

Development of wage gaps In the united states

An empirical investigation of the gender and racial wage gap from 2001 to 2011



Master Thesis

Economics of Markets, Organisations and Policy

Eric van Tol

Supervisor: **Dr. M. Bøg**

**Erasmus University Rotterdam**

**Date: 01-08-2013**

# Abstract

This paper builds upon the vast amount of literature on wage gaps and investigates the development of the wage gaps across the United States in the period 2001-2011. It extends the previous research by examining a new time period and by looking at the wage gaps on a state level. To analyze these pay gaps data from the American Community Survey is used. I use two methods to investigate the wage gaps, Oaxaca decomposition and the Juhn Murpy Pierce decomposition, with the focus being on the later. What I find is that the gender wage gap has decreased by 5.5 percent points and the racial wage gap has increased by 4.1 percent points. On a state level most states follow these directions, but with a great deal of variability. For the gender wage gap only two states don’t follow the average direction, i.e. they show an increase in the gender pay gap. For the racial wage gap seven states show a decrease in the pay gap.

***Keywords:*** *Wage gap, gender, racial, Oaxaca, JMP, United States*

Table of Contents

[Introduction 4](#_Toc362971789)

[Literature Review 7](#_Toc362971790)

[What is a wage gap 7](#_Toc362971791)

[Sources of pay gaps 7](#_Toc362971792)

[Research methods and data 10](#_Toc362971793)

[Methodology 13](#_Toc362971794)

[Methods 13](#_Toc362971795)

[Oaxaca 13](#_Toc362971796)

[Juhn Murphy Pierce 15](#_Toc362971797)

[Juhn Murphy Pierce extended for time 17](#_Toc362971798)

[Model Specification 18](#_Toc362971799)

[Areas of investigation 20](#_Toc362971800)

[Expectations 21](#_Toc362971801)

[Data 22](#_Toc362971802)

[Data Source 22](#_Toc362971803)

[Transformation of variables 23](#_Toc362971804)

[Descriptive statistics 25](#_Toc362971805)

[Results 27](#_Toc362971806)

[Male/Female overview 27](#_Toc362971807)

[Male and Female regression 27](#_Toc362971808)

[Male/Female Oaxaca decomposition 28](#_Toc362971809)

[Male/Female JMP decomposition 28](#_Toc362971810)

[White/Black overview 29](#_Toc362971811)

[White and Black regression 29](#_Toc362971812)

[White/Black Oaxaca decomposition 29](#_Toc362971813)

[White/Black JMP decomposition 30](#_Toc362971814)

[White/Black Male/Female detailed 30](#_Toc362971815)

[Yearly Male/Female and White/Black 32](#_Toc362971816)

[Wage gaps across states 35](#_Toc362971817)

[Differences in wage gaps over time and across states 38](#_Toc362971818)

[Conclusion 41](#_Toc362971819)

[Limitations and further research 43](#_Toc362971820)

[Bibliography 44](#_Toc362971821)

[Appendix 47](#_Toc362971822)

[Methodology 47](#_Toc362971823)

[Self-employed/Working for wages 47](#_Toc362971824)

[Results 50](#_Toc362971825)

[Male/Female overview 50](#_Toc362971826)

[White/Black overview 51](#_Toc362971827)

[Yearly Male/Female 53](#_Toc362971828)

[Yearly White/Black 54](#_Toc362971829)

[State Male/Female 55](#_Toc362971830)

[State White/Black 58](#_Toc362971831)

[Male/Female State over time 61](#_Toc362971832)

[White/Black state over time 63](#_Toc362971833)

# Introduction

It is the year 2000 and the imaginary company run-of-the-mill incorporated has just hired two recent graduates James and Mary. James and Mary have known each other since childhood and have always been very competitive with each other. This competitiveness has helped both of them throughout their educational careers each always trying to outperform other and if that was the case only by a small margin. James and Mary are very much alike, both went to the same schools they were even in the same classes and this similarity drove them both to apply at run-of-the-mill Inc. In the end run-of-the-mill found them both very suitable for their company and decided to hire them. There is however one difference between them James will earn $40,000 whilst Mary will only earn $30,000, even though both will be doing the same job. So why does this happen?

One could argue that there is some unobserved skill difference between the two that firm observed and hence adjusted their wage offer. However, there is a vast amount of literature that suggest that women on average earn less than men. This difference between men and women appears to be pervasive, although it has lessened over the last decades. This gender wage gap is not the only wage gap that exists, another major wage gap is a racial one the wage gap between whites and blacks[[1]](#footnote-1). And those two gaps lie at the heart of what is investigated here. I aim to investigate what has happened to these two wage gaps in the period 2001-2011 in the United States and across the different states.

This is useful to know, because although there has been much research into these gaps, most of it focusses on either one of the wage gaps, thus missing the ability to compare the two wage gaps. Also most research only looks at the United States as a whole, whilst on state level the picture could be quite different. Furthermore this research provides insight into a more recent period than the studies before and thus allows the investigation of the developments of trends described in earlier papers.

Graph 1, \* is an approximation of earnings ratio

Graph 2,\* is an approximation of wage gap

Graph 1 and 2 give a quick overview of the trends in the wage gap that have been found in previous research and how my research fits in with those. The first graph shows the average racial wage ratios and thus higher values represent more equalitarian pay. It demonstrates that from 1940’s till the 1980’s the wage ratios improved, but from the 1980’s until now the wage ratios show a decline. The second graph depicts the average raw gender gap, i.e. the unadjusted wage gap, and shows that since the 1960’s until now the wage gap has been improving, i.e. declining, although it has been at a declining rate.

To answer the main question several different aspects are investigated. Firstly, a look at the entire state of the wage gaps is taken. After which these two wage gaps are further split up to be able to look at four groups; white males, white females, black males and black females. This is done to provide greater insight in how these wage gaps differ when they are “combined”, i.e. being black and being female. After this the main research questions are investigated first by looking at the development of the two wage gaps over the entire United States and then moving to the development of these pay gaps over the different states.

All these wage gaps are also decomposed into parts that give a starting point as to where these wage gaps come from. This provides answers to questions such as: Are females less educated on average or are they paid differently for the same skills. Is the average black worker less experienced? The decomposition technique mainly used is the Juhn Murphy Pierce method, this method is used as the main method, because it does not only allow for an investigation at the mean but also for an investigation of the different percentiles of the wage distribution.

The main findings using this method are as followed. The average gender wage gap has declined over the period 2001-2011 by about 5 percent points, from about 25% to about 20%. The mean racial wage gap on the other hand has increased by about 5 percent points over the same period. The results also show that across states both wage gaps differ by large margins and even within the state the wage gaps differ considerably, i.e. states with a low gender wage gap do not necessarily have a low racial gap. They further indicate that both wage gaps have different sources and when looking at the four groups it becomes apparent that being female and being black does not have a constant effect.

The rest of this paper has the following structure. The next chapter describes what a wage-gap is, what possible sources for these wage gaps are and methods and data other research uses. The chapter after that describes in detail the methods used and to which model these methods are applied. Then the data used is described and how this data is transformed into the correct model. The results are then presented, after which this paper ends with its conclusion, its limitations and possibilities for further research.

# Literature Review

*This section will define what a wage gap is. It will also investigate the sources of these pay gaps and discuss the methods and data used by other papers.*

## What is a wage gap

A wage gap’s definition is mainly determined by what is used to measure it. It will always give an indication of how wages of two groups differ, but how they differ depends on the definition. The definitions are most abundant in the gender pay gap. The OECD defines the gender wage gap as: “The “gender wage gap” (in unadjusted form) is measured as the difference between male and female earnings expressed as a percentage of male earnings. The extent of the gap varies with the position of men and women taken as reference in the distribution of earnings”. The US Census bureau uses a similar definition: The wage gap is expressed as a percentage (e.g., in 2012, women earned 80.9% as much as men aged 16 and over) and is calculated by dividing the median annual earnings for women by the median annual earnings for men.

These two definitions only apply to the gender pay gap, but are easily modified to allow for different wage gaps. I choose to follow the OECD definition, because it allows more room for looking at different parts of the wage distribution. Thus the definition of wage gaps used in this paper is: A wage gap is the difference between two mutually exclusive groups’ earnings expressed as a percentage of the reference group’s earnings. As with the definitions of the US census and the OECD this wage gap can be determined at different levels. For the majority of this paper the wage gap is determined at the mean.

## Sources of pay gaps

When looking at wage gaps two approaches can be chosen, the individual and the institutional approach (Ferree and Mcquillon, 1994). The individual approach sees discriminatory wage gaps as being the result of isolated prejudiced individuals that decide upon wages. In this view there is only rare and non-systematic discrimination. The institutional approach however, is the more pervasive and historically the more often chosen approach. In the case of the gender pay gap it sees men as the original population of the system and females were later introduced in the system. This leads to a system where discrimination is the norm and the main source of it is not on the individual basis. Because previous research has found large wage gaps for both the gender and race wage gap, I choose to follow the institutional approach. It seems highly unlikely that these wage gaps would exist if the individual approach were to hold.

This approach still leaves much room for further investigation of why these gaps exist. Altonji and Blank (1999) provide a great overview of these possible sources and I will use them as a guide. One of the first things one can think of as a source of wage gaps is differences in qualifications. For the gender wage gap however this appears to not be the case. The Armed Forces Qualifying Text (AFQT) test-scores, which are generally viewed as a strong indicator of skill, show only minor differences between males and females. This could be because parents get both girls and boys and raise them in the same environment (Altonji and Blank, 1999). The inclusion of these AFQT test scores in a regression estimating the gender wage gap also does not considerably reduce these gaps, providing further evidence that males and females are equally skilled and thus for the gender pay gap this does not appear to be a source.

The picture becomes rather different when looking at the racial gender gap. Blacks score lower on the AFQT than whites and when this score is added to the wage gap regression. The wage gap drops from 12%, if they had same means as whites on schooling, industry and location to 4% when AFQT score is also included (O’Neill, 1990). Similar to this Neal and Johnson (1996) find that virtually all of the wage gap can be explained by differences in the mean AFQT scores of these groups. Winship and Korenman (1997) show that AFQT is strongly related to the levels of education, quality of that education and family background.

Another possible source of the wage gaps is job choice. It could be that these groups prefer or are just able to work in specific sectors or jobs and those could be in general higher/lower paying. It is difficult to determine if groups of people work in a certain sector/job, because they want to or because they have to, i.e. they are not hired in other sectors. Therefore most research into this topic brings results as: the fraction of women in an occupation has a negative relationship with the wages (Blau and Beller, 1987). Furthermore Sorensen (1990) suggest that this negative relationship explains between 15 to 30% of the male female wage gap. However research from Gerhart and Cheikh (1991) finds that this negative relationship may be due to the types of persons that choose to work in these “feminized” professions. This suggests that it is mainly a choice, but more research in this topic is needed.

For the racial wage gap the same problem of indistinguishabilty exist, but Tomaskovic-Devey (1993) finds that job-closure accounts for 38% of the racial pay gap and human capital differences for 31%. Job closure means that certain jobs are preserved for members of specific groups.

The two sources discussed before were mainly about before or right as someone enters the job market, but there are also sources that could play a role in the wage gaps once a person has been in the job market for a while such as experience and training. In terms of the gender gap a problem arises with respect to experience, because most datasets don’t contain actual work experience. A proxy is used in most research, namely potential experience[[2]](#footnote-2). Light and Ureta (1995) find that their specification with detailed control variables creates higher returns to experience, but lower returns to tenure as opposed to predominant specifications. Additionally they find that women recover faster from a drop in wage due to a career interruption and that drop is also smaller than that of males. Also they find that after nine years of work experience the wage gap starts to narrow. Wellington (1993) finds that men have a higher return to experience than women do, but accredit this to men and women acquiring different types of experience.

Women also tend to switch/quit jobs more often than men do (Becker and Lindsay, 1994; Sicherman, 1996). This in turn has a relationship with the amount of training women get in their firms. Overall they receive less training then males and this training has a significant effect on the male/female wage gap (Hill, 1995; Olsen and Sexton, 1996)

In the racial wage gap Bratsberg and Terrell (1998) do an extensive investigation into the effect of the experience. They find that experience has a higher return for whites than for blacks, but seniority has the same return for both groups. Some research even suggests that the return to seniority may be a bit higher for blacks (Altonji and Shakotko, 1987). Other literature however finds that whites and blacks have the same return to experience and tenure, but differ in the amounts they acquire of each (D’amico and Maxwell (1994). Their research however focusses only on the first five years in the labour market.

One final major source of wage gaps could be discrimination. However attributing part of the wage gap to discrimination is virtually impossible. This is because any continuous unexplained difference between groups or difference in returns could be caused by one or multiple latent variables. And thus any difference would not be discrimination, but for example unobserved skill. There is research that does suggest that there is considerable discrimination happening in the labour market. For example Goldin and Rouse (1997) use the fact that many orchestras in the 1970’s and 1980’s used blind auditions. What they found is that from 1970 to 1990 the proportion of females in those orchestras increased from 10% to 20%. Even more profound is their estimation that women had a 50% higher chance of getting through the preliminary round than when no blind auditions were used. Also sports data lends itself well to the investigation of discrimination. Kahn and Sherer (1988) use this attribute of sports data and find that non-white National basketball Association players earn less than their white counterparts with similar performance.

Overall there appear to be four main sources of the racial wage gap; differences in characteristics, differences in jobs, differences in experience and training, and finally discrimination. For gender there appear to only be three main sources. These are the same as in the racial gap, but without the differences in characteristics. With females appearing to have at least as good if not better characteristics than their male counterparts.

## Research methods and data

One of the first papers that uses a formal decomposition model is Oaxaca (1973). He develops his own model, the Oaxaca decomposition, to be able to decompose the wage gap between two groups. In his paper Oaxaca is investigating the wage gap between males and females. To find this wage gap he estimates a wage equation consisting of approximate experience and its square, age minus years of schooling minus six, years of education and its square and dummies for class of worker, industry, occupation, health problems, part-time workers, migration, marital status, size of the urban area and the region. For these estimates he uses data from the 1967 Survey of Economic Opportunity, focusing only on those that are sixteen years or older, living in an urban area and which report their race as either White or Negro. And with this data and method he reports a wage gender wage gap of 43% for whites and 30% for blacks.

Jaynes (1990) on the other hand uses earnings ratios in his investigation of the labour market status of black Americans in the period 1939-1985. Using data from the US census for the period 1940-1980 and the Current Population Survey (CPS) for 1985 he estimates the white/black earnings ratios at the mean and several percentiles. Then he further splits the data by investigating the wage ratio at either different education levels or different wage measures. He finds overall when looking at the weekly wage that the wage ratio increased from 47% in 1939 to 67% in 1984. For females the corresponding numbers are 41% to 97%.

