

NBER WORKING PAPER SERIES

THE DYNAMICS OF GENDER EARNINGS DIFFERENTIALS:
EVIDENCE FROM ESTABLISHMENT DATA

Erling Barth
Sari Pekkala Kerr
Claudia Olivetti

Working Paper 23381
<http://www.nber.org/papers/w23381>

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
May 2017, Revised July 2019

We thank Ilaria D’Angelis and Valeria Ferraro for research assistance. We received funding from the Research Council of Norway (#236770 and #179552 (Barth)). Barth thanks the Labor and Worklife Program at Harvard for hospitality and support. We are grateful to participants of SOLE 2017 in Raleigh NC for helpful comments on an earlier draft of the paper. The paper has been updated based on anonymous reviewers’ comments, but no additional data work has been done since the expiration of the Census Bureau data project. Conditional on receiving an invitation to “revise and resubmit” the manuscript to a journal, the project will be reopened allowing additional data work. 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 or the National Bureau of Economic Research. This paper was screened to ensure that no confidential data are revealed.

NBER working papers are circulated for discussion and comment purposes. They have not been peer-reviewed or been subject to the review by the NBER Board of Directors that accompanies official NBER publications.

© 2017 by Erling Barth, Sari Pekkala Kerr, and Claudia Olivetti. All rights reserved. Short sections of text, not to exceed two paragraphs, may be quoted without explicit permission provided that full credit, including © notice, is given to the source.

The Dynamics of Gender Earnings Differentials: Evidence from Establishment Data
Erling Barth, Sari Pekkala Kerr, and Claudia Olivetti
NBER Working Paper No. 23381
May 2017, Revised July 2019
JEL No. J16,J31

ABSTRACT

Despite dramatic workforce gains by women in recent decades, a substantial gender earnings gap persists and widens over the course of men's and women's careers. Since there are earnings differences across establishments, a key question is the extent to which the widening of the gender pay gap over time arises from differences in career advances within the same establishment versus differential gains from job-to-job moves across establishments. Using a unique match between the 2000 Decennial Census of the United States and the Longitudinal Employer Household Dynamics (LEHD) data, we find that both channels are important and affect workers differently by education. For the college-educated the increasing gap is for the most part due to differential earnings growth within establishment. The between-establishment component explains only 27 percent of the widening of the total gender pay gap for this group. For workers without college degree, the establishment component is the main driver of the, relatively small, widening of the gender earnings gap. For both education groups, marriage plays a crucial role in the establishment component of the increasing earnings gap.

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

Claudia Olivetti
Department of Economics
Dartmouth College
6106 Rockefeller Hall
Hanover, NH 03755
and NBER
claudia.olivetti@dartmouth.edu

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

Women have made remarkable progress in the labor market throughout the past century, resulting in clear convergence in human capital investment, employment prospects and outcomes relative to men (Blau and Kahn, 2016). However, remaining gender differences in pay still persist and increase over the working life, especially so for college graduates, even within narrowly defined occupations and even when controls are added for hours worked per week and weeks per year (see Goldin, 2014, and Bertrand, Goldin, and Katz, 2010, on MBAs).

At the same time, recent studies have shown that there are large earnings differentials across firms and establishments, and that sorting of workers into high- and low-paying establishments contributes to earnings inequality in the US¹ and other countries.² Since men are more likely to work in high-paying firms and appear to capture a larger part of the establishment premium than women, these establishment earnings differentials tend to add to the gender pay gap (Card, Cardoso, and Kline, 2016). In this paper we use a unique match between the 2000 Decennial Census of the United States and the Longitudinal Employer Household Dynamics (LEHD) data to analyze how much of the increase in the gender earnings gap over the life cycle comes from shifts in the sorting of men and women across high- and low-pay establishments over the early part of their working life and how much is due to differential earnings growth within establishment.

The novelty of our analysis with respect to other gender studies based on matched employer-employee data is its focus on the United States and on the *age dynamics* of the gender pay gap, rather than its cross-sectional average. In addition, the unique data match to the 2000 Census of the United States allows us to study how the widening of the wage gap varies by education and marital status, something that typically cannot be done with matched data, at least in the United States, but can help gain a better understanding of the potential mechanisms. The drawback of our data, as we will discuss later, is that while the matched data cover several years of an individual work history (from 1995 to 2008), key demographic variables are only observed at one point in time (in 2000). In addition, information in the LEHD is only available for the subset of states that provided UI data at the time of writing this paper. While this does not appear to impact the generalizability of our results we return to this point later.

¹ See Abowd, Creedy, and Kramarz (2002) on the large earnings variation across establishments in the US, even for the same worker. Barth, Bryson, Davis, and Freeman (2016) and Song, Price, Guvenen, Bloom, and von Wachter (2019) study the contribution of establishment and firm pay to the increase in the earnings dispersion.

² See Card, Heining, and Kline (2013) for Germany and Félix and Portugal (2016) for Portugal.

Figure I shows the predicted earnings-age profiles by gender and education based on our large panel of workers observed over time and across jobs. We group workers in two education categories: those who are not college graduates and (four-year) college graduates. Predicted earnings by age are measured relative to a 25-year-old woman without a college degree, whose earnings level is set to unity. There are striking differences in earnings profiles by education. The comparison of black and grey lines for each group shows that the gender earnings gap widens with age for both education groups, but the widening is far greater among college graduates, mainly due to the disproportionate increase in college educated men's earnings.³ While the gender earnings gap (defined as the ratio of male to female earnings) is relatively small at age 25 and does not differ much across education groups, by age 45 the gap is considerably larger, especially for the college-educated. Specifically, the gender earnings gap for the non-college-educated increases by 32 percentage points (from 15 percent to 47.4 percent) from age 25 to age 45, while for the college-educated it increases by 62 percentage points (from 13 percent to 75 percent). This finding is broadly consistent with earlier comprehensive studies using synthetic cohorts.⁴

The widening in the gender earnings gap as men and women age may be due to the combination of two different processes. The first - the pay gap between male and female workers at the same firm or establishment – arises mostly from promotions and raises that workers at a given establishment accrue over time. The second – the pay gap due to workers sorting into high- versus low-paying employers – is driven by the presence of different wage levels across firms and gender-based variation in the likelihood that a worker will both change jobs and receive a higher salary as a result of that job change.

The two processes arise from very different mechanisms. Within-employer career paths arise from different wage structures within firms, and the overall design of organizations and their job ladders. Gender differences within firms may ensue because of competition for higher positions within the hierarchy (Lazear and Rosen, 1990). The prevalence of convex reward profiles in working time, especially in some high-paying occupations, also contributes to the widening of the gender pay gap (Goldin, 2014).

³ Regressions are based on 15 million observations; see the summary table in the Appendix. The standard errors of the estimates are thus so small that we have chosen not to draw the confidence intervals around the curves.

⁴ Goldin (2014) documents that for cohorts born around 1960 and 1970 (the bulk of our sample) the gender earnings gap for college graduates increases by approximately 48 percentage points between age 25-29 and 40-44. Less than 40 percent of this increase can be explained by hours and weeks controls. Goldin's figures are for full-time (35+ hours), full year (40+ weeks) workers while our figure includes all workers (and this can explain the larger increase based on our data). Manning and Swaffield (2008) find similar results for the UK.

Between-employer career paths arise from different wage levels across firms⁵, that may be a result of frictions in the labor market, monopsonistic behavior,⁶ and/or different forms of rent sharing. In this case, gender pay differences across firms may arise if women have lower ability to extract rent (Card, Cardoso and Kline, 2016), and thus face a flatter earnings profile across firms than men, or if women have a less elastic job mobility pattern with respect to wages, so that employers may engage in monopsonistic discrimination (Barth and Dale-Olsen, 2010).⁷

Women face both between- and within-establishment gaps. Due to family and caretaking obligations, women may be less able to put in the long hours required to obtain a big promotion, or to invest in the networking and job search activities that facilitate financially advantageous job changes. These effects may be compounded if employers believe women are less likely to remain in the workforce or have higher cost of effort (Gayle and Golan, 2012; Albanesi and Olivetti, 2009) or if women are less likely to seek promotions and raises within and across firms in anticipation of needing more time or flexibility (Bronson and Thoursie, 2018) or because of family location decisions in which the career of the primary earner, usually the husbands, takes precedence (Taylor, 2007).

In this paper we quantify the importance of career movements within and across establishments using a full decomposition analysis of the establishment and individual fixed effects. Specifically, we identify directly the within-job difference in earnings profile over time between men and women. This difference arises from the combined effects of differences in wage raises and promotions within the same employer.⁸ While we cannot distinguish between internal promotions and earnings growth within job category in our data, we are able to distinguish between differential seniority profiles, which are experienced within the current job, but lost when changing employer⁹, and differential development in the within-establishment

⁵ See Groshen (1988) for an early contribution.

⁶ See Manning, 2003, Green, Machine and Manning, 1996.

⁷ Another frictional mechanism emphasized in this literature has to do with employers' disutility of hiring women (Black, 1995, Bowlus and Eckstein, 2002, Flabbi, 2010). See Barth, Bratsberg, and Raaum (2012) on the importance of establishment affiliation for the development of the immigrant-native earnings gap over time.

⁸ Booth, Francesconi, and Frank (2005) use the British Household Panel Survey, and find that women are promoted at about the same rate as men, but receive smaller wage gains with the promotions. Blau and De Varo (2007) find that among new hires in in a sample of US establishments women have lower probabilities of promotion than men, but that wage gains associated with promotions do not differ much.

⁹ Due to firm-specific versus general human capital investment (Becker, 1975) or to delayed payments as a solution to agency problems (Lazear, 1981).

earnings distribution that is not lost when changing job, e.g. as one changes from a top position in one firm to a top position in another firm.

We find that among workers without a college degree, there is only a small male advantage in the within-establishment earnings growth over the first 10 to 15 years of working life. However, by age 45 women catch up with men in terms of within-establishment earnings growth. On the other hand, for college educated workers there is a substantial difference in within-establishment earnings growth by gender: men's earnings growth is 33 log points higher than women's from age 25 to 40. Unlike the non-college educated, college women do not catch up. By the time they reach age 45 the difference is still 30 log points. A fraction of this difference is due to higher returns to seniority among men, 1 log point per year of seniority. The remaining part is due to earnings growth within establishment that is carried over in the transition to new jobs in other establishments.

The most novel results of our analysis concern the development of the establishment earnings premium over time for a given individual. For employees without a college degree, the widening of the difference in the establishment component of the earnings premium between age 25 and 45 adds 5.4 log points to the overall increasing gender gap in pay. This is the main source of the gender differential in earnings growth for the less educated. Among college-educated workers, the gender gap across establishments widens by 11 log points from age 25 to 45. Since college-educated workers experience a considerable widening of the within-establishment earnings gap, the expansion in the establishment component can explain only 27 percent of the total widening of the gender wage gap for this group.

Notably this relative more equal performance of women in the no-college degree group mainly reflects lack of growth in less-educated men's earnings over this period. Climbing the career ladder more quickly within a firm plays no role for employees in this group. The small, early career divergence between men and women's earnings among high school graduates is fully explained by men moving to better paying firms in the initial years of their careers.

We find that in the "between firms" gap marriage plays a crucial role. Indeed, married women's earning power seems to benefit very little from changing jobs. The evolution of the establishment pay premium for the never-married is very similar to that of married men. Descriptive statistics based on the publicly available 2000 Census of the United States (PUMS

data) suggest that most of the loss in earnings growth for married women, relative to married men, occurs concurrently with the arrival of children.

While we currently do not have the ability to investigate how the results of our decomposition vary by sector (or by occupation), in a companion paper (Goldin et al. 2017)¹⁰ we show that while gender differentials in sorting by occupation and industry are important, they can jointly explain only about a third of the widening of the gender earnings gap that occurs over time for these cohorts. That is, sorting into high and low paying firms matters even within narrowly defined industries and occupation (as of 2000). We also find that, for the college-educated, the magnitude and growth of the gender earnings gap look quite distinct from one sector to the next. By the peak of a person's career, the largest gender gaps can be found in the health, legal, and financial sectors (including insurance and real estate). Conversely, the widening of the gender earnings gap with age remains more modest in the tech sector.

This paper is part of a rich and currently burgeoning literature. Blau (1977) is the first to analyze the role of inter-firm wage differentials in explaining the gender wage gap, a finding further confirmed in a subsequent study by Groshen (1991). Early contributions using matched employer-employee data include the study by Barth and Mastekaasa (1996) and Bayard, Hellerstein, Neumark and Troske (2003) who show that a gender earnings differential persists even after controlling for gender differences in human capital and sex segregation within occupation, industry and establishment based on early versions of the matched data for Norway and the United States, respectively.¹¹ More recently, Card, Cardoso and Kline (2016) show that firm-specific pay premiums explain just over one-fifth of the average gender earnings gap in Portugal and interpret it in light of differential bargaining power between male and female workers.¹² Sorkin (2017) does use the LEHD data to address gender differences in sorting into high-wage firms and industries, in the cross-section. Based on a search model he argues that women tend to select into firms with better non-pay characteristics. Other authors have used matched employer/employee data to explore the role of gender segregation and pay structure

¹⁰ Goldin et al. (2017) use the data as a set of repeated cross sections of individuals rather than a panel. The longitudinal aspect of the establishment information is only used to compute mean earnings at the establishment level over the entire period.

¹¹ The paper by Bayard et al. (2003) uses an earlier version of our data that matched the 1990 Sample Edited Detail File (consisting of all household responses to the 1990 Decennial Census long form) to establishment records in the 1990 Standard Statistical Establishment List.

¹² In particular, they show that bargaining and sorting based on measured productivity account for about 80% of the overall impact of firm-specific pay premiums on the gender earnings gap.

for explaining the cross-sectional wage gap in Spain (Amuedo-Dorantes and De La Rica, 2006), Denmark (Datta Gupta and Rothstein, 2005), Germany (Heinze and Wolf, 2009), Finland (Korkeamäki and Kyryä, 2006) and a for cross-section of nine European countries (Simon, 2012). All these papers focus on the cross-sectional gender pay gap. The novelty in our paper is that we focus on the widening gender pay gap by age in the US, unlike earlier studies with a similar focus.

