

Analyses of the Gender Wage Gap by Education

Matthew Carey

The University of North Carolina at Charlotte

Data Science 1302-001

Professor Justin Grandinetti

Professor Marco Scipioni

November 18, 2023

Analysis of the Gender Wage Gap by Education Level

Introduction

Combing through the data provided in the Wages by Education dataset, there are a multitude of questions to be asked and answers to be searched for. Of those questions, the one most intriguing is given the aggregated wages of men and women separated categorically by their corresponding education level, including less than a high school degree, high school graduate, some college education, bachelor's degree graduate, or advanced degree graduate, is there a trend illustrating the reduction of wage gaps between the aforementioned genders since 1973? The initial hypothesis is that the wage gaps have diminished over time through a multitude of facilitations including, but not limited to, expanded workplace diversity initiatives, legislative measures, and a massive influx of employment opportunities. Given that a trend in the reduction of wage gaps has existed since 1973, can we predict when we could expect these gaps to be eradicated? Searching for the answers to these questions through modeling and the established literature, we will also be exploring whether or not the disparity in education levels presents more pronounced gaps between the genders represented in the dataset?

Context and Implications

The gender wage gap in the United States is a phenomenon that has been a question of study for decades. Women began joining the workforce in droves during the Second World War. According to the Bureau of Labor and Statistics (2017), “ In 1950, there were 18.4 million women in the labor force, which accounted for about one-third of the total labor force.” Around the turn of the millennia, Women’s representation in the labor force peaked at 46.8 percent. Since then, the percentage of the labor force made up of women has slightly declined, however, the number of women participating in the labor force is still increasing. (Morisi & Toossi, 2017). The expansion of women’s participation in the workforce is due to a variety of reasons. A large part of these reasons include the aspirations of young adults relative to the employment marketplace. These aspirations are linked to their desired standard of living and marital aspirations or prospects (Macunovich, 2012).

Since the dawn of women becoming a major contributor to the workforce in the United States, there has always existed a discrepancy in pay between men and women, in favor of men. Originally, this was attributed to the less laborious and skill-based employment positions women held. Since the 1950s, women have begun to compete with men for more valuable positions in the workforce. In Blau and Kahn’s article (2000), *Gender Differences in Pay*, they state:

Over the past 25 years, the gender pay gap has narrowed dramatically and women have increasingly entered traditionally male occupations. These two labor market outcomes are closely linked, since considerable research suggests that predominantly female occupations pay less, even controlling for measured personal characteristics of workers and a variety of characteristics of occupations, although the interpretation of such results remains in some dispute.

Another factor that has led to the trend in the diminishment of the gender wage gap includes the overarching increase of education amongst women. According to *The Gender Wage Gap in the New Millenium: An Analysis of the United States 2000-2020*, “Women’s education rates have increased at a significantly higher rate than their male counterparts during this time” (Storrie et al., 2023). Although the gap has narrowed dramatically, in accordance with women entering careers that were predominantly dominated by men, and having a general increase in level of education, a gap still exists and efforts need to be made to continue on course for a more equitable solution.

With these strides of advancement women have made in recent decades, why are we still experiencing the gender wage gap? According to Mandel and Rotman, two factors devalue women’s wages in the workplace, especially those with higher levels of education, the devaluation theory and the glass ceiling theory. The devaluation theory suggests that work done by women with an educated background is not perceived to be as valuable as the work done by men with similar educational experiences. The glass ceiling refers to an unofficial barrier to the top echelon of the workplace chain of command. The glass ceiling theory supports the devaluation theory because of the overrepresentation of men at the top of the corporate hierarchical structure (Mandel & Rotman, 2021).

With the massive influx of educational attainment women, in general, have experienced, the “equalizing effects of increasing gender equity in postsecondary education are not being fully developed or realized”. This can be attributed to the segregation of genders across collegiate majors. One statistic shows, “Female bachelorettes earn significantly less than their male peers even when employed full time: the marginal female deficit in yearly earnings ranges from \$4,500

among the 1993 cohort to more than \$6,800 among the 2008 cohort” (Shauman, 2016). Many majors traditionally men or women-dominated, remain dominated by their respective gender.

