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

Joanna Tyrowicz

FAME|GRAPE

University of Warsaw, and IZA

Lucas van der Velde

FAME|GRAPE

Warsaw School of Economics

Abstract

Statistical discrimination offers a compelling narrative on gender wage gaps during the early stages of the career. Expecting absences related to child-bearing and child-rearing, the employers discount productivity to adjust for the probable losses such as costs associated with finding substitutes, leaving customers, etc. If that is the case, lower and delayed fertility should imply lower discount in wages, and consequently reductions in the gender pay gap among entrants. We put this conjecture to test against the data. We provide a novel set of estimates of adjusted gender wage gaps among youth for 56 countries spanning four decades. We estimate that postponing childbirth by a year reduce the adjusted gap 2 percentage points (15%). We show that this estimate is consistent with statistical discrimination, but for some countries the estimates of AGWG imply that either statistical discrimination is not accurate or taste-based mechanisms are also at play.

JEL codes: J71, J31, J13
Key words: youth, gender wage gap, statistical discrimination

1 Introduction

It is often argued that the risk of workers being unavailable due to child-rearing is *the rational reason* for *statistical discrimination*, the notion put forward by Kenneth Arrow and Edmund Phelps (Schwab 1986, Norman 2003). However, rationality is premised by accuracy in the perception of the risk. If the incidence of child-rearing standing behind statistical discrimination is amplified, 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, studying the link between changes in fertility patterns and adjusted gender wage gaps among labor market entrants. Our study covers over fifty countries and spans nearly four decades.

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 studies (Pager and Karafin 2009, Bartoš et al. 2016, Bohren, Haggag, Imas and Pope 2019, Bohren, Imas and Rosenberg 2019). The obvious advantage of studying the link between fertility and adjusted gender wage gap is that actual average fertility decisions are perfectly observable in multiple contexts, which implies that employers receive correctly the signal about the risk of female workers' absence systematically across social domains. Obtaining adjusted gender wage gaps (AGWG) for individual characteristics using a non-parametric method (Ñopo 2008) allows reliably attributing wage differences between young men and women to gender, because within the first few years of professional activity productivity determinants are accurately proxied by the available characteristics: young workers all have low experience, their human capital is limited to educational attainment and their job seeking social network is not yet established.

In addition, there are two strands of literature relevant for our study: evidence on maternity gaps and comparative view of the long-term trends. A wide constellation of recent papers find sizable wage penalties for mothers relative to fathers, e.g. 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). This literature shows that indeed, once realized and thus observable, fertility matters for the employers in wage-setting as well as in hiring decisions. As to the long-term trends, it appears that while the raw wage gap is steadily declining (Blau and Kahn 2017), this tendency is not reflected in adjusted measures (Weichselbaumer and Winter-Ebmer 2005). An important driver of this effect may be related to composition effects, due to women systematically reducing labor supply post-fertility (Aaronson et al. 2021), much more than they themselves expect ex ante (Kuziemko et al. 2018).

This literature faces several important challenges. First, typically data limitations imply that studies cover a single country, which impedes deriving general conclusions. Second, methodological differentiation makes cross-country comparisons impossible. With these two constraints, it is difficult to study if *delay* in fertility has translated to lower adjusted gender wage gaps among labor market entrants.[1] We overcome some of the limitations of the earlier studies by developing a

---

[1]Note that the completed fertility rate declined over this period as well, but this process cannot be adequately

comprehensive collection of adjusted gender wage gaps among youth across countries and years. Our estimates are comparable between countries and over time, because we obtain them from individual-level, harmonized data and applying consistently the same estimation method. Among demographic indicators, mean maternal age at first birth is probably the most corporeal.

Overall, on aggregate, fertility rates and gender wage gaps remain unrelated in our sample. By contrast, AGWG *among young* workers declines with delay of fertility: one year of delay in mean maternal age at first birth reduces AGWG by roughly 2 percentage points, or as much as 15% of the total gap. The estimates prove to be robust across estimation methods. We also show that this estimated effect may be useful to study whether the implied extent of (adjusted) gender wage inequality is consistent with accurate statistical discrimination. Our results contribute to the debate on sharing the caring responsibilities: reducing the disproportion in time devoted to caring by women may have the same effect on their wages as delaying fertility, because at least in some countries the employers engage in accurate statistical discrimination when setting wages of young workers. Given that wage gaps at the young age perpetuate to older ages (van Staveren et al. 2018), our results have implications beyond youth.

Our paper is structured as follows. First, we discuss the relevant literature from social sciences in section 2. While our study is comprehensive relative to earlier works, we were strongly guided by the findings reported in these studies. 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. Section 4 delves into the results and discusses 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 favored and disfavored group.[2] In a nutshell, consider the following line of argumentation: as some people become parents, and parenting involves a drop in productivity, if the drop is stereotypically thought to be larger for mothers than for fathers, the employers will on average offer lower wages to women than to men when hiring young workers. This form of statistical discrimination involves averaging over all women on the one hand and all men on the other hand. The employer does not know which young worker becomes a parent during their tenure at a given business, but they expect that if that parenting occurs, the productivity drop for women will be larger than for men. If a given candidate requests a wage that ignores this statistical discounting, no employment match is formed. Female candidates, foreseeing this negative outcome, may find it optimal to lower the request in order to obtain any employment.[3]

This notion can be formalized in the following way. 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

---

studied in young populations.

[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).

care of the child.[4] 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. The assumption of $c_m < c_w$ states that the 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.

Individuals draw a procreation type as well: 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 (i.e. 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 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 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

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

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 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:

$$\begin{aligned} E_m(w|h) - E_w(w|h) &= \big(h - E_m(\pi_i)E_m(c_i)\big) - \big(h - E_w(\pi_i)E_w(c_i)\big) \\ &= E_w(\pi_i)E_w(c_i) - E_m(\pi_i)E_m(c_i) = E(\pi_i) \times (c_w - c_m) \end{aligned} \tag{2}$$

Thus, if statistical discrimination stands behind (adjusted) gender wage gaps, a decline in $E(\pi_i)$ *ceteris paribus* should imply narrowing of the AGWG. By contrast, 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. 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.