The most closely related study to ours is Bronson and Thoursie (2018) who analyze gender differences in lifecycle earnings growth using administrative data from Sweden. They find that about 70 percent of the growth in gender earnings gap can be associated with within-firm earnings growth, while sorting, work experience, tenure and field of education play much smaller roles. Interestingly, despite Sweden's generous policies towards families with young children, Bronson and Thoursie find a significant "motherhood penalty" that affects women's earnings and promotions. It is notable that the current study takes place in a very different setting, with the United States having few concessions towards new mothers in terms of paid leave, subsidized child care, and other support systems.

The distinction between job separations into non-employment versus job-to-job separations is important for the analysis of gender differences in earnings growth and job transitions. While, as discussed above, separations into non-employment may have negative effects on between-establishment earnings growth, separations to other jobs may have the effect of improving earnings. Earlier studies of separation rates find that gender differences in the probability of leaving a job tend to disappear once we control for observable characteristics (Blau and Kahn, 1981) and attachment to the labor force (Light and Ureta, 1990, 1995). Royalty (1998) shows the existence of gender differentials in the destination state. Women are more likely to leave a job for non-employment, while men are more likely to move from one job to the next (see also Manning 2003). There is also an earlier literature emphasizing differences in turnover rates (based on longitudinal data) as an important source of gender differentials in earnings growth for young workers. For example, Loprest (1992) shows that women gain less, in terms of wage growth, from switching jobs. Bowlus (1997) finds significant gender differences in quit rates "for personal reasons" that account for 20%–30% of the gender wage differential. More recently, Del Bono and Vuri (2008) identify gender difference in the returns to job mobility as the main source of gender earnings growth differentials on a sample of private sector employees in Italy. Hirsch, and Schnabel (2012) show that lower female wage-elasticity

and gender differences in the transition probability are both important for explaining the gender wage gap in German employer/employee matched data.

The remainder of the paper is organized as follows. Section I describes our data set and key variables. Section II discusses our estimation strategy. Finally, the main findings of our analysis are presented in Section III.

I. DATA AND DEFINITION OF KEY VARIABLES

Our analysis relies on a unique combination of the Longitudinal Employer-Household Dynamics (LEHD) database and the 2000 Decennial Census of Population (one in six long form). Both datasets are confidential and housed by the U.S. Census Bureau in the Research Data Centers (RDC). As the current combination of these two restricted access data sets has barely ever been used in previous empirical literature, this section will provide a detailed summary of the construction of our data platform.

The LEHD is based on quarterly, worker-level, filings by all private-sector U.S. firms in the context of the administration of state unemployment insurance (UI) benefit programs.¹³ The data identify all employees of an establishment and their quarterly compensation on a month-to-month basis. UI earnings include wages, salary and taxable bonuses and are not top-coded.¹⁴ The state UI system covers about 95% of private sector employment. Thus our analysis is fully representative of private firms within the geographical areas we study (see Hyatt, McEntarfer, McKinney, Tibbets, and Walton, 2014; Stevens, 2007). The LEHD is longitudinally linked at both the firm and employee levels, making it possible to analyze how firm employment and employee earnings evolve over time, within and across all establishments. The LEHD vintage used in this paper includes 23 states with varying initial dates of coverage, starting from 1991, and runs through 2008.

To manage the enormous number of person and firm fixed effects required to estimate our models, we focus on the largest PMSAs in the U.S. in terms of population as of 1991. Of the 50 largest U.S. PMSAs, 26 were located in 18 of the 23 LEHD covered states available to

¹³ LEHD also covers most state and local government employees, with the exception of elected officials, members of a legislative body or judiciary, and some emergency employees, Federal government employment is not covered.

¹⁴ See Abowd, Stephens, Vilhuber, Andersson, McKinney, Roemer, and Woodcock (2002) for an in-depth discussion of the benefits and shortcomings of these data.

us. Specifically, these PMSAs are located in CA, CO, GA, FL, HI, ID, IL, MD, NC, NJ, NM, OR, RI, SC, TX, UT, VA and WA. We further reduce our analysis sample by using annual data from 1995 to 2008 and selecting workers who worked more than two quarters per year and earned at least \$2,000 per quarter, on average, during the year. The last restriction removes from the sample very short and sporadic employment relationships as well as short-term contract employment arrangements, and directs our study towards the more permanent work arrangements. All dollar values are inflated to 2008 values using the all-urban-consumers-CPI published by the BLS. Finally, we focus on individuals in their prime working age. That is, for each individual, we only use observations when the person is aged 25-45.

The LEHD records limited information about workers in the individual characteristics file (ICF). This includes age, gender, race, place of birth, and citizenship status. Through the employment history files (EHF), we can also discern their earnings and job-by-job employment histories. Moreover, using the unique person identifiers (PIKs), we are able to match people in the LEHD to the individual-level records contained in the long-form responses of the 2000 Decennial Census of Population.¹⁵ The long-form was given to a random sample of 1-in-6 households and is nationally representative. This process allows us to match almost exactly 1-in-6 of our LEHD workers with added Census details from the Person File. The LEHD-Census match thus includes more detailed and comprehensive information about each individual in our sample (e.g., level of education, occupation, marital status, class of worker, etc.) and their respective families (e.g., family composition, detailed characteristics of their spouse, and household income by source). In the current study we mainly extract individual-level characteristics, that is, educational attainment, marital status and race. It is worth emphasizing that while the LEHD longitudinally follows the same individuals over time across jobs, the 2000 Census is obviously just a snapshot. To the extent that the person's marital status, education or occupation changes, either before or after 2000, we will not be able to capture that. The LEHD sample linked to the 2000 Census is over 12 million annual observations, covering about 3.3 million individual persons.

A possible concern with the LEHD data is that it tracks quarterly employment and earnings, but does not contain information on hours worked or hourly wage.¹⁶ Therefore, in

¹⁵ The Census Bureau has created the unique person identifiers (PIKs) based on Social Security Numbers (SSNs). These PIKs allow the linking of individuals across demographic surveys, censuses and administrative records.

¹⁶ Some states, most notably Washington, collect data on hours worked for hourly wage employees. We did not have access to those data or a permission to link them to the LEHD at the time of writing this paper. See Lachowska, Mas, and Woodbury (2017) for an example of a paper using the WA linked employer – employee data with hours.

theory, any age or cohort trend in the gender earnings gap could be driven by a divergence of hours worked. To explore this issue in detail, we analyze repeated cross-sections of the American Community Survey (ACS) and the Current Population Survey (CPS). With those data we can derive different earnings concepts, including annual, quarterly and weekly earnings, as well as hourly wages.

In Figure II we use CPS data to show age profiles by cohort, education and gender for college graduates measured as quarterly earnings and hourly earnings. Appendix 2 provides details of the data and estimation. It does not appear that differences in hours worked over the life cycle affect the relative shape of the age profiles across genders and cohorts. In the appendix, we show similar graphs for non-college workers, and again conclude that the age profiles are very similar irrespective of earnings measures.

In appendix 3 we also show estimated age patterns of the gender earnings gap, measured either in terms of annual earnings, weekly earnings and hourly wages using the American Community Survey (ACS). The gender earnings gap is somewhat larger when measured in terms of annual earnings than when measured in terms of hourly wages. However, the age patterns of the earnings gap are very similar regardless of the income measure (annual or weekly).¹⁷ Controlling for hours in each equation gives the expected result: the level of the gender gap varies whether it is measured in annual or weekly earnings, but the age pattern of the gap itself is unaffected. Thus although hours worked can explain part of the gender differences in the level of earnings, they do not impact the age patterns. The appendix provides a short explanation and graphical findings of this analysis based on ACS in 2001-2007.

We conclude that, as we are interested in estimating the age dynamics of the gender earnings differential, we can proceed with the LEHD-based analysis with more confidence that it is not the gender divergence in hours worked that is driving the observed age patterns.

Another possible caveat with our data is attrition from people moving out of the 26 PMSAs. This would pose a problem for us if the movements out of the PMSA were differentially selected across genders. Since our key results are obtained conditional on fixed individual and establishment effects, heterogeneity in terms of individual or establishment earnings levels is controlled for in the analysis. However, attrition bias could still affect our

¹⁷ The usual caveat regarding self-reported hours of work applies of course, see e.g. Frazis and Stewart, 2004; Abraham, Spletzer, and Stewart, 1998; Robinson and Bostrom, 1994. Surveys on hours worked may suffer from various response biases, be non-representative for salaried workers, and/or have reference weeks that are not representative of the typical monthly or annual average.

gender comparisons if there is differential selection in residual earnings growth between men and women. Although this concern cannot be confronted directly based on our data, we note that the bias could go either way. We show below that our results are driven mainly by married individuals. If married women in our sample of PMSAs are disproportionately tied stayers, this could lead to finding lower wage growth for women. But low (residual) wage growth women may as likely be tied movers, women who follow their husband as they take better job opportunities, which would bias our findings in the opposite direction. We also use the ACS and CPS to evaluate the representativeness of our PMSA sample relative to the entire United States. We find that the age patterns in the gender earnings differentials are essentially unchanged as we limit the data to the sample LEHD states instead of the entire country.¹⁸

A few final details about the Census RDC data are worth noting. First, all observation counts in the paper are disguised and rounded to the nearest 100 according to Census Bureau disclosure restrictions. Second, we generally use the establishment ID (based on the State Tax ID and the establishment number) to identify work establishments and to track them over time. The LEHD also includes firm identifiers, but unlike establishment IDs those are not fully consistent over time within a given firm. For example, if another firm acquires an establishment, the firm ID will change, even though the workers within that establishment would all continue their employment with the company. Due to the sheer size and scope of these data it would be practically impossible to try to follow each of the firms over time while taking into account all merger activities and other corporate restructuring. Moreover, we believe that the establishment-level tracking is more relevant for the current purpose because the type of job-to-job moves that is the focus of this paper may involve changes between establishments within a firm as well.

The dependent variable in our models is the natural logarithm of earnings, where earnings are measured as the average quarterly earnings during the year (over the quarters that the person was working). Another key variable is the establishment size, where the size corresponds to the LEHD reported number of employees minus the sample mean. We also include a squared-term in age, where the age refers to the person's age during the year of observation minus 35 (which is the mid-point of the age range in our sample). Since the regression specification includes establishment and person fixed effects, most of the other time-

¹⁸ These additional analyses are not included here due to space considerations. Results are available from the authors by request.

invariant person and establishment characteristics are absorbed by those fixed effects. In what follows, we also normalize establishment size to be zero at the sample mean.

II. METHODOLOGY

A. Definitions: Age-Earnings profiles and the Gender Earnings Gap

To derive age-earnings profiles of individuals within and across establishments, we utilize a simple log earnings model in the style of Abowd, Kramarz, and Margolis (1999) - AKM hereafter¹⁹ - with individual fixed effects, establishment fixed effects and time varying covariates. We also allow for a job-specific match effect, where a job is defined as a unique match between an individual and an establishment. The earnings equation is given by:

$$(1) \quad \ln w_{it} = \beta_0 + \underbrace{\beta_1^g Age_{it} + \beta_2^g Age_{it}^2 + \alpha_i}_{\text{individual component}} + \underbrace{\beta_3^g \ln Size_{j(i,t)t} + \varphi_{j(i,t)}^g}_{\text{establishment component: } \chi} + \underbrace{\xi_{ij(i,t)}}_{\text{match component}} + \gamma_t + \varepsilon_{it}$$

where g denotes gender, Age is the age of individual i at time t , α_i is the individual fixed effect, $\ln Size$ is a measure of the number of employees at the establishment where individual i works at time t , $\varphi_{j(i,t)}^g$ is the establishment fixed effect²⁰ and $\xi_{ij(i,t)}$ is an idiosyncratic match effect assumed to be orthogonal to both the individual and the establishment fixed effect, but potentially correlated with time-varying individual and establishment characteristics (e.g. age and establishment size) as well as calendar time. The equation also includes a year fixed effect, γ_t , and an error term, ε_{it} , assumed to be orthogonal to all the other variables.

The “individual component” of the earnings equation describes the expected *within-job age earnings* profile for a worker i , holding constant characteristics (e.g. cohort) over time and across employers, and conditional on establishment and match-specific fixed effects.

¹⁹ See Abowd et al (1999).

²⁰ We assume fixed establishment and match effects over the full observation period. This may be a restrictive assumption, in particular since Barth et al (2016) find an increase in the earnings distribution across establishments over this period. Upward mobility across establishments may lead to an upward bias in the estimated age-earnings profile within establishments, as workers may be more likely to be observed above the establishment mean later in the career. There are, however, unresolved trade-offs in the estimation of time varying two-way fixed effects in this context, and we leave this issue to future research.

The “establishment component” of earnings includes the establishment fixed effect and an earnings premium associated with the establishment size: $\chi_{j(i,t)t} = \beta^g_3 \ln Size_{j(i,t)t} + \varphi^g_{j(i,t)}$. Note that this term changes over time for each individual both because of changes in the establishment size over time and because $\varphi^g_{j(i,t)}$ changes as the worker moves across establishments. To capture the age earnings dynamics of this component, we define the auxiliary regression of the establishment component for individual i of gender g as a function of age and age squared as follows:

$$(2) \quad \chi_{j(i,t)t} = b^g_0 + b^g_1 Age_{it} + b^g_2 Age^2_{it} + u_{it}$$

Equation (2) defines the expected *age profile of the establishment component of earnings* for individual i of gender g . That is, it describes a worker’s life cycle earnings evolution that comes from changes in establishment size over time and from moving across establishments of different size and with different establishment fixed effects.

For ease of exposition, we use hereafter the shorthand notation of j instead of $j(i,t)$ to indicate establishment j where worker i is employed at time t . Therefore, the establishment fixed effect will be denoted by φ^g_j instead of $\varphi^g_{j(i,t)}$ and so on.

