*Journal of Demographic Economics* (2024), **90**, 659–688
doi:10.1017/dem.2024.21

RESEARCH PAPER

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

Joanna Tyrowicz[1,2,3] and Lucas van der Velde[1,4]

[1]FAME|GRAPE, Warsaw, Poland, [2]University of Regensburg, Regensburg, Germany, [3]IZA, Bonn, Germany and [4]Warsaw School of Economics, Warsaw, Poland
**Corresponding author:** Lucas van der Velde; Email: lvandervelde@grape.org.pl

(Received 18 September 2023; revised 16 September 2024; accepted 16 September 2024; first published online 20 December 2024)

**Abstract**
Statistical discrimination offers a compelling narrative on gender wage gaps among younger workers. Employers could reduce women's wages to adjust for expected costs linked to child-bearing. If this is the case, then trends toward delayed fertility should reduce the gender wage gap among young workers. We provide a novel collection of adjusted gender wage gap (AGWG) estimates among young workers from 56 countries spanning four decades and use it to test the conjecture that delayed fertility reduces gender wage inequality. We employ instrumental variables, and find that one year postponement of the first birth reduces AGWG by two percentage points (15% of the AGWG). We benchmark this estimate with the help of time-use data.

**Keywords:** gender wage gap; statistical discrimination; youth

**JEL classification:** J71; J31; J13

## 1. Introduction

We study statistical wage discrimination of young women. The risk that the workers become unavailable due to child-rearing is often cited as *the rational reason* for discrimination (a notion proposed by Kenneth Arrow and Edmund Phelps Schwab, 1986; Norman, 2003), hence the term *statistical* discrimination. Rationality is premised by accuracy in the perception of the risk. Specifically, rational employers should accommodate delayed child-bearing, as observed across many countries during the past four decades, through lowering the penalty on wages of young women relative to young men.

The existing literature shows that – once realized and thus observable – fertility matters for the employers in wage setting as well as hiring decisions (see Correll et al., 2007; Adda et al., 2017; Kleven et al., 2019b; Costa Dias et al., 2020, among others). In experimental studies, the probability of child-bearing appears to discourage employment (Baert, 2014; Becker et al., 2019; He et al., 2023) as well as assignment of high-rewards tasks (Peterson Gloor et al., 2021). Much less is known

© The Author(s), 2024. Published by Cambridge University Press in association with Université catholique de Louvain

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

about the effects of the probability of child-bearing on wages. We fill this gap. Trends toward delayed fertility change the perspective for the employers: they face a declining probability that an *average* hired young woman will become a primary care-giver during her contract. This in turn implies – if statistical discrimination is indeed the underlying cause of gender wage gaps – that wages should become more equal between young men and women.

In this paper, we test the conjecture that fertility postponement reduces gender inequality among workers aged 20–30 years. Youth is particularly interesting for two reasons. First, lack of professional experience gives rise to more ex ante gender-based stereotyping about productivity. In addition, the tenure with the current employer cannot help in resolving this uncertainty as future absences due to care-giving remain unknown until actual childbirth and can vary over time (due to changing needs of the child/children; this is not the case with, e.g., racial gaps, see Altonji Pierret, 2001, who show that employers can learn employee's productivity). Second, already observed substantial delay in fertility makes the probability of child-bearing by young workers close to negligible during a contract with an employer.[1] Thus, among this group the decline in gender wage inequality should be the steepest.

To test this conjecture, we collect individual-level data for over fifty countries, spanning nearly four decades. Using these databases, we compute comparable estimates of adjusted gender wage gaps (AGWG) among youth. These estimates account for differences in personal characteristics such as education, marital status, family composition, occupation, industry, and urban residence. We obtain them with the help of the Nopo (2008) decomposition. This method guarantees that we *compare the comparable*: It relies on the exact matching of all available characteristics to estimate the AGWG. The AGWG measures reflect the concept of equal pay for equal work, delineating gender wage equality from the improved position of women in the labor market, brought about by other secular trends, such as improved educational attainment and increased participation in the labor force, among others (see Neels et al., 2017; Blau Kahn, 2017). In general, we obtain over 1,200 comparable estimates of AGWG across countries and years. This is an unbalanced panel, which we regress against a measure of fertility timing. We use mean maternal age at first birth (MAB1). This measure portrays the risk for employers that a young woman will become a primary caregiver during the employment contract. This measure is more adequate than the completed fertility rates (CFR) or the total fertility rates (TFR).[2]

A priori, one could expect a bidirectional relationship between fertility timing and AGWG among youth (see Goldscheider et al., 2013; Baizan et al., 2016; Vignoli et al., 2020, among others). More egalitarian countries can offer better opportunities for women, which increases the opportunity costs of becoming mothers, leading to postponement of the first childbirth. Moreover, other mechanisms could affect both fertility timing and gender wage gaps among youth. To address these concerns, we propose several instruments, which are intended to reduce endogeneity bias. We also

[1]Fertility postponement does not necessitate a decline in completed fertility after the age of 30 (because women may catch up at later ages). Neither is it equivalent to an increase in childlessness (though demographic trends reveal a higher share of women above 45 years of age who have not given birth).

[2]CFR can only be computed after a given birth cohort of women exits the child-bearing age; thus it is specific to a birth cohort rather than to time or age. TFR measures are not age-specific as they refer to all women in child-bearing age in a given year.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

provide specifications with country fixed effects which address unobserved time-invariant characteristics. Thus, we overcome two important challenges faced by earlier studies. First, typically data limitations imply that studies cover a single country, which impedes deriving conclusions related to slow-moving processes such as changes in fertility patterns.[3] Second, published estimates of the gender wage gap rarely refer to youth alone, and are not directly comparable across countries and time.[4]

Our analysis suggest that delayed fertility indeed contributed to reducing gender wage gap among youth. Point estimates consistently suggest that a one year delay in fertility timing is associated with 15% reduction in gender wage gap among youth. These results confirm that the employers' perceptions of risk related with child-bearing and child-rearing follow demographic trends. However, the econometric model alone cannot inform if this estimate reflects accurately the risks related to fertility.[5] We complement our econometric analysis with a simulation exercise that recovers productivity gap associated to childbearing based on two inputs: the probability of child-bearing and the childcare time gap between men and women. We show that the marginal predictions from our econometric model exceed the simulated productivity gaps in the case of the US, the UK, and Ireland. For most continental European countries, we find that the marginal predictions are aligned with the productivity gap simulations. In these countries, the adjusted gender wage gaps among youth display magnitudes consistent with the data-driven simulations of productivity gap.

Our paper is structured as follows. First, we discuss the relevant literature from social sciences in section 2, identifying 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 3 concludes with extensive descriptive analyses documenting the correspondence between the AGWG and the fertility timing patterns. Section 4 delves into the results and discuss their robustness as well as limitations. The final section concludes with limitations and policy implications of our study.

## 2. Literature

Starting from Becker (1971), Phelps (1972), and Arrow (1973), statistical discrimination is a recognized mechanism explaining the differences in wages between a favored and a

[3]Working with individual countries has the advantage that it allows to identify the role of specific policies on both child-bearing decisions and labor market outcomes. Recent examples of single-country studies include among others Adda et al. (2017); Kleven et al. (2019b); Costa Dias et al. (2020) and Gallen (2023). Kleven et al. (2019a) takes a comparative perspective and provides tentative analysis if motherhood gaps differ systematically across the degree of traditional gender norms across countries.

[4]The estimates are not directly comparable, because typically a study relies on one specific source of data, the authors use differentiated methods and pose differentiated research questions (see Jarrell Stanley, 2004; Weichselbaumer Winter-Ebmer, 2005; van der Velde et al., 2015, for quantitative literature reviews). Our approach differs in that we explore long-term trends for a comprehensive collection of countries: we obtain estimates of adjusted gender wage gaps among youth. The full set of gender wage gaps estimates is available at [LINK REMOVED FOR ANONYMITY]. Our estimates are comparable between countries and over time, because we obtain them from individual-level, harmonized data, and we apply consistently the same estimation method.

[5]For recent literature on accurate statistical discrimination, see, e.g., Pager Karafin (2009); Bohren et al. (2019).

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

disfavored group.[6] In a nutshell, consider the argument proposed by Becker et al. (2019): when people become parents they can experience a drop in productivity, for example, due to absences or the inability to work certain hours (see also Goldin, 2014). The employer does not know whether a young worker will become a parent during their tenure at a given business, but they expect that if parenting occurs, then the productivity drop for women will be larger than for men. Rational employers will internalize this expected productivity drop by offering lower wages to women. This is a form of statistical discrimination because it involves averaging over all women on the one hand and all men on the other, regardless of their actual fertility intentions and realizations. Female candidates, foreseeing this wage discounting, may find it optimal to request lower wages in order to obtain any employment.[7]

This form of statistical discrimination is different from the racial gaps studied in the literature (Altonji Pierret, 2001), since employers cannot learn about potential productivity drops from continued interaction with the employee. Moreover, if employers expect that women will become mothers, the fact that they remain childless up to some point increases the conditional probability of having children in the future and leads to even larger wage reductions for women compared to men. Moreover, care needs might vary over time (e.g., if children catch certain diseases). So, the uncertainty surrounding productivity drops might not be resolved with childbirth.