Another aspect as to why the gap still exists comes from Goldin’s Article, *A Grand Convergence: Its Last Chapter* (2014). Goldin argues that the remaining disparity in the gender wage gap exists because:

hours of work in many occupations are worth more when given at particular moments and when the hours are more continuous. That is, in many occupations earnings have a nonlinear relationship with respect to hours. A flexible schedule often comes at a high price, particularly in the corporate, financial, and legal worlds.

This is especially apparent for highly competitive careers because more women are affected by domestic responsibility constraints placed upon women which limit the time and effort available to invest in a career (McGee et al. (2015). Workplace and schedule flexibility differs for men and women because of societal and biological reasons mainly revolving around the aspects of childbirth. Studies reveal a drop of at least eighteen percent in women’s workforce participation after childbirth for women of every level of education. This holds alongside data from Surveys of Income and Program Participation collected on mothers during two years before and after childbirth (Bailey & DiPrete, 2016).

Of the societal factors with influence on the gender wage gap, there appears to be major differences between the genders and their respective approach towards careers. In two high school-cohort longitudinal surveys, several sizable differences were illustrated between men and women. Men valued leadership and wealth nearly ten percent higher than women when selecting a career. In contrast, women ranked “Opportunities to work with people rather than

things” and “Opportunities to be helpful to others or useful to society” more than ten percent more valuable in choosing a career path than men ranked the responses (Fortin, 2008).

Despite the existence of the gender wage gap, the trends do show progress. In 2020, the WEF reported that women in the United States are “relatively well-represented” in middle and high management roles. However, this is not reflected in the top echelon of roles in business. As for entrepreneurial opportunities, “lack of care infrastructure” and “lack of access to capital” barricade opportunities for women. The WEF also reports that the gap may take 257 years to close (Gharehgozli & Atal, 2020). There is still much work to be done on this front.

Our analysis involves looking at broad data centering on the gender pay gap concerning the differing educational levels. The stakeholders of our analysis include both young or working class, men and women in the United States. Women are the primary demographic impacted, due to the fact women are those most negatively impacted by the disparity in pay. Men are also a stakeholder in this analysis because a shift in the landscape of wages will affect the marketplace of wages.

Our analysis could result in both potentially negative or positive outcomes. Negative outcomes include the interpretation of the wage gap closing causing a withdrawal or reduction in support of many organizations working towards eliminating the wage gap. This outcome could cause a reverse trend by increasing the current wage gap as well as eliminating jobs involving the infrastructure built to eliminate the gap. Positive outcomes could be the restructuring of support into areas that will create the most impact on reducing the wage gap. This could include creating more resources for women of higher education to be better equipped to compete in the job market.

We are evaluating our project through the deontological framework. There has been a tremendous amount of effort invested by many organizations and researchers to get involved in reducing the gender wage gap. A person's opportunities, career, and wage should not differ based on gender or gender-dependent qualities. According to deontological ethics, wages should be based on fairness, to do anything else would be discriminatory. Furthermore, the mere definition of this ethical lens suggests wages should be universal and not based on gender. This is the lens through which we are approaching this project in researching the trends of the wage gap since 1973 and the supporting literature.

Measurement

We will be using linear regression to assess the Wages by Education dataset because the data is best illustrated as a linear trend when given the year and the corresponding wage. We need to assess the dataset to determine the subset of data relevant to our study. Our study focuses on the disparity in wages between men and women since 1973 and analysis of the trends of the dataset. "Our main research question is as follows: Given the aggregated wages of men and women separated categorically by their corresponding education level, is there a trend illustrating the reduction of wage gaps between the aforementioned genders since 1973?"

