Effect of Education on Wage Earning

Group Members: Quentin Talley, Thomas Wang, Geoff Zaski

Abstract

The scope of this project includes individuals aged 18-65 who finished their education and do not have significant disadvantages in seeking employment selected from a survey of Georgia residents regarding wage income, education, and other factors. By obtaining estimates of the multiple linear regression coefficients of education, experience, hours worked per week on education, sex, marital status, and employment type, a ceteris paribus effect of educational level on wage income of individuals. It was found that keeping other variables, increasing 1 year of education increases individual income by nearly $8000 per year. This figure can be important in policy making in terms of the revenue effect of increasing funding for education to boost educational level and in individual decisions by providing quantitative indicators of the tradeoffs of furthering education.

1. Introduction

This project aims to explore the effects on income the education level of an individual would have and analyze other variables that help delineate the ceteris paribus effects of education on income. Education, in theory, could be hypothesized to have a positive impact on the income of an individual, holding other variables constant, since having a higher level of education means a higher amount of human capital, which increases productivity and correspondingly income. Other factors that have some extent of collinearity with education could be reasonably expected to impact income, such as experience. To more completely study the ceteris paribus of education on income, such variables were included in a multiple linear regression. Experience could be expected to have a negative correlation with education, and therefore ignoring them would lead to a positive bias in the estimate of the least-squares coefficient education has on income, since experience could only plausibly positively correlate with individual income. Other variables, such as hours of work per week, while not being necessarily collinear with education levels, would contribute highly to explaining the large standard deviation of income at every level of education. Analysis of the impact of education on income using multiple linear regression could, however, have the problem of violating the homoskedasticity assumption of the regression model as with higher education, the field of study of individuals would become increasingly divergent, leading to a higher variance of the error term.

Uncovering the ceteris paribus effect of education on income has significance in economic policies, particularly educational spending. If higher levels education could, in reasonable confidence, be concluded to increase the average income of individuals, then increases of funding for education can be in the long-term revenue-neutral without increasing rate of taxation as the total income to tax would increase. Such hypotheses are consistent with macroeconomic theories relating to human capital and how it increases productivity per worker. There are, however, limitations to the extent this conclusion is comprehensive in whether any increase in funding for education would be beneficial, as increased funding for education with the intent to boost the average level of education does not necessarily entail meaningful increases in quality of education, which cannot easily be formatted into analyzable data.

2. Literature Review

One of the papers reviewed was “The Effects of Education Quality on Income Growth and Mortality Decline”. Although the paper evaluated education’s effect on both income growth and mortality decline, solely the effect on income growth was evaluated as it pertains to the study. The paper used data from up to 62 countries, with data being collected at 10 year intervals from 1960-2000. The study found a

significant correlation between improved education and income growth rates. One test showed that one standard deviation higher test performance yielded 0.5-0.9 percent higher annual income growth rates. While the study supports the link between increased education and increased income growth, it also points out the significant impact of economy openness on bolstering this association, as more open economies show greater increases in productivity as a result of education. The paper also explains that income quality, instead of solely years of education, has a significant effect on income growth per capita. This can likely be addressed by policy changes and devising plans to improve teacher quality. The implication of this is that focuses on education quality may be more cost effective than increasing years of schooling as a means of affecting income growth per capita.

Another paper reviewed was “The impact of education on income inequality and intergenerational mobility”. The paper analyses the effects of innate ability, compulsory education (grades 1-9), and non-compulsory education (grades 10+) on income inequality. It’s discovered that investment in education has a significant impact on income inequality, with the gap growing as level of education increases. It was concluded that investment in early education is the most important driver of income inequality. This difference is revealed by the difference in innate ability being 1.36 between the bottom and top quantiles, with the difference increasing to 2.35 and 2.89 at the end of compulsory and non-compulsory education, respectively. The paper suggests that the ability of high-income families to invest more into early childhood education serves to widen the income gap. The most effective proposed solution of the paper is to subsidize low-income families’ investment in early education. This would mitigate budget concerns of low-income families and serve to reduce the innate ability gap created by different levels of investment in education by families, and especially in early education. By reducing this innate ability gap, income inequality would presumably be reduced, or at the least the rate of increase of income inequality would decrease.