Experiments show that in short-term interactions, statistical discrimination against women in stereotypically male tasks (programming) shrinks when information on performance is provided, but it does not vanish entirely (Bohren et al., 2019). While this result is a cause for some optimism, updating beliefs about actual workers might be insufficient to address 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 Karafin, 2009). On the other hand, attention discrimination effectively limits the updating of beliefs, thus reinforcing inaccurate premises upon which statistical discrimination is based (Bartoš et al., 2016).

Correspondence and audit studies have found that both the actual birth of a child (and parenthood status, see Petit, 2007; Baert, 2014; Bygren et al., 2017; Hipp, 2020; Kleven et al., 2019a, among others) and the potential parent status matters for the call back rates, which are usually lower for mothers and the potential mothers (Baert, 2014; Becker et al., 2019; He et al., 2023; Wang Chen, 2023). Besides call backs, vignette experiments suggest that the type of positions (tasks) offered to women of childbearing age differ from offers to men and other women (Peterson Gloor et al., 2021).

Observational studies have consistently found that realized fertility leads to wage declines for women both compared to childless women and to men (Gangl Ziefle, 2009; Adda et al., 2017; Fuller Cooke, 2018; Kleven et al., 2019b; Costa Dias et al., 2020; Cukrowska-Torzewska Lovasz, 2020). These studies find that wage gaps related to motherhood are prevalent across countries (Kleven et al., 2019a), and that wage gaps tend to exceed productivity gaps between parents and non-parents (Gallen,

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

[7]For a thorough review of relevant social theories as well as an experimental exploration see Auspurg et al. (2017), within economics Cortés Pan (2023) offer an excellent summary of research on childbearing and gender inequality.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

2023).[8] Descriptive evidence suggests that the delayed fertility is beneficial for wages, ceteris paribus. However, the extent to which women can benefit from postponing childbirth is heterogeneous across education levels (Taniguchi, 1999) and occupations (Landivar, 2020). These studies do not identify the presence of bias against women *before* child-bearing, rather they confirm strong bias against mothers, and occasionally also a preference for fathers (e.g., Yu Hara, 2021).

Finally, demographers have provided voluminous evidence of reverse link, i.e., that economic conditions and gender equality affect fertility decisions. Job instability and uncertainty have been linked to fertility postponement and foregoing (e.g., Vignoli et al., 2012; Wood Neels, 2017; Vignoli et al., 2020). Okun Raz-Yurovich (2019); Goldscheider et al. (2013) and González et al. (2019) study the role of gender equality within households and subsequent fertility with mixed evidence.

To the best of our knowledge, this is the first study to focus on changes in fertility timing and gender inequality in wages among young workers using a comparative approach. While economists are concerned about causal identification of the effects of parenthood once it occurs, demographers and sociologists devote attention to households/couples and their fertility decisions, which left this question of paramount policy relevance somewhat orphan. In addition, typically data limitations imply that studies cover a single country, which impedes deriving general conclusions, whereas methodological differentiation makes cross-country comparisons of the estimates available in the literature a challenge.

Our study aims to fill the existing gap in several ways. First, we provide a novel and comprehensive collection of AGWG among young workers. We have harmonized nearly 1,200 individual-level data sets and obtained comparable estimates of AGWG among youth. This large collection of estimates allows us to purposefully ignore time-invariant country-specificity, such as culture, legal context, or social norms. We discuss the details in the next section. Second, we explore the link which has so far slipped from the radar of social scientists of many disciplines: we study whether AGWG among young workers declines, as employers receive sufficiently informative signals about delayed fertility of young women. We utilize several instruments to address the issue of endogeneity and thus provide causal estimates. Our results are discussed in detail in sections 4 and 4.2.

## 3. Data and methods

The research question at hand requires estimates of the AGWG among young workers, across countries, and for subsequent birth cohorts. These estimates constitute the dependent variable of the main analysis. Given that no such data set exists, we collected individual-level data sets, harmonized them, and obtained over 1,200 comparable estimates of adjusted gender wage gaps among youth. In this section, we describe the availability of individual-level data across countries and years. In section 3.1, we discuss the harmonizing of the acquired data, while in section 3.2 we describe the method used to measure the AGWG. The key explanatory variable of interest is mean maternal age at first birth. We discuss in detail the sources for this variable in section 3.3. We describe instruments for fertility timing in section 3.4. Section 3.6 presents descriptive statistics for the sample used in the study. Finally, section 3.5 describes the estimation strategy.

[8]A notable exception is the early work of Amuedo-Dorantes Kimmel (2005), who showed a motherhood premium among women with tertiary studies.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

### 3.1. Data on gender wage gaps among youth

We collected a large number of individual-level databases. 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 have 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 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, albeit smaller, samples. These cross-country sources were subsequently complemented by individual-level data obtained from central statistical offices or analogous institutions around the world. We obtain panel data for Canada, Germany, Korea, Russia, Sweden, Ukraine, the UK, and the US. We obtain labor force survey data or household budget survey data from Albania, Argentina, Armenia, Belarus, Chile, Croatia, France, Italy, Poland, Serbia, the UK, and Uruguay. This selection of countries was driven by the availability of hourly wage data. Finally, the World Bank in cooperation with local statistical offices provides Living Standards Measurement Survey for several countries around the world, including Albania, Bosnia and Herzegovina, Bulgaria, Kazakhstan, Kyrgyzstan, Serbia, and Tajikistan. Appendix A discusses in detail each of the data sources.

Overall, we collected data for 56 countries spanning the last four decades. These databases were harmonized in order to obtain comparable estimates of adjusted gender wage gaps. The dependent variable in the decomposition is hourly wage, which is derived based on usual hours worked and total pay without bonuses. The sample is restricted to individuals aged 20–30 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 variables include indicators for the presence of small children,[9] partnership status,[10] and degree of urbanization.[11]

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

### 3.2. Measuring the adjusted gender wage gaps

We decompose wage differences using the approach proposed by Nopo (2008). This method is based on exact matching, and it is not affected by gaps generated pertaining non-matchable individuals. Moreover, given that the approach is non-parametric, the resulting estimates are less sensitive to inaccurate model specification than regression-based decompositions. This feature is particularly

[9]Data permitting, we used the threshold for the school age.
[10]It equals one if the individual is married or cohabiting, and zero in the remaining cases.
[11]It equals one for city dwellers, and zero for individuals living in rural areas.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

important given that we apply the decomposition to a large collection of highly heterogeneous countries. Prior research found that estimates of the adjusted gap obtained using Nopo (2008) decomposition prove robust to the inclusion of additional control variables (Goraus et al., 2017), which is useful given that the full set of controls is not always available for each country and data source.

The measures of the AGWG account for three sets of variables: personal (age and education), household (urban residence, marital status, and having children), and job characteristics (industry and occupation). This set of controls is common in the literature (Weichselbaumer Winter-Ebmer, 2005). It is also the largest set of controls feasible jointly in our collection of the individual data sets. With age and education we can adjust for human capital. With household characteristics we adjust for the type of the labor market as well as possible flexibility constraints and incentives (children, marital status). Finally, occupation and industry help us to account for job characteristics.

For each harmonized data set, we identify the availability of control variables and obtain AGWG for the most comprehensive set of controls, but also for subsets controls, excluding the third set, job characteristics. On the one hand, increasing the number of controls raises the internal validity of our estimates (young men and women, for whom we compute the gap, are indeed comparable). On the other hand, Nopo (2008) decomposition can only be estimated within common support, i.e., the procedure excludes men and women without statistical twins from the estimation. A large proportion of individuals outside the common support limits the external validity of the AGWG estimate as representative sample of the population. One major advantage of Nopo (2008) decomposition is that one can compute the share of (*un*)matched men and women. We use this information to determine which of the battery of estimates for each country and year is preferable.

In order to strike a balance between comprehensiveness of AGWG measure and the comprehensiveness of sample on which it was computed, we compare several estimates in each sample for combination of control variables. We 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.[12] As a robustness check, we replicate all the estimates with the additional restriction that the set of controls must include occupation and industry.[13]

### 3.3. Data on fertility timing

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

We combine multiple sources to collect data for fertility timing for the countries and years covered by the individual-level data. The data for most European countries are

[12]Table B1 in Appendix B shows how often we adjust for each control in the final sample.

[13]Conditional on industry and occupation – the share of matchable men and women might be higher in countries with more progressive gender norms than in countries with more traditional attitudes, because the extent of gender segregation could lead to differences in selection across countries and over time. This would undermine validity of Nopo (2008) decomposition only in the case of perfect segregation. To err on the cautious side, we provide results for the AGWG estimates with and without conditioning on industry and occupation.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

provided by the Eurostat (variable AGEMOTH1). United Nations Economic Commission for Europe (UNECE). The Organization for Economic Cooperation and Development (OECD) extends this data source to include some non-EU members and reports a full time series from 1960 onward.[14] In addition, Human Fertility Database and Human Fertility Collection report maternal age at first birth for some developing countries around the world.[15] Bongaarts Blanc (2015) provide data for a large collection of countries using Demographic and Health Surveys program. Data for China come from He et al. (2019). Australian Bureau of Statistics provides full extent of first birth data by the age of the mother spanning 1975–2019, which we use to calculate the means for each year. The central statistical office from South Africa provides extensive birth data based on 2011 census. Last, Population Bulletin of the United Nations reports data for selected years in the case of Brazil.

