The Wage Gap in Young Government Workers

Alice Gindin -Senior Thesis

4/28/2017

Abstract

In this paper I use data collected from the US Office of Personnel Management (OPM) in 2010 to examine the wage gap between new male and female government employees in their twenties. For most of these men and women, this is their first job out of school. In fact, the average level of postgraduate work experience is about four months. I test for the presence of a gender wage gap amongst these new employees. I use OLS to regress gender onto wage and use different covariates such as human capital, industry controls, and stem occupations to help explain the gap. I find that with the full specification of controls there still exists a 2-3% wage gap that is unexplained. This is less than the 6-11% that is still unexplained over the whole population, yet as the sample was younger than usual, this was expected. Hopefully from this work we can see how much of the wage gap Economists normally present is comprised of factors that apply to women just entering the workforce and how much cannot be attributed to those factors.

1) Introduction

The gender wage gap, in essence, is the difference between the average woman and the average man's salary. As more and more Economists have attempted to explain why the wage gap exists, and why it is as large as it is, a robust literature surrounding the gender wage gap has emerged. The heart of the literature is dedicated to determining what portion of the wage gap is caused by markets or firms discriminating against women, and what portion of the gap is caused by anything else.

The pay gap is an important policy issue for multiple reasons. First, if economists are ever able to definitively identify discrimination, then they could empower government to address the issue. Second there are some policies that people use to explain the wage gap, like failure to bargain or demand for more flexible hours so they can care for children. Policies could be put in place to discourage salary negotiations or encourage firms to allow more flexible hours to their employees or encourage telecommuting.

Most Economists agree that the wage gap becomes more pronounced as women enter their thirties [Blau and Kahn, (2016)]. As a result, when the literature studies the wage gap for women of a specific age, authors tend to focus on women in their thirties or later, trying to explain the reason for the increased gap [see Smithson, (2004); McCrate (2005)]. This has left a gap in the literature to investigate what the gender wage gap looks like for men and women in their twenties. Looking at the wage gap for younger workers has a useful purpose; when Economists try to identify factors that might be driving the unexplained wage gap, some of these factors would affect women starting the moment they enter the workforce, whereas other factors would not begin affecting women until their early thirties. By determining how large the wage gap is for women in their twenties is, Economists can determine the extent to which the factors that should affect women regardless of age are affecting the wage gap. Conversely, by comparing the gender wage gap for women in their twenties and the wage gap for women in their thirties, Economists can determine how much of the wage gap is stemming from factors that are not relevant until women enter their thirties.

In this paper I use data collected from the US Office of Personnel Management (OPM) in 2010 to examine the wage gap between new male and female government employees in their twenties. For most of these men and women, this is their first job out of school. In fact, the average level of postgraduate work experience is about four months. I test for the presence of a gender wage gap amongst these new employees. I use OLS to regress gender onto wage and use different covariates such as human capital, industry controls, and stem occupations to help explain the gap. I find that with the full specification of controls there still exists a 2-3% wage gap that is unexplained. This is less than the 6-11% that is still unexplained over the whole population, yet as the sample was younger than usual, this was expected. Hopefully from this work we can see how much of the wage gap Economists normally present is comprised of factors that apply to women just entering the workforce and how much cannot be attributed to those factors.

2) Review of the Literature

According to Blau and Kahn (2016), the completely uncorrected gender wage gap as of 2010 was 20.7%, or as people often put it colloquially, women make 79 cents to a man's dollar. There are four general categories of factors Economists use to explain portions of this gap; human capital, different industries/STEM jobs, labor force participation/time flexibility, and soft skills/personality traits. The first two categories (human capital and industry) currently account for another 10-15% of the gap leaving 6-11% unexplained [Blau and Kahn, 2016 & Dept. of Labor].

