Marguerite Morgan

4/29/16

Wealth Inequality in the United States

**Introduction:**

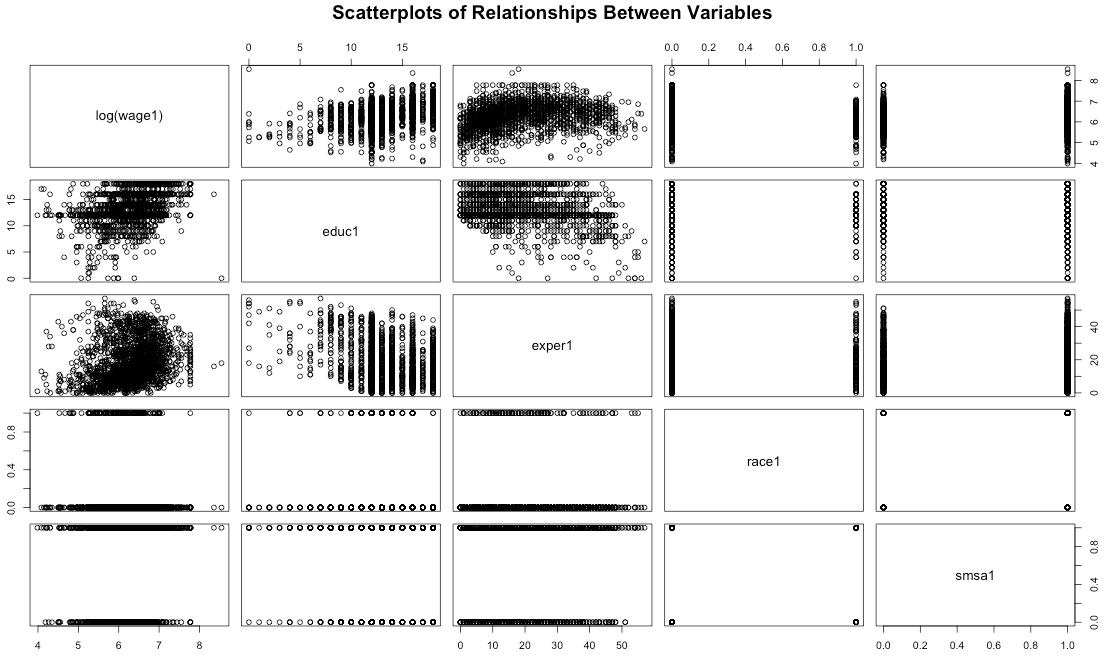
Wealth inequality in the United States has reached alarming levels in recent years. As of 2012, 1% of Americans held 40% of the nation’s wealth while the bottom 80% had only 7% of the country’s wealth.[[1]](#footnote-1) Income inequality falls along racial lines as well: a typical black household today only has 6% of the wealth of a typical white household.[[2]](#footnote-2) This disparity has become so extreme that in 2015, President Obama called economic inequality “the defining challenge of our time.”[[3]](#footnote-3) The dataset “uswages,” which contains information on the distribution of wages for 2,000 American males as of 1988 is thus of particular interest and importance in our current climate. The dataset includes data on weekly wage, years of education, years of experience, race (coded 1 for black and 0 for white), geographic location in the country, full versus part-time work, and whether someone lives in a standard metropolitan statistical area i.e. a city.

The overall research question of interest is how wages for full-time male workers in the United States are related to years of education, years of experience, race, and whether or not someone lives in a metropolitan area. In particular it would be interesting to see whether there is a significant difference in wages between blacks and whites when educational attainment is controlled for. Can the wage gap between blacks and whites in the U.S. be attributed to the fact that black Americans are less likely to go to college? Further, I will investigate whether educational attainment or years of experience has a bigger impact on wage. How important is a college education? Based on prior knowledge, I would expect that wages will increase with years of education and years of experience, they would be higher for white men compared to black men and would be higher for those living in cities. I would also hypothesize that a racial wage gap exists regardless of educational attainment and that amount of experience has a larger positive impact on wage than years of education.

**Methods and Results:**

To begin the analysis, I first cleaned up the dataset by sub-setting the data to only focus on full-time workers as opposed to part-time workers. I noticed there were 13 observations with negative values for years of experience. Because it does not make sense to have a negative value for years of experience, I deleted these values from the dataset leaving my final dataset with 1,802 observations. I then looked at the distribution of wages. There is a wide range of wages: the minimum wage value is $53.83 (about $10.77 a day) and the maximum wage is $5,144.00 (about $1,028.80 a day). The distribution of wages is highly right skewed as seen in Figure B. Therefore even in this sample of 1,802 males there is obvious evidence of high levels of income inequality.