Mean maternal age at first birth is especially well suited for our study (MAB1, see OECD, 2024, for a definition). It is obtained from vitality statistics and thus it is less frequently available across countries and years than TFR. However, this measure isolates first-time mothers, not confounding the postponement of the first childbirth with the spacing of children at older ages. The advantage of using MAB1 data is that it focuses exactly on fertility timing. For example, two countries can report the same TFR or CFR, but have highly differentiated patterns of mothers' age at the first birth. Also, it is independent of the number of children per woman. Age-specific fertility rates are scarcely available across countries and years. Moreover, age-specific fertility rates include higher parity births, for which the relation to the gender wage gap is less intuitive than for the first parity.

### 3.4. Instrumenting for fertility measures

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

*Pill authorization.* As the main source of identification we use the authorization of contraceptive pills. Authorization is a purely administrative procedure. It does not imply automatically access to the drug. Authorization stems from a procedure verifying if a given product fulfills the public health criteria established by regulatory authorities. These criteria are determined by the authorities independently across countries.[16] Likewise, authorization as a medication does not preclude that it is dispensed at all.[17]

Thus, authorization does not necessitate adoption or utilization, but lack of it almost certainly precludes them. Specifically, in many countries the pill was authorized but it

[14]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 data sources. UNECE distributes data online as https://w3.unece.org/

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

[16]For example, in some countries the contraception pill was authorized as a medication against hormonal disorders unrelated to contraception.

[17]One recent example concerns the varying rules of mifepristone availability in the USA, as documented by Food and Drug Administration (FDA): http://FDA.gov.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

was not adopted. Finlay et al. (2012) report wide dispersion concerning the channels of distribution and administering. For example, in some countries, once authorized, contraceptive pills were initially solely distributed as treatment for hormonal disorders, whereas in other countries it was available only to married women.[18]

The literature on pill and women's labor supply decisions is rich.[19] To the best of our knowledge, there is no literature studying the role of pill in forming employers' beliefs. Most of our AGWG estimates correspond to cohorts joining the labor market in the mid to late 1990s rather than cohorts directly exposed to the introduction of the pill. This feature distinguishes our approach from the earlier literature: for the cohorts covered by AGWG estimates the link between the pill and fertility timing is mediated by a number of channels. First, there appears to be improvement in the quality of parents, which may strengthen the shift in general social norms toward gender equality in reaching professional aspirations (Ananat Hungerman, 2012). Second, inter-generational transmission of norms between mothers and daughters is a recognized phenomenon (Booth 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).

With this instrument, concerns for reverse causality are negligible. The pill was authorized (the focus of our instrument, rather than adoption) in most countries of our sample in early to late 1970s. The birth cohorts studied in our estimation the procreative age and the labor market at the earliest in the 1990s. There is no reason to expect that gender wage gaps for youth active as of 1990s influenced the timing of the technical authorization procedure for the pill several decades earlier. A violation of the exclusion restriction would require that conditional on fertility timing, employers varied their expectation concerning other characteristics of women, and that this variation is linked to when the pill became authorized.

*Compulsory education.* The reforms in compulsory education have been previously demonstrated to causally affect both fertility timing and fertility level (Black et al.,

[18]In some countries (such as Puerto Rico), firms conducted initial trials to test long-term consequences of the pill, as well as the possibility of becoming pregnant after discontinuing the use. We do not consider these initial experiments as the year in which the pill was available. Countries of the Warsaw Pact as well as the USSR have authorized pill early on, but it was unavailable to women due to international trade restrictions. Likewise, the UK has authorized the pill early on, but for a nearly decade the law permitted prescribing the pill only to married women to treat hormonal disorder. Norway was among the latest to authorize the pill despite progressive politics and almost immediate adoption. The timing of authorization does not seem to have direct effects on adoption: Poland and Czech Republic authorized pill in the same year, but adoption is currently 14% and 48%, respectively – the former is ranked among the lowest and the latter among the highest in terms of adoption rates among European countries.

[19]Women postponed marital decisions and were more likely to become professionals in non-traditional sectors (Goldin Katz, 2002), women worked more hours at both extensive and intensive margins (Bailey, 2006). Accordingly, a significant portion of the reduction of the *unadjusted* gender wage gap between 1980 and 1990s 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 child-bearing until finishing education, which increased quality of parenting without reduction in completed fertility (Ananat Hungerman, 2012). Fertility declined for women without tertiary education (Bailey, 2010).

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

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

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

The exogeneity of the instrument hinges on the lack of gender-specific consequences from these reforms. The exclusion restriction would be violated if the marginal increase in the years of compulsory education was differentiated between boys and girls. The evidence is scarce. Devereux Hart (2010) study the reform in Great Britain, showing similar take up rate. They show some minor differences in terms of returns (positive for boys, imprecise estimates for girls). Pischke von Wachter (2008) find no evidence of returns to marginal increase in compulsory education in Germany, neither for boys nor for girls.

*Military conscription.* The third instrument corresponds to the length of military conscription in months. The length of military conscription can drive mean age of women at first birth through several channels. On the one hand, longer military conscription could lead to postponing fertility. First, longer periods of conscription could lead men to postpone family formation until the conscription is over. Second, even if individuals find their partners before enlisting, they could be deployed or relocated, thus facing obvious obstacles to conceiving children. Accordingly, compulsory military service may rise mean maternal age at first birth. On the other hand, military service provides stable, guaranteed income, which may reduce earnings uncertainty and thus encourage child-bearing. Similarly, military service can provide skills relevant for future employers, thus raising earnings potential among men and encouraging child-bearing. Given that military conscription can work in either way, we do not hypothesize on the sign of the coefficient in the first stage regression. We use data on the number of months as provided by Mulligan Shleifer (2005) and extend it for time and countries using the same data source, that is the *Military Balance* which is published annually. We supplement this database with the records of War Resisters' International and the World Factbook.[20] This variable exhibits variation over time and countries. This variable ranges from zero, i.e., there is no conscription, to 48 months in Israel. Given that in some countries the duration is established as a range, models include two instruments: one for the lower bound and one for the upper bound.

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

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Empirical evidence tentatively suggests that conscription does not affect gender earnings gap, because its effects on earnings remain small. This makes compulsory military service a suitable instrument for fertility timing in AGWG regressions. Mouganie (2019) finds that lifting conscription in France did not have a significant impact on wages or employment later on, even though eligibility increases the years of education (as people studied to avoid conscription). Similarly, Grenet et al. (2011) also fail to find significant returns to conscription in another regression discontinuity design. Card Cardoso (2012) find no wage effects of military conscription in Portugal (except for men with only primary education).[21]

*Fertility among mothers' generation.* The fourth instrument utilizes data on total fertility rate in the generation of mothers of the individuals in our sample. For example, if a sample for a given country comes from 2000, and we restrict the individuals used in the estimation to between 20 and 30 years of age, we take the data for 1980 (=2000 − 20) for that country. The data total fertility rate of mothers' generation comes from The World Bank. There is broad evidence for the inter-generational transmission of fertility norms covering both the demographic transition of the 19th century (Pearson Lee, 1899) and the current demographic changes (Steenhof Liefbroer, 2008; Kolk, 2015), which makes this instrument plausibly correlated with mean maternal age at first birth measured contemporaneously. Clearly, there is no room for reverse causality with this instrument (the fertility of the mothers' generation has already been realized and is well known). Likewise, there is little reason to expect that it has effects on gender wage gaps other than via forming the employers' beliefs about the contemporaneous timing of fertility.

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

### 3.5. The estimation method

We model the link between fertility timing (*FT*) and adjusted gender wage gap among youth (*AGWG*):

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

where $i$ denotes country, $s$ denotes data source, and $t$ denotes time. A negative and statistically significant coefficient implies that AGWG follows adjustments in fertility timing. In other words, objective risks associated with the costs of child-bearing and child-rearing to the employer transmit to the wage gaps imposed by those employers. Finding evidence for statistical discrimination does not imply that it is the only reason for unequal pay for equal work.

