An Empirical Research on the Gender Wage Gap

Kha Do, Thao Phan, My Nguyen

College of Arts and Sciences, University of South Florida

December 1, 2023

Abstract

This paper investigates the impacts of various factors in determining wages, emphasizing whether gender is the only attribute of the gender wage gap. Using the Panel Study Income Dynamics (PSID), we conduct three econometric specification models and the log-linear regression to test if surveyed factors are statistically significant in earnings and specifically run the F-test to test if occupations and industries are good indicators of earnings. The study finds out that gender, levels of education, years of experience, region, occupation, industry, and unionization play an important role in setting wages. On the other hand, race does not emerge as a significant indicator of earnings.

Keywords: gender wage gap, unadjusted model, human-capital model, full model, ln(wage)

1. Introduction

The wage gap between men and women has proven to be consistently intriguing in the United States since the 1860s when a historical event "Equal Pay for Equal Work" occurred. Although the two activists advocating for women's rights - Susan Anthony and Elizabeth Stanton - successfully carried out a case and gained approval in the 19th Amendment to the Constitution in 1920, the wage gap has remained compelling until now. According to Figure 1 of the U.S. Census Bureau, full-time women employees working throughout the year earned 60 percent as much as men did on average in 1960. This female-to-male earnings ratio kept increasing radically and reached 83.7% after 60 years. Although the gap has become narrower over time, Figure 2 shows that the median weekly earnings of females in seven main age groups were always lower than those of males, with the highest difference of $1,098 in the 55-to-64 age group in 2017. This implies that the differences in wages of men and women significantly exist and might lead to unexpected practical results for women. A prominent example is that the gap caused 13% of 18-64-year-old American women to be financially constrained compared to only 9% of men (Fontenot et al., 2018), and the domino effects on the salaries in the new jobs and less welfare from the community. Interestingly, A Pew Research Center survey in October 2022 revealed that nearly 50% of US adults attribute the different treatment of employers for women to the main drivers behind the wage gap. Although those findings align with our intuition, they may mislead people to the assumption that females earn less than males in almost all situations and that gender is the only attribute in their wage.

This presumption is not true because multiple factors might also be counted to determine the wage for an employee, including years of education, years of experience, occupation, industry, unionization, race, and so on. Therefore, it is important to investigate the weighted effect of gender on wages. If gender is one of the significant factors in establishing the salary, then this means there

exists gender discrimination and the employers should eliminate this practice in the process of setting the wage. In this paper, we try to implement different factors that might be the determinants of the wage and test if they are statistically significant in determining the salary.

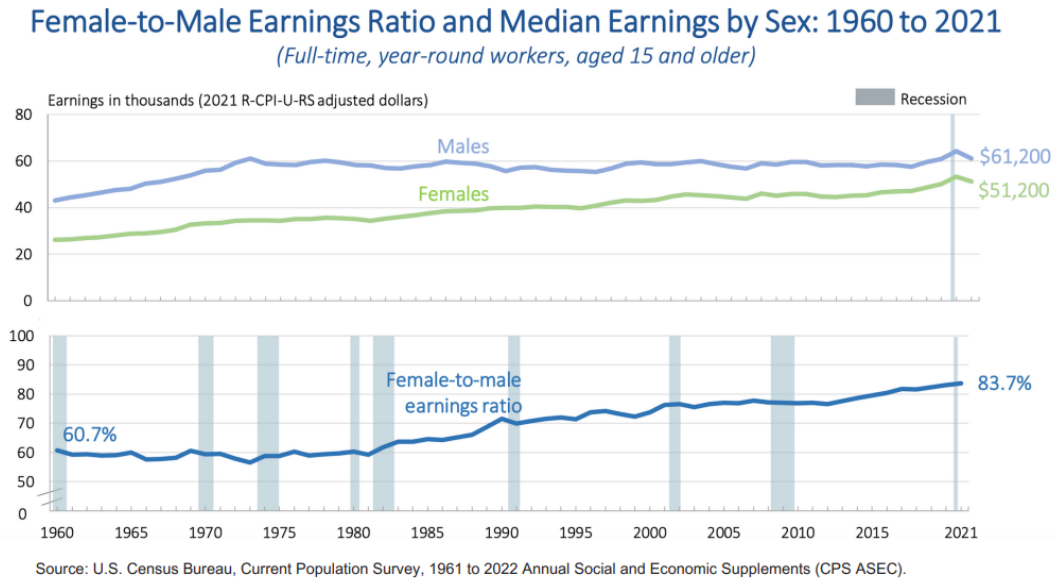

Figure 1

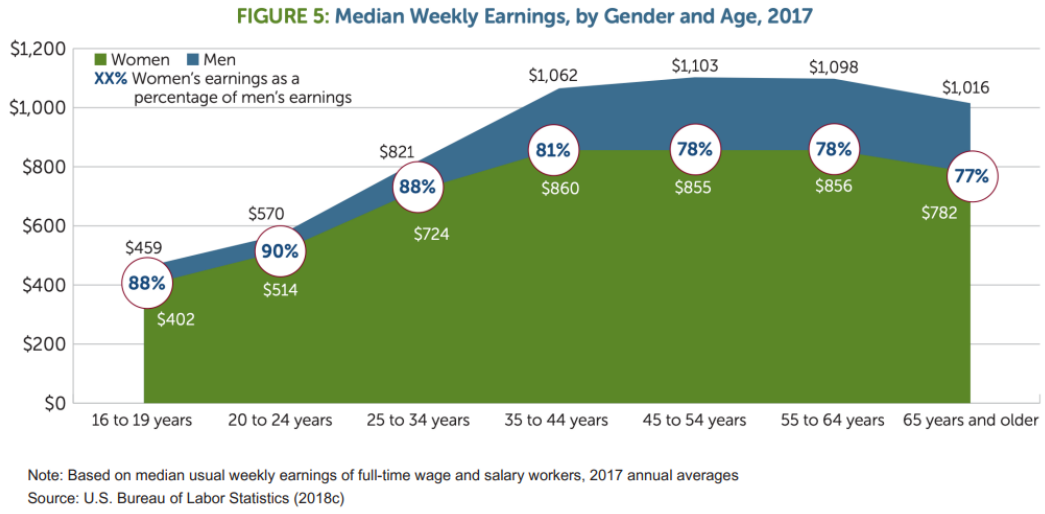

Figure 2

2. Literature Review

The Equal Pay Act of 1963 prevents sexual discrimination from influencing the wage rate for men and women. However, the earnings ratio has been still biased for male workers over time. According to the U.S. Department of Labor, the gender wage gap is defined as the difference in the earnings of male and female employees who work full-time all year to control any other differences in years of education, years of experience, and so on.

Huang (1999) shows that gender and education are positively and interactively correlated with income. By this, Huang means that, if keeping all other factors constant, males might earn higher than females even if they have the same capabilities in the same position. Moreover, the higher levels of education workers attain regardless of sex, the higher the wage rate they will be paid. However, his findings show that women with junior college level or higher degrees will earn a larger salary increase compared to men. This implies that women might be able to narrow the wage gap by having additional years of education beyond the junior college level.

