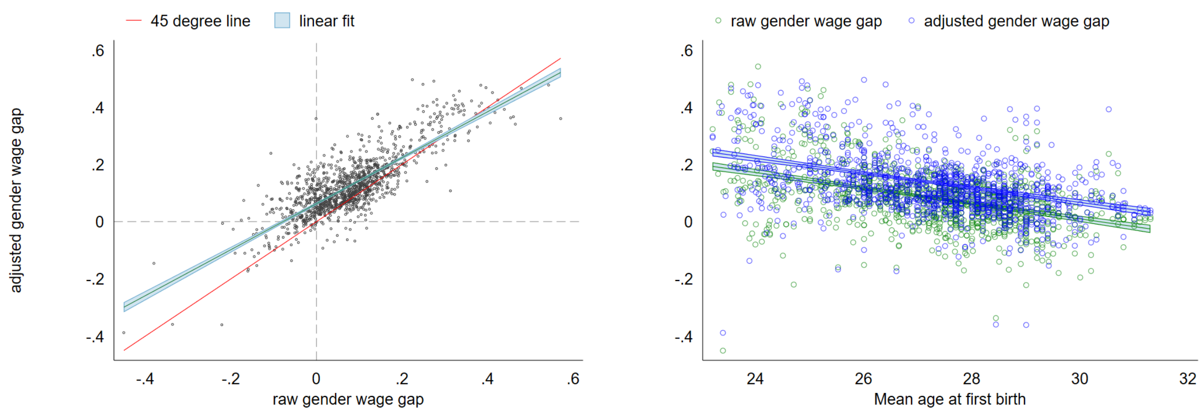


Figure 1: Raw and adjusted gender wage gap among youth and mean maternal age at first birth

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. Models estimated with country fixed effects and source fixed effects, standard errors clustered at the level of country and data source. Given that for each year and country more than one source may be available, we account for multiple observations through weights: $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of estimates for given country in a given year.

The left panel of Figure 1 presents the raw and adjusted gender wage gaps among youth in our sample. The right panel of Figure 1 portrays the overall correlation between mean maternal age at first birth and gender wage inequality measured by raw wage gaps and adjusted wage gaps. Adjusted gender wage gaps are located above the 45 degree line in the left panel, which signifies that the adjusted gender wage gap is higher than the raw gender wage gap. This is true for the majority of samples in our data, and is confirmed at the mean (12.2% versus 7.9%, respectively) and at the median (10.3% versus 7.0%, respectively); both differences are highly statistically significant. Adjusted gender wage gap in excess of the raw gap implies that if men and women were rewarded equally, women's wages would have been higher than they are in the observational data. The raw and the adjusted gender wage gaps appear fairly similar and closely correlated (the correlation coefficient is 0.80, $p = 0.000$). We rely on adjusted gender wage gap, as using the raw value could lead to underestimating the effects.

Note that the levels of raw and adjusted GWG in our study are different than the ones reported in reviews by e.g., Blau and Kahn (2017). This is because our study reports gaps for youth, as opposed to individuals in working age in general. Given that we have individual level data for individuals across all age groups, we additionally compute the gaps for all working age individuals as well, and these prove to be comparable to estimates reported in earlier studies. They also exhibit similar time trends. Table B1 reports the time evolution of raw gender wage gap in our sample, adjusted gender wage gap as well as mean maternal age at first birth. The portrayed time trends adjust for data and country composition and thus are not driven by the differentiated data availability across countries. In columns (1) and (2) we report the time trend for the working-age populations, whereas in columns (3) and (4) we report analogous estimates for the youth. Earlier literature reports essentially no time trend for the adjusted gender wage gap at the country level (Weichselbaumer and Winter-Ebmer 2005), which we replicate in our data spanning nearly two decades more. We add evidence on time trends for adjusted gender wage gaps among young individuals. This trend is negative, but the decline is slow: 0.16 of a percentage point each year. Given an average adjusted gap of roughly 12%, it would take better part of a century for the gap to disappear. By contrast, time trend on mean maternal age at first birth is large and positive. On average, it rises by a full year, every twelve years. Notwithstanding, the trend towards a higher mean maternal age at first birth is not universal. For numerous countries in our sample this statistic falls over time. We portray the full distribution of our GWG data and mean maternal age data in Figure B1 in the Appendix.

4 Results

The model of interest is the model of fertility timing (FT) and adjusted gender wage gap among youth ($AGWG$). It is given by

$$AGWG_{i,s,t} = \alpha + \beta \times FT_{i,t} + \gamma time + \xi s + \epsilon_{i,s,t}, \tag{1}$$

where i denotes country, t denotes time, and s denotes data source and specification. The coefficient β is subject to endogeneity bias, hence the IV approach. We estimate a two-stage IV model of the form:

$$AGWG_{i,s,t} = \alpha + \beta^{IV} \times \widehat{FT}_{i,t} + \gamma time + \xi s + \epsilon_{i,s,t} \tag{2}$$

$$FT_{i,t} = \phi + \theta PILL_{i,t} + \varrho EDU_{i,t} + \mu CONSCR_{i,t} + \varsigma M\_FERT_{i,t} + \varepsilon_{i,t}, \tag{3}$$

$PILL$ denotes the instrument obtained from contraceptive pill authorization, EDU corresponds to compulsory schooling duration, $CONSCR$ is based on military conscription data. Finally, M_FERT utilizes variation in total fertility rates from the period when the mothers generation was in reproductive age. Due to the nature of our instruments, we utilize the random effects estimator proposed by Baltagi (1981) with fixed effects for data source and specification used to obtain $AGWG_{i,s,t}$. As is conventional in the literature (Heckman and Navarro-Lozano 2004, Mogstad and Torgovitsky 2018), the estimates of θ, ϱ, μ

and ς are vectors accounting for cross-sectional component of each variable, time-varying component of each variable; for levels and basis functions up to the forth polynomial for each variable. Our preferred specification utilizes all available instruments, but for robustness we provide estimates that use subsets of instruments. The time trend is estimated as common across countries, otherwise we would not be able to identify β parameter separately from γ parameter. Estimating the model given by equations (2) and (3), we cluster model disturbances at the level of country, data source and AGWG specification controls.[11] For the sake of comparison, we estimate also the models for fertility levels (FL) and adjusted gender wage gaps in the overall working age population.

We report the results in two substantive parts. First, we focus on our empirical exercise, reporting the estimates from panel regressions and instrumental variables estimation in section 4.1. These results inform about the strength of the statistical relationship between mean maternal age at first birth and the estimates of the adjusted gender wage gap. Based on these results we could reject the hypothesis of fully rational statistical discrimination if there was no decline of AGWG with the rise of mean maternal age at first birth. However, finding a negative and statistically significant coefficient does not imply that *only* statistical discrimination exist. To this end, in section 5 we provide a theory-disciplined data-driven benchmark for the obtained coefficients.