“The impact of education on household income and expenditure inequality” examined the distribution of returns to education at the household level using a representative household survey in Portugal. The estimated coefficients for the model measured the impact of number of years of education of the household’s head on the logarithm of income. This revealed log values below 0.07 at the low end, up to 0.095 at the high end, which is statistically significant. The paper also examined a scenario in which years of education was replaced by discrete levels of educations, here defined as 0 years, 4 years, 6 years, 9 years, 12 years and university degree. This breakdown revealed that income distribution is driven mostly by the upper education levels. The conclusion of the paper is that education contributes to increase income, as well as income inequality in Portugal. This article is limited in that it only draws on data from

Portugal, but lends insight that can be compared with other nations and presents an avenue for further study and evaluation.

The final paper evaluated is titled “Does the Attained Level of Education Affect the Income Situation of Households?” Data for the analysis is based on a survey conducted between 2005-2009 of households in the Czech Republic. The level of education was based on the level obtained by the household member with the highest income. Levels were defined as primary/no education, lower secondary education, full secondary education, and higher (tertiary) education. The study found that the most vulnerable group to poverty comprises households with primary or no education-the lowest education bracket. Interestingly, the study concludes that higher education level of the head of household is no guarantee of a lower risk of poverty. This was because when the study examined disposable income per household it found that households headed by a person with the highest level of education had the lowest disposable income. This is somewhat surprising, however may be unique to the Czech Republic as perhaps university graduates are not as easily implemented or fully utilized in the marketplace. The paper concludes by saying that higher education level generally leads to a better situation in terms of risk of poverty, however this is unless they lose their job or stable source of income. In this case they are more significantly impacted than households with lower level of education, likely because of higher spending habits, the difficulty in finding a new job, as well as the failure to qualify for social benefits.

This paper contributes to published literature in several ways. Firstly, this paper examines United States population data. This differs from the literature reviewed in which one focused on a multitude of countries, with the others focusing on China, Portugal, and the Czech Republic, respectively. The United States has the world’s highest nominal GDP and it will be interesting to see how education relates to income in such a large and developed economy. Another difference is that the paper written will focus solely on education level’s impact on income level, whereas the articles examined often focused on additional factors along with income level, or income inequality as opposed to income. This is distinctive because although the papers had a lot of interesting implications for education’s impact on income level, many didn’t provide direct evidence or establish a correlation between education level and income level.

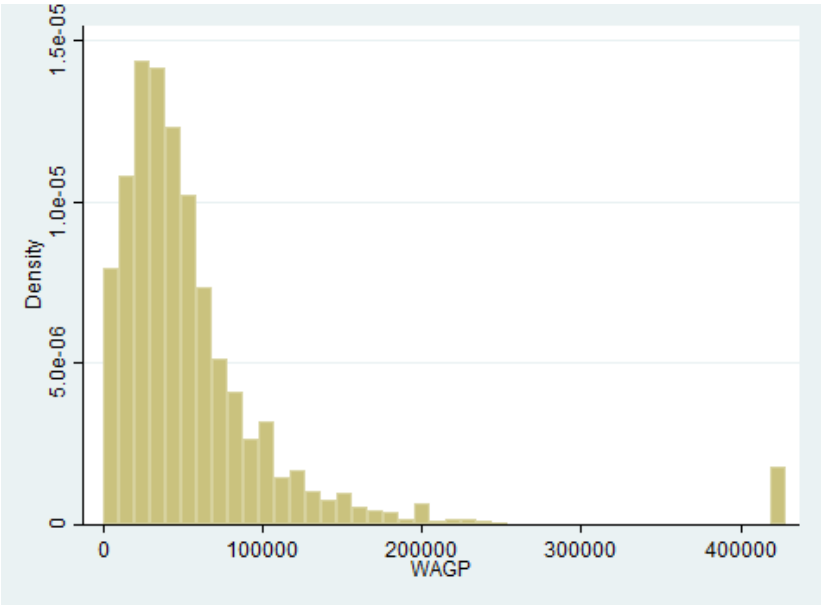
3. Data

In order to complete this analysis, different variables that we believed would influence a person's wage were included. With a person’s income/wage acting as the dependant variable, there are plenty of factors which could possibly affect a person’s income so obviously not all were included. Some of the