Given (1) and (2), we define the *age-specific gender earnings gap* as the difference, for given age, between the expected earnings of a woman and a man living in the same PMSA and with the same education:

$$(3) \quad \Gamma_{Age} = E(\alpha_i|f) - E(\alpha_i|m) + (\beta^f_1 - \beta^m_1)Age + (\beta^f_2 - \beta^m_2)Age^2 + E(\chi^f|Age) - E(\chi^m|Age).$$

where, $E(\alpha_i|g)$, $g = f, m$ is the average individual fixed effect for gender g , and $E(\chi^g|Age)$ is the expected establishment earnings component for a person of gender g , as defined above. Using the auxiliary regression (2) that describes the evolution of the establishment component by age, we decompose the change in the age specific gender earnings gap between any two ages, for instance between age 35 and age 25, as:

$$(4) \quad \Delta \Gamma_{Age} = \underbrace{(\beta^f_1 - \beta^m_1)\Delta Age + (\beta^f_2 - \beta^m_2)\Delta Age^2}_{\text{change in the individual component}} + \underbrace{(b^f_1 - b^m_1)\Delta Age + (b^f_2 - b^m_2)\Delta Age^2}_{\text{change in the establishment component}}$$

where Δ denotes the difference in the variables as measured between the two ages. Equation (4) provides a decomposition of the change in the earnings gap between men and women for any age during their career.²¹ The first bracket of equation (4) reflects the change in the *individual component* of the gender earnings gap by age, arising from different earnings growth over time among workers who remain in the same establishment, conditional on establishment size. We explain below how both $(\beta^f_1 - \beta^m_1)$ and $(\beta^f_2 - \beta^m_2)$ may be identified from the estimation of (1) using fixed job effects to absorb individual, establishment, and match specific fixed effects, and adding interaction terms with a gender dummy when appropriate.

The second bracket of equation (4) reflects the changes in the *establishment component of the* gender earnings gap, arising from different sorting of men and women across establishments by age, and from changes in establishment size over time. We discuss below how both $(b^f_1 - b^m_1)$ and $(b^f_2 - b^m_2)$ may be identified from a second stage regression of the predicted establishment component of equation (1) on age and age square as outlined by equation (2).

Notably, the widening of the gender earnings gap estimated using a simple OLS regression, as illustrated in figure I, encompasses both the individual and the establishment components, as well as any composition effects stemming from different individuals entering or leaving the sample over time. Since composition effects in terms of individual characteristics will be taken care of by the inclusion of individual fixed effects in all our models, we focus our discussion below on the estimation of the individual and the establishment components.

B. Estimation: Identifying the Widening Earnings Gap by Age

We include fixed individual effects in our earnings model, and thus control for cohort fixed effects in the most flexible way. However, it is well known that it is not possible to separately identify linear cohort, time, and age effects in any panel model as they will be collinear. For each individual (or year-cohort), the variables time and age are collinear in the data, and it is not possible to distinguish between them. We may identify the coefficients of the second order term for age, β^g_2 , but we cannot identify the linear term, β^g_1 directly.

²¹ Note that the change in the earnings gap over time does not rely on differences in the expected individual fixed effects for each gender or any fixed difference in the establishment components as reflected by the constant term in equation (2), as they are constant over time and thus differenced out.

However, as our focus is on the change in the gender earnings gap by age, assuming common time effects γ_t across genders allows us to identify the *difference* in the linear terms, $(\beta^f_1 - \beta^m_1)$. According to (4) this is what we need in order to obtain an estimate of the *change of the individual component of the earnings gap*:

Step 1:

Summing the individual, establishment, and match fixed effects into a job fixed effect allows us to rewrite equation (1) as

$$(1') \quad \ln w_{it} = \beta_0 + \beta^g_1 Age_{it} + \beta^g_2 Age^2_{it} + \beta^g_3 \ln Size_{jt} + \psi^g_{ij} + \gamma_t + \varepsilon_{it}$$

where $\psi^g_{ij} = \alpha_i + \varphi^g_j + \xi_{ij}$ is the job-specific fixed effect. We estimate (1') by pooling the data for men and women and run a model where all the time varying covariates are interacted with a female dummy variable F, and we may write all terms as for instance for age: $\beta^m_1 Age_{it} + (\beta^f_1 - \beta^m_1) F Age_{it}$.

Since age is perfectly correlated with calendar time for each individual, we exclude the linear part of Age (the first term with the coefficient β^m_1) from the equation²². We may however, retain the interaction term between gender and Age, which is not collinear with calendar time since it is the age for women but zero for men.

Excluding the term for age, $\beta^m_1 Age_{it}$, from the estimation has the consequence that the linear age effect is absorbed partly by the year dummies and partly by the individual fixed effects. The new year dummies are given by $\tilde{\gamma}_t = \gamma_t + \beta^m_1 (Age_{it} - Age_{i0}) = \gamma_t + \beta^m_1 t$, where Age_{i0} is age at the base year (start of the panel) and t is the number of years since the start of the panel. $\tilde{\gamma}_t$ is thus fixed across all observations for each calendar year. The new individual fixed effects are given by: $\tilde{\alpha}_i = \beta^m_1 Age_{i0} + \alpha_i$, which is fixed across all observations for each individual. We easily see that we have absorbed the linear term for age as $\tilde{\alpha}_i + \tilde{\gamma}_t = \alpha_i + \gamma_t + \beta^m_1 Age_{it}$. The model we estimate in step 1 is thus:

²² This point is not consistently acknowledged in the literature using fixed effects for individuals and establishments, and many researchers tend to include a linear age term in models, where it cannot be separately identified, simply as a matter of habit. This practice may not be of much consequence if age and time are just control variables, but it is crucial for the discussion of earnings dynamics over time. Moreover, as it will be clear below, the way in which we treat the linear age term is crucial for the interpretation of the individual fixed effects.

$$(1'') \quad \ln w_{it} = \tilde{\beta}_0 + (\beta^f_1 - \beta^m_1) FAge_{it} + \beta^m_2 Age^2_{it} + (\beta^f_2 - \beta^m_2) FAge^2_{it} + \beta^m_3 \ln Size_{jt} + (\beta^f_3 - \beta^m_3) F \ln Size_{jt} + \tilde{\psi}^G_{ji} + \tilde{\gamma}_t + \varepsilon_{it},$$

where $\tilde{\psi}^G_{ji} = \tilde{\alpha}_i + \varphi^g_j + \xi_{ij}$ is the job fixed effect. $\tilde{\beta}_0$ will be determined by the normalization that the job fixed effects are set to zero for the average person in the estimation sample.

The estimation of equation (1'') in step 1 provides us with estimates of all parameters needed to calculate the change in the individual component of the expected gender gap by age, as defined in equation (4).

Step 2:

In the second step we use the parameters from step one to obtain an estimate of the establishment component, χ^g_{jt} . To begin the process, we retrieve the fixed job effects, $\tilde{\psi}^G_{ji}$, from equation (1''), and decompose them into an individual effect and an establishment effect by the following AKM-type of decomposition:

$$(5) \quad \tilde{\psi}^G_{ji} = \tilde{\alpha}_i + \varphi^g_j + u_{it},$$

which is estimated on the full sample of observations by gender within the PMSA²³. Combining the term for the logarithm of size from the step 1 with the estimated establishment fixed effects provides us with an estimate of the time varying establishment component of the earnings premium $\chi^g_{jt} = \beta^g_3 \ln Size_{jt} + \varphi^g_j$ for each observation in the data.

Recall that the establishment component may change over time for each individual both as firm size changes in a given establishment, conditional on the worker staying, and as the individual changes jobs into new establishments. The establishment component of the gender pay gap may thus change both from differential development of expected firm size and from differential sorting of men and women into different establishments. To obtain estimates of b^g_s ($s=1,2$) describing the dynamics of the establishment component of earnings by age, we run the auxiliary regression (2) separately by gender and education.

For both steps of the analysis, we estimate the parameters of this model by PMSA and report weighted average estimates over all PMSAs. All parameters are estimated separately

23 We use the REG2HDFE procedure in Stata to do the decomposition. See Guimaraes and Portugal (2009).

for each education group, gender and PMSA, with the exception of the time effects that are assumed to be equal across genders within each education/PMSA cell, and establishment fixed effects that are assumed to be equal for each education group, but allowed to vary across gender and PMSA. In addition to the orthogonality of the error term²⁴ ε_{it} , this is all we require to identify the change in the gender earnings gap by age for each education group and to decompose it into its individual and establishment components.

The key assumption for the identification of the widening of the individual component of the gender earnings gap by age is that the time effects are set to be equal across genders. Technically, this assumption ensures that the time fixed effects drop out of equation (3), which defines the gender earnings gap by age. This assumption implies that any year-to-year changes in labor market conditions affect men and women in the same birth cohort and education group and PMSA similarly over the period (1995-2008). We argue that this assumption is reasonable in our application.²⁵ The alternative, a common cohort effect across genders, seems questionable, given the gender convergence in labor market outcomes across the 33 cohorts included in our study.²⁶ As documented by Goldin (2016), women have made major inroads in terms of labor market participation, occupational segregation, and earnings across these cohorts, surpassing men in educational attainment. In contrast, such changes have been much more modest within specific birth cohorts as they aged during the observation period. Finally, while the gender pay gap shrank drastically in the 1970s and 1980s (affecting the earliest cohorts in our analysis), it has remained practically unchanged over our period.

Comparing the gender pay gap by age for different cohorts over time may help assess whether the assumption is realistic. Consider the alternative to our key assumption, namely the case where the trend over calendar time is less steep for women than for men and where the widening of the gender pay gap that we observe within each cohort is attributable to a widening over calendar time rather than by age within each cohort. In this case, the gender pay gap measured at different ages would be increasing in calendar time. For instance, the earnings gap for 27 year old in 2005 would be wider than the earnings gap observed for 27 year old in 2000, and so on. Figure III, which presents the development of the average gender earnings gap over time for different birth cohorts, does not seem to support such a scenario. In

²⁴ See Card et al. (2016) for a discussion of this assumption in a two-way fixed effect model of workers and establishments.

²⁵ While the Great Recession of 2009 (outside our sample period) has been noted to have differential effects on employment by gender, the 2000 recession (the main recession within our sample) was very gender neutral (Mishel et al. 2003: table 3.7; BLS 2007; BLS 2009).

²⁶ Our data comprises workers in age group 25-45 during the 13 years from 1995 to 2008, they are thus born from 1950 to 1983.

the figure, each line represents the gap in average earnings for each cohort as they are observed in subsequent cross sections of the CPS.²⁷ Cohorts are labelled by the year when they reach 25 to 29 years of age.²⁸

We make three observations. First, the differences between cohorts measured at a particular age and thus at different calendar times, seem small relative to the earnings growth by age for each cohort, particularly among workers with college. Second, it is not the case that the earnings gap, measured at a given age, is consistently smaller in the earlier cohorts that are observed at earlier points in time. Ranking the gender earnings gap at age 27 in ascending order, we observe the smallest gap for 27 year old in 2010, followed by 1995 and 2005. The largest gender earnings gap at age 27 is observed in 2000. Third, the ranking of the cohorts is not consistent across age groups. For instance, at age 37, the 2000 cohort has the lowest gender earnings gap. It is, of course, possible to construct combinations of gender specific trends offset by time-varying cohort effects that would partially reconcile these observations, but there is nothing in the raw data that suggests that differences in calendar time trends across genders is a noteworthy contributor to the widening of the gender earnings gap over the life course.

Finally, note that even if the common time effects assumption is not valid, our decomposition into individual and establishment components nevertheless would remain valid and meaningful. However, the external validity and interpretation of the results would be different: If a differential time trend adds to the widening of the gender earnings gap over time, we would expect new cohorts to come in with a higher gap, and we would not be able to distinguish between developments over the observation period that affect men and women of the same birth cohort differentially versus differential developments over the career purely due to age.

C. Extension: Age-Earnings Profiles and the Return to Seniority

So far, we have been concerned with the widening of the gender earnings gap by age. However, if we want to identify separate age-earnings profiles for each gender, we also need to identify β^m_1 , the baseline linear part of earnings growth for men. As explained above, when removing the linear term for age in equation (1''), the individual fixed effect may be reformulated as $\tilde{\alpha}_i =$

²⁷ The figure shows gender earnings gaps at different ages calculated from the data illustrated in figure II and discussed in appendix 2.

²⁸ Details of the construction – reference to earlier figures.

$\beta^m_1 Age_{i0} + \alpha_i$ where α_i incorporates cohort effects. To identify an age effect as distinct from a linear trend in the cohort effects, we need an additional assumption. We have chosen to represent the cohort effects for men by their decade of birth, assuming that among men the cohort effect is the same for all men born during the same decade²⁹. We thus obtain an estimate of β^m_1 from the equation:

$$(6) \quad \tilde{\alpha}_i = \alpha_c + \beta^m_1 Age_{i0} + \alpha_i$$

where α_c represents cohort fixed effects associated with decade of birth. In practice, we retrieve the individual fixed effects from the AKM decomposition of the job fixed effects (equation 4) in step 2 and regress them on “age at the beginning of panel” and the cohort fixed effects. Note that we do not need to treat the cohort effects for women similarly, since we already have an unbiased estimator for the difference between the linear age effects for men and women. We think an assumption of decade-by-decade cohort effects would be less realistic for women since the evolution of their labor force participation and pay was more dramatic even within a cohort decade. We thus obtain an estimate of the linear part of women’s earnings profile β^f_1 by adding the estimated coefficient for males from the step 1 estimate (equation 1’’) of the difference ($\beta^f_1 - \beta^m_1$).

An alternative would be to follow the procedure used by Card et al. (2013) and assume that the earnings profile of men reaches its apex at 40 years of age.³⁰ We believe that this assumption would also seem somewhat arbitrary in our application, since the linear part of women’s earnings profile might top at a later age than 40 due to career interruptions during the prime age years. To sum up, our preferred estimate comes from equation (6), but when presenting results on earnings growth by group, we do report numbers using both of the alternative procedures.

Another extension arises from the possibility that the within-establishment earnings growth may contain elements that are carried on from one job to the next (i.e. a general component), and other elements that are unique to the current job but lost once the individual changes establishment (i.e. an establishment-specific seniority premium). If this is the case the

29 Note that since we include job fixed effects in our first step, the individual fixed effects and thus the cohort effects are estimated conditional on establishment fixed effects. Any differences that allocate some cohorts into better paying firms or establishments over time is thus controlled for, and what we are addressing here is cohort effects on the distribution of earnings within firms.