4.1 The effects of gradually delayed fertility on AGWG

Delayed fertility implies lower adjusted gender wage gap among labor market entrants. We estimate that a rise in mean maternal age at first birth by one year leads to a reduction of the adjusted gender wage gap among labor market entrants of roughly 2-3 percentage points. We provide estimates for a broad array of specifications in Table 1. In columns (1)-(3) we report the panel IV estimations, whereas column (4) reports analogous results from a panel OLS estimation. Column (1) reports the coefficient for mean maternal age at first birth, when all four instruments are utilized. Subsequent two columns report estimates for specifications with the pill authorization as the only instrument in column (2) and for the remaining three instruments in column (3). Panel A reports the results for all obtained AGWG estimates. As a robustness check, in Panel B we report the estimates analogous to Panel A, but on a sample restricted to AGWG specifications adjusting occupation and industry. This restriction implies a drop of from 1100+ samples (country $\times$ year $\times$ data source) to 800+ samples. The estimates remain essentially unaffected by the restriction.

The estimated effect of 2-3 percentage points implies that a change in fertility timing amounts to a decline between 20% and 30% of the average inequality in wages adjusted for differences in individual characteristics. This magnitude is robust to the inclusion of multiple control variables. The $F-statistics$ reported in Table 1 are large, well above the conventionally assumed thresholds (Lee et al. 2020) which speaks to relevance of the utilized instruments. Admittedly, with the Baltagi (1981) estimator, the $F-statistics$ explores both the cross-section and the time in the variation of instrumental variables.[12]

Consistent with our conjectures, there are no effects if we use overall fertility rather than fertility timing. Columns (5)-(8) report panel estimations when we use level of fertility (TFR) rather than timing of fertility (mean maternal age at first birth). The estimates in column (5) are directly comparable to column (4): the dependent variable is the same and the specification is identical, but we use TFR rather than mean age at first birth. The coefficient becomes lower a much less precisely estimated, it is positive and insignificant for AGWG estimates across all age groups. Similarly insignificant is obtained in specifications for TFR and raw

[11] Note that we utilize the "best" specification for a given country, year and source. The "best" specifications differ across countries and sources, depending on data availability in the original data set and on the statistical properties of the obtained AGWG estimates.

[12] The Baltagi (1981) IV estimator does not have weak instruments test. We provide the $F-statistics$ as a result of a Wald test for an auxiliary estimation of the Baltagi (1981) model, excluding the controls for AGWG estimation controls and data source dummies, as these vary at the level of the dependent variable, but not along mean age at first birth.

Table 1: The effect of delayed fertility on AGWG

AGWG estimates estimates	Youth, MAB, AGWG IV β^{IV}			OLS β	Youth TFR, AGWG, OLS β	All	Youth TFR, RGWG, OLS	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Panel A: *Full sample*								
Fertility timing	-0.025***	-0.033***	-0.025***	-0.019**				
	(0.0062)	(0.0093)	(0.0065)	(0.0097)				
Fertility level					-0.016	0.013	0.0026	0.019
					(0.026)	(0.022)	(0.036)	(0.033)
Observations	1106	1161	1114	1170	1241	1301	1241	1301
R-squared	0.28	0.28	0.28	0.76	0.74	0.83	0.62	0.80
F-statistic	25286.6	461.1	15667.1					
Panel B: *Sample restricted to AGWG specifications adjusting for occupation and industry*								
Fertility timing	-0.026***	-0.032***	-0.025***	-0.020***				
	(0.0051)	(0.0069)	(0.0053)	(0.0060)				
Fertility level					-0.0045	-0.011	0.018	0.033
					(0.025)	(0.018)	(0.045)	(0.038)
Observations	825	864	834	873	938	939	938	939
R-squared	0.28	0.28	0.28	0.62	0.57	0.80	0.67	0.83
F-statistic	7058.1	379.5	3894.7					
Clustering SE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: IV specifications using Baltagi (1981) estimator with time varying and time-invariant component, 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. All specifications include time trends and adjust for type of data source and AGWG model specification. The OLS specifications adjust for $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of sources for given country in a given year. All columns report estimates for labor market entrants (under 30 years of age), except columns (6) and (8), which report estimates obtained for all working-age birth cohorts. Columns (4)-(8) adjust for country and specification fixed effects and time fixed effects. Standard errors clustered at country-source-controls level. In Panel A, for each country, year and data source we utilize one estimate, 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 impose an additional restriction that the AGWG estimates adjust for occupation and industry. 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.

gender wage gaps, among youth and across all age groups: the sign is flipped to positive and insignificant, despite a larger sample, see columns (7) and (8).

4.2 Robustness

While Baltagi (1981) is particularly well suited to our data, we study the sensitivity of the overall conclusions to alternative instrumental variable estimators. The OLS estimate of β reported in column (4) of Table 1 adjust for country fixed effects, which we replicate with in columns (1) and (2) of Table 2. We employ conventional fixed effects IV in column (1) of Table 2, subsequently high-dimensional fixed effects IV estimator in column (2). Both prove to deliver highly statistically significant estimates of β^{IV} and of similar magnitude as the IV estimates reported in Table 1.

Indeed, whether we use the Baltagi (1981) estimator or its alternatives, the estimates prove similar to the OLS in economic terms. In statistical terms, the IV specifications yield estimates roughly 25% higher than in the OLS specifications, yielding the difference from roughly 2 percentage points to 2.5-3 percentage points. Ishimaru (2022) proposes to decompose the OLS-IV gap into the weights implied by covariates, the weights implied by treatment, as residual, the endogeneity bias. When we apply this approach to the estimates obtained in column (1) of Table 2, each of the three components proves insignificant for the full sample of AGWG estimates. In the restricted sample, there is some role for the weights implied by treatment levels. Likewise the endogneity bias remains insignificant. In other words, the reverse causality bias is not large in statistical terms: in a sample covering over fifty countries spanning four decades the timing of the first child does not appear to be driven by the prevailing adjusted gender wage gaps, whereas the weights implied by treatment levels exhibit an effect roughly 25% higher than the OLS estimates.

We also employ a quantile IV estimator to study the magnitude of the effects of fertility along the distribution of AGWG. The results are presented in columns (3)-(5) for the 25th percentile, median and 75th percentile, respectively. This estimation uses unconditional quantiles of AGWG via recentered influence function (Firpo et al. 2009) and model specification analogous to column (1) from Table 1. The estimated effect of fertility timing on AGWG amounts to roughly -0.02 for the 25th and 50th percentile, it appears to be somewhat higher for the 75th percentile, but still within the ballpark of Baltagi (1981) estimates reported in Table 1.

Table 2: The effect of delayed fertility on AGWG - alternative estimators of β^{IV}

Fertility timing	FE 2SLS (1)	HDFE (2)	Quantile Regression Q25 (3)	Quantile Regression Q50 (4)	Quantile Regression Q75 (5)	Heterogeneous fertility Intercepts (6)	Heterogeneous fertility Slopes (7)
FT	-0.022 ***	-0.038 ***	-0.023 ***	-0.022 ***	-0.032 ***		
	(0.008)	(0.004)	(0.004)	(0.004)	(0.007)		
FT < Q25						0.139 ***	-0.018
						[0.07,0.20]	[-0.05,0.02]
FT ∈ [Q25, Q75]						0.030	-0.019
						[-0.02,0.08]	[-0.05,0.01]
FT > Q75							-0.019
							[-0.05,0.01]

Notes: Standard errors 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 1. In column (1) we use fixed effects IV estimator. In column (2) we use HDFE IV estimator. In columns (3)-(5) we utilize Firpo et al. (2009) recentered influence function transformation of the estimator from Column (1) in Table 1, the transformation is for the 25th, 50th and 75th percentile of AGWG. In columns (6) and (7) we provide estimates of quantile IV estimator, accounting for the quantiles 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. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from first and second stage regressions is available upon request. Analogous set of estimates for sample restricted as in Panel B of Table 1 is reported in Table C1 in Appendix C.