Moving forward in time Blau and Kahn (1997 and 2004) use data from the Panel Study of Income Dynamics (PSID) and the CPS, focusing on the PSID, to investigate the gender pay gap. Using only data from full-time, non-agricultural employees between the ages of 18 and 65 and excluding those people that are self-employed. They use the Juhn Murpy Pierce decomposition (JMP), which allows for decomposition at different percentiles as opposed to only at the mean and can be extended for differences over time. This JMP method is applied to two regressions. First a wage equation with a race dummy, education variables (years of schooling and dummies for college and graduate degree) and experience variables (part-time and full-time experience and their squares) is estimated. The second equation includes these variables and variables indicating collective bargaining power, occupation dummies and industry dummies. They find that the gender wage gap declined from about 46% in 1979 to 23% in 1998.

Kim (2010) takes the Smith and Welch (1989) approach in investigating the racial wage gap across the period 1980-2005. This approach extends the Oaxaca decomposition by a time factor and thus allows for a comparison of wage gaps across time. This method is applied to a mincerian wage equation, with age and dummies for attained education levels. The wage equation further controls for region, metropolitan residence and the marital status. The wage equations are based of the CPS-Merged Outgoing Rotation Group (CPS-MORG), using individuals that are 25 to 65 years old males with positive income and not self-employed or in the military. Using this sample it is found that the male racial wage gap increased by 2 percent points.

Continuing with some recent international research, Simon and Russel (2005) use both the Oaxaca and the JMP decomposition techniques to investigate the sources of the gender pay gap in the European Union. For the Oaxaca method they use wage equations that control for education, type of contract, tenure plus its square, potential experience and its square and cube, logarithmic number of hours, part-time dummy, occupation dummies, management dummy and premium pay. The JMP method uses the same estimated wage equations. To estimate the equations they use data from the European Structure of Earnings Survey (ESES). What they find is that females tend to segregate into low-paying jobs and that differences in the magnitude of the wage gap and the wage structure are the main sources of differences across countries.

Fearon and Wald (2011) us the Brown Moon Zoloth (BMZ) method to look into the racial wage gap of Canada. The BMZ decomposition extends the traditional Oaxaca method to include the decomposition of occupational differences. They use data from the 2006 Canadian Census Public Use Microdata File (PUMF), restricting themselves to individuals between 18 and 64 years, who worked full-time and had positive earnings and removing those that are self-employed. In their wage equation they control for immigrant status, native English speaker, marital status, years of school, highest level of schooling, work experience and its square, province residing and if one resides in a metropolitan area. What they find is that wage discrimination and occupational segregation account of the majority of the racial pay gap, with differences in endowments only accounting for a small portion.

# Methodology

*In this section the choice of the methods used to investigate the wage gaps is explained, how these specific methods work, the implementation of these methods and the results expected with these methods.*

## Methods

To investigate the question; how have wage-gaps developed over the period 2001-2011 in the USA and across states? Two decomposing methods were selected, namely Oaxaca and Juhn Murphy Pierce (JMP). The Oaxaca method was selected, because it allows for a detailed decomposition of the wage gap. It also was the main driver for research into wage gaps in the early years of this field and continuous to be a leading decomposition method.

The Juhn Murphy Pierce method appears in many regards very similar to the Oaxaca method and it does generate the very similar results when looking at the mean. However, due to a small change in the decomposition method it allows for investigation of wage gaps at different percentiles, whilst the Oaxaca method does not. This is the main reason this method was selected. This flexibility does come at the cost of losing the detailed decomposition. These two methods hence appear to be perfect compliments, whilst having the same foundation. On top of this the JMP-method also lends itself for a decomposition over time.

### Oaxaca

The Oaxaca method was first published in the article; male-female wage differentials in urban labour markets, by Ronald Oaxaca in 1973. And was one of the first formal decomposition methods, before that research was mainly done on the ratio of the wages between the two groups. This decomposition method can be done in two ways, a two-fold and a three-fold decomposition.

In words the method, for the three-fold decomposition, does the following, first it estimates linear regressions for both of the groups of interest. It then compares the average wage for both of these regressions and calls that the difference. It then decomposes this difference further into three categories, difference in endowments, difference in coefficients, i.e. differences in returns, and interaction of these two differences. For the endowments effect the method sets the second group to have the same characteristics as the first group. Then by taking the difference between the second group’s actual average wage and the second group’s average wage based on the first groups characteristics it identifies the endowments effect. In a similar fashion it determines the coefficient effect. It now sets the second group to have the same coefficients as the first group. It again takes the difference between the second group’s actual average wage and the average wage based on the same coefficients as the first group and this is the coefficients effect. Then finally these two effects together also have an impact and is identified as the interaction effect.

To get more insight into this method the following graphical illustration is used.



This image represent the two-fold decomposition. The for the regression lines are estimated using the individual observations of each group. The total wage gap is then represented by the difference between and both of which are based on the average characteristics. This is then decomposed in a characteristics effect and the unexplained effect. The characteristics effect is calculated by taking the difference between the average characteristics of the two groups evaluated at the from the first group. The unexplained effect is the difference between the of the two groups evaluated at the average characteristics of the second group.

After this the three-fold decomposition is easier to understand, but it requires a different graphical representation.



The same linear regressions as before are used for the determination of the . The entire surface of the square represents the average wage earned by first group. And using this it becomes easier to determine “why” group 2 does not earn the same as group 1. The endowments effect represents the difference between the individual characteristics, . The coefficient effect represents the different pay-offs that the two groups get for their individual characteristics, . The interaction effect shows the contribution of both different individual characteristics and different prices,

### Juhn Murphy Pierce

The Juhn Murphy Pierce method was first used in the chapter; Accounting for the slowdown in black-white wage convergence, in Workers and Their Wages, by Kosters in 1991. The method was largely inspired by the Oaxaca method however it does use a different approach to allow for a percentile overview of the decomposition.

The basic JMP method in the mean results in the similar outcomes as the Oaxaca models and is as well a threefold decomposition. The names used by JMP however are Quantities, Prices and Unobservables.

Mathematically the basic JMP method goes as followed (Jann, 2006). The method starts by estimating wage equations for both groups.

With subscripts i and j representing the individual and the group respectively. And and being vectors of individual characteristics and their estimated coefficients.

Then let represent the cumulative distribution of the residuals for each group, so that

With representing the individual’s percentile in the residual distribution of their respective groups. Then can be defined as followed

With representing the inverse of the cumulative distribution function. Next assume that is residual distribution of the reference group and represents the coefficients from the reference group. Then 2 counterfactual wage equations can be estimated. First one in which quantities differ but the prices and residual distribution are fixed.

In the second counterfactual wage both quantities and prices, i.e. coefficients, differ but the residuals are fixed.

The final equation is the same as the original equations because quantities, prices and residual distributions are allowed to vary.

These three equations are then used to create a three-fold decomposition. In this decomposition capital letters stand for the summary statistic of the distribution of the variable denoted by the corresponding lower-case letter. For example at the 25th or 75th percentile, then the total wage gap can be decomposed in the following manner, this decomposition depends on what is chosen as a reference group, in this case the first group is used as a reference group.

Where the first part of the equation represents the quantity effect, i.e. endowment effect. The second part is the prices effect, i.e. coefficient effect. The third part is due to differences in unobservable quantities and prices.

### Juhn Murphy Pierce extended for time

When extending the JMP method to include further decomposition in our case time. The start off point is the two-fold decomposition of JMP (Jann, 2008). The basic two-fold decomposition of the JMP starts with the following linear regression

With being a vector of outcomes, a vector of individual characteristics, a vector of estimated coefficients and is the error term. This model can be rewritten in the following manor

With being a vector of standard deviation of the residuals and a vector of the standardized residuals. Then the following two-fold decomposition can be attained where the capital letters represent the summary statistic of the distribution of the variable denoted by the corresponding lower-case letter, for example at the 25th or 75th percentile.

With being the difference between the two groups so . In this two fold decomposition the wage gap is divided between differences in observed quantities and differences in residual gap (unobserved quantities, unobserved prices and “discrimination”) .

Then the decomposition between two different time periods can be obtained by taking the difference of the difference of each year :

With subscripts 11 standing for the values of the first group for the first period and 12 for the first group in the second period. This can be re-written as

The first part represents differences in the predicted gap between the two time periods and the second part represent differences in the residual gap of the two periods. This can then further be decomposed to become a four-fold decomposition. This four-fold decomposition depends on what is chosen as a benchmark, in this case the first group.

The first term

Represents differences in differences of observed characteristics between to the two time periods.

The second term

Consists of the differences in differences of the observed coefficients between the two time periods

The third term

Represents differences in differences of the relative wage positions of the two groups of interest in between the two time periods. This may represent differences in unobserved characteristics

The fourth term

Reflects differences in difference in residual inequality in between the two time periods. It is also named the unobserved difference in skills.

## Model Specification

In the regressions the basic JMP method is the method on which the focus is. This method was selected because in the mean it allows for virtually the same estimation as the Oaxaca method and it also allows for percentile estimation. This does come at the loss of the detailed decomposition option, but since the object of interest is the development of wage-gaps, not the sources of wage gaps I believe this is a justified sacrifice. The time enhanced version of the JMP was not selected as the main focus because for the time overview the basic JMP equation is estimated for each of the years individually. This allows for a tracking of the wage gap over time, instead of just comparing wage gaps between two years.

The regressions performed will follow the mincer model. Starting with using the natural logarithm of an income indicator, in this case approximate hourly wage. This indicator was chosen because it gives in my opinion the most conservative outcomes, with regards of any of the income indicators available. As is common with the mincer model indicators of work experience and education are included as explanatory variables.

For the work experience indicator usually age minus years of education minus some constant is used. However in this data set education was not listed in years of education, but in highest attained level. Therefore age and age squared are used to model work experience.

As stated above education only exist in the highest attained level form and hence dummies are used to have all the levels of education included in our regression.

The mincer equation is expanded with the following variables; race, sex and number of children.

Race and sex are added in the equations where these are not the groups of interest. Thus when looking at the wage gap between White and Black a dummy to control for sex is added and vice versa.

Finally number of children and its square are used in the regression. This was to catch some of the negative work experience that children represent. Even though society has become more and more equalitarian, it still holds that women tend to take more time off when having children. Thus with this variable I hope to take this effect out of the work experience variable and thus get a better picture of how work experience differs across the groups. The square of number of children was added because I believe that the negative work experience effect reduces with the increase of children.

Adding all these variables together the following regression is used to investigate the wage gaps:

Where Educ’ is a vector of dummies for each of the education levels and being a vector of coefficients corresponding to those education levels. Race’ is a vector of the race categories registered and its corresponding coefficient vector. As stated before Female and Race’ are only added in regressions where possible.

This high number of dummies included in the regressions poses a problem for the detailed Oaxaca decomposition. This is because the reference group in the detailed decomposition is “arbitrary” and this choice of reference group influences the coefficients of the included groups (Oaxaca & Ransom, 1999). To get around this problem the solution of Yun (2005) is used. The idea behind the solution is that by estimating the regression with each of the dummies being the reference group once and then taking the average of the estimated coefficients the true contribution to the wage gaps can be retrieved.

## Areas of investigation

This basic regression is used to look at a variety of different topics. Firstly an overview of the two wage gaps of main interest; male/female and white/black. This will just look at the wage gaps over the entire period and no further splitting of the groups. For this overview all three methods will be used.

After that a closer look at each of these wage gaps is taken by further splitting them up into four groups, but still looking over the entire time-period. The four comparisons are as followed; white male/female, black male/female, white male/black male, white female/ black male. These comparisons and any further comparisons will be estimated with the basic JMP method looking at the mean, 25th percentile and the 75th percentile.

Then one of the main interests of this paper is examined by looking at the two comparison groups over time for the entire United States. This will show how the wage gaps of these groups have developed over the past 11 years. To achieve this regressions are estimated for each of the individual years and then the JMP method is applied.

Finally, the individual states become the object of interest, firstly looking at the states across the entire time period and after that applying the JMP method extended with the time factor for the period 2001-2011. [[3]](#footnote-3)

## Expectations

Given the areas of the wage gaps investigated here and the literature review, certain expectations about the findings come about. Firstly and most obviously is that there is a male/female wage gap and also a white/black wage/gap. This is something that is also expected to be present in these findings. Furthermore the literature suggest that there are certain trends going on with these pay gaps. For the male/female gap the trend over the past 40 years has been to decrease the gap albeit at a slower and rate over time. So here we also expect there to be a reduction in the gender wage gap, although given the trend this my only be a small decrease. For the racial wage gap the trend in the past 20 years has been that of a small increase.

# Data

*In this section the data source, the data itself and the transformations applied to create a workable dataset are described in detail.*

## Data Source

To investigate our research question I use data from the American Community Survey (ACS). This survey is conducted every year and funded by the government of the United States of America. The main purpose of the survey is to provide data to communities, state governments and federal programmes so that they can determine their investments and their services. Overall this dataset is used every year to determine how to distribute over 400 billion dollars over federal and state funds.

To be able to determine this division of money the ACS asks questions in the following areas; demographics, income and work, health insurance, education, veteran status and disabilities.

The dataset is relatively new and is a derivation from the US census long form, testing of the survey started in 1996. The first data produced by the survey was that for 2000. As of 2005 the data set was fully implemented for populations over 65,000. In 2010 it released its first dataset covering all population sizes and all areas, by combining data from the 5 previous years.

The ACS collects its data in the following way. It first sends out invitations to selected households throughout the U.S. Those selected can fill out the questionnaire either on paper or online. Those that did not fill out the questionnaire are then contacted, either by phone or in person, by a member of the Census staff. Then all the data is collected by the U.S. Census Bureau. Which then combines all the data, after that it offers the data to the public and creates its own statistics from this data. By using this method a response rate of 66 % was reached in 2009.

The survey consists of three different versions. All of these versions rely on the same data, but differ mainly in the population areas they represent. The three versions are 1%, 3% and 5% samples. The 1% sample represents all areas with a population greater than 65,000 and is released every year. The 3% sample covers areas with a population over 20,000, it combines the data from three years and thus is only released once every three years. The 5% sample represents areas of all population sizes, this is achieved by combining the data of 5 years. The names of the versions indicate how much of the population of the US they represent.

Not all of the data in the survey is used in the estimations, first of all I opted for using the 1% samples that are released every year. This sample was chosen because this has yearly data and since the smallest population of interest is on a state level, I do not believe that the population areas covered in this sample (65,000+) will be a problem. Furthermore I limited my research to only people that are working and also receive income. After this the dataset was limited to only include the years 2001-2011, Although the years 2001-2004 only have half the observations of the other years (about 500,000 vs 1,200,000), I still believe they have added value and especially in the yearly estimations will not be influenced by the higher number of observations in the other years. This leaves a total observations of 10,814,916 in the basic dataset. The basic data-set covers all working and income earning observations in the 2001-2011 period and was aggregated and unified by IPUMS.

