

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

BY CLAUDIA GOLDIN, SARI PEKKALA KERR, CLAUDIA OLIVETTI AND ERLING BARTH

Acknowledgments. Kerr and Barth thank the Research Council of Norway (grants #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 have been revealed. A shorter version of this paper was presented in the "Gender Agenda" session at the American Economic Association Meetings in Chicago, January 6, 2017 and is published in the *American Economic Review, Papers and Proceedings*, May 2017, 107(5), pp. 110-14.

The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census
Claudia Goldin, Sari Pekkala Kerr, Claudia Olivetti, and Erling Barth
June 2017
JEL No. J16, J22, J24, J31, J71

ABSTRACT

The gender earnings gap is an expanding statistic over the lifecycle. We use the LEHD-Census 2000 to understand the roles of industry, occupation and establishment 14 years after school leaving. The gap for college graduates from the time they are 26 to 39 years old expands by 34 log points. Most of the widening occurs in the first seven years. About 44 percent of the expanding gender earnings gap is due to disproportionate shifts by men into higher-earning positions, industries and firms and about 56 percent to differential advances by gender within firms. Widening is greater for married individuals and for those in certain sectors. High school graduates without college degrees experience less widening but with similar patterns.

Claudia Goldin
Department of Economics
Harvard University
Cambridge MA 02138
and NBER
cgoldin@harvard.edu

Claudia Olivetti
Department of Economics, Maloney Hall
Boston College
Chestnut Hill MA 02467
and NBER
claudia.olivetti@bc.edu

Sari Pekkala Kerr
Wellesley College
106 Central Street
Wellesley MA 02481
skerr3@wellesley.edu

Erling Barth
Institute for Social Research
P.O. Box 3233 Elisenberg
0208 Oslo Norway
and NBER
erling.barth@socialresearch.no

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, for example, Bertrand, Goldin, and Katz 2010 on MBAs). It exists even when controls are added for hours worked per week and weeks per year.¹

The widening of the gender earnings gap 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 c.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 c.1960 between the same ages (from -0.133 to -0.343). Excluding hours and weeks controls, the gap expanded by more: 39 log points for both cohorts. Although there is evidence that the gender earnings gap narrows after cohorts are in their late 40s, our focus here is on the widening portion of the shifting statistic.

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 establishments 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 relative to men are disproportionately in lower wage establishments. We also want to know if women move with less frequency than do men to higher wage establishments and whether women 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

¹ When we discuss the gender earnings gap we generally mean the log gap, thus the ratio of the earnings of women to men.

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 given sector and occupation.

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 age 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, from ages 26 to 32. Given (ln) mean establishment earnings (*MEE*), industry and occupation (in 2000), the gender earning gap increased 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 (given industry and occupation) and 56 percent is due to differential earnings growth 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 gender earnings gap expands less than for the college group, which is not surprising given the longer right tail of the college graduate earnings distribution. But the percentage impact of including *MEE*, and thus the between and within percentages, are about the same for the two educational groups. We also explore similar changes among college graduates within several important sectors.

The 2000 census gives information on whether the person was ever-married. Substantial differences in the widening of the gender earnings gap exist by that measure of marital status somewhat, even though it is an incomplete measure for the full period. We find that the widening is far greater for those who were ever-married by 2000.

The bottom line is that the widening of the gender earnings gap is concentrated in the first half of the age range, mainly when young people get married and have their children. In addition, the widening is due somewhat more to what happens within establishments given industry and

occupation, although much occurs because of differential moves by gender across establishments.

I. Exploring the Expanding Gender Earning Gap using the LEHD-Census

To understand some of the factors behind the increase in the gender earnings gap with age, we use the 2000 Census (one in six long form) linked to the LEHD database. The version of the LEHD employed here covers the years 1995 to 2008. It 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 brief.

Demographic and some economic information (e.g., education, marital status, occupation) come from the 2000 U.S. 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. One of the advantages 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 during each year from 1995 to 2008.