[21]Recently Bingley et al. (2020) point to the opportunity cost of military service in Denmark, but this case concerns a randomized draft, thus the estimates concern men drafted relative to men not drafted rather than women.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

The variable FT is potentially endogenous in the presence of time-varying unobserved variables that affect both the AGWG and fertility choices. If that is the case, $\beta$ is biased as an estimate of the link between fertility timing and gender inequality. Hence, we also estimate the relationship using an IV approach. Our two-stage IV model takes the form:

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

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

where *PILL* denotes the instrument obtained from contraceptive pill authorization, *EDU* corresponds to compulsory schooling duration, *CONSCR* is based on military conscription data. Finally, *M_FERT* utilizes variation in total fertility rates from the period when the mothers generation was in reproductive age. Due to the nature of our instruments, we utilize the random-effects estimator proposed by Baltagi (1981) with fixed effects for data source and specification used to obtain $AGWG_{i,s,t}$. As is conventional in the literature (Heckman Navarro-Lozano, 2004; Mogstad Torgovitsky, 2018), the estimates of $\theta$, $\varrho$, $\mu$, and $\varsigma$ are vectors accounting for cross-sectional component of each variable, time-varying component of each variable; for levels and basis functions up to the fourth polynomial for each variable. Our preferred specification utilizes all available instruments, but for robustness we provide estimates that use subsets of instruments. The time trend is estimated as common across countries, otherwise we would not be able to identify $\beta^{IV}$ parameter. Estimating the model given by equations (2) and (3), we cluster model disturbances at the level of country, data source, and AGWG specification controls.[22]

### 3.6. Descriptive statistics

In total, we obtained 1,233 unique estimates of adjusted gender wage gaps computed for individuals aged 18–30 years old.[23] This collection of estimates covers 55 countries for 37 years. The specific number of estimates for each country and data source is reported in Table A1 in Appendix A. This number of estimates reflects the unique combinations of country, data source, and year. The final sample includes estimates for 1,106 countries, years, and data sources for estimates of AGWG among youth, as for some country and years fertility timing and/or our instruments are missing.

The average AGWG suggests that at the beginning of their careers men earn 12.6% more than women (7.9% without adjustment). The gender gaps are quite spread. In some cases, we observe negative values, which indicate that in a given country × year × data source women earn more than men. More broadly, adjusted estimates tend to be larger, as seen also in the last percentile. This is also illustrated by Fig. B1 in Appendix B, where we plot adjusted and unadjusted estimates against each other.

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

[23]This number corresponds to estimates that meet the quality criteria outlined in section 3.2: percentage matched of each gender above 75%, at least 100 observations of each gender, and gaps below 100% in absolute value. Relaxing all the restrictions would raise the number of estimates by 150 estimates.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Most estimates lie above the 45 degree line, which indicates that the adjusted gaps are larger than the unadjusted ones.

Note that the levels of AGWG in our study are different than those reported in reviews by, e.g., Blau Kahn (2017). This is because our study reports gaps for youth, as opposed to individuals in working age in most other studies. Given that we have data for individuals across all age groups, we additionally compute the gaps for all working-age individuals, and these prove to be comparable to estimates reported in earlier studies. They also exhibit similar time trends. Table B3 in Appendix B reports the time evolution of both gender wage gaps as well as mean maternal age at first birth. The portrayed time trends adjust for data source and country composition and thus are not driven by the differentiated data availability across countries. In columns (1) and (2) we report the time trend for the working-age populations, whereas in columns (3) and (4) we report analogous estimates for the youth. Earlier literature reports essentially no time trend for the AGWG at the country level (Weichselbaumer Winter-Ebmer, 2005), which we replicate in our data spanning nearly two decades more. We add evidence on time trends for AGWG 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 almost a century for the gap to disappear. By contrast, the 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 toward 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 youth AGWG data and mean maternal age at first birth in Fig. B2 in Appendix B.

As discussed earlier, our measure of fertility timing is the mean maternal age at first birth. Table 1 reports the average value in our sample at around 27 years old (see also Fig. B2 in Appendix B). Figure 1 displays two measures of association between between fertility timing and the AGWG for young workers. The figure on the left presents raw data, and as such is a measure of the unconditional correlation. The figure on the right is obtained after detrending both MAB1 and the youth AGWG.

Table 1 reports the unadjusted descriptive statistics for the same instruments for all observations. We report the relationship between fertility timing (MAB1), and each of the instrumental variables in Fig. B3, also residualizing the time trends. All proposed instruments display strong correlation to fertility timing. Most links tend to be non-linear. We report variation in the year in which the pill was authorized in each country. Authorization, as we emphasized, is no evidence for adoption in any given country. It only states that the pill was legally admissible in a given country. The time since the introduction of the pill is associated with lower MAB1 for most part of the distribution. We report also the descriptive statistics for the duration of compulsory education, the duration of military conscription, and mothers' generation fertility rate. More years of education are accompanied by an increase in the mean maternal age at first birth. By contrast, higher values of past fertility rate and months of mandatory conscription are associated with lower mean maternal age at first birth.

## 4. Results

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 sub section 4.1. These results are indicative of the strength of the statistical relationship

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

**Table 1.** Sample characteristics

| | Mean | SD | P10 | P90 | N |
|---|---|---|---|---|---|
| Adjusted gender wage gap (youth) | 0.126 | 0.107 | 0.019 | 0.264 | 1233 |
| Unadjusted gender wage gap (youth) | 0.079 | 0.108 | −0.040 | 0.203 | 1233 |
| Mean maternal age at first birth | 27.18 | 1.820 | 24.56 | 29.30 | 1144 |
| Instruments | | | | | |
| Year of pill authorization | 1963.6 | 8.159 | 1960 | 1978 | 1213 |
| Years of compulsory education | 10.12 | 1.511 | 8 | 12 | 1193 |
| Fertility among mothers' generation | 2.062 | 0.811 | 1.400 | 2.965 | 1205 |
| Duration of conscription (minimum, months) | 5.358 | 7.159 | 0 | 14 | 1191 |
| Duration of conscription (maximum, months) | 6.641 | 9.271 | 0 | 18 | 1197 |

between fertility timing and the estimates of the AGWG. The second substantive part of our analysis is presented in section 4.2, where we provide a theory-disciplined data-driven benchmark for the obtained coefficients.

### 4.1. The effects of gradually delayed fertility on AGWG

Delayed fertility implies lower AGWG among young workers. Our estimates indicate that a rise in mean maternal age at first birth by one year leads to a reduction of the AGWG among young workers of roughly 2–3 percentage points. We provide estimates for a broad array of specifications in Table 2. In columns (1)–(3) we report panel IV estimations, whereas column (4) reports analogous results from a panel OLS estimation. Column (1) reports the coefficient for mean maternal age at first birth, when all four instruments are utilized. Subsequent two columns report estimates for specifications with the pill authorization as the only instrument in column (2) and for the remaining three instruments in column (3).

The estimated effect of 1.9–3.3 percentage points implies that the observed delay in fertility amounts to a decline between 20% and 30% of the inequality in wages adjusted for differences in individual characteristics. This magnitude is robust to the inclusion of multiple control variables. The *F*-statistics reported in Table 2 are large, well above the conventionally assumed thresholds (Lee et al., 2020) which speaks to relevance of the utilized instruments. Admittedly, with the Baltagi (1981) estimator, the *F*-statistics explores both the cross-section and the time in the variation of instrumental variables.[24]

Table 2 displays three panels. Panel A reports the results for the best AGWG estimates, as defined earlier. Panel B reports estimates analogous to panel A, but on a sample restricted to AGWG estimates that do not adjust for occupation and industry. This sample selection reflects the premise that potentially the process

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

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

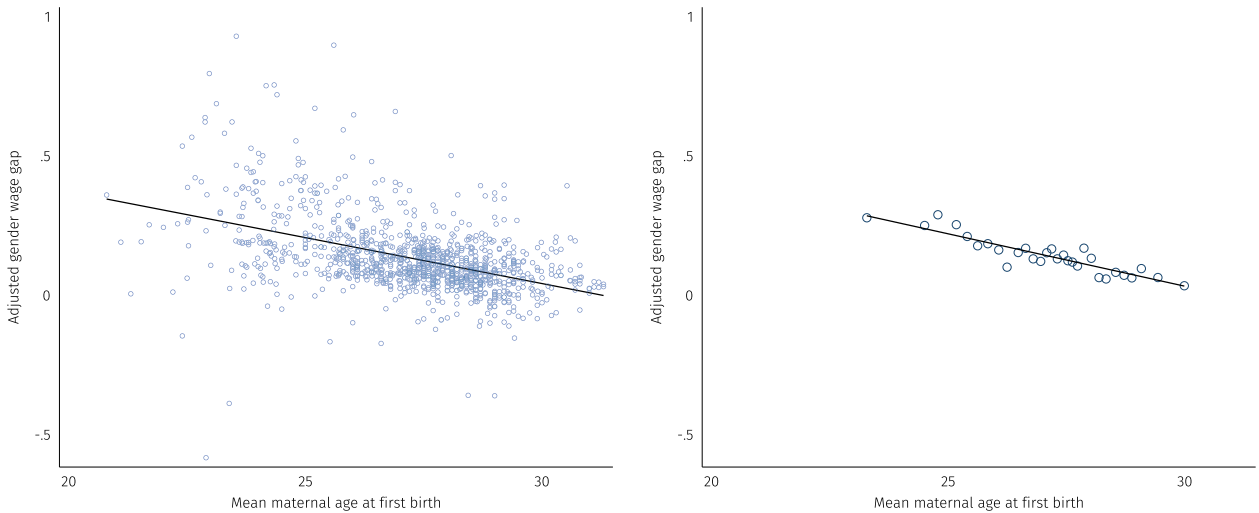



**Figure 1.** Fertility timing and the adjusted gender wage gap among youth.
*Notes:* For each country, year, and data source we utilize one estimate, that with the maximum number of available control factors subject to the constraint that at least 75% of individuals of each gender find a match among the opposite gender. Given that for each year and country more than one data source may be available, we account for multiple observations through weights: $weight = 1/N_{c,y}$, where $N_{c,y}$ denotes a number of estimates for given country in a given year. The left panel presents the raw data. The right panel presents the binned scatter plot adjusted for time trend and data source fixed effects (with 30 bins).