Furthermore only the following variables are used out of the pool of variables available within the dataset. Income from wage, Age, Sex, Race, Education, Number of children, Year, State, Class of worker, Weeks worked and Hours worked.

The basic dataset is only used for the comparison between self-employed persons and those working for wages. For all the other regressions an even stricter dataset is used. The stricter dataset, being my main dataset, only contains persons working for wages. This was done because I expecte the wage structure for the self-employed to be different from that for those working for wages. And I did find evidence of that fact in our regressions. Any further tables in this paper will be based off the data in the main dataset, only the tables and figures pertaining to the self-employed will make us of the basic dataset.

## Transformation of variables

In order to make the dataset workable first dummies for Race and Education were created. The following race categories are available.

|  |  |  |  |
| --- | --- | --- | --- |
| **Race [general version]** | **Freq.** | **Percent** | **Cum.** |
| White | 8,257,330 | 80.26 | 80.26 |
| Black/Negro | 915,357 | 8.9 | 89.16 |
| American Indian or Alaska Native | 77,509 | 0.75 | 89.91 |
| Chinese | 109,908 | 1.07 | 90.98 |
| Japanese | 31,197 | 0.3 | 91.28 |
| Other Asian or Pacific Islander | 328,441 | 3.19 | 94.48 |
| Other race, nec | 403,834 | 3.93 | 98.4 |
| Two major races | 151,938 | 1.48 | 99.88 |
| Three or more major races | 12,491 | 0.12 | 100 |
| **Total** | **10,288,005** | **100** |  |

Table 1, Frequency-table Race

As is clear in the table the vast majority (80%) of the respondents identifies with the white race, whilst 9% percent identifies with the black race and the remaining 11% is a mix of different races and multiples of races.

For education dummies are used, because the dataset does not record years of schooling, it rather records highest level of schooling attended. The following levels of schooling are registered in the survey.

|  |  |  |  |
| --- | --- | --- | --- |
| **Educational attainment** | **Freq.** | **Percent** | **Cum.** |
| N/A or no schooling | 51,678 | 0.5 | 0.5 |
| Nursery school to grade 4 | 38,896 | 0.38 | 0.88 |
| Grade 5, 6, 7, or 8 | 197,230 | 1.92 | 2.8 |
| Grade 9 | 122,858 | 1.19 | 3.99 |
| Grade 10 | 198,662 | 1.93 | 5.92 |
| Grade 11 | 264,008 | 2.57 | 8.49 |
| Grade 12 | 3,633,972 | 35.32 | 43.81 |
| 1 year of college | 1,654,510 | 16.08 | 59.89 |
| 2 years of college | 901,506 | 8.76 | 68.66 |
| 4 years of college | 2,051,077 | 19.94 | 88.59 |
| 5+ years of college | 1,173,608 | 11.41 | 100 |
| **Total** | **10,288,005** | **100** |  |

Table 2, Frequency-table Education

The table shows that about 8% of the population never graduates from high school, 35% finishes high school but follows no further education, and the final 57% attains at least some college-level education.

After that a fairer representation of income was constructed, to achieve this income was adjusted for the usual hours worked in a week and an approximation of the weeks worked. Approximation of the weeks worked is used, because in the dataset only records intervals of weeks worked are available. Therefore first the weeks worked were estimated by the middle value of each of the intervals. After which the following variable was created.

The usage of the approximate weeks works, compared to using income adjusted only for hours, reduces on average the raw wage gaps by 10-15%.

## Descriptive statistics

The next page shows descriptive statistics for the main dataset, together with two graphs regarding the hourly wage and educational attainment of the four groups that are the main interest of this research; white, black, male, female.

The summary statistics shows that the average person in this research is a 42 year old male, having attained one year of college level education. He earns 41,800 $ by working 40 weeks a year and working 39 hours in each of those weeks and supports 1 child with these earnings.

Furthermore it shows that the youngest person working for wages is 16 years old, this is by design of the survey, and the oldest person is 95 years of age. The maximum amount of children that any one of the respondents has is 9, but at least 50% has no children at all.

It is also visible that by moving from Income to the natural logarithm of approximate hourly wage both the skewness and the peakedness of the income variable is greatly improved. Overall Income, Age, Sex and Number of Children are at least to some degree skewed to the right and ln(Income), ln(Hourly wage), Education, Hours worked and Weeks worked are skewed to the left.

Graph 1 depicting hourly wage, shows that Whites earn the most, closely followed by the Male group. The Females and Black are the worst paid groups by far, although women appear to be closing the gap. This strengthens the idea that white males are the non-‘discriminated’ group.

The educational attainment graph paints a rather different picture Women are now the top performers together with white. Males are now in the middle and closing the ranks is black. It should be noted however that the graph starts at an educational attainment of 6.4 years and thus the differences between the groups appear larger than they are.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | ***LN income*** | ***LN hourly wage*** | ***Income*** | ***Age*** | ***Sex*** | ***Education*** | ***Hours Worked*** | ***Weeks worked*** | ***Number of Children*** |
| **Mean** | 6.573302 | 2.779478 | 41773.56 | 41.67786 | 1.49086 | 7.47808 | 39.30495 | 46.92879 | 0.7896621 |
| **Median** | 6.656441 | 2.787877 | 31200 | 42 | 1 | 7 | 40 | 51 | 0 |
| **Minimum** |  | -7.140651 | 4 | 16 | 1 | 0 | 1 | 7 | 0 |
| **Maximum** | 13.04115 | 10.88877 | 666000 | 95 | 2 | 11 | 99 | 51 | 9 |
| **Standard Deviation** | 0.903013 | 0.7543788 | 45478.11 | 13.72268 | 0.4999165 | 2.271741 | 11.83801 | 10.24519 | 1.092362 |
| **Skewness** | -0.958633 | -0.063539 | 4.490224 | 0.10853 | 0.036565 | -0.1296378 | -0.0630787 | -2.792608 | 1.419578 |
| **Kurtosis** | 6.108495 | 4.756228 | 35.4898 | 2.30206 | 1.001337 | 2.836368 | 5.543367 | 9.937058 | 5.098261 |
| **Observations** | 10288005 | 10288005 | 10288005 | 10288005 | 10288005 | 10288005 | 10288005 | 10288005 | 10288005 |

Table 3, Descriptive Statistics

Graph 3 Graph 4

# Results

*In this section the results of the regressions is described and interpreted. Starting off with the regressions of the two comparisons and then working down the list that was stated in the methodology section, finally ending in an overview of the changes across the states in the period 2001-2011.*

## Male/Female overview

### Male and Female regression

The results discussed here can be found in in the appendix. First the linear regressions of males and females for the entire period are of interest. Examining these equations it becomes visible that for both groups additional education almost always leads to a higher wage. For men attaining nursery school till grade 4 decreases their wage and attaining grade 5-8 does not add additional value over having no schooling. For females the negative association with education goes from attaining nursery school till grade 9. After this also for women additional educational attainment has benefits compared to having to no education.

The coefficients for the race dummies show that most races earn less than whites on average. In the male category only those identifying as Japanese earn more than whites. Whilst for females all that identify as Asian earn more on average and those with three or more race identifications earn the same on average.

The age variable has a positive effect in the beginning, but this positive effect declines over time. Using partial derivatives we find that for males after turning 53 the positive effect starts to decline and once they turn a 105, age actually has a negative impact on wages. For females these values are 52 years and 103 years. But given that the oldest person in the sample is 95 years old, age always has a positive effect and since this value is also so far outside the sample it should not be seen as too indicative.

Looking at number of children there are much larger differences between the two groups. Males “benefit” the most from 3 children after which the benefit starts to decline and after 6 children there is no associated benefit left. Whilst for women the benefit starts to decline after the first child and after the second child all associated benefit to the hourly wage is disappeared.

Examining the magnitudes of the coefficients it becomes apparent that age, i.e. experience, and education are the major determinants of the hourly wage, as is expected with a mincer type wage equation.

### Male/Female Oaxaca decomposition

Moving to the decomposition of the wage gap between these groups. It is visible that the average male earns almost 18 dollars an hour and the average woman earns 14.4 dollars an hour. This leaves a raw wage gap of 22%. Using the Oaxaca method this raw wage gap is decomposed in the following manner; -3% endowments, 25% coefficients, i.e. returns and 0.3% interaction. Clear from these numbers is that the main component of the raw wage gap is differences in returns.

The more detailed decomposition gives insight into which variables are the main contributors to each of these effects. For endowments experience appears to be the main driver, although it should also be noted that all levels of education are in favour of the women, albeit with small magnitudes. In terms of the coefficient effect again experience looks to be a major contributor now together with the constant. The contribution of the constant however is positive one i.e. it reduces the wage-gap by roughly 20 percent points. In the interaction effect again experience has a big impact.

### Male/Female JMP decomposition

The Juhn Murphy Pierce method reports that at the 25th percentile the wage gap is 19% with the quantity effect, i.e. differences in endowments, contributing -4%, the prices effect, i.e. differences in returns, contributing 23% and the unobserverables accounting for -0.7%. For the 75th percentile a raw difference of 25% is found, accounted for by -0.7% of quantity effect, 26% of a prices effect and 0.3% of unobservables. This shows that increasing the percentile increases the raw wage gap between males and females. Further on this shows that the main determinant of the wage gap are differences in prices.

The JMP method extended with a time factor shows that on average the male female wage-gap has declined by 5.7 percent points, this is accounted for by a 1.7 point improvement in observable quantities, a 0.5 point decrease in differences of observable prices, a 5 point decrease unobservable quantities and 1.6 point increase in unobservable prices.

So far we found that the male and female wage equations are mainly driven by age and education. On average a woman earns 22% less than the male counter parts and most of this wage difference is due to differences in returns between the two groups. Furthermore the wage gap declined by 5.7 percent points in the period 2001-2011. This wage gap decline is in line with what previous research has found.

## White/Black overview

### White and Black regression

The results discussed here can be found in the appendix. Starting with the regressions, education for whites attaining nursery till grade 8 has a negative effect on the hourly wage. For the blacks attaining nursery till grade 11 has a negative effect on hourly wage.

Comparing the female dummy something interesting shows up for whites being female leads on average to a wage that 26% lower, whilst for black females the difference on average is only 15%. This finding seems to imply that being female does not have a constant effect, i.e. it differs if you are white or black.

Again comparing the different age factors using partial derivatives we find that for whites the highest benefit is possible at 53 and has no benefit after the age of 104 years old. And for blacks these values are exactly the same however the highest possible benefit is lower than that of whites 1.83 versus 1.6 log hourly wages.

For number of children a similar picture arises whites with 3 children earn on average the most whilst after the 5th child children no longer have a positive impact on wage compared to no children at all. Blacks with 2 children earn the most on average and after the 4th child no wage benefit is associated with children anymore. Overall however the number of children does not appear to be a great contributor towards the wage.

Looking at the magnitudes, again age and education are main determinants of wage, followed by sex.

### White/Black Oaxaca decomposition

Turning towards the Oaxaca method, the average white earns $16.55 an hour, whilst the average black earns $13.84. This results in a raw wage gap of 18%. This gap is created for 7% by endowments, 10% prices and 1% interaction.

An inspection of the detailed composition shows that a major part of the endowment effect is explained by differences in education, especially in the highest level of education blacks appear to be under represented. In terms of coefficient effect age again is a major contributor with the constant acting as a mitigating factor. Also higher levels of education are rewarded less for blacks. The interaction effect does not show a main driver.

### White/Black JMP decomposition

Delving further in these results the JMP method shows that at the 25th percentile the wage gap is 15%, with quantity representing 6% prices 9% and unobservables 1%. For the 75th percentile the wage gap increases to 17.6% which is lower than the average, but only by a small margin. The quantity effect contributes 8% the prices 11% and the unobservables reduce it by 1%. With these numbers it becomes clear that for the white/black wage gap both quantity and prices are main determinants.

Using the JMP extend form it becomes apparent that over the period 2001-2011 the wage gap has increased by 4.1%. The main driver here is the unobservable quantities with 2.9%. Unobservable prices add 0.6%. Observable quantities have reduced the wage gap with 0.1 percent and observable prices increased it by 0.7%.

Overall the white/black wage gap appears to be quite different from the male/female gap. The gap is first of all lower with its 18% on average and also has two drivers, quantity and prices, as opposed to one. Also the white/black gap has increased over the past 11 years, which was expected given the trend the 20 years before.

## White/Black Male/Female detailed

For this section the comparisons are split up to get a total of 4 comparisons. Allowing for two paths to move from the standard group; white males, to what is thought of as the most “discriminated” group, black females. This splitting up assists in investigating if being female and if being black have a constant effect. Meaning that moving from male to female has the same effect for whites as blacks and moving from white to black has the same effect for both male and female. The results described here can be found on the next page in table 4.

Starting off with the white male/black male comparison the raw wage gap ranges from 20% to 25%. Decomposing that further quantity has a range from 7 to 9%, price from 13 to 18% and finally unobservables range from -1% to 0%. Overall the raw difference is higher than that of just being black, but there is a similar split of the wage gap between prices and quantities, although here more of an emphasis is on prices.

Continuing with the white female/black female comparison quite a different picture is being depicted. The raw wage difference now only ranges from 9 to 10%. The quantity effect here ranges from 6 to 7%, price effect only accounts for 4% and unobservables account for -1 to 1%. Overall this comparison shows a smaller and less dispersed wage-gap. At all of the percentiles the contribution of the decomposition factors is very similar. Here also quantity appears to play a bigger role than prices do. This already shows that effect of moving from white to black is not constant across males and females.

The white male/white female comparison shows a very similar picture as the white/black male raw wage gap, ranging from 21 to 25%. However the decomposition of this wage gaps is quite different. White females appear to be better qualified than their male counterparts, with a quantity effect of -4 to -2%. The price effect though completely outweighs this advantage with a range from 25 to 28%. The unobservables again add little to the raw wage gap with values from -1 to 0%. So for the white male/female decomposition virtually all of the effect comes from the prices effect.

Finally the black male/black female comparison shares attributes from the white/black female comparison and the white male/female comparison with a raw wage gap from 9 to 12% and a quantity effect ranging from -6 to -4%. The price effect is 15-16% and the unobservables range from -1 to 1%. So again the price effect is the main determinant that is being constrained by the negative effect from the qualifications of black woman compared to their male counterparts.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***White/Black Male*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **p25** | 0.195954 | 0.066801 | 0.129133 | 0.000020 |
| **mean** | 0.245322 | 0.085796 | 0.159484 | 0.000042 |
| **p75** | 0.244197 | 0.072242 | 0.181160 | -0.009205 |
|  |  |  |  |  |
|  |  |  |  |  |
| ***White/Black Female*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **p25** | 0.099530 | 0.056842 | 0.035999 | 0.006689 |
| **mean** | 0.093334 | 0.056253 | 0.037053 | 0.000027 |
| **p75** | 0.096730 | 0.068399 | 0.037593 | -0.009263 |
|  |  |  |  |  |
|  |  |  |  |  |
| ***White Male/Female*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **p25** | 0.213812 | -0.034696 | 0.256129 | -0.007621 |
| **mean** | 0.245105 | -0.021625 | 0.266716 | 0.000014 |
| **p75** | 0.244197 | -0.038875 | 0.281840 | 0.001232 |
|  |  |  |  |  |
|  |  |  |  |  |
| ***Black Male/Female*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **p25** | 0.117388 | -0.038825 | 0.158926 | -0.002712 |
| **mean** | 0.093117 | -0.056811 | 0.149902 | 0.000026 |
| **p75** | 0.096730 | -0.058452 | 0.149655 | 0.005526 |

Table 4, White/Black Male/Female JMP decomposition

Combining all this information we see that moving from white male to female and to black male have the same effect on the raw wage gap and thus naturally also going from black male to black female and from white female to black female have the same effect. This shows that the effect of moving from male to female is not stable. This also means that the white to black relationship is not a stable one.