There are a few limitations of the data. Occupation and marital status are recorded just in the census data for 2000. Another restriction is that only the 23 states that provided state UI data can be included in the LEHD. In addition, because of computational reasons, we restrict our sample to firms in the 50 largest Primary Metropolitan Statistical Areas (PMSA), of which there are 26. These PMSAs, moreover, are found in 18 of the 23 states. Only those whose contemporaneous employer is in one of the 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 those 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 data set in 1995 is chosen so that individual respondents would probably have completed their highest grade by then yet be young enough to be in their earliest jobs. We picked

ages by education level (as of year 2000) as follows. Those who were not high school graduates are observed at ages 20-22 in 1995. Those who were high school graduates but not college graduates are observed at ages 23-25 in 1995. College graduates are 25-27 years old in 1995. About 40 percent of the sample women are college graduates and 33 percent of the men are; 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 $\ln(\text{mean quarterly earnings}) = \text{sum of quarterly earnings in the main job in year } t \text{ divided by the number of quarters in the main job that year}$. The earnings data come from the UI records that are linked to those working in each establishment. Note that there is no information in the UI records, and thus 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:

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

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 eq. (2) we add I , the (three-digit SIC level) industry (there are about 450) of the establishment and (time invariant) three-digit occupation dummies, O , from the 2000 census. We also 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 calculated from the entire sample and not only those in our analysis sample).

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

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 eq. (1). The panel A estimates include the basic individual controls. In panel B, we graph the same for eq. (2), which adds industry and occupation, and in panel C we add $\ln(MEE)$. Because of the large sample sizes, standard errors are small and have not been included in the figures. The horizontal axis for all panels in Figure 1 is year. But for each of the education groups, the axis can be scaled as age where the college graduates begin around age 26, the high school graduates without a college degree start at age 24 and the high school drop outs begin at age 21.

Although the LEHD is a longitudinal data set, individuals can exit the data if they shift their place of employment from one of the 26 PMSAs or if they leave the labor force, among a host of other reasons. Because of the possibility that some individuals can be in the data set for a very short time, we use observations for individuals present at least half the years covered. We therefore use the LEHD as a set of repeated cross sections, although the longitudinal aspect of the establishment information is used. The value of MEE is computed over the entire period, for all firms, and is not time variant.

Data from the LEHD-Census clearly show a widening gender gap in earnings for all education groups, except the lowest (high school dropouts), in the 14 years considered (Figure 1 panel A). 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, between 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.

It is important to note that for the two education groups that have the largest widening of the gender gap with age, the increase occurs mainly in the initial five to seven years. In fact, at the half-way point in age 80 percent of the widening already occurred for the college graduates and 90 percent for high school graduates without a college degree.

The results for eq. (2) in Figure 1 panel B add industry and occupation. The widening is reduced and the levels are increased. The difference between panel A and panel B demonstrate the importance of industry and occupation. For college graduates, the initial difference is reduced from 10.9 to 3.2 log points and the final difference is diminished from 44.6 to 28.5 log points. Therefore, the widening is reduced by 8.4 of the initial 33.7 log points.

The addition of the establishment variable (*MEE*) shifts the earnings gap function for the two higher-level education groups so that the gender earnings gap in Figure 1 panel C widens somewhat less than it did in panel B. For college graduates, the initial gap with *MEE* is about zero (-0.017) but in 14 years the gap expands to -0.205 or by 18.9 log points.

The difference between panel A and panel C shows the importance of establishment given industry and occupation. The remaining increase of the gap in panel C (-0.189) reveals the role of within establishment changes.

Recall that the gender earnings gap (with the basic individual controls) expanded by 33.7 log points during the 14 years. Of that increase, 14.9 log points (44 percent) are due to differential shifts of male and female workers *across* establishments and 18.9 log points (56 percent) are due to the expansion of the gender earnings gap *within* establishments, given industry and occupation. Men move into higher paying establishments relative to women and men disproportionately move to higher paying jobs within establishments. It should be noted that the relative role of within establishment effects is increased to about 75 percent if the comparison is to a widening given industry and occupation.

Because the widening is largest for the college graduate group, we also explore the evolving gender earnings gap for college graduates within various economic sectors. In Figure 2 we graph the evolution of the estimated φ_{jt} from eq. (2) for college graduates in the (two-digit SIC) Finance, Insurance and Real Estate (FIRE) sector and in the Technology (Tech) sector (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.²

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 -0.220 in Technology and -0.234 in FIRE, showing that around 24 to 30 percent of the widening is the result of differential changes in earnings by sex between establishments (given sector, industry and occupation). Most of the expansion (70 to 76 percent) is due to the greater advancement of men than women within establishments.³

As in the aggregate analysis, there is a greater shift of men than women into higher paying establishments in FIRE and Tech. The widening across establishments occurs during the full 14 years in both sectors, although it is less steep in Tech. Note that the *MEE* adjusted lines (dashed) for both FIRE and Tech level off after a widening in the gender earnings gap. That is, differential progression within establishments occurs only in the first six to eight years.