Clearly, individual-level productivity (h) is unobservable ex ante to the employer, but the information about individual-level effort and talent gradually becomes less asymmetric, as the

---

[4]Implicitly we assume that the parental type maps one-to-one to the productivity costs c_i.

employer-employee relationship evolves. Altonji and Pierret (2001) study if employers adjust wages in reaction to observed worker performance as tenure increases for racial gaps in the US. In the case of racial gaps in the US, the underlying mechanism for statistical discrimination was stereotyping over human capital: on average the employers assumed lower skill for Blacks than for Whites and with due time this gap has narrowed. This approach, however, has less applicability to the context of gender, because rather than human capital differences, employers push lower wages for young women because of child-bearing and potentially higher absenteeism related to child-rearing. First, not having children until given time censure during the contract window, may actually *raise* the probability of procreation until the end of that window in a sense that not bearing the costs of parenthood until the middle of the contract window is associated, ceteris paribus, with increased probability of incurring this cost in the second half of the contract window. Hence, the learning mechanism works actually in the opposite direction in the case of the gender gaps, as compared to racial gaps. Second, even once parenthood has already occurred, the employers still experience the uncertainty associated with the productivity decline c_i, because the parenting style can change. Hence, the event of childbearing does not resolve the uncertainty on the side of the employer, the only period of certainty is actually *before* child-bearing. For both these 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.[5]

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).

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. Correspondences 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 (e.g. 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 appears to be that in some specific contexts, institutional and cultural

---

[5]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 (e.g. 2011), Ewens et al. (e.g. 2014), Flage (e.g. 2018), Morse and Pence (e.g. 2020). In addition, statistical discrimination appears to intensify with increasing prevalence of algorithms Kleinberg et al. (e.g. 2018), Raghavan et al. (e.g. 2020).

factors may reduce the bias against mothers (Baert 2018). In the current circumstances, it is more a matter of exception than a rule.

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).[6] These studies do not have the identification to talk about the presence of bias against women *before* child-bearing, but they confirm strong bias against mothers. While this issue has been much less extensively studied, some results suggest that the wage gap at a young age actually exacerbates over career (van Staveren et al. 2018).

Studies exploiting the data for 2000s by design ignore 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. What has changed substantially is the expectations of young women concerning their future labor supply: 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 true across many countries.

Much less has been said about the link of fertility patterns 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. Actually, 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, especially given the statistical regularity that in a high fraction of heterosexual unions the man has a higher educational attainment than the woman despite opposite proportions in population. 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 orphaned. In addition, typically data limitations imply that the studies cover

---

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

a single country, which impedes deriving general conclusions, whereas methodological differentiation makes cross-country comparisons a challenge.

Our study aims to fill these 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 datasets and obtained comparable estimates of AGWG amonth 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 if AGWG among labor market entrants declines, as employers receive strong signals about reduced and delayed fertility of young women. Our results are discussed in detail in section 4. Third, we enrich the standard panel analysis with an attempt at causal identification. Finally, our approach helps also to address an important puzzle stemming from experimental studies, namely in setups such as Bohren, Imas and Rosenberg (2019), Bohren, Haggag, Imas and Pope (2019), the employment relationship is by design short and parenthood should be irrelevant for all intents and purposes in the contractual exchange. Effectively, in these experiments the fear of child-related absenteeism at work cannot explain the gap in compensations to workers. We take this puzzle to observational data.

Summarizing, unequivocally, 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 wage differences – that wages should become more equal.

3 Data and methods

For this study, we need measures of adjusted gender wage gaps among labor market entrants, across countries and for subsequent birth cohorts as the explained (dependent) variable. No such data set exist, hence 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, subsequently 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 of obtaining 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. It could be the case gender equality in wages resulted in fertility at 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 samples with fewer observations. These cross-country sources were subsequently complemented by individual-level data obtained from central statistical offices or analogous institutions around the world. Second, we obtain panel data for Canada, Germany, Korea, Russia, Sweden, Ukraine, the UK and the US. Third, 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. Data were acquired from multiple sources; Appendix A discusses in detail each of them.

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. Wage is measured as hourly wage. We use the usual hours worked and total pay without bonuses. Age is measured in years, 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.[7] 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 match one-digit International Standardized Classification of Occupations. For consistency with estimates distributed by LIS, 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 by Ñopo (2008). This method is based on exact matching and thus is non-parametric. Consequently, the estimates account for wages of *comparable* men and women and are less sensitive to inaccurate model specification than regression-based decompositions. This feature is particularly important given the large collection of highly

---

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

heterogeneous countries on which we apply the decomposition. Moreover, prior research also found that estimates of the adjusted gap obtained using Ñopo (2008) decomposition prove robust to the inclusion of additional control variables (Goraus et al. 2017), which is very fortunate given that for some countries and sources, controls such as occupation or industry are unavailable.

One final advantage of Ñopo (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. In estimations, we may focus on those countries, for whom sufficiently high fraction of men and women are matched, i.e. 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. However, intuitively the more comprehensive the controls, the lower the probability that a person with such characteristics exists in the subsamples of men and women. 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 and 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.

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. 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 decisions, and employers would believe that it is less likely that women below the former mean age at first birth would bear children. This measure is preferred over mean age at birth, since this last measure accounts for second and subsequent births as well, thus confounding the postponement with the spacing of children at older ages.

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.[8] In addition, Human Fertility Database and Human Fertility Collection report maternal age at first birth for some developing countries around the world.[9] 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,

---

[8]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/

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

Population Bulletin of the United Nations reports additionally data for selected years in the case of Brazil.

Obviously, there are alternative measures to mean maternal age at first birth by year, such as fertility rate, fraction of childless women, age specific fertility rates and the like. Yet, we argue that these alternatives are not better at capturing the process that we study. Total fertility rate measures the number of birth in a given year over the number of women aged 18 to 40 in that year. This measure captures well the decline in fertility, but understates the role of delay in fertility. Fraction of childless women, can only be computed for those birth cohorts who completed fertility already (reached the age of 40 years), hence it does not refer to youth in a given year. Age-specific fertility rates are available even more scarcely across countries and years than mean maternal age at first birth. Moreover, age-specific fertility rates includes higher parity births, for which the relation to the gender wage gap is less intuitive.

