THE DYNAMICS OF GENDER EARNINGS DIFFERENTIALS: EVIDENCE FROM ESTABLISHMENT DATA*

Erling Barth. Institute for Social Research, Oslo and NBER

Sari Pekkala Kerr, Wellesley College and NBER

Claudia Olivetti, Dartmouth College and NBER

February 2021
[First draft: October 31, 2016]

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 whether the widening of the gender pay gap arises from differences in career advances within the same establishment or from 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 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.

JEL codes: J16, J31

Keywords: gender earnings gap, establishment, wage age-profiles

*Barth: Institute for Social Research, Munthesgate 31, 0208 Oslo, Norway (e-mail: erling.barth@socialresearch.no) (phone: +4793091410). Kerr: WCW, Wellesley College, 106 Central Street, Wellesley, MA 02481 (e-mail: skerr3@wellesley.edu); Olivetti: Department of Economics, Dartmouth College, 6106 Rockefeller Hall, Hanover, NH 03755 (e-mail: claudia.olivetti@dartmouth.edu). 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. The research was conducted while Kerr and Barth were Special Sworn Status researchers of the US Census Bureau, Boston Census Research Data Center. Any views expressed are those of the authors and not those of the U.S. Census Bureau. The Census Bureau's Disclosure Review Board and Disclosure Avoidance Officers have reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 786. (CBDRB-FY20-P786-R8785). This research uses data from the Census Bureau's Longitudinal Employer Household Dynamics Program, which was partially supported by the following National Science Foundation Grants SES-9978093, SES-0339191 and ITR-0427889; National Institute on Aging Grant AG018854; and grants from the Alfred P. Sloan Foundation.

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 US[1] 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. We introduce a methodology for the identification and decomposition of gender gaps along the age-earnings profile into an individual component and an establishment component. In addition, our methodology separately identifies the contribution to the individual component of earnings growth from establishment-specific seniority, controlling for both individual- and match-specific fixed effects. The unique data match to the 2000 Decennial Census allows us to study how the widening of the wage gap varies by education, marital status, and occupation, something that typically cannot be done with matched data, at least in the United States, but that 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

[1] See Abowd, Creecy, 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.

(1995-2008), 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.

Figure 1 shows the predicted age earnings profiles by gender and education based on a quadratic specification, estimated on our large panel of workers observed over time and across jobs. [3] 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. 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 workers with no college degree 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 by 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 pay raises that workers at a given establishment may accrue over time. The second – the pay gap due to workers sorting into high- versus low-paying establishments – is driven by the presence of different wage levels

[3] A quadratic age-earnings profile seems to be reasonable in our case, since our sample is restricted to individuals, aged 25 to 45. As illustrated by Murphy and Welch (1990), the main reason to add higher order polynomials in earnings equations is to accommodate for deviating profiles (from a quadratic specification) for younger and older workers. Adding higher order terms for age does not change the predicted age-earnings profiles of Figure 1 in any discernable way. Note that our regressions are based on 15 million observations. 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.

across firms and gender-based variation in the likelihood that a worker will change jobs and receive a higher salary as a result of that job change.

The two processes are the results of very different mechanisms. Within-employer career paths may depend on differences in compensation 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).

Between-employer career paths arise from different wage levels across firms[5], 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 rents (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).

Here we quantify the importance of career movements within and across establishments using a full decomposition analysis of 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 differential wage raises and

[5] See Groshen (1988) for an early contribution.

[6] See Manning (2003) and Green, Machin, 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.

promotions in the same firm.[8] While we cannot distinguish between internal promotions and earnings growth within job category in our data, we do distinguish between differential seniority profiles (experienced in the current job, but lost when changing employer)[9], and differential development in the within-establishment 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.

Our analysis reveals the importance of selection for the widening of gender earnings gap, at least for some groups. Based on our specification, the gender earnings gap for non-college workers widens by 10 log points between age 25 and 35; it then narrows by 8 log points past age 35. In contrast, the widening of the predicted gender earnings gap in Figure 1, which is based on the OLS specification, is twice as large between age 25 and 35, and continues to increase, albeit at a slower pace between age 35 and 45. This is not the case for college educated workers. For this group, the gender earnings gap increases by about 40 log points between age 25 and 45, irrespective of the specification.

Among workers without a college degree, we find a small male advantage in the individual (or within-establishment) component of earnings growth between age 25 and 35. It explains 64 percent of the 10-log point gender gap widening over this period. However, over the next 10 years, women's within-establishment earnings growth catches up and surpasses that of men. Overall, non-college educated men and women have similar within-establishment earnings growth between age 25 and 45, 25 and 28 log points respectively.

Conversely, for college educated workers there is a substantial difference in the individual component of earnings growth by gender. Conditional on establishment fixed effect, the age-earnings profiles are generally much steeper than the age-earnings profiles of workers with no college. The gender differential is particularly striking and is the main driver of the widening gender pay gap for this group. For a college educated man, within establishment earnings grow by 54 log points from age 25 to 45, while they grow only by 23 log points for a college educated woman. The individual component can explain about 70 percent of the widening of the gender earnings gap for this group. Unlike the non-college educated, college educated women never catch up.

[8] Booth, Francesconi, and Frank (2005) use the British Household Panel Survey, and find that women are promoted at the same rate as men, but receive smaller wage gains with promotions. Blau and De Varo (2007) find that among new hires in US establishments women have lower probabilities of promotion than men, but associated wage gains 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).

The most novel results of our analysis concern the development of the establishment earnings premium over time. For employees without a college degree, the widening of the establishment component of the earnings premium between age 25 and 45 adds 5 log points to the overall gender gap in pay. This is the main source of the widening of the gender pay gap for this group. Were it not for the establishment component the gender pay gap would actually *decline*.

Among college educated workers, the gender gap across establishments widens by 11 log points from age 25 to 45. This accounts for about 30 percent of the overall widening of the gender wage gap for this group.[10] The importance of the establishment component increases over time. It accounts for 16 percent of the increasing gap between age 25 and 35 but for 41 percent of the additional widening after age 35.

The analysis by broad occupational categories characterized according to key traits related to gender career differences identified by Goldin (2014), reveals that the patterns by education can be explained by the type of occupations these different workers are in. The overall gender pay gap grows the most, 32 log points from age 25 to 45, in professional and managements occupations, while its growth is considerably lower, around 18 log points in service and sales occupations, and very modest, 4 log points, in production and transportation occupations. The within establishment pay gap is the largest driver of the widening gender gap in management and professional occupations, while in sales, and production and transportation occupations, it is the between establishment component that plays the major role. The great majority of workers in management and professional occupations have a college degree, while at least 80 percent of workers in sales, service and production occupations do not have a college degree.

Notably the more equal performance of women with no-college degree mainly reflects the lack of growth in less-educated men's earnings. Employees in this education group disproportionately work in occupations where climbing the career ladder within a firm does not play a big role. 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. After

[10] While it is beyond the scope of the current study, it is worth emphasizing that the cost of career interruption varies by gender and education, and could lead women to select jobs and occupations that allow career breaks without imposing a large wage penalty (e.g. Spivey, 2005; Bertrand, Goldin, and Katz, 2010). Those types of jobs typically have a much flatter age-earnings profile. The occupation-specific analysis below sheds some light on this issue.

age 40 the total earnings gap narrows for both groups as women catch up within establishments, while men keep their advantage across establishments.

Finally, we show that for college educated women marital status is an important determinant of the within establishment earnings growth. This is not the case for non-college women. However, marital status matters regardless of education for earnings growth due to the between establishment component: married women's earning power seems to benefit very little from changing jobs. 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. With the caveat that the results by marital status should not necessarily be viewed as causal, given the potential endogenous selection into marriage, especially among men (e.g. Nakosteen and Zimmer, 1987; Korenman and Neumark, 1991), we discuss in the concluding section how these observed patterns may arise from the interaction between reward structures within and across firms in different types of occupations and family obligations.

This paper is part of a rich and currently burgeoning literature. Blau (1977) was 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) uses 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 for explaining the cross-sectional wage gap in Spain (Amuedo-Dorantes and De La Rica,

[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.

2006), Denmark (Datta Gupta and Rothstein, 2005), Germany (Heinze and Wolf, 2009), Finland (Korkeamäki and Kyyrä, 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.

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

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 childcare, 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

[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.

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 employees' 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 Primary Metropolitan Statistical Areas (PMSA) in the U.S. in terms of population as of 1991, and estimate the main models separately by PMSA and schooling.[15] Of the 50 largest U.S. PMSAs, 26 were located in 18 of the 23 LEHD covered states available to 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, during the observation year, worked more than two quarters and earned at least $2,000 per quarter, on average. 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.[16] 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

[15] For the U.S. geographical classification, see https://www2.census.gov/geo/pdfs/reference/GARM/Ch13GARM.pdf

[16] 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.

household income by source). In the current study we mainly extract individual-level characteristics, that is, educational attainment, marital status and occupation. 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.

Table 1 provides basic descriptive statistics by gender (Panel A) and by occupation (Panel B). While men in our data are more likely to be married (the shares are 60 percent and 54 percent, respectively), women in the sample are slightly more likely to have a college education (32 percent vs. 29 percent). Women work in larger establishments[17], while men have higher earnings and a longer average tenure in their jobs. Men and women's distribution across occupations differ most notably in sales and office occupations and in production and transportation. Women are twice as likely as men to be employed in sales and office occupations (the shares are 33 percent and 16 percent, respectively), while men are more than twice as likely to work in production and transportation occupation (17 percent and 7 percent, respectively). The first row of panel B reiterates this point: 63 percent of workers in sales and office occupation are women, while only 25 percent of workers in production occupations are. On the other hand, professional and management occupation, as well as service occupations are evenly held by men and women. The share of workers with a college degree is very high in the professional and management occupations (over 60 percent), but much lower in service and office occupations (21 percent), services (11 percent) and transportation, and production occupations (about 6 percent). Occupational gender earnings differences are largest (about 40 log points) in the professional and service occupations, as well as the transportation and production occupations, the two occupation groups with the highest and lowest share of college graduates.

A possible concern with the LEHD is that it tracks quarterly employment and earnings but does not contain information on hours worked or hourly wage.[18] Therefore, in theory, age

[17] That women work in larger establishments may seem at odds with studies finding that men work in larger and more productive firms. Note, however, that there is a difference between establishment and firm size. A large firm may have many establishments, and there may be differences across male and female dominated industries, such as manufacturing and health, with respect to the number of establishments per firm. Note that Ln(size) is defined as deviation from the mean in our data.

[18] 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

or cohort trends 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 derive different earnings concepts, including annual, quarterly and weekly earnings, as well as hourly wages.

In Figure 2 we use CPS data to show age profiles by cohort, education and gender for college graduates measured as quarterly earnings and hourly wages (see Appendix 2 for details). It does not appear that differences in hours worked over the life cycle significantly 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 addition, in Appendix 3 we show estimated age patterns of the gender earnings gap, measured in terms of annual earnings, weekly earnings and hourly wages using the American Community Survey (ACS). Not surprisingly, the gender earnings gap is somewhat larger when measured in terms of annual earnings than when measured with hourly wages. However, the age patterns of the earnings gap are very similar regardless of the income measure (annual or weekly).[19] Controlling for hours in each equation gives the expected result: the level of the gender gap is reduced whether it is measured in annual or weekly earnings, but the gap's age dynamics is unaffected. Thus, although hours worked 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 if the movements out of the PMSA were differentially selected across genders, exacerbating potential biases due to selection into the labor force over the life cycle. Since our key results are obtained conditional on fixed individual and establishment effects, heterogeneity in terms of individual earnings ability or establishment

Lachowska, Mas, and Woodbury (2017) for an example of a paper using the WA linked employer – employee data with hours.

[19] The usual caveats regarding self-reported hours of work apply (Frazis and Stewart, 2004; Abraham, Spletzer, and Stewart, 1998; Robinson and Bostrom, 1994). Surveys on hours worked may suffer from response bias, be non-representative for salaried workers, and/or have reference weeks that are not representative of the typical monthly or annual average.

earnings levels (i.e. sorting) is controlled for in the analysis. However, attrition bias or selection bias could still affect our gender comparisons if there is differential selection in the residual earnings growth between men and women. Although this concern cannot be confronted directly based on our data, we note that both biases could go either away. For example, 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 of women may just as likely be from tied movers, women who follow their husband as they take better job opportunities, which would bias our findings in the opposite direction.

To assuage some of these concerns, at least partially, we perform two robustness analyses. First, we use ACS and CPS data to evaluate the representativeness of our PMSA sample relative to the entire United States. The age patterns in the gender earnings differentials are essentially unchanged as we limit the data to the LEHD states instead of the entire country.[20] In addition, the main analysis is repeated on an LEHD sample that includes only individuals who are present (in the labor force and in our PMSAs) for at least half of the years covered. The results are not discernibly different from those presented in Section III.

A few final details about the Census RDC data are worth noting. First, to disclose results based on the restricted access LEHD data, Census Bureau requires that no exact observation counts be reported. Instead, all observation counts (N) are rounded to the nearest round number (e.g. N=1,021 becomes N=1,000). 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.