We are also able 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 observe the impact of the mean establishment earnings variable given industry and occupation. We noted before that the widening of the gender earnings gap has been found, in other research, to be greater for women who are or have been ever-married. We find similar results.

Individuals who were ever-married in 2000, will always be ever-married going forward in the data, by definition. Although we do not know when they married, their ages at the start of the analysis were chosen to be beyond the usual school leaving age for each education group. Therefore, the college graduate group would have been 30-32 years old in year 2000. Individuals

² We also explored a similar analysis for the health sector. But because men are disproportionately in occupations (e.g., medical doctor) that require more years of education many are not in the earnings data until they are somewhat older. Therefore, the gender earnings gap is positive (women earn more than men) for about the first four years and then becomes negative and resembles the levels for FIRE.

³ The reason for the difference between this analysis and the aggregate is that we are holding sector, industry and occupation constant in the initial regressions. As noted, the within establishment effect is equally large if the comparison is to the widening with industry and occupation.

who were never-married in 2000 can be married after that date, but we would not know of that change. Thus, some of the difference between the two groups post 2000 can result from an unobserved change in marital status for those never-married by 2000.

For those who were ever-married in 2000 the gender earnings gap for college graduates widened by 38.6 log points from 1995 to 2008, whereas it widened by 21.3 log points during the same period for those who were never-married in 2000. As seen in Figure 3, about 87 of the widening for those who were ever-married in 2000 occurred during the first seven years of the period. The widening to year 2000 was 29.4 logs points for those ever-married by 2000 but just 12.3 for the never-married by 2000. That is, far more of the widening for the never-married group by 2000 occurred after they could transition into the ever-married status. The impact of adding controls for industry, occupation and establishment are similar to those for the aggregate analysis.⁴

These are suggestive results that family responsibilities for women are an important factor in the widening of the gender earnings gap. But we note the limitations of the data concerning marital status and the absence of information on the presence of children.

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 U.S.). 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

⁴ The results by marital status in 2000 for the high school graduates who did not graduate college are similar. The widening is less, as was the case in the aggregate data, but the differences between the ever-married and never-married (by 2000) groups are the same. Interestingly, there is no difference between the two marital status groups for the high school drop-outs for which the gender earnings gap levels are higher at the start but there is almost no widening.

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 in year 2000.

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 given industry and occupation. 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 the age ranges we have chosen and these are the ages when families are formed.

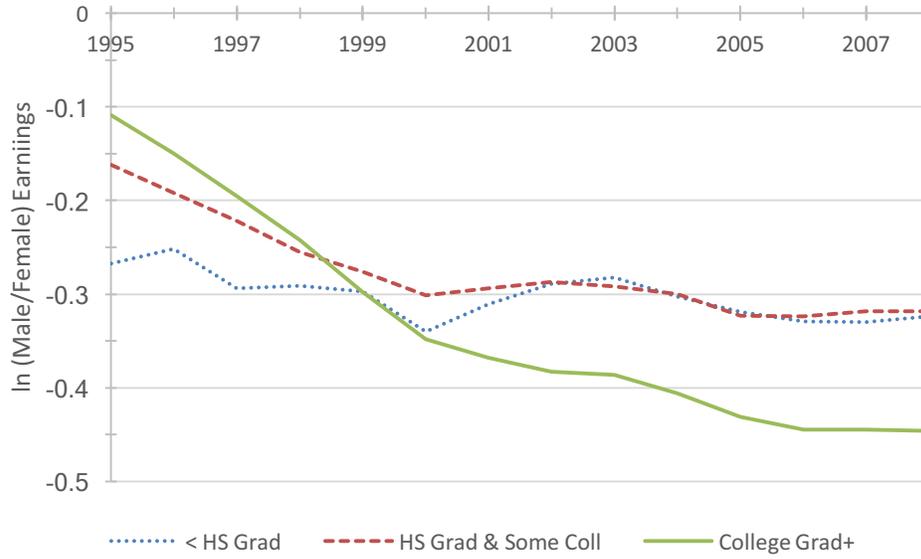
Our bottom line is that the widening is split between men's greater ability or preference to move to higher paying firms and positions and their better facility to advance within firms. Given industry and occupation, far more is due to men's better advancement within firms. Women's greater family responsibilities appear to be largely responsible for both the between and within firm differences. These data cannot tell us what part of the differences are due to women's choice to work in less demanding, and thus lower paying, industries, firms and occupations when they have greater family responsibilities. But they hint to that as an important explanation.