3.4 Instrumenting for fertility measures

We instrument for maternal age at first birth using four variables. First, we rely on drivers of family formation: the length of compulsory schooling, and the military conscription. In addition, we use the authorization of contraceptive pills. Fourth, we use fertility rate in the generation of mothers, i.e. lagged by 25 years. We describe these instruments and data sources below.

The reforms in compulsory education have been previously demonstrated to causally affect fertility (Black et al. 2008, Cygan-Rehm and Maeder 2013). We use data on the number of years in compulsory education provided universally by UNESCO as of 1998. Data for earlier years base on the subsequent reforms have been documented by Brunello et al. (2009), Murtin and Viarengo (2011), Fenoll and Kuehn (2017) and we utilize these sources to obtain data for 1980s and 1990s.[10] Naturally, one could be tempted to use birth-cohort specific data on compulsory education. However, most of the reforms have affected cohorts born in 1950s and 1960s, hence not the cohorts that were used to compute mean maternal age at first birth, nor gender wage gaps among the young labor market participants.

The military conscription can drive mean maternal age at first birth through several channels. First, longer the period of conscription implies family formation at higher age by men. Second, with compulsory military service, married men have more obstacles to conceive children, thus delaying mean age at first birth among their partners. By contrast, military service provides stable, guaranteed income, which may reduce earnings uncertainty and thus encourage child-bearing. Finally, in some cases military service can facilitate obtaining 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 provided by Mulligan and Shleifer (2005) and extend it for time

---

[10]The data do not include pre-primary education. To fill the gaps for several countries missing from those sources we utilize additional sources such as country legislation reported in reports of Right to Education Initiative. In Canada the length of education is set at the level of provinces, we use estimates by Oreopoulos (2005). Across countries there are substantial differences as to 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, i.e. 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.

and countries using the same source, i.e. the *Military Balance* which is published annually. We supplement this data with the records of War Resisters' International and the World Factbook. [11]

As the third 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 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, separate 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 distributed solely as medicament treating hormonal disorders, whereas in other countries it was available only to married women.

To the best of our knowledge, there is no literature studying the role of pill in forming employers' beliefs, but the literature on pill and women's labor supply decisions is rich. 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).

Our approach differs from that pursued by the previous literature. We do not study the behaviors of cohorts directly exposed to the introduction of the pill, rather we focus on cohorts joining the labor market in the mid to late 1990's. For these cohorts, the link between the year when the pill was introduced and fertility is likely mediated by a number of channels. First, there appears to be improvement in the quality of the parents which may strengthen shift in 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).

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 age, we take the data for 1956 (= 2000 - 24 - 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 C (Pearson and Lee 1899) and the current demographic changes (Steenhof and Liefbroer 2008, Kolk 2015), whereas fertility decisions by mothers are clearly exogenous to the contemporaneous hiring of young workers.

Among the four instruments in our study, military conscription, compulsory schooling and

---

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

lagged fertility have country-by-year variation, whereas the pill authorization is essentially one year for each country. For some specifications we construct for each country-year sample in our data an indicator measuring how many years have passed since the introduction of the pill. However, in our preferred specifications, we use Baltagi (1981) estimator. This estimator is efficient in a setup including time-invariant and time-varying instruments. This flexibility comes at the cost of lacking a pure $F - statistic$ for the first-stage regression. We utilize the $F - statistic$ on both time-variant and time-invariant component of the instrument.

3.5   Descriptive statistics

In total, we obtained 1,133 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 1,041 countries, years and data sources. Given that for each year and country several sources may be available, we construct weights that account for multiple observations. Specifically, we construct our weights as $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of estimates for given country in a given year.

Figure 1: Raw and adjusted gender wage gap among youth and mean maternal age at first birth
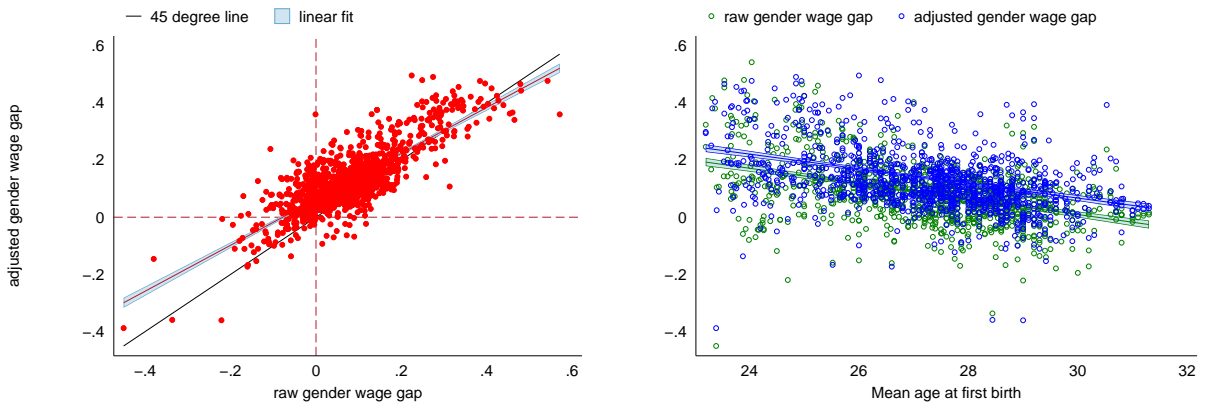
Notes: For each country, year and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender. Models estimated with country fixed effects and source fixed effects, standard errors clustered at the level of country and data source.

The left panel of Figure 1 presents the raw and adjusted gender wage gaps among youth in our sample. Intuitively, adjusted gender wage gaps are located above the 45 degree line, which signifies that for the majority of samples in our data the adjusted gender wage gap is higher than the raw gender wage gap. This 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, statistically significant), there is a level effect of 2 percentage points. We thus rely on adjusted gender wage gap, as using the raw value could lead to underestimating the effects.

Table 1: Time trends in gender wage gaps and mean maternal age at first birth

| | All age groups | | Youth | | Mean age at first birth |
	Raw GWG (1)	Adjusted GWG (2)	Raw GWG (3)	Adjusted GWG (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.

