

The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census[†]

By CLAUDIA GOLDIN, SARI PEKKALA KERR, CLAUDIA OLIVETTI, AND ERLING BARTH*

The gender earnings gap is a shifting statistic. It widens with age, especially in the 15 to 20 years after school leaving. It expands more for those who are married and even more for those with young children. The increased gap is far greater among college graduates than others and in sectors known to penalize shorter hours, job flexibility, and time off (see Bertrand, Goldin, and Katz 2010 on MBAs). It exists even when controls are added for hours worked per week and weeks per year.

The widening over the lifecycle, moreover, is substantial. Using the March Current Population Survey (CPS), Goldin (2014) shows that the gender earnings gap for college graduates born circa 1970 (given hours, weeks, and education controls) widened by 21.4 log points (from -0.10 to -0.314) from ages 25–29 to 40–44 and similarly for those born circa 1960 between the same ages (from -0.133 to -0.343). Excluding hours and weeks controls, the gap expanded by 39 log points for both cohorts. Although there is evidence that the gender earnings gap has narrowed after cohorts are in their late 40s, our focus here is on the widening portion of the shifting statistic.

* Goldin: Department of Economics, Harvard University, Cambridge, MA 02138 (e-mail: cgoldin@harvard.edu); Kerr: Wellesley College, WCW 106 Central Street, Wellesley, MA 02481 (e-mail: skerr3@wellesley.edu); Olivetti: Boston College, 104 Commonwealth Avenue, Chestnut Hill, MA 02147 (e-mail: olivettc@bc.edu); Barth: Institute for Social Research, Postboks 3233 Elisenberg 0208 Oslo, Norway (e-mail: erling.barth@socialresearch.no). Kerr and Barth thank the Research Council of Norway (#236770 and #179552) for funding. The research was conducted while Kerr and Barth were Special Sworn Status researchers of the US Census Bureau, Boston Census Research Data Center. Research results and conclusions are the authors' and may not reflect the views of the Census Bureau. This paper was screened to ensure that no confidential data are revealed.

[†] Go to <https://doi.org/10.1257/aer.p20171065> to visit the article page for additional materials and author disclosure statement(s).

An extensive literature exploring the factors responsible for the widening has emphasized the role of children, tied movers and stayers, greater demand by women for work amenities such as flexibility, and less internal advancement for mothers. Given these possibilities, we use the Longitudinal Employer-Household Dynamics (LEHD) database linked to the 2000 census to explore how much of the widening occurs within versus between establishments by mean establishment earnings (see Bayard et al. 2003 for an analysis of the cross-sectional gender pay gap using earlier firm-level data).

We seek to know if women are in lower wage establishments, more than men, and move with less frequency to higher wage ones, and whether they are less apt to change their relative position within establishments. That is, we ask how much of the expanding gender earnings gap in the decade and a half after schooling ends is due to shifting employment by men and women across establishments differing in mean earnings and how much is due to differential wage growth within establishments, for narrowly defined age and education groups.

Our main findings are that the gender earnings gap among college graduates in the LEHD, given basic demographic controls, expanded by 33.7 log points from ages 26 to 39 (from 1995–2008 for those born around 1970) and that about 80 percent of the increase occurred in the first half of the period. Given (ln) mean establishment earnings (*MEE*), industry, and occupation (in 2000), the gender earnings gap increases less: by 18.9 log points. Thus, 44 percent of the gap widening for college graduates is due to differential mobility between establishments by gender and 56 percent is due to differential earnings changes within establishments by gender (see also Barth, Kerr, and Olivetti 2017).

For those who graduated high school but not college, the gap grows from ages 23 to 36 (1995–2008) by 15.6 log points but expands

by just 8.1 log points given *MEE* (plus industry and occupation). The gap expands less than for the college group, but the percentage impact of including *MEE*, and thus the between and within percentages, are about the same.

We also explore changes among college graduates in several important sectors (Finance, Insurance, and Real Estate—henceforth, FIRE—and Technology) as well as differences by marital status. The bottom line is that the widening is concentrated in the first half of the age range and is due somewhat more to what happens within establishments, although much occurs because of differential moves by gender across establishments.