Finally, we explore the heterogeneity along the distribution of mean age at first birth. We do it for the intercepts of the link between mean age at first birth and AGWG and for the slopes. We split the sample into low (below the 25th percentile), medium and high (above 75th percentile) mean maternal age at first birth. For the intercepts, we take a set of dummies for medium and high mean maternal age at first birth as endogenous variables. For the slopes, there are three endogenous variables: the first takes on the value of mean maternal age at first birth if it is low and zero otherwise, whereas the second and the third take on the medium and high values, respectively. In all other respects, these specifications are analogous to column (1) from Table 1. The level effect appears to be the highest when mean maternal age is within the first quartile, but this observation may partly be a consequence of the fact that the fourth quartile in terms of fertility timing is very close to the upper boundary on the age in the estimation of adjusted gender wage gaps.[13]

4.3 Discussion

The correlation as well as the causal estimates of delayed fertility on gender wage gaps at young age reveal an effect of roughly 2-3 percentage points. Given that mean maternal age at first birth in our sample increased from 25.96 in 1990s to 28.16 in 2010s, these estimates imply that essentially a third of the decline in AGWG among labor market entrants can be attributed to delayed fertility.

The instruments used in our study have been previously demonstrated to affect various margins of fertility, labor supply and human capital in individual level investigations. It is thus not warranted that

[13] Indeed, Q25 threshold is 26 years of age and Q75 threshold is 28 years, whereas we estimate AGWG until the age of 30. There may be too little variation in MAB to yield a statistically significant intercept in Q2-Q4 of the MAB distribution.

fertility timing is the only channel through which our instruments affect the adjusted gender wage gaps. Ours is a country-level study, however and to the best of our knowledge, there is no evidence utilizing our instruments in the comparative studies of labor supply decisions of women, tertiary enrollment, etc. Notwithstanding, we study if and to what extent our instruments could operate directly, that is in addition to instrumenting for fertility timing. We explore three direct channels: employment, education and (contemporaneous) fertility. In addition, we account for GDP per capita as a measure of labor productivity (akin to the opportunity cost of not working). In Table C2 we show that once we account for these direct channels, the IV estimates remain in the same ballpark (despite considerable reduction in the countries and years due to data limitations).

Our approach rests upon using several instruments simultaneously. In practice this requires that, conditional on all other instruments, no remaining instrument should affect the endogeneous variable in opposite directions across countries or years (partial monotnicity, see Mogstad et al. 2021). Our results remain robust to including one instrument at the time and combination of instruments, which is somewhat reassuring that partial monotonicity is not violated in our case.[14].

It the next section we provide data-driven benchmarks for these estimates.

5 Benchmarking statistical gender discrimination

To establish a benchmark for our estimates, we consider the following framework. Parents draw a parental type c_i from a distribution, with $E_m(c_i) = c_m < E_w(c_i) = c_w$, which denotes how involved they will be in the care of the child.[15] In principle, c_i can take any real value. While for most parents c_i would be positive, some parents might have no cost of caring, and some might even see positive productivity spillovers from having children, e.g. if having a child improves their motivation to work. The assumption of $c_m < c_w$ states that costs are on average higher for women. Parental type is private information, and it cannot be communicated to the employer ex ante (before becoming a parent), whereas ex post (after becoming a parent) credible communication is costly, e.g., it depends on child characteristics which may vary over time.

Individuals draw a procreation type: with probability π_i an individual becomes a parent, bearing the associated productivity costs within the window of contract duration. With the complementary probability they remain without children for the duration of the contract. The procreation type is unobservable ex ante and for the sake of simplicity it is unknown to the worker (that is the procreation intentions may turn out impossible to implement, or lack thereof can still be associated with unintended procreation). Becoming a parent (π_i) is independent of the productivity change associated with parenting costs (c_i) and it is drawn from a distribution common for both genders.

Conditional on human capital h, workers' productivity equals $h - c_i$ if they bear the costs of parenthood c_i which occurs with probability π_i or h if they have not become parents yet. The employer can observe human capital h, but before parenting, the employer cannot know individual π_i. Moreover, even after becoming a parent, the employers cannot costlessly observe c_i. The expected productivity of the worker and thus the wage w is given by

$$E(w) = E(\pi_i(h - c_i)) + E((1 - \pi_i)h) = h - E(\pi_i c_i) = h - E(\pi_i)E(c_i), \tag{4}$$

where the last equality follows from the fact that $cov(\pi_i, c_i) = 0$.

While statistical discrimination may be viewed as an injustice, as much as any form of group responsibility, it is also conceived as economically rational under information asymmetry when no credible separating

[14]The full set of estimates available upon request
[15]Implicitly we assume that the parental type maps one-to-one to the productivity costs c_i.

equilibrium exists. The upside of statistical discrimination is that this hypothesis relies on rationality of employers: it is the cost $E_w(c_i) - E_m(c_i) > 0$ rather than distaste or nonexistent differences in h that explains differences in wages between women and men. Under statistical discrimination, the employers offer wages in expectation of individual productivity, averaging over groups. With probability π_i of bearing the parenthood costs, the adjusted gender wage gap becomes:

$$\begin{aligned} AGWG = E_m(w|h) - E_w(w|h) &= \big(h - E_m(\pi_i)E_m(c_i)\big) - \big(h - E_w(\pi_i)E_w(c_i)\big) \\ &= E_w(\pi_i)E_w(c_i) - E_m(\pi_i)E_m(c_i) = E(\pi_i) \times (c_w - c_m) \end{aligned} \tag{5}$$

Thus, if statistical discrimination stands behind (adjusted) gender wage gaps, a decline in $E(\pi_i)$ *ceteris paribus* should imply narrowing of the AGWG. Note that even if π_i is the same across all individuals, there is still averaging due to unobservable differences in productivity (c_i) between parents and non-parents as well as between mothers and fathers.[16]

Equation (5) takes the employers' perspective and portrays the link between *objective* differences in productivity across genders and the (adjusted) gender wage gap. We aim to obtain analogs of $c_w - c_m$ and $E(\pi_i)$ from observational data to compare our estimates of $E(w_m|h) - E(w_w|h)$. We then juxtapose $E(\pi_i) \times (c_w - c_m)$ to empirical estimates of $E(w_m|h) - E(w_w|h)$. Rationality imposes that the employers form expectations based on the observed probability $E(\pi_i) = \pi$ of incurring the cost c, thus the two should be equal if statistical discrimination is an important reason for gender wage gaps among youth. Such employers should also internalize the actual cost in terms of productivity loss c. Indeed, productivity loss is akin to a reduction in productivity endowment, with endowment identical across men and women, with and without children. Clearly, child-bearing is not the only reduction to the time endowment. Individuals may have other caring responsibilities and social norms may be driving the gender distribution of those functions.