more important factors that we looked were a person's years of education, their years of experience, age, weeks worked in the past year, english proficiency, and if a person is disabled. We made the main focus on a person's years of education as we believed that that would be one of the more impactful variables on a person's income. Because the original data on education was categorical by degree, some liberties were taken and the degrees were converted into years of education by taking the average amount of years it takes in order to complete said degree; so 2 years for an associates degree, 4 years for a bachelors, etc. Another variable that we looked at was a person's years of work experience. This is another factor that we believed would influence a person's income that they receive. Because the experience variable was not included in the raw data set, we have attempted to approximate the years of experience a person had using a person's age and their years of education giving us an idea of the time that they were out of school. For the purposes of this paper we are making the assumption that most of these people spent their time out of school working and building up experience. Age was also an important factor due to its relation with experience and education; for this variable ages below 18 and greater than 65 were excluded as we took the assumption that most people below 18 were either not working or only had small part time jobs due to still being in high school, and many people over 65 have retired. Another independent variable we looked at was the number of weeks worked in the past year, giving us people who were working all year round and not doing seasonal work or had to take time off due to extreme circumstances. English proficiency was also looked at to see how the regression changed accounting for people who are not fluent in English due to the fat that it may be more difficult for them to find steady work. The final independent variable we looked at showed whether or not a person was disabled, as we believe that a person's disability could easily have an effect on their potential wage.

2. What is the source of data?

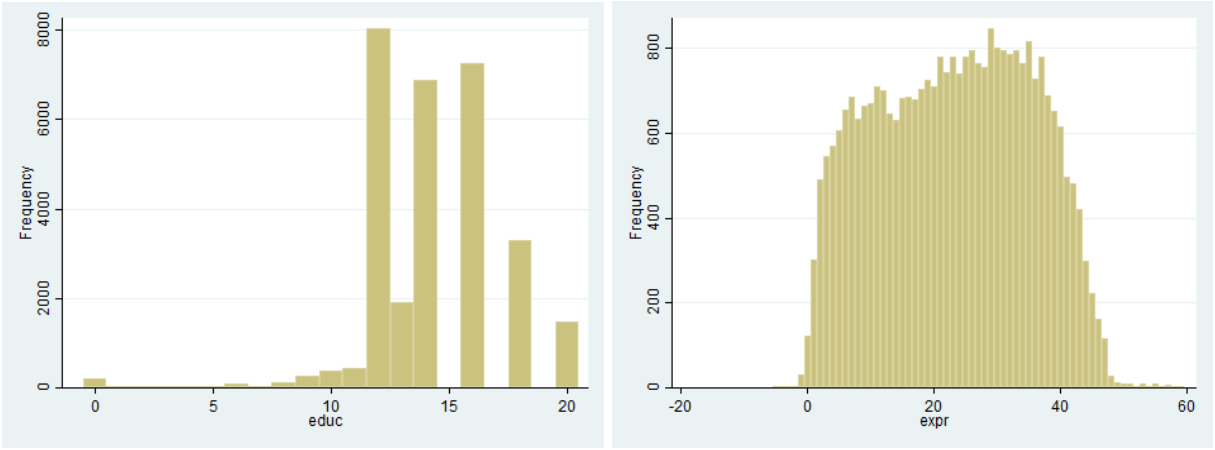
The data was mostly taken from Public Use Microdata Sample files obtained from the American Community Survey. The American Community Survey is a survey conducted by the US Census Bureau that takes data from a sample of households across the country and looks at various variables including a person's age, weight, marital status, employment status, etc. The data used was taken from the 2016 year's data and we decided to only use data from the state of Georgia. We decided to limit our area of data to a single state due to differences in education that can exist across the country, so a more localized data set will give us a better estimate. The use of data from 2016 also allows us to view more relevant data than that from older years.

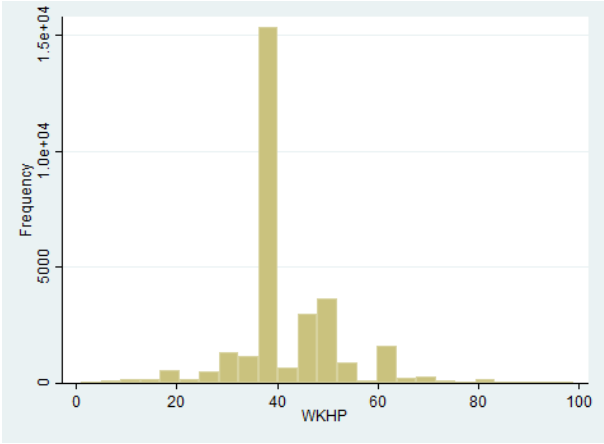


The distribution of incomes of individuals in the State of Georgia that satisfy the following criteria: (1) aged 18-65; (2) not having attended school in the last 3 months; (3) proficient in English; (4) working a normal amount of weeks (48-52 weeks) per year; and (5) not disabled, is as demonstrated on the histogram to the left. The mean annual wages is $57007.93, with a standard deviation of $63240.27.