The dataset is split into many columns including a column for the year. This will be the independent variable for our linear regression models. The remainder of the dataset's columns are split up to give a wage value that is representative of the population of men or women in the workforce with a given education level. These wages are the dependent variables of our models. The education levels define the maximum level of education the men or women have achieved and these include those with less than a high school degree, high school graduates, some college education, Bachelor's degree graduates, and advanced degree graduates. We will be looking to

analyze the wage gaps or the difference between the men's wages and women's wages. The trends we are looking for would include whether or not the gaps are increasing or decreasing from 1973 to 2022.

Subsequent questions include can we predict when we could expect these gaps to be eradicated? and whether or not the disparity in education levels presents more pronounced gaps between the genders represented in the dataset? We consider the eradication of a wage gap as being the point where men's and women's wages are equal. This will be determined by the intersection of the models generated. We will also be observing the size of the wage gaps to determine the more pronounced gaps between the genders. This will be done by looking at how each model relates to its counterpart at each education level. The gaps with the highest difference between men's and women's wages will present the most pronounced gaps.

Data

Using the Wages by Education CSV, we initially divided the data set into sections. These sections were differentiated by the variables used to calculate a representative wage. These sub-datasets included datasets representing a timeline since 1973 with corresponding wages by:

- education level
- education level and gender
- education level and race
- education level, gender, and race

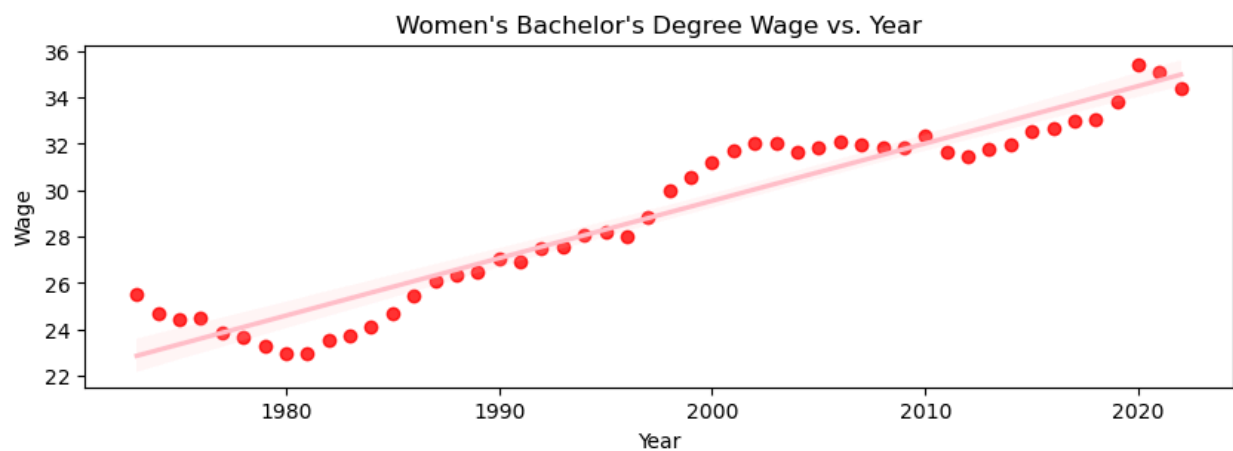
Our analysis and research question focuses on the differences in wages by gender, so the wages by education level and gender was the dataset we chose to analyze.

To begin working with the dataset, we created a pandas dataframe and subsequently divided that dataframe into two data frames representing data for men and women respectively.