**Table 2.** The effect of fertility timing on AGWG among youth (aged 30 years or less)

| | IV ($\beta^{IV}$) | | | OLS ($\beta$) |
|---|---|---|---|---|
| AGWG estimates | (1) | (2) | (3) | (4) |
| Panel A: *full sample* (all youth AGWG estimates) | | | | |
| Fertility timing | −0.025*** | −0.033*** | −0.025*** | −0.019** |
| | (0.0062) | (0.0093) | (0.0065) | (0.0097) |
| Observations | 1106 | 1161 | 1114 | 1170 |
| $R^2$ | 0.28 | 0.28 | 0.28 | 0.76 |
| *F*-statistic | 25286.6 | 461.1 | 15667.1 | |
| Panel B: *youth AGWG estimates* ***without*** *occupation and industry of employment* | | | | |
| Fertility timing | −0.020*** | −0.030*** | −0.020** | −0.0076 |
| | (0.0076) | (0.0098) | (0.0079) | (0.013) |
| Observations | 1103 | 1158 | 1111 | 1167 |
| $R^2$ | 0.26 | 0.27 | 0.26 | 0.74 |
| *F*-statistic | 20910.9 | 494.9 | 16448.2 | |
| Panel C: *youth AGWG estimates* ***with*** *occupation and industry of employment* | | | | |
| Fertility timing | −0.026*** | −0.032*** | −0.025*** | −0.020*** |
| | (0.0051) | (0.0069) | (0.0053) | (0.0060) |
| Observations | 825 | 864 | 834 | 873 |
| $R^2$ | 0.28 | 0.28 | 0.28 | 0.62 |
| *F*-statistic | 7058.1 | 379.5 | 3894.7 | |
| Clustering SE | Yes | Yes | Yes | Yes |
| Time trends | Yes | Yes | Yes | Yes |

*Notes:* AGWG stands for adjusted gender wage gap. IV specifications using Baltagi (1981) estimator with time-varying and time-invariant component, random-effects models, include year, specification, and data source fixed effects. Column (1) with all instruments jointly. Column (2) with the pill authorization as the only instrument. Column (3) with all instruments *but* the pill authorization. In all IV specifications, we include linear term and base functions up to a fourth polynomial. We report the reduced form estimates in Fig. B4

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

All specifications include a time trend, data source fixed effects, and adjust for the AGWG model specification. The OLS specifications adjust for *weight* = $1/N_{c,y}$, where $N_{c,y}$ denotes a number of data sources for given country in a given year. All columns report estimates for young workers (between 20 and 30 years of age).

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

behind industry and occupation choice is not gender neutral and the returns to those characteristics should be included in the unexplained component. Finally, panel C reports estimates obtained from a sample where all AGWG measures adjust occupation and industry, which corresponds to a narrow definition. Recall that for data points (country × year) it is possible that no estimates for panel C are available,

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

but all countries and years included in panel C, will be included in panels A and B. By focusing on AGWG estimates that meet some criteria, the sample size drops from panel A to either B or C. The sample size decreases more in panel C, which reflects both the fact that some databases do not collect information on occupation or industry, and that the inclusion of these variables reduces the common support, i.e., in some database the percentage of matched individuals of each fender drops below the 75% threshold. The restriction in panel C implies a drop from 1100 + samples (country × year × data source) to 800 + samples. In spite of differences across samples, estimates of the relationship between fertility timing and the AGWG remain essentially unaffected.

#### 4.1.1. Robustness

While the Baltagi (1981) estimator is particularly well suited to our data, we study the sensitivity of the overall conclusions to alternative instrumental variable estimators. The OLS estimate of $\beta$ reported in column (4) of Table 2 adjust for country fixed effects, which we replicate in an IV setting in columns (1) and (2) of Table 3. We employ conventional fixed effects IV in column (1) of Table 3, subsequently high-dimensional fixed effects IV estimator in column (2). Both prove to deliver highly statistically significant estimates of $\beta^{IV}$ and of similar magnitude as the IV estimates reported in Table 2.

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

Table 3 tests of heterogeneous effects along two dimensions. We also employ a quantile IV estimator to study the magnitude of the effects of fertility along the distribution of AGWG. The results are presented in columns (3)–(5) for the 25th percentile, median, and 75th percentile, respectively. This estimation uses unconditional quantiles of AGWG via recentered influence function (Firpo et al., 2009) and model specification analogous to column (1) from Table 2. The estimated effect of fertility timing on AGWG amounts to roughly −0.02 for the 25th and 50th percentile, it appears to be somewhat higher for the 75th percentile, but still within the ballpark of Baltagi (1981) estimates reported in Table 2.

Next, we explore the heterogeneity along the distribution of fertility timing. We conduct this analysis for the intercepts of the link between fertility timing 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

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

**Table 3.** The effect of fertility timing on youth AGWG – alternative estimators of $\beta^{IV}$

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

*Notes:* IV estimates of the relation between adjusted gender wage gaps (AGWG) and fertility timing. Standard errors clustered at the country–data source level in parentheses in columns (1)–(5). Confidence intervals (95%) in brackets in columns (6) and (7). We report estimations analogous to column (1) from Table 2. In column (1) we use fixed-effects IV estimator (FE 2SLS). In column (2) we use high-dimensional fixed-effects IV estimator (HDFE). In columns (3)–(5) we utilize Firpo et al. (2009) recentered influence function transformation of the estimator from column (1) in Table 2, the transformation is for the 25th, 50th, and 75th percentile of AGWG. In columns (6a) and (6b) we provide estimates of quantile IV estimator, accounting for the quantiles of mean maternal age at first birth (intercepts and slopes).

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

All specifications include time trends, data source fixed effects, and adjust for the AGWG model specification. Asterisks ***, **, and * denote significance at 10%, 5%, and 1%, respectively.

Analogous set of estimates for sample restricted as in panel C of Table 2 is reported in Table C1 in Appendix C. Full set of estimates from first and second stage regressions is available upon request.

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. The level effect appears to be the highest when mean maternal age at first birth is within the first quartile, but this observation may partly be a consequence of the fact that the fourth quartile in terms of fertility timing is very close to the upper boundary on the age in the estimation of adjusted gender wage gaps.[25]

The OLS and the IV estimates of delayed fertility on gender wage gaps at young age reveal an effect of roughly 2-3 percentage points. Given that mean maternal age at first birth in our sample increased from 25.96 in 1990s to 28.16 in 2010s, these estimates imply that essentially a third of the decline in AGWG among young workers can be attributed to delayed fertility. Note that this delay in fertility is of paramount importance from the perspective of the employer–employee relationship: Cazes Tonin (2010) show that the mean employment tenure of European workers below 30 years old in 2006 was between 1.5 and 2.5 years. A year of fertility delay is a sizable chunk of the average tenure in these countries.[26]

#### 4.1.2. Discussion

The instruments used in our study have been previously demonstrated to affect various margins of fertility, labor supply, and human capital in individual-level investigations. It is thus not warranted that fertility timing is the only channel through which our instruments affect the AGWG.[27] We study if and to what extent our instruments could operate through other channels, that is in addition to instrumenting for fertility timing. We explore three distinct channels: employment, education, and (contemporaneous) fertility rates. In addition, we account for GDP per capita as a measure of labor productivity (akin to the opportunity cost of not working). Table C2 in C shows that once we account for these direct channels, the IV estimates remain in the same ballpark (despite considerable reduction in the countries and years due to data limitations). Unfortunately, time-varying information concerning gender norms is not available for a large group of countries in our sample, so we cannot study if they are affected by our instruments and at the same time could affect both women's and men's wages.

Further, we study if the effects on youth, as identified in Table 2 are relevant for the labor market in general. To this end, Table C3 in Appendix C reports estimations analogous to Table 2, where the dependent variable is AGWG for all age groups. To run this analysis, we re-estimate AGWG from individual-level data without age restriction (including in the AGWG estimation the controls for age groups). We then re-estimated the specifications to obtain analogous results for all working-age groups. The point estimates are around half the size, but they remain negative and

[25]Indeed, Q25 threshold is 26 years of age and Q75 threshold is 28 years, whereas we estimate AGWG until the age of 30. There might not be enough variation in fertility timing to yield a statistically significant intercept in Q2–Q4 of the distribution of mean maternal age at first births.

[26]In terms of probability of child-bearing, Table D1 Appendix D shows that in European countries even a small delay in fertility involved substantial increases in the probability of being childless: among women aged 25, the increase in childlessness was of around 20 percentage points, still holding in two-digits for among women aged 30.

[27]Ours is a country-level study, however, and to the best of our knowledge, there is no evidence utilizing our instruments in the comparative studies of adjusted gender wage gaps.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

statistically significant. This corroborates the intuition that fertility timing has persistent effect on gender inequality among age groups.