There are however a few relationships that do appear to be stable across these groups. It for example is clear that females always have better qualifications than their male counterparts do. Looking at the male/female wage gap we also see for both white and black that the main determinant of that gap are differences in coefficients. In the white/black comparisons it becomes clear that blacks are always underqualified compared to their white counterparts. On top of that the returns to those qualifications, i.e. the coefficients, are always lower too, exacerbating the wage-gap between these two groups.

## Yearly Male/Female and White/Black

Graph 5 to 10 give an overview of how the mean wage gaps in the two comparisons have developed over the period 2001-2011 and in the appendix the underlying data can be found. It also shows the development in two additional percentiles, the 25th and 75th. Furthermore, it shows the decomposition of these wage gaps and their development.

Focusing on the male/female gap it is visible that over the different percentiles the tendency is for the wage gap to decrease. Looking at the actual numbers the 25th percentile gap dropped from 24% to 15%, the mean from 26% to 20 % and the 75th percentile from 25% to 21%. The graphs also indicate that the main causes of these drops are differences in quantities and in prices.

Further investigation shows that overall the quantity effect improved by 2 percent points, meaning that the “advantage” that women had in terms of endowments became even larger. The prices effect dropped by 6 percent points for the 25th percentile, 4 percent points at the mean and 3 percent points at the 75th percentile.

The unobservables changed very little over this time period, this is visible in the graphs but is more apparent in the raw data. For both the 25th percentile and the mean the values of the unobservables did not change significantly. The 75th percentile saw a small increase, i.e. a widening effect on the wage gap, but because of the small magnitudes the overall impact of this is severely limited.

Moving to the white/black comparison the picture changes significantly. Firstly, none of the decomposition lines is above the raw wage gap anymore. Meaning that no factor has a major limiting effect. It is also visible that the main drivers of the wage gap in general here are quantity and prices as opposed to the singe driver in the male/female comparison. Another major change is the fact that there does not appear to be a down-wards trend, it actually appears to be an upward trend.

Looking at the raw wage gaps over all the percentiles we find that at the 25th percentile the gap increases from 15% to 17%, at the mean from 16% to 20% and the 75th percentile increase from 16% to 20%. From the graphs the main driver appears to be changes in the price differential. Quantity in this case appears to stay stable, although it does show some shocks.

A more detailed inspection of the data confirms this observation. Price differentials increase with 2 percent points at the 25th percentile and increase with 4 percent points at both the mean and the 75th percentile. The quantity effect did not change over the entire period at the 25th percentile, but did fluctuate by as much as 2 percent points and the general tendency was to lower the quantity effect. At the mean the quantity effect increased 1 percent point and remained stable within a 1 percent bandwidth, the same holds for the 75th percentile. Also here the unobservables did not vary much over the time period and given its small magnitudes it did not add much to the raw differential to begin with.

Overall we find that the wage gap for the male/female comparison follows the trend it has been following for the past 40 years and is still declining. Much of this decline is due to further improvements in the endowments and despite the price effect declining, there still appears to be a lot of room for improvement. The white/black comparison follows the trend of the past 20 years and appears to be increasing. This increase appears to mainly be due to increases in the price difference. Taken over the whole both the quantity and price difference still allow considerable room for improvement.

Graph 5 Graph 6 Graph 7

Graph 8 Graph 9 Graph 10

## Wage gaps across states

Comparing the wage differences across states, shows many differences and also shows differences for the two comparisons within the same state. To get better insight into these differences figure 1 to 6 show maps for both comparisons of the raw differences, price effect and quantity effect at the mean[[4]](#footnote-4). Unobservables are not shown because the magnitudes are so small that they virtually have no effect. In the appendix the data-points for the mean can be found and also a graph of the raw difference for each state compared to the average over the entire sample.

Beginning with the male/female comparison, no clear divide is visible between the states. However it is observed that Wyoming and Utah have the highest values of all states, with a raw difference of 34 and 32 percent respectively. Louisiana being third with a wage gap of 30%. Out of the three of these Utah has the most interesting find, with it being the only state together with the District of Columbia, which has a positive coefficient for the quantity effect. Meaning that women there have lower characteristics than males on average.

The three best performing states are District of Columbia, California and Vermont, with 14, 17 and 17 percent of a raw wage gap each. District of Columbia is included despite it not being a state, because it is available in the dataset and it is one of the more important areas in the United States. It is however excluded from the maps, because its values influenced the scale too much for a place that is not visible on them.

Looking at the states the overall the raw wage gap has a dispersion of 20 percent points, the quantity effect ranges from -0.04 to 0.04 and the prices effect from 11% to 35%. To get a better insight of how the states perform over the entire wage distribution a ranking was created. The ranking was constructed by ranking each of the states at the 25th, 50th and 75th percentile on raw difference, quantity effect and price effect. Combining these results creates a ranking for each of these parts for each state that can range from 3 being the lowest coefficients to 153 being the highest coefficients. The full table of these ranks can be found in the appendix.

Inspecting this rank table we only focus on one variable namely the raw wage gap, this is because there is only one main determinant for this wage gap. From the table we find that the District of Columbia scores the best with a rank of 5, followed by Vermont and California with 12 and 14 respectively. The worst performing are again Wyoming, Utah and Louisiana with each ranking 153, 150 and 145. The only change here with the findings of the mean is that Vermont and California have switched places.

Switching the focus to the white/black comparison, the maps do show a visible difference in states geographically. First of all the south eastern states appear to have high values over all categories. Furthermore the more east one moves the bigger the quantity effect becomes.

The actual data shows surprisingly that District of Columbia now has the highest wage gap with 47%. Followed by Mississippi and Louisiana with 32% each. The best performing states are Vermont, Kansas and Arizona, with 4, 10 and 11% respectively. Vermont deserves a special note, because it is the only state with an observation of 0% wage gap, this is in the 75th percentile. With all these results and the following results it should also be noted that certain states have a low percentage of residents identifying as black. Therefore certain states have few observations for the black category and thus the observed differences are not necessarily representative. But I have no indication that these results are biased in any particular direction and therefore I do report them.

Overall the raw wage difference ranges from 4% to 47%, the quantity effect from -3% to 20% and the price effect from 3% to 27%. For the quantity effect only three states have a negative coefficient, meaning that blacks on average have better characteristics than whites. The unobservables here are bigger in magnitude and do have an impact on the results with a range from 0 to 6%. However only 14 states have an unobservables effect of 1% or higher.

The rankings table gives a similar overview of the states. District of Columbia, Mississippi and Louisiana are the states with the highest ranking, 153, 148 and 146. District of Columbia actually scores the highest possible rank on each of the decomposition variables. In the states with the lowest ranks Vermont, Pennsylvania and Kansas have ranks 7, 12 and 12.5. Arizona that was third when looking at the mean but drops to 10th place.

We find overall no general pattern across the states for the male/female comparison. For the white/black comparison there is an observable relationship between the geographical location of the state and its wage gap, although this is mainly for the south eastern states. This is in line with Smith and Welch (1989) who find that part of the reduction in racial wage gap from the 1940’s to 1980’s was due to migration of blacks to the northern states, suggesting that southern states pay less to blacks than the northern states. It however contradicts with Vigdor (2006), who finds that the racial wage gap has converged between the northern and southern states over the period 1960 to 2000. He does however use adjusted wage gaps as opposed to

Figure 1, Mean Male/Female raw wage gap Figure 2,Mean Male/Female quantity effect Figure 3, Mean Male/Female prices effect

  

Scale: 0.17 – 0.34 Scale: -0.04 – 0.04 Scale: 0.20 – 0.36

Figure 4, Mean White/Black raw wage gap Figure 5, Mean White/Black quantity effect Figure 6, Mean White/Black prices effect

  

Scale: 0.04 – 0.32 Scale: -0.03 – 0.17 Scale: 0.03 – 0.21

the raw wage gaps here. Furthermore the District of Columbia appears to be a very unique location with both the lowest wage gap in the male/female comparison and the highest in white/black comparison. Overall it appears that in all the states there are still considerable wage gaps

## Differences in wage gaps over time and across states

Finally we investigate how the wage differences have changed over the period 2001-2011 across the different states. The results of this are found in the appendix together with the ranked version of these results. The ranked tables in this case are only based on the mean estimation and thus only assists in seeing where a state is positioned compared to the others. Below graphs 11 and 12 show the change in raw wage gaps with the average over the entire sample.

The male/female graph shows that for virtually all states the wage gap has declined, exceptions to this are Utah and West Virginia. Those have a wage gap increase of 1% and 0.5%. For both differences in unobservables has mainly contributed to this increase. Although West Virginia also has differences in observables contributing. It is only one of 7 states that has an increasing effect of the observables.

The graph further shows that Michigan, Iowa and Montana have experienced the greatest reduction in wage gaps, with decreases of 13, 11 and 10 percent points respectively. For Michigan and Iowa the main contributor to these decreases are the differences in unobservables with 10 and 9 percent respectively. For Montana both observables and unobservables contribute similar amounts 4 and 6 percent points.

The difference in wage gap ranges from -13 percent points to 1 percent point, difference in observables from -5 to 1 percent point and differences in unobservables from -10 to 3 percent points. Investigating these both further shows that for both of these the quantity factor is the main driver and for unobservables, prices effect actually has a positive coefficient for most states. This means that it unobservable prices mainly added to the wage gap, whilst the general tendency overall was a decrease in the wage gap.

Moving to the white/black comparison a big change is visible, because the differences are now mainly positive and with several states spiking high above the rest. However some if not all of these spikes are caused by a problem described in the section above. Namely that there is not as much data on blacks as there is on the other groups and now we only use data from 2 years exacerbating the problem. The low number of observations even made it impossible to do

Graph 11

Graph 12

estimations for Montana and therefore it is also excluded from the graph. This limitation should be kept in mind when going through the results.

From the data it becomes clear that Idaho, District of Columbia and Arkansas have experienced the greatest reduction in wage gap with 8, 6 and 6 percent points. Most of this change is due to changes in the observables. Although for Arkansas there is a split between observables and unobservables. In total there are 8 states with at least some reduction in wage difference.

Looking at which states have the highest increase we find West Virginia, Maine and Kentucky, this is the case if we exclude states with an increase of more than 15 percent points. I chose 15 as the cut-off point because the male/female comparison did not show any changes of more than 15 points. For the states with increases in the wage gap the drivers are unclear, because they differ a lot across the different states. A closer inspection of the observables also does not show a clear determinant. Although the quantity effect appears to be dominantly negative and the price effect positive. The unobservables also do not appear to have a clear determinant, but now both effects appear to be primarily positive.

In short we found that there was great variability in the changes in wage-gaps. For male/female those are predominantly reductions in wage gaps. The main driver for these changes are the changes in the unobservables. For the white/black comparison the changes in wage gaps are predominantly positive, with there not being any particular source for these changes.

# Conclusion

The main purpose of this paper was to investigate how wage gaps have developed across the period 2001 to 2011 in the United States and across its states. By applying three different techniques, Oaxaca, JMP and JMP enhanced for time, this question can now be answered.

First of all for the average gender wage gap it is found that the wage gap declined from 25.5 percent in 2001 to 19.9% in 2011, a reduction of 5.5 percent points. This reduction was mainly due to changes in the characteristics. For the average racial wage gap an increase of 4.1 percent points was observed, moving from 16.1% to 20.1%. This change was also mainly due to changes in differences of characteristics.

Overall it was found that the male/female gap across the different percentiles and over the different years was mainly determined by differences in returns, whilst differences in characteristics was a small mitigating factor for it. The gap over the different percentiles also showed similar trends of females improving their characteristics and also an improvement in the returns of those characteristics. The racial pay gap across the percentiles and the years is determined by both differences in characteristics and prices. In terms of trends the 25th percentile does not show the smooth trend that both the mean and the 75th percentile have, but does follow in the same direction and magnitude.

The splitting up of the groups into four decompositions demonstrated that the effect of either being black or a female is not constant. Black males and white females face similar wage gaps when compared to white males. Aspects that are constant are that females tend to have better characteristics than males. Whilst blacks always appear to generally have lower characteristics than their white counterparts and have lower returns to these characteristics. For females the main source of the wage gap across the percentiles is differences in returns.

Moving to the comparison of the wage gaps across the different states it became apparent that there is a great deal of variation across the states in a particular wage gap, but also within the same state between the two wage gaps. A good example of this variability being the District of Columbia, achieving the lowest gender wage gap and the highest racial wage gap. What is fairly stable across states is the rank of the state in terms of wage gap over the wage distribution, i.e. states achieve a similar rank when looking at the 25th percentile as in the 75th percentile. Also stable here is the source of the wage gap for each of the two decompositions. Furthermore geographical location shows a relationship with the racial wage gap, with the highest white/black wage gaps clustering in the south-east part of the US. Additionally, the difference in characteristics is higher the further east one looks.

By looking at the states across time it becomes clear that the general variability across states also continues in the trends. For the male/female pay gap all states except for two states, Utah and West Virginia, have a declining gap. The racial wage gap predominantly shows increases, with only 8 states showing at least some reduction in the gap. For the gender wage gap the change across states varied by 14 percent points and for the racial wage gap 39 percent points, but this is mainly due to the small sample of blacks for certain states.

In conclusion this paper found that the wage gaps, gender and racial, have continued the trends that were found in previous research. The gender pay gap has closed further as it has done in the past 40 years and the racial wage gap has widened as it has been doing since the 1980’s. These finding suggest that there is still much room for improvement in the racial wage gap, but also for the gender gap, because although it has been closing there still is on average a gap of 20%. This research should help in finding states that need extra attention in reducing the wage gaps and also shows which states are performing best in this reduction, which perhaps could be used as examples for the worst performing.