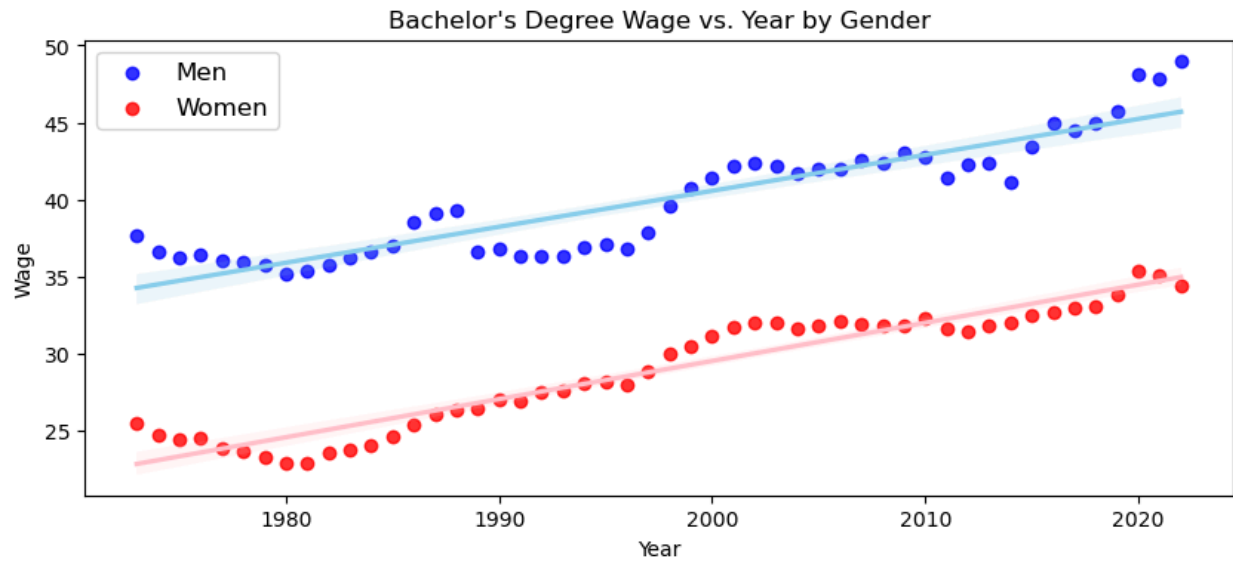
We used these datasets to create linear regression models for each gender, concerning the given educational levels. These education levels included less than a high school degree, high school graduate, some college education, bachelor's degree graduate, and advanced degree graduate.

The parameters of the created model were then stored to be used later in the analysis.

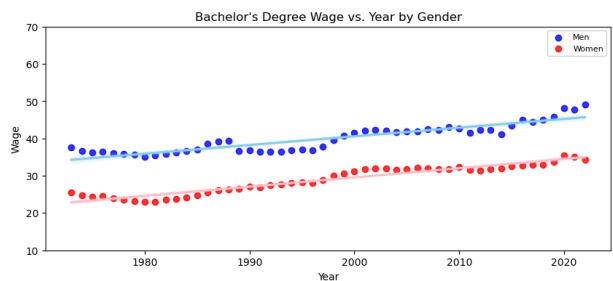
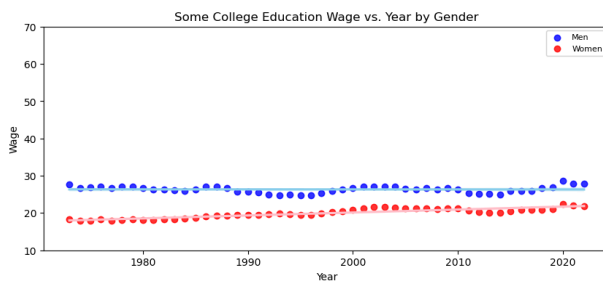
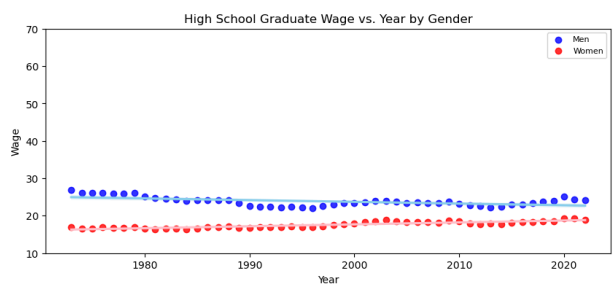
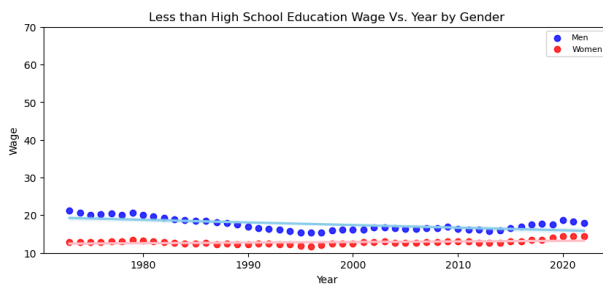
After creating the ten models representing each gender (five for men and five for women), we used seaborn to visualize the data. We created all ten visualizations for each model. Below is an example of one of these visualizations representing Bachelor's degree-educated women's wages since 1973.