30 In our case this is done by calculating $\beta^m_1 = -10\beta^m_2$, since we have normalized age to be zero at 35 in our data.

within-establishment age earnings profiles can be decomposed into a seniority profile and a general age-earnings profile, even when conditioning on the establishment fixed effect.

Empirically, it is difficult to identify the effect of seniority on earnings, in particular because a job match entailing a positive earnings premium for the worker is likely to last longer than matches with lower pay. However, in our case this common concern is taken care of by the introduction of job fixed effects. With the fixed effects, though, it is not possible to separate the linear terms for age and seniority within jobs, even if we assume the common time effects for each gender³¹. To separately identify the age and seniority profiles for each gender we rely on individuals with more than one job observed during the sample period. To simplify the following exposition, we ignore all variables that vary independently within each job, such as age squared and firm size. Consider the simplified within-job earnings equation:

$$(7) \quad \ln w_{it} = \beta_0 + \beta^g_1 Age_{it} + \beta^g_3 S_{ijt} + \psi^g_{ij} + \gamma_t + e_{it},$$

where S is the number of years in current establishment. We follow the same procedure as above, representing the model with a gender interaction term and removing the linear part of age. Note that $Age_{it} - Age_{i0} = S_{ijt} - S_{ij0} = t$, where S_{ij0} is seniority for individual i in establishment j at the beginning of the panel. We may then rewrite (7) as follows:

$$(7') \quad \ln w_{it} = \beta_0 + (\beta^f_1 + \beta^f_3 - \beta^m_1 - \beta^m_3) F Age_{it} + \check{\psi}^g_{ij} + \check{\gamma}_t + e_{it},$$

where it is clear that we can identify the linear term $(\beta^f_1 + \beta^f_3 - \beta^m_1 - \beta^m_3)$ only. We have $\check{\gamma}_t = \gamma_t + (\beta^m_1 + \beta^m_3)t$, $\check{\psi}^g_{ji} = \tilde{\alpha}_i + \beta^g_2 S_{ij0} + \varphi^g_j + \xi_{ij}$ and as before $\tilde{\alpha}_i = \beta^m_1 Age_{i0} + \alpha_i$. We may thus identify β^m_3 and β^f_3 from the gender specific equations³².

$$(8) \quad \check{\psi}^g_{ji} = \tilde{\alpha}_i + \beta^g_2 S_{ij0} + \varphi^g_j + \omega_{ij}$$

where the identification of β^g_3 comes from individuals who hold two or more jobs only. Again we may obtain the benchmark linear term for men using (6)³³.

³¹ It is possible, however, to directly identify the non-linear parts of the seniority profile in the within job specification (1'). This will be added in the next revision of the paper, once we can apply renewed access to the data based on an invitation to revise and resubmit the paper to an academic journal (as per Census Bureau regulations).

³² Again, we use REG2HDFE.

³³ An added complication arises since we only observe seniority from the beginning of our panel. This means that the Age0 variable captures the effect of seniority at the beginning of the panel as well. We make a correction for this by estimating for each gender-education group how much seniority grows by age using survey data (CPS March 2000) containing information on seniority (job tenure), and calculate $\hat{b}_1 = (\beta_1 + \hat{\beta}_3 \hat{\tau})$, where b is coefficient for Age0 in (6), and τ is the coefficient for age in a regression of seniority on age from the survey. The first part is the coefficient for age in the model, β_1 , that we are after, whereas the second part is the contribution from seniority, $\hat{\beta}_3 \hat{\tau}$, up to Age0. The coefficient τ is estimated to be 0.31 for no-college women, 0.34 for college women, and 0.41 for men, irrespective of education. We use the

To sum up: The widening of the within establishment gender earnings gap by age is estimated directly from a within job log earnings regression for each gender, using an assumption of common time effects to identify the age profile of the gap. The widening of the establishment component of the gender earnings gap by age is estimated by regressing the predicted establishment component from the same regression on age, separately by gender. We make two extensions: To obtain the gender specific age-earnings profiles within establishments that we show in figure III below, we add an estimate of the linear age term for men by regressing the individual fixed effects on age at the beginning of the panel and cohort effects defined by decade of birth. Finally, we add an estimate of the seniority part of the age-earnings profile using individuals observed with two different jobs in the panel.

III. RESULTS

III A. A large part of the increasing earnings gap occurs within establishments

Figure IV shows the predicted age-earnings profiles obtained from separate regressions with fixed individual and establishment effects as outlined above. The models include year effects as well. The figure illustrates the evolution of earnings over time for the same individual, keeping the contribution from the establishment component unchanged, relative to a 25-year-old of the same gender and educational group. Note that, given the large sample size, all coefficients are estimated very precisely. This is why we do not report standard errors in the figures and tables (see data section for details).

Once again, we find a much steeper profile for college-educated workers, in particular for men. An average college educated man improves his earnings by 57 log points going from age 25 to age 45, while a college educated woman improves her earnings only by 27 log points. The first takeaway of our analysis is thus that for college-educated workers, a considerable part of the growth in the gender earnings gap occurs within establishments.

For workers without a college degree, however, the within establishment earnings gap increases over the first 10 years but then narrows considerably over time as the earnings profile for men flattens out. By age 45 women without a college degree more than catch up to similarly

CPS Job Tenure data for the March 2000 CPS supplement provided by the CEPR (<http://ceprdata.org/cps-uniform-data-extracts/cps-job-tenure/cps-job-tenure-data/>) and estimate the model for workers between 25 and 45 years of age.

educated men. Their earnings grow by 28 log points over the 20-year period, while men's earnings increase only by 25 log points.

Table I quantifies the change in the gender earnings gap within establishments for different age intervals. Among the non-college educated, we find that men have faster earnings growth within establishments until the age of 35, but women catch up in terms of earnings growth after that. By age 45 their accumulated earnings growth is 3 log points higher than men's. Among the college educated, men have faster earnings growth up to age 40, followed by some catching up by women between ages 40 and 45 as the male earnings profile flattens out. However, by age 45 college educated men have a 30 log points advantage in accumulated earnings growth relative to women in the same education group.

III.B. Seniority and age profiles

Next, we decompose the within-establishment earnings growth into the component that is purely due to seniority within a firm, i.e. the part of earnings growth that is lost when a person changes jobs, and the component of within establishment earnings growth that is retained also when changing jobs. Table II displays the estimated returns to seniority, measured as the relative earnings gain from one year of additional tenure within the same establishment. We find that men have a higher return to seniority than women, particularly among the college educated.

The remaining part of the within-establishment earnings growth by age, that is retained when changing jobs, is found by subtracting the average contribution from the return to seniority, calculated using average seniority by age, education and gender³⁴, from the overall within-establishment earnings profile as displayed in Figure II. Conditional on seniority and the establishment fixed effect, the age profile measures the improvement in the relative earnings within the establishment that a person retains even when changing jobs. An example of a positive within-establishment earnings change would be a person low in the earnings hierarchy of one establishment who changes jobs to a position higher up in the earnings hierarchy of another establishment. Figure V illustrates how the within-establishment seniority and age

³⁴ See the methodology section for details. As discussed there, it is possible also to identify non-linear terms for the seniority profile in the within job specification (1''). The current results are based on a linear specification in seniority, and the distribution of earnings growth allocated to age and seniority may be sensitive to this restriction.

profiles interact for a person who changes job every fifth year, using college educated men and women as an example.

Both the seniority and age profiles contribute to the widening of the within-establishment earnings gap over time, particularly for college educated workers. In Table III we show the contribution to the development of the gender earnings gap over time from the within-establishment age and seniority profiles, separately. The seniority wage contribution is constant over time as the returns to seniority only has a linear term, whereas the age term is allowed to change over time. The contribution of gender differentials in returns to seniority to the widening of the pay gap is twice as large among the college-educated than among workers without a college degree.

In the beginning of the career, the age profile is steeper for men than for women, and dominates as a source of increasing pay gap, particularly for college educated workers. This means that men, in addition to having steeper earnings growth within establishments, also gain more in terms of relative earnings within establishments when they change jobs. As people age, the age profile in particular for men, dips to negative, and contributes to a narrowing of the gender earnings pay gap.

III.C. The establishment component of the gender earnings gap increases with age

The establishment component of earnings is the sum of the coefficient of the fixed establishment effect and the component of earnings that is attributable to firm size. It measures an establishment earnings premium, or how much more (or less) the employer pays an individual worker over and above the average establishment in the economy. The widening of the gap in the establishment component as individuals age provides a key finding in this paper: It shows the differential contribution for men and women arising from changes in establishment characteristics and job-to-job changes across employers with different establishment earnings premiums during their prime working age years. Figure VI shows the development of the establishment earnings component over time, compared to a person of 25 years of age with the same gender and education.

We find a rising age-earnings profile in the establishment component for both college and non-college workers. An average college educated man gains 21 log points from age 25 to 45 in terms of the establishment earnings component, while a college educated woman on

average gains 10 log points. Non-college men experience a growth of 17 log points, whereas the number for women is 12 log points.

Table IV shows the contribution to the widening gender earnings gap that comes from the shifting distribution of pay across establishments. For non-college educated workers, the establishment contribution is largest in the beginning, suggesting that men disproportionately move into higher paying establishments early in their career. For college educated workers, the difference is larger and more evenly spread across age intervals. The gender gap in the average establishment earnings premium increases by 5 log points going from age 25 to age 45 for non-college workers and 11 log points for college workers.

Next, we quantify the contribution of the individual and establishment component to the widening of the gender gap. Table V reports the total change in the gender earnings gap by age (column 1 and 3) and the share attributable to the establishment component (column 2 and 4). Among workers without a college degree, the total earnings gap widens from age 25 to 30, and from age 30 to age 35. The share arising from an increasing establishment earnings premium is 28 and 57 percent, respectively. For college workers, the total gender earnings gap increases in all age intervals up to 40 years of age. The share of the establishment contribution increases from 13 percent during the first 5 years to 41 percent between age 35 and 40.

Up to age 40, about 56 percent of the accumulated gender gap in earnings growth for non-college educated workers, and about 20 percent of the accumulated gender gap for the college-educated is due to differential growth in the establishment component of pay. The remaining parts are due to differential earnings growth in the individual component.

We further investigate the role of marital status in explaining the widening of the gender earnings gap arising from earnings differentials across establishments. Figure VII displays the predicted establishment earnings premium by age for married and never married (as of 2000) men and women. Interestingly, although not surprisingly, the bulk of the increase in the establishment earnings gap over the life cycle is driven by the behavior of married men and women. Consider college educated workers first. The earnings path for married and non-married men is very similar, but so is also the earnings path of non-married women. The difference in the growth of the establishment component of pay seems to be entirely driven by the lower growth among married women. For non-college workers, the pattern is similar, but less pronounced. Again, the dominant part of the difference between men and women arises from lower growth among married women.

III.D. Additional considerations

Given that the LEHD data are sparse in many variables that are often available in survey data sets such as the CPS, we cannot easily quantify how much of the gender pay gap increase is due to gender differences in hours worked, prevalence of part-time work, differential accumulation of work experience with age and so on. Instead, our findings relate to the overall gap in earnings having controlled for worker and firm fixed effects. We acknowledge that any and all of these factors might contribute to the pay gap patterns documented in this paper. For example, to the extent that women take more career breaks around the time of family formation, that would help explain why they experience less of a benefit from changing jobs. While we cannot directly tackle this issue we experimented with alternative selection rules that limit the sample to workers who are more strongly attached to the labor market. For example, we required workers to be present at least 50% or 80% of the time or restricted the sample to individuals observed in the data for 5 or 10 consecutive years. In all cases, we found that the increase in the earnings gap over time (and the pattern between the individual and establishment components) were very similar regardless of the sample selection criteria.³⁵

Finally, previous studies have shown that the number of hours worked can explain a large part of the overall gender pay gap, and in many occupations can be an important determinant of earnings (Goldin, 2014). Since we only have data on usual hours worked as of 2000 we are unable to tease out the importance of hours in our dynamic analysis. However, in cross-sectional analyses for year 2000 we found that the usual hours worked are positively related to earnings and can explain about 20 to 25 percent of the gender pay gap in the cross-section, depending on specification and occupations included.³⁶

IV. DISCUSSION AND CONCLUSIONS

The gender earnings gap widens considerably during the first two decades of working life. This widening is much stronger among college-educated workers than among workers without a college degree. Also, when controlling for cohort effects, the earnings growth from age 25 to 45 of men and women with a college degree is similar to the development depicted in figure I.

³⁵ The trade-off is that the sample size shrinks considerably the stricter the criteria used.

³⁶ We also checked how much of the 2008 pay gap the usual hours in the 2000 Decennial Census could explain, and found that the estimated gender pay gap was reduced by about 16 percent when including a control for the usual work hours 8 years previous.

College educated men on average experience a wage growth of 110 percent, while women experience 39 percent earnings growth over the same period. In comparison, the difference between the earnings growth of men and women without a college degree is minimal: Controlling for cohort effects, men without a college degree on average experience an earnings growth of 52 percent, very similar to women's earnings growth of 49 percent. This deviates from the pattern for non-college workers in figure I, which shows a widening of the gender earnings pay gap by age groups in the cross sections. This different pattern suggests that younger cohorts of non-college women have improved their position relative to the older cohorts of non-college women, in comparison to men with the same age and education.

In order to assess the extent to which the widening of the pay gap over a worker's career is associated with the organization of firms and advancements in job ladders within firms, or with different earnings distributions and job-to-job mobility patterns across establishments, we decomposed it into an individual and an establishment component.

Over time, an individual may improve his or her position within the establishment by climbing the corporate ladder. When changing employer, this improvement may be transferred over to the new employer or even enhanced, or it may be lost as in a game of chutes and ladders. The seniority profile captures the earnings growth within a job that is lost when changing employer, while the within-establishment age profile, conditional on seniority, captures the within-establishment earnings growth that is maintained even when changing employer. In terms of earnings growth within establishments, college educated men stand out with a significantly higher earnings growth than all the other groups. Relative to women, men have both steeper seniority profiles, on average longer seniority, and experience a steeper age-earnings profile within establishments, conditional on seniority. Men's relative advantage in terms of earnings advances within the establishment is particularly large during the early career; after 10 years the earnings gap has widened by 29 log points. From 40 to 45 women catch up some of the advantage, amounting to 5.4 log points.