REFERENCES

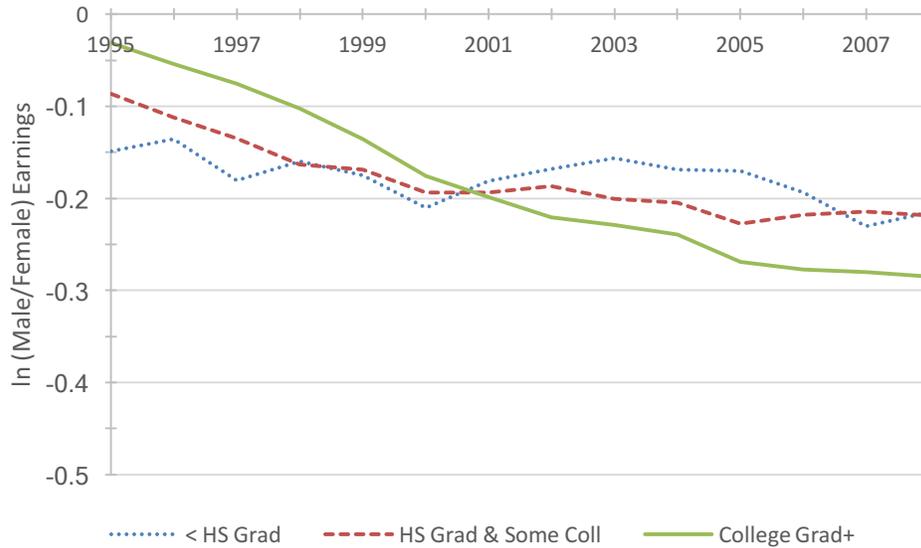
- Barth, Erling, Alex Bryson, James C. Davis, and Richard B. Freeman.** 2016. "It's Where You Work: Increases in Earnings Dispersion across Establishments and Individuals in the U.S." *Journal of Labor Economics* 34 (S2, Part 2): 67-97.
- Barth, Erling, Sari Pekkala Kerr, and Claudia Olivetti.** 2017. "The Dynamics of Gender Wage Differentials: Evidence from Establishment Level Data in the United States." NBER Working Paper No. 23381 (May).
- 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." Working Paper.

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

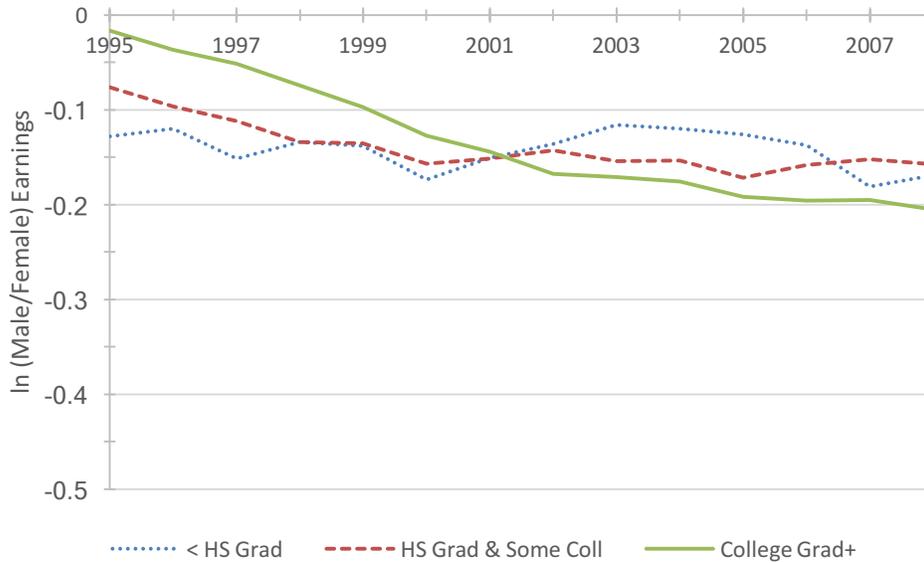
Panel A. Coefficients on female, φ_{jt} , from eq. (1).



Panel B. Coefficients on female, φ_{jt} , from eq. (2) including three-digit SIC industry and occupation (from 2000 Census).