The quantitative independent variables that are selected as regressors are education, experience, and hours worked per week. The table below displays the summary statistics associated with these variables and the histograms below show the distribution of said variables of individuals per the restrictions of the scope of the study.

Variable	Mean	Std. Dev.	Min	Max
annual wage income	57007.93	63240.27	0	429000
years of education	14.31052	2.80816	0	20
years of experience	23.13641	12.22362	-5	59
hours of work per week	42.61362	10.30011	1	99





The linear parameter assumption is satisfied by simply using STATA as the analytical tool, and the random sampling assumption is satisfied by the methodology of the American Community Survey. The no perfect collinearity assumption can be verified by performing linear regression between the regressors, and in each case R^2 value is smaller than 0.05 (0.0406 between education and experience, 0 between experience and weekly hours of work, and 0.0113 between education and weekly hours of work). Somewhat predictably, the sample data cannot be assumed to be exhibit homoskedasticity, since the standard deviation of income largely increases by years of education ($28324.1 for 0 years, $115538.5 for 20 years/doctorate or professional degree).

4. Results

i. Simple Regression

$$Annual\ Wage\ Income = -55032.62 + 7829.315 * Years\ Of\ Education$$

Specific data about the regression are shown below:

Source	SS	df	MS	Number of obs =		30,372
				F(1, 30370) =		4175.36
Model	1.47E+13	1	1.47E+13	Prob > F =		0
Residual	1.07E+14	30,370	3.52E+09	0R-squared =		0.1209
				Adj R-squared =		0.1208
Total	1.21E+14	30,371	4.00E+09	Root MSE =		59296
wagp	Coef.	Std. Err.	t	P>t	[95% Conf.Interval]	
educ	7829.315	121.165	64.62	0.0000	7591.827	8066.804
_cons	-55033.62	1767	-31.15	0.0000	-58497.01	-51570.22

ii. Multiple Regression

$$\textit{Annual Wage Income} = -137224 + 7940.944\ \textit{Years of Education} + 778.862\ \textit{Years of Experience} + 1468.372\ \textit{Hours Worked Per Week}$$

Source	SS	df	MS	Number of obs =		30,372
				F(3, 30368) =		2544.25
Model	2.44E+13	3	8.13E+12	Prob > F =		0
Residual	9.71E+13	30,368	3.20E+09	R-squared =		0.2009
				Adj R-squared =		0.2008
Total	1.21E+14	30,371	4.00E+09	Root MSE =		56536
wagp	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
educ	7940.944	118.6486	66.93	0.000	7708.388	8173.5

expr	778.862	27.10354	28.74	0.000	725.738	831.9861
wkhp	1468.378	31.68415	46.34	0.000	1406.275	1530.48
_cons	-137224	2257.583	-60.78	0.000	-141649	-132799.1

iii. Analysis

Both the simple and multiple regression equations show a very large positive coefficient on wage income, meaning that an increase in 1 year of education can have on average, an increase of more than $7800/year of wage income based on both equations. From the R^2 value of the simple regression, we see that years of education alone can explain 12.09% of wage income variation within the sample. Both equations have a very large negative intercept because the vast majority of individuals have 9 or more years of education.

By comparing the differences of simple and multiple regression, we can see that years of experience and hours worked per week also both have a significant positive coefficient on wage income. More specifically, increasing 1 year of work experience increases the wage income on average by $778.86/year, and increasing hours of work per week by 1 increases the wage income on average by $1468.38/year.

The coefficient on years of education increased by about 110, which is a result of negative bias when ignoring years of experience, a variable which has a negative coefficient as a regressor of years of education ($\delta_1 < 0$, as shown in equation below),

$$Years\ of\ Experience = 35.68884 - 0.8771472\ Years\ of\ Education$$

a positive coefficient as a regressor of wage income ($\beta_2 > 0$).