The establishment earnings premium also changes over time for each individual. This change adds significantly to the widening of the gender earnings gap, especially for the college educated. Men climb the establishment ladder faster than women. From 25 to 45 years of age, the establishment earnings component grows by 21 log points among college educated men, and by 10 log points for college educated women. This part of the increase of the gender earnings gap is entirely different from the individual component. While the individual component originates from the combination of the management determined job structure within

firms, and the workers' competition to climb the ladder, the establishment component originates from the combination of options that arise in the marketplace and the workers' job mobility patterns. Gender differences in the growth of the establishment component arise from differences in the probability of job-to-job transitions and differences in the earnings gain associated with such job changes. A voluntary job move is likely to be associated with earnings gains, whereas involuntary or "tied" moves are more likely to be associated with an earnings loss. Speculatively, if women are disproportionately 'tied movers' (i.e. change jobs/location to follow the "primary earner" in the family) this could explain the widening of the gap that arises from the between component (see e.g. Cooke et al., 2009). What we have accounted for in this paper is the combined result of such differences.

We find that almost all the difference in the growth of the establishment component of pay across genders is due to differences by marital status. Whereas the difference between men and women in the growth of the establishment component is small among non-married employees, there is a large widening of the gender gap for the married. This pattern is particularly strong among college educated workers: At 45 years of age, a married college educated man has gained more than 20 log points of earnings, whereas a married college educated woman gains on average less than 5 log points. These results suggest that among married couples, the household division of labor tend to limit women's career choices with respect to job-to-job changes and this is an important determinant of the widening of the gender earnings gap, especially for college educated women who are more likely to be in occupations with steep age-earnings profiles. Evidence based on the publicly available 2000 Census (not reported) suggests that this is linked to the arrival of children, which is especially costly for couples where both partners are college educated. Our linked employer-employee data allow further exploration and (potentially) identification of the underlying mechanisms because unlike in other linked data, spouses (as of 2000) are matched, providing data on the demographic characteristics of both spouses and the presence and age of children. Because of limitations in data access we are currently unable to investigate this further, but plan to do that in future work.

REFERENCES

Abowd, John, Francis Kramarz, and David Margolis (1999) "High Earnings Workers and High Earnings Firms." *Econometrica* 67(2): 251-333.

Abowd, John M., Bryce E. Stephens, Lars Vilhuber, Fredrik Andersson, Kevin L. McKinney, Marc Roemer, and Simon Woodcock. (2002) "The LEHD Infrastructure Files and the Creation of the Quarterly Workforce Indicators." Longitudinal Employer-Household Dynamics Technical Papers 2002-05, Center for Economic Studies, U.S. Census Bureau

Abowd, John M., Robert H. Creecy, and Francis Kramarz. (2002) "Computing Person and Firm Effects Using Linked Longitudinal Employer-Employee Data." Longitudinal Employer-Household Dynamics Technical Papers 2002-06, Center for Economic Studies, U.S. Census Bureau.

Abraham, Katharine R., James R. Spletzer, and Jay C. Stewart (1998) "Divergent Trends in Alternative Wage Series." in John Haltiwanger, Marilyn E. Manser, and Robert Topel (editors), *Labor Statistics Measurement Issues*, NBER Studies in Income and Wealth, Vol. 60, Chicago, University of Chicago Press, pp. 293–324.

Albanesi, Stefania and Claudia Olivetti. (2009) "Home Production, Market Production and the Gender Earnings Gap: Incentives and Expectations." *Review of Economic Dynamics* 12(1): 80-107.

Altonji, Joseph and Rebecca Blank. (1999) "Race and Gender in the Labor Market." In Orley Ashenfelter and David Card (editors), *Handbook of Labor Economics* Vol. IIIc. Amsterdam: Elsevier, pp. 3143-3259.

Amuedo-Dorantes, Catalina, and Sara De la Rica. (2006) "The Role of Segregation and Pay Structure on the Gender Wage Gap: Evidence from Matched Employer-Employee Data for Spain." *The BE Journal of Economic Analysis & Policy* 5(1).

Bayard, Kimberley, 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.

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 and Arne Mastekaasa (1996) "Decomposing the Male/Female Earnings Gap: Within and Between Establishment Differences." *Labour*, 10: 339–356.

Barth, Erling and Harald Dale-Olsen. (2009) "Monopsonistic Discrimination, Worker Turnover and the Gender Wage Gap." *Labour Economics* 16:5 89-597

Barth, Erling, Bernt Bratsberg and Oddbjørn Raaum. (2012) "Immigrant Earnings Profiles Within and Between Establishments." *Labour Economics* 19(4): 541-556.

- Bayard, Kimberly, Judith Hellerstein, David Neumark and Kenneth Troske, (2003) "New Evidence on Sex Segregation and Sex Differences in Earnings from Matched Employee-Employer Data." *Journal of Labor Economics* 21(4): 887-922.
- Becker, Gary S. (1975) *Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.* Second Edition NBER Books, National Bureau of Economic Research, Inc.
- Bertrand M, Goldin C, Katz LF. (2010) "Dynamics of the Gender Gap for Young Professionals in the Financial and Corporate Sectors." *American Economic Journal: Applied Economics* 2(3): 228-255.
- Black, Dan A. (1995) "Discrimination in an Equilibrium Search Model." *Journal of Labor Economics* 13: 309–34.
- Blau, Francine D. (1977). *Equal Pay in the Office.* Lexington Books, D.C. Heath and Company Lexington, Massachusetts Toronto
- Blau, Francine D. and Lawrence M. Kahn. (1981) "Race and sex differences in quits by young workers." *Industrial and Labor Relations Review, ILR Review*, Cornell University, ILR School 34(4): 563-577.
- Blau, Francine D. and Lawrence M. Kahn. (2000) "Gender Differences in Pay." *Journal of Economic Perspectives* 14(4): 75-99.
- Blau, Francine D., and Lawrence M. Kahn. (2006) "The U.S. Gender Pay Gap in the 1990s: Slowing Convergence." *Industrial and Labor Relations Review* 60 (1): 45–66.
- Blau, Francine D. and Jed DeVaro. (2007) "New Evidence on Gender Differences in Promotion Rates: An Empirical Analysis of a Sample of New Hires." *Industrial Relations.*
- Blau, Francine D. and Lawrence M. Kahn. (2016) "The Gender Wage Gap: Extent, Trends, and Sources." *Journal of Economic Literature* (forthcoming). Available as NBER WP 21913.
- BLS (2007) News: the employment situation May 2007. Bureau of Labor Statistics Washington: United States Department of Labor.
- BLS (2009) News: the employment situation February 2009. Bureau of Labor Statistics Washington: United States Department of Labor.
- Booth, Alison L., Marco Francesconi, and Jeff Frank. (2003) "A Sticky Floors Model of Promotion, Pay, and Gender." *European Economic Review*, 47:295-322.
- Bowlus, Audra J. (1997) "A Search Interpretation of Male-Female Earnings Differentials." *Journal of Labor Economics* 15(4): 625-657.
- Bowlus, Audra J. and Zvi Eckstein. (2002) "Discrimination and Skill Differences in an Equilibrium Search Model." *International Economic Review* 43(4): 1309-1345.

- Bosworth, Barry, Gary Burtless and C. Eugene Steuerle. (2001) "Lifetime earnings patterns, the distribution of future social security benefits, and the impact of pension reform." *Social Security Bulletin*, Vol. 63. No. 4.
- Bronson, Mary Ann and Peter Skogman Thoursie. 2018. "The Wage Growth and Internal Mobility of Men and Women: New Evidence and Theory." Manuscript, Georgetown University.
- 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.
- Card, David, Jörg Heining and Patrick Kline. (2013) "Workplace Heterogeneity and the Rise of West German Wage Inequality." *The Quarterly Journal of Economics* 128(3): 967-1015.
- Cooke, Thomas, Paul Boyle, Kenneth Couch, and Peteke Feijten. (2009), "A Longitudinal Analysis of Family Migration and the Gender Gap in Earnings in the United States and Great Britain." *Demography* 46(1): 147-167.
- Datta Gupta, Nabanita and Donna S. Rothstein. (2005) "The Impact of Worker and Establishment-Level Characteristics on Male-Female Wage Differentials: Evidence from Danish Matched Employee-Employer Data." *LABOUR* 19(1): 1-34.
- Del Bono, Emilia, and Daniela Vuri. (2011) "Job Mobility and the Gender Wage Gap in Italy." *Labour Economics* 18(1): 130-142.
- Félix, Sónia and Portugal, Pedro. (2016) "Labor Market Imperfections and the Firm's Wage Setting Policy." IZA Discussion Papers 10241, Institute for the Study of Labor (IZA).
- Flabbi, Luca. (2010) "Gender Discrimination Estimation in a Search Model with Matching and Bargaining." *International Economic Review* 51(3): 745-783.
- Francois, Patrick. (1998) "Gender Discrimination without Gender Difference: Theory and Policy Responses." *Journal of Public Economics* 68(1): 1-32.
- Francois, Patrick and Jan van Ours. (2000) Gender Earnings Differentials in a Competitive Labor Market: The Household Interaction Effect. IZA Discussion Papers 202, Institute for the Study of Labor (IZA).
- Frazis, Harley, and Jay Stewart. (2004) "What Can Time-Use Data Tell Us About Hours of Work?" *Monthly Labor Review*, December 2004: 3-9.
- Gayle, George-Levi and Limor Golan. (2012). "Estimating a Dynamic Adverse-Selection Model: Labour-Force Experience and the Changing Gender Earnings Gap 1968–1997." *Review of Economics and Statistics*, 79(1): 227-267.
- Goldin, Claudia. (2014) "A Grand Gender Convergence: Its Last Chapter." *American Economic Review*, 104(4): 1–30.

- Goldin, Claudia, Sari Kerr, Claudia Olivetti and Erling Barth. (2017) "The Expanding Gender Earnings Gap: Evidence from the LEHD-2000 Census." *American Economic Review P&P*, 107(5): 110-114.
- Green, Francis, Stephen Machin and Alan Manning. (1996) "The Employer Size-Earnings Effect: Can Dynamic Monopsony Provide an Explanation?" *Oxford Economic Papers* 48(3): 433-55.
- Groshen, Erica L. (1988) "Why Do Earnings Vary Among Employers?" *Economic Review*, Federal Reserve Bank of Cleveland, issue Q I: 19-38.
- Groshen, Erica L. (1991) "The Structure of the Female/Male Wage Differential: Is it Who You Are, What You Do, or Where You Work?" *Journal of Human Resources*, Vol. 26(3): 457-472.
- Guimaraes, Paulo and Portugal, Pedro. (2010) "A Simple Feasible Alternative Procedure to Estimate Models with High-Dimensional Fixed Effects." *The Stata Journal* 10(4): 625-649.
- Heinze, Anja and Wolf, Elke. (2010) "The Intra-Firm Gender Wage Gap: A New View on Wage Differentials Based on Linked Employer–Employee Data." *Journal of Population Economics* 23(3): 851-879.
- Hirsch, Boris, and Claus Schnabel. (2012) "Women Move Differently: Job Separations and Gender." *Journal of Labor Research* 33(4): 417-442.
- Hyatt, Henry, Erika McEntarfer, Kevin McKinney, Stephen Tibbets, and Doug Walton. (2014) "JOB-TO-JOB (J2J) Flows: New Labor Market Statistics From Linked Employer-Employee Data." Working Papers 14-34, Center for Economic Studies, U.S. Census Bureau.
- Korkeamäki, Ossi, and Tomi Kyrrä. (2006) "A Gender Wage Gap Decomposition for Matched Employer–Employee Data." *Labour Economics* 13(5): 611-638.
- Lazear, Edward P. (1981) "Agency, Earnings Profiles, Productivity, and Hours Restrictions." *American Economic Review* 71(4): 606-620.
- Lazear, Edward P. and Sherwin Rosen. (1990) "Male-Female Earnings Differentials in Job Ladders." *Journal of Labor Economics* 8:1, Part 2: Essays in Honor of Albert Rees. (January): S106-S123.
- Light, Audrey and Manuelita Ureta. (1995) "Early-Career Work Experience and Gender Earnings Differentials." *Journal of Labor Economics* 13(1): 121-54
- Light, Audrey and Manuelita Ureta. (1990) "Gender Differences in Earnings and Job Turnover Among Continuously Employed Workers." *American Economic Review* 80: 293-97.
- Loprest, Pamela J. (1992) "Gender Differences in Wage Growth and Job Mobility." *American Economic Review*, 82(2): 526-32.
- Manning, Alan. (2003) *Monopsony in Motion*. Princeton: Princeton University Press.
- Manning, Alan, and Joanna Swaffield. (2008) "The Gender Gap in Early-Career Earnings Growth." *Economic Journal* 118(530): 983–1024.

- McCue, Kristin. (1996) "Promotions and Wage Growth." *Journal of Labor Economics* 14(2): 175-209.
- Meng, Xin. (2004) "Gender Earnings Gap: The Role of Firm Specific Effects." *Labour Economics* 11: 555-573.
- Mishel, Lawrence., Jared Bernstein, And Sylvia Allegretto. (2003) *The State of Working America 2002/2003*. An Economic Policy Institute Book. Ithaca, N.Y.: ILR Press, an imprint of Cornell University Press.
- Robinson, John, and Ann Bostrom. (1994) "The Overestimated Workweek? What Time Diary Measures Suggest." *Monthly Labor Review*, August 1994: 11–23.
- Royalty, Anne Beeson. (1998) "Job-to-Job and Job-to-Nonemployment Turnover by Gender and Education Level." *Journal of Labor Economics* 16(2): 392-443.
- Simón, Hipólito. (2012) "The Gender Gap in Earnings: An International Comparison with European Matched Employer–Employee Data." *Applied Economics* 44(15): 1985-1999.
- Song, Jae, David J. Price, Fatih Guvenen, Nicholas Bloom, and Till von Wachter. (2019) "Firming Up Inequality" *Quarterly Journal of Economics*, 134(1): 1-50.
- Sorkin, Isaac. (2017). "The Role of Firms in Gender Earnings Inequality: Evidence from the United States." *American Economic Review Papers and Proceedings*, 107(5):384-387.
- Taylor, Mark. (2007). "Tied Migration and Subsequent Employment: Evidence from Couples in Britain." *Oxford Bulletin of Economics and Statistics*, 69(6): 795-818.

