Identifying Changes to the Gender Wage Gap

Ryan McPherson

March 2021

Abstract

Explore gender pay gap through CPS data to observe changing trends and lasting effects of policy and social movements. With CPS data we explore the change in percentage of men's median salary that the women's median salary makes up before and after 2009. This provides an estimate of the effects of the Wage Fairness Act Passed in 2009. After performing this discontinuity analysis for the median salary of all individuals age 15 and above, we separate the medians by age group and view the varying levels of change in wage gap before and after 2009. It becomes evident that the effect was more pronounced in younger individuals, which allows for the exploration of statistical significance. It is found that none age grouping discontinuities produce statistically significant results. Finally we discuss possible reasons the effects may not be clearly stated and hypothetical data that may allow for more meaningful results to be displayed.

1 Historical Significance

There is a documented disparity in pay between men and women doing identical jobs. This systematic difference in pay is commonly referred to as the gender wage gap. The inequality in pay likely originated alongside the origin of the antiquated unequal view of men and women. Focusing on the United States of America, we see the gender wage gap emerging as a political issue in the 1860s [1]. While the fight for equal rights continued, eventually in 1920 women's suffrage lead to the right to vote, however the wage gap was not addressed. It would not be until 1963 with the Equal Pay Act, that legislature would address the difference in pay [2]. The Equal Pay Act stated that men and women are to receive equal pay for jobs requiring equal skill, effort, responsibilities and working conditions. The guidelines of this legislation left room for companies to maintain dissimilar pay for men and women in similar

roles by claiming they were not identical. In 1980 it was estimated that women age 25 to 34 earned 77 cents for every dollar earned by men. While efforts never ceased to bring proper pay equity, little was passed by governing bodies. The next major legislative break through came in 2009. In 2009 congress passed the Wage Fairness Act which addressed the gender wage gap further through targeting pay equity. This piece of legislature attempted to properly instill the equity designed in the Equal Pay Act of 1963.

Today the gender wage gap continues to plague the United States. With reports publishing that as of 2020 women earned approximated 84% of men's salaries on average across the country [3]. While a noticeable increase from the the 77% that was seen in the 80's, the question is raised as to what methods made noticeable change. Today the United States has a significantly different social climate, with 40 years of changes many different factors are likely at play. This paper will set out to find if any actionable efforts made large strides in decreasing the gender wage gap.

2 Introduction

Using United States Current Population Survey data I will explore the changes in gender wage gap over time. Changes in the wage gap can be attributed to a variety of factors, ranging from social movements to political policy. A focus will be put on measuring the effects of legislation passed to regulate the gender wage gap. The legislature of interest in this paper is the Wage Fairness Act of 2009. This will act as our treatment for a regression discontinuity approach. Using data on income from the Current Population Survey, the effects over time to the median wage of men and women will be able to be measured, and possibly attributed to the Wage Fairness Act. The Current Population Survey, or CPS, allows for the comparisons of median wages for United States citizens of various gender and ages from a range of years. Identifying the years in which legislation was passed and put into effect we can begin to estimate the effects of legislation on the gender wage gap. We can

observe the trends in change of the gender wage gap, then isolate the effects of legislation by observing either a discontinuity or escalation in change at the years following the regression.

3 Data

United States Current Population Survey data provides the basis for our experimentation. The data used comes from the historic listings of median salaries for men and women. The data set being used contains the recorded median salaries for both sexes by year ranging from 1947 to 2020. We also have a column of median income for both sexes adjusted for 2020 dollars. We will use this for our regressions as to eliminate the disparity in incomes that we observe from inflation. The age grouping of the data contains the median for all individuals 15 and above, and then separately median incomes for ranges of 15-24, 25-34, 35-44, 45-54, and 55 and up. The data is originally formatted in an easy to digest excel sheet, with lots of text descriptions and summary information. For our use case we will need to reformat the data into a more usable standard. Columns for year, number of observations, median mens salary, median women's salary, both salaries adjusted to 2020 dollars were created and saved into the data set. Each column was created for the ranges of ages and one set of columns was created for the 15 and up category. After estimating effects of legislature using the age 15 and above data, we can then dive into the effects of legislature in the breakdown of wage gap by age range. A downside to the data is the large bin size, containing a year's median as a single point. However we do not have to worry about reporting bias as the data is collected from employers, meaning that incorrect salaries are not reported.