Mincer (1974), a labor economist, was the first to suggest using human capital variables to try and explain the gender wage gap. That same year, Mincer and Polachek (1974) wrote a paper that allowed for more flexible measures of human capital. Up until this paper Economists used age as a proxy for experience, but Mincer and Polachek realized that women's age could not proxy for their experience in the labor force because many of them had taken time off to raise children. They proposed a more natural way of determining their experience. Economists still use these human resources controls [see Blau and Kahn, 2016]. At the same time, human capital controls have become less explanatory of the wage gap over time because women have surpassed men in terms of educational attainment. Instead it is useful to control for what industry women and men are working in when studying the wage gap [Albrecht, 2003].

Most Economists agree that the gender wage gap widens as women get older [Canon and Golan, 2016]. The following two sections

2.1) Factors that should Contribute to the Wage Gap from Date of First Hire

The following factors, if they affect the gender wage gap at all should affect women just entering the workforce, and as a result affect the women in this paper's sample:

STEM - One factor that contributes to the gender wage gap is the number of men that go into STEM (Science, Technology, Engineering, and Math) vis-a-vis women. This trend is especially pronounced in the more mathematically intensive parts of STEM, as women participate in these fields far less often than in non-mathematically oriented parts of STEM (Ceci et. al 2014). This trend of women shying away from mathematics starts early but is not innate. Fryer and Levitt (2010) found that although the mean math scores for kindergarten boys and girls are the same, by the time those students are in 5th grade, the mean female math score is .2 standard deviations below the mean male math score. Equivalently the average female fifth grader is scoring in the 42nd percentile for male fifth graders on standardized math tests. This result holds across all racial categories and all regions of the country [Fryer & Levitt, 2010]. Goldin et. al. found that although women started closing the gap vis-a-vis men in math in the 1980s, they haven't made much progress since 1990. Paglin and Rufolo (1990) showed that there was a connection between mathematical ability and what field people chose to go into. They then demonstrated that this had an effect on male-female earning differentials.

Soft Skills - Some Economists have been examining how aspects of women's personality affect her future earnings. Betrand (2011) discusses psychological studies which give evidence that women have difficulty bargaining over salaries which hurts their earnings as the average man has no such qualms. Croson and Gneezy (2009) find that women are more risk averse which means they take fewer risks that could lead them to management positions.

Discrimination - finally if there is actual discrimination playing into the wage gap then it should affect women who are 25 just as much as it affects women who are 35.

2.2) Factors that Should Contribute After 30 but not Before

The gender wage gap widens as workers age at the median, yet this effect is far more pronounced at the top of the income distribution. In 2012, the top 1% of American earners was 89% male [Guvenen et. al., 2014]. This male predominance is especially pronounced in executive positions. According to Wolfers (2006), between 1990 and 2005 only 1.3% of ExecuComp CEOs were female. Despite the very low participation rates, when women actually attain high level executive positions there is very little evidence of a pay gap. Gayle et. al. (2012) actually found that when they controlled for human capital (education, labor market experience) and executive rank within a company, female executives earned more, were promoted at a higher rate, and had less job security than their male counterparts. The reason that women in the executives sample were making less than men was primarily because female executives were often lower on the executive ladder than men with the same qualifications.

Economists believe that much of this has to do with women in their thirties requesting more time flexibility or being able to dictate when a person works the hours required of them. Canon found that although women are starting to work more and more inflexible hours, they still lagged behind men in that respect and this was affecting their wages. Smithson found that within the accounting profession

I do not expect these factors would be present in the sample. First, the women and men in this sample are 25-29 and still starting their careers. This means that women and men alike have not been up for many promotions. Moreover motherhood is presumably not yet a factor for many of these women; most of these women have at least a college degree according to Pew Research Center, the median age women with bachelor’s degrees have their first child is 28 and for women with master's degrees it is 30 years old [Livingston 2015]. Without children women are far less likely to need flexible hours.

Why Pay Inequality Hurts Firms Too

Identifying factors that drive the wage gap is important for reasons other than a notion of equality or justice. As Gary Becker (1957) put it "When an act of discrimination occurs he must either pay or forfeit something for that privilege." Moreover, from a firm standpoint, often times pay inequality can be destructive. This applies whether or not the reason for pay inequality can be explained by human capital or personality traits.