In another research, Blau and Kahn (2017) found patterns in the gaps between men's and women’s surveyed wages over the period from 1980 to 2010 and used the Oaxaca-Blinder decompositions of OLS regressions to explain the difference in the mean wage of two groups. One notable finding is that although the decline in gender differences in job choices has partly narrowed the wage gap, the discrepancies in occupations and industries remain important in contributing to the wage gap. Furthermore, the impact of gender roles and the division of labor are highlighted to exert an impact on the gap, with the salient examples of motherhood wage penalty and marriage premium perks. More importantly, although it has been not officially proved, the unexplained gender wage gap implies that labor-market discrimination oversees the same pattern as the gender wage gap. In other words, from 1980 to 2010, when the gender wage gap was narrowed, labor-

market discrimination also decreased. This has raised the question of whether gender differences might lead to wage differences.

Olga and Coral (2023) also reveal that the impact of gender on the gender wage gap is independent of race, ethnicity, and class variables. By this, their study shows that, regardless of race or ethnicity, women earn lower wages than men do even if they are in the same group. Moreover, Black, and Hispanic employees, particularly Black women, suffer the most from wage disadvantages, which is partly explained by the class hierarchy. The research also finds out that, irrespective of class, one of the main reasons that Black women earn less is due to the choice of occupations, while this kind of decision exerts no effect on White and Hispanic women.

3. Data

3.1 Data Sources

To test if gender is an important indicator of wage controlling other variables, we require detailed micro-level data that comprehensively captures wage information for both male and female individuals. The Panel Study of Income Dynamics (PSID) stands as a data source, becoming a nationwide representative sample of over 18,000 individuals across 5,000 households. Initiated in 1968, the PSID has maintained a consistent approach to data collection of PSID families and residents over time. This long-term framework offers the temporal depth required to examine the impact of gender on wages.

Originally, our dataset covers the wage from 1980 to 2011. However, in this empirical research, we choose several specific calendar years to study how gender differences affect the gender wage gap in different time periods. In particular, we pick the 1980-, 1998-, and 2010-year data for testing the hypothesis on whether or not sex significantly determine one’s earning. Besides, our main focus is on men and women aged between 25 and 64 who worked full-time in

any non-farm-related field for at least 26 weeks during the preceding year. These selections of full-time workers with substantial employment engagement throughout the year are designated to ensure the corresponding levels of labor-force commitments. After that, the design turns out 13,317 observations that are found to be satisfied with these restrictions, in which 3,795 records from 1980, 4,673 from 1998, and 4,849 from 2010.

3.2 All Variables Used in the Analysis

The target variable is the natural logarithm of wage $ln(wage_i)$ instead of the actual wage values. There are two main reasons why log transformation is necessary in regression analysis. Firstly, the transformation improves the model fit by altering the scale and obtaining the *wage* variable – which initially exhibits a right-skewed distribution, more normally distributed. Secondly, the interpretability of coefficients is enhanced when implementing the log wages. In this case, the regression coefficients will be elasticities and it is interpreted as a one-unit change in dependent variables (X) would lead to a β% increase/decrease in wage (Y).

Since our objective is to study if sex is a significant determinant of wage controlling other variables, we proceed in three stages. First, the wage is estimated by only the gender variable to measure the mean difference in pay between males and females which is called the "Unadjusted" pay gap. Second, we enhance this model with control variables of education, experience, region, and race/ethnicity. We name this regression as "Human-capital specification" because the human-capital variables – education and experience – are included in addition to the sex variable. Third, a series of unionization, industry, and occupation dummy variables are augmented into the model, resulting in what we refer to as the "Full specification" model.

Regarding education features, we control the number of schooling years and dummy variables to differentiate between those with only a bachelor's degree and those with an advanced

degree. Our second model measures full-time and part-time labor-market experience, including the respective squared terms. Region variables are controlled for utilizing four mutually exclusive categories: northeast, northcentral, south, and west (the excluded class). Race/ethnicity variables are mutually exclusive dummy variables indicating if a person is white non-Hispanic, black non-Hispanic, other non-Hispanic, and Hispanic. In the full specification model, we further control for fourteen industries and twenty occupations dummy variables, along with a dummy variable indicating if a company offers union coverage or not.

3.3 Data Summary Statistics

The summaries of the significant determinants are listed in Table 1 (see Appendix Table 1), and these determinants pertain to three specific study years: 1980, 1998, and 2010. The mean value of the sex feature is around 0.54 which implies that the filtered dataset is balanced containing an almost equal number of samples from male or female classes. The log real wage of all employees ranges from 0.725 in 1980 to 6.255 in 2010 which can be explained by the inflation and productivity growth. Moreover, the mean number of years of schooling is 13.47 which means that they are currently in their freshmen and sophomore years of college. While the average number of years of experience for full-time employees exceeds 16 years, part-time labor accounts for nearly one-fourth, equivalent to 2.33 years. In terms of ethnicity, the majority of the employed individuals are recorded as white non-Hispanic, constituting approximately 64.1%. Lastly, one more thing we could tell from the table is that 44% of the workers from the dataset is from the southern region, followed by 25.7% from the northcentral, 15.4% from the west, and 14.9% from the northeast.

4. Econometric model specification

4.1 Exploratory Data Analysis

Here is the histogram for hourly wage.

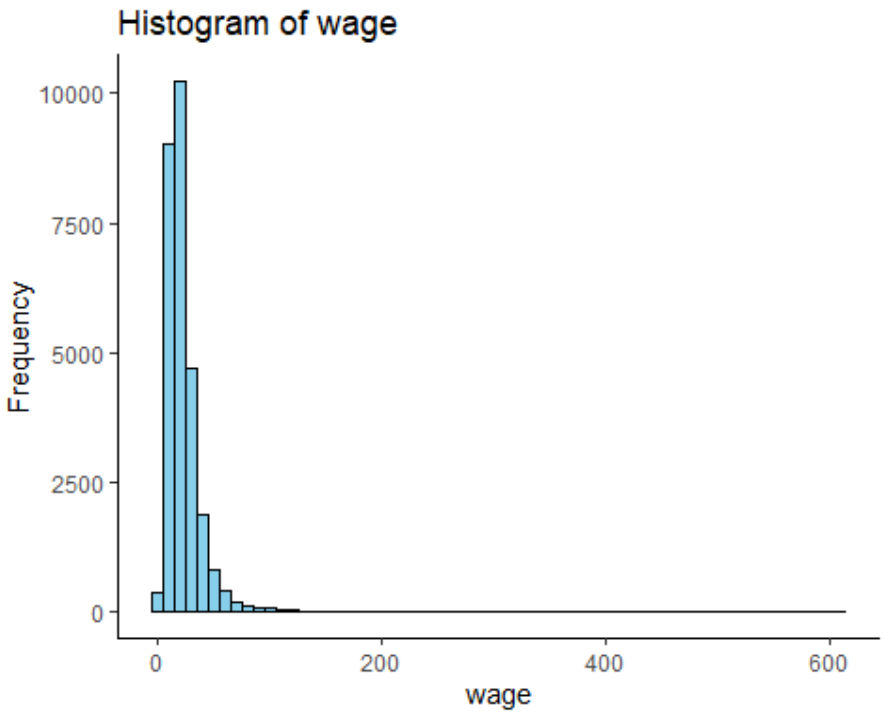

Figure 3

Figure 3 illustrates a highly skewed distribution of wages, a characteristic commonly observed in this type of data due to the presence of high-income individuals, resulting in numerous outliers. This skewness is a typical feature, prompting a common practice of utilizing the natural logarithm of wages as a more suitable representation.