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

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 over the whole time period. This normalization simply allows us to interpret the establishment fixed effect as a deviation from the average establishment. 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-invariant person and establishment characteristics are absorbed by those fixed effects.

II. METHODOLOGY

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

Our log earnings model includes a job-spell specific match component in addition to individual fixed effects, establishment fixed effects and time varying covariates usually included in fixed effects models in the style of Abowd, Kramarz, and Margolis (1999) - AKM hereafter. In our framework, a job is defined as a unique match between an individual and an establishment. Formally, we estimate the following fixed effects earnings equation:

$$(1) \qquad lnw_{it} = \underbrace{\alpha_i + \beta^g{}_1 Age_{it} + \beta^g{}_2 Age^2{}_{it}}_{individual\ component} + \underbrace{\beta^g{}_3 lnSize_{j(i,t)\,t} + \varphi^g{}_{j(i,t)}}_{establishment\ component:\ \chi} + \underbrace{\xi_{ij(i,t)}}_{match\ component} + \gamma_t + \varepsilon_{it}$$

where *Age* is the age of individual *i* at time *t*, α_i is the individual fixed effect, *lnSize* is the log of the number of employees at the establishment where individual *i* works at time *t*, measured as deviation from the sample mean, $\varphi^g{}_{j(i,t)}$ is the establishment fixed effect,[21] 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

[21] We assume fixed establishment and match effects over the full period. This may be a somewhat 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, which are outside the scope of this paper.

establishment characteristics (e.g. age and establishment size) as well as calendar time. The superscript *g* indicates parameters that differ by gender. The equation also includes year fixed effects, γ_t, and an error term, ε_{it}, assumed to be orthogonal to all the other variables.

The "individual component" of earnings describes the expected *within-job-spell age earnings* profile for a worker *i*, for given observable or unobservable characteristics (e.g. gender, birth cohort, education or ability) that are constant over time and across employers, and conditional on establishment and match-specific fixed effects.

The "establishment component" of earnings includes the establishment fixed effect and an earnings premium associated with establishment size: $\chi_{j(i,t)t} = \beta^{g}{}_{3} lnSize_{j(i,t)\,t} + \varphi^{g}{}_{j(i,t)}$. Note that this term can change over time both because of changes over time in the size of the establishment an individual works for (irrespective of whether he/she stays with the same employer) and because $\varphi^{g}{}_{j(i,t)}$ changes when a worker moves across establishments. To capture the age earnings dynamics of the establishment component at the individual level, we define the auxiliary regression for an individual *i* of gender *g* as a (quadratic) function of age as follows:

(2) $$\chi_{j(i,t)t} = b_0^g + b^{g}{}_{1}Age_{it} + b^{g}{}_{2}Age^{2}{}_{it} + u_{it}$$

Equation (2) defines the expected *age profile of the establishment component of earnings* for an individual *i* of gender *g*. That is, it describes the evolution of a worker's life cycle earnings stemming from changes over time in the size of the establishment he/she works for and from changes that occurs as he/she changes job, moving across establishments of different size and with different establishment fixed effects. This equation is estimated separately by gender.

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 $\varphi^{g}{}_{j}$ 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, at a given age, between the expected earnings of a woman and a man living in the same PMSA and with the same level of educational attainment:

(3) $$\Gamma_{Age} = E(\alpha_i|f) - E(\alpha_i|m) + \left(\beta^{f}{}_{1} - \beta^{m}{}_{1}\right)Age + \left(\beta^{f}{}_{2} - \beta^{m}{}_{2}\right)Age^{2} + E(\chi^{f}|Age) - E(\chi^{m}|Age).$$

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 defined above. Using the auxiliary regression (2), we decompose the change in the gender earnings gap between any two ages, for instance between age 25 and age 35, as:

$$(4)\qquad \Delta\Gamma_{Age} = \underbrace{\left(\beta^f{}_1-\beta^m{}_1\right)\Delta Age + \left(\beta^f{}_2-\beta^m{}_2\right)\Delta Age^2}_{change\ in\ the\ individual\ component} + \underbrace{(b^f{}_1-b^m{}_1)\Delta Age + (b^f{}_2-b^m{}_2)\Delta Age^2}_{change\ in\ the\ establishment\ compontent}$$

where Δ denotes the difference between the two ages. Equation (4) provides a decomposition of the changing gender earnings gap along the age-profile.[22] The first bracket of equation (4) reflects the change in the *individual component* of the gender earnings gap by age, arising from differential earnings growth by gender among workers who remain in the same establishment, conditional on establishment size. We explain below how both $\left(\beta^f{}_1-\beta^m{}_1\right)$ and $\left(\beta^f{}_2-\beta^m{}_2\right)$ may be identified from the estimation of (1) using fixed job-spell 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 changes in the *establishment component* of the gender earnings gap, arising from differential 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 a quadratic in age as outlined by equation (2).

It is important to note that the estimated widening of the gender earnings gap using a simple OLS regression, as in Figure 1, may be biased. In addition to the sum of the individual and establishment component of earnings change, the OLS estimates may include considerable selection effects related to both individual and establishment characteristics. A typical selection on individual characteristics would arise if older cohorts of women have a lower earnings potential than younger cohorts, for instance from lower labor market attachment among earlier generations. Another source of selection would be if women with low-wage characteristics (i.e. low labor market attachment) disproportionately re-enter the labor market at later ages (in their

[22] Note that the change in the earnings gap does not rely on differences in the expected individual fixed effects by gender or any fixed difference in the establishment components as reflected in the constant term of equation (2). These are constant over time and thus differenced out.

thirties and forties) after extensive periods of non-participation. Such negative correlations between labor market attachment and age would induce an upward bias in the widening of the gender earnings gap by age.

Similar selection issues may arise related to the establishment component. The distribution of establishments is constantly changing, for instance from structural and technological change, or from the evolution of business cycles. A typical selection on establishment characteristics would arise if younger cohorts get access to entry jobs in high paying firms with new technology or are in high demand. Such a negative correlation between establishment effects for each cohort and age would induce a downward bias in the age-earnings profile facing each individual. Whether such effects affect men and women the most is open for speculation, but it remains clear that they may contaminate the OLS estimates of the widening of the gender earnings gap as well.

Adding individual fixed effects to the estimated age profiles of the individual and establishment components controls effectively for such composition effects. Thus, differences in the estimated changes in the earnings gap over time between the OLS specification and the ones estimated from equation (4) are due to these composition effects. In the following, we outline the estimation strategy used to identify the key parameters in equation (4).

B. *Estimation: Identifying the Components of the Widening Earnings Gap by Age*

A key estimation challenge is related to the identification of age, cohort, and time effects. As it is well known, because of collinearity, it is not possible to identify separately linear birth-cohort, calendar time, and age fixed effects in a panel regression with individual fixed effects. Thus, we cannot identify the coefficients of the linear terms, ${\beta^g}_1$ directly for each gender, though we can identify the coefficients of the second order terms for age, ${\beta^g}_2$.

Importantly, we can also identify the *difference* in the linear terms, $\left({\beta^f}_1 - {\beta^m}_1\right)$ by assuming a common time effects γ_t across genders. Based on equation (4) this common calendar time trend assumption across genders ensures that we can estimate the *change of the individual component of the earnings gap*. Appendix I outlines the estimation procedure for $\left({\beta^f}_1 - {\beta^m}_1\right)$ in *Step 1.* This involves running a pooled regression for men and women where all the time varying covariates are interacted with a female dummy. The key point is that while a linear term for age is collinear with the time dummies, and thus excluded from the fixed job-spell specification, the interaction term between age and female is not collinear with time, and thus can be identified from the comparison of job spells across the genders.

The assumption of equal time effects across genders is key for identifying the widening of the individual component of 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.[23] The alternative, a common cohort effect across genders, seems questionable, given the well-documented gender convergence in the labor market outcomes across the 33 cohorts included in our study.[24] As shown by Goldin (2014), 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 the individuals 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), the gap 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 a 27-year old in 2005 would be wider than the earnings gap observed for a 27-year old in 2000, and so on. Furthermore, we would expect the earnings gap of all age groups between cohorts to be widened in parallel, providing a consistent ranking of cohorts by age. Appendix Figure 2B, which presents the development of the gender earnings gap over time for different birth cohorts, does not seem to support such a scenario. In the figure, each line represents the gap in average earnings for each cohort as they are observed in subsequent cross sections of the CPS. Cohorts are labelled by the year when they reach 25 to 29 years of age.

[23] 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).
[24] Our data comprises workers in age group 25-45 during the 13 years (1995-2008), and they are thus born between 1950 and 1983.

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 education. Second, it is not the case that the earnings gap, measured at a given age, is consistently smaller in the earlier cohorts. On the contrary, particularly among workers without college, the gender earnings gap is larger among the earlier cohorts. Ranking the gender earnings gap at age 27 for workers without college degree in ascending order, we observe the smallest gap for 27-year old in 2010 and 2005, followed by the older cohorts from 1995 and 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 are a noteworthy contributor to the widening of the gender earnings gap over the life course.

Consider next the estimation of the between-establishment component, $\chi^{g}{}_{jt}$. In *Step 2* we first retrieve the fixed job-spell effects from Step 1, and decompose them into an individual fixed effect, an establishment fixed effect and an orthogonal match specific component, using a standard AKM-decomposition. Next, we combine the term for the logarithm of establishment size from Step 1 with the estimated establishment fixed effects. This provides us with an estimate of the time varying establishment component of the earnings premium $\chi^{g}{}_{jt} = \beta^{g}{}_{3} lnSize_{j\,t} + \varphi^{g}{}_{j}$ for each observation in the data. As noted above, the establishment component changes over time both due to differential firm size growth 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 for individuals by age, we run the auxiliary regression (2) separately by gender and education, including individual fixed effects to account for fixed factors such as the quality of the person's entry-job into the labor market, which may vary between cohorts who enter the labor market under different conditions.

For both Step 1 and Step 2 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 for each education group[25], gender and PMSA, with the exception of the time effects that are assumed equal across genders within each education-PMSA cell, and establishment fixed effects that are assumed equal for each education group but allowed to vary across gender and PMSA. In addition to the orthogonality of the error term 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.[26]

C. *Extension: Age-Earnings Profiles by Gender and Returns 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 ${\beta^m}_1$; the baseline linear part of earnings growth for men. As shown in Appendix I, removing the linear term for age in equation (1) in Step 1 implies that the individual fixed effect is given by $\tilde{\alpha}_i = {\beta^m}_1 Age_{i0} + \alpha_i$, where Age_{i0} is the age at the beginning of the panel, and α_i the individual fixed effect in (1). Notably, α_i incorporates the cohort effects. In order to identify the linear age effect for men separately from a potential 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 decade.[27] We thus obtain an estimate of ${\beta^m}_1$ from the equation:

(5) $$\tilde{\alpha}_i = \alpha_c + {\beta^m}_1 Age_{i0} + \alpha_i$$

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-spell fixed effects in step 2 and regress them on the "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. In our view, an assumption of decade-by-decade cohort effects would be less realistic for women since the evolution of their labor force participation and pay has been more dramatic even within a ten-year birth cohort. We thus obtain an estimate of the linear part of women's earnings

[25] All references to education groups in this paragraph apply also for occupation or marital status when included in the analyses.
[26] See Card et al. (2016) for a discussion of this assumption in a two-way fixed effect model of workers and establishments.
[27] Note that since we include job-spell 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 are thus controlled for, and what we are addressing here is cohort effects on the distribution of earnings within firms.

profile $\beta^{f}{}_{1}$ by adding the estimated coefficient for males from (5) into the Step 1 estimate of the difference $\left(\beta^{f}{}_{1} - \beta^{m}{}_{1}\right)$.

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 age 40.[28] We believe that this assumption would be somewhat arbitrary in our application, since the linear part of women's earnings profile might peak at a later age than 40, due to career interruptions during the prime working age years. In fact, for the younger cohorts who postpone fertility, women could plateau around age 40 as they drastically reduce work hours and weeks worked because of childcare responsibilities (see Goldin, 2014). Our preferred estimates thus come from equation (5) and they imply that the earnings profiles reach their apex at 42 and 43 years of age for non-college and college educated men, respectively. This is not very different from the Card et al. assumption. It is noteworthy that even if this difference implies different slopes on the age earnings profiles, the development of the gender gap by age does not change across specifications, since gender difference in the age earnings profiles are obtained directly from the interaction terms in the fixed job-spell models in step 1.

Another issue 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 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-spell fixed effects. With the fixed effects, though, it is not possible to separate the linear terms for age and seniority within jobs, even with the assumption of common time effects across genders. 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. Specifically, we add the seniority at the beginning of the panel for each job-spell into the AKM-decomposition:

[28] In our case this would be done by calculating $\beta^{m}{}_{1}$= -10$\beta^{m}{}_{2}$, since we have normalized age to be zero at 35 in our data.

(6) $\check{\psi}^{g}{}_{ji} = \tilde{\alpha}_i + \beta^{g}{}_{s} S_{ij0} + \varphi^{g}{}_{j} + \omega_{ij}$

where the identification of $\beta^{g}{}_{S}$ comes from individuals who hold two or more jobs only (see Appendix I for details). Separate terms for seniority squared by gender may be identified directly in the within-job-spell specification in Step (1), as they are not collinear with age squared or the calendar year effects. The procedure is thus to run Step 1 including seniority squared terms, and then add the seniority at the beginning of the panel for each job-spell in the subsequent AKM-decomposition of the job-spell fixed effects (6).