Within the Behavioral Economics literature, there is a focus on the theory that relative wage can matter just as much or more than real wage to a worker's utility. The reasoning is that the worker retains a certain amount of utility from their real wage affording them some amount of purchasing power, however wage inequality can separately cause disutility to a worker from either jealousy of others or the resentment of others depending on whether the worker is on the low end or high end of the relative wage spectrum. In some cases the disutility of that jealousy or having others resent a worker is enough that said worker would rather work at a lower absolute wage. In essence the disutility from jealousy is reasonably sizable. Breza, Kaur, and Shamdasani (2016) found that workers were willing to give up 9.3% of their earnings to work in a factory block where every worker was paid the same wage for the same job.

3) DATA USED

The data used is a compilation of five different Office of Personnel Management (OPM) sources - The 2010 Government Accessions Data Set and the March, June, September, and December 2010 OPM Employment Cubes. The Accessions data set provided individual level observations containing an individual's salary, gender, age, years of experience, location, date of hire, occupation, and occupational category (meaning the OPM's assessment of the difficulty of the given occupation plus the level of education required for each job).

The March 2010 OPM Employment Cube listed which occupations are designated as STEM. These occupations were broken down into four categories - Science, Engineering, Information Technology (IT), and Mathematics. From this, I created a STEM designation dummy variable for each observation. Additionally, most observations in the 2010 Government Accessions data set were also included in either the March, June, September, or December 2010 Employment Cubes with their initial hire data (same salary and pay-grade adjustments). From this, via hire dates, salary, and other identifying data, I matched each observation in the Accessions Data Set to their education level listed in the Employment cubes. I did this by restricting the relevant Employment Cube to those who were hired in the correct stretch of time and then using the merge function in Stata to insert each Accession Data Set observation's corresponding educational level. For the 32,119 entries that were not matched to an education level, I created a dummy variable, edmatch, which took on the value 1 if the educational level had been matched to an observation, and a 0 otherwise.

From the given educational variable, I created 4 different variables to line up with models from the literature: a variable that represents years of education - *edyrs*, an indicator variable for vocational training - *voc*, an indicator variable for a bachelor's degree - *bac*, and an indicator variable for a master's degree - *mas*. The full coding for the educational variables can be found in Appendix 2.

Next, for the full specification model, I coded the industry dummies. There were 10 industry categories in total whose breakdown can be seen in Appendix 1. They were based off of the industry dummies in Blau and Kahn (2016). Of the 10 total, the omitted industry was miscellaneous. Note that *stem*, one of the 10 industries, was already in the data set, so there were only eight new variables created (there is no variable for the category Miscellaneous as it would cause issues with multicollinearity).

The 2010 Government Accessions data set contained 305,918 observation. Each observation was an employee who had either been hired by or transferred within the US Government in the fiscal year 2010. 143 had no occupation designated and were dropped. 1183 of these employees did not have a listed salary and were dropped from the sample. Of the remaining employees, 13879 were simply being transferred between branches of the government and were hence dropped from the sample. 39918 employees were part time workers and were dropped. Finally, I dropped all employees under the age of 20 and over the age of 29 from the sample leaving 84,933 total observations. Of these observations, 52,814 were successfully matched to the Employment Cubes and have their corresponding education level.

3.1 SAMPLE MEANS

Of the 84,933 observations in the sample, the average salary was $39,387. Breaking this down by gender the average male salary was $39,502 and the average female salary was 39,235$. The first thing to note is that the male and female average salaries are very close to one another. The female mean is only lagging behind by $125 each year. Additionally, the percent of the sample holding masters degrees is almost identical, 41.8% of women and 41.9% of men hold them. Women are more likely to hold at least a Bachelor's degree in this sample; 62.% of women have completed their Bachelor's whereas only 57.8% of men have. Men are slightly more likely to have received vocational training that woman at 1.17% to 1.09%. For Years of Education and holding STEM jobs, the differences are much larger. The average woman has 7 months more education than the average man. Both male and female years of education have double peaked distributions, and both distributions peak at 12 years and 16 years (corresponding to a high school and college diploma in most cases). Yet looking at Figure 1, it's notable that the larger of the two peaks for the male distribution is concentrated around 12 years, or a high school diploma whereas the larger of the two peaks for the female distribution is concentrated around 16 years, or a college diploma. At the same time, men were more than twice as likely to hold a STEM job than women.