I. Exploring the Expanding Gender Earnings Gap Using the LEHD-Census

To understand some of the factors behind the increase in the gender earnings gap, we use the 2000 census (one in six long form) linked to the LEHD database. The LEHD used here covers the years 1995 to 2008 and is a rich database of private-sector firms containing administrative earnings data from state unemployment insurance (UI) records. UI earnings include wages, salary, and taxable bonuses and are not top-coded. The virtues and deficiencies of the LEHD have been described in detail by others (see, e.g., Barth et al. 2016), thus we will be succinct.

Demographic and some economic information (e.g., education, marital status, occupation) come from the 2000 US Population census long-form, whereas some others (e.g., sex, age) come from the UI records. Individuals from the long-form records are matched, when possible, to their employer in the LEHD. A virtue of these data is that, in principle, all individuals working in the private sector who filled out the 2000 census long form can be linked to the establishments at which they were employed in each year from 1995 to 2008. A limitation is that occupation and marital status are recorded only for 2000. Only the 23 states that provided state UI data can be included in the LEHD. We are, in addition, using firms only in the 50 largest Primary Metropolitan Statistical Areas (PMSA), of which there are 26 in 18 of the 23 states. Only those whose contemporaneous employer is in one of these 26 PMSAs can be tracked. To make certain that we are tracking individuals in our data for a sufficient

period, we construct an analysis sample of individuals present in the sample (and having earnings that exceed ten hours per week at the minimum wage) for at least half the quarters considered.

The analysis sample is divided into three education groups. For each group, age at the start of the dataset in 1995 is chosen so that individual respondents would probably have completed their highest grade yet be young enough to be in their earliest jobs. Those who are not high school graduates are 20–22 years in 1995; those who are high school graduates but not college graduates are 23–25 years in 1995; and college graduates are 25–27 years in 1995. About 40 percent of the women and 33 percent of the men are college graduates; 54 percent of both sexes are high school graduates but not college graduates, and the rest are not high school graduates.

Our dependent variable is the $\ln(\text{mean quarterly earnings}) = \text{sum of quarterly earnings in the main job in year } t \text{ divided by the number of quarters in the main job that year}$. Note that there is no information in the LEHD on hours or weeks worked.

II. Establishment Effects and the Expanding Gender Earnings Gap

We analyze the data by individual (i) for each year (t) from 1995 to 2008. Given our data construction, calendar year also tracks the age of the individuals in each of the three education groups (j). We estimate the following equations for the 42 education-year ($j \times t$) groups:

$$(1) \quad \ln(y_{ijt}) = \alpha_{jt} + \sum_k \beta_{kjt} X_{ijkt} + \varphi_{jt} F_{ij} + \varepsilon_{ijt},$$

where y is mean quarterly earnings for individual i of education level j in year t ; X is a vector of k individual characteristics (time invariant: race, exact education; time varying: exact age, state) and F indicates female.

In equation (2) we add $\ln(MEE)$, log of mean establishment earnings, which is the (constant dollar) mean for each establishment across the full period for all employees (thus does not vary with time and is not calculated from only those in our analysis sample). We also add I , the (three-digit SIC level) industry (there are about 450) of the establishment and (time invariant)

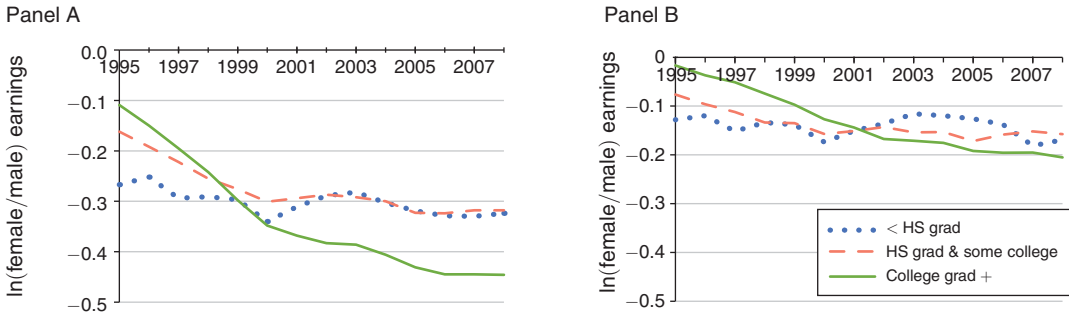


FIGURE 1. GENDER EARNINGS GAP FOR THREE EDUCATION GROUPS: 1995–2008

Notes: Panel A gives coefficients on female, φ_{jt} , from equation (1). Panel B gives coefficients on female, φ_{jt} , from equation (2), which includes $\ln(MEE)$ (mean establishment earnings), three-digit SIC industry and occupation (from 2000 census). < HS grad are 20–22 years old in 1995; HS grads and some college are 23–25 in 1995; college grad+ are 25–27 years old in 1995. Estimation uses LEHD-census 2000. See text for analysis sample selection.