To sum up: The widening of the within establishment gender earnings gap by age is estimated directly from a within job-spell 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. To obtain the gender specific age-earnings profiles within establishments that we show in the figures 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 with at least two different jobs in the panel.

III. RESULTS

We start discussing the results by education. Table 2 summarizes the results by gender and education, comparing the OLS estimates of the gender gap to those obtained based on our model in equation (4). Figure 3 displays the individual and establishment age-earnings profiles based on our decomposition.

III A. The OLS specification exaggerates the widening of the gender earnings gap by age for non-college workers, but not for college workers.

The results from the OLS specification underlying Figure 1 are presented in columns (1) and (2) in Table 2. Consider first workers without a college degree. The estimated gender earnings gap increases by 24.6 log points from age 25 to 45. The gap increases the most between age 25 and 30, by 10.6 log points. The gap then grows at a decreasing rate with a minimum growth of 1.7 log points between ages 40 and 45. In Figure 1 this is visible from the differential growth in predicted earnings by gender in the earlier years, while both profiles flattens after age 40.

Column (3) presents the estimated change in earnings gap based on equation (4). For non-college workers the difference with respect to the OLS estimates is striking and shows the importance of selection. The estimated gap controlling for fixed effects still widens by 10.2 log points between age 25 and 35. However, there is a considerable catching up of women's earnings, by 7.9 log points, between age 35 and 45. The difference between the OLS and the fixed effects estimates, suggests that the OLS specification understates the earnings growth of women relative to men in later years. As discussed in the methodology section, this could be the case if the group of working women in their thirties and forties includes a disproportionate number of low labor market attachment individuals, with extensive periods of non-participation due to childcare responsibilities. This pattern may also be produced by increasing cohort effects for women over time, which may be particularly prominent for women without a college degree. Women in older cohorts may have a lower earnings potential than women belonging to the younger cohorts, possibly due to lower labor market attachment. This in turn may skew the estimated OLS age-earnings profiles for women downwards. Such a scenario is supported by the observation in Appendix Figure 2B, where we study the gender earnings gap observed in the CPS for 7 different 5-years cohorts from age 25 to 45. For non-college educated workers, there is a clear pattern of a larger earnings gap among the older cohorts relative to the younger cohorts. It may also be the case that the entry jobs available for older cohorts of women were to a greater extent in lower paying establishments, than those available to younger cohorts.

The pattern is very different for college educated workers. Here the OLS shows a widening of the gender earnings gap by 43.8 log points, while the gap widens by 41.6 log points in the fixed effects model. For college educated workers, there appears to be less correlation between the fixed part of the individual and establishment components of earnings and age than for non-college workers. One possible explanation is that college educated women postpone fertility. In this case we would be more likely to see selection effects in the forties and fifties (and indeed in Table 2 we see the OLS and fixed effect estimates starting to diverge between age 40 and 45). Another explanation is that the labor market attachment of older and younger cohorts of women (in terms of age span and time period) differs less for college educated than for non-college educated, simply because older cohorts of non-college educated workers include a larger share of women with marginal labor market attachment, while this pattern is no longer true for college educated workers. Again, this scenario is supported by the observation in Appendix Figure 2B, where there seem to be no consistent pattern in the gender earnings gap between different cohorts. We now proceed to consider the decomposition of equation (4) into an individual and an establishment component.

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

The upper panels of Figure 3 show the predicted age-earnings profiles obtained from separate regressions with fixed individual and establishment effects as outlined above. The models also include year effects. The figures illustrate the evolution of the predicted individual component of earnings over time for the same worker, keeping the contribution from the establishment component unchanged, relative to a 25-year-old of the same gender and education 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.

Once again, we find a much steeper profile for college educated workers, particularly for men. An average college educated man improves his earnings by 54 log points from age 25 to 45, while a college educated woman improves her earnings only by 23 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 as the earnings profile for men flattens out. By age 45, women without a college degree have more than caught up to similarly educated men. Their earnings grow by 28 log points over the 20-year period, while men's earnings increase only by 25 log points.

Column (4) in Table 2 quantifies the change of the gender gap in the individual earnings component within establishments for different age intervals. Among the non-college educated, we find that men have faster earnings growth within establishments until age 35, but women catch up in terms of earnings growth after that. By age 45 their accumulated earnings growth is 3.1 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 point advantage in accumulated earnings growth relative to women in the same education group in terms of the individual component.

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

The establishment component of earnings is the sum of the fixed establishment effect and the earnings premium that is attributable to firm size. It measures an establishment earnings premium, or how much more (less) the employer pays an individual worker over and above the average establishment in the economy. We estimate the relationship between a quadratic in age and the establishment component as outlined above. The widening of the gender gap in the establishment component as individuals age provides another key finding: 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. The bottom panels of Figure 3 show the development of the establishment earnings component by age, compared to a 25-year old person with the same gender and education.

We find a rising age-earnings profile in the establishment component for both college and non-college educated 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 educated men experience a growth of 17 log points, whereas the number for women is 12 log points. Column (5) of Table 2 shows the widening of the 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 evenly spread across age intervals.[29] The gender gap in the average establishment earnings premium increases by 5.4 log points from age 25 to 45 for non-college educated workers and 11.2 log points for college educated workers.

Consider next the relative contribution of the individual versus the establishment component to the total widening of the gender gap. Among workers without a college degree, the total earnings gap widens from age 25 to 30 by 7 log points, and from age 30 to 35 by 3 log points. The share arising from an increasing establishment earnings premium is 29 and 55 percent, respectively. From age 35 onwards, the establishment component continues to widen the gap, but the total change is dominated by the larger catching up by women in the individual component. For college educated workers, the total gender earnings gap increases in all age intervals up to age 40, totaling 42 log points. The share of the establishment contribution increases from 13 percent during the first 5 years to 41 percent between age 35 and 40. Also

[29] The numbers are the same for each age interval as the coefficients for the second order term of the quadratic in age are estimated to be the same for men and women for this education group.

among college educated workers, the establishment component continues to widen the gap after age 40 but is offset by a catching up in terms of the individual component within establishments.

Up to age 40, about 56 percent of the accumulated gender gap in earnings growth for non-college educated, and about 20 percent 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. After age 40, the total earnings gap narrows for both groups as women catch up with men within establishments, which more than counteracts the continued disadvantage across establishments.

The differences between the age-earnings patterns of high and low-skilled labor markets are striking. The remuneration schemes within establishments for college educated workers are characterized by delayed payment, of which women appear to receive less. At the same time the between establishment component of pay appears to benefit men the most, regardless of education. In the next section, we show how these patterns arise within occupations with different educational requirements.

III.D. The individual component differs greatly across occupations

The LEHD does not offer any details on individual jobs, beyond the number of quarters worked, quarterly earnings, and the type of establishment. On the other hand, the Decennial Census long form tells us, for each individual, the 3-digit Census occupation they held in 2000. As it is not practical to generate decomposition analyses at that level of occupational detail, we chose to combine occupations based on their task composition, as done e.g. by Acemoglu and Autor (2011) and Goldin (2014). We used the O*Net database for information on occupational tasks and a crosswalk that links the Census occupational classification to the Standard Occupational Classification (SOC), which is the basis of the O*Net.

Based on Goldin's 2014 analysis, we consider five O*Net characteristics related to gender career differences. These include time pressure, maintaining interpersonal relationships, contact with others, freedom to make decisions, and how structured the work is. We normalize each characteristic to have a mean of zero and a standard deviation of one. Using regression analysis, we identified 6 groups of occupations that were distinctly different from each other along those five dimensions. As two of the identified categories were either very sparsely covered by the large PMSAs (Farming, Fishing, and Forestry Occupations), or employed very few women (Construction, Extraction and Maintenance Occupations), those were excluded from the following analyses. Instead, we chose to focus on four groups of occupations: 1) Professional, Management, and Related Occupations, 2) Sales and Office Occupations, 3) Service Occupations, and 4) Production, Transportation, and Material Moving Occupations. The LEHD sample by occupation group is summarized in Panel B of Table 1. The O*Net regression analysis results are in Appendix Table 4A. As can be seen in the table, Production and Transportation occupations score far below the others on most of the five measures (except for time pressure) and in some cases the differences are almost one standard deviation lower. In other words, in comparison with the other occupations, those in production have far fewer client and worker contacts and far fewer working relationships with others. On the other hand, management and professional occupations have far more need for establishing and maintaining interpersonal relationships, have more structured jobs and more discretion in making decisions than any of the other occupations.

The estimation of the individual and establishment components follows the same procedure as for education. It should be noted that occupation here is considered a "fixed trait" in the sense that we only observe it once (in 2000) for each individual. In reality, and in our panel data, people may of course shift across occupations, but that is likely less prevalent across the broad groups utilized here. It is nevertheless a feature of our data that should be kept in mind when interpreting the results.

Table 3 summarizes the results by occupation category, and Figure 4 displays the individual and establishment age-earnings profiles by gender and occupation. Several interesting patterns emerge from the occupation analysis. First, the overall gender gap grows the most in the professional and management occupations (32 log points between age 25 and 45), while the growth is quite modest in the production and transportation occupations (only 4 log points). This reflects the typical career paths available in those occupations, and the fact that men are more likely to be promoted in the professional and managerial jobs. Indeed, 67 percent of the overall growth in the gender gap can be attributed to the individual component in those types of occupations. For sales and service occupations, about 55 percent of the overall growth stems from the individual component, while the rest comes from the differential mobility and sorting between establishments.

Interestingly, the establishment component is solely responsible for any change in the gender gap in the production and transportation occupations, while the individual component actually reduces the gap. In all occupation groups, the gap begins to narrow in the 40-45 age range due to the individual component. As Figure 4 shows, the narrowing comes from the flattening of the male age profile of the individual component, while the female age profile remains more or less linear. The narrowing is most notable in the service occupations where the female age profile of the individual component is essentially flat and the male profile peaks at a relatively low level before age 40. No catch-up happens in the establishment component for any of the occupations, though.

These results highlight the importance of different career paths between occupations driving the evolution of the gender earnings gap by age. Occupations with high educational requirements such as professionals, managers, and related occupations are characterized by career ladders and delayed payment schemes within establishments, most likely due both to extensive continued learning on the job, and to various agency problems and incentive schemes arising when effort, performance, and ability are difficult to measure. In occupations with lower educational requirements, training requirements may also be lower, performance may be easier to monitor, and payment schemes more in line with spot labor markets. For example, the age-earnings profile of women in service occupations is strikingly flat.

On the contrary, the age-profiles of the establishment component display similar patterns across occupations. The processes of job-to-job mobility appear to be similar, even if the individual components within establishments are rather different. This suggests that the underlying mechanisms, most likely various frictions in the labor market, are quite similar regardless of the skill content of jobs or internal remuneration schemes. We also note that the mechanisms producing a widening of the gender pay gap across establishment seem to occur across the board as well.

III. E. The role of marital status

We now move to investigate the role of marital status in explaining the widening of the gender earnings gap. For data availability reasons, "married" is defined by marital status in the year of 2000 and considered to be a fixed attribute of the individual in our analysis. Since the marital status pertains to 2000, it is quite possible that some non-married individuals marry after 2000, while some married persons divorce. This would tend to bias the observed differences downward.[30] Table 4 provides the main results by marital status. Columns (1) and (4) report

[30] The median age at first marriage in the United States has remained relatively constant in 1995-2008: 24.5 to 26.0 for women, and 26.9 to 28.0 for men (U.S. Census Bureau), and the gap between college educated and non-college educated practically disappeared (Pew Research Center, based on Decennial Censuses and 2008 American Community Survey (ACS) Integrated Public Use Micro Samples (IPUMS)).

the total change in gender earnings gap by 5-year interval, and the total widening from age 25 to 45. For both college and non-college educated workers, the gender earnings gap widens considerably more when comparing married individuals than when comparing individuals who are not married. For all groups we find a pattern of widening earnings gap during the earliest years, then catching up in the later ages, but the gap is larger among married workers, and the catching up comes later and is weaker.

The upper panel of Figure 5 illustrates the profiles for the individual earnings component within establishments. Consider college educated workers first. Non-married men have a considerably lower earnings growth during the career than married men. Non-married women have a slightly steeper age-earnings profile up until age 35, crossing the 10-log point line at age 29, while married women cross it at age 32. However, married women appear to catch up relative to non-married women after age 35 and have a significant advantage by age 45 (30 log points versus 16, relative to own group at age 25). All in all, the gender earnings gap increases significantly more for married workers than for non-married ones in the early stages of the career, and by age 45 the gender earnings gap has widened twice as much for the married workers relative to non-married.

For workers with no college the differences between men and women are smaller both among married and non-married workers, although married workers of both genders have a steeper career profile than their non-married counterparts, ending up at an increase over 30 log points (versus 15 log points for the non-married). Again, married women display a convex age-earnings profile, while non-married women show a concave profile.