5.1 Implementation

Denote $p(a)$ to signify age (a) specific fertility rates for the first parity. Then, the probability of becoming a parent during a contract window (conditional on not being a parent at the moment of hiring) is given by

$$E(\pi_i|childless) = \int_{a=20}^{a=30} p(a)da \Big/ \Big(1 - \int_{a=18}^{a=20} p(a)da\Big). \tag{6}$$

Effectively, equation (6) an upper bound expectation in a sense that the contract duration with a given employee can be shorter than the age brackets $a \in [20, 30]$ (in other words: the actual probability faced by the cannot be higher).[17] The age brackets of between 20 and 30 are set to be consistent with the age groups for which adjusted gender wage gap was estimated. We construct this indicator using information about the number of first births to women of a given age in a given year. The data comes from the Eurostat for European countries and from the Human Fertility Database for the US.

To capture $c_w - c_m$, we resort to the time-use data.[18] In the time-use surveys, household members report the time spent on caring. We distinguish between parents in households with children and independent

[16] If changes in $E(\pi_i)$ over time are accompanied by commensurate or even more pronounced changes in $c_w - c_m$ of the opposite direction (Kuziemko et al. 2018), AGWG will not change or may even increase despite employers adjusting their expectation $E(\pi_i)$. For this reason we need to obtain potentially time-varying measures of $c_w - c_m$ from the data.

[17] This measurement assumes implicitly that within the ten-year window, age specific fertility rates for the age groups of interest do not vary substantially.

[18] We obtain time-use surveys from the Center on Time Use Research at University College London. The center provides Multinational Time Use Study, which is an effort to harmonize the available time-use surveys. Time-use surveys are implemented rarely (typically once every ten years) and their harmonization is a challenge, which makes MTUS a unique source of time-use data. Given the lag in making the data available on MTUS, in several cases we were able to complement this data source with the time-use data distributed by IPUMS at University of Michigan.

adults in households without children. We isolate parents in households with children and independent adults in households without children. For each person we obtain the measure of time spent on caring. We then apply formula (7) to obtain the measures of $c_w - c_m$. We aggregate this individual level data to recover information on caring by men and women, with and without child-rearing responsibilities, aged between 20 and 30 years old. We construct four mean (median) measures: for men without kids, men with kids, women without kids and women with kids. Based on these data, we compute the reduction in the time endowment of T hours per week by the mean (median) number of hours spent on caring and we obtain $c_w - c_m$ as:

$$c_w - c_m = \frac{\big((T - t_{w,k}) - (T - t_{w,\sim k})\big) - \big((T - t_{m,k}) - (T - t_{m,\sim k})\big)}{T}, \tag{7}$$

where t denotes time spent caring, w and m denote women and men, respectively, whereas k and $\sim k$ denote with and without children, respectively. Conventionally, we set $T = 80$ hours per week. The reduction in time endowment proxies the reduction in production capacity. A potential concern is related to potentially disproportionate allocation of household chores among mothers, when compared to childless men and men with children. To address this concern, we provide a set of analogous results with $c_w - c_m$ measure augmented with household chores.

We combine data availability of data-driven proxy of $c_w - c_m$ from the time use data with the value of π computed using the Human Fertility Database and the Eurostat data. The obtained measures of $c_w - c_m$ and π are available only for a small selection of the countries included in our sample: these are wealthy countries with several decades of rolling out several gender equality policies. Thus, we do this exercise to test if the ballpark of estimates obtained as our main results in Table 1 lies close to the simulated $(c_w - c_m) \times \pi$ from observational data. If that is the case, then it seems that the extent of adjusted gender wage gap coincides with accurate statistical discrimination. If, however, $(c_w - c_m) \times \pi$ falls short of the estimated AGWG, then either statistical discrimination is inaccurate or additional stereotyping is involved. The results are portrayed in Figure 2.

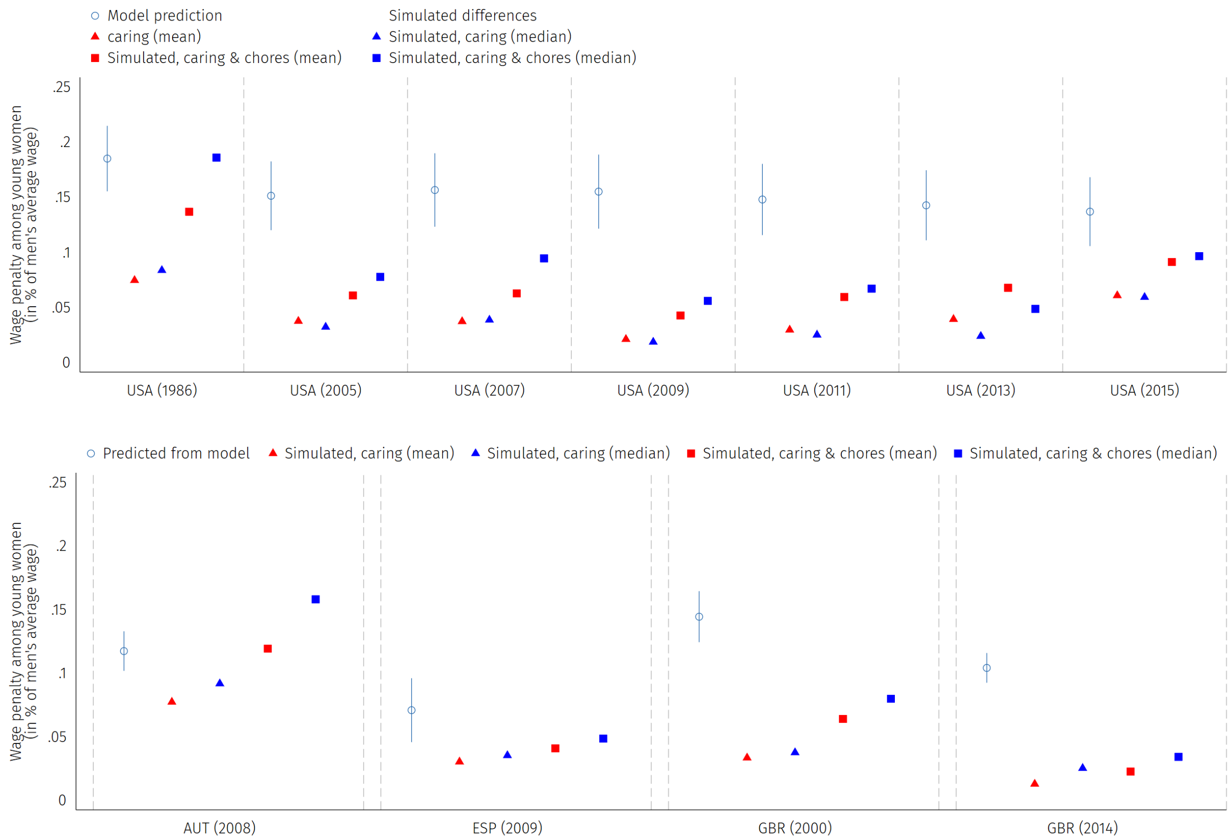
In all four countries, the estimates of AGWG are higher than most of the simulated values implied by $(c_w - c_m) \times \pi$. We infer that the existing adjusted gender wage gaps among youth are higher than the benchmarks implied by the data on the timing of fertility and actual reduction in time endowment due to caring obligations. The gap is rather small for Spain, but much higher for the UK and the US. In these two countries, availability of multiple waves of time-use permits tracing dynamics over time and the higher values of $(c_w - c_m) \times \pi$ tend to be associated with higher values of AGWG estimates, *ceteris paribus*.