Table I: Change in the individual component of the gender earnings gap by education and age

Age intervals	No college	College
25-30	0.052	0.181
30-35	0.013	0.111
35-40	-0.028	0.041
40-45	-0.068	-0.029
25-45	-0.030	0.304

Notes: Log differences. A positive number indicates an increase in the earnings gap between men and women. Calculated from predicted values from separate regressions of ln earnings by education, see Figure II.

Table II: Returns to seniority by gender and education

No college			College		
Women	Men	Difference	Women	Men	Difference
0.0139	0.0172	0.0033	0.0112	0.0213	0.0100

Notes: Estimated coefficient for years of seniority in the full model specification estimated from job-effect equations (equation (7)). Dependent variable ln earnings.

Table III Change in the individual component of the gender earnings gap by age interval.
Contributions from seniority and age by gender and education.

Age interval	No college		College	
	Age	Seniority	Age	Seniority
25-30	0.038	0.014	0.156	0.025
30-35	-0.002	0.014	0.086	0.025
35-40	-0.042	0.014	0.016	0.025
40-45	-0.082	0.014	-0.054	0.025
25-40	-0.086	0.056	0.205	0.099

Notes: Change in gender earnings gap due to different returns to seniority and age within establishments. A positive number indicates an increase in the earnings gap between men and women. Log differences calculated from equation (1') and (6).

Table IV: Change in the establishment component of the gender earnings gap by age and education

Age interval	No College	College
25-30	0.021	0.028
30-35	0.016	0.028
35-40	0.011	0.028
40-45	0.006	0.028
25-45	0.054	0.110

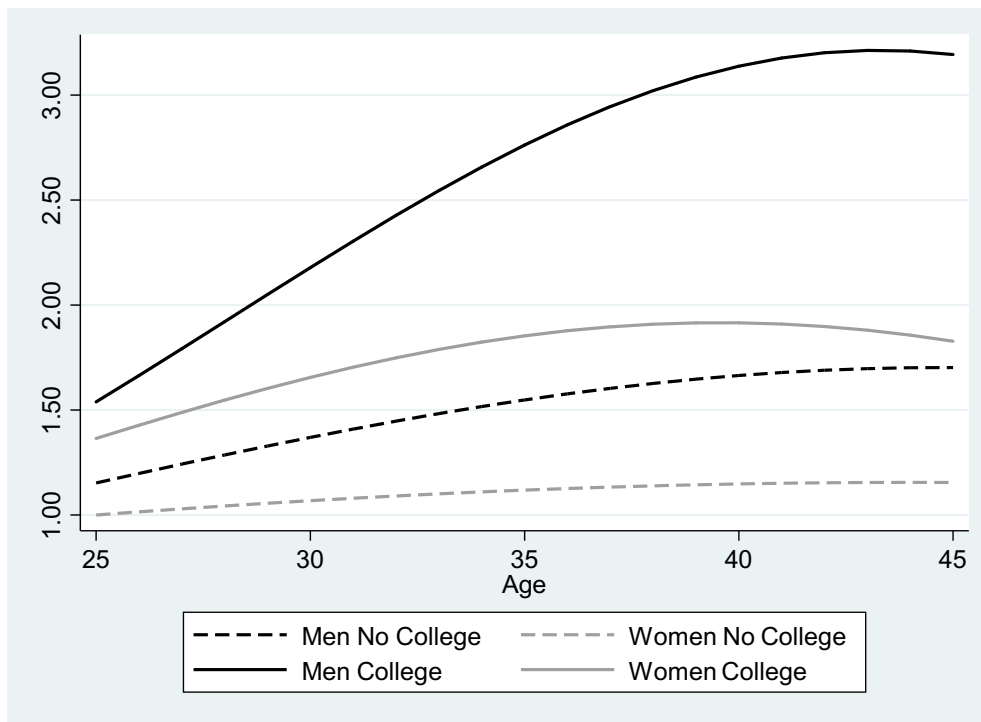
Notes: Log differentials calculated from predicted values from separate regressions of log earnings by education, see Figure IV. A positive number indicates an increase in the earnings gap between men and women.

Table V: Share of growth in earnings gap due to the establishment component by age interval

Age interval:	<u>Non-college</u>		<u>College</u>	
	Total change	% Establishment	Total change	% Establishment
25-30	0.074	28 %	0.209	13 %
30-35	0.028	57 %	0.138	20 %
35-40	-0.016		0.069	41 %
40-45	-0.062		-0.002	.
25-45	0.024	225%	0.414	27 %

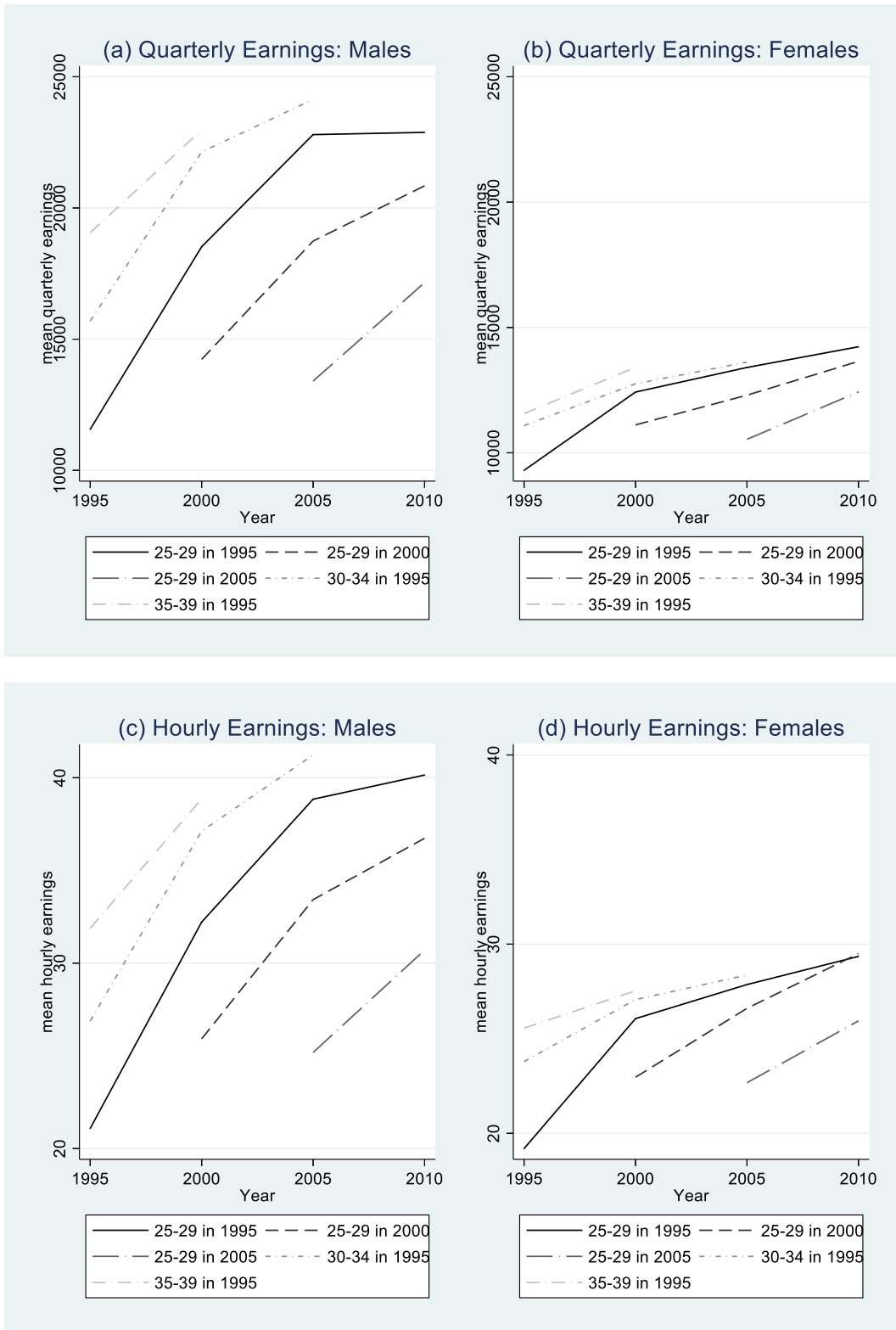
Notes: The change in the gender gap is obtained as the sum of the age/education specific entries from Tables I and IV. % between is obtained as the ratio of the establishment component (in Table IV) to the total.

Figure I: Predicted relative earnings-age profiles by education and gender.
 Reference: 25-year-old woman without a college degree



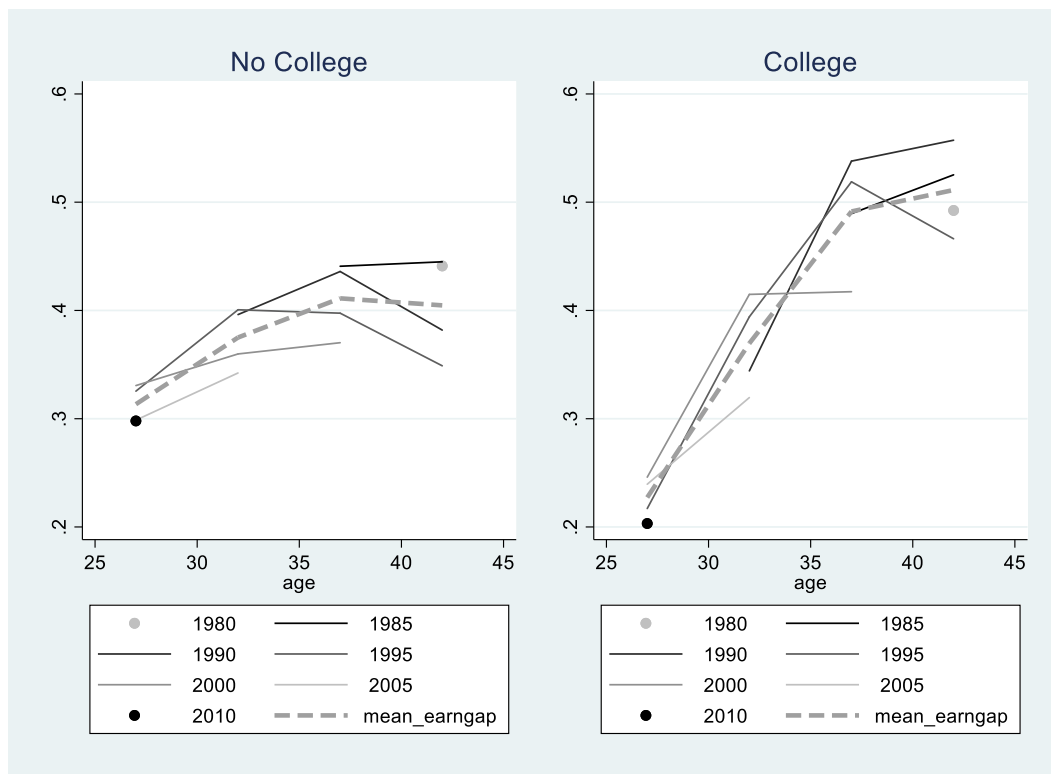
Notes: The lines show the relative earnings of an average individual of each gender and education group by age, normalized by the earnings of a 25-year-old woman with no college degree. These statistics are calculated from predicted values of separate regressions of log earnings on age and its square for each gender/education bin. All models include time dummies. *Sources* Average profiles are calculated on the matched sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. See data section for details on sample construction. Education is obtained from the matched 2000 Decennial Census long-form data.

Figure II. Age-profiles by gender, cohort and earnings concept: College Graduates



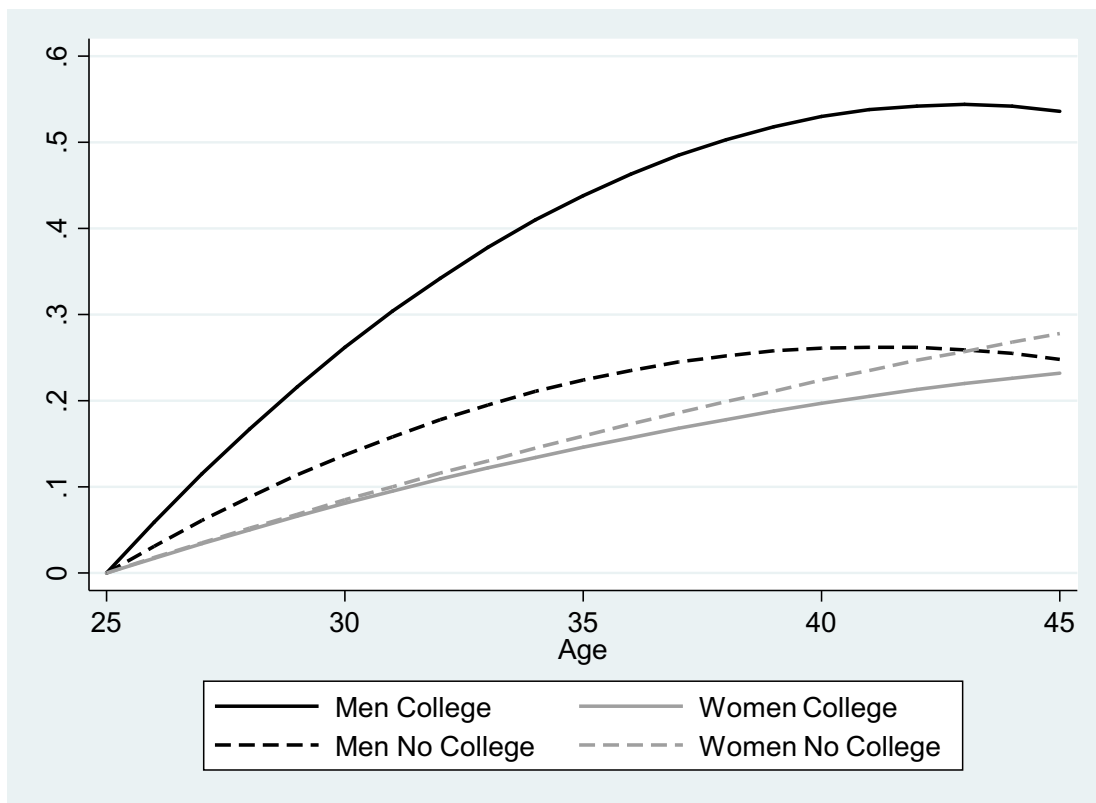
Notes: Data from CPS. See appendix 2 for details.