The negative intercept is even larger for the multiple regression can be explained by the fact that people who are working for a wage on average work significantly more than 0 hours per week.

iv. Statistical Inference

The t-values obtained for all variables in both our simple and multiple regression models were found to be statistically significant in all tests conducted from a 10% level of significance to a 1% level of significance. The large t-values and the statistical significance given from them in the regression models give a similar effect to the p-values obtained where the show significance at all levels due to the low

p-values we obtained from the model. A similar story is told from the confidence intervals that were calculated at a 95% significance level. Here we can see how each variable and its respective confidence interval significantly differs from zero showing a large effect on a person's income. With the results from each of these tests it is safe to say that our independent variables of Education, Experience, and Hours Worked per Week, have a large significant effect on the dependent variable, a person's income.

5. Extension

i. Robustness

For the simple regression, the regression is overall significant at a 1% level, at an overall F-statistic of 4175.36 (critical F(1, 30370)). The multiple regression model is also overall significant at a 1% level with an F- statistic of 2544.25 (critical F(3, 30368)). The 1% significance F-statistic values for the simple and multiple regression models are 6.63 and 3.78, respectively. As 4175.36 is significantly greater than 6.63, and 2544.25 is significantly greater than 3.78, the null hypothesis is rejected and the explanatory variables are statistically significant in both the simple regression model, and in the multiple regression model in which the three variables are jointly significant.

ii. Functional Form

One of the functional forms that was briefly contemplated for this analysis was changing the annual wage into logarithmic form. The logarithmic wage is constructed by the natural log of (1+wage). However, the multiple regression model without dummy variables gives a poor R^2 compared to the linear model even though it does give a positive coefficient for a model where negative income is impossible and has coefficients with statistically significant t-values. The multiple regression model with logarithmic wage is shown below.

Source	SS	df	MS	Number of obs =	30,372
				F(3, 30368) =	644.11
Model	9663.1475	3	3221.04917	Prob > F =	0.0000

Residual	151864.041	30,368	5.00079164	R-squared =	0.0598
				Adj R-squared =	0.0597
Total	161527.188	30,371	5.31846788	Root MSE =	2.2362

lwagp	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
educ	.1404714	.004693	29.93	0.000	.1312728	.1496699
expr	.0038309	.0010721	3.57	0.000	.0017296	.0059321
wkhp	.0359062	.0012532	28.65	0.000	.0334498	.0383626
_cons	6.56585	.0892968	73.53	0.000	6.390825	6.740876

iii. Dummy Variables

Categorical variables that could have a plausible effect on wage income were selected from the data source - sex, marital status, and class of worker. It was also hypothesized that marriage would have different effects on wages for the sexes and therefore we consider these categories jointly rather than separately. Three binary dummy variables were created - married male, married female, and unmarried female, so that unmarried males are the control group. The multiple regression model with these additional dummy variables is shown below.

Source	SS	df	MS	Number of obs =	30,372
				F(6, 30365)=	1522.89
Model	2.8096e+13	6	4.6826e+12	Prob > F =	0.0000

Residual	9.3368e+13	30,365	3.0749e+09	R-squared =		0.2313
				Adj R-squared =		0.2312
Total	1.2146e+14	30,371	3.9993e+09	Root MSE =		55451
wagp	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
educ	7965.394	118.7393	67.08	0.000	7732.66	8198.128
expr	659.794	27.49553	24.00	0.000	605.9016	713.6864
wkhp	1204.968	32.10091	37.54	0.000	1142.049	1267.888
malemarr	19052.12	955.0461	19.95	0.000	17180.19	20924.05
femmarr	-4272.568	1011.003	-4.23	0.000	-6254.177	-2290.959
femunmarr	-8216.79	1044.141	-7.87	0.000	-10263.35	-6170.23
_cons	-127546.6	2267.48	-56.25	0.000	-131990.9	-123102.2

The other categorical distinction is whether an individual is employed by a private company, by a government institution, or self-employed. For this category, we created two binary variables - government worker and self-employed, so that the private company employee is the control group. The multiple regression model with only these two dummy variables added is as shown below:

Source	SS	df	MS	Number of obs =	30,372
				F(5, 30366) =	1742.42
Model	2.7079e+13	5	5.4158e+12	Prob > F =	0.0000

Residual	9.4385e+13	30,366	3.1082e+09	R-squared =		0.2229
				Adj R-squared =		0.2228
Total	1.2146e+14	30,371	3.9993e+09	Root MSE =		55752
wagp	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
educ	8431.71	118.4498	71.18	0.000	8199.543	8663.876
expr	858.6207	26.89847	31.92	0.000	805.8986	911.3429
wkhp	1488.531	31.25205	47.63	0.000	1427.276	1549.786
gov	-20428.78	878.5019	-23.25	0.000	-22150.68	-18706.88
self	-23591.85	1106.993	-21.31	0.000	-25761.6	-21422.09
_cons	-141291.4	2231.372	-63.32	0.000	-145665	-136917.9