In Appendix D, we present analogous estimates using data from International Social Survey Program (ISSP). Using the ISSP data allows us to expand the list of countries to be studied (in total: fourteen out of 56 countries in our estimations), but self-reported time-use data from ISSP may be subject to measurement error, which could be gendered. Moreover, sample sizes for individuals aged 20 to 30 years old are relatively small, when split between with and without children in ISSP. Notwithstanding, the replication of the benchmarking exercise for a broader group of countries reveals a similar conclusion: for all countries at least one of the measures falls in the confidence interval of the AGWG estimates.

Finally, in the case of the United States, the fertility evolution is highly volatile. Specifically, over the available period the π first declined substantially, to then rise and decline again. This volatility implies that the employers who would – as we propose – want to adjust the wage offer to the evolution in π received mixed signals about the actual probability of childbearing over the analyzed period. In Appendix E we discuss the special case of the United States, demonstrating the link between our findings and earlier work of Kuziemko et al. (2018).

Note that the measurement of AGWG in our study varies with the changes in adjusted gaps for each subgroup in the labor market, as well as with the composition change: the share of a given group in the

Figure 2: Obtaining a benchmark for statistical gender discrimination – time use surveys



Notes: the data come from the available time-use surveys obtained from MTUS project and I-PUMS. Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (2), for the year of MTUS availability for each country. Analogous simulations for the ISSP data are reported in Appendix D.
MTUS for the United Kingdom is available for 2000 and 2014. Data availability for the US is described in Appendix E. If more than one time-use survey is available for a given country, we use the most recent available survey. Simulations at the mean and at the median utilize Eurostat and Human Fertility Database age-specific fertility rates.

total. For example, as higher fraction of men have a university degree, the contribution of the gap between men and women with this level of education to the total AWGW becomes larger.[19] Consequently, a decline in AGWG may imply both a decline in gaps within specific subgroups and a shift of workers towards those subgroups where the gaps are lower. Likewise, the employers may hold heterogeneous beliefs about $(c_w - c_m) \times \pi$ across subgroups of workers. It is thus possible that the match between $(c_w - c_m) \times \pi$ and AGWG is driven by composition effects. Further research disentangling composition effects from within group changes is called for.

6 Conclusions

Statistical discrimination – regardless of its legal status and ethical consequences – stems from the idea that rational employers internalize productivity gaps when maximizing the expected payoff from hiring a worker. Consequently, hiring workers who are expected to deliver lower productivity, the employers discount that fact in wages. For statistical discrimination to be consistent with the data, employers need to adjust their expectations concerning the productivity gaps. A delay in fertility observed around the world over the past decades provides a convenient context for evaluating if (adjusted) gender wage gaps among labor market

[19] Specifically, Nopo (2008) is a weighted average of all the gaps, where the weights are given by the share of a given men's subgroup in total population.

entrants are consistent with the hypothesis of statistical discrimination. In this study we provide estimates of adjusted wage gaps between young men and women from 56 countries around the world, spanning four decades and compare those estimates with the evolution of mean maternal age at first birth.

We find significant effect of delayed fertility on adjusted gender wage gaps among youth. This result proves robust to the estimation method. The effect estimated through instrumental variables amounts to roughly 2 percentage points decline in AGWG per one year delay in the first parity. This effect is sizable, amounting to 15% of the overall youth adjusted gender wage gap and about 30% of the observed decline in AGWG among young workers over the past decades.

The fact that AGWG for young workers declines with delayed fertility is not proving that entire gap is due to statistical discrimination: employers could adjust slowly their expectations to increases in the mean maternal age at first birth, or they may view the productivity costs implied by motherhood as higher than they actually are. Both inaccuracies would imply stereotypes and heuristics inconsistent with the hypothesis of statistical discrimination. To address this issue we provide simulations for the productivity gap of young women, relative to men. For some countries, the implied productivity cost of parenting is well aligned with the range of AGWG implied by our model. We also illustrate that for a group of European countries the range of AGWG estimates largely exceeds the productivity costs implied by the data-driven probability of first parity among young women active in the labor market.

Our study contributes by demonstrating that in general the employers correctly receive the signal about the the changes to the probability of child bearing and adjust downwards AGWG in the light of delayed fertility. This adjustment is accurate in terms of magnitude for some countries, whereas in others we are able to show that the estimates of AGWG are in excess of what would be justifiable given the observed distribution of age at first birth and the costs associated with motherhood. This may explain why audit studies on motherhood penalty finds such conflicting evidence: from strong discrimination against would-be mothers in some countries to virtually no differences in call-back rates.

While our study is able to bridge several gaps in the existing literature, caution is needed in interpreting the results. In terms of data, our study covers 56 countries, some of which have yet to undergo the second demographic transition. Although extending the study to comprise other countries is currently impossible, our estimates do not need to apply to employers in countries where individuals aged 20 to 30 years old have children with near certainty. Data limitations in terms of individual level wage data and demographic data on first parity constrained our ability to study countries with high levels of fertility rates and low age at first birth.

In terms of methodology, we introduce four instruments to identify the causal effect of delayed fertility on adjusted gender wage gaps. The statistical properties of the first stage regression appear satisfactory, yet the IV estimates are qualitatively very similar to the linear model. More research is called for to determine the magnitude of the reverse causality bias, that is to study the role of gendered labor market inequality in the timing of child birth.

In terms of policy implications, our study shows that probability of childbearing is reflected in wage offer for young women relative to men. This hints that greater equality sharing of the care between mothers and fathers can help the labor market position of young mothers and especially would-be mothers. Exploring further the role of sharing the care is a promising avenue for future research.

References

Aaronson, D., Dehejia, R., Jordan, A., Pop-Eleches, C., Samii, C. and Schulze, K.: 2021, The effect of fertility on mothers' labor supply over the last two centuries, *The Economic Journal* **131**(633), 1–32.

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.

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, Princeton University Press, pp. 3–33.

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.: 2018, Hiring discrimination: An overview of (almost) all correspondence experiments since 2005, pp. 63–77.

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.

Baltagi, B. H.: 1981, Simultaneous equations with error components, *Journal of Econometrics* **17**(2), 189–200.

Bartoš, V., Bauer, M., Chytilová, J. and Matějka, F.: 2016, Attention discrimination: Theory and field experiments with monitoring information acquisition, *American Economic Review* **106**(6), 1437–75.

Becker, G. S.: 1971, *The economics of discrimination*, University of Chicago Press.

Becker, S. O., Fernandes, A. and Weichselbaumer, D.: 2019, Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment, *Labour Economics* **59**, 139–152.

Bergmann, B.: 1973, Sex discrimination in wages: comment, Princeton University Press, pp. 152–154.

Black, S. E., Devereux, P. J. and Salvanes, K. G.: 2008, Staying in the classroom and out of the maternity ward? the effect of compulsory schooling laws on teenage births, *Economic Journal* **118**(530), 1025–1054.