Figure 1: Sample Means

VARIABLES	Sample Mean	Female Sample Mean	Male Sample Mean
Salary	39,387	39235.53	39501.89
Experience	0.503	.3861	.593
Fraction in STEM	0.0772	0.0449	.102
Fraction with Vocational Certification	0.0114	.0109	.0117
Fraction with Bachelor's	0.597	0.622	.578
Fraction with Masters	0.418	0.418	.419
Education in Years	14.19	14.55	13.907

It is also important to compare the observations in our set that have been matched to their education variables to the observations which have missing values for education variables. We must ensure that the observations that are missing data are distributed similarly to those observations which were successfully matched. Looking at these categories with respect to salary, the mean for the observations with values for education variables was \$39,700 and the mean for the observations missing education variables was \$38,600. The variances are 14323 and 14319 respectively, and both are right skewed. Appendix 3 has the the histograms of salary for the sample with and without the missing education variables, and the two seem to be distributed rather similarly. As such, running regressions with those observations with missing values dropped should not bias the sample. To be safe I run two phases of regressions, one with the observations with missing educational data dropped, and one where they are kept.

4) The Model

This paper uses OLS models to determine what portion of the wage gap is not explicable by other controls. I run two versions of each model. The first drops all observations that are missing educational level (roughly 40% of the sample). The second follows Allison's method for missing data points in which I replace each missing value for education with a zero and then add a covariate *edmatch* to any regression below that includes educational variables. Allison warns against this method if the data is not missing at random, however in the previous section I describe how education missing from an observation is uncorrelated with salary. Moreover, the way that these observations failed to match was a missing data point in a later set. This is not due to non-response on the part of the observational unit, it is due to clerical errors on the part of the OPM when transferring files or the original entry would have been dropped from the file.

In this paper I will run 4 models using OLS to get the coefficient estimates for covariates. Table 2 contains the coefficients for these regressions in the model where we drop those with missing education variables.

1. Model 1 - No Covariates - this specification regresses salary onto gender alone. It simply presents the total gender wage gap with no explanation. It is written out as:

$$log(salary)=\beta_0+\beta_1 female+\varepsilon$$

2. Model 2 - Human Capital Covariates - this specification includes the human capital specifications that Mincer and Polachek (1974) suggest to better predict the wages of

women in the labor force who often take time off and whose experience is not purely a function of age. It is written as:

$$log(salary)=\beta_0+\beta_1 female+\beta_2 age+\beta_3 exp+\beta_4 exp^2+\beta_5 edyrs+\beta_6 bac$$
$$+\beta_7 mas+\beta_8 voc+\varepsilon$$

3. Model 3 - Human Capital Covariates and STEM - this specification adds a STEM control to Model 2 and is written as follows:

$$log(salary)=\beta_0+\beta_1 female+\beta_2 age+\beta_3 exp+\beta_4 exp^2+\beta_5 edyrs+\beta_6 bac+$$
$$\beta_7 mas+\beta_8 voc+\beta_9 stem+\varepsilon$$

4. Model 4 - Full Specification - This specification adds on the industry controls that Blau and Kahn (2016) suggest. It is written as:

$$log(salary)=\beta_0+\beta_1 female+\beta_2 age+\beta_3 exp+\beta_4 exp^2+\beta_5 edyrs+\beta_6 bac+\beta_7 mas+$$
$$\beta_8 voc+\beta_9 stem+\beta_{10} soc+\beta_{11} admin+\beta_{12} med+$$
$$\beta_{13} bus+\beta_{14} law+\beta_{15} art+\beta_{16} skl+\beta_{17} con+\varepsilon$$

Table 2:

VARIABLES	(1) Bivariate	(2) Human Capital Specification	(3) Human Capital & STEM Specification	(4) Full Specification
Female	- 0.00851***	-0.0489***	-0.0312***	-0.0244***
	(0.00230)	(0.00238)	(0.00233)	(0.00234)
Age		0.0173***	0.0177***	0.0149***
		(0.000272)	(0.000263)	(0.000258)
Experience		-0.0468***	-0.0395***	-0.0158***
		(0.00287)	(0.00278)	(0.00271)
Experience Squared		0.0109***	0.00952***	0.00557***
		(0.000520)	(0.000504)	(0.000490)

Education in Years		0.0350***	0.0339***	0.0333***
		(0.000943)	(0.000913)	(0.000911)
Bachelor's Degree		0.150***	0.116***	0.104***
		(0.00409)	(0.00400)	(0.00390)
Master's Degree		0.0770***	0.0755***	0.0830***
		(0.00535)	(0.00518)	(0.00510)
Vocational Degree		0.0329***	0.0321***	0.00823
		(0.00891)	(0.00863)	(0.00836)
STEM Job			0.273***	0.353***
			(0.00464)	(0.00472)
10 Industry Controls?	No	No	No	Yes
Constant	10.53***	9.576***	9.564***	9.554***
	(0.00151)	(0.0137)	(0.0133)	(0.0133)
Observations	52,814	52,814	52,814	52,814
R-squared	0.000	0.322	0.364	0.411

Standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

This table contains the four models listed above taken over the sample of all observations with complete Education data.

Table 3 replicates Table 2 except that it includes all observations and contains an additional dummy variable as a covariate each time *edyrs* is a covariate.

Table 3:

VARIABLES	(1) Bivariate	(2) Human Capital Specification	(3) Human Capital & STEM Specification	(4) Full Specification
Female	-0.00851***	-0.0372***	-0.0145***	-0.0201***
	(0.00230)	(0.00203)	(0.00195)	(0.00198)
Age		0.0197***	0.0198***	0.0171***
		(0.000229)	(0.000219)	(0.000216)
Experience		-0.0546***	-0.0408***	-0.0174***
		(0.00233)	(0.00223)	(0.00218)
Experience Squared		0.0119***	0.00965***	0.00578***
		(0.000420)	(0.000402)	(0.000393)
Years of Education		0.0342***	0.0331***	0.0295***
		(0.00102)	(0.000970)	(0.000959)
Bachelor's Degree		0.147***	0.107***	0.0925***
		(0.00442)	(0.00424)	(0.00414)
Master's Degree		0.0712***	0.0702***	0.0777***
		(0.00577)	(0.00551)	(0.00541)
Vocational Training?		0.0294***	0.0290***	0.00178
		(0.00962)	(0.00919)	(0.00893)
Was Education Matched		-0.295***	-0.300***	-0.252***
		(0.0189)	(0.0181)	(0.0178)
STEM			0.330***	0.413***
			(0.00366)	(0.00372)

10 Industry Controls?	No	No	No	Yes
Constant	10.53***	9.822***	9.817***	9.804***
	(0.00151)	(0.00894)	(0.00854)	(0.00836)
Observations	84,933	84,933	84,933	84,933
R-squared	0.000	0.236	0.303	0.347

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

This table contains the four models listed above taken over the sample of all observations in the Sample.

As a note, many Economists, including Blau and Kahn (2016) argue that it is better to use the Oaxaca-Blinder decomposition because the coefficients should be more accurate as they are tailored only to the men or only to the women. This is because the Oaxaca-Blinder decomposition runs a given OLS regression on men and women separately and then uses the different coefficients to find the difference between the groups However, Cotton (1998) argues that it is better to use the pooled coefficient model because if the market were not discriminatory (which is the null hypothesis in most papers using the Oaxaca-Blinder decomposition) then the coefficients on covariates should be the same for both men and women. This is because when the market treats men and women the same, the effect of an extra year of schooling, or holding a STEM job should be the same whether regardless of gender. As a result I did not run the Oaxaca-Blinder decomposition on this data.