In order to initially observe the relationships between the variables, I created scatterplots and noticed that the wage data points appeared to be clustered around the bottom of the plots. I thus performed a log transformation on wage in order to better visualize the relationships. As seen in Figure A, it appears that there is a roughly positive, linear relationship between log(wage) and years of education: as years of education increases, log(wage) increases as well. The relationship between experience and log(wage) is more difficult to interpret from this plot. There may be a possible positive, linear relationship between experience and log(wage) but we cannot draw any definitive conclusions. There appears to be a larger range of wages for whites than for blacks and black wages appear to be generally lower than white wages. There also appears to be a larger range of wages for those that live in metropolitan areas compared to those who do not and those who live outside cities appear to have generally lower wages. Years of education and years of experience do not appear to have any obvious correlation because their scatterplot has no clear pattern. Blacks seem to have generally fewer years of education except for one outlier. Furthermore, blacks appear to have fewer years of experience generally compared to whites.

*Figure A*

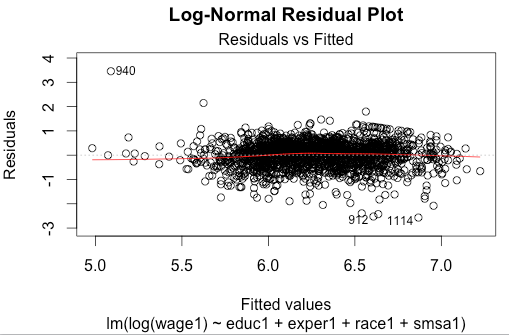
From the scatterplot above, it appears that number of years of education impacts the variance of wage and thus the dispersion of wage may vary depending on educational attainment. Therefore it makes sense to try jointly modeling the mean and dispersion. However, the original log-normal model and the joint mean and dispersion model have almost identical residual plots and extremely similar R^2 values. Thus, joint modeling of the mean and dispersion is not useful.

Now, we turn to modeling the effect of education, experience, race, and living in a metropolitan area on wage. First, I ran a linear model with a log transformation on wage. Since all wage values are positive and the distribution of wage appears highly right-skewed as seen in Figure B, I also ran a gamma GLM using the log link and then compared these two models. Figure C shows the residual plots for each regression. Although the residual plots look very similar, the log-normal residual plot shows that the residuals are slightly fanning outwards as they increase whereas the residuals in the gamma GLM plot are more evenly distributed.

*Figure B*



*Figure C*



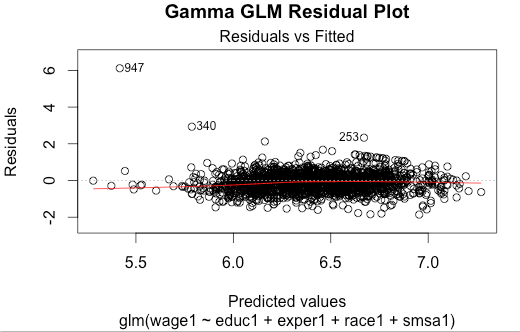
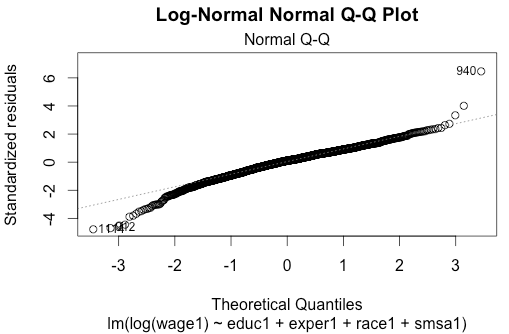
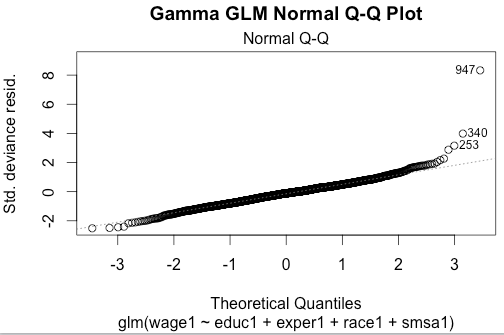