Table 1 portrays 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 overall populations across working ages, 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.

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. Note that this graph confuses cross-country dispersion (countries with higher inequality are characterized by lower mean age at first birth) and within-country trends over time. We interpret this figure as tentative evidence that there is a relationship to explore. Section 4 provides a rigorous analysis of this relationship.

4 Results

We estimate the model in the form

$$AGWG_{i,s,t} \quad = \quad \alpha + \beta \times time + \gamma \widehat{MAB}_{i,t} + \xi s + \epsilon_{i,s,t} \tag{3}$$

$$MAB_{i,t} \quad = \quad \phi + \theta PILL_{i,t} + \varrho EDU_{i,t} + \mu CONSCR_{i,t} + \varsigma M\_FERTILITY_{i,t} + \varepsilon_{i,t} \tag{4}$$

where i denotes country, t denotes time, and s denotes data source and specification.[12] $PILL$ denotes the instrument obtained from contraceptive pill authorization, EDU corresponds to compulsory schooling duration, $CONSCR$ is based on military conscription data. Finally, $MOTH\_FERT$ utilizes variation in total fertility rates of the mothers generation. To obtain panel estimates with the instrumental variable we utilize Baltagi (1981) with standard errors clustered at the level of country, data source and controls. The estimates of θ, ϱ, μ and ς are vectors, which account for cross-sectional component of each variable, time-varying component of each variable; for levels, second, third and fourth powers of the variable. Our preferred specification utilizes all available instruments, but for robustness we provide the estimates which include only EDU and $CONSCR$, only $PILL$ and only $M\_FERTILITY$ as instruments. Time trend is estimated as common across countries, otherwise we would not be able to identify θ parameter separately from β parameter.

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 4.2 we provide a 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. In columns (1)-(4) we report the IV estimations, whereas columns (8) and (9) report analogous results from a fixed effect panel estimation. Meanwhile, there are no effects of overall fertility: columns (5)-(7) report fixed effects panel estimations when we use TFR as a measure of fertility rather than mean maternal age at first birth. The relationship is much weaker for TFR, in some regressions even positive and imprecisely estimated, amounting to no statistical significance, regardless of whether we look at adjusted gender wage gaps among youth in column (5), adjusted gap for all birth cohorts in column (6) or raw gap for all cohorts in column (7).

We provide estimates for a broad array of specifications in Table 2. Column (1) reports the coefficient for mean maternal age at first birth, when all four instruments are utilized. Subsequent three columns report estimates for specifications with one instrument at a time: compulsory schooling and military service in column (2), pill authorization in column (3), mothers' generation fertility in column (4).

---

[12]Note that we utilize the "best" specification for a given source, so there is only one estimate for each source, but the "best" specifications may differ across countries and sources, depending on data availability in the original data set and on the statistical properties of the obtained AGWG estimates.

The magnitude of the estimated IV specifications is not different in statistical terms from the panel OLS specification. There are two ways to interpret this finding. The first interpretation states that reverse causality bias is small in statistical terms: the decisions about the timing of the first child is not strongly driven by the prevailing adjusted gender wage gaps. This interpretation essentially implies that OLS estimates are reliable for causal inference. The second interpretation states the opposite: that our instruments are too weak to provide reliable measure of exogenous variation in mean maternal age at first birth – i.e. independent of the prevailing AGWG. Under this interpretation neither IV specifications nor OLS specifications can be used for causal inference. The $F-statistics$ reported in Table 2 are large, well above the conventionally assumed thresholds (Lee et al. 2020). Admittedly, with the Baltagi (1981) estimator, the $F-statistics$ explores both the cross-section and the time-trend in the variation of instrumental variables, which may inflate the obtained test statistic.

Table 2: The effect of delayed fertility on AGWG

Gender wage gap	Youth, MAB, AGWG				OLS	Youth All TFR, AGWG, OLS		Youth All TFR, RGWG, OLS	
	IV								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Fertility	-0.025***	-0.041***	-0.033***	-0.024***	-0.023***	-0.050	0.025	0.077	-0.016
	(0.006)	(0.011)	(0.009)	(0.007)	(0.006)	(0.038)	(0.023)	(0.051)	(0.037)
R-squared	0.276	0.282	0.278	0.275	0.620	0.559	0.830	0.880	0.612
F-statistic	25,843	9,368	265.0	306.9					
Observations	1,106	1,121	1,161	1,142	1,170	1,243	1,303	1,303	1,243
Clustering	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time trends	Yes	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 compulsory schooling and military conscription as instruments. Column (3) with the pill authorization as instrument. Column (4) with lagged fertility as instrument. In all IV specifications, we include linear term, quadratic term, up to the fourth power of the instrumental variable. All IV specifications include time trends adjust for type of data source. No weights employed in IV specifications. All columns report estimates for labor market entrants (under 30 years of age), except columns (7) and (8), which report estimates for all birth cohorts. Column (5) adjust for country and specification fixed effects and time fixed effects. In addition to time, specification and data source fixed effects, columns (6)-(9) adjust also for country fixed effects. Standard errors clustered at country-source-controls level, thus leaving only time variation. 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. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from first and second stage regressions is available upon request.

We estimated that a rise in mean maternal age at first birth by one year is associated with a reduction of the adjusted gender wage gap among labor market entrants of roughly 2-3 percentage points, which 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, covariates.

We provide additional insights on the magnitude of the estimated effects by testing additional model specifications. Specifically, we employ high-dimensional fixed effects instrumental variable estimator instead of the Baltagi (1981) estimator in column (1) of Table 3. 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 (2)-(4) for the 25th percentile, median and 75th percentile. This estimation uses unconditional quantiles of AGWG via recentered influence function (Firpo et al. 2009) and model specification analogous to column (1) from Table 2.

Finally, we also 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 2.