If all the dummy variables are incorporated into the model, it would be as follows:

Source	SS	df	MS	Number of obs =	30,372
				F(8, 30363) =	1292.85
Model	3.0862e+13	8	3.8578e+12	Prob > F =	0.0000
Residual	9.0601e+13	30,363	2.9839e+09	R-squared =	0.2541
				Adj R-squared =	0.2539

Total	1.2146e+14	30,371	3.9993e+09	Root MSE =	54625
wagp	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]
educ	8419.274	118.221	71.22	0.000	8187.556	8650.993
expr	738.3745	27.22559	27.12	0.000	685.0112	791.7378
wkhp	1223.646	31.63785	38.68	0.000	1161.634	1285.657
malemarr	19629.2	941.2798	20.85	0.000	17784.25	21474.15
femmarr	-3623.602	997.4321	-3.63	0.000	-5578.611	-1668.593
femunmarr	-8364.314	1029.576	-8.12	0.000	-10382.33	-6346.302
gov	-18919.49	862.7546	-21.93	0.000	-20610.52	-17228.45
self	-26276.48	1087.373	-24.17	0.000	-28407.78	-24145.19
_cons	-131337.3	2239.845	-58.64	0.000	-135727.4	-126947.1

The robustness tests conducted for the regression models with dummy variables included were shown to be jointly significant. In the case of the multiple regression model including dummy variables for married male, married female, and unmarried female, with unmarried males as the control group had an F-statistic of 1522.89. This is significantly greater than $F(6, 30365) = 2.80$. The other dummy variable model included government worker and self-employed dummy variables, such that private company employee was the control group. This regression model had an F-statistic of 1742.42. This is significantly greater than $F(5, 30366) = 3.02$. This shows that the inclusion of these dummy variables have joint significance. Lastly, all of the dummy variables mentioned previously were added to the multiple regression model. This model had an F-statistic of 1292.85, which is significantly greater than F(8,

30363) = 2.51. This shows that all of the created dummy variables have joint significance and play a role in wage earnings that deserves consideration.

6. Conclusion

In debating whether or not a person's years of schooling would have an effect on their yearly income, we found the results to clearly dictate the significance that education has on income. We can see how only one extra year of education can potentially increase a person's income by over $8000 showing the importance of earning a good education. However, it is not as clear cut as it may seem, and as one would expect a person's education and hours worked per week also play a role in income earnings. These help confirm many common and well known assumptions that are seen in today's society. Another common assumption that is a major talking point in many circles is the accusation of female wage earnings being lower than men's. With the inclusion of our dummy variables we can see that in both the cases of married and unmarried women, they tend to have a lower income than both married and unmarried men; however, whether this is due to sexism or other factors, such as the industry in which they work, cannot be evaluated through our regression. While dealing with sectors, we can see how on average people tend to earn the most in the private sector of employment, while self-employed people tend to earn lower wages than both private sector and government employees on average. Although, even with all these factors a person's education still seems to have the most significant effect on income.

This relationship does lead to one major issue, the problem of opportunity cost. Even though one could earn over $8000 by going to school for another year, that is potentially one year without wages, or working only part time. There then arises the cost of schooling where people will spend well over a eight thousand dollars to go to school for one more year. This dilemma can cause many issues both on the personal level, and the federal level when debating what the best choice to make is, choose one more year of work, or one more year of school.

References

Jamison, E. A., Jamison, D. T., & Hanushek, E. A. (2006). *The Effects of Education Quality on Income Growth and Mortality Decline*. Retrieved from http.//www.nber.org/papers/w12652.

Yang, J., Qiu, M. (2015). The impact of education on income inequality and intergenerational mobility. *China Economic Review, 37*, 110-125

Nuno, A. (2012). The impact of education on household income and expenditure inequality.*Applied Economics Letters, 19*(10), 915-919. doi: 10. 1080/13504851.2011.607125

Turčínková, J., Stávková, J. (2012). Does the Attained Level of Education Affect the Income Situation of Households? *Procedia-Social and Behavioral Sciences, 55*, 1036-1042