We begin by viewing the age 15 and up median incomes for both men and women over the length of the data. Pictured in figure 1, we have a graph displaying the trends of median wages for all individuals age 15 and over. This graph displays the movement of the median wages, showing both an upward trend and a decrease in distance. However in this graph it becomes instantly clear that the median salary for women is below the median salary for men

when observing individuals age 15 and above. A variable, $percentage\_of\_mens$ is generated for each year, containing the percentage of mens median salary that womens median salary makes up. We have $percentage\_of\_mens_{year} = \frac{womens\_median_{year}}{mens\_median_{year}}$, an important point is this does not account for weighting of observations at the given median. Figure 2 shows the corresponding percentage of men's wages that women's median wage account for. Important points of observation of figure 2 come with the beginning upward trend soon after 1963 and the large change in slope that occurs soon after 2009. These points represent the changes that we will explore being attributed to legislature. When observing the wage gap for all individuals ages 15 and up, the salary of women appears to be much lower than commonly reported, finishing below 70% in 2020.

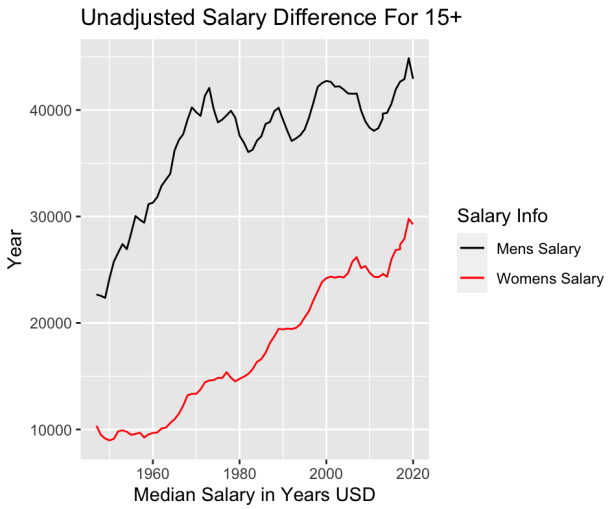

Figure 1: Both men and women's median salary by year for 15 and over

Figure 2: Percentage of mens median salary that women's median salary makes up by year for 15 and over

To explore the seemingly overstated difference we can plot the percentage of the median men's salary that the women's salary makes up for the available age groups. We can raise questions about the average age of entry to the workforce and possibly average age of exit of the workforce. We would expect that younger men and women's median salary should be the closest. This would be due to their recent entry to the workforce and lower amounts of observations away from minimum wage. To generate the percentage of we have

$percentage\_of\_mens_{(age\_group)(year)} = \frac{womens\_median_{(age\_group)(year)}}{mens\_median_{(age\_group)(year)}}$. This will also give us important insight into the possibility of a heteroskedastic relationship within the data.

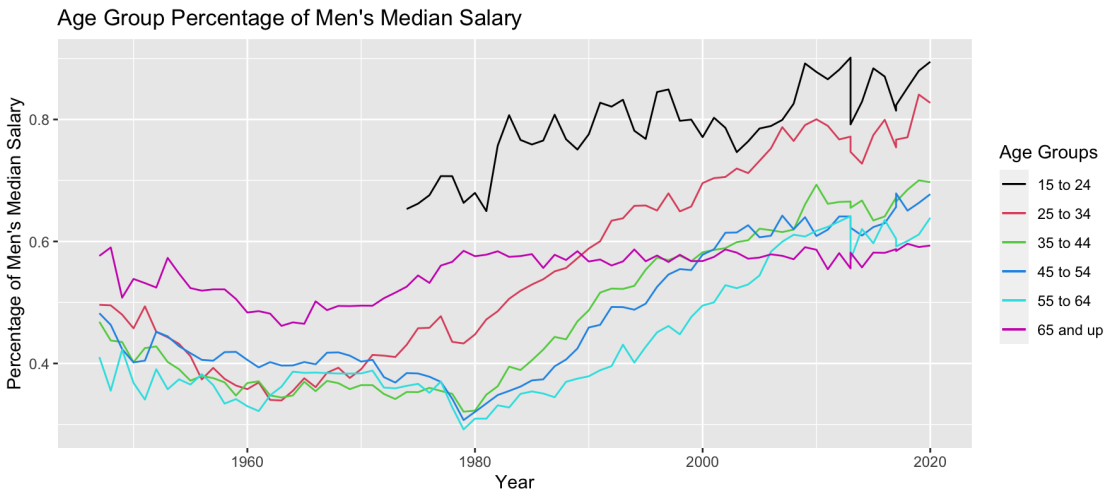

Figure 3: Percentage of lines for all age groups

While the observations for age group 15 to 24 only begin at 1974, our expectations are nearly met. In figure 3 we see an increase in the wage gap as age increases, represented by the downward jump of the lines of higher age groups. Additionally, age groups 15 to 24 and 25 to 34 see the lowest difference in gender wage gap. However these observations are by no means flawless, but serve as a starting point for our regressions. Factors such as years of experience, profession and hours worked may act as unobservable in our data which could account for a fundamental difference in wages which is not attributed to companies systematically paying women less.