In the main results, *fertility timing* is captured by the mean maternal age at first birth. Table C4 in Appendix C explores the relation between gender wage gaps and fertility *level* (as measured by TFR). This check amounts to replacing MAB1 with TFR as the endogenous variable in equations (2) and (3), and re-estimating the IV regressions. Consistent with our earlier discussion on how to capture fertility timing, there is no evidence of a relationship between TFR and the estimated gender wage gaps.[28] These estimates should be taken with caution, as one cannot prove the lack of a relationship. Moreover, delayed fertility can in principle imply fewer children and hence lower fertility rates. However, these patterns are not visible in the data, whereas the patterns for fertility timing are strong and robust.

Our approach rests upon using several instruments simultaneously. In practice this requires partial monotonicity, that is conditional on all other instruments no remaining instrument affects the endogenous variable in opposite directions across countries or years (partial monotonicity, see Mogstad et al., 2021). Our results remain robust to including one instrument at the time as well as a combination of instruments, which is somewhat reassuring that partial monotonicity is not violated in our case. These estimates are available in Tables C5 and C6 in Appendix C. Table C5 explores changes related to the inclusion of each set of instruments one by one, whereas C6 includes only one instrument at a time.

Our estimates of AGWG adjust for the fact of having children: when comparing the wages of men and women, child-rearing is taken into account. However, one way to think about our results would be to focus only on young individuals with children. However, such exercise could be misleading in a sense that estimates of AGWG among parents alone ignore the marginal parent, whereas those obtained for parents and non-parents average among the marginal parent. Furthermore, this exercise would further reduce sample sizes for many countries. This would be especially acute for the sample coming from ISSP, thus reducing substantially the non-European variation in our sample.

Nopo (2008) ensures that gender wage gaps are computed as a weighted average of differences among men and women of identical characteristics. It is possible that the relative importance of high wage gap groups declines and share of low wage gap groups increases. For example, as higher fraction of individuals have a university degree, the contribution of the gap between men and women with this level of education to the total AWGW becomes larger. This could be particularly relevant in our case if changes in mean maternal age at first birth were also driven primarily by composition changes. For example, groups with early fertility would become less numerous whereas groups with delayed fertility would become more numerous in a given country in a given year. One should expect no adjustment in the beliefs of the employers *among their potential pool of employees*, but weighted average of wage gaps and weighted average of mean maternal age at first birth would display co-movement. In such case, our estimates would reflect composition changes rather than adjustment of the beliefs of specific employers. We cannot rule this channel out in our estimations. However, in Table C2 in Appendix C we adjust the estimates for a broad array of measures of the composition of workers and their variation over time. Our results remain unaffected.

[28]Note that TFR is in general more broadly available across countries and years, hence the estimates on Table C4 use samples with roughly 200 more country–year observations than in Table 2.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

### 4.2. Benchmarking statistical gender discrimination

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 the mechanism relies on rationality of employers: it ought to reflect the expected productivity gap between young men and women. To establish an empirical analog of the productivity gap, we consider the following framework.

Young adults draw a parental type $c_i$ from a distribution, with $E_m(c_i) = c_m < E_w(c_i) = c_w$, which denotes how involved they will be in the care of the child.[29] In principle, $c_i$ can take any real value. While for most parents $c_i$ would be positive, some parents might have no cost of caring, and some might even see positive productivity spillovers from having children, e.g., if having a child improves their motivation to work. The assumption of $c_m < c_w$ states that costs are on average higher for women (see Becker et al., 2019). Parental type is private information, and it cannot be communicated to the employer *ex ante* (before becoming a parent), whereas *ex post* (after becoming a parent) credible communication is costly, e.g., it depends on child characteristics which may vary over time.[30]

Individuals draw a procreation type: with probability $\pi_i$ 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. Assume that the probability of becoming a parent ($\pi_i$) is independent of the productivity change associated with parenting costs ($c_i$) and it is drawn from a distribution common for both genders.

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

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

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

Under statistical discrimination, the employers offer wages in expectation of individual productivity, averaging over groups. With probability $\pi_i$ of bearing the parenthood costs, the AGWG becomes:

$$\textit{Youth} \quad AGWG: E_m(w|h) - E_w(w|h) = \big(h - E_m(\pi_i)E_m(c_i)\big) - \big(h - E_w(\pi_i)E_w(c_i)\big) \tag{5}$$

$$= E_w(\pi_i)E_w(c_i) - E_m(\pi_i)E_m(c_i) = E(\pi_i) \times (c_w - c_m)$$

Thus, if statistical discrimination stands behind (adjusted) gender wage gaps, a decline in $E(\pi_i)$ *ceteris paribus* should imply narrowing of the AGWG. Note that even if $\pi_i$ is the same across all individuals, there is still averaging due to unobservable differences in

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

[30]A simple and feasible extension is to have the type drawn from a distribution in each period $t$ with distribution for women first-order dominating the distribution for men.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

productivity ($c_i$) between parents and non-parents as well as between mothers and fathers.[31]

Equation (5) takes the employers' perspective and portrays the link between *objective* differences in productivity across genders and the (adjusted) gender wage gap. We aim to obtain analogs of $c_w - c_m$ and $E(\pi_i)$ from observational data to compare our estimates of $E(w_m|h) - E(w_w|h)$. We then juxtapose $E(\pi_i) \times (c_w - c_m)$ to empirical estimates of $E(w_m|h) - E(w_w|h)$. Rationality imposes that the employers form expectations based on the observed probability $E(\pi_i) = \pi$ of incurring the cost $c$.

#### 4.2.1. Implementation

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

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

Effectively, equation (6) an upper bound expectation in a sense that the contract duration with a given employee can be shorter than the age brackets $a \in [20,\ 30]$ (in other words: the actual probability faced by the employer cannot exceed this value).[32] The age brackets of between 20 and 30 are set to be consistent with the age groups for which adjusted gender wage gap was estimated. We construct this indicator using information about the number of first births to women of a given age in a given year. The data come from the Eurostat for European countries and from the Human Fertility Database for the US.

*Productivity cost.* Productivity loss is akin to a reduction in productivity endowment. This endowment is identical for both men and women, regardless of whether they have children or not. To capture $c_w - c_m$, we resort to the time-use data. 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.[33]

In the time-use surveys, household members report the time spent on caring. Child-bearing is not the only reduction to the time endowment. Individuals may have other caring responsibilities and social norms may be driving the gender distribution of those functions. We distinguish between parents in households with children and independent adults in households without children. For each person we obtain the measure of time spent on caring. We then apply formula (7) to obtain the

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

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

[33]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 the University of Michigan.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

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

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

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

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

In the case of the US and the UK, the estimates of AGWG are higher than most of the simulated values implied by $(c_w - c_m) \times \pi$. In these two countries, availability of multiple waves of time-use permits tracing dynamics over time and the higher values of $(c_w - c_m) \times \pi$ tend to be associated with higher values of AGWG estimates, *ceteris paribus*.[34] We infer that the existing AGWG among youth far exceed the productivity gap implied by the data on the timing of fertility and actual reduction in time endowment due to caring obligations. The gap is rather small for Spain and in the ball park of the productivity gap estimates for Austria.

In Appendix E, we present analogous estimates using data from International Social Survey Program (ISSP). Using the ISSP data allows us to expand the list of countries to be studied (in total: 14 out of 55 countries in our estimations). This replication of the benchmarking exercise for a broader group of countries reveals a similar conclusion: for all countries at least one of the measures falls in the confidence interval of the AGWG

[34]The United States exhibits highly volatile evolution of fertility patterns. Specifically, over the available period the $\pi$ 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 $\pi$ received mixed signals about the actual probability of child-bearing over the analyzed period. In Appendix F we discuss the special case of the United States, demonstrating the link between our findings and earlier work of Kuziemko et al. (2018).