three-digit occupation dummies, O , from the 2000 census:

$$(2) \quad \ln(y_{ijt}) = \alpha_{jt} + \sum_k \beta_{kjt} X_{ijk} + \gamma_{jt} \ln(MEE)_{ijt} + \xi_{jt} I_{ijt} + \omega_{jt} O_{ij} + \varphi_{jt} F_{ij} + \varepsilon_{ijt}.$$

We graph, in Figure 1, panel A, the evolution of the estimated φ_{jt} , which gives the gender earnings gap by year (or age) and education group from equation (1). In Figure 1 panel B we graph the same for equation (2), which includes $\ln(MEE)$, industry and occupation. Because of the large sample sizes, standard errors are small and have not been included in the figures.

We note that although the LEHD is a longitudinal dataset, individuals can exit the data if they shift their employment away from one of the 26 PMSAs and for a host of other reasons. Because of that, we use observations for individuals present at least half the years and we use the LEHD as a set of repeated cross sections. We use the longitudinal aspect of the establishment information; MEE is computed over the entire period and is not time variant.

Data from the LEHD-census clearly show a widening earnings gender gap for all education groups except the lowest in the 14 years considered. For the college graduate group, the gender earnings gap around age 26 is 10.9 log points but widens somewhat continuously to 44.6 log points by age 39 or by 33.7 log points.

The expansion is a bit below that from the CPS we cited earlier, without controls for hours and weeks worked, but our analysis sample from the LEHD imposes a minimum earnings level. The widening for the middle education group from around ages 23 to 37 is from 16.2 to 31.8 log points. The lowest education group begins with the largest gender earnings gap of 26.7 log points but it increases only slightly to 32.4.

Because the results for equation (2), excluding industry and occupation, are almost identical to those including them, we will more concisely refer to equation (2) as adding $\ln(MEE)$. With the addition of $\ln(MEE)$, all gender earnings gap functions shift up, indicating that men are disproportionately employed in higher wage establishments.

More to the point, the addition twists the two higher-level education functions so that the gender earnings gap in Figure 1, panel B, widens far less than in panel A. For college graduates, the initial gap with MEE is about zero (-0.017) but in 14 years it expands by 18.9 log points to -0.205 . The basic gap had initially expanded by 33.7 log points, so that 14.9 log points (44 percent) are due to differential shifts of male and female workers across establishments by mean earnings levels and 18.9 log points (56 percent) to an expansion of the gender earnings gap within establishment. About 80 percent of the widening occurs from ages 26 to 33 years.

Because the widening is largest for the college graduate group, we also explore the evolving gender earnings gap for college graduates

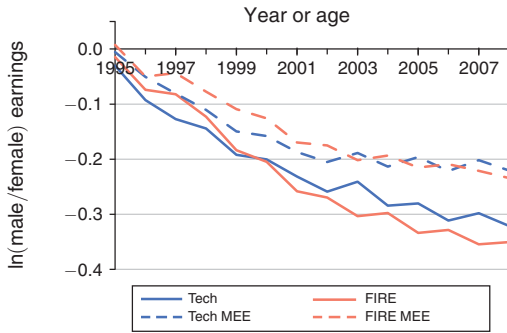


FIGURE 2. GENDER EARNINGS GAP FOR COLLEGE GRADUATES IN TWO SECTORS: 1995–2008 (Ages 25–27 to 38–40)

Notes: Solid lines give coefficients on female, φ_{jt} , from equation (2) with three-digit SIC industry and occupation. Dashed lines also include *MEE* (mean establishment earnings). College graduates are 25–27 years old in 1995 and 38–40 years old in 2008. FIRE = financial, insurance, and real estate; Tech = technology industries as defined by the NSF. Estimation uses LEHD-census 2000. See text.

in various sectors. In Figure 2 we graph the evolution of the estimated φ_{jt} from equation (2) for college graduates in the (two-digit SIC) FIRE sector and in Technology (using the NSF-defined four-digit industries). Three-digit industry and occupation (for 2000) dummies are added so that gender differences within sectors are not solely due to industry and occupational differences between men and women.