4 Discontinuity Approach

We will begin with attempting to measure a discontinuity at 2009. 2009 was the year which the Wage Fairness Act was passed by congress. Before we spend time creating a linear model for estimating the discontinuity, we must ensure the data is not heavily biased. First we want to ensure that we have proper balance on both sides of our regression discontinuity year, 2009. We will observe linear regressions for the number of observations with a cut off

at 2009. We create a graph that shows the median salary for men and women at each year ranging from 4 years before 2009 to 4 years after 2009. On this graph we calculate linear models of the data, where the number of observations of the median salary are estimated in terms of years. We create a linear regression for the left hand side, from 2005 to 2009, and the right hand side, 2009 to 2013. We then overlay these linear estimate line segments and observe if there is a significant discontinuity at 2009.

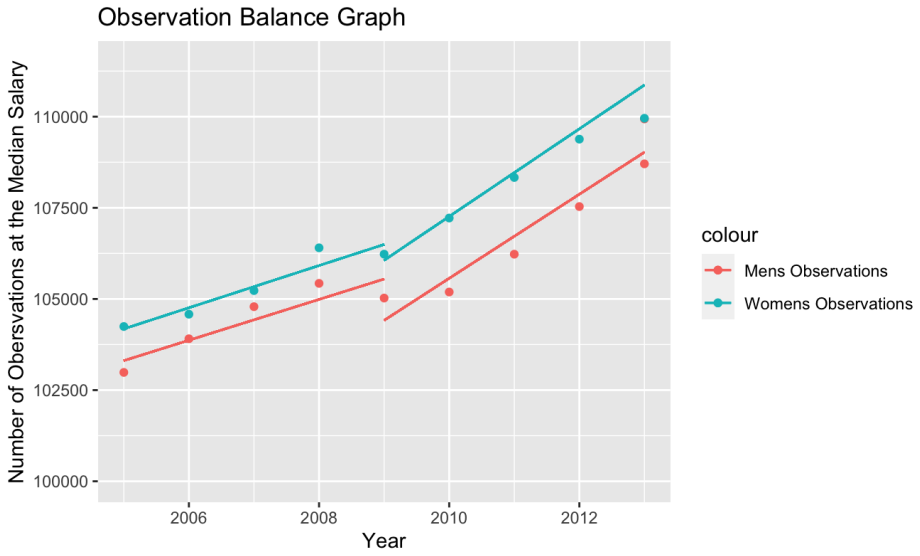

Figure 4: Balance Graph of Observations

As displayed in figure 4, the balance graph shows linear estimates to appear appropriately close. This confirms that there is not an unexpected increase in observations that could represent a skew or misreporting of data prior to or following the passing of legislature. It would be ideal to have multiple non-correlated characteristics, however we are limited by the characteristics contained in the CPS data. We can now continue to estimate the regression discontinuity that occurs at 2009. We use a linear model to estimate the median wage for men and women with a discontinuity at 2009. Additionally a discontinuity regression for the percentage of male median wage that female median wage makes up. Upon running the regressions we can create the following tables.

Evidently, figure 5 shows very little change in 2009. This indicates that for the median salary of men and women age 15 and up, legislature in 2009 did not have a great effect on

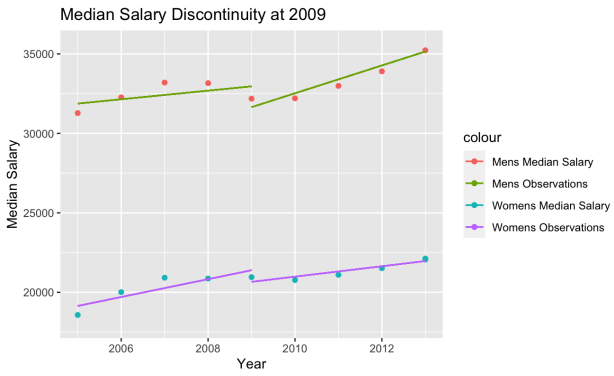

Figure 5: Linear estimates of the median salaries by gender, discontinuity at 2009