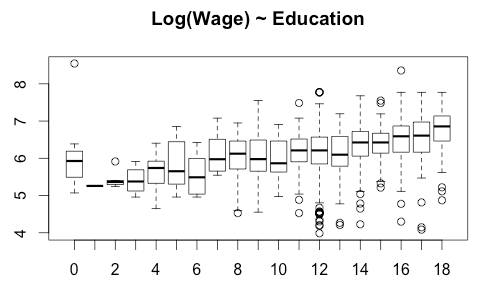
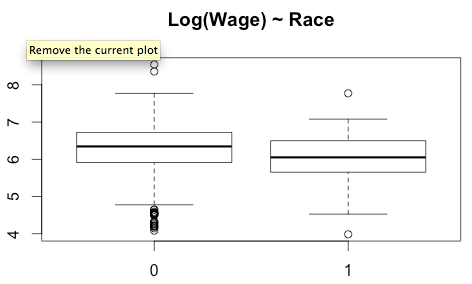
Figure D shows the Normal Q-Q plots for each regression. The Normal Q-Q plot for the gamma GLM appears to be more linear than the log-normal plot. Thus we conclude that the gamma GLM is a more suitable model than the log-normal model. Furthermore, the variance of the log-normal model is about 0.29398 whereas the MLE for the variance for the gamma GLM is about 0.275459. Thus, the variance for the gamma GLM is slightly smaller and our intuition to proceed with the gamma GLM with the log link is supported.

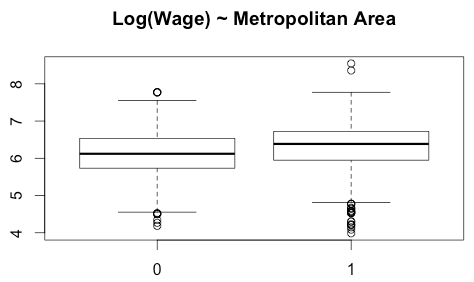
*Figure D*



Looking at the boxplots of each variable in Figure E, it appears there are some clear outliers. For example, there is one individual who has zero years of education yet has a wage very close to the maximum wage in the dataset ($5,114.00). Although removing these outliers may improve the fit of our model it is important to include them because they contain valuable information about the fact that wages can sometimes be random. Despite the fact that wage is often associated with variables such as education and amount of experience, there are many cases where individuals without many years of education or experience obtain high-paying jobs through personal connections or pure luck. These outliers represent that possibility and thus will be kept in the model.

*Figure E*







Now we compare different gamma GLM models in order to see which variables and interactions to include in our model. Intuitively I would suspect there to be possible interaction between years of education and years of experience. I also am interested to see if there is an interaction between years of education and race, and years of experience and race. I will thus compare the full model that regresses education, experience, race, metropolitan area, the interaction between education and experience, the interaction between race and education, and the interaction between experience and education on wage to each of its nested models. In order to compare these 8 models (including the null model), I ran an ANOVA using the F-test. The results as shown in Table 1 indicate that Model 4, which includes education, experience, race, and metropolitan area is the best model. None of the interactions were significant.

*Table 1*

ANOVA table:

|  |  |  |  |
| --- | --- | --- | --- |
| Model | Deviance | F | P-value |
| Wage~1 |  |  |  |
| Wage~educ | 86.226 | 160.7930 | <0.0001 |
| Wage~educ+exper | 60.020 | 111.9240 | <0.0001 |
| Wage~educ+exper+race | 4.710 | 8.7829 | 0.003081 |
| Wage~educ+exper+race+smsa | 10.833 | 20.2020 | <0.0001 |
| Wage~educ+exper+race+smsa+educ:exper | 0.197 | 0.3682 | 0.544080 |
| Wage~educ+exper+race+smsa+educ:exper+educ:race | 0.078 | 0.1460 | 0.702394 |
| Wage~educ+exper+race+smsa+educ:exper+educ:race+exper:race | 0.050 |  |  |

The final model chosen is thus the model that regresses education, experience, race, and whether someone lives in a metropolitan area on wage using the gamma GLM with the log link. This model is summarized below in Table 2. Every independent variable is significant. Since the MLE for the dispersion is less than 0.54, it appears the model is slightly underdispersed meaning the observed variability is slightly less than the expected variability. The underdispersion could be due to a variety of reasons including an incorrect model, a small sample size, or clustered sampling rather than random sampling. In this case, the sample size is relatively large (greater than 1,000) and we are unsure how the data was sampled. Because the residual plot appears to be a random scatter about zero and the normal Q-Q plot is roughly linear, we will not assume that this underdispersion signifies an incorrect model and will proceed with Model 4.