5) Results

5.1 The Pooled OLS Results - Unmatched Data Dropped from Sample

Model 1, the Bivariate Specification, shows how small the unexplained gender wage gap is within the sample. The model finds a very small, although negative coefficient on the female dummy variable, only -.00851, meaning there is only a .85% wage gap in the first place. Looking at the sample means, which were only $125 apart, this result is unsurprising. In fact because this is the bivariate specification, the coefficient, -.00851, is simply the difference between the mean female and mean male logged salary.

Moving onto Model 2, the Human Capital Specification, the coefficient estimate for the female dummy variable has increased nearly ten-fold. This is in large part because of the differences in education between men and women and because women have less average experience. Work experience is always expressed as a quadratic because it has two components; the first is that additional work experience is helpful to the worker and makes them better at their job this should increase their salary. However, the second component, which is relevant when workers are young, is that the more experienced workers have under their belt, the less time they could have gone to school because they could not be both a full time student and a full time worker [Mincer, 1974]. As a result the model uses two terms, an experience term that dominates at the start of a worker's career which is traditionally negative, and a squared experience term that begins to dominate as the worker's years of experience become more relevant than their educational background. In this case we find that both *experience* and *experience squared* are statistically significant beyond the 99% confidence level. We note that *experience* is both negative and slightly more than four times the magnitude of *experience squared* which is positive. This mean *experience squared* has to grow to be four times as large as *experience* for the positive effect from *experience squared* to become dominate, or another way, for the first four years out of school additional experience is a hinderance to a worker's wage. After about the four year mark, worker's salaries begin to get a net benefit from their work. This aligns with Mincer's predictions of the relationship between *experience* and *experience squared.*

As most of these observations in the set are new hires with very little work experience (the mean for women in the set is about 3 months whereas the mean for men in the set is around 5 months), women's mean two months lesser experience, combined with their mean seven months more education should mean that women have higher wages when all else is equal. As a result, since women in fact have very slightly lower wages than men instead of higher wages, Model 2 estimates an unexplained wage gap of 4.89%.

Model 3 reduces the gap introduced by Model 2 by adding a STEM covariate to Model 2. The STEM coefficient is .273 and is very statistically significant. This means that getting hired into a STEM job leaving all else equal will increase a worker's pay by 27%. Since men in the sample are more than twice as likely as women to hold a STEM job, the bump from working in STEM affects far more men in the sample than women. As a result the unexplained wage gap went from 4.89% to 3.12% of average male earnings. This

Model 4 adds the industry controls to Model 3 (and also to Model 2 since STEM is one of the industries we control for), reducing the unexplained gap even farther. Blau and Kahn (2016) found that when they introduced industry controls to their human capital specification, creating a "full specification" like Model 4, the explained gap shrunk by about 50% according to their 2010 data. In this paper (using different 2010 data) the unexplained gap shrinks from 4.89% to 2.44% between the Human Capital and Full Specifications. Thus the results from Model 4 are consistent with what was expected from the literature.

5.2 The Pooled OLS Results - Unmatched Data in Sample with Dummy

In this case all results continue to be statistically significant, only they are smaller in the new sample. Model 1 comes out essentially the same since the difference between men and women's salaries in both the restricted and unrestricted samples are essentially the same. In Model 2 the coefficient is smaller, only indicating a 3.7% gap and in Model 3, after adding in STEM the estimated coefficient drops to 1.4%. Surprisingly, in the final model, adding in the industry controls actually increases the unexplained wage gap back to 2.01% which is closer to the dropped sample's model 4 which gave an unexplained gap of 2.44%. Because these coefficients are inconsistent, I am more inclined to trust the coefficients from the dropped regression. One possible reason for the difference is the the dummy variable interacting with some industry variables.