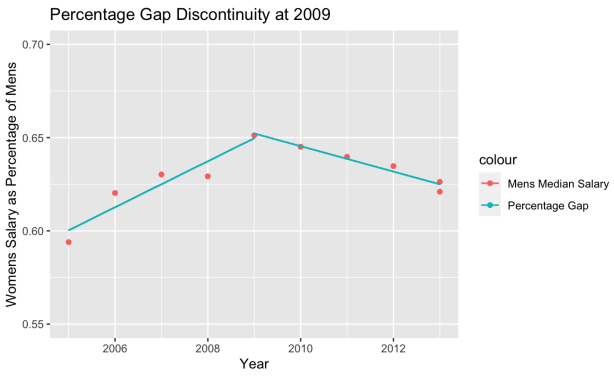

Figure 6: Linear estimates of the percentage wage gap, discontinuity at 2009

the trend of their median salary. This is backed up by the nearly touching regressions in the figure 6 graph. With such low differences in discontinuity it is safe to assume that there was not a large shift in the wage gap that could be attributed to the legislature passed in 2009. One potential concern would be capturing a change in median salary attributed to the 2008 financial crisis. The benefit of observing the median salary instead of the mean salary means that the 2008 financial crisis was unlikely to reveal itself in the median. While we did not see a large change in the median salary for all individuals age 15 and up, we may see changes that are overlooked in the age groups. We shall explore the various age groups to see if the results may have been more pronounced for some grouping of individuals. We will recreate the discontinuity of percentage wage gap for each age grouping. Then we can combine the plotting of all the discontinuities to view if any have a large gap at 2009.

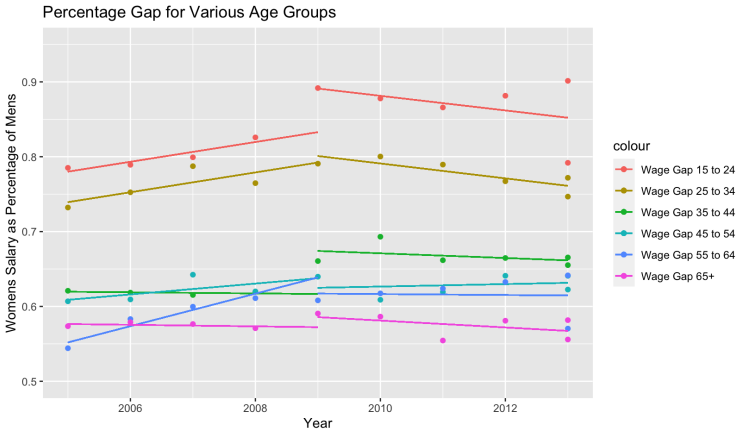

Figure 7: Wage gap discontinuity for all age groups

From this we can graph the difference's in the points at 2009 to better observe what the measurable distance change is between age groups. This will assist in a visual representation of the effectiveness of the 2009 legislature. Following the graphing of the discontinuity, we will compute the statistical significance of the change in gender wage gap. The graph of the changes in intercepts as a bar chart as follows.