# Limitations and further research

Although the results in this paper are derived using the most relevant methods and careful consideration of the statistical rules, it does face several limitations which also indicate interesting and valuable areas for further research. One of the biggest limitation of this paper and many like it, is the availability of certain variables and/or indicators. This paper could have benefitted greatly from information on the actual work experience of the subjects and also an indication of skill, e.g. AFQT scores, would have added much value to the final results. Unfortunately this data was not available in this dataset and thus not included. Any further research that is able to add these variables will allow for even more accurate estimates of the wage gaps and their sources. The other major limitation is the limited data on the black group, this limitation does not matter for the general estimations of this paper. However once this data is split up to states and years, it does become a problem for the states with a small black population. This limitation can easily be solved in the future when the ACS has more years of observations and if the 3% samples are used, although that would not allow for a yearly comparison.

Other areas of further research that are not necessarily related to the limitations of this paper, are the self-employed and reversals. The topic of the self-employed is only briefly mentioned in this paper and is only explored in the appendix, it does however present a unique sub-sample of the working population that can provide some unique insights. With the area of reversals I mean an investigation of professions where the wage gaps may be reversed, i.e. professions where women earn more than males or blacks more than whites, one profession in which this could be the case is modelling. If these reversals exist, they could provide insight into how to reduce the wage gaps found in most jobs.

# Bibliography

Altonji, J. G., & Blank, R. M. (1999). ‘Race and gender in the labor market’. *Handbook of labor economics*, *3*, 3143-3259.

Altonji, J. G., & Shakotko, R. A. (1987). ‘Do wages rise with job seniority? ’. *The Review of Economic Studies*, *54*(3), 437-459.

Becker, E., & Lindsay, C. M. (1994). ‘Sex differences in tenure profiles: Effects of shared firm-specific investment’. *Journal of Labor Economics*, 98-118.

Blau, F. D., & Beller, A. H. (1987). ‘Trends in earnings differentials by gender 1971-1981’. *Indus. & Lab. Rel. Rev.*, *41*, 513.

Blau, F. D., & Kahn, L. M. (1997). ‘Swimming upstream: Trends in the gender wage differential in the 1980s’. *Journal of Labor Economics*, 1-42.

Blau, F. D., & Kahn, L. M. (2004). *The US gender pay gap in the 1990s: Slowing convergence* (No. w10853). National Bureau of Economic Research.

Bratsberg, B., & Terrell, D. (1998). ‘Experience, tenure, and wage growth of young Black and White men’. *Journal of Human Resources*, 658-682.

Chandra, A. (2000). ‘Labor-market dropouts and the racial wage gap: 1940-1990’. *The American Economic Review*, *90*(2), 333-338.

D'amico, R., & Maxwell, N. L. (1994). ‘The Impact of Post‐School Joblessness on Male Black‐White Wage Differentials’. *Industrial Relations: A Journal of Economy and Society*, *33*(2), 184-205.

Fearon, G., & Wald, S. (2011). ‘The Earnings Gap between Black and White Workers in Canada: Evidence from the 2006 Census’. *Relations industrielles/Industrial Relations*, *66*(3), 324-348.

Ferree, M. M., & McQuillan, J. (1998). ‘GENDER-BASED PAY GAPS Methodological and Policy Issues in University Salary Studies’. *Gender & Society*,*12*(1), 7-39.

Filer, R. K. (1993). ‘The usefulness of predicted values for prior work experience in analyzing labor market outcomes for women’. *Journal of Human Resources*, 519-537.

Gerhart, B., & Cheikh, N. E. (1991). ‘Earnings and percentage female: A longitudinal study’. *Industrial Relations: A Journal of Economy and Society*,*30*(1), 62-78.

Goldin, C., & Rouse, C. (1997). *Orchestrating impartiality: The impact of" blind" auditions on female musicians* (No. w5903). National Bureau of Economic Research.

Gunderson, M. (1989). ‘Male-female wage differentials and policy responses’. *Journal of Economic Literature*, *27*(1), 46-72.

Hill, E. T. (1995). ‘Labor market effects of women's post-school-age training’. *Industrial and Labor Relations Review*, 138-149.

Hundley, G. (2000). ‘Male/female earnings differences in self-employment: The effects of marriage, children, and the household division of labor’. *Indus. & Lab. Rel. Rev.*, *54*, 95.

Jann, B. (2006). ‘JMPIERCE: Stata module to perform Juhn-Murphy-Pierce decomposition’. *Statistical Software Components*.

Jann, B. (2008). ‘A Stata implementation of the Blinder-Oaxaca decomposition’. *Stata Journal*, *8*(4), 453-79.

Jann, B. (2008). ‘JMPIERCE2: Stata module to compute trend decomposition of outcome differentials’. *Statistical Software Components*.

Jaynes, G. D. (1990). ‘The labor market status of black Americans: 1939-1985’.*The Journal of Economic Perspectives*, *4*(4), 9-24.

Juhn, C., Murphy, K. M. and Pierce, B. (1991) Accounting for the slowdown in black-white wage convergence, in Workers and Their Wages (Ed.) M. H. Kosters, American Enterprise Institute Press, Washington, DC, pp. 107–43

Kahn, L. M., & Sherer, P. D. (1988). ‘Racial differences in professional basketball players' compensation’. *Journal of Labor Economics*, 40-61.

Kim, C. (2010). ‘Decomposing the change in the wage gap between White and Black men over time, 1980-2005: An extension of the blinder-Oaxaca decomposition method’. *Sociological Methods & Research*, *38*(4), 619-651.

Light, A., & Ureta, M. (1995). ‘Early-career work experience and gender wage differentials’. *Journal of labor economics*, *13*(1), 121-54.

Neal, D. A., & Johnson, W. R. (1996). ‘The Role of Premarket Factors in Black-White Wage Differences’. *The Journal of Political Economy*, *104*(5), 869-895.

Oaxaca, R. (1973). ‘Male-female wage differentials in urban labor markets’. *International economic review*, *14*(3), 693-709.

OECD (2012), OECD Family Database, OECD, Paris ([www.oecd.org/social/family/database](http://www.oecd.org/social/family/database))

Olsen, R. N., & Sexton, E. A. (1996). ‘Gender differences in the returns to and the acquisition of on‐the‐job training’. *Industrial Relations: A Journal of Economy and Society*, *35*(1), 59-77.

O'Neill, J. (1990). ‘The role of human capital in earnings differences between black and white men’. *The Journal of Economic Perspectives*, *4*(4), 25-45.

Ruggles, S., Alexander, J. T., Genadek, K., Goeken, R., Schroeder, M. B., & Sobek, M. (2010). Integrated Public Use Microdata Series: Version 5.0 (Machine-readable database). Minneapolis: University of Minnesota, 2006 American Community Survey.

Sicherman, N. (1996). *Gender differences in departure from a large firm* (No. w4279). National Bureau of Economic Research.

Simón, H., & Russell, H. (2005). ‘Firms and the gender pay gap: A cross-national comparison’. *Pay Inequalities and Economic Performance Working Paper*, *15*, 1-43.

Smith, J. P., & Welch, F. R. (1986). *Closing the Gap: Forty Years of Economic Progress for Blacks*. Rand Corporation, 1700 Main Street, PO Box 2138, Santa Monica, CA 90406-2138.

Smith, J. P., & Welch, F. R. (1989). ‘Black economic progress after Myrdal’. *Journal of Economic Literature*, *27*(2), 519-564.

Sorensen, E. (1990). ‘The Crowding Hypothesis and Comparable Worth’. *Journal of Human Resources*, *25*(1), 55-89.

Tomaskovic-Devey, D. (1993). ‘The gender and race composition of jobs and the male/female, white/black pay gaps’. *Social Forces*, *72*(1), 45-76.

Vigdor, J. L. (2006). *The new promised land: Black-white convergence in the American South, 1960-2000* (No. w12143). National Bureau of Economic Research.

Wellington, A. J. (1993). ‘Changes in the male/female wage gap’, 1976-85.*Journal of Human Resources*, 383-411.

Winship, C., & Korenman, S. (1997). ‘Does staying in school make you smarter? The effect of education on IQ in The Bell Curve’. *Intelligence, genes, and success: Scientists respond to The Bell Curve*, 215-34.

Yun, M. S. (2005). ‘A simple solution to the identification problem in detailed wage decompositions’. *Economic Inquiry*, *43*(4), 766-772.

# Appendix

## Methodology

### Self-employed/Working for wages

|  |  |  |  |
| --- | --- | --- | --- |
|  | ***Self-Employed Male*** | ***Self-Employed Female*** | ***Wage working Female*** |
| **age** | 0.043347 | \*\*\* | 0.038020 | \*\*\* | 0.061839 | \*\*\* |
| **age\_sq** | -0.000349 | \*\*\* | -0.000308 | \*\*\* | -0.000600 | \*\*\* |
| **nchild** | 0.121027 | \*\*\* | 0.063693 | \*\*\* | 0.017708 | \*\*\* |
| **nchild\_sq** | -0.015368 | \*\*\* | -0.010213 | \*\*\* | -0.007733 | \*\*\* |
| **educ\_\_2** | -0.104621 | \*\* | -0.004464 |   | -0.091703 | \*\*\* |
| **educ\_\_3** | -0.100564 | \*\*\* | -0.083997 |   | -0.060973 | \*\*\* |
| **educ\_\_4** | -0.062350 | \* | 0.063780 |   | -0.020336 | \*\*\* |
| **educ\_\_5** | 0.014570 |   | 0.056423 |   | 0.033784 | \*\*\* |
| **educ\_\_6** | 0.040306 |   | 0.101628 | \* | 0.067454 | \*\*\* |
| **educ\_\_7** | 0.092951 | \*\*\* | 0.227826 | \*\*\* | 0.240024 | \*\*\* |
| **educ\_\_8** | 0.209853 | \*\*\* | 0.318367 | \*\*\* | 0.378043 | \*\*\* |
| **educ\_\_9** | 0.147996 | \*\*\* | 0.329989 | \*\*\* | 0.533919 | \*\*\* |
| **educ\_\_10** | 0.464554 | \*\*\* | 0.563292 | \*\*\* | 0.752154 | \*\*\* |
| **educ\_\_11** | 0.895485 | \*\*\* | 0.882671 | \*\*\* | 0.994209 | \*\*\* |
| **race\_\_2** | -0.197767 | \*\*\* | -0.115976 | \*\*\* | -0.037140 | \*\*\* |
| **race\_\_3** | -0.206917 | \*\*\* | -0.128763 | \*\*\* | -0.072393 | \*\*\* |
| **race\_\_4** | -0.174888 | \*\*\* | 0.016490 |   | 0.084283 | \*\*\* |
| **race\_\_5** | 0.026417 |   | 0.102469 | \*\* | 0.063946 | \*\*\* |
| **race\_\_6** | -0.122793 | \*\*\* | 0.063524 | \*\*\* | 0.049902 | \*\*\* |
| **race\_\_7** | -0.092248 | \*\*\* | -0.044479 | \*\* | -0.035449 | \*\*\* |
| **race\_\_8** | -0.120203 | \*\*\* | -0.101923 | \*\*\* | -0.014540 | \*\*\* |
| **race\_\_9** | -0.292545 | \*\*\* | -0.040400 |   | 0.003105 |   |
| **\_cons** | 1.373019 | \*\*\* | 1.114648 | \*\*\* | 0.776374 | \*\*\* |
|  |   |   |   |   |   |   |
| **Adjusted R^2**  | 0.0977 |   | 0.0572 |  | 0.2723 |  |
| **Number of Obs** | 371361 |   | 155595 |  | 5049973 |  |

As explained before this comparison is meant to indicate if self-employed woman earn similar to their male counterpart and thus by being self-employed are able to avoid the “discrimination”. The reason for this expectation being that self-employed people set their own wage and thus we would not expect women to “discriminate” against themselves.

Before delving into the results a word of caution is necessary. All other regressions in this paper and also that of wage working females have an adjusted R-square of about 25 to 30 percent. The self-employed groups however only have an adjusted R-square of 6 - 10%. This indicates that the wage determinants for the self-employed appear to be quite different from those working for wages. On top of that there are compared to the other group relatively few observations for the self-employed. This finding also encourages our restriction for the main results of only looking at those working for wages.

Keeping that in mind with results displayed in table on the next page. The biggest wage gap thus far is found, the wage gap between self-employed male/females is 34 to 38% and as expected with a male female wage gap the prices are the main driver again with a range from 32 to 35 percent. Surprisingly not all women here appear to be better qualified than their male counter parts, quantities account for 0 to 4 percent of the wage gap. The unobservables keep playing an insignificant role of -1 to 1 percent.

Moving to the raw difference of self-employed males/wage working women a big dispersion is observed ranging from 21 to 54%. Combined with this bigger dispersion, another contradictory find is that opposed to all the male/female comparisons before both quantity and price have a big impact here. On top of that now also the unobservables play a big role and are the main reason for the big dispersion. Quantity effect explains 15 – 21 % of the wage gap, prices effect accounts for 15-16% and unobservables range from -18 to 24%.

Comparing the two wage gaps no clear conclusions can be drawn. The big dispersion in the second decomposition would suggest that lower qualified are better of working for wages, whilst the highest qualified women are better off being self-employed. Better off meaning less of a wage-gap, it could still hold that either working for wages or self-employment earns a higher wage despite a bigger wage gap.