The lower panel of Figure 5 shows the age profiles of the establishment component. Consider again the college educated workers first. While the age-earnings profile across establishments for men is quite similar regardless of marital status, the age-earnings profile of married women is significantly more compressed than for the non-married. While non-married women have gained 13 log points by age 45, married women have only gained 7 log points. The difference is most pronounced in the early career; non-married women cross 5 log points at age 29 while married women do so first at age 35. Workers without a college degree display a similar pattern: married and non-married workers chiefly differ in terms of women's age-earnings profiles. Non-married women display a steeper age-earnings profile than married women, and the earnings gap widens more among non-married workers.

To sum up: The main driver behind the widening of the gender gap in the individual component within establishments is the steeper career profile of college educated married men relative to college educated married women. Married women with college education show an earnings growth similar to married women without college education. The main difference is that married women show a convex growth pattern with some catching up after age 35, while non-married women show a concave profile with steeper growth before age 35. Men show a steeper growth early in the career within all groups. An important caveat here is that one reason for the non-married women's slowdown in earnings growth after age 35 could be an artifact of our data. We observe marital status in 2000 so women classified as non-married might actually marry past year 2000. Likewise, the main driver behind the widening of the gender gap in the establishment component is the slower career growth across establishments for married women. This is the case for both education groups. While marital status seems to matter very little for men, it actually matters more than the level of education for women.

It is unclear if the results with respect to career profiles of married versus non-married college educated men are due to selection or "home resources". The difference between married and non-married women is consistent with an unequal distribution of home tasks, in particular in connection to children (between ages 25 and 35). It is feasible that both career advances (for the college educated workers) and job-to-job changes (for all groups) get constrained by the division of household-work, childcare activities in particular.

Seniority profiles

As discussed in the methodology section, the individual earnings component, conditional on the establishment earnings premium, can further be decomposed into a component purely due to seniority within a firm, i.e. an establishment specific earnings growth that is lost when a person changes to a new establishment, and a component of individual earnings growth that is retained also when changing jobs. For instance, when a person low in the earnings hierarchy of one establishment changes jobs to a position higher up in the earnings hierarchy of another establishment. Traditionally, a distinction is made between the returns to experience, reflecting general human capital accumulation that is retained across firms, and the returns to seniority, reflecting job specific human capital or firm specific remuneration or promotion schemes that is lost when changing jobs.[31]

[31] In our set up, return to experience arises from a combination of the age profile of the establishment component, and the remaining change in the individual component, conditional on experience.

We estimate a specification including a quadratic seniority term in addition to age, as outlined in the methodology section. To illustrate the results, Figure 6 shows the predicted earnings for men and women in a scenario where individuals change jobs every 5 years. It is important to note that the prediction is made conditional on the establishment component, so the figure shows, by age, the development of the individual component of earnings only.

Among college educated workers, the seniority profiles contribute considerably to the widening of the gender earnings gap. In particular, in the beginning of the career, up to age 35, the seniority slopes of married men are significantly steeper than those of women. But, we also note that the underlying age profile regardless of seniority is steeper for married men than for married women with a college education: men improve their relative position within the firm more than women when changing jobs. The difference between married and non-married men lies mostly in the returns to age, conditional on seniority. When changing jobs, married men appear to jump to a higher within-job trajectory, while non-married men gain very little in terms of relative position within the new firm as they age.

Among non-college educated workers, a similar development of the individual components between men and women is visible as well. The convexity of the age profile for married women is clear both in terms of the returns to seniority and to age; they lose relative to men in the beginning of the career, but gain after age 35. For non-married workers, women's age profiles remain rather constant, while the men lose out both in terms of the return to seniority and age after 35.

To sum up: In the beginning of the career, the age and seniority profiles are steeper for men than for women, and dominate 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 non-married men, dips to negative, and contributes to a narrowing of the gender earnings pay gap within establishments.

III.F. 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. [32]

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.[33]

IV. DISCUSSION AND CONCLUSIONS

The gender pay gap widens by age even when we control for selection on individual observed and unobserved characteristics. The widening occurs along two dimensions: Differential earnings growth along the internal earnings ladders within establishments, and differential earnings growth from job mobility across establishments with different levels of pay. Slicing up the earnings-profiles along these dimensions reveals key mechanisms underlying the development of the gender pay gap by age.

We find that within establishments, the widening of the gender pay by age is occurring entirely among college-educated workers. Conditioning on individual, establishment, and job match fixed effects, college-educated men have significantly steeper age-earnings profiles than

[32] The trade-off is that the sample size shrinks considerably the stricter the criteria used.

[33] 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.

women, whereas for non-college educated workers, the within establishment age-earnings profiles are less steep and display very little divergence by gender.

College educated workers are typically found in career-oriented jobs characterized by competition over promotions and delayed compensation schemes (Lazear 1981), and by what Goldin (2014) identifies as "non-linear (convex) payment structure with regard to hours worked". In contrast, non-college educated workers are more often in jobs with a linear payment structure with regard to hours in labor markets that look more like "spot-markets". Within establishments, the gender pay gap widens more in occupations with certain characteristics. In particular, it widens more in occupations with less "temporal flexibility", "substitutability", and more "autonomy" according to Goldin's (2014) classification.

There are striking differences by marital status between the relative earnings growth within establishments of men and women. The widening of the gender pay gap by age among college-educated workers occurs between married, rather than between non-married, men and women. In fact, married college educated workers appear to be the group behind almost all of the observed widening of the gender pay gap within establishments by age in our data. Their typical career paths within establishments require competitive efforts. As emphasized by Albanesi and Olivetti (2009), the utility cost of effort is likely to be increasing in home hours. Specialization within the household, available for married couples, and the greater productivity of home hours arising from parenthood, create wedges in the marginal utility cost of effort between married men and women relative to non-married men and women. Consistent with such a view, there is a large difference between married and non-married workers in the degree of widening of the gender gap in college-level labor markets, characterized by non-linear reward structures in terms of hours, whereas there is very little difference in the earnings profiles of men and women regardless of marital status in for non-college educated workers.

Between establishments, the picture looks very different. The establishment component contributes to earnings growth by age for both college and non-college educated workers. For men, there is very little difference between the age-earnings profile regardless of their education or marital status, implying that different career patterns within firms and other job characteristics are not key to the evolution of earnings across firms, and that the mechanisms of search and frictions that generate the rising age-earnings profiles across establishments are similar across occupations. For women, however, there are large differences between married and non-married workers. While married women display a significantly flatter age-earnings profile across establishments than men, non-married women display much more similar age-earnings profiles to those of men. This difference adds up to a considerable contribution to the widening of the gender pay gap by age for college educated workers, and to the entire widening of the gender pay gap by age for non-college educated workers.

Our conjecture is that the marginal utility costs of effort, in this case search costs and the efforts and uncertainties related to beginning a new job, are also rising with home hours, creating a similar divide by marital status as the one we observed for the internal careers of college workers. In this case, however, the mechanism operates regardless of occupation or structure of remuneration.

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.

Korenman, Sanders, and David Neumark. (1991) "Does Marriage Really Make Men More Productive?" *Journal of Human Resources*, 26(2):282-307.

Korkeamäki, Ossi, and Tomi Kyyrä. (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.

Murphy, Kevin M. and Finis Welch (1990). Empirical Age-Earnings Profiles. *Journal of Labor Economics, 8*(2): 202-229.

Nakosteen, Robert A., and Michael A. Zimmer. (1987) “Marital Status and Earnings of Young Men: A Model with Endogenous Selection.” *The Journal of Human Resources* 22(2), 248–268.

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 Economics Review Papers and Proceedings*, 107(5):384-387.

Spivey, Christy. (2005) “Time off at What Price? The Effects of Career Interruptions on Earnings.” *Industrial and Labor Relations Review,* 59(1): 119–140.

Taylor, Mark. (2007). “Tied Migration and Subsequent Employment: Evidence from Couples in Britain.” *Oxford Bulletin of Economics and Statistics*, 69(6): 795-818.

Table 1: Descriptive Statistics for the LEHD Regression Sample

Panel A: Overall Sample

	Men		Women	
	Mean	St. Dev.	Mean	St. Dev.
Quarterly earnings (logs)	9.33	0.72	9.00	0.64
Age	35.62	5.88	35.43	6.04
Share ever married	0.60	0.49	0.54	0.50
Share with College	0.29	0.45	0.32	0.47
Establishment size (logs)	0.55	2.80	0.89	2.91
Tenure	2.17	2.64	2.01	2.51
Share in:				
Professional/Management occupations	0.32	0.46	0.38	0.48
Service occupations	0.10	0.30	0.12	0.33
Sales/Office occupations	0.16	0.37	0.33	0.47
Production/Transport occupations	0.17	0.38	0.07	0.25
Other Jobs (n.e.c.)	0.08	0.27	0.09	0.29
Number of observations	7,304,000		6,080,000	

Panel B: Four Broad Occupational Categories

	Professional & Managerial		Sales & Office		Services		Production & Transport	
	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.	Mean	St. Dev.
Share Female	49.99%		63.27%		50.74%		24.48%	
Share with College	60.18%		21.48%		11.28%		5.62%	
Quarterly earnings (logs)								
Males	9.67	0.75	9.31	0.72	8.99	0.64	9.10	0.54
Females	9.25	0.65	8.95	0.56	8.67	0.55	8.70	0.50
Establishment size (logs)	1.06	2.94	0.50	2.92	0.62	2.77	0.65	2.61
Number of observations	4,397,000		3,082,000		1,427,000		1,622,000	

Notes: Sample of men and women aged 25 to 45. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education, occupation and marital status are obtained from the 2000 Decennial Census long-form data. See data section for details.

Table 2: Gender Earnings Gap Dynamics and Decomposition Results by Education

	(1)	(2)	(3)	(4)	(5)
	No-College				
	OLS estimates		Fixed Effect Decomposition		
	Gender Gap (logs)	Change (logs)	Change (logs)	Individual Component	Establishment Component
25	0.142				
30	0.249	0.106	0.073	0.052	0.021
35	0.325	0.077	0.029	0.013	0.016
40	0.372	0.047	-0.017	-0.028	0.011
45	0.388	0.017	-0.062	-0.068	0.006
Change 25-45		0.246	0.023	-0.031	0.054
	College				
	OLS estimates		Fixed Effect Decomposition		
	Gender Gap (logs)	Change (logs)	Change (logs)	Individual Component	Establishment Component
25	0.124				
30	0.276	0.152	0.209	0.181	0.028
35	0.400	0.124	0.139	0.111	0.028
40	0.495	0.095	0.069	0.041	0.028
45	0.562	0.067	-0.001	-0.029	0.028
Change 25-45		0.438	0.416	0.304	0.112

Notes: The first column reports OLS estimates of the gender earnings gap by age calculated from predicted values from separate regressions by education and gender (see notes to Figure 1). The second column reports the log change between subsequent ages (30 to 25, 35 to 30 etc.) A positive (negative) number indicates an increase (decrease) in the gender earnings gap. Column (3) to (5) report the results of the fixed effect decomposition in Equation (4). The log change in the predicted gender earnings gap between any two ages is in column (3). Column (4) and (5) reports individual and establishment component, respectively. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education is obtained from the 2000 Decennial Census long-form data. See data section for details.

Table 3: Fixed Effect Decomposition of Gender Earnings Differential by Age: Broad Occupational Categories

	(1)	(2)	(3)	(4)	(5)	(6)
	Professional and Management Occupations			Sales and Office Occupations		
	Change (logs)	Individual Component	Establishment Component	Change (logs)	Individual Component	Establishment Component
25-30	0.178	0.143	0.035	0.133	0.106	0.027
30-35	0.112	0.083	0.029	0.073	0.051	0.022
35-40	0.046	0.023	0.023	0.012	-0.004	0.016
40-45	-0.020	-0.038	0.017	-0.048	-0.059	0.011
25-45	0.315	0.210	0.105	0.170	0.094	0.076
	Service Occupations			Production and Transportation Occupations		
	Change (logs)	Individual Component	Establishment Component	Change (logs)	Individual Component	Establishment Component
25-30	0.116	0.097	0.019	0.054	0.035	0.020
30-35	0.069	0.052	0.017	0.024	0.009	0.015
35-40	0.022	0.007	0.015	-0.006	-0.016	0.010
40-45	-0.024	-0.038	0.014	-0.035	-0.041	0.005
25-45	0.183	0.118	0.065	0.037	-0.012	0.049

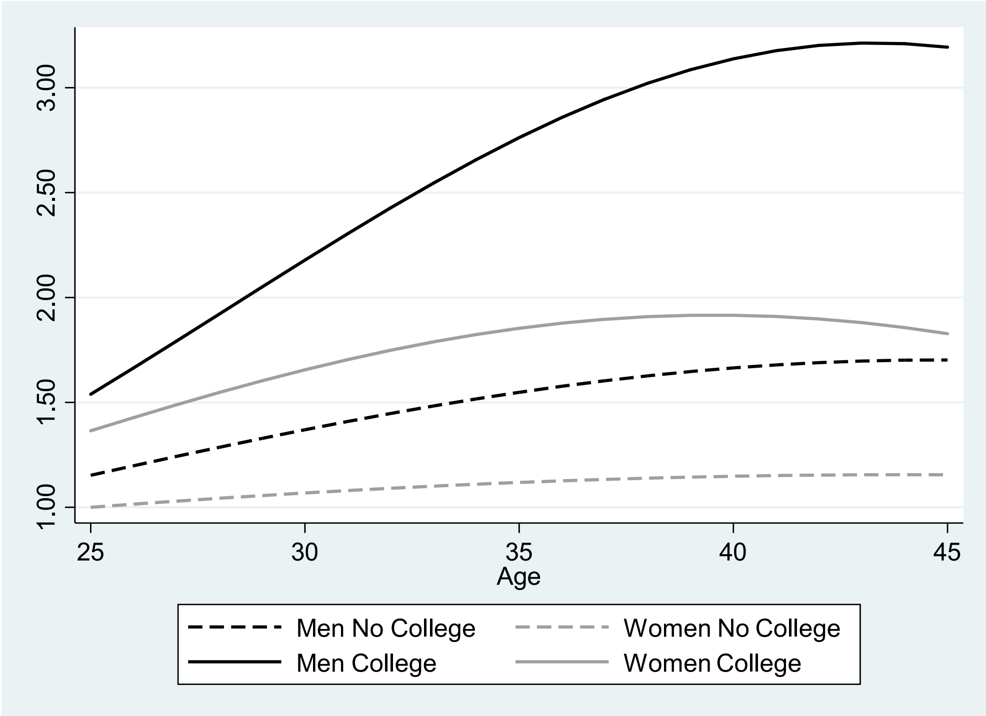
Notes: Entries are based on the fixed effect decomposition in Equation (4) by broad occupational categories. Column (1) and (4) report the log change in the predicted gender earnings gap between any two ages (30 to 25, 35 to 30 etc.) Column (2) and (5) reports the individual component, while column (3) and (6) the establishment component. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education and occupation are obtained from the 2000 Decennial Census long-form data. See data section for details.