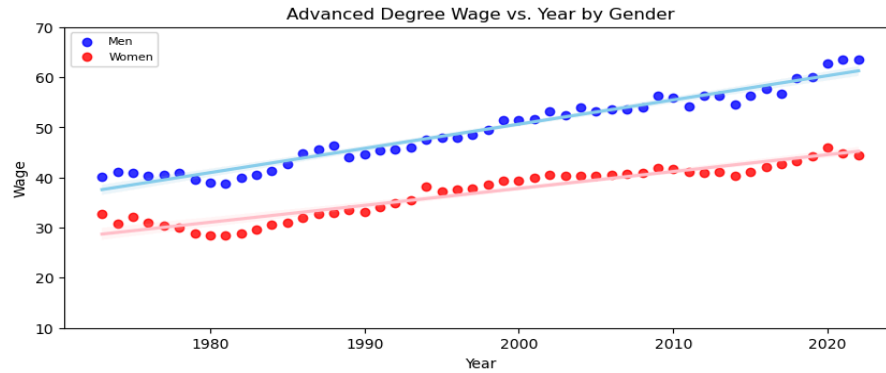


We then created a visualization that illustrates the disparity between men's and women's wages for each educational level (five combined visualizations) at the end of each section of our notebook. Below is the corresponding visualization representing both models created for men and women with a Bachelor's degree education.



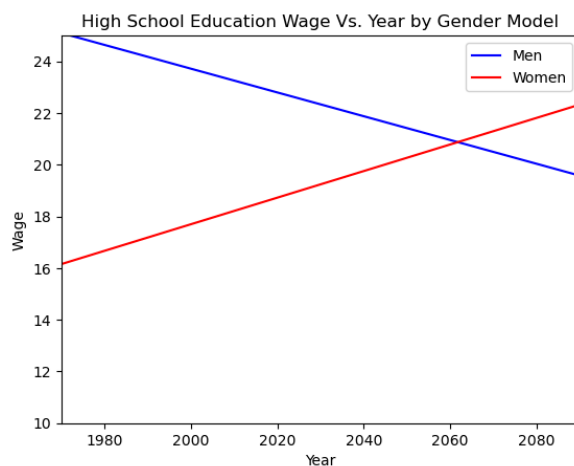
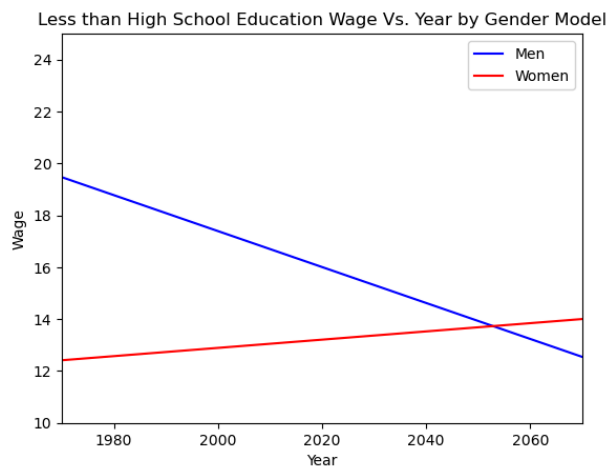
The next step in our process was to provide a snapshot of the five combined visualizations to better illustrate the disparities between men, women, and education levels. We did this by creating five more visualizations whose axes were equivalent. This was a necessary step to make the data more consumable from a holistic approach. These five models are represented below (in a larger format available in the corresponding Jupyter Notebook).