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press



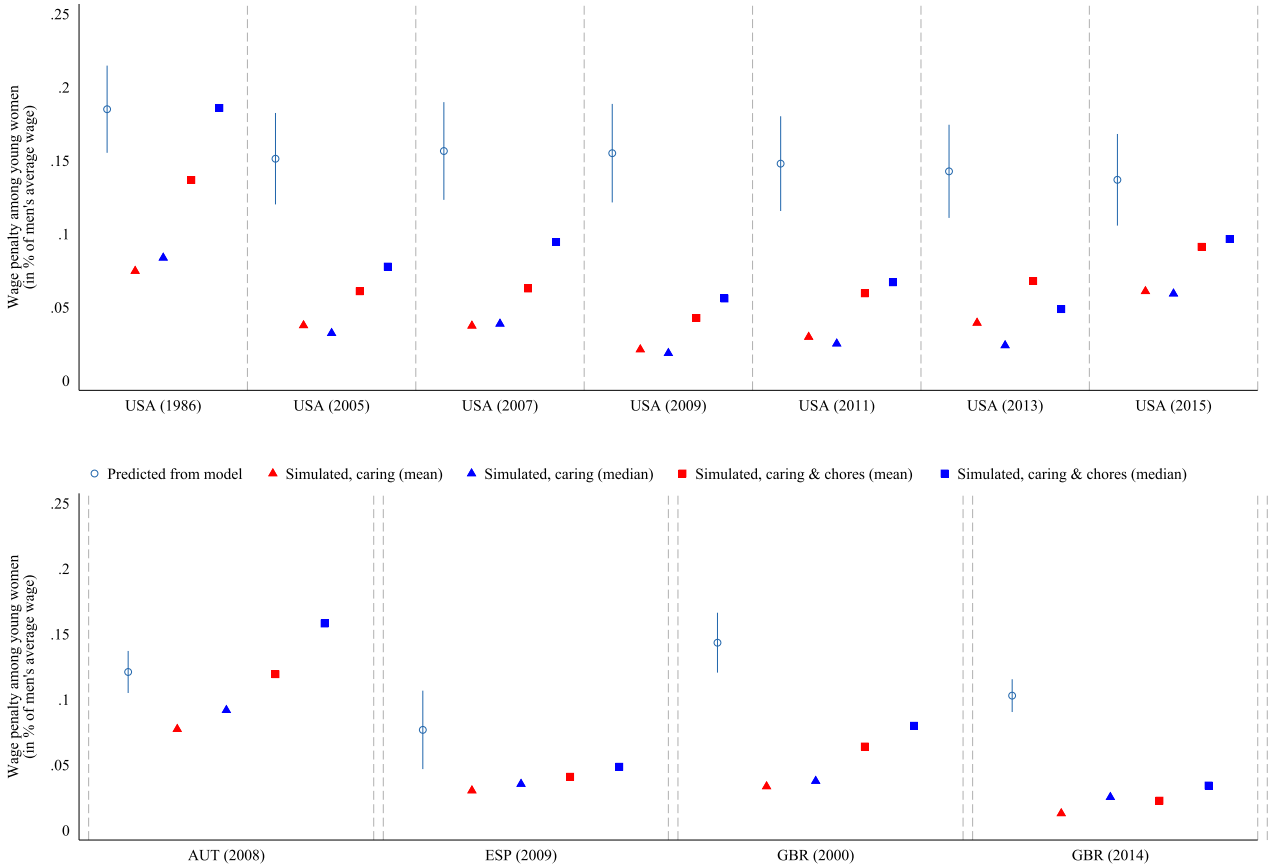



**Figure 2.** Obtaining a benchmark for statistical gender discrimination – time-use surveys.
*Notes:* the data come from the available time-use surveys obtained from MTUS project and I-PUMS. Estimates from the model obtained as marginal predictions from the estimates (3), adjusting for (2), for the year of MTUS availability for each country. Analogous simulations for the ISSP data are reported in Appendix E. MTUS for the United Kingdom is available for 2000 and 2014. Data availability for the US is described in Appendix F. 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.

estimates.[35] Our results are consistent with the findings of Gallen (2023) for Denmark, who shows that the pay gap for mothers is entirely explained by productivity, whereas the gap for non-mothers is not.

## 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, when 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 in line with the data. A delay in fertility observed around the world over the past decades provides a convenient context for evaluating if (adjusted) gender wage gaps among young workers are consistent with the hypothesis of statistical discrimination. In this study, we provide estimates of adjusted wage gaps between young men and women from 56 countries around the world, spanning four decades and compare those estimates with the evolution of mean maternal age at first birth.

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

The fact that AGWG for young workers declines with delayed fertility is not proving that entire gap is due to statistical discrimination: employers could adjust slowly their expectations to changes in fertility timing, 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 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.

[35]Note that self-reported time-use data from ISSP may be subject to measurement error, which could be gendered. Moreover, sample sizes for individuals aged 20–30 years old are relatively small, when split between with and without children in ISSP.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

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

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

In terms of policy implications, our study shows that probability of child-bearing 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.

Given our results, future research could focus on study whether and to what extent the shifts in the gender wage gap by age follow the same pattern as the shifts in the age-profile of delivering a first child. To the best of our knowledge, such analysis could be performed only for a few advanced economies, of which one – the United States – exhibits patterns very different from the rest of the world. More research is needed to understand the origins of these country-specific developments and extracting further insights common across countries.

**Supplementary material.** The supplementary material for this article can be found at https://doi.org/10.1017/dem.2024.21.

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

**Competing interests.** Joanna Tyrowicz and Lucas van der Velde acknowledge the support received by the National Science Centre (Narodowe Centrum Nauki) through the project grant #2017/27/L/HS4/03219. The authors do not hold any position that could create a conflict of interest.

## References

Adda, J., C. Dustmann, & K. Stevens (2017). The career costs of children. *Journal of Political Economy*, *125*(2), 293–337. https://doi.org/10.1086/690952

Altonji, J. G., & C. R. Pierret (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, *116*(1), 313–350. https://doi.org/10.1162/003355301556329

Amuedo-Dorantes, C., & J. Kimmel (2005). The motherhood wage gap for women in the United States: The importance of college and fertility delay. *Review of Economics of the Household*, *3*(1), 17–48. https://doi.org/10.1007/s11150-004-0978-9

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Ananat, E. O., & D. M. Hungerman (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. https://doi.org/10.1162/REST_a_00230

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

Auspurg, K., T. Hinz, & C. Sauer (2017). Why should women get less? Evidence on the gender pay gap from multifactorial survey experiments. *American Sociological Review*, *82*(1), 179–210. https://doi.org/10.1177/0003122416683393

Baert, S. (2014). Career lesbians. Getting hired for not having kids? *Industrial Relations Journal*, *45*(6), 543–561. https://doi.org/10.1111/irj.2014.45.issue-6

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. https://doi.org/10.1257/aer.100.1.98

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

Baizan, P., B. Arpino, & C. E. Delclós (2016). The effect of gender policies on fertility: The moderating role of education and normative context. *European Journal of Population*, *32*(1), 1–30. https://doi.org/10.1007/s10680-015-9356-y

Baltagi, B. H. (1981). Simultaneous equations with error components. *Journal of Econometrics*, *17*(2), 189–200. https://doi.org/10.1016/0304-4076(81)90026-9

Bartoš, V., M. Bauer, J. Chytilová, & F. Matějka (2016). Attention discrimination: Theory and field experiments with monitoring information acquisition. *American Economic Review*, *106*(6), 1437–75. https://doi.org/10.1257/aer.20140571

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

Becker, S. O., A. Fernandes, & D. Weichselbaumer (2019). Discrimination in hiring based on potential and realized fertility: Evidence from a large-scale field experiment. *Labour Economics*, *59*, 139–152. https://doi.org/10.1016/j.labeco.2019.04.009

Bergmann, B. (1973). *Sex discrimination in wages: Comment*. Princeton University Press, pp. 152–154.

Bingley, P., P. Lundborg, & S. V. Lyk-Jensen (2020). The opportunity costs of mandatory military service: Evidence from a draft lottery. *Journal of Labor Economics*, *38*(1), 39–66. https://doi.org/10.1086/704495

Black, S. E., P. J. Devereux, & K. G. Salvanes (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. https://doi.org/10.1111/j.1468-0297.2008.02159.x

Blau, F. D., & L. M. Kahn (2017). The gender wage gap: Extent, trends, and explanations. *Journal of economic literature*, *55*(3), 789–865. https://doi.org/10.1257/jel.20160995

Boelmann, B., A. Raute, & U. Schönberg (2020). Wind of change?: Cultural determinants of maternal labor supply. CREAM Discussion Paper 2020–20.

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

Bongaarts, J., & A. K. Blanc (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. https://doi.org/10.1186/s12963-015-0058-9

Booth, A. L., & H. J. Kee (2009). Intergenerational transmission of fertility patterns. *Oxford Bulletin of Economics and Statistics*, *71*(2), 183–208. https://doi.org/10.1111/obes.2009.71.issue-2

Brunello, G., M. Fort, & G. Weber (2009). Changes in compulsory schooling, education and the distribution of wages in Europe. *The Economic Journal*, *119*(536), 516–539. https://doi.org/10.1111/j.1468-0297.2008.02244.x

Bygren, M., A. Erlandsson, & M. Gähler (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. https://doi.org/10.1093/esr/jcx051

Card, D., & A. R. Cardoso (2012). Can compulsory military service raise civilian wages? Evidence from the peacetime draft in portugal. *American Economic Journal: Applied Economics*, *4*(4), 57–93.

Cazes, S., & M. Tonin (2010). Employment protection legislation and job stability: A European cross-country analysis. *International Labour Review*, *149*(3), 261–285. https://doi.org/10.1111/ilr.2010.149.issue-3

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Correll, S. J., S. Benard, & I. Paik (2007). Getting a job: Is there a motherhood penalty? *American Journal of Sociology*, *112*(5), 1297–1338. https://doi.org/10.1086/511799

Cortés, P., & J. Pan (2023). Children and the remaining gender gaps in the labor market. *Journal of Economic Literature*, *61*(4), 1359–1409. https://doi.org/10.1257/jel.20221549

Costa Dias, M., R. Joyce, & F. Parodi (2020). The gender pay gap in the UK: Children and experience in work. *Oxford Review of Economic Policy*, *36*(4), 855–881. https://doi.org/10.1093/oxrep/graa053

Cukrowska-Torzewska, E., & A. Lovasz (2020). The role of parenthood in shaping the gender wage gap – a comparative analysis of 26 European countries. *Social Science Research*, *85*, 102355. https://doi.org/10.1016/j.ssresearch.2019.102355

Cygan-Rehm, K., & M. Maeder (2013). The effect of education on fertility: Evidence from a compulsory schooling reform. *Labour Economics*, *25*, 35–48. https://doi.org/10.1016/j.labeco.2013.04.015

Devereux, P. J., & R. A. Hart (2010). Forced to be rich? Returns to compulsory schooling in Britain. *The Economic Journal*, *120*(549), 1345–1364. https://doi.org/10.1111/j.1468-0297.2010.02365.x

Fenoll, A. A., & Z. Kuehn (2017). Compulsory schooling laws and migration across European countries. *Demography*, *54*(6), 2181–2200. https://doi.org/10.1007/s13524-017-0615-x

Finlay, J., D. Canning, & J. Po (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., N. M. Fortin, & T. Lemieux (2009). Unconditional quantile regressions. *Econometrica*, *77*(3), 953–973. https://doi.org/10.3982/ECTA6822

Fuller, S., & L. P. Cooke (2018). Workplace variation in fatherhood wage premiums: Do formalization and performance pay matter? *Work, Employment and Society*, *32*(4), 768–788. https://doi.org/10.1177/0950017018764534

Gallen, Y. (2023). Motherhood and the gender productivity gap. BFIRE Working Paper 2018-41 updated from the author's website, Becker Friedman Institute for Research in Economics.