Table 4: Fixed Effect Decomposition of Gender Earnings Differentials by Age: Education and Marital Status

	(1)	(2)	(3)	(4)	(5)	(6)
	Married					
	College			No College		
	Change (logs)	Individual Component	Establishment Component	Change (logs)	Individual Component	Establishment Component
25-30	0.289	0.238	0.051	0.088	0.062	0.026
30-35	0.185	0.144	0.041	0.038	0.017	0.021
35-40	0.081	0.051	0.030	-0.012	-0.029	0.016
40-45	-0.024	-0.043	0.020	-0.063	-0.074	0.011
25-45	0.532	0.390	0.142	0.051	-0.024	0.075
	Never Married					
	College			No College		
	Change (logs)	Individual Component	Establishment Component	Change (logs)	Individual Component	Establishment Component
25-30	0.143	0.123	0.021	0.056	0.047	0.009
30-35	0.093	0.075	0.018	0.021	0.014	0.008
35-40	0.043	0.028	0.015	-0.013	-0.020	0.007
40-45	-0.007	-0.020	0.012	-0.048	-0.053	0.006
25-45	0.272	0.206	0.066	0.017	-0.012	0.029

Notes: Entries are based on the fixed effect decomposition in Equation (4) by education and marital status (as observed in year 2000). Column (1) and (4) report the log change in the predicted gender earnings gap between any two ages (30 to 25, 35 to 30 etc.) Column (2) and (5) reports the individual component, while column (3) and (6) the establishment component. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education and marital status are obtained from the 2000 Decennial Census long-form data. See data section for details.

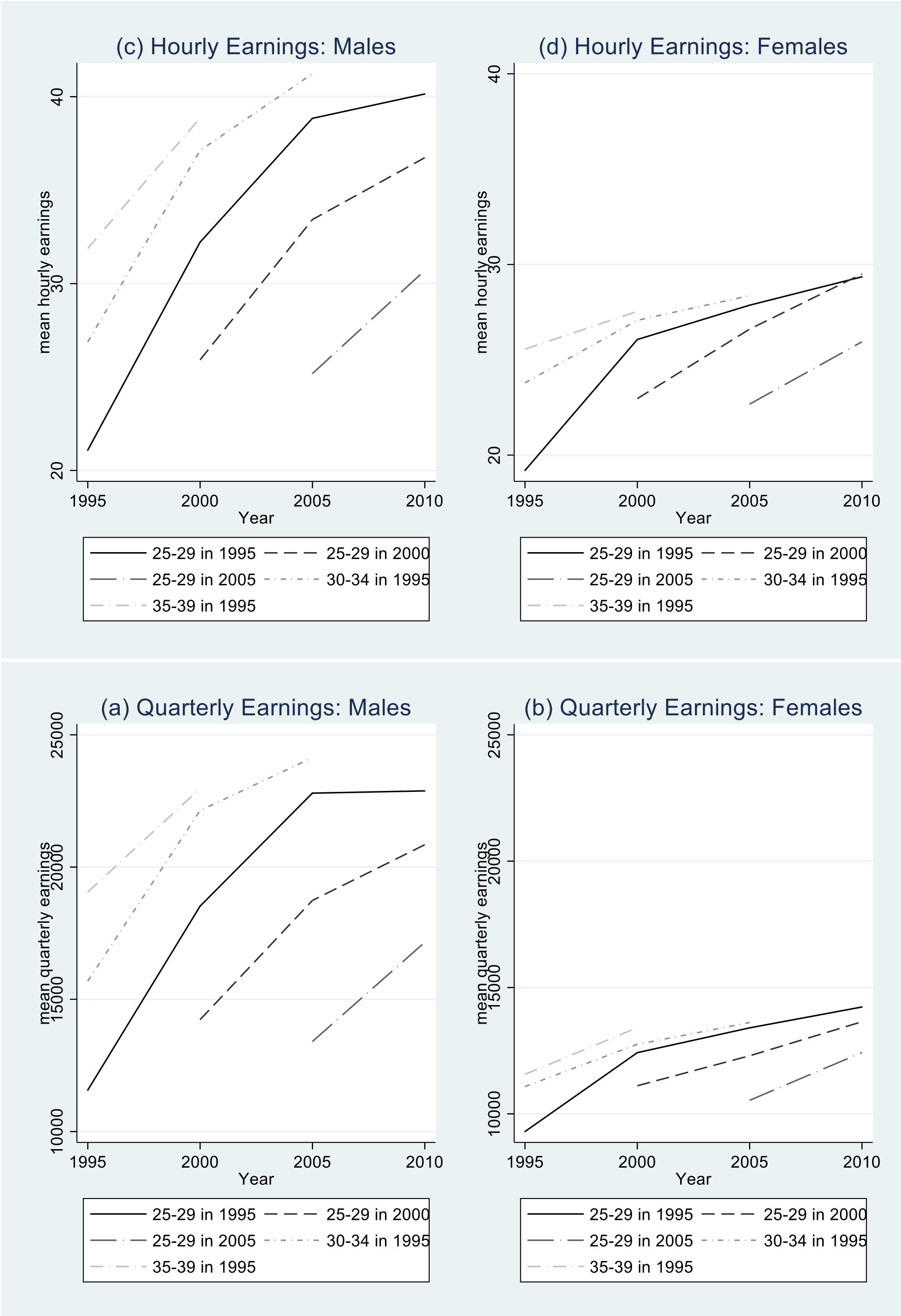
Figure 1: Predicted Relative Age-earnings Profiles by Education and Gender.
Reference: 25-year-old woman without a college degree



Notes: The lines show the relative earnings normalized by the earnings of a 25-year-old woman with no college degree. 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. *Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education is obtained from the 2000 Decennial Census long-form data. See data section for details.

Figure 2. Age-Earnings Profiles by Gender, Cohort and Earnings Concept:

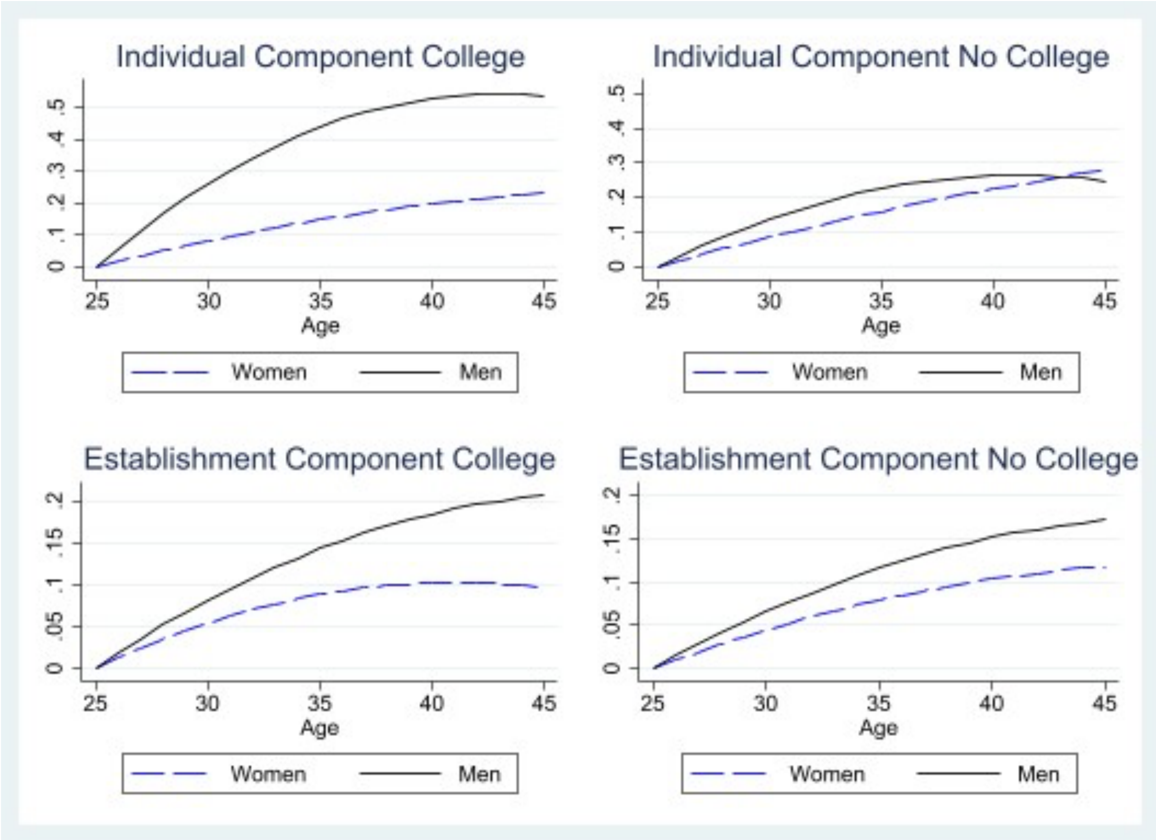
College Graduates



Notes: Average earnings by age-cohort-gender-education computed using individual weights for the ASEC sample. *Source:* 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). See appendix 2 for details.

Figure 3. Age-Earnings Profiles: Individual and Establishment Components

Log differences by gender and education.



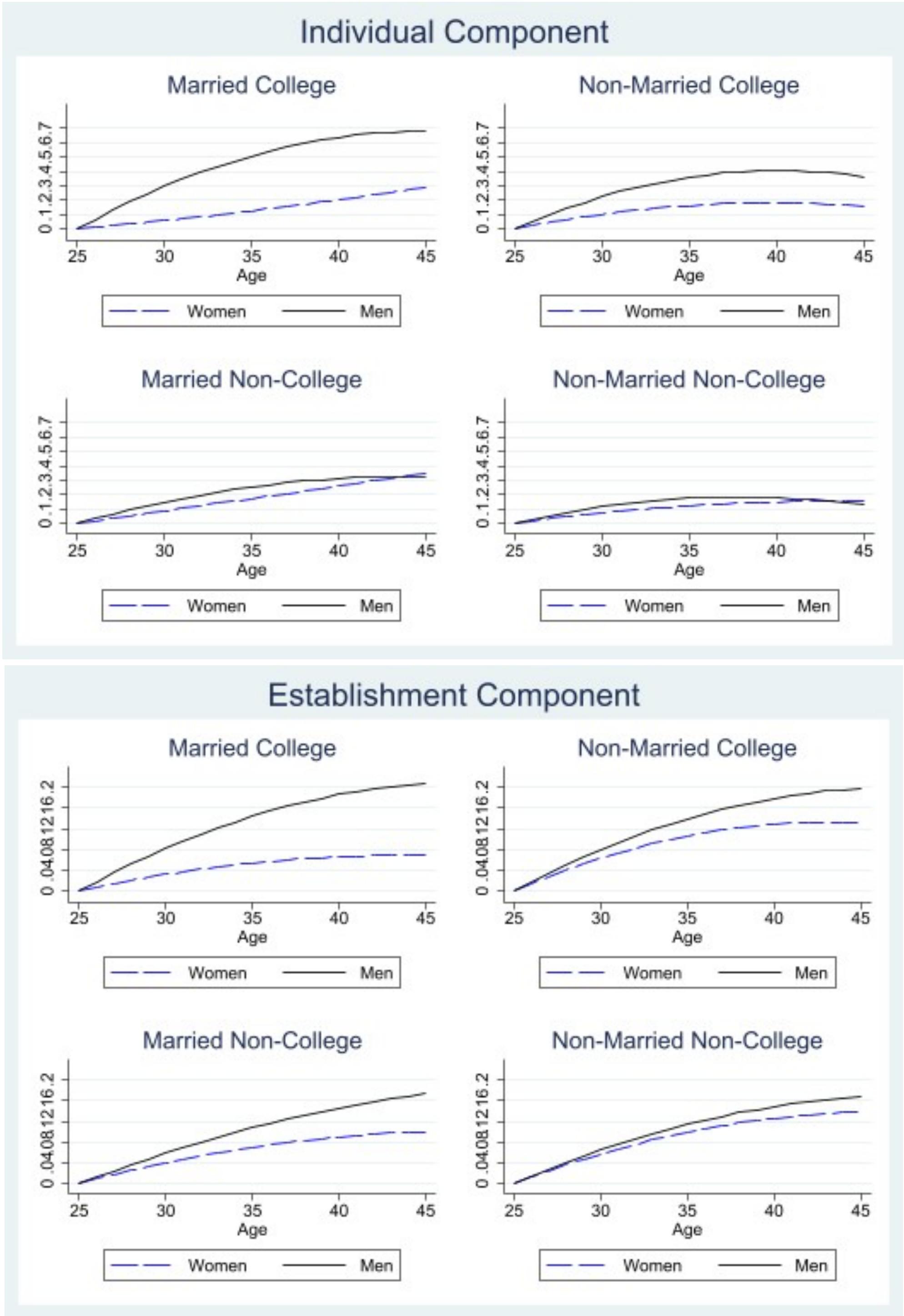
Notes: Log differences relative to a group-specific 25-year-old. Note that the scales are different between the upper and lower panels. *Individual Component*: Predicted values from separate regressions by education (equation 1). The models include job-spell fixed effects and the log of establishment size with gender interactions. The models also include individual fixed effects. See Section 3 for details. *Establishment component*: Predicted values from separate regressions of establishment earnings premiums for each gender and educational level (equation 2). The establishment earnings premium includes the fixed establishment effect and the earnings premium due to firm size. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education is obtained from the 2000 Decennial Census long-form data. See data section for details.