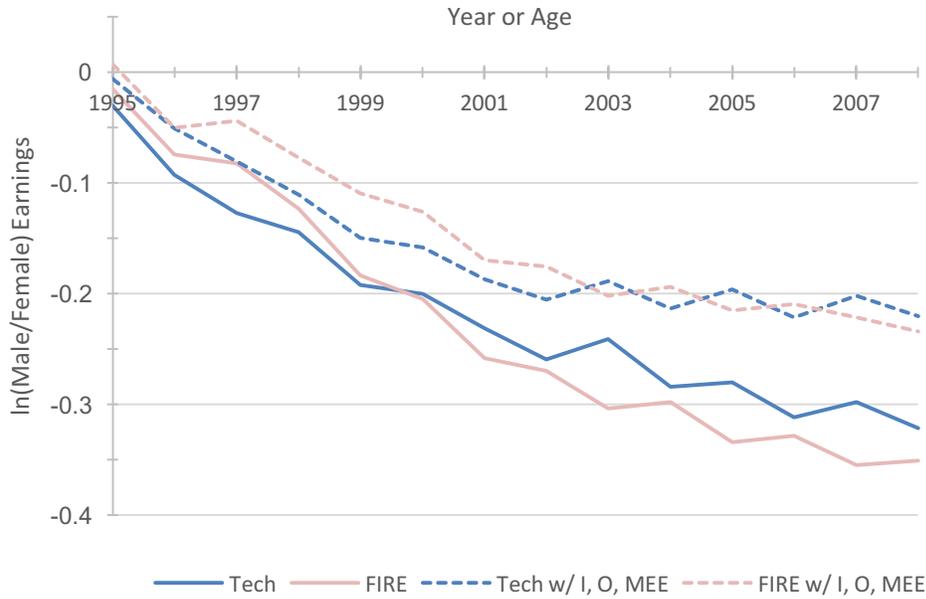


Panel C. Coefficients on female, φ_{jt} from eq. (2) including $\ln MEE$ (mean establishment earnings).



Notes: Panel A gives coefficients on female, φ_{jt} , from eq. (1). Panel B gives coefficients on female, φ_{jt} , from eq. (2) including three-digit SIC industry and occupation (from 2000 Census). Panel C gives coefficients on female, φ_{jt} from eq. (2) and includes $\ln MEE$ (mean establishment earnings). All panels include the basic controls (exact education and cohort, race, state). < HS Grad are 20-22 years old in 1995; HS Grads & Some Coll 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.

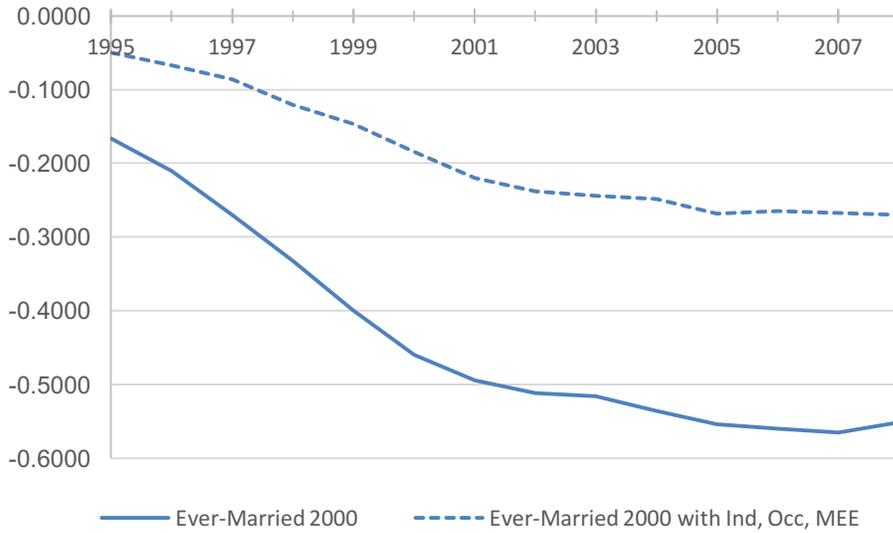
FIGURE 2. GENDER EARNINGS GAP FOR COLLEGE GRADUATES IN TWO SECTORS: 1995-2008 (AGES 25-27 TO 38-40)



Notes: Solid lines give coefficients on female, φ_{jt} , from eq. (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.

FIGURE 3. GENDER EARNINGS GAP FOR COLLEGE GRADUATES BY MARITAL STATUS IN 2000: 1995-2008 (AGES 25-27 TO 38-40)

Panel A: Ever-married in 2000



Panel B: Never-married in 2000



Notes: Solid lines give coefficients on female, φ_{jt} , from eq. (1) with basic controls. Dashed lines give coefficients on female, φ_{jt} , from eq. (2) with basic, three-digit SIC industry, occupation and *MEE* (mean establishment earnings) controls. College graduates are 25-27 years old in 1995 and 38-40 years old in 2008. Estimation uses LEHD-Census 2000. See text.