The gender gap in earnings is just -0.030 log points in Technology and -0.015 in FIRE around age 26 (within narrow industry and occupation groups) but grows to -0.321 in Technology and -0.350 in FIRE during the next 14 years. Adding *MEE* reduces the gap at 14 years to -0.22 in Technology and -0.234 in FIRE, showing that around 24 to 30 percent of the initial widening was the result of differential changes in earnings by sex between establishments (given industry and occupation). Most of the expansion is due to the greater advancement of men than women within establishments. But it is also the case that in the first half of the period there is a greater shift of men than women into higher paying establishments in FIRE and Tech and the impact on wage growth of being in a high wage establishment also increases. Note that the *MEE* adjusted lines (dashed) for both FIRE and Tech level off after a widening in the gender earnings gap.

Yet another way of cutting the data is to compare the evolution of the estimated φ_{jt} for those ever-married in 2000 (using census data) versus those never-married in 2000 and then to observe the impact of the mean establishment earnings variable. We noted before that the widening of the gender earnings gap has been found to be greater for women who are or have ever been married. Using the LEHD-census data the gender earnings gap for college graduates widened by 39 log points from 1995 to 2008 for those who were ever-married in 2000 whereas it widened by 22 log points during the same period for those who were never-married in 2000. These are suggestive results given the limitations of the data.

III. Discussion and Conclusions

An important literature has developed around the increase in earnings inequality across establishments and the increase in the sorting of workers by firms (see Barth et al. 2016 and Song et al. 2016 on the United States). High wage establishments are employing relatively more high wage workers. Because higher wage firms engage in more rent sharing with their employees, greater sorting by wages has increased earnings inequality (Card, Cardoso, and Kline 2016 analyze rent sharing, segmentation, and gender gaps in Portugal).

In our analysis, the question is whether we are measuring, as we would like to believe, the growth of the gender earnings gap with age or whether we are picking up an increase in inequality. Because age and year are collinear in our analysis, the answer is not immediately clear. We test the possibility by estimating the gender earnings gap for the same age group but changing the year and vice versa. We find that the gender earnings gap during the 1995 to 2008 period increased with age but not by year. That is, the gender earnings gap was virtually the same for each age independent of the year and these results are maintained with the establishment and industry controls. Thus, we have demonstrated that the gender earnings gap widens considerably during the first decade and a half after schooling ends particularly for college graduates, those in certain sectors, and those who were ever-married.

More revealing is that we also show the degree to which the gap widens given mean establishment earnings. Somewhat more than

40 percent of the increased gender gap in earnings is because men disproportionately shift into higher paying establishments. Around 60 percent can be attributed to women's lesser capability to advance their earnings within firms. Finally, most of the widening occurs in the first seven years of our age range just when families are being formed.

Our bottom line is that the widening is split between men's greater ability or preference to move to higher paying firms and positions and their better facility to advance within firms. Both factors increase with women's greater family responsibilities.

REFERENCES

- Barth, Erling, Alex Bryson, James C. Davis, and Richard Freeman.** 2016. "It's Where You Work: Increases in the Dispersion of Earnings across Establishments and Individuals in the United States." *Journal of Labor Economics* 34 (2): S67–97.
- Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti.** 2017. "The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data." Boston College Working Paper 923.
- Bayard, Kimberly, Judith Hellerstein, David Neumark, and Kenneth Troske.** 2003. "New Evidence on Sex Segregation and Sex Differences in Wages from Matched Employee-Employer Data." *Journal of Labor Economics* 21 (4): 887–922.
- Bertrand, Marianne, Claudia Goldin, and Lawrence F. Katz.** 2010. "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2 (3): 228–55.
- Card, David, Ana Rute Cardoso, and Patrick Kline.** 2016. "Bargaining, Sorting, and the Gender Wage Gap: Quantifying the Impact of Firms on the Relative Pay of Women." *Quarterly Journal of Economics* 131 (2): 633–86.
- Goldin, Claudia.** 2014. "A Grand Gender Convergence: Its Last Chapter." *American Economic Review* 104 (4): 1091–119.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter.** 2016. "Firming Up Inequality." National Bureau of Economic Research Working Paper 21199.