Gangl, M., & A. Ziefle (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. https://doi.org/10.1353/dem.0.0056

Goldin, C. (2014). A grand gender convergence: It's last chapter. *American Economic Review*, *104*(4), 1091–1119. https://doi.org/10.1257/aer.104.4.1091

Goldin, C., & L. F. Katz (2002). The power of the pill: Oral contraceptives and women's career and marriage decisions. *Journal of Political Economy*, *110*(4), 730–770. https://doi.org/10.1086/340778

Goldscheider, F., E. Bernhardt, & M. Brandén (2013). Domestic gender equality and childbearing in Sweden. *Demographic Research*, *29*, 1097–1126. https://doi.org/10.4054/DemRes.2013.29.40

González, M. J., C. Cortina, & J. Rodríguez (2019). The role of gender stereotypes in hiring: A field experiment. *European Sociological Review*, *35*(2), 187–204. https://doi.org/10.1093/esr/jcy055

Goraus, K., J. Tyrowicz, & L. van der Velde (2017). Which gender wage gap estimates to trust? A comparative analysis. *Review of Income and Wealth*, *63*(1), 118–146. https://doi.org/10.1111/roiw.2017.63.issue-1

Grenet, J., R. A. Hart, & J. E. Roberts (2011). Above and beyond the call. Long-term real earnings effects of British male military conscription in the post-war years. *Labour Economics*, *18*(2), 194–204. https://doi.org/10.1016/j.labeco.2010.09.001

He, D., X. Zhang, Z. Wang, & Y. Jiang (2019). China fertility report, 2006–2016. *China Population and Development Studies*, *2*(4), 430–439. https://doi.org/10.1007/s42379-019-00022-9

He, H., S. X. Li, & Y. Han (2023). Labor market discrimination against family responsibilities: A correspondence study with policy change in China. *Journal of Labor Economics*, *41*(2), 361–387. https://doi.org/10.1086/719966

Heckman, J., & S. Navarro-Lozano (2004). Using matching, instrumental variables, and control functions to estimate economic choice models. *Review of Economics and Statistics*, *86*(1), 30–57. https://doi.org/10.1162/003465304323023660

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. https://doi.org/10.1093/esr/jcz056

Ishimaru, S. (2022). Empirical decomposition of the IV-OLS gap with heterogeneous and nonlinear effects. *Review of Economics and Statistics*, 1–45.

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Jarrell, S. B., & T. D. Stanley (2004). Declining bias and gender wage discrimination? A meta-regression analysis. *The Journal of Human Resources*, *39*(3), 828. https://doi.org/10.2307/3558999

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

Kleven, H., C. Landais, & J. E. Søgaard (2019b). 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. https://doi.org/10.1080/00324728.2013.819112

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. https://doi.org/10.4054/DemRes.2015.32.51

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

Landivar, L. C. (2020). First-birth timing and the motherhood wage gap in 140 occupations. *Socius: Sociological Research for a Dynamic World*, *6*, 237802312093942.

Lee, D. L., J. McCrary, M. J. Moreira, & J. Porter (2020). Valid t-ratio inference for IV. Preprint arXiv:2010.05058.

Mogstad, M., & A. Torgovitsky (2018). Identification and extrapolation of causal effects with instrumental variables. *Annual Review of Economics*, *10*, 577–613. https://doi.org/10.1146/economics.2018.10.issue-1

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

Mouganie, P. (2019). Conscription and the returns to education: Evidence from a regression discontinuity. *The Scandinavian Journal of Economics*, *122*(3), 1112–1139. https://doi.org/10.1111/sjoe.v122.3

Mulligan, C. B., & A. Shleifer (2005). Conscription as regulation. *American Law and Economics Review*, *7*(1), 85–111. https://doi.org/10.1093/aler/ahi009

Murtin, F., & M. Viarengo (2011). The expansion and convergence of compulsory schooling in Western Europe, 1950–2000. *Economica*, *78*(311), 501–522. https://doi.org/10.1111/j.1468-0335.2009.00840.x

Neels, K., M. Murphy, M. Ní Bhrolcháin, & E. Beaujouan (2017). Rising educational participation and the trend to later childbearing. *Population and Development Review*, *43*(4), 667–693. https://doi.org/10.1111/padr.2017.43.issue-4

Nopo, H. (2008). Matching as a tool to decompose wage gaps. *Review of Economics and Statistics*, *90*(2), 290–299. https://doi.org/10.1162/rest.90.2.290

Norman, P. (2003). Statistical discrimination and efficiency. *Review of Economic Studies*, *70*(3), 615–627. https://doi.org/10.1111/1467-937X.00258

OECD (2024). Fertility rates (indicator).

Okun, B. S., & L. Raz-Yurovich (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. https://doi.org/10.1111/padr.2019.45.issue-1

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

Pager, D., & D. Karafin (2009). Bayesian bigot? Statistical discrimination, stereotypes, and employer decision making. *Annals of the American Academy of Political and Social Science*, *621*(1), 70–93. https://doi.org/10.1177/0002716208324628

Pearson, K., & A. Lee (1899). On the inheritance of fertility in mankind. *Philosophical Transactions of the Royal Society of London. Series A*, *192*, 257–330. https://doi.org/10.1098/rsta.1899.0006

Peterson Gloor, J. L., T. G. Okimoto, & E. B. King (2021). 'Maybe baby?' The employment risk of potential parenthood. *Journal of Applied Social Psychology*, *52*(8), 623–642. https://doi.org/10.1111/jasp.v52.8

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. https://doi.org/10.1016/j.labeco.2006.01.006

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

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press

Pischke, J.-S., & T. von Wachter (2008). Zero returns to compulsory schooling in germany: Evidence and interpretation. *Review of Economics and Statistics*, *90*(3), 592–598. https://doi.org/10.1162/rest.90.3.592

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

Steenhof, L., & A. C. Liefbroer (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. https://doi.org/10.1080/00324720701788616

Taniguchi, H. (1999). The timing of childbearing and women's wages. *Journal of Marriage and the Family*, *61*(4), 1008. https://doi.org/10.2307/354020

van der Velde, L., J. Tyrowicz, & J. Siwinska (2015). Language and (the estimates of) the gender wage gap. *Economics Letters*, *136*, 165–170. https://doi.org/10.1016/j.econlet.2015.08.014

Vignoli, D., S. Drefahl, & G. De Santis (2012). Whose job instability affects the likelihood of becoming a parent in Italy? A tale of two partners. *Demographic Research*, *26*, 41–62. https://doi.org/10.4054/DemRes.2012.26.2

Vignoli, D., V. Tocchioni, & A. Mattei (2020). The impact of job uncertainty on first-birth postponement. *Advances in Life Course Research*, *45*, 100308. https://doi.org/10.1016/j.alcr.2019.100308

Wang, P., & J. Chen (2023). Fertility-based employment discrimination and family policy: Evidence from a field experiment. *Social Policy & Administration*, *57*(7), 1135–1149. https://doi.org/10.1111/spol.v57.7

Weichselbaumer, D., & R. Winter-Ebmer (2005). A meta-analysis of the international gender wage gap. *Journal of Economic Surveys*, *19*(3), 479–511. https://doi.org/10.1111/j.0950-0804.2005.00256.x

Wood, J., & K. Neels (2017). First a job, then a child? Subgroup variation in women's employment-fertility link. *Advances in Life Course Research*, *33*, 38–52. https://doi.org/10.1016/j.alcr.2016.09.003

Yu, W.-h., & Y. Hara (2021). Motherhood penalties and fatherhood premiums: Effects of parenthood on earnings growth within and across firms. *Demography*, *58*(1), 247–272. https://doi.org/10.1215/00703370-8917608

**Cite this article:** Tyrowicz, J., & van der Velde, L. (2024). Statistical gender discrimination: evidence from young workers across four decades and 56 countries. *Journal of Demographic Economics* **90**, 659–688. https://doi.org/10.1017/dem.2024.21

https://doi.org/10.1017/dem.2024.21 Published online by Cambridge University Press