*Table 2*

Summary of Model 4: wage ~ educ + exper + race + smsa, family=Gamma, link=log

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | T-value | P-value |
| Intercept | 4.945737 | 0.095293 | 51.900 | <0.0001 |
| Educ | 0.080812 | 0.006080 | 13.291 | <0.0001 |
| Exper | 0.015629 | 0.001438 | 10.869 | <0.0001 |
| Race | -0.204559 | 0.065601 | -3.118 | 0.00185 |
| SMSA | 0.184419 | 0.040703 | 4.531 | <0.0001 |

Null deviance: 672.25 on 1801 degrees of freedom

Residual deviance: 510.46 on 1797 degrees of freedom

Dispersion = 0.5426491

MLE for the dispersion is about 0.2711124.

In order to determine whether obtaining higher education significantly impacts wages, I created a new variable, “college,” which is coded as 1 for all individuals with more than 12 years of education and 0 for all individuals with 12 or less years of education. I then created a new regression model, “Education Model,” using this new variable in place of education. The results of the regression are found in Table 3; again all of the variables are significant.

*Table 3*

Summary of Education Model: wage ~ race + smsa + college + exper, family=Gamma, link=log

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Estimate | Standard Error | T-value | P-value |
| Intercept | 5.864747 | 0.038402 | 152.720 | <0.0001 |
| Race | -0.216363 | 0.053276 | -4.061 | <0.0001 |
| Smsa | 0.176404 | 0.033110 | 5.328 | <0.0001 |
| College | 0.424361 | 0.028984 | 14.641 | <0.0001 |
| Exper | 0.013305 | 0.001139 | 11.684 | <0.0001 |

**Conclusions:**

Based on Model 4, which regresses education, experience, race, and metropolitan area on wage using a gamma GLM with a log link, we see that all of the independent variables included in the model are significantly associated with wage. Education is significantly positively associated with wage meaning that as education increases by 1 year, average wage increases by e^0.080307= $1.08 when controlling for experience, race, and whether someone lives in a city. As experience increases by 1 year, average wage increases by e^0.015743= $1.02 when controlling for education, race, and whether someone lives in city. Black men on average have wages that are e^-0.204115=$0.82 less than wages for white men when controlling for education, experience, and whether someone lives in a city. Those that live in a city have wages that are on average e^0.184936= $1.20 more than those that do not live in a city when controlling for education, experience and race.

These findings support the hypotheses that as years of experience and education increase average wage increases, blacks have lower wages than whites regardless of educational attainment, and those that live in cities have higher wages than those that do not. It appears that years of education and years of experience have a very similar impact on wage; there is only a 6-cent difference in their effect on wage. This is interesting considering that in today’s work environment, having valuable work experience is often said to be more important than amount of education.

The Education Model explores how obtaining higher education versus not obtaining higher education impacts wage. The new variable college is significantly, positively associated with wage. Those who obtain higher education have on average, wages that are e^0.424361 = $1.53 higher than whose with 12 or fewer years of education when controlling for race, living in a city, and work experience. In this model, work experience has a similar impact on wage as it did in Model 4. As experience increases by 1 year, average wage increases by e^0.013305 = $1.01 when controlling for education, race, and whether someone lives in a city. Thus, attending college has a larger impact on wage than work experience. Note that here we are assuming that those with less than 12 years of education did not obtain any higher education.

In general, from this analysis we can conclude that wages are disproportionately distributed in the United States with the majority of people receiving very low wages and a small minority receiving exceptionally high wages. Regardless of where someone lives, their educational attainment, or their amount of work experience, black males have significantly lower wages than white males on average. Educational attainment and work experience have similar impacts on wage. However, when educational attainment is categorized as whether or not someone obtains higher education, educational attainment has a greater impact on average wage than work experience. It is important to keep in mind that the circumstances of work experience may greatly impact this variable’s association with wage. Work experience in a particular field may lead to significantly higher wages at a job in that field than work experience in an unrelated industry. Because we do not know how this question was asked, it is difficult to definitively determine the monetary impact of work experience on wage. This dataset is from 1988 and the economic and racial climate during that time period is different than it is now. Economic inequality and racial disparities have increased over the years so it would be interesting to compare this analysis to an analysis of present-day data.

1. *Wealth Inequality in America*. N.p., 20 Nov. 2012. Web. [↑](#footnote-ref-1)
2. Shin, Laura. "The Racial Wealth Gap." *Forbes*. N.p., 26 Mar. 2015. Web. [↑](#footnote-ref-2)
3. Fitz, Nicholas. "Economic Inequality: It's Far Worse Than You Think."*Scientific American*. N.p., 31 Mar. 2015. Web. [↑](#footnote-ref-3)