Table 3: The effect of delayed fertility on AGWG - robustness checks

| | HDFE | Quantile Regression | | | Heterogeneous fertility | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
		Q25	Q50	Q75	Intercepts	Slopes
MAB	-0.012 ***	-0.023 ***	-0.022 ***	-0.032 ***		
	[-0.02,-0.00]	[-0.03,-0.01]	[-0.03,-0.01]	[-0.04,-0.02]		
MAB < Q25					0.133 ***	-0.018
					[0.07,0.20]	[-0.05,0.02]
MAB ∈ [Q25, Q75]					0.027	-0.019
					[-0.03,0.08]	[-0.05,0.01]
MAB > Q75						-0.019
						[-0.05,0.01]

Notes: Specifications analogous to column (1) from Table 2. In column (1) we use HDFE IV estimator. In columns (2)-(4) we utilize Firpo et al. (2009) recentered influence function transformation for AGWG of the model at 25th, 50th and 75th percentile, respectively. In columns (5) and (6) 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. Confidence intervals (95%) are included in square brackets. Asterisks ***, **, and * denote significance at 10%, 5% and 1%, respectively. Full set of estimates from first and second stage regressions is available upon request.

The robustness checks in Table 3 reveal that the estimated effect of roughly -0.02 is relatively homogeneous along the distribution of both AGWG and mean maternal age at first birth. The level effect appears to be the 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 MAB is very close to the upper boundary on the age in the estimation of adjusted gender wage gaps.[13] 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 decline in AGWG among labor market entrants can be attributed to delayed fertility. It the next section we provide data-driven benchmarks for these estimates.

4.2 Benchmarking statistical gender discrimination

For establishing the benchmark for our estimates, recall equation (3). This equation portrays the link between *objectivized* differences in productivity from the employers perspective and the (adjusted) gender wage gap. Our objective is to obtain analogs of $c_w - c_m$ and $E(\pi)$ from observational data to compare our estimates of $E(w_m|h) - E(w_w|h)$. Rationality imposes that the employer forms expectations based on observed probability π of incurring the cost c, with $p(a)$

---

[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.

denoting age (a) specific fertility rates for the first parity and

$$E(\pi|childless) = \frac{\int_{a=20}^{a=30} p(a)da}{1 - \int_{a=18}^{a=20} p(a)da} \tag{5}$$

This is an upper bound expectation in a sense that the contract duration with a given employee can be shorter than the age brackets $a \in [20, 30]$, hence the actual probability faced by the employer is no higher than given by equation (5). We construct the share of women who have had their first child by cohort using the data on the number of first-borns to women of a given age in a given year.

Such an employer 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 functions and social norms may be driving the gender distribution of those functions.

To capture c, we resort to the time-use data. 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 (e.g. 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.[14]

Utilizing this individual level data, we obtain information on caring by men and women, with and without child-rearing responsibilities, aged between 20 and 30 years old. We obtain four mean (median) measures: for men without kids, men with kids, women without kids and women with kids. Based on this 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:

$$c_w - c_m = \frac{\left((T - t_{w,k}) - (T - t_{w,\sim k})\right) - \left((T - t_{m,k}) - (T - t_{m,\sim k})\right)}{T}, \tag{6}$$

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 productivity endowment to the extent to which workers are not free to increase productivity per time unit. A potential concern could be related to the fact that in addition to caring, also household chores can consume disproportionate amount of time from 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

---

[14]Alternative data source is time-use surveys. In those surveys, household members report the time spent on caring. We distinguish between parents in households with children and independent adults in households without children. 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 (e.g. the previous week) and are expected to report the start and end hours by themselves. This differentiation makes data from the time-use surveys hardly comparable across countries, sometimes even within-countries and across periods. 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.

augmented with household chores.

We combine data availability of data-driven proxy of $c_w - c_m$ for 2012 from the ISSP with the value of pi computed using Eurostat Data. Clearly, the obtained simulation of $c_w - c_m$ and π is available only for a small selection of the countries included in our sample and likely this selection is not generally representative. Indeed, these are rather wealthy European countries with a long tradition of gender equality at least at the narrative level. Thus, we do not treat this exercise as a general approach, rather, we test if the ballpark of estimates obtained as our main results in Table 2 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 it appears that either statistical discrimination is inaccurate or additional stereotyping is involved.

The results are portrayed in Figure C2. Note that equation (3) has no adjusting variables except 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 C2 is based on the mean maternal age at first birth as implied by the equation (4) and time. Thus, the dispersion in Figure C2 follows solely from the variation in fertility patterns across these countries.

Figure 2: Benchmarking statistical gender discrimination
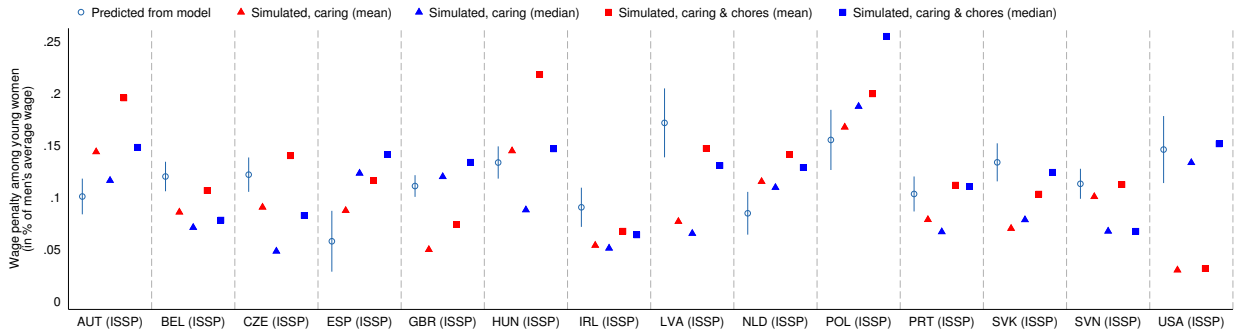
Notes: data comes from International Social Survey Program (data from 2012) Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (4), for $year = 2012$. None of the wage gaps predictions are based on ISSP data (i.e. for no country in this sample ISSP has proven 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. Analogous simulations for the available time-use data are reported in Appendix C.