.

Figure 4. Age-Earnings Profiles by Occupation

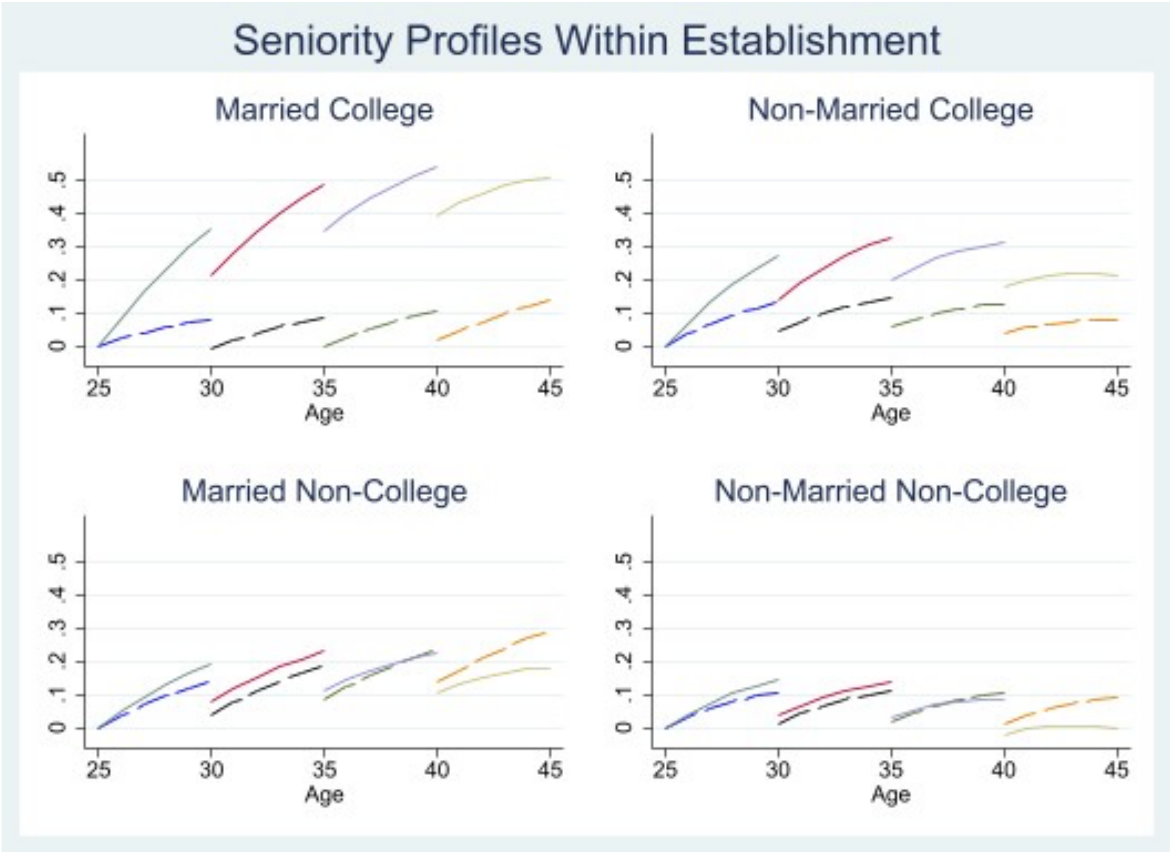
Notes: Log differences relative to group specific 25-year-old. *Individual Component*: Predicted values from separate regressions by education (equation 1). The models include job-spell fixed effects and the log of establishment size with gender interactions. The models also include individual fixed effects. See Section 3 for details. *Establishment component*: Predicted values from separate regressions of establishment earnings premiums for each gender and educational level (equation 2). The establishment earnings premium includes the fixed establishment effect and the earnings premium due to firm size. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education and occupation are obtained from the 2000 Decennial Census long-form data. See data section for details.

Figure 5. Age-Earnings Profiles by Marital Status



Notes: Log differences relative to group specific 25-year-old. Marital status defined as of year 2000. *Individual Component*: Predicted values from separate regressions by education (equation 1). The models include job-spell fixed effects and the log of establishment size with gender interactions. The models also include individual fixed effects. See Section 3 for details. *Establishment component*: Predicted values from separate regressions of establishment earnings premiums by gender and education (equation 2). The establishment earnings premium includes the establishment fixed effect and the earnings premium due to firm size. *Data Source*: Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education and marital status are obtained from the 2000 Decennial Census long-form data. See data section for details.

Figure 6. The Individual Component of Earnings with Establishment Specific Seniority Profiles. Example with establishment change every 5 years.



Note: Log differences relative to group specific 25-year-old. Marital status defined as of year 2000. Each line follows relative expected earnings from the beginning of a new employment relationship until its end after 5 years. The level differences between the lines beginning points show the development of expected earnings at the beginning of an employment relationship by age. Each line combines returns to both age and seniority within the job-spell, while the level differences between the end of a job and the beginning point of a new job show the seniority premium associated with 5 years seniority in the old job. Seniority profiles estimated from equation (6) (linear term) and equation (1) (w/second order term), and age profiles derived from equation (5) and (6) (linear term) and (1) (second order term). The models also include job-spell fixed effects and the log of establishment size with gender interactions, see Section 3 for details. *Data Source:* Matched LEHD sample from 26 of the top-50 largest PMSAs in the US that are located in LEHD-covered states. Education and marital status are obtained from the 2000 Decennial Census long-form data. See data section for details.

Appendix I: Details of the estimation method

The estimation of our earnings model described in equation (1) requires several steps.

Each of these steps is described in detail below.

a) *Step 1: Job-spell fixed effects*

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

(A1) $$lnw_{it} = \beta^{g}{}_{1}Age_{it} + \beta^{g}{}_{2}Age^{2}{}_{it} + \beta^{g}{}_{3}lnSize_{j\,t} + \psi^{g}{}_{ij} + \gamma_t + \varepsilon_{it}$$

where $\psi^{g}{}_{ij} = \alpha_i + \varphi^{g}{}_{j} + \xi_{ij}$ is the job-spell fixed effect. We estimate (A1) 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, for instance for Age: $\beta^{m}{}_{1}\,Age_{it} + \left(\beta^{f}{}_{1} - \beta^{m}{}_{1}\right) FAge_{it}$.

Since age is perfectly correlated with calendar year time for each individual, we exclude the linear part of Age (the first term with the coefficient $\beta^{m}{}_{1}$) from the equation[34]. We may however, retain the interaction term between gender and Age, which is not collinear with calendar year 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 may be written as: $\tilde{\gamma}_t = \gamma_t + \beta^{m}{}_{1}(Age_{it} - Age_{i0}) = \gamma_t + \beta^{m}{}_{1}t$, 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 may be written as: $\tilde{\alpha}_i = \beta^{m}{}_{1}Age_{i0} + \alpha_i$, 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:

(A2) $$lnw_{it} = \tilde{\beta}_0 + \left(\beta^{f}{}_{1} - \beta^{m}{}_{1}\right) FAge_{it} + \beta^{m}{}_{2}Age^{2}{}_{it} + \left(\beta^{f}{}_{2} - \beta^{m}{}_{2}\right) FAge^{2}{}_{it} + \beta^{m}{}_{3}\, lnSize_{jt} + \left(\beta^{f}{}_{3} - \beta^{m}{}_{3}\right) FlnSize_{jt} + \tilde{\psi}^{G}{}_{ji} + \tilde{\gamma}_t + \varepsilon_{it} ,$$

[34] 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 may be 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.

The coefficient $\tilde{\beta}_0$ will be determined by a normalization where the job-spell fixed effects are set to zero for the average person in the estimation sample, and where the year effect is set to zero at the beginning of the panel.

The estimation of equation (A2) 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).

b) Step 2: The establishment component

In the second step we use the parameters from Step 1 to obtain an estimate of the establishment component, $\chi^{g}{}_{jt}$. To begin the process, we retrieve the fixed job-spell 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:

(A3) $$\tilde{\psi}^{G}{}_{ji} = \tilde{\alpha}_{i} + \varphi^{G}{}_{j} + u_{it},$$

estimated on the full sample of observations by gender within the PMSA[35]. Combining the term for the logarithm of firm 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 $\chi^{g}{}_{jt} = \beta^{g}{}_{3} lnSize_{j\,t} + \varphi^{g}{}_{j}$ for each observation in the data. To obtain estimates of $b^{g}{}_{s}$ (s=1,2) in equation (4), describing the dynamics of the establishment component of earnings, $\chi^{g}{}_{jt}$ by age, we run the auxiliary regression (2) separately by gender and education.

c) Estimating the effect of seniority on earnings

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 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. In order 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

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

simplify the following exposition, we ignore all variables that vary within each job, such as age squared and firm size. Consider the simplified within-job earnings equation where we allow for a separate effect of the number of years in the current job-spell:

(A4) $$lnw_{it} = \beta_0 + \beta^g{}_1 Age_{it} + \beta^g{}_s S_{ijt} + \psi^g{}_{ij} + \gamma_t + e_{it},$$

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[36]. We may then rewrite (A4) as follows:

(A5) $$lnw_{it} = \beta_0 + \left(\beta^f{}_1 + \beta^f{}_s - \beta^m{}_1 - \beta^m{}_s\right) FAge_{it} + \check{\psi}^g{}_{ij} + \tilde{\gamma}_t + e_{it},$$

Where $\tilde{\gamma}_t = \gamma_t + (\beta^m{}_1 + \beta^m{}_s)t$, $\check{\psi}^g{}_{ji} = \tilde{\alpha}_i + \beta^g{}_s S_{ij0} + \varphi^g{}_j + \xi_{ij}$, and $\tilde{\alpha}_i = \beta^m{}_1 Age_{i0} + \alpha_i$. It is clear that we can identify the linear term $\left(\beta^f{}_1 + \beta^f{}_3 - \beta^m{}_1 - \beta^m{}_3\right)$ only. However, w may identify $\beta^m{}_s$ and $\beta^f{}_s$ from the gender specific equations:

(A6) $$\check{\psi}^g{}_{ji} = \tilde{\alpha}_i + \beta^g{}_s S_{ij0} + \varphi^g{}_j + \omega_{ij}$$