Overall no support is found that women in self-employed have less of a wage-gap than those working for wages. A possible explanation for this is that men select to work in self-employment to earn higher wages, whilst women choose self-employment to facilitate household production (Hundley, 2000). Meaning that women in self-employment pool at the lower wages and therefore even comparing women on the high end of the female self-employed wage distribution have low wages compared to the males, resulting in the higher wage gap found at the 75th percentile.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***White Self-employed Male/Female*** | ***Raw difference*** | ***Quantity*** | ***Prices*** | ***Unobservables*** |
| **p25** | 0.377294 | 0.038689 | 0.324417 | 0.014188 |
| **Mean** | 0.373530 | 0.031171 | 0.342290 | 0.000069 |
| **p75** | 0.338975 | -0.001660 | 0.353217 | -0.012581 |
|  |  |  |  |  |
|  |  |  |  |  |
| ***White Self-employed Male/Wage working Female*** | ***Raw difference*** | ***Quantity*** | ***Prices*** | ***Unobservables*** |
| **p25** | 0.206949 | 0.216361 | 0.166052 | -0.175464 |
| **Mean** | 0.364824 | 0.205499 | 0.159310 | 0.000015 |
| **p75** | 0.542574 | 0.152400 | 0.152666 | 0.237508 |

## Results

### Male/Female overview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***OLS*** | ***Male*** |  | ***Female*** |  |
| **age** | 0.075373 | \*\*\* | 0.061839 | \*\*\* |
| **age\_sq** | -0.000720 | \*\*\* | -0.000600 | \*\*\* |
| **nchild** | 0.118661 | \*\*\* | 0.017708 | \*\*\* |
| **nchild\_sq** | -0.018500 | \*\*\* | -0.007733 | \*\*\* |
| **educ\_\_2** | -0.080634 | \*\*\* | -0.091703 | \*\*\* |
| **educ\_\_3** | 0.006520 |   | -0.060973 | \*\*\* |
| **educ\_\_4** | 0.070262 | \*\*\* | -0.020336 | \*\*\* |
| **educ\_\_5** | 0.115842 | \*\*\* | 0.033784 | \*\*\* |
| **educ\_\_6** | 0.145316 | \*\*\* | 0.067454 | \*\*\* |
| **educ\_\_7** | 0.304881 | \*\*\* | 0.240024 | \*\*\* |
| **educ\_\_8** | 0.436513 | \*\*\* | 0.378043 | \*\*\* |
| **educ\_\_9** | 0.509854 | \*\*\* | 0.533919 | \*\*\* |
| **educ\_\_10** | 0.794599 | \*\*\* | 0.752154 | \*\*\* |
| **educ\_\_11** | 1.019997 | \*\*\* | 0.994209 | \*\*\* |
| **race\_\_2** | -0.159912 | \*\*\* | -0.037140 | \*\*\* |
| **race\_\_3** | -0.135751 | \*\*\* | -0.072393 | \*\*\* |
| **race\_\_4** | -0.043980 | \*\*\* | 0.084283 | \*\*\* |
| **race\_\_5** | 0.055672 | \*\*\* | 0.063946 | \*\*\* |
| **race\_\_6** | -0.046834 | \*\*\* | 0.049902 | \*\*\* |
| **race\_\_7** | -0.106502 | \*\*\* | -0.035449 | \*\*\* |
| **race\_\_8** | -0.069460 | \*\*\* | -0.014540 | \*\*\* |
| **race\_\_9** | -0.071866 | \*\*\* | 0.003105 |   |
| **\_cons** | 0.604179 | \*\*\* | 0.776374 | \*\*\* |
|  |   |   |   |   |
| **Adjusted R^2**  | 0.3272 |   | 0.2723 |  |
| **Number of Obs** | 5238032 |   | 5049973 |  |

|  |  |
| --- | --- |
| ***Oaxaca overall*** | ***Coefficients*** |
| **Male** | 2.889374 | \*\*\* |
| **Female** | 2.665489 | \*\*\* |
| **difference** | 0.223884 | \*\*\* |
| **endowments** | -0.026625 | \*\*\* |
| **coefficients** | 0.247363 | \*\*\* |
| **interaction** | 0.003147 | \*\*\* |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***JMP*** | ***Raw Difference*** | ***Quantity*** | ***Prices*** | ***Unobservables*** |
| **p25** | 0.192560 | -0.035171 | 0.234681 | -0.006951 |
| **Mean** | 0.223884 | -0.023488 | 0.247363 | 0.000009 |
| **p75** | 0.254892 | -0.007877 | 0.259750 | 0.003019 |

|  |  |  |  |
| --- | --- | --- | --- |
| ***JMP Time*** | ***Raw Difference*** | ***Observable*** | ***Unobservable*** |
| ***Quantity*** | ***Prices*** | ***Quantity***  | ***Prices*** |
| **2001-2001** | -0.056555 | -0.017349 | -0.004508 | -0.050694 | 0.015995 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***Oaxaca overall*** | ***Coefficients*** |  |  |  |
| **group 1** | 2.806605 | \*\*\* |  |  |  |
| **group 2** | 2.627802 | \*\*\* |  |  |  |
| **difference** | 0.178804 | \*\*\* |  |  |  |
| **endowments** | 0.074243 | \*\*\* |  |  |  |
| **coefficients** | 0.095200 | \*\*\* |  |  |  |
| **interaction** | 0.009360 | \*\*\* |  |  |  |
|  |  |  |  |  |  |
| ***JMP*** | ***Raw Difference*** | ***Quantity*** | ***Prices*** | ***Unobservables*** |  |
| **p25** | 0.147230 | 0.055424 | 0.085171 | 0.006635 |  |
| **Mean** | 0.178804 | 0.083586 | 0.095200 | 0.000017 |  |
| **p75** | 0.176456 | 0.083575 | 0.106285 | -0.013403 |  |
|  |  |  |  |  |  |
| ***JMP Time*** | ***Raw Difference*** | ***Observable*** | ***Unobservables*** |
| ***Quantity*** | ***Prices*** | ***Quantity***  | ***Prices*** |
| **2001-2011** | 0.041174 | -0.000965 | 0.006822 | 0.029066 | 0.006251 |

### White/Black overview

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***OLS*** | ***White*** |  | ***Black*** |  |
| **age** | 0.070316 | \*\*\* | 0.061222 | \*\*\* |
| **age\_sq** | -0.000673 | \*\*\* | -0.000583 | \*\*\* |
| **nchild** | 0.072847 | \*\*\* | 0.053973 | \*\*\* |
| **nchild\_sq** | -0.012624 | \*\*\* | -0.011605 | \*\*\* |
| **educ\_\_2** | -0.102280 | \*\*\* | -0.059899 | \*\*\* |
| **educ\_\_3** | -0.022555 | \*\*\* | -0.052612 | \*\*\* |
| **educ\_\_4** | 0.033920 | \*\*\* | -0.047546 | \*\*\* |
| **educ\_\_5** | 0.095795 | \*\*\* | -0.044248 | \*\*\* |
| **educ\_\_6** | 0.128774 | \*\*\* | -0.023639 | \*\* |
| **educ\_\_7** | 0.289196 | \*\*\* | 0.169244 | \*\*\* |
| **educ\_\_8** | 0.425195 | \*\*\* | 0.317917 | \*\*\* |
| **educ\_\_9** | 0.541033 | \*\*\* | 0.425215 | \*\*\* |
| **educ\_\_10** | 0.794882 | \*\*\* | 0.666439 | \*\*\* |
| **educ\_\_11** | 1.019836 | \*\*\* | 0.906400 | \*\*\* |
| **female** | -0.268164 | \*\*\* | -0.145606 | \*\*\* |
| **\_cons** | 0.756258 | \*\*\* | 0.927626 | \*\*\* |
|  |   |   |   |   |
| **Adjusted R^2**  | 0.3171 |   | 0.2386 |  |
| **Number of Obs** | 8257330 |   | 915357 |  |

|  |  |  |  |
| --- | --- | --- | --- |
| ***Oaxaca detailed*** | ***Endowments*** | ***Coefficients*** | ***Interaction*** |
| **age** | 0.057583 | \*\*\* | 0.374857 | \*\*\* | -0.004405 | \*\*\* |
| **age sq** | -0.053312 | \*\*\* | -0.169460 | \*\*\* | 0.003436 | \*\*\* |
| **nchild** | -0.002695 | \*\*\* | 0.015297 | \*\*\* | 0.000447 | \*\*\* |
| **nchild sq** | 0.002197 | \*\*\* | -0.001948 | \*\*\* | -0.001442 | \*\*\* |
| **educ 1** | 0.000419 | \*\*\* | -0.000457 | \*\*\* | -0.000093 | \*\*\* |
| **educ 2** | -0.000103 | \*\*\* | -0.000283 | \*\*\* | -0.000079 | \*\*\* |
| **educ 3** | -0.000529 | \*\*\* | -0.000710 | \*\*\* | 0.000247 | \*\*\* |
| **educ 4** | 0.000086 | \*\*\* | -0.000049 |   | 0.000245 | \*\*\* |
| **educ 5** | 0.001282 | \*\*\* | 0.001267 | \*\*\* | 0.000159 | \*\*\* |
| **educ 6** | 0.003305 | \*\*\* | 0.002540 | \*\*\* | 0.000115 | \*\*\* |
| **educ 7** | 0.001639 | \*\*\* | 0.013564 | \*\*\* | 0.000411 | \*\*\* |
| **educ 8** | -0.004800 | \*\*\* | 0.004272 | \*\*\* | -0.000162 | \*\*\* |
| **educ 9** | 0.001327 | \*\*\* | 0.002511 | \*\*\* | 0.001688 | \*\*\* |
| **educ 10** | 0.028500 | \*\*\* | 0.006128 | \*\*\* | 0.000030 | \*\* |
| **educ 11** | 0.027941 | \*\*\* | 0.002113 | \*\*\* | 0.000035 | \*\*\* |
| **male** | 0.005703 | \*\*\* | 0.026719 | \*\*\* | 0.001154 | \*\*\* |
| **female** | 0.005703 | \*\*\* | -0.034560 | \*\*\* | 0.001402 | \*\*\* |
| **cons** |   |  | -0.146600 | \*\*\* |   |   |
|  |   |   |   |   |   |   |
| **Total** | 0.074243 |   | 0.095200 |   | 0.003187 |   |

### Yearly Male/Female

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***P25*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.239891 | -0.022055 | 0.266777 | -0.004831 |
| **2002** | 0.233615 | -0.022903 | 0.259856 | -0.003337 |
| **2003** | 0.217873 | -0.028628 | 0.246526 | -0.000025 |
| **2004** | 0.200670 | -0.031719 | 0.241682 | -0.009292 |
| **2005** | 0.205181 | -0.036138 | 0.245738 | -0.004419 |
| **2006** | 0.201404 | -0.034704 | 0.240180 | -0.004073 |
| **2007** | 0.199887 | -0.039737 | 0.241136 | -0.001512 |
| **2008** | 0.204794 | -0.026641 | 0.235130 | -0.003695 |
| **2009** | 0.194744 | -0.042178 | 0.244639 | -0.007717 |
| **2010** | 0.182322 | -0.043879 | 0.233312 | -0.007112 |
| **2011** | 0.154150 | -0.042136 | 0.212291 | -0.016004 |
|  |  |  |  |  |
| ***Mean*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.255768 | -0.010375 | 0.266101 | 0.000042 |
| **2002** | 0.247370 | -0.012113 | 0.259436 | 0.000048 |
| **2003** | 0.235313 | -0.015434 | 0.250707 | 0.000041 |
| **2004** | 0.237656 | -0.014420 | 0.252033 | 0.000043 |
| **2005** | 0.226494 | -0.021867 | 0.248338 | 0.000023 |
| **2006** | 0.225078 | -0.023869 | 0.248926 | 0.000022 |
| **2007** | 0.222872 | -0.026261 | 0.249111 | 0.000022 |
| **2008** | 0.228928 | -0.023955 | 0.252860 | 0.000023 |
| **2009** | 0.227555 | -0.023819 | 0.251352 | 0.000023 |
| **2010** | 0.211297 | -0.027782 | 0.239056 | 0.000023 |
| **2011** | 0.199212 | -0.032216 | 0.231402 | 0.000027 |
|  |  |  |  |  |
| ***P75*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.252131 | -0.021392 | 0.276158 | -0.002635 |
| **2002** | 0.262364 | -0.019305 | 0.282729 | -0.001060 |
| **2003** | 0.254375 | -0.017412 | 0.269172 | 0.002615 |
| **2004** | 0.253915 | -0.024286 | 0.271281 | 0.006920 |
| **2005** | 0.231112 | -0.026178 | 0.259203 | -0.001913 |
| **2006** | 0.249942 | -0.019257 | 0.266619 | 0.002580 |
| **2007** | 0.223144 | -0.029463 | 0.257402 | -0.004796 |
| **2008** | 0.223144 | -0.044601 | 0.269144 | -0.001400 |
| **2009** | 0.225405 | -0.031915 | 0.256365 | 0.000956 |
| **2010** | 0.231112 | -0.029965 | 0.255686 | 0.005390 |
| **2011** | 0.212174 | -0.041630 | 0.247143 | 0.006661 |

### Yearly White/Black

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***P25*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.153274 | 0.074826 | 0.072097 | 0.006351 |
| **2002** | 0.162519 | 0.084611 | 0.067754 | 0.010154 |
| **2003** | 0.134348 | 0.067919 | 0.062472 | 0.003957 |
| **2004** | 0.118784 | 0.048329 | 0.067862 | 0.002593 |
| **2005** | 0.148593 | 0.056330 | 0.083577 | 0.008686 |
| **2006** | 0.146603 | 0.062294 | 0.080811 | 0.003498 |
| **2007** | 0.148887 | 0.066612 | 0.076143 | 0.006133 |
| **2008** | 0.159389 | 0.068490 | 0.088501 | 0.002397 |
| **2009** | 0.165985 | 0.069304 | 0.089115 | 0.007566 |
| **2010** | 0.148702 | 0.059105 | 0.087334 | 0.002264 |
| **2011** | 0.170818 | 0.070124 | 0.093674 | 0.007020 |
|  |  |  |  |  |
| ***Mean*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.160665 | 0.082074 | 0.078345 | 0.000246 |
| **2002** | 0.161177 | 0.082506 | 0.078427 | 0.000245 |
| **2003** | 0.151200 | 0.077390 | 0.073605 | 0.000205 |
| **2004** | 0.161069 | 0.074581 | 0.086246 | 0.000242 |
| **2005** | 0.181960 | 0.085021 | 0.096850 | 0.000089 |
| **2006** | 0.183785 | 0.086948 | 0.096750 | 0.000087 |
| **2007** | 0.180242 | 0.088478 | 0.091680 | 0.000084 |
| **2008** | 0.191693 | 0.083599 | 0.108014 | 0.000080 |
| **2009** | 0.184614 | 0.080368 | 0.104150 | 0.000097 |
| **2010** | 0.184388 | 0.080277 | 0.104028 | 0.000084 |
| **2011** | 0.201839 | 0.088075 | 0.113662 | 0.000102 |
|  |  |  |  |  |
| ***P75*** | ***Raw difference*** | ***Quantity*** | ***Price*** | ***Unobservables*** |
| **2001** | 0.161212 | 0.079106 | 0.092025 | -0.009918 |
| **2002** | 0.173953 | 0.098198 | 0.091749 | -0.015993 |
| **2003** | 0.179048 | 0.083319 | 0.106493 | -0.010764 |
| **2004** | 0.174354 | 0.073582 | 0.105525 | -0.004754 |
| **2005** | 0.190951 | 0.087059 | 0.112811 | -0.008919 |
| **2006** | 0.186394 | 0.088717 | 0.108502 | -0.010825 |
| **2007** | 0.196168 | 0.091272 | 0.109190 | -0.004294 |
| **2008** | 0.212175 | 0.080874 | 0.132932 | -0.001632 |
| **2009** | 0.185649 | 0.078603 | 0.112891 | -0.005845 |
| **2010** | 0.186394 | 0.083934 | 0.115510 | -0.013051 |
| **2011** | 0.209039 | 0.085936 | 0.132943 | -0.009840 |