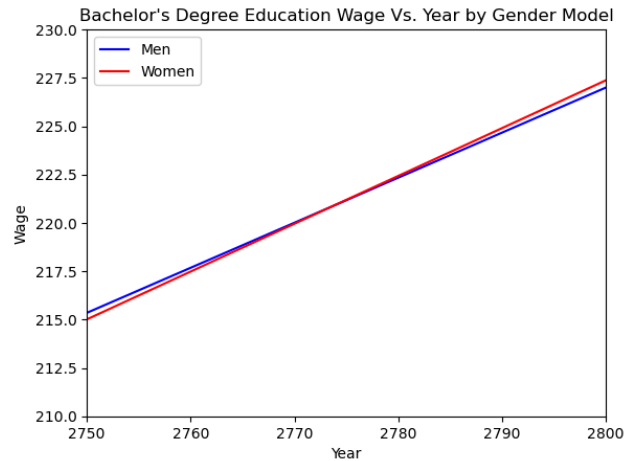




Looking at the above visualizations, a reduction of the wage gaps exists for the less than high school, high school graduates, and some college education models. We can also see that the most disparity exists at the Bachelor's and advanced degree levels of education.

Finally, we wanted to take our generated models and use them to predict when the current wage gaps were on trend to be eradicated. To do this, we built a function that created two line segments that were calculated based on the parameters we stored on each model's creation. This allowed us to use the Shapely library to find the intersection of these line segments. The x coordinate of the intersection gave us our predicted year that the wage gaps would disappear. We were able to calculate this for every education level except for the advanced degree. The corresponding visualizations of the intersections are below.





Using the function written to calculate the intersections, we were able to retrieve precise predictions correlating to the date the wage gap would dissipate per education level for the three lowest levels of education measured. For those with less than a high school degree, we found that the trend shows the wage gap ending in 2052. For high school graduates, the trend shows the gap disappearing in 2061. Lastly, for those with some college education, we calculated that the trend shows the gap ending in 2077.

The trend at the advanced degree level of education shows men's wages increasing at a rate faster than the women's wages, therefore the lines would not intersect. The bachelor's degree category also had an issue. The estimated year the gap would disappear was calculated to be 2773. This is due to the near equivalency of the weight the year holds on predicting the wage in this category. The men's wage increased at a rate of 0.2330 per year, while the women's wage increased at a rate of 0.2473 per year. The prediction of 2773 does not hold much value, but these two categories (bachelor's degree and advanced degree) do tell us that more focus and resources need to be allocated towards these categories, to reduce the wage gap at a faster rate. The bachelor's degree and advanced degree also illustrate the most significant wage gaps between men and women in the workforce.

The last important piece of information to look over is the summary of each of our linear models that have been discussed throughout this data section. The key statistic to look into is the R-squared values associated with each respective model. While looking at the less high school education level, the linear model for men had an R-squared of 0.380 while the model pertaining to women had an R-squared value of 0.168. This shows the linear relationship between less high school and wage is not necessarily strong, it is stronger for men as opposed to women. The next education level is high school and for this education level, men had an R-squared value of 0.299 while women had an R-squared value of 0.769. This shows the model with women has a much stronger correlation as opposed to the model with men. The same can be said for the education level of some colleges, in this model, men have an R-squared value of 0 while women have an R-squared value of 0.812. The analysis revealed no correlation for men, indicating that this model is not an effective predictor for men who have some college education. When looking at the models about men and women with bachelor's degrees, men have an R-squared value of 0.827 and women have an R-squared value of 0.919, this shows that at this education level, there is a stronger correlation for women as opposed to men. Finally, for the last education level of master's degree, men and women both had strong R-squared values. For men, the R-squared value was 1 and for women, the R-squared value was 0.908. The analysis of R-squared values suggests that the effectiveness of the linear models improves with higher levels of education.