Our benchmarking exercise reveals that for many countries, the ballpark implied by our model estimates is indeed close to $(c_w - c_m) \times \pi$. as evidenced by triangles, except for the time-use data for the UK. For example, in the case of Belgium, the Netherlands, Hungary or Poland, confidence intervals of the AGWG predicted by equation (3) encompass simulated $(c_w - c_m) \times \pi$. By contrast, countries such as Latvia and Slovakia, the estimates of AGWG substantially exceed $c \times \pi$ proxies. If taken at face value, these results imply that the employers have excessively high perception of $(c_w - c_m) \times \pi$ – this would be consistent with either inaccurate statistical discrimination or stereotyping (or a mixture of both). Thus, the statistical discrimination hypothesis on its own is not sufficient to explain the existing adjusted gender wage gaps.

Once we adjust the $c_w - c_m$ measure for differences household chores, portrayed by squares, for majority of countries the estimated costs are higher. Often adding differences in household chores leads to cost estimates that exceed the range implied by AGWG estimates. The implication of our findings is that the employers do not discount differences in household chores in wage

offer if disproportion in caring time is amplified by the disproportion in household chores. This is particularly true for Hungary, Poland and, to a lesser extent, Spain.

Finally, there may indeed be a disparity between the accuracy of statistical discrimination and rationality in a sense that the perception of the underlying data is accurate, but the operator which the employers apply to this data is not the expected value, but rather alternatives. Blue diamonds portray medians rather than means of $c_w - c_m$. This would be consistent with prospect theory or with employers being risk averse rather than risk neutral. Interesting cases are provided by the Netherlands, Spain and Ireland on the one hand and the UK on the other hand. For the former three countries, AGWG estimates is consistent with the mean, but not with the median $(c_w - c_m) \times \pi$. For the UK, by contrast, the AGWG is higher than $(c_w - c_m) \times \pi$ implied at the mean, but consistent with the median.

Finally, our benchmarking exercise reveals also that for Austria the estimates of AGWG are substantially lower than the values implied by $(c_w - c_m) \times \pi$. If taken at face value, these results imply that the extent of adjusted gender wage gap inequality among youth in these two countries is *smaller* than a rational employer would impose. Kleven et al. (2020) argue that strong convergence in wages across genders in this country cannot be explained by child-related absences. They utilize administrative data covering six decades and show that raising the costs to the employers had virtually no effect on systematic closing of the gender wage gap in this country. Given low overall labor force participation of women in Austria, relative to other EU countries, one potential explanation of our findings for this country may be that it is actually optimal to pay a premium to the disfavored group if labor market participation is an informative signal of productivity (Blair and Chung 2021).

These results should be interpreted with caution, because the sample of countries analyzed in the benchmarking exercise is smaller than in the estimation: only 14 countries vs 56 in the estimation. Likewise, 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. Availability of data on age specific fertility rates over time is also limited. We provide analogous estimates across the countries using time-use surveys in Appendix C. The number of countries for which benchmarking is possible is considerably smaller with the time-use data, but the results are broadly in line with those obtained using ISSP data.

Finally, for one country – the United States – data availability permits tracing the time evolutions of c, π and the estimates of AGWG. Unfortunately, the fertility evolution in the US 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 D we discuss the special case of the United States, demonstrating the link between our findings and earlier work of Kuziemko et al. (2018).

5 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 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 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.

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, but relatively fewer countries in our sample have yet to undergo the second demographic transition. 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. 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.

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, i.e. 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

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

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.

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.

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.

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.

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.
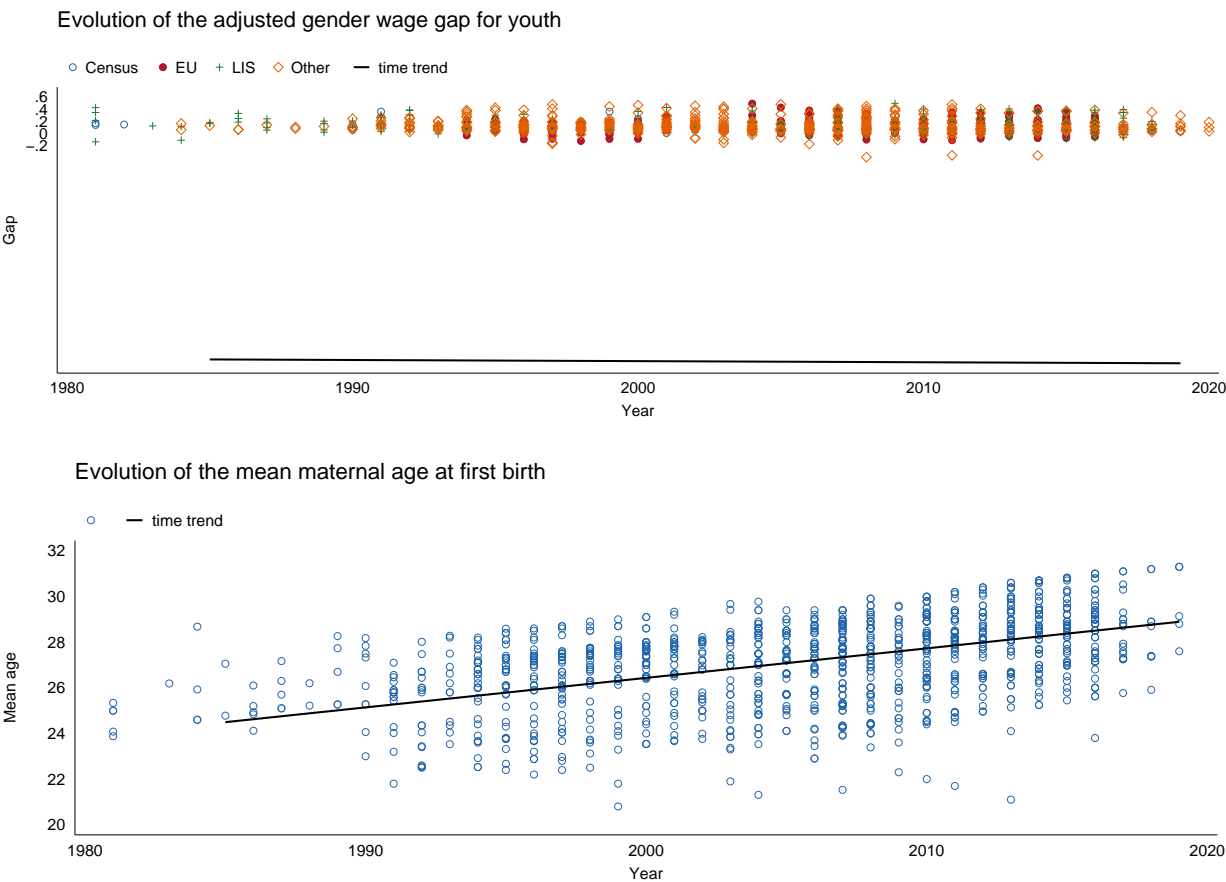
Figure A1: Number of countries across years
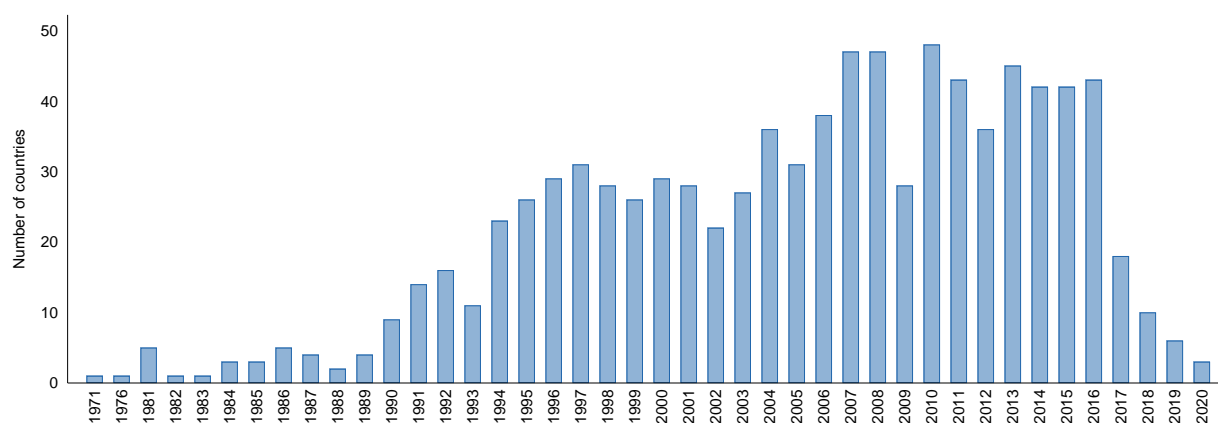
Notes: For each country, year and data source we utilize one estimate, with maximum number of available control factors subject to the constraint that 75% of individuals find a match among the opposite gender. If no specification reaches 75% of individuals matched, this country is not included in the analyses.