Blair, P. Q. and Chung, B. W.: 2021, A model of occupational licensing and statistical discrimination, *AEA Papers and Proceedings* **111**, 201–05.

Blau, F. D. and Kahn, L. M.: 2017, The gender wage gap: Extent, trends, and explanations, *Journal of economic literature* **55**(3), 789–865.

Boelmann, B., Raute, A. and Schönberg, U.: 2020, Wind of change?: Cultural determinants of maternal labor supply, *CREAM Discussion Paper 2020-20*.

Bohren, J. A., Haggag, K., Imas, A. and Pope, D. G.: 2019, Inaccurate statistical discrimination, *NBER Working Paper 25935*, National Bureau of Economic Research.

Bohren, J. A., Imas, A. and Rosenberg, M.: 2019, The dynamics of discrimination: Theory and evidence, *American Economic Eeview* **109**(10), 3395–3436.

Bongaarts, J. and Blanc, A. K.: 2015, Estimating the current mean age of mothers at the birth of their first child from household surveys, *Population Health Metrics* **13**(1), 1–6.

Booth, A. L. and Kee, H. J.: 2009, Intergenerational transmission of fertility patterns, *Oxford Bulletin of Economics and Statistics* **71**(2), 183–208.

Brunello, G., Fort, M. and Weber, G.: 2009, Changes in compulsory schooling, education and the distribution of wages in Europe, *The Economic Journal* **119**(536), 516–539.

Bygren, M., Erlandsson, A. and Gähler, M.: 2017, Do employers prefer fathers? evidence from a field experiment testing the gender by parenthood interaction effect on callbacks to job applications, *European Sociological Review* **33**(3), 337–348.

Correll, S. J., Benard, S. and Paik, I.: 2007, Getting a job: Is there a motherhood penalty?, *American Journal of Sociology* **112**(5), 1297–1338.

Costa Dias, M., Joyce, R. and Parodi, F.: 2020, The gender pay gap in the uk: children and experience in work, *Oxford Review of Economic Policy* **36**(4), 855–881.

Cukrowska-Torzewska, E. and Lovasz, A.: 2020, The role of parenthood in shaping the gender wage gap–a comparative analysis of 26 European countries, *Social Science Research* **85**, 102355.

Cygan-Rehm, K. and Maeder, M.: 2013, The effect of education on fertility: Evidence from a compulsory schooling reform, *Labour Economics* **25**, 35–48.

Engelhardt, H., Kögel, T. and Prskawetz, A.: 2004, Fertility and women's employment reconsidered: A macro-level time-series analysis for developed countries, 1960–2000, *Population Studies* **58**(1), 109–120.

Ewens, M., Tomlin, B. and Wang, L. C.: 2014, Statistical discrimination or prejudice? a large sample field experiment, *Review of Economics and Statistics* **96**(1), 119–134.

Fenoll, A. A. and Kuehn, Z.: 2017, Compulsory schooling laws and migration across european countries, *Demography* **54**(6), 2181–2200.

Finlay, J., Canning, D. and Po, J.: 2012, Reproductive health laws around the world, *PGDA Working Papers 9612*, Program on the Global Demography of Aging, Harvard T. H. Chan School of Public Health.

Firpo, S., Fortin, N. M. and Lemieux, T.: 2009, Unconditional quantile regressions, *Econometrica* **77**(3), 953–973.

Flage, A.: 2018, Ethnic and gender discrimination in the rental housing market: Evidence from a meta-analysis of correspondence tests, 2006–2017, *Journal of Housing Economics* **41**, 251–273.

Fuller, S. and Cooke, L. P.: 2018, Workplace variation in fatherhood wage premiums: Do formalization and performance pay matter?, *Work, Employment and Society* **32**(4), 768–788.

Gangl, M. and Ziefle, A.: 2009, Motherhood, labor force behavior, and women's careers: An empirical assessment of the wage penalty for motherhood in Britain, Germany, and the United States, *Demography* **46**(2), 341–369.

Goldin, C. and Katz, L. F.: 2002, The power of the pill: Oral contraceptives and women's career and marriage decisions, *Journal of Political Economy* **110**(4), 730–770.

Goldscheider, F., Bernhardt, E. and Brandén, M.: 2013, Domestic gender equality and childbearing in sweden, *Demographic Research* **29**, 1097–1126.

González, M. J., Cortina, C. and Rodríguez, J.: 2019, The role of gender stereotypes in hiring: a field experiment, *European Sociological Review* **35**(2), 187–204.

Goraus, K., Tyrowicz, J. and van der Velde, L.: 2017, Which gender wage gap estimates to trust? a comparative analysis, *Review of Income and Wealth* **63**(1), 118–146.

He, D., Zhang, X., Wang, Z. and Jiang, Y.: 2019, China fertility report, 2006–2016, *China Population and Development Studies* **2**(4), 430–439.

Heckman, J. and Navarro-Lozano, S.: 2004, Using matching, instrumental variables, and control functions to estimate economic choice models, *Review of Economics and statistics* **86**(1), 30–57.

Hipp, L.: 2020, Do hiring practices penalize women and benefit men for having children? experimental evidence from germany, *European Sociological Review* **36**(2), 250–264.

Hsieh, C.-T., Hurst, E., Jones, C. I. and Klenow, P. J.: 2019, The allocation of talent and US economic growth, *Econometrica* **87**(5), 1439–1474.

Ishimaru, S.: 2022, Empirical decomposition of the iv-ols gap with heterogeneous and nonlinear effects, *Review of Economics and Statistics* pp. 1–45.

Kleinberg, J., Ludwig, J., Mullainathan, S. and Sunstein, C. R.: 2018, Discrimination in the age of algorithms, *Journal of Legal Analysis* **10**, 113–174.

Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimüller, J.: 2019, Child penalties across countries: Evidence and explanations, *AEA Papers and Proceedings*, Vol. 109, pp. 122–26.

Kleven, H., Landais, C., Posch, J., Steinhauer, A. and Zweimüller, J.: 2020, Do family policies reduce gender inequality? Evidence from 60 years of policy experimentation, *NBER Working Paper 28082*, National Bureau of Economic Research.

Kleven, H., Landais, C. and Søgaard, J. E.: 2019, Children and gender inequality: Evidence from denmark, *American Economic Journal: Applied Economics* **11**(4), 181–209.

Kolk, M.: 2014, Multigenerational transmission of family size in contemporary Sweden, *Population Studies* **68**(1), 111–129.

Kolk, M.: 2015, The causal effect of an additional sibling on completed fertility: An estimation of intergenerational fertility correlations by looking at siblings of twins, *Demographic Research* **32**, 1409–1420.

Kuziemko, I., Pan, J., Shen, J. and Washington, E.: 2018, The mommy effect: Do women anticipate the employment effects of motherhood?, *NBER Working Paper 24740*, National Bureau of Economic Research.

Lee, D. L., McCrary, J., Moreira, M. J. and Porter, J.: 2020, Valid t-ratio inference for iv, *preprint arXiv 2010.05058*, arXiv.

Matysiak, A. and Vignoli, D.: 2008, Fertility and women's employment: A meta-analysis, *European Journal of Population/Revue européenne de Démographie* **24**(4), 363–384.