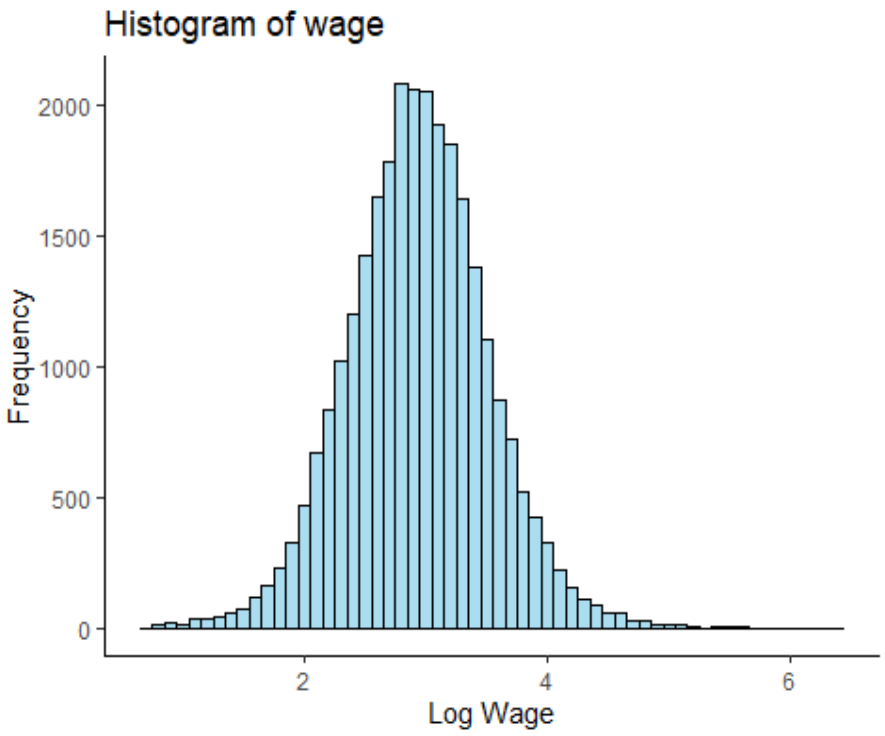

Figure 4

In Figure 4, Opting for log-transformed wages instead of raw wages in a linear regression model brings an added advantage—enhanced interpretability. The coefficients in this context become more straightforward to interpret, representing percentage changes in wages rather than absolute changes, thereby providing a more intuitive understanding of the model outcomes.

There are many factors that affect wages. One of the most important variables is education.

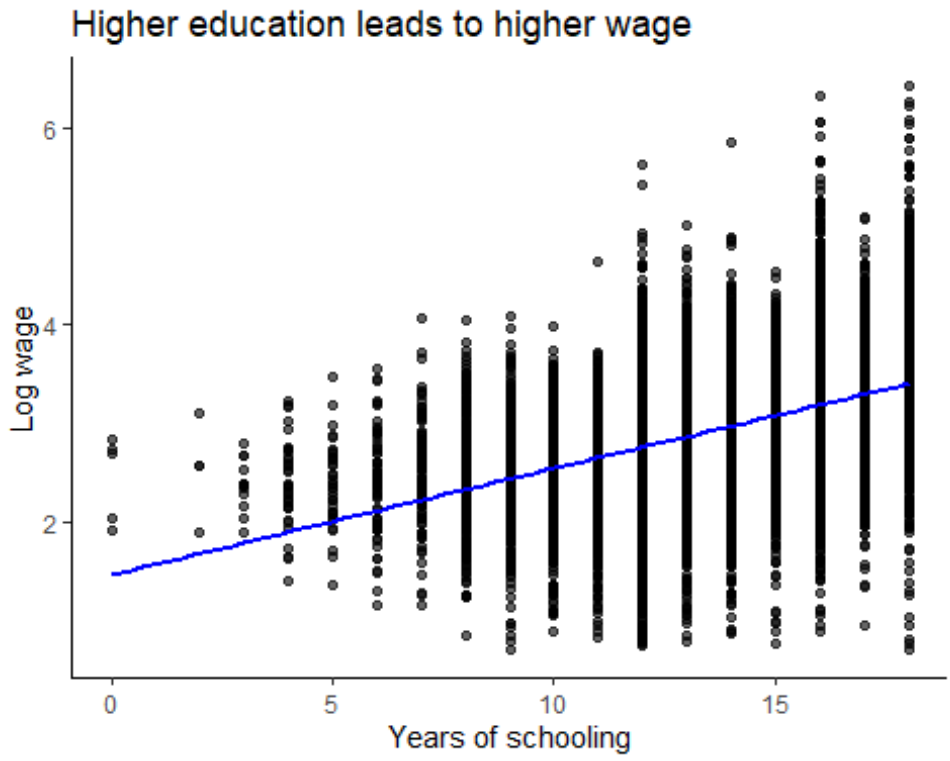

Figure 5

In Figure 5, The graph indicates a clear positive trend: as education years rise, so does log wage. This aligns with the logical expectation that higher education enables individuals to pursue more skilled and specialized roles. Notably, an extra year of education could potentially result in a 10% wage increase (Psacharopoulos and Patrinos 2004). Additionally, log wage variability increases at higher education levels.

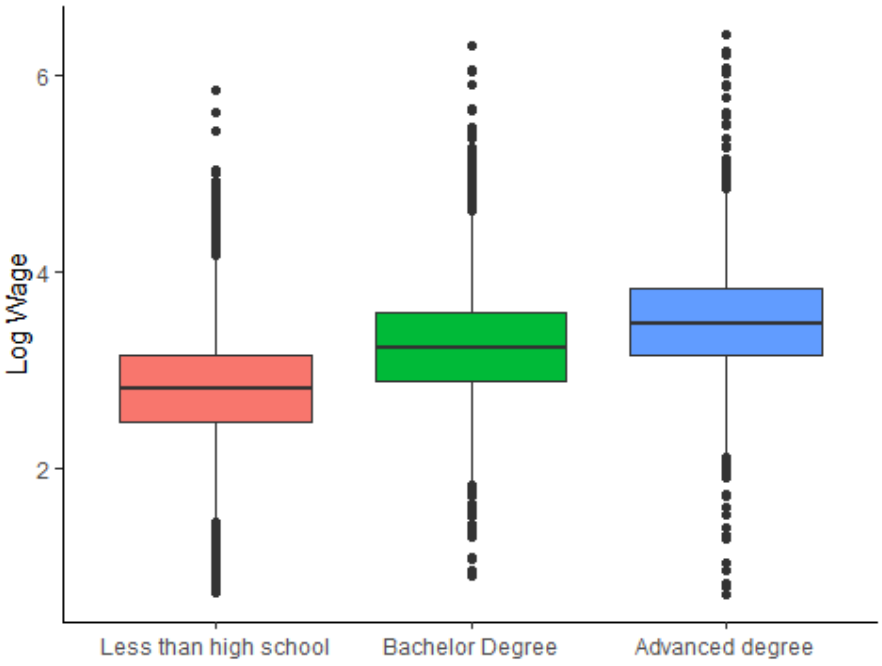

Figure 6

The connection between higher education and increased wages shown in Figure 6, becomes more pronounced when considering the specific degree individuals hold, as evident in the boxplot analysis. Notably, those with less than a high school degree earn the least, while individuals with advanced degrees earn the most. The disparity is further emphasized by a noticeable increase in the number of outliers at the advanced degree level.