### State Male/Female

|  |
| --- |
| Mean |
|
| **State code** | **State** | **Raw Difference** | **Quantity** | **Prices**  | **Unobservables** |
| AL | Alabama | 0.260172 | -0.023980 | 0.284019 | 0.000133 |
| AK | Alaska | 0.213277 | -0.012034 | 0.224807 | 0.000504 |
| AZ | Arizona | 0.190294 | -0.031258 | 0.221448 | 0.000104 |
| AR | Arkansas | 0.214986 | -0.037634 | 0.252457 | 0.000163 |
| CA | California | 0.166805 | -0.039244 | 0.206028 | 0.000021 |
| CO | Colorado | 0.220895 | -0.015845 | 0.236633 | 0.000107 |
| CT | Connecticut | 0.239038 | -0.005675 | 0.244611 | 0.000103 |
| DE | Delaware | 0.199271 | -0.020071 | 0.219018 | 0.000323 |
|  | District of Columbia | 0.141546 | 0.030786 | 0.110457 | 0.000303 |
| FL | Florida | 0.187014 | -0.021705 | 0.208678 | 0.000041 |
| GA | Georgia | 0.211734 | -0.027602 | 0.239261 | 0.000075 |
| HI | Hawaii | 0.184119 | -0.027165 | 0.210968 | 0.000316 |
| ID | Idaho | 0.275083 | -0.000904 | 0.275669 | 0.000317 |
| IL | Illinois | 0.257695 | -0.011613 | 0.269260 | 0.000047 |
| IN | Indiana | 0.265581 | -0.017088 | 0.282580 | 0.000089 |
| IA | Iowa | 0.229034 | -0.026747 | 0.255632 | 0.000149 |
| KS | Kansas | 0.249302 | -0.020646 | 0.269765 | 0.000184 |
| KY | Kentucky | 0.230691 | -0.042169 | 0.272734 | 0.000126 |
| LA | Louisiana | 0.295272 | -0.028032 | 0.323173 | 0.000131 |
| ME | Maine | 0.193286 | -0.032327 | 0.225378 | 0.000235 |
| MD | Maryland | 0.195638 | -0.008328 | 0.203906 | 0.000061 |
| MA | Massachusetts | 0.234881 | -0.003321 | 0.238136 | 0.000065 |
| MI | Michigan | 0.261331 | -0.012150 | 0.273414 | 0.000067 |
| MN | Minnesota | 0.226291 | -0.018596 | 0.244814 | 0.000073 |
| MS | Mississippi | 0.245290 | -0.029416 | 0.274603 | 0.000103 |
| MO | Missouri | 0.243055 | -0.019025 | 0.262015 | 0.000064 |
| MT | Montana | 0.240008 | -0.021294 | 0.260993 | 0.000309 |
| NE | Nebraska | 0.230222 | -0.026479 | 0.256573 | 0.000127 |
| NV | Nevada | 0.192133 | -0.023158 | 0.215178 | 0.000113 |
| NH | New Hampshire | 0.262231 | -0.010533 | 0.272406 | 0.000358 |
| NJ | New Jersey | 0.246918 | -0.000542 | 0.247406 | 0.000055 |
| NM | New Mexico | 0.203494 | -0.043187 | 0.246409 | 0.000272 |
| NY | New York | 0.184864 | -0.036255 | 0.221092 | 0.000027 |
| NC | North Carolina | 0.190444 | -0.040171 | 0.230501 | 0.000114 |
| ND | North Dakota | 0.255834 | -0.030388 | 0.285813 | 0.000409 |
| OH | Ohio | 0.245962 | -0.009000 | 0.254904 | 0.000057 |
| OK | Oklahoma | 0.246715 | -0.029001 | 0.275556 | 0.000161 |
| OR | Oregon | 0.208631 | -0.015061 | 0.223528 | 0.000164 |
| PA | Pennsylvania | 0.248744 | -0.015118 | 0.263810 | 0.000052 |
| RI | Rhode Island | 0.208229 | -0.012043 | 0.220049 | 0.000223 |
| SC | South Carolina | 0.219473 | -0.026541 | 0.245875 | 0.000139 |
| SD | South Dakota | 0.192900 | -0.033437 | 0.226068 | 0.000269 |
| TN | Tennessee | 0.228204 | -0.025558 | 0.253650 | 0.000112 |
| TX | Texas | 0.240082 | -0.028112 | 0.268164 | 0.000030 |
| UT | Utah | 0.321458 | 0.036266 | 0.285040 | 0.000152 |
| VT | Vermont | 0.170614 | -0.037460 | 0.207732 | 0.000343 |
| VA | Virginia | 0.251252 | -0.005649 | 0.256832 | 0.000069 |
| WA | Washington | 0.251799 | -0.004713 | 0.256432 | 0.000080 |
| WV | West Virginia | 0.274280 | -0.038331 | 0.312354 | 0.000257 |
| WI | Wisconsin | 0.233296 | -0.022172 | 0.255403 | 0.000064 |
| WY | Wyoming | 0.342934 | -0.015819 | 0.358228 | 0.000526 |

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Rank Raw Difference** | **Rank Quantity** | **Rank Prices** |
| **Alabama** | 119 | 44 | 141 |
| **Alaska** | 59 | 87 | 47 |
| **Arizona** | 31 | 45 | 27 |
| **Arkansas** | 59.5 | 41 | 76 |
| **California** | 14 | 40 | 14 |
| **Colorado** | 69.5 | 96 | 54 |
| **Connecticut** | 74 | 118 | 66 |
| **Delaware** | 44.5 | 99 | 32 |
| **District of Columbia** | 5 | 150 | 3 |
| **Florida** | 15 | 51 | 12 |
| **Georgia** | 38 | 40 | 51 |
| **Hawaii** | 45 | 81 | 18 |
| **Idaho** | 136 | 125 | 127 |
| **Illinois** | 117.5 | 111 | 114 |
| **Indiana** | 128 | 99 | 134 |
| **Iowa** | 86 | 65 | 96 |
| **Kansas** | 100 | 59 | 116 |
| **Kentucky** | 84.5 | 28 | 112 |
| **Louisiana** | 145 | 67 | 149 |
| **Maine** | 25 | 44 | 35 |
| **Maryland** | 31 | 108 | 7 |
| **Massachusetts** | 73 | 128 | 48 |
| **Michigan** | 135 | 126 | 122 |
| **Minnesota** | 80 | 96 | 69 |
| **Mississippi** | 97.5 | 63 | 125 |
| **Missouri** | 103.5 | 95 | 95 |
| **Montana** | 116 | 118 | 98 |
| **Nebraska** | 83 | 62 | 95 |
| **Nevada** | 47 | 86 | 26 |
| **New Hampshire** | 119 | 85 | 122 |
| **New Jersey** | 97 | 140 | 63 |
| **New Mexico** | 51 | 21 | 71 |
| **New York** | 19 | 23 | 36 |
| **North Carolina** | 21 | 17 | 41 |
| **North Dakota** | 105 | 47 | 135 |
| **Ohio** | 114 | 119 | 90 |
| **Oklahoma** | 97 | 51 | 128 |
| **Oregon** | 42 | 95 | 25 |
| **Pennsylvania** | 103.5 | 97 | 102 |
| **Rhode Island** | 47.5 | 105 | 33 |
| **South Carolina** | 59 | 39 | 74 |
| **South Dakota** | 38 | 62 | 47 |
| **Tennessee** | 54 | 25 | 76 |
| **Texas** | 78 | 46 | 104 |
| **Utah** | 150 | 151 | 129 |
| **Vermont** | 12 | 27 | 18 |
| **Virginia** | 113 | 82 | 96 |
| **Washington** | 113.5 | 123 | 90 |
| **West Virginia** | 140 | 54 | 146 |
| **Wisconsin** | 90 | 81 | 90 |
| **Wyoming** | 153 | 116 | 153 |

### State White/Black

|  |
| --- |
| Mean |
|
| **State code** | **State** | **Raw Difference** | **Quantity** | **Prices**  | **Unobservables** |
| AL | Alabama | 0.256271 | 0.096014 | 0.160027 | 0.000229 |
| AK | Alaska | 0.225241 | 0.077712 | 0.141311 | 0.006219 |
| AZ | Arizona | 0.110122 | 0.010840 | 0.097962 | 0.001321 |
| AR | Arkansas | 0.201732 | 0.062835 | 0.138152 | 0.000744 |
| CA | California | 0.110372 | 0.021200 | 0.088974 | 0.000198 |
| CO | Colorado | 0.175768 | 0.072215 | 0.101917 | 0.001636 |
| CT | Connecticut | 0.284878 | 0.167307 | 0.116833 | 0.000738 |
| DE | Delaware | 0.168229 | 0.077672 | 0.089301 | 0.001255 |
|  | District of Columbia | 0.474358 | 0.207456 | 0.266529 | 0.000374 |
| FL | Florida | 0.221540 | 0.103579 | 0.117840 | 0.000121 |
| GA | Georgia | 0.221085 | 0.089906 | 0.131068 | 0.000111 |
| HI | Hawaii | 0.228597 | 0.119432 | 0.102273 | 0.006892 |
| ID | Idaho | 0.133849 | -0.018998 | 0.126330 | 0.026516 |
| IL | Illinois | 0.146375 | 0.074820 | 0.071339 | 0.000216 |
| IN | Indiana | 0.124127 | 0.053240 | 0.070370 | 0.000517 |
| IA | Iowa | 0.165818 | 0.064759 | 0.094621 | 0.006438 |
| KS | Kansas | 0.102820 | 0.058914 | 0.042029 | 0.001877 |
| KY | Kentucky | 0.156402 | 0.066116 | 0.089539 | 0.000748 |
| LA | Louisiana | 0.318973 | 0.113584 | 0.205136 | 0.000253 |
| ME | Maine | 0.228511 | 0.088481 | 0.116473 | 0.023557 |
| MD | Maryland | 0.144534 | 0.100142 | 0.044232 | 0.000160 |
| MA | Massachusetts | 0.263806 | 0.146473 | 0.116618 | 0.000715 |
| MI | Michigan | 0.118397 | 0.071522 | 0.046543 | 0.000332 |
| MN | Minnesota | 0.186388 | 0.079940 | 0.104881 | 0.001567 |
| MS | Mississippi | 0.323189 | 0.109911 | 0.213031 | 0.000247 |
| MO | Missouri | 0.114723 | 0.059881 | 0.054330 | 0.000512 |
| MT | Montana | 0.186814 | 0.015111 | 0.112114 | 0.059590 |
| NE | Nebraska | 0.121776 | 0.047984 | 0.069501 | 0.004291 |
| NV | Nevada | 0.140061 | 0.022031 | 0.116985 | 0.001044 |
| NH | New Hampshire | 0.186488 | 0.033610 | 0.141307 | 0.011571 |
| NJ | New Jersey | 0.233285 | 0.123868 | 0.109170 | 0.000247 |
| NM | New Mexico | 0.133080 | 0.041353 | 0.086297 | 0.005430 |
| NY | New York | 0.143665 | 0.113645 | 0.029912 | 0.000107 |
| NC | North Carolina | 0.252377 | 0.097584 | 0.154639 | 0.000153 |
| ND | North Dakota | 0.124863 | 0.036493 | 0.050171 | 0.038199 |
| OH | Ohio | 0.159069 | 0.074125 | 0.084687 | 0.000256 |
| OK | Oklahoma | 0.169086 | 0.059817 | 0.107941 | 0.001328 |
| OR | Oregon | 0.131259 | 0.044445 | 0.082338 | 0.004476 |
| PA | Pennsylvania | 0.122100 | 0.083504 | 0.038226 | 0.000370 |
| RI | Rhode Island | 0.264441 | 0.114991 | 0.144076 | 0.005374 |
| SC | South Carolina | 0.293567 | 0.112108 | 0.181190 | 0.000270 |
| SD | South Dakota | 0.188981 | 0.055117 | 0.112851 | 0.021012 |
| TN | Tennessee | 0.153300 | 0.070581 | 0.082413 | 0.000306 |
| TX | Texas | 0.162200 | 0.054188 | 0.107883 | 0.000129 |
| UT | Utah | 0.199139 | 0.033250 | 0.156280 | 0.009609 |
| VT | Vermont | 0.043341 | -0.029158 | 0.025083 | 0.047416 |
| VA | Virginia | 0.276900 | 0.149435 | 0.127260 | 0.000205 |
| WA | Washington | 0.148835 | 0.042847 | 0.104395 | 0.001593 |
| WV | West Virginia | 0.136357 | 0.016047 | 0.116446 | 0.003864 |
| WI | Wisconsin | 0.158649 | 0.087189 | 0.070058 | 0.001402 |
| WY | Wyoming | 0.200232 | 0.019303 | 0.158508 | 0.022421 |

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Rank Raw Difference** | **Rank Quantity** | **Rank Prices** |
| **Alabama** | 132 | 115 | 132 |
| **Alaska** | 127 | 95 | 133 |
| **Arizona** | 33 | 11 | 80 |
| **Arkansas** | 102 | 59 | 113 |
| **California** | 21 | 16 | 56 |
| **Colorado** | 85 | 76 | 69 |
| **Connecticut** | 135 | 148 | 100 |
| **Delaware** | 73 | 82 | 52 |
| **District of Columbia** | 153 | 153 | 153 |
| **Florida** | 103 | 98 | 96 |
| **Georgia** | 112 | 99 | 116 |
| **Hawaii** | 110 | 127 | 78 |
| **Idaho** | 59.5 | 9 | 124 |
| **Illinois** | 27.5 | 62 | 33 |
| **Indiana** | 35.5 | 39 | 38 |
| **Iowa** | 67 | 83 | 57 |
| **Kansas** | 12.5 | 64 | 16 |
| **Kentucky** | 74 | 84 | 58 |
| **Louisiana** | 146 | 133 | 147 |
| **Maine** | 110 | 109 | 91 |
| **Maryland** | 44 | 115 | 12 |
| **Massachusetts** | 131 | 140 | 102 |
| **Michigan** | 27 | 76 | 13 |
| **Minnesota** | 94 | 82 | 84 |
| **Mississippi** | 148 | 130 | 150 |
| **Missouri** | 23 | 52 | 26 |
| **Montana** | 72 | 48 | 81 |
| **Nebraska** | 30.5 | 52 | 29 |
| **Nevada** | 57.5 | 17 | 102 |
| **New Hampshire** | 76 | 35 | 108 |
| **New Jersey** | 114 | 129 | 86 |
| **New Mexico** | 50 | 30 | 60 |
| **New York** | 42 | 124 | 9 |
| **North Carolina** | 122 | 90 | 130 |
| **North Dakota** | 73 | 110 | 22 |
| **Ohio** | 56 | 64 | 51 |
| **Oklahoma** | 73 | 53 | 78 |
| **Oregon** | 30 | 34 | 45 |
| **Pennsylvania** | 12 | 75 | 8 |
| **Rhode Island** | 136 | 128 | 124 |
| **South Carolina** | 144 | 119 | 144 |
| **South Dakota** | 62 | 62 | 94 |
| **Tennessee** | 45 | 66 | 41 |
| **Texas** | 53 | 32 | 77 |
| **Utah** | 104 | 63 | 129 |
| **Vermont** | 7 | 13 | 11 |
| **Virginia** | 132 | 145 | 103 |
| **Washington** | 65 | 49 | 76 |
| **West Virginia** | 48 | 18 | 87 |
| **Wisconsin** | 78 | 105 | 38 |
| **Wyoming** | 111 | 60 | 116 |