Mogstad, M. and Torgovitsky, A.: 2018, Identification and extrapolation of causal effects with instrumental variables, *Annual Review of Economics* **10**, 577–613.

Mogstad, M., Torgovitsky, A. and Walters, C. R.: 2021, The causal interpretation of two-stage least squares with multiple instrumental variables, *American Economic Review* **111**(11), 3663–98.
URL: *https://www.aeaweb.org/articles?id=10.1257/aer.20190221*

Morse, A. and Pence, K.: 2020, Technological innovation and discrimination in household finance, *NBER Working Paper 26739*, National Bureau of Economic Research.

Mulligan, C. B. and Shleifer, A.: 2005, Conscription as regulation, *American Law and Economics Review* **7**(1), 85–111.

Murtin, F. and Viarengo, M.: 2011, The expansion and convergence of compulsory schooling in Western Europe, 1950–2000, *Economica* **78**(311), 501–522.

Nopo, H.: 2008, Matching as a tool to decompose wage gaps, *Review of Economics and Statistics* **90**(2), 290–299.

Norman, P.: 2003, Statistical discrimination and efficiency, *Review of Economic Studies* **70**(3), 615–627.

Okun, B. S. and Raz-Yurovich, L.: 2019, Housework, gender role attitudes, and couples' fertility intentions: Reconsidering men's roles in gender theories of family change, *Population and Development Review* **45**(1), 169–196.

Oreopoulos, P.: 2005, Canadian compulsory school laws and their impact on educational, *Analytical Studies Branch Research Paper Series 11F0019MIE*, Statistics Canada.

Pager, D. and Karafin, D.: 2009, Bayesian bigot? statistical discrimination, stereotypes, and employer decision making, *Annals of the American Academy of Political and Social Science* **621**(1), 70–93.

Pearson, K. and Lee, A.: 1899, On the inheritance of fertility in mankind, *Philosophical Transactions of the Royal Society of London. Series A* (192), 257–330.

Petit, P.: 2007, The effects of age and family constraints on gender hiring discrimination: A field experiment in the French financial sector, *Labour Economics* **14**(3), 371–391.

Phelps, E. S.: 1972, The statistical theory of racism and sexism, *American Economic Review* **62**(4), 659–661.

Pope, D. G. and Sydnor, J. R.: 2011, What's in a picture? evidence of discrimination from Prosper.com, *Journal of Human Resources* **46**(1), 53–92.

Raghavan, M., Barocas, S., Kleinberg, J. and Levy, K.: 2020, Mitigating bias in algorithmic hiring: Evaluating claims and practices, *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pp. 469–481.

Schwab, S.: 1986, Is statistical discrimination efficient?, *American Economic Review* **76**(1), 228–234.

Steenhof, L. and Liefbroer, A. C.: 2008, Intergenerational transmission of age at first birth in the Netherlands for birth cohorts born between 1935 and 1984: Evidence from municipal registers, *Population Studies* **62**(1), 69–84.

van Staveren, I., Tyrowicz, J. and van der Velde, L.: 2018, Does age exacerbate the gender wage gap? New method and evidence from Germany 1984–2014, *Feminist Economics* **24**(4), 108–130.

Vignoli, D., Drefahl, S. and De Santis, G.: 2012, Whose job instability affects the likelihood of becoming a parent in Italy? A tale of two partners, *Demographic Research* **26**, 41–62.

Vignoli, D., Tocchioni, V. and Mattei, A.: 2020, The impact of job uncertainty on first-birth postponement, *Advances in Life Course Research* **45**, 100308.

Weichselbaumer, D. and Winter-Ebmer, R.: 2005, A meta-analysis of the international gender wage gap, *Journal of Economic Surveys* **19**(3), 479–511.

Wood, J. and Neels, K.: 2017, First a job, then a child? Subgroup variation in women's employment-fertility link, *Advances in Life Course Research* **33**, 38–52.

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, Kyrgizstan, 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.

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

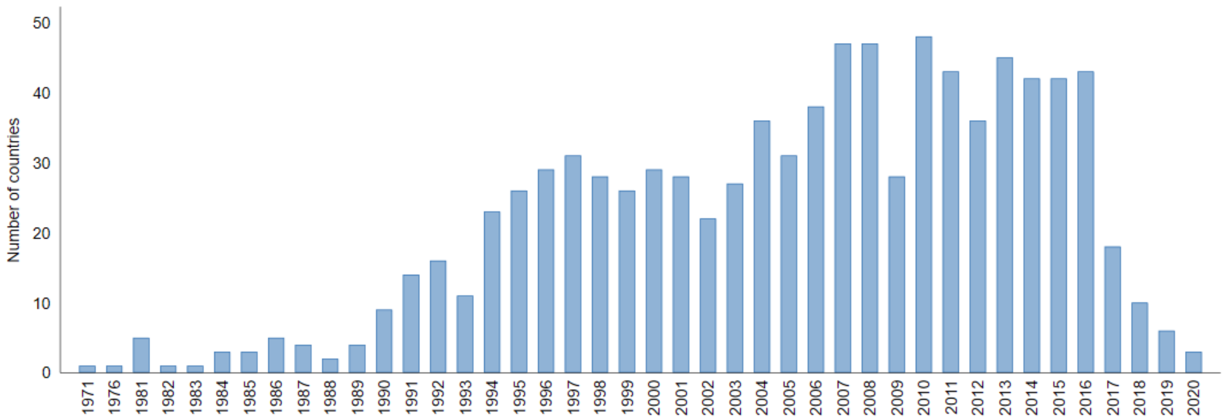


Figure A1: Number of countries across years

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. If no specification reaches 75% of individuals matched, this country is not included in the analyses.

B Descriptive statistics

Table B1: Time trends in gender wage gaps and mean maternal age at first birth

	All age groups		Youth		Mean age
	Raw GWG	Adjusted GWG	Raw 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: Models estimated with country fixed effects and source fixed effects, standard errors clustered at the level of country and data source. 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.