In addition to education, years of experience play a pivotal role in determining wages. The expectation is that as individuals accumulate more years of experience, their wages are anticipated to rise accordingly. This underscores the multifaceted nature of factors influencing wage levels, where both educational attainment and professional experience contribute significantly.

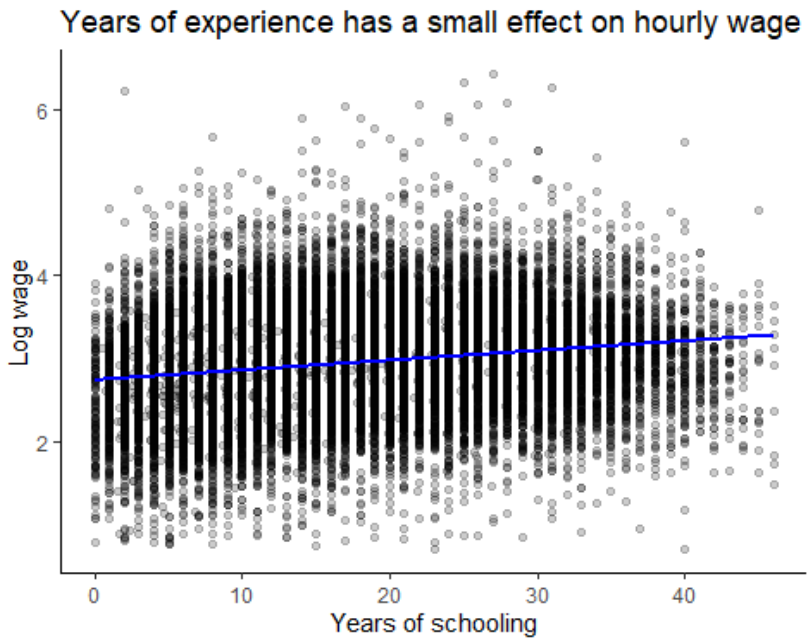

Figure 7

As shown in the Figure 7, we see a small relationship between between years of experience and wage.

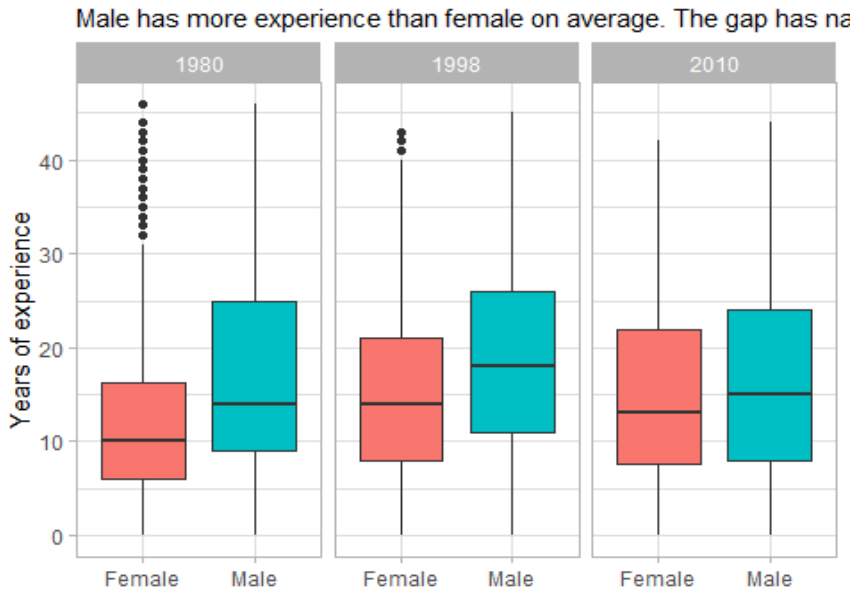

Figure 8

This factor could also result in the gender wage gap. In all the years, male has more experience than female on average. The gap has narrowed over time as shown in Figure 8.

4.2 Econometric Model Specification

There are many other factors that could also determine wage and correlate with our variable of interest, sex. To capture the effect of gender on wages, we need to build linear regression models with control of different variables to ensure that the effect of sex on wages is as unbiased as possible. Therefore, we build 3 econometrics models with different variables to examine how including different variables affects the effect of sex on wages. Our models are:

- Unadjusted model: $\ln(wage_i) = \beta_0 + \beta_1 \text{Sex}_i + \epsilon_i$
- Human-capital specification model: $\ln(wage_i) = \hat{\beta}_0 + \hat{\beta}_1 \text{Sex}_i + \mathbf{H}^T \widehat{\boldsymbol{\beta}}_{\mathbf{human\_capital}} + \epsilon_i$
- Full model: $\ln(wage_i) = \hat{\beta}_0 + \hat{\beta}_1 \text{Sex}_i + \mathbf{H}^T \widehat{\boldsymbol{\beta}}_{\mathbf{human\_capital}} + \mathbf{J}^T \widehat{\boldsymbol{\beta}}_{\mathbf{job}} + \epsilon_i$

Where:

- **H** is the variable vector related to human capital including Education variables, Experience variables, Region variables, and Race variables
- **J** is the variable vector related to industry including Unionization variables, Industry variables, and Occupation variables
- $\widehat{\boldsymbol{\beta}}_{\mathbf{human\_capital}}$ is the coefficient vector for variables in H
- $\widehat{\boldsymbol{\beta}}_{\mathbf{job}}$ is the coefficient vector for variables in J

The three models were run utilizing data from the year 2010. Subsequently, we plan to construct two additional models for the years 1980 and 1998. This approach aims to explore how the impact of sex on wages has evolved while accounting for the same set of control variables in each model.

5. Results

The impact of the sex variable is consistently negative across all models, demonstrating a statistically significant coefficient at a 0.01 significance level. In Model 1, it is revealed that female earnings are estimated to be 22.2% less than male earnings; however, it is important to note that Model 1 may suffer from omitted variable bias. As we progress to Model 2 and 3, introducing control variables leads to a decrease in the sex coefficient by 0.01 and 0.08, respectively. These findings align with those presented in "The Gender Wage Gap: Extent, Trends, and Explanations" paper by Blau and Kahn.

The education variables (sch, ba, adv) in Model 2 and 3 emerge as significant indicators of earnings. Notably, higher levels of education are associated with higher wages. For instance, individuals with a bachelor's degree experience a noteworthy increase in earnings by 16.3% in Model 3, underscoring the positive correlation between education levels and income. When using the f-test for the education variables, we reject the null hypothesis at 1% significance level. (See appendix)

Years of experience also plays a role in determining the wages. As examining during the preliminary analysis, experience has a small effect on wage. This reflects in the result of our model. For full-time workers, the coefficient for years of experience is 0.040. This implies that for every additional unit of experience, wages are expected to increase by approximately 4%, underscoring the incremental effect of accumulated professional expertise on earnings.

In both Model 2 and Model 3, two out of three coefficients associated with regions (Northeast, North Central, South) demonstrate statistical significance. Specifically, the coefficients for North Central and South are negative, suggesting a potential lower wage for

individuals residing in these regions. Conversely, the coefficient for the Northeast is positive, indicating a potential higher wage, although the evidence supporting this observation is not statistically robust. Employing the f-test for the region variables, we reject the null hypothesis at a 1% significance level, as detailed in the appendix.

Contrary to expectations, race does not emerge as a significant indicator in Model 2 and 3 when considering individual t-statistics. However, when testing for the joint hypothesis, we reject the null hypothesis at 1% confidence interval as detailed in the appendix. this could imply that while individual race may not have a significant impact on the outcome, the combination of the 3 races is collectively influential in explaining variations in wage.