Figure III. Average gender earnings gap by age and cohort



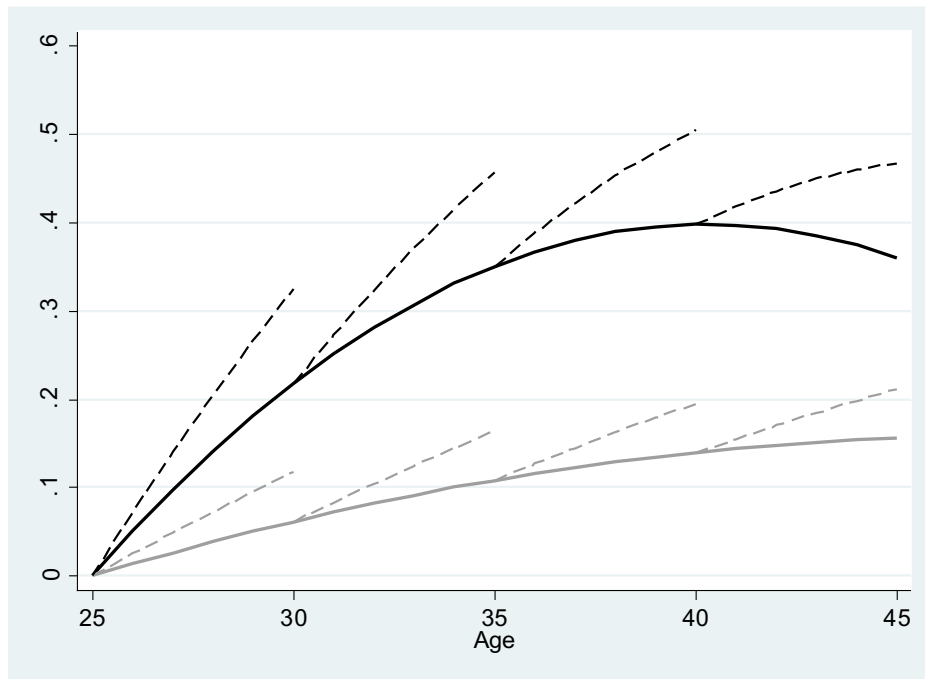
Notes: Calculated from CPS years. See appendix 2 for details.

Figure IV: Growth in the individual component of earnings by age



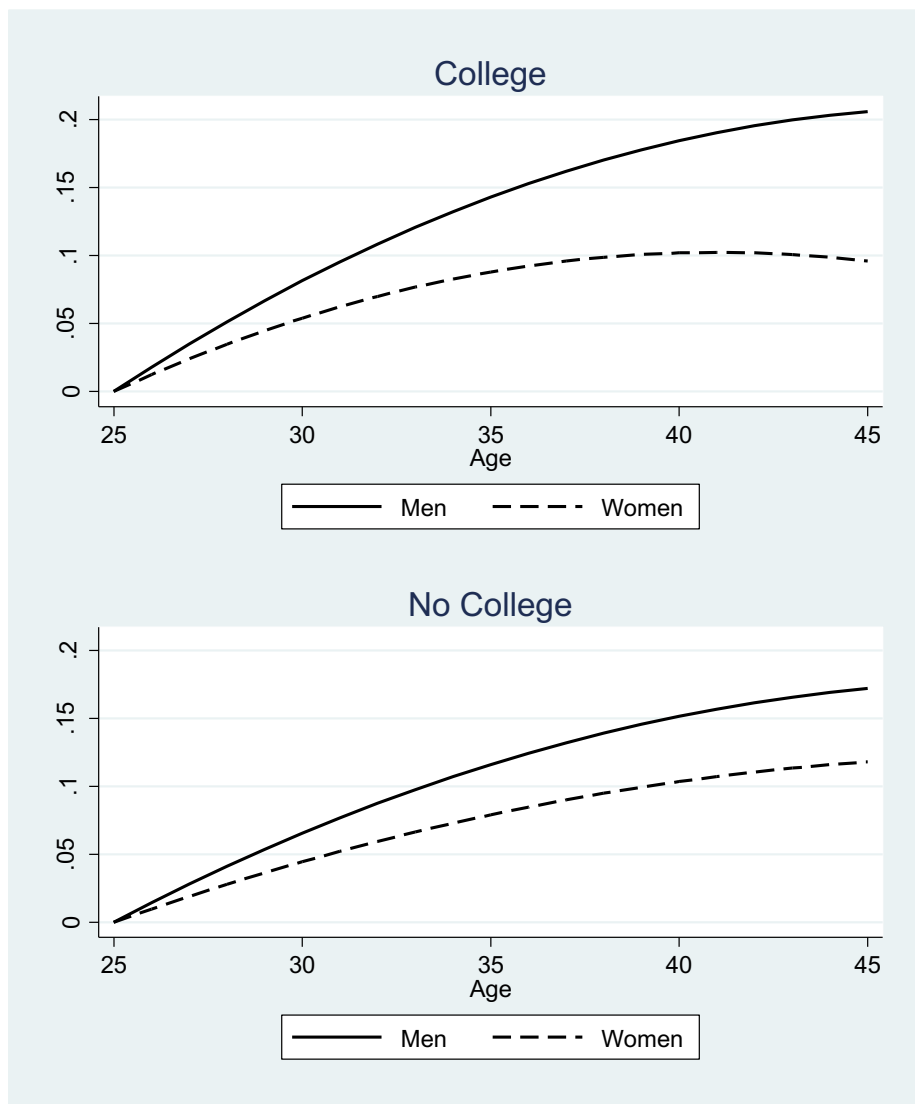
Notes: Log differences relative to group specific 25-year-old. Predicted values from separate regressions by education (equation (1')). The model also includes individual and establishment fixed effects and the log of establishment size with gender interactions, see Section 3 for details. Average profiles calculated on the matched sample from 26 of the 50 largest PMSAs located in LEHD states.

Figure V: Log earnings-differential by seniority and age. College educated workers.



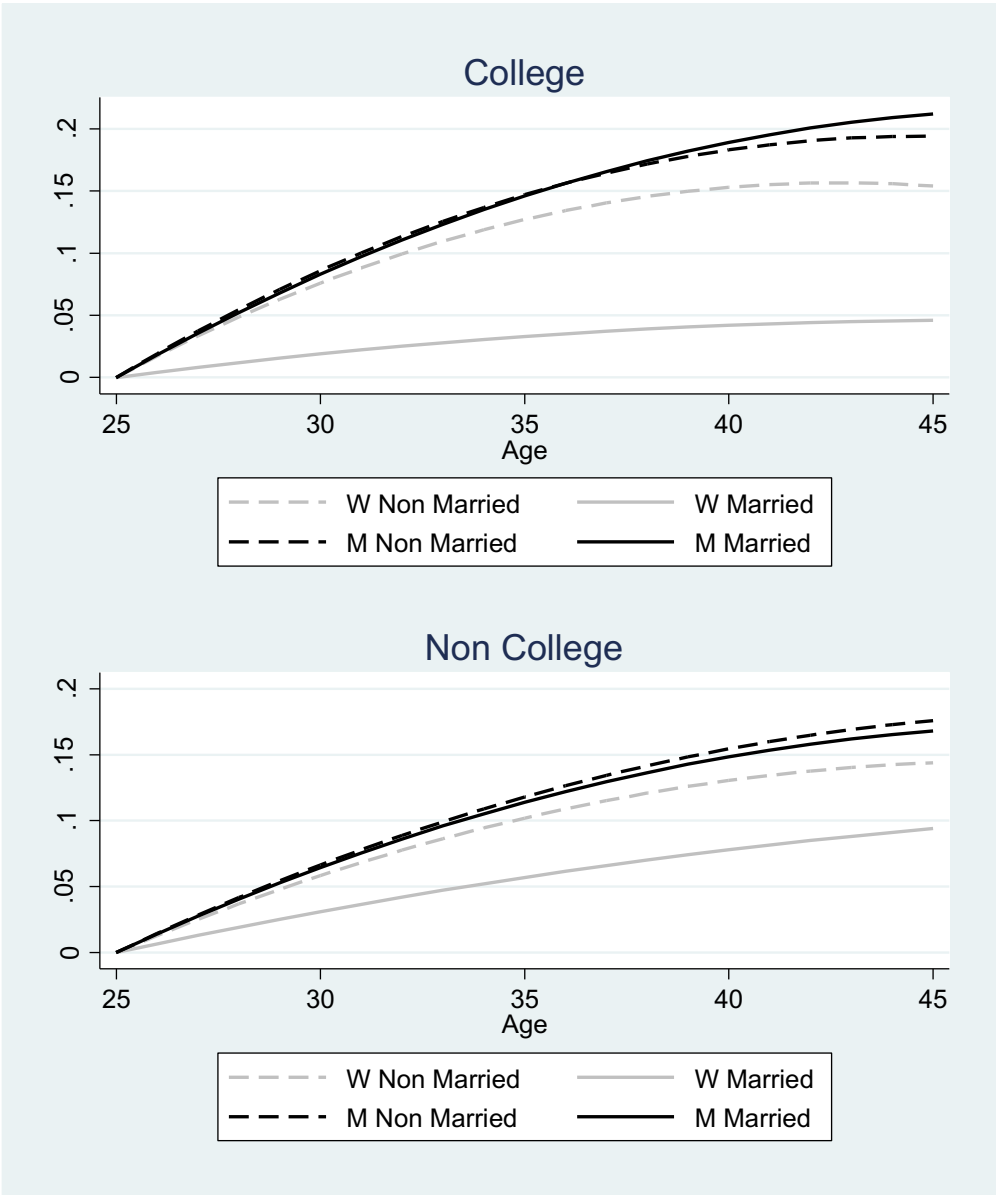
Notes: Calculated within-establishment earnings profiles for men and women with a college degree who changes jobs every 5 years. Black: Men, Grey: Women. Solid lines show predicted age-earnings profiles, conditional on seniority, and dotted lines show the additional returns obtained in each new job. Average profiles calculated on the matched sample from the largest 26 PMSA's in the US.

Figure VI Predicted establishment component of earnings by age



Notes: Predicted establishment earnings component by age relative to a 25-year-old with the same education and gender. Predicted values from separate regressions for each gender and educational level (equation 5 – model with seniority). The establishment earnings premium includes the fixed establishment effect and the earnings premium due to firm size. The model also includes individual fixed effects, see methods section for details on the estimation method. Average profiles calculated on the matched sample from the largest 26 PMSA's in the US.

Figure VII: Predicted establishment earnings by age, gender, education, and marital status



Notes: Predicted values from separate regressions for each gender and educational level. Dependent variable: Establishment earnings premium. Average profiles calculated on the matched sample from the largest 26 PMSA's in the available LEHD States, see data section for details.

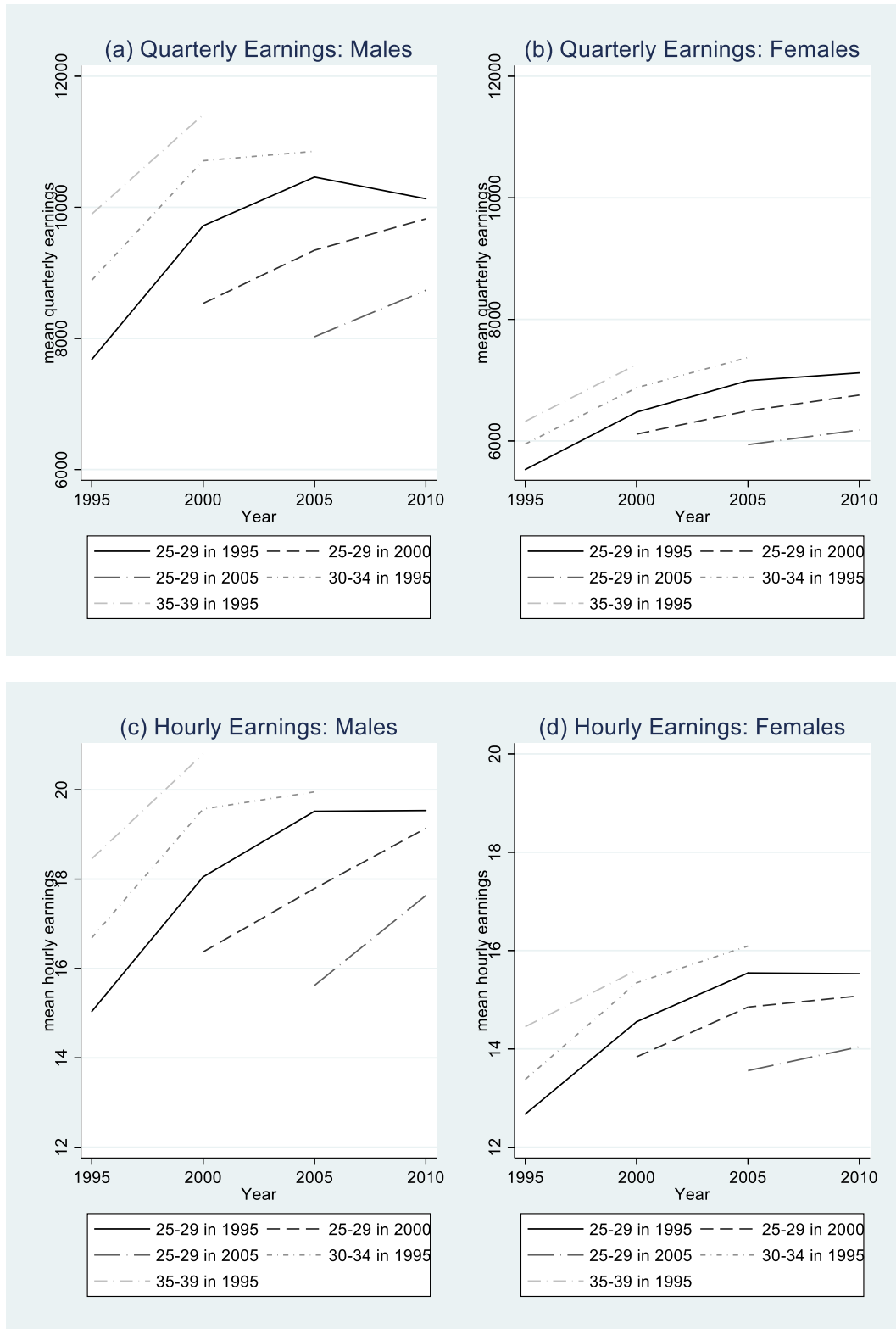
Appendix 1: Descriptive Statistics for the LEHD Sample