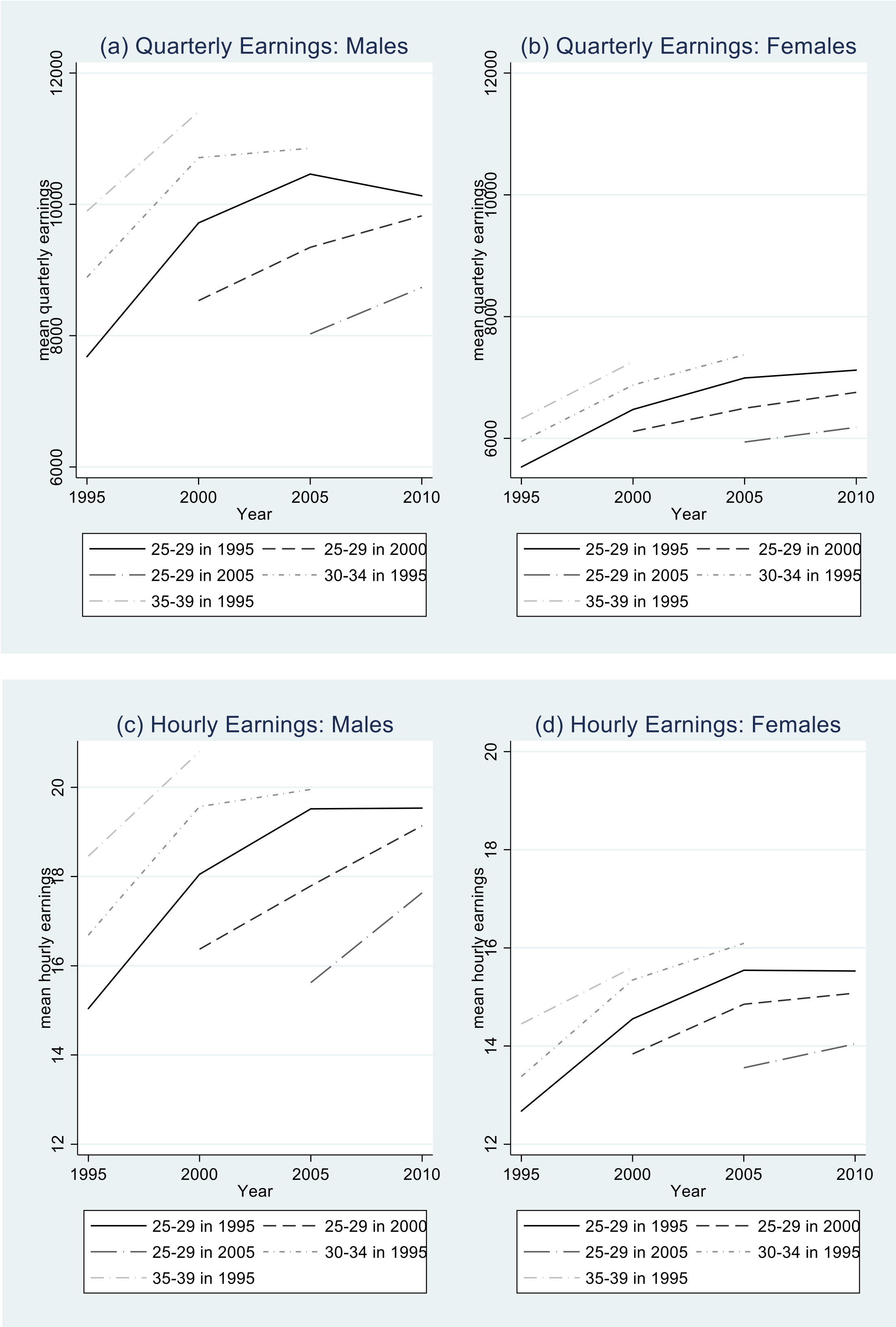
Since S_{ij0} is constant for individuals holding only one job, identification of $\beta^g{}_s$ comes from individuals who hold two or more jobs only.

[36] Note that for jobs starting after the first year in the panel, S_{ij0}<0.

Appendix 2: Current Population Survey (CPS) Based Analyses

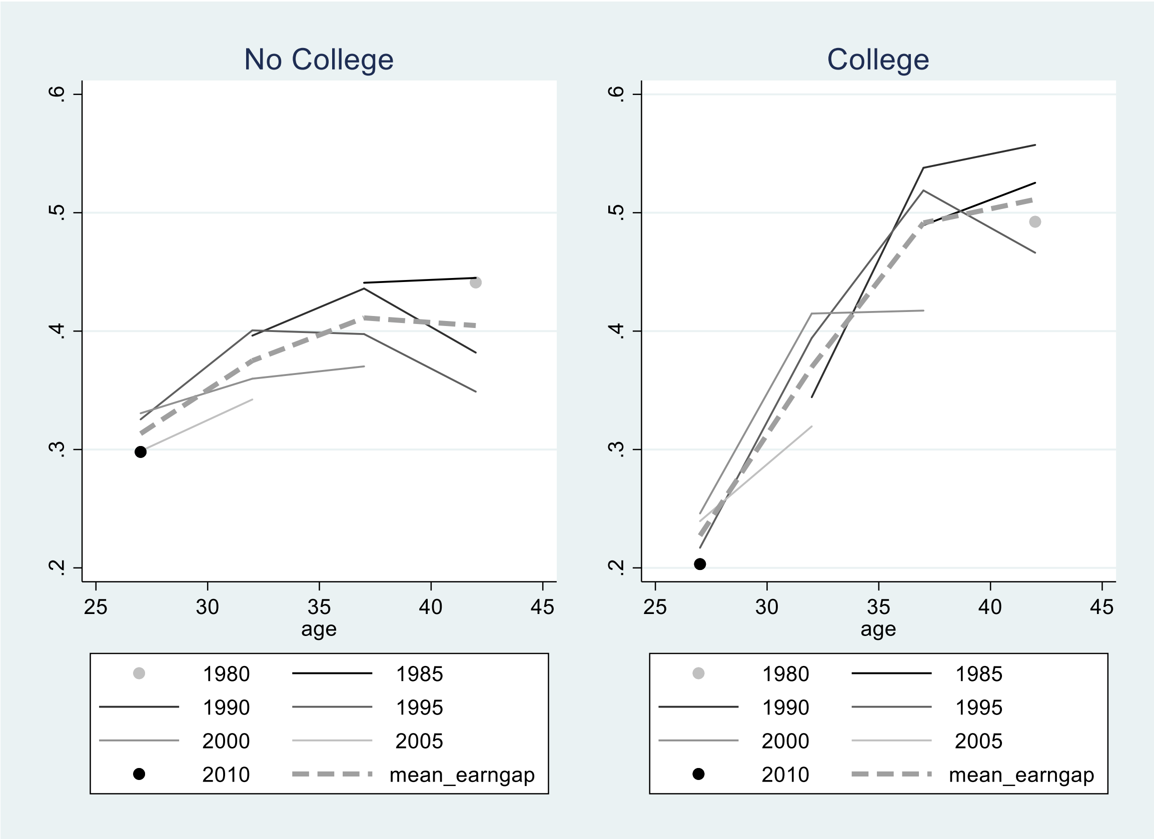
To study the impact of hours worked on the age-wage profile by gender, cohort and educational attainment we use data from the March Current Population Survey (CPS) augmented with information on annual earnings from wage and salary, usual hours worked per week and weeks worked per year from the Annual 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 appendix Figure 2A 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 2 in the main text. We also show the age profiles for the gender difference in log quarterly earnings by cohort and education in appendix Figure 2B, which is discussed in Section IIB.

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



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

Appendix Figure 2B: Gender earnings gap age-profiles by cohort and education



Note: Difference between men and women in log quarterly earnings by age, cohort, and education. Averages calculated using individual weights from the ACEC sample. *Source:* 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).

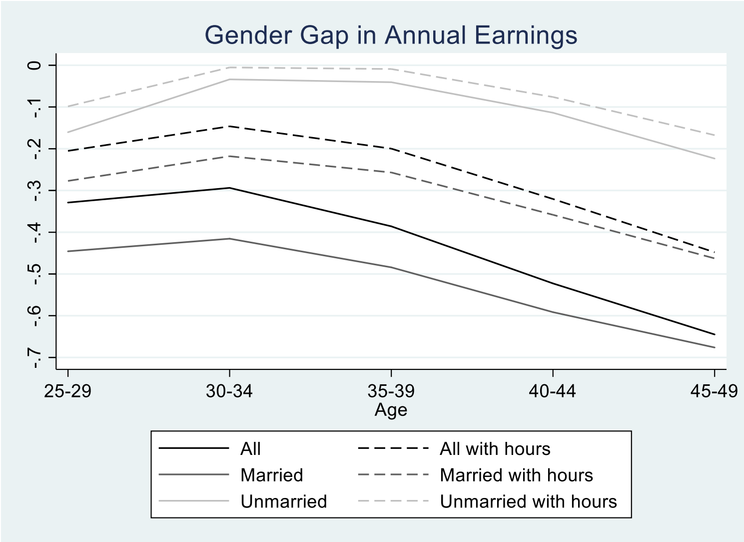
Appendix 3: American Community Survey (2001-2007) Based Analyses

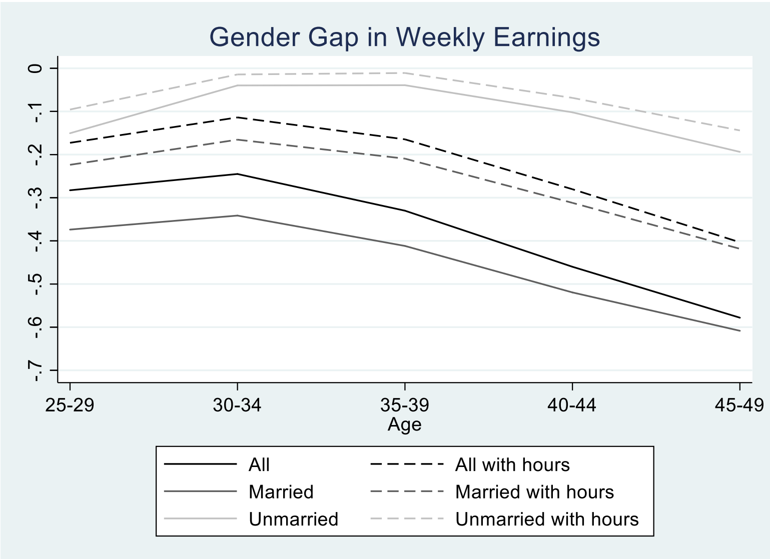
To study the impact of hours worked on the age pattern of gender pay gaps, we use 2001-2007 American Community Survey (ACS) data. Those are the years for which weeks worked are available in a continuous (not categorical) variable in the ACS. We limit the sample to individuals aged 25-50, who are employed, 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.

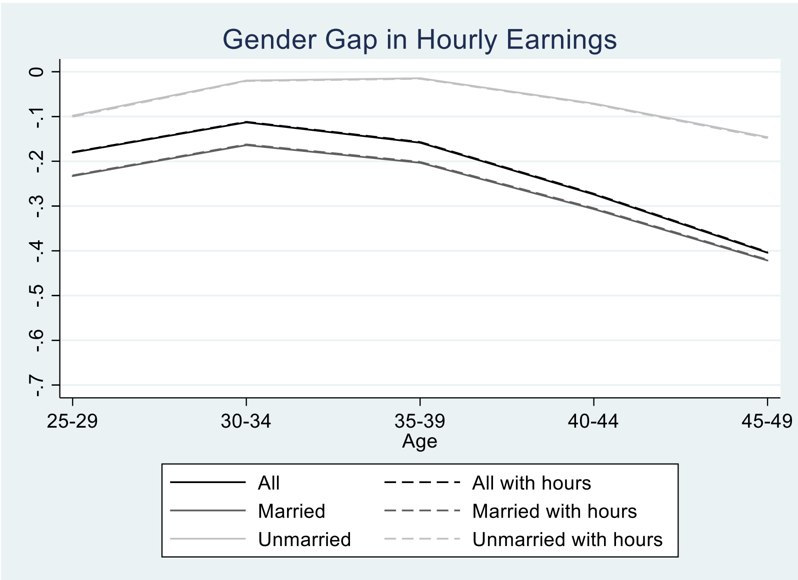
We use three main earnings concepts: annual earnings, weekly earnings, and hourly wages. All are expressed in natural logarithms. We estimate models where we interact gender with age group dummies, and control for time and state fixed effects. Since the ACS does not have information on the establishment 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.