Table A1: Databases used in this study

Country	Census	EU	HBS/LFS	ISSP	LISSY	LSMS	longitudinal	Total
Albania						4		4
Argentina			12	1				13
Armenia			4					4
Austalia				2	9			11
Austria		18			7			25
Belarus			1					1
Belgium		21		2	19			42
Brazil	2			2	5			9
Bulgaria		8				4		12
Canada	4			1	16			21
Chile					13			13
China				3	1			4
Colombia					5			5
Cote d'Ivore					2			2
Croatia		7	12	3				22
Cyprus		11						11
Czechia		2		6	8			16
Denmark		14						14
Dominican Republic					1			1
Egypt					1			1
Estonia		12			5			17
Finland		10		5	5			20
France		20	30	1	4			55
Georgia					2			2
Germany		18		9	25		32	84
Greece		15			7			22
Guatemala					3			3
Hungary		10		5	8			23
Iceland		2		1	2			5
India					2			2
Ireland		15			20			35
Israel	1				19			20
Italy	9	18	5	1	12			45
Japan					2			2
Kyrgyzstan			3			3		6
Latvia		11	4	7				22
Lithuania		7			9			16
Luxembourg		14			8			22
Malta		7						7
Mexico	3			3	14			20
Netherlands		21			8			29
Norway		5		8				13
Panama					4			4
Paraquay					6			6
Peru					4			4
Philippines				3				3
Poland		11	28	4	4			47
Portugal		20						20
Romania		7						7
Russia				13	4		24	41
Serbia		4	9		1	3		17
Slovakia		11		3	10			24
Slovenia		10		4	6			20
South Africa					4			4
South Korea							14	14
Spain		20			7			27
Sweden		8		12	3		7	30
Switzerland		7		1	12			20
Taiwan				1	10			11
Tajikistan						1		1
Turkey				2				2
Ukraine				2			3	5
United Kingdom		20	28	2	27		18	95
United States	19			1	29			49
Uruguay			29	2	5			36
Venezuela	1			2				3
Vietnam					1			1

B   Descriptive statistics

Figure B1: Gender wage gap and fertility - descriptive data

Evolution of the adjusted gender wage gap for youth


Evolution of the mean maternal age at first birth

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 Time-use survey data for benchmarking statistical gender discrimination

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.

In the time-use surveys, household members report the time spent on caring. We isolate parents in households with children and independent adults in households without children. For each person we obtain the measure of time which they spent on caring. We then apply formula (6) to obtain the measures of c. We obtain the measure of π using Eurostat for EU Member States and Human Fertility Database for the US.

Figure C2: Benchmarking statistical gender discrimination – time use surveys
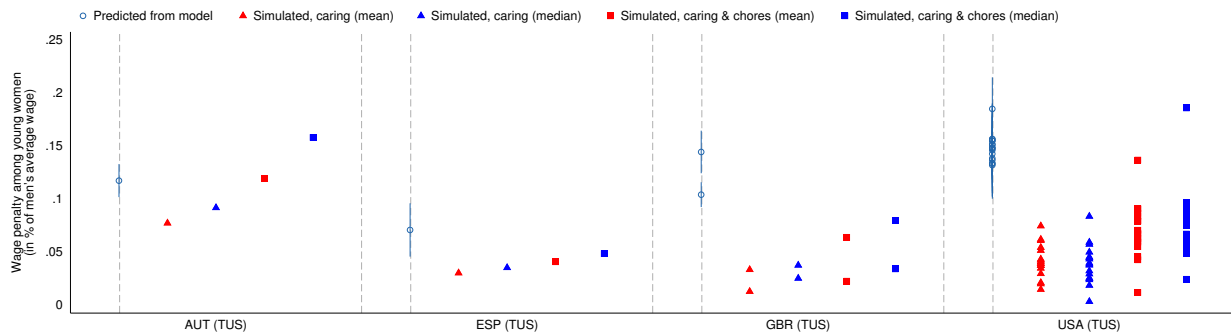
Notes: data comes from the available time-use surveys obtained from MTUS project and I-PUMS. Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (4), for the year of MTUS availability for each country. MTUS for the United Kingdom is available for 2000 and 2014. Data availability for the US is described in Appendix D. 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.