	In 2000 Sample mean	In 1995-2008 Sample mean
ln(earnings)	9.2122	9.1928
female	0.4449	0.4524
age	0.5785	0.5283
agesq	34.7137	35.6591
evermarried	0.6135	0.5715
college	0.3334	0.3112
ln(size)	0.7409	0.7002
jobage	1.9023	2.0445
age0	-4.4214	-5.9002
age00	-3.0976	-4.3840
year	2000.0000	2001.4290
Observations	1,176,400	15,043,800

Appendix 2: Current Population Survey (CPS) Based Analyses

To study the impact of hours worked on the age-wage profile by gender, cohort and educational attainment we use data from the March Current Population Survey (CPS) augmented with information on annual earnings from wage and salary, usual hours worked per week and weeks worked per year from the Annual Social and Economic Supplement (ASEC). The ASEC information is for the previous fiscal year. Therefore, in keeping with our main sample, we select individuals aged 26 to 45 in the years 1995 to 2012. We further select the sample to include only observations with non-missing values for all the variables of interest. Individuals are grouped in two education categories (College, No College) based on whether the higher level of education completed is a bachelor degree or more. We consider two earnings concepts: Quarterly earnings (annual earnings divided by four) and an hourly wage, which is obtained by dividing annual earnings by the product of usual hours worked per week and weeks worked per year. Earnings are converted in 2008 dollars using the Bureau of Labor Statistics Consumer Price Index (for all urban consumers) for the years 1995-2012. Averages by age-cohort-gender-education are computed using individual weights for the ASEC sample. In this appendix we compare the age profiles by earnings concept, gender and cohort for individuals who did not complete college. The statistics for college graduates are presented in Figure II in the main text.

Appendix Figure 2. Age-profiles by gender, cohort and earnings concept: No College



Notes: Data from CPS. See appendix 2 for details.

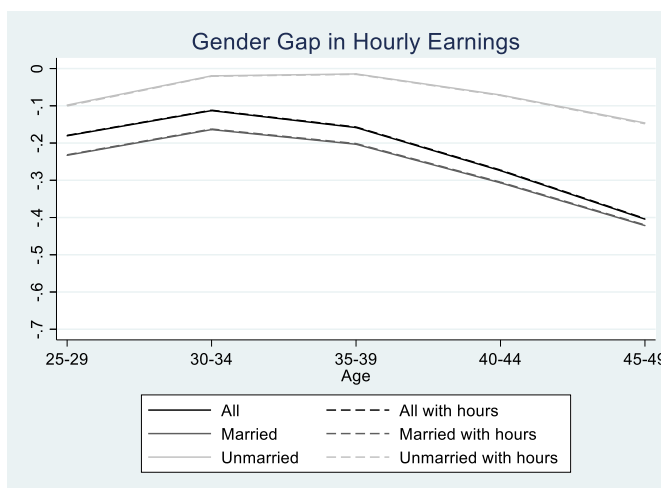
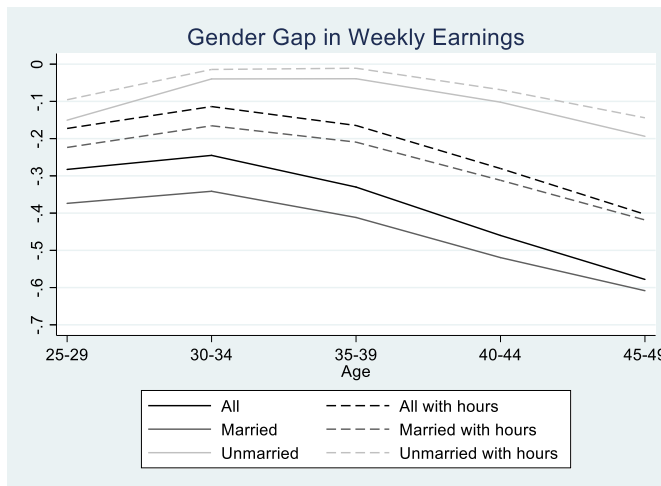
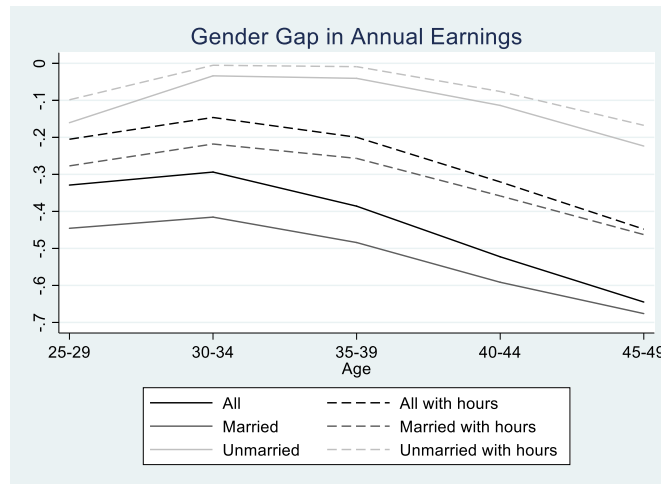
Appendix 3: American Community Survey (2001-2007) Based Analyses

To study the impact of hours worked on the age pattern of gender pay gaps, we use 2001-2007 American Community Survey (ACS) data. Those are the years for which weeks worked are available in a continuous (not categorical) variable in the ACS. We limit the sample to individuals aged 25-50, who are employed, work for wages, and earn at least \$2,000 per year. We drop individuals who have annual earnings greater than one million dollars per year. Just like in the main analyses reported in this paper, we separately analyze college educated individuals and those who do not have a college degree.

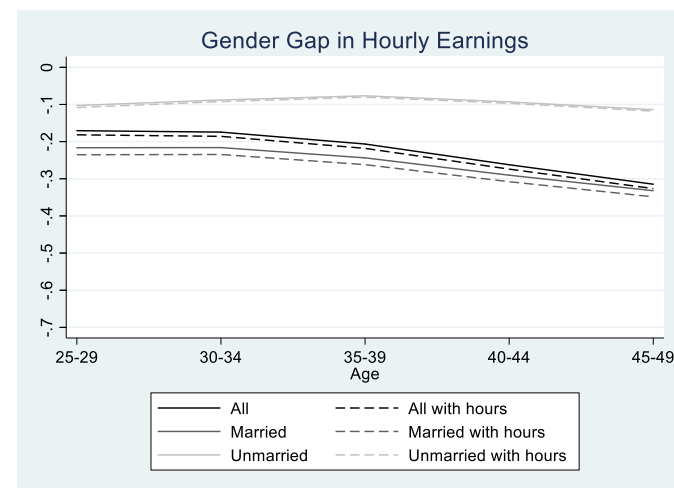
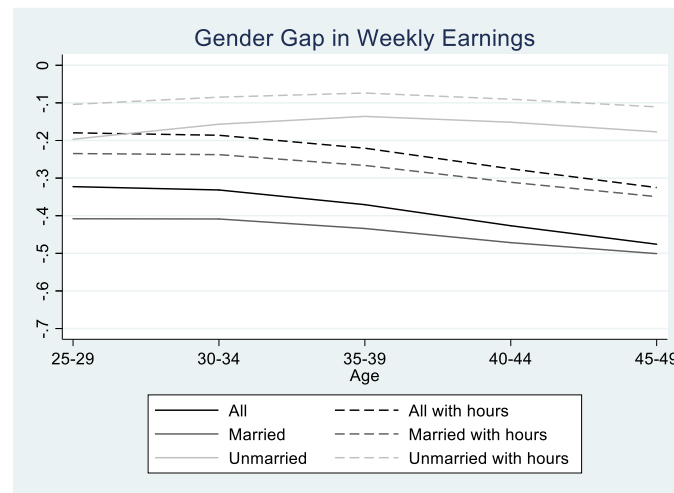
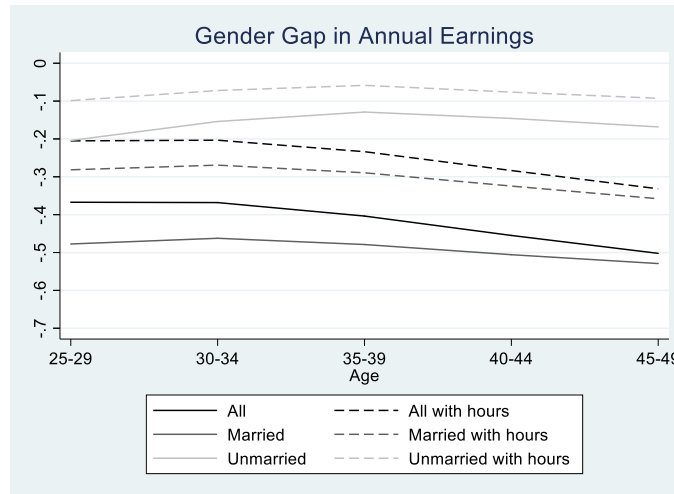
We use three main earnings concepts: annual earnings, weekly earnings, and hourly wages. All are expressed in natural logarithms. We estimate models where we interact gender with age group dummies, and control for time and state fixed effects. Since the ACS does not have information on the establishment each person works at we instead estimate a version where we also include controls for the 3-digit NAICS industry and 2-digit occupation. The figures below contrast the estimates for gender earnings gap when no control for hours worked is included in the models versus a version with the hours worked control. We report the results for the overall sample and by marital status (married, never married).

Not very surprisingly, one key observation is that the usual hours worked affect the estimated level of the gender earnings gap in annual earnings and weekly wages. Men work, on average, a greater number of hours than the average female employee, and this explains part of the gender pay gap. A much smaller impact from hours worked is evident in the hourly wage graph. However, regardless of the income concept used, the shape of the age pattern in the gender pay gap is *not* impacted by the inclusion / exclusion of the hours worked variable. This is true for both the college-educated and non-college educated sample. In other words, we see a widening gender pay gap with age, whether or not we control for the hours worked and regardless of the income concept adopted.

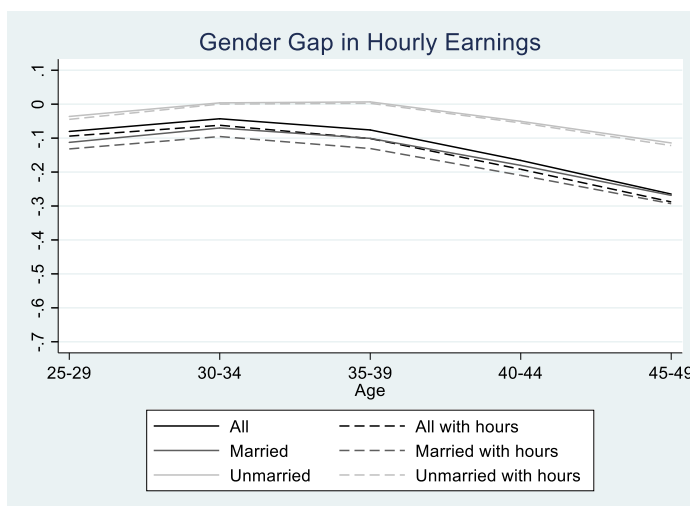
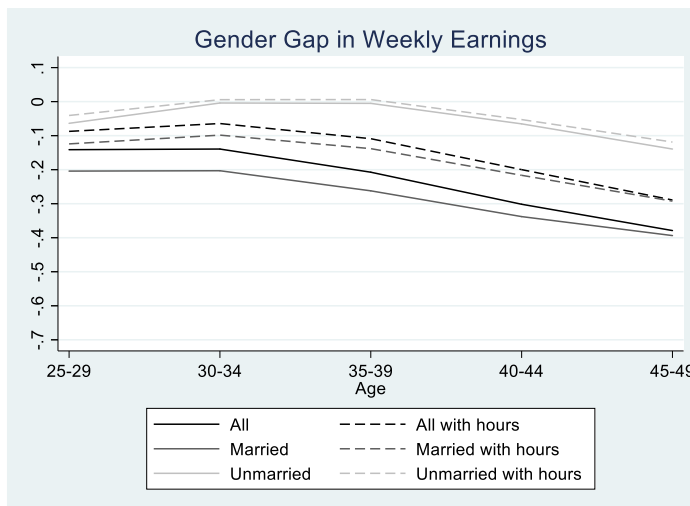
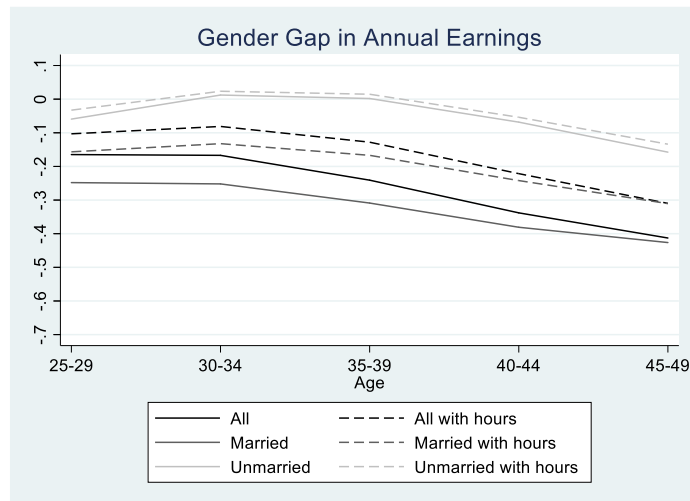
Appendix Figure 3A: Models without industry or occupation controls. College graduates



Appendix Figure 3B: Models without industry or occupation controls. No College



Appendix Figure 3C: Models with industry and occupation controls. College graduates



Appendix Figure 3D: Model with industry and occupation controls. No College