6) Conclusion

I found that for newtly hired US government workers in their twenties, the unexplained wage gap was much smaller than it is for the population as a whole. Although these results are unsurprising, it is useful to know that this is true. For one if thewage gap was as sizable as it is for women in the population at large, then many of the explanations Economists currently use for the wage gap would be rendered incorrect. Since this data is generalizable to US Government employees, this evidence suggests that only about 2% of the gender wage gap can be attributed to factors that would affect women in their twenties and thirties alike. As gender based discrimination is likely to affect both groups in a similar manner, this result indicates that at least within the US Government, gender based discrimination presumably does not account for more than 2% of the wage gap within the US Government. This is positive as the US Government at any given time employs more than 2,000,000 citizens.

The most natural extension to this paper would be to test what the characteristics of the gender wage gap are at large to see if these results are applicable to the country as a whole, if granted access to the full OPM database. Another option would be to use data like the CPS and PSID as Blau and Kahn did to see if these results hold for young new hires at large.

APPENDIX 1 : List of Variables and Sources

Variable Name	Description	Source
salary	An individual's annual salary (not adjusted in the case of unpaid leave)	2010 Accessions
female	Dummy - 1 for female, 0 for male	2010 Accessions
edlvl	How much education a person has/whether or not it is vocational training. Coded 1-22	Mar-Dec 2010 Fedscope Employment Cubes
edyrs	Years of education	Coded from edlvl
voc	Dummy - 1 if received vocational training, 0 otherwise	Coded from edlvl
bac	Dummy - 1 if received bachelor's degree, 0 otherwise	Coded from edlvl
mas	Dummy - 1 if received master's degree, 0 otherwise	Coded from edlvl
age	Age in years	2010 Accessions
exp	Experience in years	2010 Accessions
exp2	Experience in years squared	Coded from exp
occup	Occupation title	2010 Accessions
stem	Dummy - 1 if occupation is designated STEM (science, technology, engineering, math), 0 otherwise	Mar-Dec 2010 Fedscope Employment Cubes
loc	Location in country	2010 Accessions
agysub	Agency and subagency code	2010 Accessions
bus	Personnel Mgmt & Industrial Relations, Business and Industry, Accting and Budget	Mar-Dec 2010 Fedscope Employment Cubes
STEM	Nat Resource Mgmt and Bio, physical sci math and stats, IT, Engineering	Mar-Dec 2010 Fedscope Employment Cubes
soc	SS psych and welfare, Education	Mar-Dec 2010 Fedscope Employment Cubes

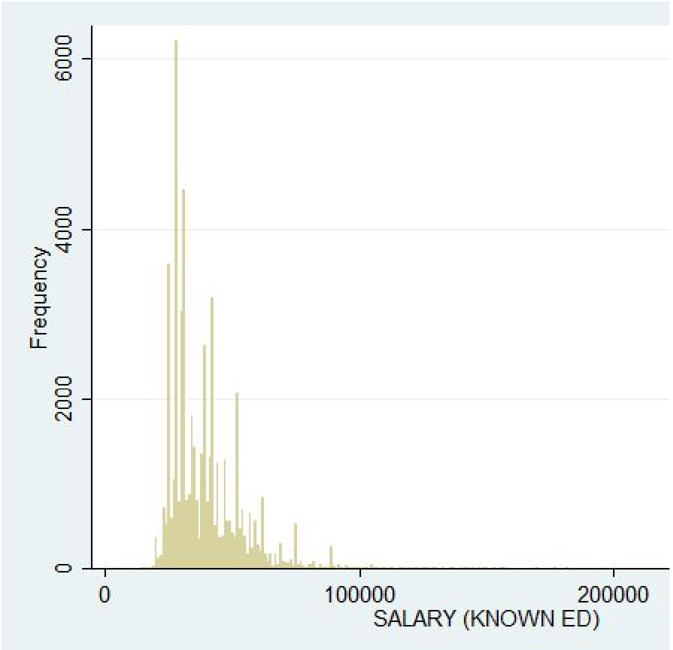
cler	General Administration and Office services, Equipment, Facilities and Services, Supply, Transportation	Mar-Dec 2010 Fedscope Employment Cubes
med	Medicine, Hospital, Dental, Public health, Veterinarians	Mar-Dec 2010 Fedscope Employment Cubes
Law	Copyright, Investigations and Inspections	Mar-Dec 2010 Fedscope Employment Cubes
art	Info and Arts, Library Archives	Mar-Dec 2010 Fedscope Employment Cubes
skl	Installation adn maint, Machine Operation	Mar-Dec 2010 Fedscope Employment Cubes
con	Construction, Painting, Transport, Industry Equipment Operations	Mar-Dec 2010 Fedscope Employment Cubes
edmatch	Dummy - 1 if edyrs exists, 0 otherwise	From edyrs