In Model 3, variables associated with industry and occupation prove to be valuable indicators of earnings. We used f-test to test if industry and occupation is statistically significant and we can both result the null hypothesis at 1% significant level. The results is presented in the Appendix. This underscores the importance of considering the specific industry and occupation when assessing wage differentials. The inclusion of these variables provides a more nuanced understanding of the factors influencing earnings.

Unionization emerges as a significant factor impacting earnings, with average wages estimated to be 19.7% higher in companies with union coverage. This underscores the role of collective bargaining and union representation in influencing wage levels, providing insights into the broader dynamics affecting compensation in the studied context. ## How the effect of sex on wage in the full model changes over time. In 1980, men earned approximately 24.7% more than women, as reflected by the sex coefficient of 0.247. Over the years, with more women joining the workforce, this gap has narrowed significantly to 15.6% in 1998 and further to 14.4% in 2010 (see Appendix Tables 5, 6, and 7). Notably, thechange from 1998 to 2010 is relatively small, raising

the possibility that it could be influenced by the 2008 Recession. Further research is needed to confirm this hypothesis.

6. Conclusion

The objective of the research question is to study how gender differences contribute to the gender wage gap and, consequently, how this understanding could facilitate feasible solutions to tackle this nerve-racking problem. For example, what actions the companies should take to mitigate gender discrimination in the workplace regarding the salary negotiation process.

From the work we have done, the regression model consistently shows that there is a negative and statistically significant coefficient of gender variable across all three models at 1% significance level. In specific, for the most recent year of 2010, the sex coefficients are 0.222, 0.210, and 0.144 for unadjusted, human-capital specification, and full specification models respectively. It implies that the average earnings of male individuals are estimated to be 22.2%, 21.0%, and 14.4% more than those of female individuals. The decrease in gender coefficient can be explained by the control variables indicating education, race, region, industry, occupation, and unionization in the second and third models.

Furthermore, there is a positive correlation between education and wages, which aligns with Blau and Kahn's findings. In other words, the higher the education earned, the higher salary an employee could get. Race variables, in this case, are not significant determinants of wage differences. Regional variables, on the other hand, show mixed impacts on this matter.

The full specification model points out the industry and profession are of importance to explain the wage differentials between men and women. Finally, unionization is considered a significant features where companies obtaining union coverage offer a 19.7% higher average wage.

However, although our findings are mostly consistent with the literature review, there are four limitations that can be improved in further research. Firstly, the omitted variable bias might exist because even though control variables are included in Models 2 and 3, many other potential attributes correlate with the existing variables and are the determinants of wage, including psychological attributes, noncognitive skills, social norms, and so on. Secondly, the study needs to take causation and correlation into consideration. This is because correlation does not imply causation. Specifically, in this research, although unionization is positively correlated with wages, this does not mean that higher unionization causes higher wages. Instead, the relationship between unionization and wages might be influenced by many other external factors. Thirdly, in contrast to the findings of Olga and Coral (2023), our study concludes race not to be a good indicator of earnings. This result requires us to consider race more broadly and capture more comprehensive effects of different races that affect wages. Finally, it comes to the temporal factors. Although we conducted research for three different separate years, we have not considered the potential trends in changing wage determinants over time. This leads to the fact that the results might be biased in certain years and not apply to all periods.

Despite the aforementioned limitations, the study provides insightful findings into the effect of gender, human capital, and job-related variables in wage determination, laying a strong foundation for further investigation and regulation release.

References

Arpita, C., Jim, M. & Hiren N. (2019). Gender-Based Pay Disparity Study. *Women's Bureau - U.S. Department of Labor*.

Blau, F. D., & Kahn, L. M. (2017). The gender wage gap: Extent, trends, and explanations. *Journal of Economic Literature*, 55(3), 789–865.

Daugherty, G. (2023, March 1). *Gender and Income Inequality: History and Statistics*. Investopedia. https://www.investopedia.com/history-gender-wage-gap-america-5074898#:~:text=Though%20the%20gender%20wage%20gap,Equal%20Pay%20for%20Equal%20Work.%E2%80%9D

Evelyn. *"Gender Pay Gap Comparisons with Regression Analysis."* Medium, Medium, 9 July 2023, medium.com/@evelyn.eve.9512/gender-pay-gap-comparisons-with-regression-analysis-45223cd3ed13.

Fischer, J., & Hayes, J. (2013). The importance of Social Security in the incomes of older Americans: Differences by gender, age, race/ethnicity, and marital status. Washington, DC: Institute for Women's Policy Research.

Fontenot, K., Semega, J., & Kollar, M. (2018). Income and Poverty in the United States: 2017. Washington: U.S. Census Bureau.

Hlavac, M. (2018a, January 3). Package "Oaxaca" for Blinder-Oaxaca decomposition in R - part 1: Introduction and theory. Medium. https://medium.com/@MarekHlavac/package-oaxaca-for-blinder-oaxaca-decomposition-in-r-part-1-introduction-and-theory-b8dafb085a31

Hlavac, M. (2018, January 3). *Package "Oaxaca" for Blinder-Oaxaca decomposition in R - part 2: Features and example.* Medium. https://medium.com/@MarekHlavac/package-oaxaca-for-blinder-oaxaca-decomposition-in-r-part-2-features-and-example-12222e9a886d

Huang, T. (1999), "The impact of education and seniority on the male-female wage gap: is more education the answer?", International Journal of Manpower, Vol. 20 No. 6, pp. 361-374. https://doi.org/10.1108/01437729910289710

Full article: Gender, Race, and Class in an Intersectional Framework: Occupations and Wages in the United States (tandfonline.com)

Stock, James H. and Mark W. Watson (2007), Introduction to Econometrics, 2nd Edition, London: Pearson Education.

Pew Research Center. (2023, March 1). *Gender pay gap remained stable over past 20 years in US | Pew Research Center*. https://www.pewresearch.org/short-reads/2023/03/01/gender-pay-gap-facts/#:~:text=The%20gender%20gap%20in%20pay,%2D%20and%20part%2Dtime%20workers.

Pew Research Center. (2023b, March 1). *Gender pay gap remained stable over past 20 years in US | Pew Research Center*. https://www.pewresearch.org/short-reads/2023/03/01/gender-pay-gap-facts/#:~:text=The%20gender%20gap%20in%20pay,%2D%20and%20part%2Dtime%20workers.