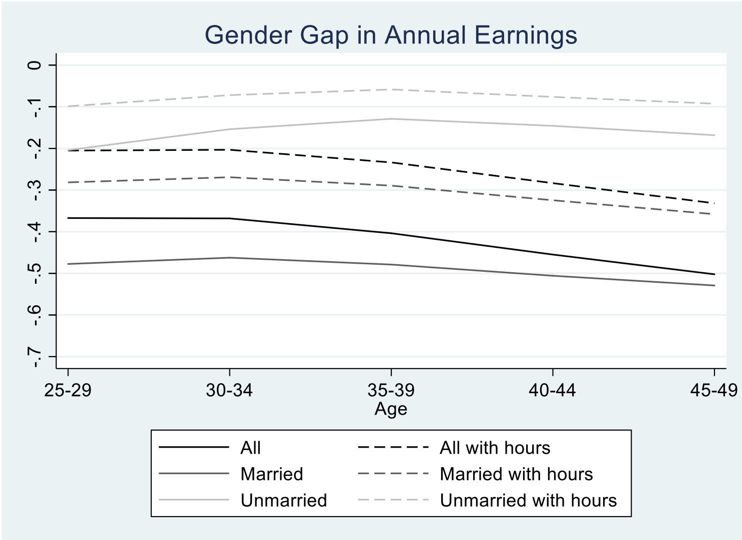
Appendix Figure 3A: Models without industry or occupation controls. College graduates

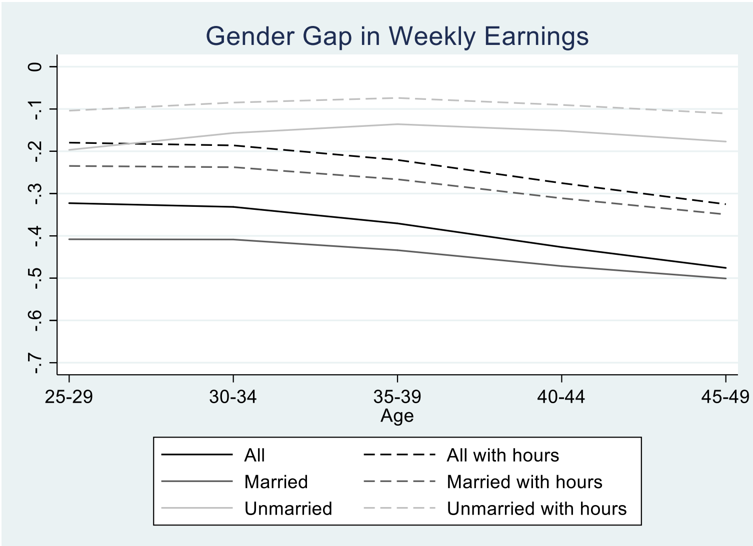


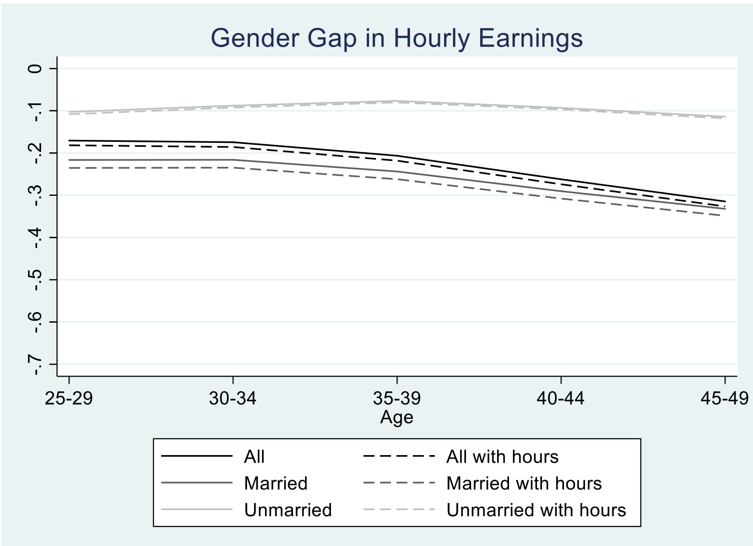




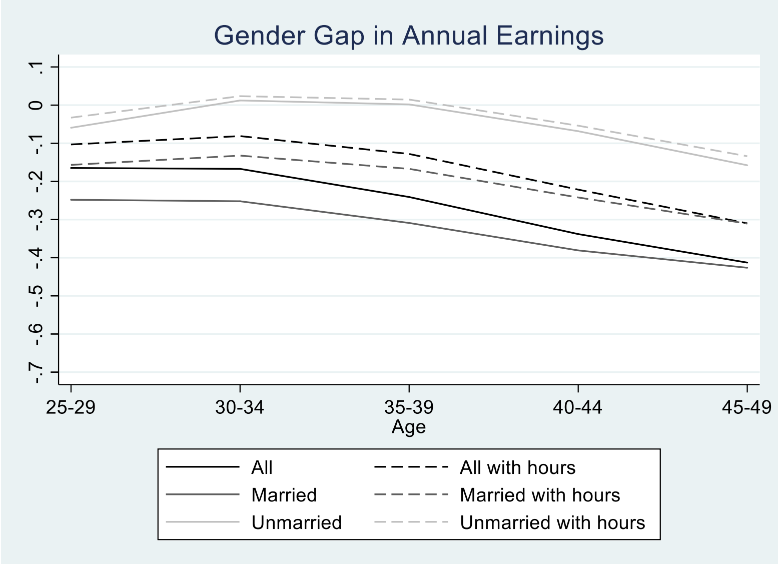
Appendix Figure 3B: Models without industry or occupation controls. No College

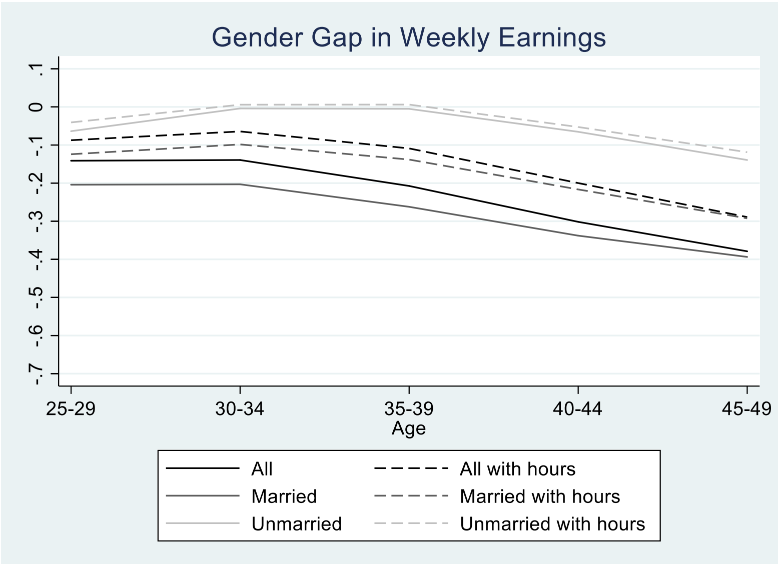


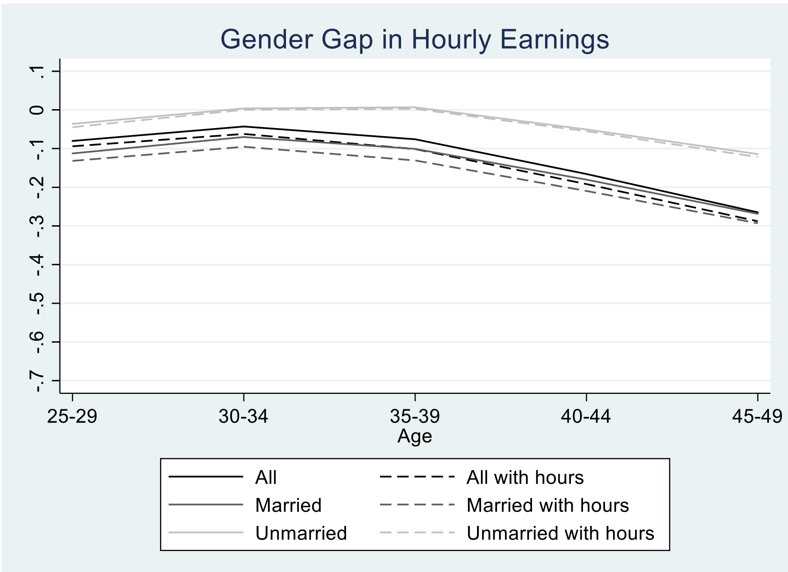




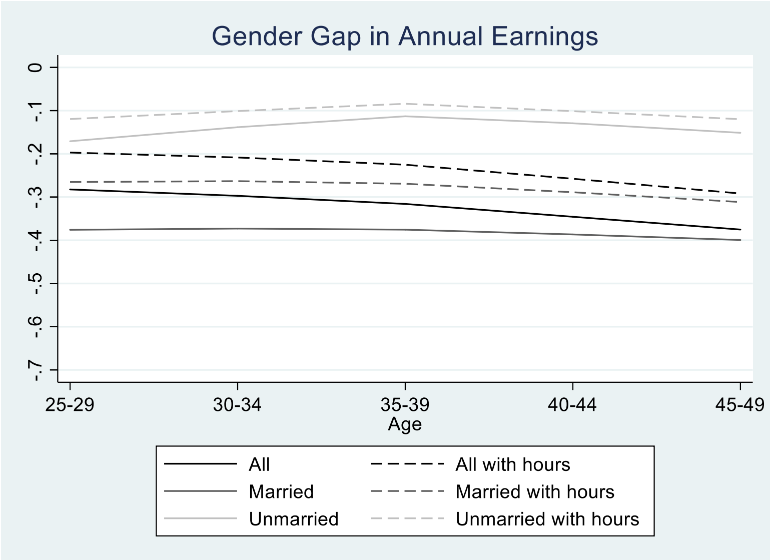
Appendix Figure 3C: Models with industry and occupation controls. College graduates

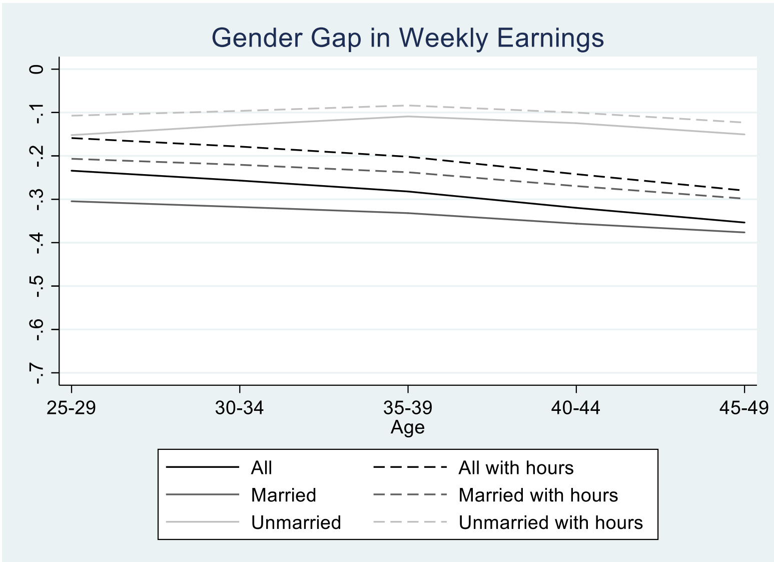


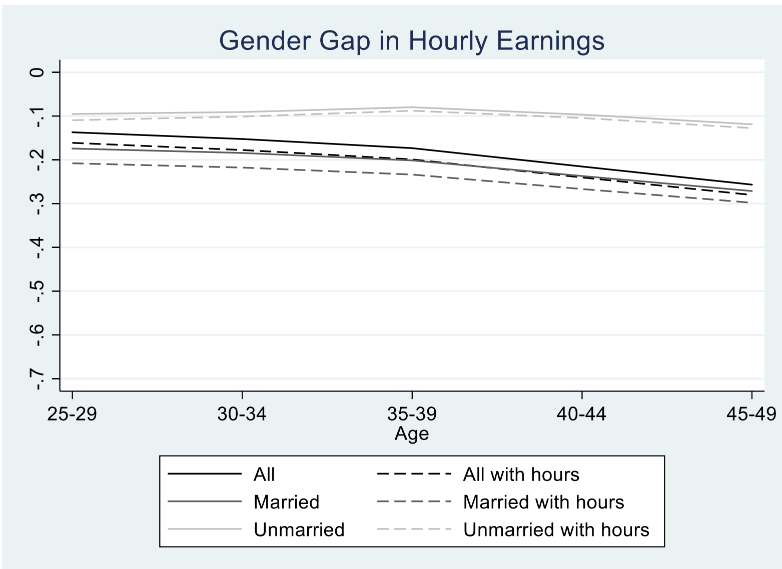




Appendix Figure 3D: Model with industry and occupation controls. No College







Appendix 4: O*Net Analysis

Appendix Table 4A—O*Net Characteristics: Means (Normalized) by Occupational Group

Occupational Group:	Time pressure	Contact with others	Interpersonal relationships	Structured work	Freedom to make decisions	Number of occupations
Management & Professionals	0.045	0.214	0.697	0.505	0.410	157
Production &Transportation	0.192	-0.726	-0.892	-0.658	-0.458	98
Sales & Office	0.234	0.625	0.342	0.034	-0.334	63
Service Occupations	-0.874	0.457	0.075	-0.171	-0.167	55

Notes: Each of the O*Net characteristics has been normalized to have a mean of 0 and a standard deviation of 1. Following Goldin (2017) the work setting characteristics and questions we use in the analysis are:

1. Time pressure: *How often does this job require the worker to meet strict deadlines?* Lower pressure means worker does not have to be around at particular times.

2. Contact with others: *How much does this job require the worker to be in contact with others* (*face-to-face, by telephone, or otherwise*) *in order to perform it?* Less contact means greater flexibility.

3. Establishing and maintaining interpersonal relationships: *Developing constructive and cooperative working relationships with others, and maintaining them over time.* The more working relationships, the more workers and clients the employee must be around.

4. Structured versus unstructured work: *To what extent is this job structured for the worker, rather than allowing the worker to determine tasks, priorities, and goals?* If the job is highly structured to the worker, there would be a lower chance that the worker would have close substitutes.

5. Freedom to make decisions: *How much decision-making freedom, without supervision, does the job offer*. Generally means that the worker determines what each client should receive, rather than being given a specific project, and thus workers are poorer substitutes for each other the greater are these freedoms.

Source: O*Net Online, http://www.onetonline.org/ (version 18.1).