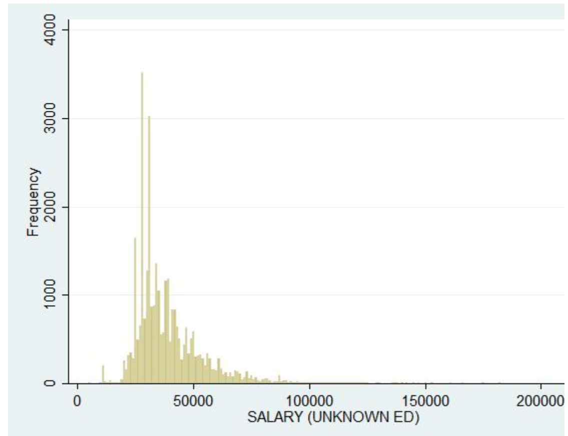
APPENDIX 2 : Coding for Education Variables

Value of edlvl	edyrs	voc	bac	mas
1	6	0	0	0
2	8	0	0	0
3	10	0	0	0
4	12	0	0	0
5	10	1	0	0
6	12	1	0	0
7	12.5	0	0	0
8	13	0	0	0
9	14	0	0	0
10	14	0	0	0

11	15	0	0	0
12	16	0	0	0
13	16	0	1	0
14	17	0	1	0
15	17	0	1	0
16	17	0	1	0
17	17.5	0	1	1
18	18	0	1	1
19	18	0	1	1
20	20	0	1	1
21	20	0	1	1
22	22	0	1	1

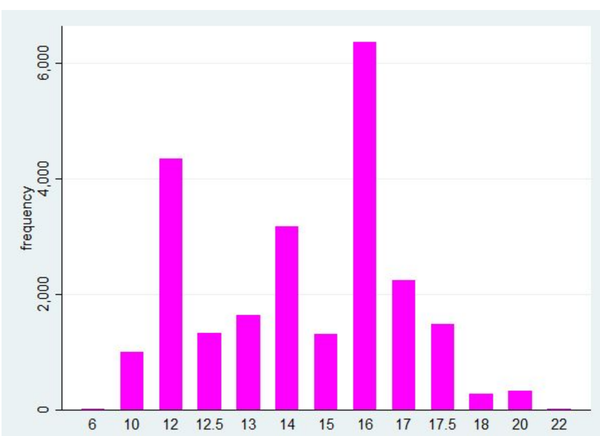
APPENDIX 3: Graph of Salaries for Observations with and without known Education Levels

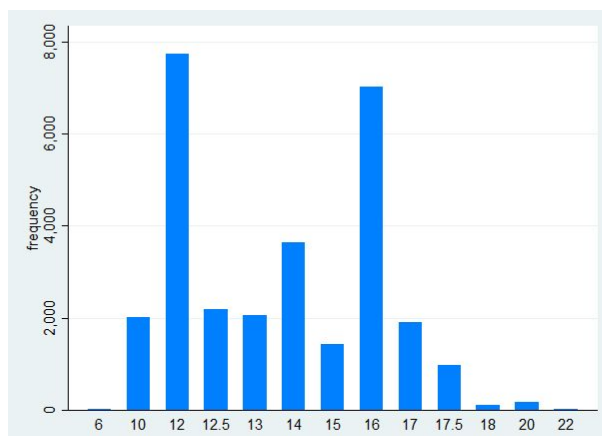




Note these graphs are on different scales so I resized them to be on nearly identical scales. See how the distributions essentially line up.

APPENDIX 4: Years of Education for Women and Men in the Sample





Women are on the left, men are on the right.