Table 1: Summary Statistics for 1980, 1998, and 2010

Statistic	N	Mean	St. Dev.	Min	Max
sex	13,317	0.539	0.499	0	1
lnrealwg	13,317	2.915	0.579	0.725	6.255
wave	13,317	1,997.240	12.008	1,980	2,010
ba	13,317	0.190	0.392	0	1
adv	13,317	0.090	0.287	0	1
northeast	13,317	0.149	0.356	0	1
northcentral	13,317	0.257	0.437	0	1
south	13,317	0.440	0.496	0	1
white	13,317	0.641	0.480	0	1
black	13,317	0.320	0.467	0	1
hisp	13,317	0.025	0.156	0	1
yrsftexpfz	13,317	16.129	9.741	0.000	46.000
yrsptexpfz	13,317	2.334	3.356	0.000	37.592
yrsftexpfzsq	13,317	355.024	382.724	0.000	2,116.000
yrsptexpfzsq	13,317	16.712	57.889	0.000	1,413.151
sch	13,308	13.473	2.491	0	18
miningconstruction	13,317	0.059	0.235	0	1
durables	13,317	0.124	0.329	0	1
nondurables	13,317	0.081	0.273	0	1
Transport	13,317	0.056	0.230	0	1
Utilities	13,317	0.016	0.125	0	1
Communications	13,317	0.029	0.167	0	1
retailtrade	13,317	0.075	0.264	0	1
wholesaletrade	13,317	0.033	0.180	0	1
finance	13,317	0.062	0.242	0	1
hotelsrestaurants	13,317	0.033	0.179	0	1
Medical	13,317	0.118	0.322	0	1
Education	13,317	0.106	0.308	0	1
professional	13,317	0.067	0.249	0	1
publicadmin	13,317	0.079	0.270	0	1
manager	13,317	0.106	0.308	0	1
business	13,317	0.023	0.149	0	1
financialop	13,317	0.018	0.135	0	1
computer	13,317	0.023	0.150	0	1
architect	13,317	0.027	0.163	0	1
scientist	13,317	0.012	0.109	0	1
socialworker	13,317	0.019	0.137	0	1
postseceduc	13,317	0.007	0.081	0	1
legaleduc	13,317	0.058	0.234	0	1
artist	13,317	0.009	0.097	0	1
lawyerphysician	13,317	0.009	0.094	0	1
healthcare	13,317	0.043	0.203	0	1
healthsupport	13,317	0.028	0.166	0	1
protective	13,317	0.028	0.166	0	1
foodcare	13,317	0.040	0.196	0	1
building	13,317	0.032	0.176	0	1
sales	13,317	0.076	0.264	0	1
officeadmin	13,317	0.158	0.365	0	1
constructextractinstall	13,317	0.089	0.285	0	1
production	13,317	0.119	0.324	0	1
wtrgov	13,278	0.252	0.434	0	1
unjob	13,317	0.222	0.416	0	1

Table 2: Summary Statistics 2010

Statistic	N	Mean	St. Dev.	Min	Max
sex	4,849	0.492	0.500	0	1
lnrealwg	4,849	3.018	0.610	0.779	6.255
wave	4,849	2,010.000	0.000	2,010	2,010
ba	4,849	0.228	0.419	0	1
adv	4,849	0.114	0.317	0	1
northeast	4,849	0.146	0.353	0	1
northcentral	4,849	0.264	0.441	0	1
south	4,849	0.434	0.496	0	1
white	4,849	0.627	0.484	0	1
black	4,849	0.322	0.467	0	1
hisp	4,849	0.039	0.193	0	1
yrsftexpfz	4,849	15.796	9.543	0.000	44.000
yrsptexpfz	4,849	2.516	3.290	0.000	37.592
yrsftexpfzsq	4,849	340.546	357.285	0.000	1,936.000
yrsptexpfzsq	4,849	17.153	54.483	0.000	1,413.151
sch	4,849	14.189	2.232	2	18
miningconstruction	4,849	0.055	0.229	0	1
durables	4,849	0.087	0.282	0	1
nondurables	4,849	0.052	0.222	0	1
Transport	4,849	0.053	0.224	0	1
Utilities	4,849	0.015	0.123	0	1
Communications	4,849	0.029	0.167	0	1
retailtrade	4,849	0.077	0.266	0	1
wholesaletrade	4,849	0.035	0.183	0	1
finance	4,849	0.066	0.249	0	1
hotelsrestaurants	4,849	0.042	0.200	0	1
Medical	4,849	0.146	0.353	0	1
Education	4,849	0.110	0.313	0	1
professional	4,849	0.075	0.263	0	1
publicadmin	4,849	0.091	0.288	0	1
manager	4,849	0.097	0.296	0	1
business	4,849	0.026	0.159	0	1
financialop	4,849	0.025	0.157	0	1
computer	4,849	0.027	0.163	0	1
architect	4,849	0.022	0.146	0	1
scientist	4,849	0.014	0.117	0	1
socialworker	4,849	0.027	0.163	0	1
postseceduc	4,849	0.007	0.081	0	1
legaleduc	4,849	0.070	0.255	0	1
artist	4,849	0.010	0.099	0	1
lawyerphysician	4,849	0.011	0.105	0	1
healthcare	4,849	0.054	0.226	0	1
healthsupport	4,849	0.032	0.176	0	1
protective	4,849	0.037	0.189	0	1
foodcare	4,849	0.052	0.223	0	1
building	4,849	0.027	0.162	0	1
sales	4,849	0.078	0.268	0	1
officeadmin	4,849	0.168	0.374	0	1
constructextractinstall	4,849	0.079	0.270	0	1
production	4,849	0.070	0.256	0	1
wtrgov	4,814	0.260	0.439	0	1
unjob	4,849	0.180	0.384	0	1

Table 3: Summary Statistics 1998

Statistic	N	Mean	St. Dev.	Min	Max
sex	4,673	0.532	0.499	0	1
lnrealwg	4,673	2.916	0.563	0.789	5.848
wave	4,673	1,998.000	0.000	1,998	1,998
ba	4,673	0.194	0.396	0	1
adv	4,673	0.085	0.279	0	1
northeast	4,673	0.143	0.350	0	1
northcentral	4,673	0.274	0.446	0	1
south	4,673	0.434	0.496	0	1
white	4,673	0.669	0.471	0	1
black	4,673	0.302	0.459	0	1
hisp	4,673	0.010	0.098	0	1
yrsftexpfz	4,673	16.847	9.367	0.000	45.000
yrsptexpfz	4,673	2.609	3.639	0.000	34.000
yrsftexpfzsq	4,673	371.546	360.298	0.000	2,025.000
yrsptexpfzsq	4,673	20.048	65.114	0.000	1,156.000
sch	4,665	13.557	2.204	0	18
miningconstruction	4,673	0.055	0.228	0	1
durables	4,673	0.124	0.329	0	1
nondurables	4,673	0.080	0.272	0	1
Transport	4,673	0.056	0.230	0	1
Utilities	4,673	0.015	0.120	0	1
Communications	4,673	0.031	0.173	0	1
retailtrade	4,673	0.074	0.262	0	1
wholesaletrade	4,673	0.033	0.180	0	1
finance	4,673	0.065	0.247	0	1
hotelsrestaurants	4,673	0.030	0.170	0	1
Medical	4,673	0.108	0.311	0	1
Education	4,673	0.114	0.318	0	1
professional	4,673	0.074	0.263	0	1
publicadmin	4,673	0.080	0.272	0	1
manager	4,673	0.122	0.327	0	1
business	4,673	0.025	0.157	0	1
financialop	4,673	0.018	0.131	0	1
computer	4,673	0.027	0.162	0	1
architect	4,673	0.031	0.173	0	1
scientist	4,673	0.012	0.109	0	1
socialworker	4,673	0.015	0.123	0	1
postseceduc	4,673	0.006	0.077	0	1
legaleduc	4,673	0.061	0.240	0	1
artist	4,673	0.010	0.100	0	1
lawyerphysician	4,673	0.010	0.100	0	1
healthcare	4,673	0.041	0.199	0	1
healthsupport	4,673	0.023	0.150	0	1
protective	4,673	0.027	0.161	0	1
foodcare	4,673	0.032	0.176	0	1
building	4,673	0.032	0.177	0	1
sales	4,673	0.089	0.284	0	1
officeadmin	4,673	0.150	0.357	0	1
constructextractinstall	4,673	0.083	0.277	0	1
production	4,673	0.115	0.319	0	1
wtrgov	4,669	0.259	0.438	0	1
unjob	4,673	0.204	0.403	0	1