D   Time evolutions for 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, π and adjusted gender wage gaps. Figure D3 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 D3: Evolution of fertility patterns in the United States
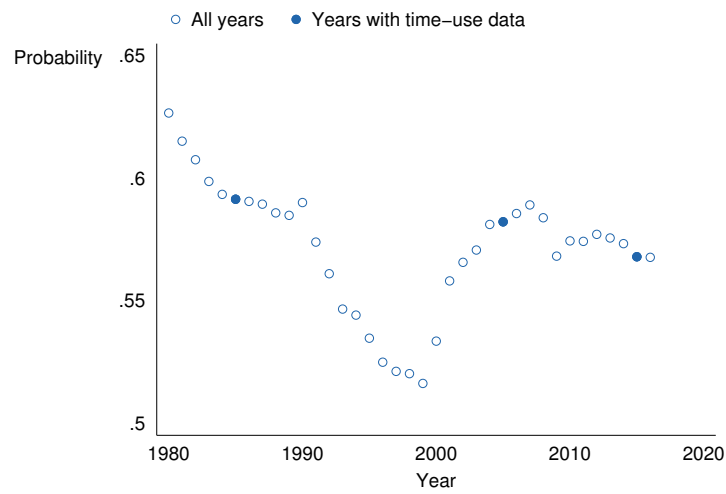
Notes: Data on age-specific fertility rates comes from Human Fertility Database. Full markers denote years for which time-use data is available. Data between 2005 and 2015 are available on an annual basis, but reveal similar picture. Hence, for clarity, we portray these two data points: 2005 and 2015.

Figure D4 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 (c 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.

Finally, Figure D5 reports the evolution of marginal predictions of AGWG and $c \times \pi$ in the US. The differences in AGWG estimates are not statistically significant, because the IV estimates have relatively low precision. However, tentatively, adjusted gender wage gaps declined between 1985 and 2005. During this period, the first parity was substantially delayed and subsequently returned to relatively more frequent events. During this period, time allocated to caring has declined for women and increased for men. Consequently, $c \times \pi$ has lower levels in 2005 and 2015 compared

Figure D4: Evolution of caring time in the United States
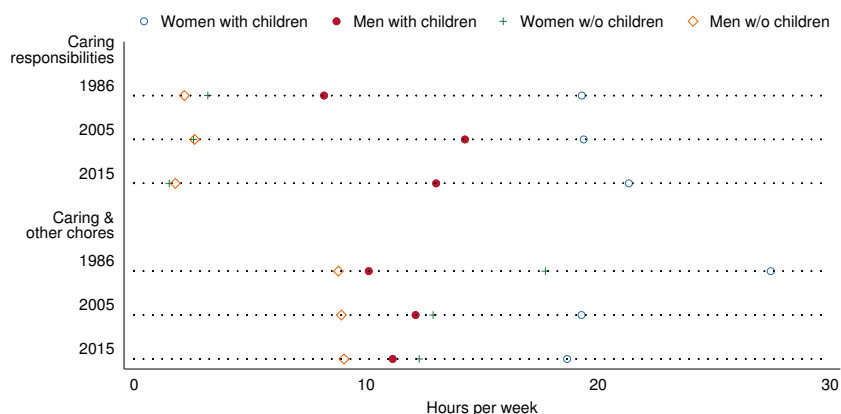
Notes: Data from American Time Use Survey (ATUS).

Figure D5: Benchmarking statistical gender discrimination – changes over time for the USA
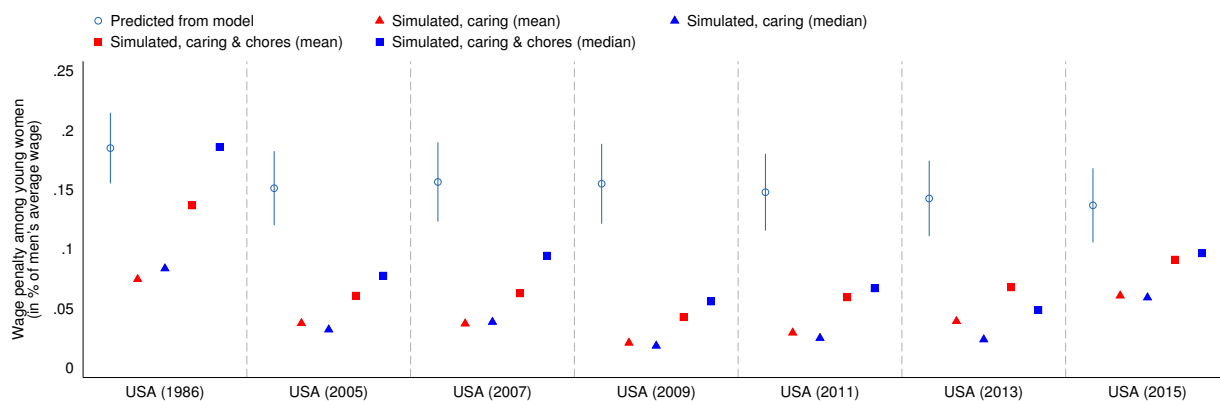
Notes: Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (4), for the years of American Time-Use data availability. Simulations at the mean and at the mean utilize ISSP data for c and Human Fertility Database age-specific fertility data for π. The data for the USA refer to MTUS from 1986 and AGWG estimates for 1985.

to 1985. These tendencies have been the subject of research (Kuziemko et al. 2018) and make the US a special case relative to other countries experiencing systematic delay in fertility. Across the available periods, the AGWG in the US are substantially higher than implied by average $c \times \pi$. Notwithstanding, in the 1980s, the gaps observed were consistent in terms of magnitude with the median $c \times \pi$.