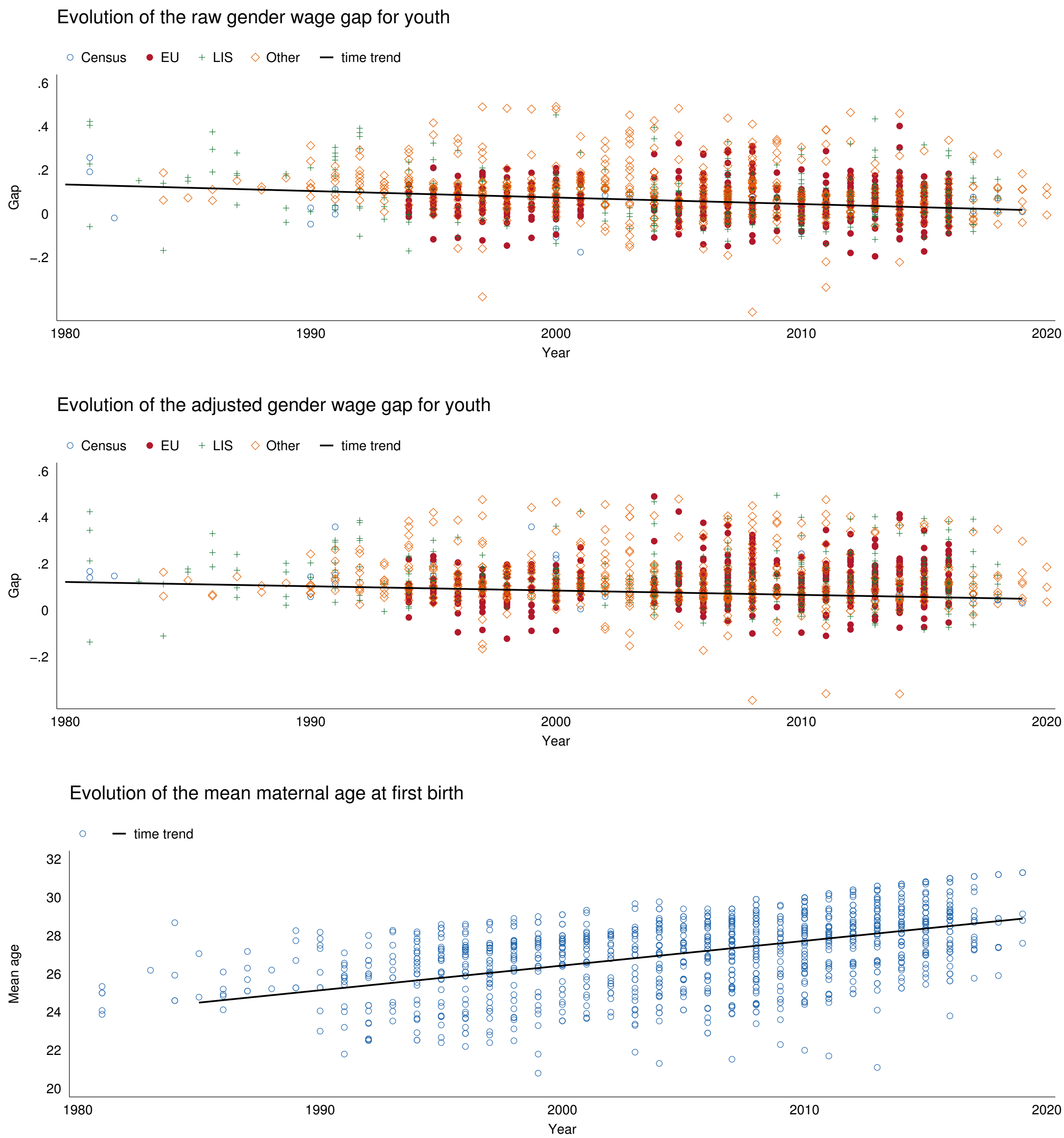


Figure B1: Gender wage gap and fertility: descriptive data

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.

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)	(6)	(7)
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 $\in$ [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 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 1. In column (1) we use fixed effects IV estimator. In column (2) we use HDFE IV estimator. 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 (6) and (7) 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 controls must include industry and occupation. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from first and second stage regressions is available upon request. Analogous set of estimates for the full sample is reported in Appendix 2.

Table C2: The effect of delayed fertility 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. Column (1) above replicates Column (1) in Table 1 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. Standard errors 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 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 source 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 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.[20]

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 there are 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 raw wages spanning several decades across genders in this country. They 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

[20] 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 $(c_w - c_m)$.

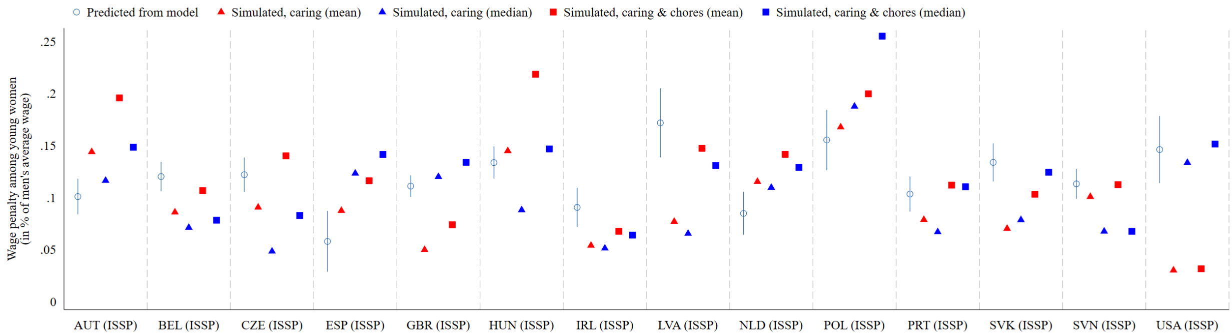


Figure D1: Benchmarking statistical gender discrimination

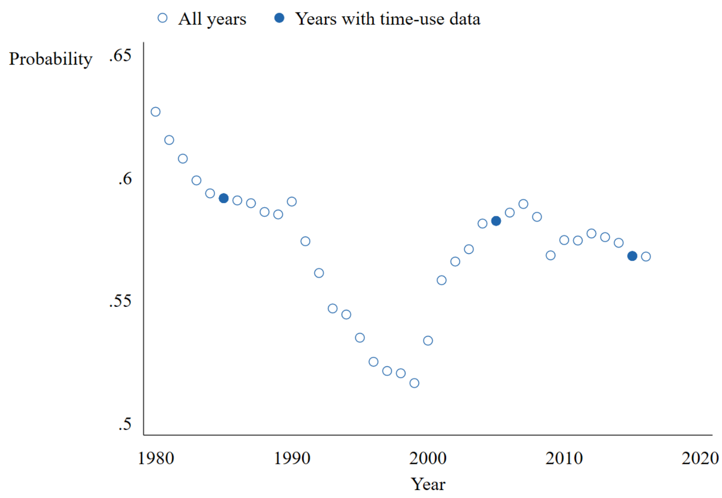
Notes: data comes from International Social Survey Program (data from 2012) Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (2), for $year = 2012$. None of the wage gaps predictions are based on ISSP data (for no country in this sample ISSP has proven to 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 age at first birth, a time trend and 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.

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).

E 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 E1 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 the maternal age at first birth implies that the 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.

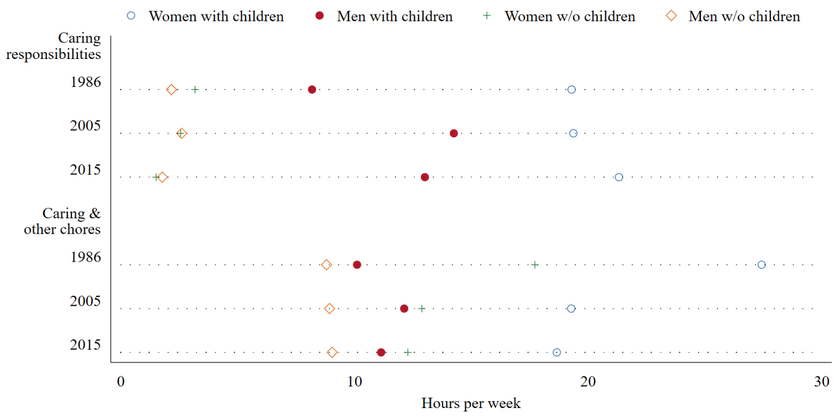
Figure E1: Evolution of fertility patterns in the United States



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 E2 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 E2: Evolution of caring time in the United States



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