### Male/Female State over time

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State code** | **State** | **Raw Difference** | **Differences in Observables** | **Differences in Unobservables** |
| AL | Alabama | -0.038244 | -0.009700 | -0.028544 |
| AK | Alaska | -0.057296 | -0.042546 | -0.014750 |
| AZ | Arizona | -0.060265 | -0.030302 | -0.029964 |
| AR | Arkansas | -0.021544 | -0.047154 | 0.025610 |
| CA | California | -0.051165 | -0.031100 | -0.020065 |
| CO | Colorado | -0.031602 | -0.019498 | -0.012104 |
| CT | Connecticut | -0.041217 | -0.023781 | -0.017436 |
| DE | Delaware | -0.056139 | -0.022400 | -0.033738 |
|  | District of Columbia | -0.029279 | 0.008963 | -0.038241 |
| FL | Florida | -0.021260 | -0.012428 | -0.008832 |
| GA | Georgia | -0.046965 | -0.033062 | -0.013903 |
| HI | Hawaii | -0.019592 | -0.033977 | 0.014385 |
| ID | Idaho | -0.085827 | -0.028232 | -0.057595 |
| IL | Illinois | -0.088898 | -0.038161 | -0.050737 |
| IN | Indiana | -0.061334 | -0.018436 | -0.042898 |
| IA | Iowa | -0.108700 | -0.017964 | -0.090736 |
| KS | Kansas | -0.000818 | -0.018589 | 0.017771 |
| KY | Kentucky | -0.049007 | 0.010533 | -0.059539 |
| LA | Louisiana | -0.020707 | -0.012955 | -0.007752 |
| ME | Maine | -0.036145 | -0.024570 | -0.011574 |
| MD | Maryland | -0.069861 | 0.004500 | -0.074361 |
| MA | Massachusetts | -0.044811 | 0.002341 | -0.047152 |
| MI | Michigan | -0.125712 | -0.027651 | -0.098062 |
| MN | Minnesota | -0.065073 | -0.023721 | -0.041352 |
| MS | Mississippi | -0.051595 | -0.003277 | -0.048319 |
| MO | Missouri | -0.053940 | -0.017692 | -0.036248 |
| MT | Montana | -0.100339 | -0.044648 | -0.055691 |
| NE | Nebraska | -0.070777 | 0.005094 | -0.075872 |
| NV | Nevada | -0.079155 | -0.010251 | -0.068904 |
| NH | New Hampshire | -0.052502 | -0.012228 | -0.040274 |
| NJ | New Jersey | -0.054321 | -0.025161 | -0.029160 |
| NM | New Mexico | -0.029389 | 0.000674 | -0.030062 |
| NY | New York | -0.073708 | -0.034584 | -0.039124 |
| NC | North Carolina | -0.083453 | -0.030038 | -0.053415 |
| ND | North Dakota | -0.034513 | -0.017741 | -0.016772 |
| OH | Ohio | -0.074256 | -0.012503 | -0.061752 |
| OK | Oklahoma | -0.031966 | -0.021625 | -0.010341 |
| OR | Oregon | -0.023628 | -0.003914 | -0.019714 |
| PA | Pennsylvania | -0.044773 | -0.018644 | -0.026129 |
| RI | Rhode Island | -0.072526 | -0.020224 | -0.052302 |
| SC | South Carolina | -0.045191 | -0.018612 | -0.026579 |
| SD | South Dakota | -0.050805 | -0.045529 | -0.005275 |
| TN | Tennessee | -0.066498 | -0.033116 | -0.033382 |
| TX | Texas | -0.036138 | -0.028373 | -0.007765 |
| UT | Utah | 0.009804 | -0.002303 | 0.012106 |
| VT | Vermont | -0.079921 | -0.017335 | -0.062587 |
| VA | Virginia | -0.027397 | -0.020414 | -0.006983 |
| WA | Washington | -0.030088 | -0.016037 | -0.014051 |
| WV | West Virginia | 0.005140 | 0.003706 | 0.001435 |
| WI | Wisconsin | -0.046844 | -0.000664 | -0.046180 |
| WY | Wyoming | -0.000053 | -0.017309 | 0.017256 |

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Rank Raw difference** | **Rank Observables** | **Rank Unobservables** |
| Alabama | 33 | 40 | 28 |
| Alaska | 18 | 4 | 35 |
| Arizona | 17 | 11 | 26 |
| Arkansas | 44 | 1 | 51 |
| California | 24 | 10 | 31 |
| Colorado | 38 | 24 | 38 |
| Connecticut | 32 | 18 | 33 |
| Delaware | 19 | 20 | 23 |
| District of Columbia | 41 | 50 | 21 |
| Florida | 45 | 37 | 41 |
| Georgia | 27 | 9 | 37 |
| Hawaii | 47 | 7 | 48 |
| Idaho | 5 | 14 | 9 |
| Illinois | 4 | 5 | 13 |
| Indiana | 16 | 28 | 17 |
| Iowa | 2 | 29 | 2 |
| Kansas | 48 | 27 | 50 |
| Kentucky | 26 | 51 | 8 |
| Louisiana | 46 | 35 | 43 |
| Maine | 34 | 17 | 39 |
| Maryland | 13 | 48 | 4 |
| Massachusetts | 30 | 46 | 15 |
| Michigan | 1 | 15 | 1 |
| Minnesota | 15 | 19 | 18 |
| Mississippi | 23 | 42 | 14 |
| Missouri | 21 | 31 | 22 |
| Montana | 3 | 3 | 10 |
| Nebraska | 12 | 49 | 3 |
| Nevada | 8 | 39 | 5 |
| New Hampshire | 22 | 38 | 19 |
| New Jersey | 20 | 16 | 27 |
| New Mexico | 40 | 45 | 25 |
| New York | 10 | 6 | 20 |
| North Carolina | 6 | 12 | 11 |
| North Dakota | 36 | 30 | 34 |
| Ohio | 9 | 36 | 7 |
| Oklahoma | 37 | 21 | 40 |
| Oregon | 43 | 41 | 32 |
| Pennsylvania | 31 | 25 | 30 |
| Rhode Island | 11 | 23 | 12 |
| South Carolina | 29 | 26 | 29 |
| South Dakota | 25 | 2 | 45 |
| Tennessee | 14 | 8 | 24 |
| Texas | 35 | 13 | 42 |
| Utah | 51 | 43 | 47 |
| Vermont | 7 | 32 | 6 |
| Virginia | 42 | 22 | 44 |
| Washington | 39 | 34 | 36 |
| West Virginia | 50 | 47 | 46 |
| Wisconsin | 28 | 44 | 16 |
| Wyoming | 49 | 33 | 49 |

### White/Black state over time

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **State code** | **State** | **Raw Difference** | **Differences in Predicted Gap** | **Differences in residual gap** |
| AL | Alabama | 0.028696 | 0.023800 | 0.004896 |
| AK | Alaska | 0.309853 | 0.160827 | 0.149026 |
| AZ | Arizona | 0.017048 | 0.017913 | -0.000865 |
| AR | Arkansas | -0.056556 | -0.033659 | -0.022896 |
| CA | California | 0.011587 | -0.031973 | 0.043561 |
| CO | Colorado | 0.100298 | 0.014709 | 0.085590 |
| CT | Connecticut | 0.096579 | 0.027886 | 0.068693 |
| DE | Delaware | -0.012169 | -0.036090 | 0.023921 |
|  | District of Columbia | -0.063653 | -0.071345 | 0.007692 |
| FL | Florida | 0.032765 | 0.009944 | 0.022821 |
| GA | Georgia | 0.040969 | 0.008537 | 0.032432 |
| HI | Hawaii | 0.261988 | -0.018992 | 0.280980 |
| ID | Idaho | -0.076586 | -0.122644 | 0.046058 |
| IL | Illinois | 0.052230 | -0.004682 | 0.056912 |
| IN | Indiana | 0.026223 | 0.020045 | 0.006178 |
| IA | Iowa | 0.231405 | 0.017884 | 0.213521 |
| KS | Kansas | 0.223034 | 0.002748 | 0.220286 |
| KY | Kentucky | 0.107487 | 0.046430 | 0.061057 |
| LA | Louisiana | 0.083543 | 0.016014 | 0.067529 |
| ME | Maine | 0.132582 | 0.226765 | -0.094183 |
| MD | Maryland | 0.099698 | 0.038741 | 0.060957 |
| MA | Massachusetts | 0.032371 | 0.044180 | -0.011809 |
| MI | Michigan | 0.053292 | -0.026317 | 0.079609 |
| MN | Minnesota | -0.022225 | 0.028201 | -0.050426 |
| MS | Mississippi | 0.042277 | -0.015899 | 0.058176 |
| MO | Missouri | 0.021615 | -0.027259 | 0.048874 |
| MT | Montana |  |  |  |
| NE | Nebraska | -0.018200 | -0.037973 | 0.019774 |
| NV | Nevada | 0.015818 | -0.008550 | 0.024368 |
| NH | New Hampshire | 0.069940 | 0.125786 | -0.055846 |
| NJ | New Jersey | 0.017345 | -0.001819 | 0.019165 |
| NM | New Mexico | 0.048818 | 0.126415 | -0.077597 |
| NY | New York | 0.015468 | 0.010957 | 0.004511 |
| NC | North Carolina | 0.028881 | 0.030687 | -0.001806 |
| ND | North Dakota | 0.214400 | -0.073402 | 0.287802 |
| OH | Ohio | 0.075863 | 0.006884 | 0.068980 |
| OK | Oklahoma | 0.106642 | 0.008310 | 0.098332 |
| OR | Oregon | 0.236482 | 0.050582 | 0.185899 |
| PA | Pennsylvania | 0.030962 | 0.003026 | 0.027937 |
| RI | Rhode Island | 0.035181 | 0.098248 | -0.063067 |
| SC | South Carolina | 0.030624 | 0.022067 | 0.008557 |
| SD | South Dakota | 0.188157 | -0.001455 | 0.189612 |
| TN | Tennessee | 0.070425 | 0.021088 | 0.049337 |
| TX | Texas | -0.033704 | -0.017728 | -0.015976 |
| UT | Utah | 0.270863 | 0.060714 | 0.210149 |
| VT | Vermont | 0.183476 | 0.169846 | 0.013630 |
| VA | Virginia | -0.009779 | -0.001112 | -0.008666 |
| WA | Washington | -0.004598 | 0.012726 | -0.017324 |
| WV | West Virginia | 0.148010 | -0.022548 | 0.170557 |
| WI | Wisconsin | 0.216094 | 0.056623 | 0.159471 |
| WY | Wyoming | 0.016277 | -0.182737 | 0.199014 |

|  |  |  |  |
| --- | --- | --- | --- |
| **State** | **Rank Raw difference** | **Rank Observables** | **Rank Unobservables** |
| Alabama | 18 | 35 | 14 |
| Alaska | 50 | 48 | 40 |
| Arizona | 14 | 31 | 12 |
| Arkansas | 3 | 7 | 6 |
| California | 10 | 8 | 26 |
| Colorado | 36 | 28 | 38 |
| Connecticut | 34 | 36 | 35 |
| Delaware | 7 | 6 | 22 |
| District of Columbia | 2 | 4 | 16 |
| Florida | 23 | 25 | 21 |
| Georgia | 25 | 24 | 25 |
| Hawaii | 48 | 12 | 49 |
| Idaho | 1 | 2 | 27 |
| Illinois | 28 | 16 | 30 |
| Indiana | 17 | 32 | 15 |
| Iowa | 46 | 30 | 47 |
| Kansas | 45 | 20 | 48 |
| Kentucky | 38 | 41 | 33 |
| Louisiana | 33 | 29 | 34 |
| Maine | 39 | 50 | 1 |
| Maryland | 35 | 39 | 32 |
| Massachusetts | 22 | 40 | 9 |
| Michigan | 29 | 10 | 37 |
| Minnesota | 5 | 37 | 5 |
| Mississippi | 26 | 14 | 31 |
| Missouri | 16 | 9 | 28 |
| Montana | #N/A | #N/A | #N/A |
| Nebraska | 6 | 5 | 20 |
| Nevada | 12 | 15 | 23 |
| New Hampshire | 30 | 46 | 4 |
| New Jersey | 15 | 17 | 19 |
| New Mexico | 27 | 47 | 2 |
| New York | 11 | 26 | 13 |
| North Carolina | 19 | 38 | 11 |
| North Dakota | 43 | 3 | 50 |
| Ohio | 32 | 22 | 36 |
| Oklahoma | 37 | 23 | 39 |
| Oregon | 47 | 42 | 43 |
| Pennsylvania | 21 | 21 | 24 |
| Rhode Island | 24 | 45 | 3 |
| South Carolina | 20 | 34 | 17 |
| South Dakota | 42 | 18 | 44 |
| Tennessee | 31 | 33 | 29 |
| Texas | 4 | 13 | 8 |
| Utah | 49 | 44 | 46 |
| Vermont | 41 | 49 | 18 |
| Virginia | 8 | 19 | 10 |
| Washington | 9 | 27 | 7 |
| West Virginia | 40 | 11 | 42 |
| Wisconsin | 44 | 43 | 41 |
| Wyoming | 13 | 1 | 45 |

1. The terms whites and blacks are used here because the dataset used in this paper uses this terminology and their definition of blacks encompasses more people than just African-Americans. Thus to avoid confusion the terms white and black are used instead of Caucasian and African-American. [↑](#footnote-ref-1)
2. This proxy however tends to overestimate the experience of women (Filer, 1993). He suggest to use predicted experience as opposed to potential experience and finds a decrease in the returns of schooling of about 20%. He however only uses the predicted experience for women and does not check if this predicted experience has a similar effect for males. [↑](#footnote-ref-2)
3. As an area of personal interest, differences between self-employed and those working for wages are investigated. This is a little beyond the scope of the subject, but is thought to perhaps have some interesting insights. First self-employed males are compared with self-employed females, followed by a Self-employed male comparison with females working for wages. Those two decompositions will then be compared. The reason behind this investigation is to see how much of the male female wage gap is due to females being paid differently. The logic behind this being that a self-employed woman should pay herself according to her worth, since she determines her own wage. Whilst a woman working for wages has to accept what her boss is willing to pay. Naturally the self-employed woman can still have someone that hires her and thus there still maybe some difference but it is expected it to be less. The racial wage gap is not investigated, because data on the self-employed is scarce and focusing just on blacks, which is about 10% of the data, would reduce the available data too much. Overall I find that the wage gap varies more over the percentiles for the self-employed and this is a positive variation at the low percentiles but negative at the higher percentiles. A detailed investigation of the results is given in the appendix. [↑](#footnote-ref-3)
4. One should note that the scale on each of the images is different and thus the same colour across maps does not represents the same value and darker values represent higher numbers. [↑](#footnote-ref-4)