Conclusion

Our findings conclude that at all of the education levels except for those with advanced degrees the wage gaps between men and women are closing. However, the trend showing wage growth for men and women with a Bachelor's degree is nearly equal, with women having a slight

edge over men. We were able to predict the eradication of the wage gap for those with less than a high school degree in the year 2052, those with a high school degree in the year 2061, and those with some college education in the year 2077. As stated, the prediction for the end of the wage gap for those with a Bachelor's degree is centuries away and the gap for those with an advanced degree is increasing. We also illustrated how the gaps corresponding to a Bachelor's degree and advanced degree education represent the most pronounced gaps in the dataset.

In the future, this study can be updated with new yearly values and the trends can be analyzed. The findings can be used to prioritize where to allocate resources or create new opportunities to reduce the wage gap between men and women. As of now, the areas with higher levels of education are not on course to correct the discrepancies in wages, whereas the gaps at the lower levels of education are on track to dissipate.

There are many limitations to this study and variables not explored. The aggregated wages by education per gender alone overlook many extraneous variables that would lend towards a non-discriminatory wage gap including, but not limited to, career fields, experience level, majors, and hours worked. We also did not have the numbers that could have been used to further manipulate the dataset including the percentage of representation in the workforce of men versus women by year or the percentage of representation in the workforce of the men broken down by education level by year and women broken down by education level by year.

References

- Bailey, M. J., & DiPrete, T. A. (2016). Five Decades of Remarkable but Slowing Change in U.S. Women's Economic and Social Status and Political Participation. *RSF : Russell Sage Foundation Journal of the Social Sciences*, 2(4), 1–32.
<https://doi.org/10.7758/rsf.2016.2.4.01>
- Blau, F. D., & Kahn, L. M. (2000). Gender Differences in Pay. *The Journal of Economic Perspectives*, 14(4), 75–99. <https://doi.org/10.1257/jep.14.4.75>
- Fortin, N. M. (2008). The Gender Wage Gap among Young Adults in the United States: The Importance of Money versus People. *The Journal of Human Resources*, 43(4), 884–918.
<https://doi.org/10.1353/jhr.2008.0006>
- Gharehgozli, O., & Atal, V. (2020). Revisiting the gender wage gap in the United States. *Economic Analysis and Policy*, 66, 207–216. <https://doi.org/10.1016/j.eap.2020.04.008>
- Goldin, C. (2014). A Grand Gender Convergence: Its Last Chapter. *The American Economic Review*, 104(4), 1091–1119. <https://doi.org/10.1257/aer.104.4.1091>
- Macunovich, D. J. (2012). Relative Cohort Size, Relative Income, and Married Women's Labor Force Participation: United States, 1968-2010. *Population and Development Review*, 38(4), 631–648. <https://doi.org/10.1111/j.1728-4457.2012.00530.x>
- Mandel, H., & Rotman, A. (2021). Revealing the Concealed Effect of Top Earnings on the Gender Gap in the Economic Value of Higher Education in the United States, 1980–2017. *Demography*, 58(2), 551–570. <https://doi.org/10.1215/00703370-9009367>
- McGee, A., McGee, P., & Pan, J. (2015). Performance pay, competitiveness, and the gender wage gap: Evidence from the United States. *Economics Letters*, 128, 35–38.
<https://doi.org/10.1016/j.econlet.2015.01.004>

Shauman, K. A. (2016). Gender Differences in the Early Career Outcomes of College Graduates:

The Influence of Sex-Type of Degree Field Across Four Cohorts. *RSF : Russell Sage*

Foundation Journal of the Social Sciences, 2(4), 152–193.

<https://doi.org/10.7758/rsf.2016.2.4.06>

Storrie, C. L., Lee, T. W., & Matzel, J. M. (2023). The Gender Wage Gap in the New

Millennium: An Analysis of the United States 2000-2020. *Journal of Business Diversity*,

23(2), 26–39. <https://doi.org/10.33423/jbd.v23i2.6165>

Toossi, M. & Morisi, T. L. (2017, July). Women in the Workforce Before, During, and After the

Great Recession : Spotlight on Statistics. *U.S. Bureau of Labor Statistics*.

<https://www.bls.gov/spotlight/2017/women-in-the-workforce-before-during-and-after-the-great-recession/home.htm>