Table 4: Summary Statistics 1980

Statistic	N	Mean	St. Dev.	Min	Max
sex	3,795	0.607	0.489	0	1
lnrealwg	3,795	2.782	0.528	0.725	4.982
wave	3,795	1,980.000	0.000	1,980	1,980
ba	3,795	0.135	0.342	0	1
adv	3,795	0.067	0.251	0	1
northeast	3,795	0.160	0.367	0	1
northcentral	3,795	0.228	0.420	0	1
south	3,795	0.455	0.498	0	1
white	3,795	0.625	0.484	0	1
black	3,795	0.340	0.474	0	1
hisp	3,795	0.027	0.162	0	1
yrsftexpfz	3,795	15.670	10.376	0	46
yrsptexpfz	3,795	1.764	2.989	0	33
yrsftexpfzsq	3,795	353.180	436.486	0	2,116
yrsptexpfzsq	3,795	12.040	52.077	0	1,089
sch	3,794	12.455	2.780	0	18
miningconstruction	3,795	0.068	0.252	0	1
durables	3,795	0.170	0.376	0	1
nondurables	3,795	0.119	0.323	0	1
Transport	3,795	0.060	0.238	0	1
Utilities	3,795	0.018	0.134	0	1
Communications	3,795	0.027	0.161	0	1
retailtrade	3,795	0.076	0.264	0	1
wholesaletrade	3,795	0.032	0.176	0	1
finance	3,795	0.054	0.227	0	1
hotelsrestaurants	3,795	0.026	0.159	0	1
Medical	3,795	0.093	0.291	0	1
Education	3,795	0.091	0.288	0	1
professional	3,795	0.047	0.211	0	1
publicadmin	3,795	0.063	0.243	0	1
manager	3,795	0.097	0.296	0	1
business	3,795	0.016	0.124	0	1
financialop	3,795	0.011	0.105	0	1
computer	3,795	0.013	0.113	0	1
architect	3,795	0.030	0.171	0	1
scientist	3,795	0.010	0.100	0	1
socialworker	3,795	0.013	0.115	0	1
postseceduc	3,795	0.007	0.086	0	1
legaleduc	3,795	0.040	0.196	0	1
artist	3,795	0.008	0.090	0	1
lawyerphysician	3,795	0.004	0.067	0	1
healthcare	3,795	0.032	0.175	0	1
healthsupport	3,795	0.030	0.171	0	1
protective	3,795	0.019	0.137	0	1
foodcare	3,795	0.034	0.181	0	1
building	3,795	0.038	0.190	0	1
sales	3,795	0.056	0.230	0	1
officeadmin	3,795	0.156	0.363	0	1
constructextractinstall	3,795	0.109	0.311	0	1
production	3,795	0.187	0.390	0	1
wtrgov	3,795	0.233	0.423	0	1
unjob	3,795	0.298	0.457	0	1

Table 5: Linear regression table for year 2010

	Dependent variable:		
	lnrealwg		
	(1)	(2)	(3)
sex	0.222*** (0.017)	0.210*** (0.015)	0.144*** (0.016)
ba		0.190*** (0.027)	0.163*** (0.026)
adv		0.240*** (0.040)	0.250*** (0.038)
northeast		0.029 (0.026)	0.030 (0.024)
northcentral		−0.150*** (0.023)	−0.130*** (0.021)
south		−0.128*** (0.022)	−0.099*** (0.021)
white		0.083 (0.065)	0.055 (0.060)
black		−0.096 (0.066)	−0.068 (0.061)
hisp		0.030 (0.074)	0.008 (0.068)
yrsftexpfz		0.045*** (0.003)	0.040*** (0.003)
yrsptexpfz		−0.004 (0.004)	−0.003 (0.004)
yrsftexpfzsq		−0.001*** (0.0001)	−0.001*** (0.0001)
yrsptexpfzsq		0.00005 (0.0002)	0.0002 (0.0002)
sch		0.085*** (0.006)	0.055*** (0.006)
miningconstruction			0.244*** (0.041)
durables			0.312*** (0.037)
nondurables			0.320*** (0.041)
Transport			0.281*** (0.042)
Utilities			0.353*** (0.060)
Communications			0.262*** (0.048)
retailtrade			0.010 (0.038)
wholesaletrade			0.154*** (0.046)
finance			0.306*** (0.038)
hotelsrestaurants			0.070 (0.046)
Medical			0.185*** (0.035)
Education			0.137*** (0.039)
professional			0.250*** (0.036)
publicadmin			0.328*** (0.039)
manager			0.512*** (0.038)
business			0.377*** (0.051)
financialop			0.262*** (0.053)
computer			0.423*** (0.051)
architect			0.458*** (0.054)
scientist			0.278*** (0.064)
socialworker			0.098* (0.053)
postseceduc			0.083 (0.089)
legaleduc			0.077 (0.049)
artist			0.341*** (0.075)
lawyerphysician			0.756*** (0.073)
healthcare			0.392*** (0.047)
healthsupport			−0.001 (0.053)
protective			0.233*** (0.049)
foodcare			−0.062 (0.048)
building			−0.068 (0.051)
sales			0.338*** (0.039)
officeadmin			0.121*** (0.034)
constructextractinstall			0.201*** (0.039)
production			−0.016 (0.040)
wtrgov			−0.081*** (0.021)
unjob			0.197*** (0.019)
Constant	2.909*** (0.012)	1.300*** (0.111)	1.376*** (0.110)
Observations	4,849	4,849	4,814
R^2	0.033	0.338	0.458
Adjusted R^2	0.033	0.336	0.453
Residual Std. Error	0.600 (df = 4847)	0.497 (df = 4834)	0.451 (df = 4763)
F Statistic	166.268*** (df = 1; 4847)	176.213*** (df = 14; 4834)	80.629*** (df = 50; 4763)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 6: Linear regression table for year 1998