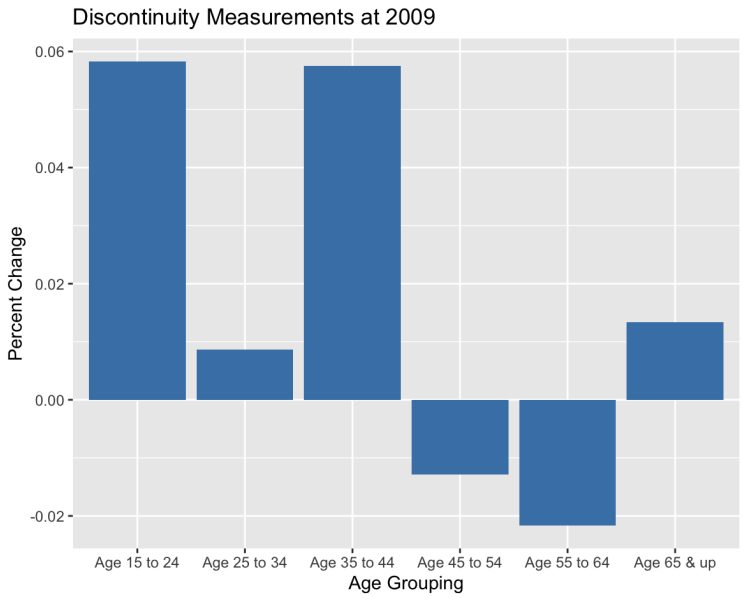

Figure 8: Chart of differences in estimates at 2009

Now we move to addressing the statistical significance of the change observed. Using the summary command in R to observe the p value of regressions coefficients with a control variable for a year being before or after 2009. The estimates produce a p value for the cut off, or our changes in gender wage gap. None of the regressions resulted in the post-2009 variable to appear statistically significant. This indicates the change in wages does not appear to be statistically correlated to the year 2009. While evidence that a strong correlation was unable to be established, it does not mean the legislature was completely ineffective. But the main question being asked, if the gender wage gap decrease after 2009 can be attributed to events of 2009, has been answered as no.

5 Identifying Potential Years of Large Change

Working backward from identifying if a single year created causal effects, we can attempt to observe years which did have large effects. If we are able to identify any, we can then research events that occurred during the years in an attempt to view these events possible role in changing the gender wage gap. We will use the 15 and older median wages as it is representative of the country's entire work force. This opens the review to contain years 1947 to 2020. We will begin by finding the largest discontinuity between wage gaps for all given years. We find that 2013 to 2014, 2004 to 2005, and 1979 to 1980 have the largest increases. We will continue with the discontinuity process for each prior year, 1979, 2004, 2013. We start by generating a linear regression using the left 5 years and right 5 years. The t-statistics produced are not statistically significant for 2004 and 2013. While it appears that the regression performed at 1980 appears to be highly statistically significant, we can explain the misconception with our own data. Observations for the age group 15-24 begin at 1976, which is captured within our regression. At this point, we are generating an improper estimate and the discontinuity is tied to the increase in reporting of data. After finding this, we can conclude that no individual year had a statistically significant positive impact to the gender wage gap.

6 Conclusion

Exploring the effects of the Fair Wage Act on the gender wage gap produces somewhat dull results. With our findings indicating that no individual year had a statistically significant positive change to the following years gender wage gap. While this does not allow us to conclude if the legislature was effected or not, we do observe a positive trend in the percentage of mens salary that women's salary make up. Albeit this observation is by no means original, it enforces the idea that the topic could be more properly explored in an alternative way. We can now explore places where our methods and data may have fallen short, or could

be improved up. Starting with the data, we view a lack of observations for the median salary. With a bin size of one year, we lack the finer detail necessary to display more jarring adjustments to the gender wage gap. When our regressions are relying on very few points, only a couple years as to not include the effects of too many other years in our analysis, we see a large amount of inaccuracy. Having data that contained the median salary on a monthly basis would open the door to a much more fine tuned approach. Additionally, it would have been nice to have a continuous age variable. Instead of a grouping of 10 years, a continuous age variable would have allowed for a histogram to display the distribution of wage gap changes. I believe this would have appeared as more easily understandable and give more insight than the bar chart. Finally, if the data were to contain more characteristic variables beyond male and female salaries, it would pose for better control variables in the regressions. Controlling for years of experience, education and location would allow for even more meaningful insights on the effectiveness of legislature.

As far as methods used, I believe there is room for improvement, however the I believe the techniques employed were acceptable. Producing no causal effects or highly statistically significant figures would be better than forcing or misinterpreting information for the sake of producing findings. I was interested in running a regression discontinuity for nearly every year, however was stopped by the implications of multiple inference. If we were to observe 65 regressions we would assume that, at a 95 confidence interval, we would have a 96% chance of revealing a false positive. The equation $1 - (1 - 0.05)^6 5 \approx .96$, backs up our statement. With this in mind, it appeared dishonest or disingenuous to perform such a set of regressions all in the hopes of essentially finding a false positive. I would have also really liked to make use of a difference in differences approach, however the lack of a control observation makes this difficult. I am believe I approached the questions raised with a proper method given the information and data at hand.

References

[1] Daugherty, Greg. "The History of the Gender Wage Gap in America." Investopedia, https://www.investopedia.com/history-gender-wage-gap-america-5074898. Accessed 27 Nov. 2021.

[2] Equal Pay for Equal Work — U.S. Department of Labor. https://www.dol.gov/agencies/oasam/centers-offices/civil-rights-center/internal/policies/equal-pay-for-equal-work. Accessed 27 Nov. 2021.

[3] Gender Pay Gap in U.S. Held Steady in 2020 — Pew Research Center. https://www.pewresearch.org/fact-tank/2021/05/25/gender-pay-gap-facts/. Accessed 29 Nov. 2021.