	Dependent variable:		
	lnrealwg		
	(1)	(2)	(3)
sex	0.279*** (0.016)	0.205*** (0.014)	0.156*** (0.016)
ba		0.131*** (0.025)	0.120*** (0.024)
adv		0.268*** (0.035)	0.300*** (0.035)
northeast		0.001 (0.025)	−0.001 (0.024)
northcentral		−0.087*** (0.022)	−0.073*** (0.021)
south		−0.171*** (0.021)	−0.125*** (0.020)
white		0.077 (0.050)	0.075 (0.046)
black		−0.011 (0.051)	0.004 (0.048)
hisp		−0.110 (0.085)	−0.045 (0.079)
yrsftexpfz		0.030*** (0.003)	0.026*** (0.002)
yrsptexpfz		−0.009** (0.004)	−0.006* (0.004)
yrsftexpfzsq		−0.0005*** (0.0001)	−0.0004*** (0.0001)
yrsptexpfzsq		0.0004* (0.0002)	0.0003 (0.0002)
sch		0.075*** (0.005)	0.054*** (0.005)
miningconstruction			0.144*** (0.040)
durables			0.250*** (0.034)
nondurables			0.212*** (0.037)
Transport			0.264*** (0.039)
Utilities			0.390*** (0.059)
Communications			0.308*** (0.046)
retailtrade			−0.001 (0.036)
wholesaletrade			0.143*** (0.044)
finance			0.252*** (0.037)
hotelsrestaurants			−0.082* (0.048)
Medical			0.122*** (0.036)
Education			−0.012 (0.039)
professional			0.234*** (0.036)
publicadmin			0.171*** (0.041)
manager			0.287*** (0.033)
business			0.274*** (0.049)
financialop			0.232*** (0.056)
computer			0.439*** (0.048)
architect			0.293*** (0.046)
scientist			0.182*** (0.065)
socialworker			0.130** (0.061)
postseceduc			0.163* (0.092)
legaleduc			0.133*** (0.048)
artist			0.138* (0.071)
lawyerphysician			0.524*** (0.073)
healthcare			0.319*** (0.047)
healthsupport			−0.031 (0.054)
protective			0.050 (0.051)
foodcare			−0.004 (0.049)
building			−0.083* (0.045)
sales			0.257*** (0.036)
officeadmin			0.074** (0.032)
constructextractinstall			0.158*** (0.035)
production			0.044 (0.033)
wtrgov			−0.007 (0.022)
unjob			0.214*** (0.018)
Constant	2.768*** (0.012)	1.478*** (0.084)	1.465*** (0.086)
Observations	4,673	4,665	4,661
R^2	0.061	0.336	0.432
Adjusted R^2	0.061	0.334	0.426
Residual Std. Error	0.546 (df = 4671)	0.460 (df = 4650)	0.427 (df = 4610)
F Statistic	303.815*** (df = 1; 4671)	168.264*** (df = 14; 4650)	70.214*** (df = 50; 4610)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 7: Linear regression table for year 1980

	Dependent variable:		
	lnrealwg		
	(1)	(2)	(3)
sex	0.419*** (0.016)	0.327*** (0.015)	0.247*** (0.016)
ba		0.097*** (0.026)	0.057** (0.026)
adv		0.123*** (0.036)	0.109*** (0.037)
northeast		−0.009 (0.025)	−0.013 (0.023)
northcentral		−0.024 (0.023)	−0.028 (0.021)
south		−0.144*** (0.021)	−0.084*** (0.020)
white		0.040 (0.078)	0.028 (0.072)
black		−0.063 (0.079)	−0.049 (0.073)
hisp		−0.014 (0.088)	−0.039 (0.081)
yrsftexpfz		0.030*** (0.003)	0.027*** (0.002)
yrsptexpfz		−0.015*** (0.005)	−0.012*** (0.004)
yrsftexpfzsq		−0.0005*** (0.0001)	−0.0004*** (0.0001)
yrsptexpfzsq		0.001*** (0.0002)	0.001*** (0.0002)
sch		0.060*** (0.004)	0.046*** (0.004)
miningconstruction			0.218*** (0.039)
durables			0.277*** (0.034)
nondurables			0.180*** (0.035)
Transport			0.330*** (0.039)
Utilities			0.281*** (0.054)
Communications			0.294*** (0.049)
retailtrade			0.067* (0.038)
wholesaletrade			0.172*** (0.046)
finance			0.202*** (0.040)
hotelsrestaurants			−0.011 (0.051)
Medical			0.165*** (0.038)
Education			0.097** (0.041)
professional			0.115*** (0.041)
publicadmin			0.278*** (0.042)
manager			0.342*** (0.033)
business			0.385*** (0.056)
financialop			0.423*** (0.065)
computer			0.419*** (0.061)
architect			0.318*** (0.044)
scientist			0.306*** (0.068)
socialworker			0.157** (0.064)
postseceduc			0.226*** (0.084)
legaleduc			0.312*** (0.051)
artist			0.147** (0.074)
lawyerphysician			0.460*** (0.101)
healthcare			0.280*** (0.049)
healthsupport			−0.024 (0.050)
protective			0.035 (0.054)
foodcare			−0.076 (0.048)
building			−0.116*** (0.041)
sales			0.167*** (0.038)
officeadmin			0.092*** (0.029)
constructextractinstall			0.114*** (0.031)
production			0.047* (0.027)
wtrgov			−0.032 (0.020)
unjob			0.242*** (0.016)
Constant	2.527*** (0.013)	1.601*** (0.094)	1.457*** (0.092)
Observations	3,795	3,794	3,794
R^2	0.151	0.381	0.490
Adjusted R^2	0.151	0.379	0.483
Residual Std. Error	0.486 (df = 3793)	0.416 (df = 3779)	0.379 (df = 3743)
F Statistic	673.789*** (df = 1; 3793)	166.427*** (df = 14; 3779)	71.860*** (df = 50; 3743)

Note: *p<0.1; **p<0.05; ***p<0.01

F test

```
## [1] "ft_education"

## Linear hypothesis test
##
## Hypothesis:
## ba = 0
## adv = 0
## sch = 0
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df     F    Pr(>F)
## 1   4766
## 2   4763  3 164.7 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] "ft_experience"

## Linear hypothesis test
##
## Hypothesis:
## yrsftexpfz = 0
## yrsptexpfz = 0
## yrsftexpfzsq = 0
## yrsptexpfzsq = 0
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
```

```
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1   4767
## 2   4763  4 87.251 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "ft_region"
```

```
## Linear hypothesis test
##
## Hypothesis:
## northeast = 0
## northcentral = 0
## south = 0
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1   4766
## 2   4763  3 25.189 3.656e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "ft_race"
```

```
## Linear hypothesis test
##
## Hypothesis:
## white = 0
## black = 0
## hisp = 0
```

```
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1   4766
## 2   4763  3 20.388 4.005e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## [1] "ft_industry"

## Linear hypothesis test
##
## Hypothesis:
## miningconstruction = 0
## durables = 0
## nondurables = 0
## Transport = 0
## Utilities = 0
## Communications = 0
## retailtrade = 0
## wholesaletrade = 0
## finance = 0
## hotelsrestaurants = 0
## Medical = 0
## Education = 0
## professional = 0
## publicadmin = 0
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
```

```
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1   4777
## 2   4763 14 13.847 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
## [1] "ft_occupation"
```

```
## Linear hypothesis test
##
## Hypothesis:
## manager = 0
## business = 0
## financialop = 0
## computer = 0
## architect = 0
## scientist = 0
## socialworker = 0
## postseceduc = 0
## legaleduc = 0
## artist = 0
## lawyerphysician = 0
## healthcare = 0
## healthsupport = 0
## protective = 0
## foodcare = 0
## building = 0
## sales = 0
## officeadmin = 0
## constructextractinstall = 0
## production = 0
##
## Model 1: restricted model
## Model 2: lnrealwg ~ (sex + wave + ba + adv + northeast + northcentral +
##     south + white + black + hisp + yrsftexpfz + yrsptexpfz +
##     yrsftexpfzsq + yrsptexpfzsq + sch + miningconstruction +
##     durables + nondurables + Transport + Utilities + Communications +
##     retailtrade + wholesaletrade + finance + hotelsrestaurants +
##     Medical + Education + professional + publicadmin + manager +
##     business + financialop + computer + architect + scientist +
##     socialworker + postseceduc + legaleduc + artist + lawyerphysician +
##     healthcare + healthsupport + protective + foodcare + building +
##     sales + officeadmin + constructextractinstall + production +
##     wtrgov + unjob) - wave
##
## Note: Coefficient covariance matrix supplied.
##
##   Res.Df Df      F    Pr(>F)
## 1   4783
## 2   4763 20 27.841 < 2.2e-16 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```