Effect of Education and Experience on Wages

Tao Wu

August 9, 2023
Professor Salma Asif
ECON 382: Introduction to Econometrics

1 Abstract

This paper sets out to investigate the degree of correlation to which changes in education and work experience affect the changes in workers' hourly wages in New York State. The scope of this project includes working adults aged 23-40 who concluded their academic journey and were selected from *Current Population Survey* of New York State residents regarding wage income, years of formal education, and other factors. By applying econometric techniques, we analyze the impact of education and work experience on wages while controlling for other relevant factors such as age and smoker status. The findings from this study will contribute to our understanding of economic advancement and its crucial components, as well as shed light on enabling individuals to realize their economic potential as productive members of society.

2 Introduction

The ancient Greeks believed that virtues, wisdom, and happiness are acquired through education, imparting a sense of purpose on the learners. In today's information age, education remains an indispensable cornerstone of one's well-lived life. Moreover, it is seen and recognized as the single most important determinant of people's economic outcomes in the modern world. Leading to better job opportunities and higher wages, education, along with work experience, will be the focus of this study and examined through its impact on wage growth in the state of New York. Thus, this project aims to explore and understand the relationship between education, work experience, and wages in this context and to analyze other closely relevant factors that help to define the ceteris paribus effects of each variable in the models, which is constructive for informed policymaking and guiding individuals' decisions regarding their educational pursuits.

3 Literature Review

The existing literature has extensively probed into the relationship between education and wages, with a large number of studies providing research evidence of a positive correlation between educational attainment and earnings. Three pieces of literature included here are germane to our study. "Schooling and work experience account for about half of the variation in earnings, wage trends of the workforce as a whole have been influenced by changes in its composition, " Kosters wrote. [Kosters(1990)][Dickson(2013)] The author's finding not only enhances our intuition about education boosting earnings but also introduces us another factor of work experience, though correlated with education, that is necessary to be included for our econometric analysis.

In a separate paper titled, "Human Capital vs. Signalling Explanations of Wages", the author write, "Better-educated workers are not a random sample of workers: they have lower propensities to quit or to be absent, are less likely to smoke, drink or use illicit drugs, and are generally healthier." [Weiss(1995)] This piece is centered on human capital formation, personal attributes considered useful in the production process, and its various implications. Nevertheless, when it comes to wages, it is sensible to consider personal characteristics in our analysis as well, namely years of age and smoking history, especially with the supportive results of another paper titled "The Obese Smoker's Wage Penalty." Essentially, its writers conclude the inverse relationship between smoking and wages based on the comprehensive data he gathered.[Baum et al.(2006)Baum, Ford, and Hopper]

4 Data

Data for our study comes from the American Community Survey, a premier source for information about America's changing population, housing, and workforce, hosted by the United States Census Bureau, one of our country's recognized providers of quality data about its people and economy. In order to perform an in-depth analysis of wages, we set our sights on multiple aforementioned factors such as education levels, work experience, age, and smoking history to be included as independent variables in our econometric model. Despite tremendous efforts, it was difficult for us to find quantitative data on education levels. Our thinking behind this is that circumstances differ widely from person to person. Categorical data that represent educational attainment would not be accurate enough to indicate the correlation between wages and education. Plus, some people drop out of school without finishing their degrees and the already acquired knowledge they carry could still play a role in their future income. With the help of peers that study computer science, we are able to scrape the internet and find a survey that suits our needs and contains hourly wages information as well as data on education levels collected on New York State residents aged between 23 and 40 with years of experience in the workforce, who self-identify as smokers or nonsmokers.

5 Econometric Model

For a population of people in the workforce of New York State, let y (dependent variable) = wage, where wage is measured in dollars per hour; Let x1 (independent variable) = educ denote years of formal education; Let x2(independent variable) = age denote ages between 23 and 40; Let x3(independent variable) = exper denote years of work experience; let x4(independent dummy variable) = smokr denote the binary state of smoker status.

5.1 Simple Regression

$$wage = a + \beta educ + u$$

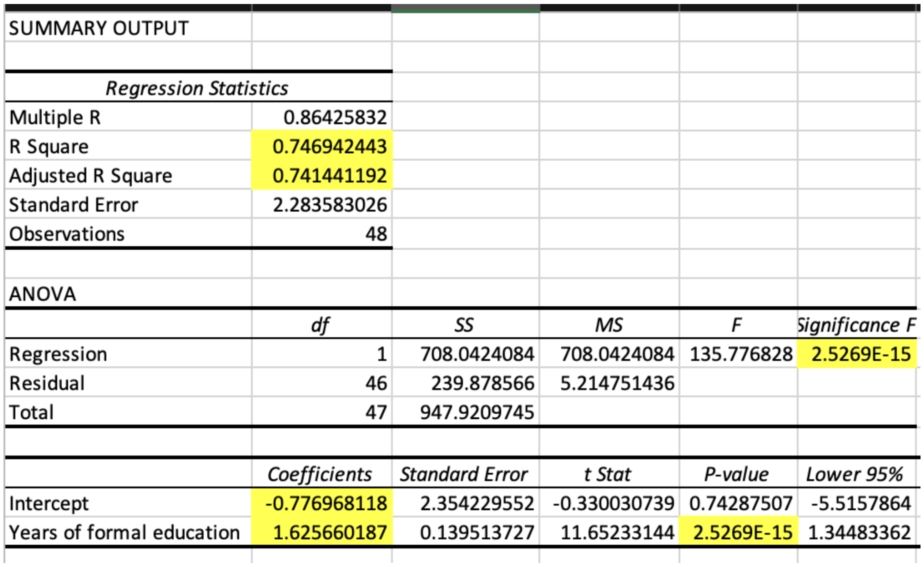
where,

wage = hourly wage of an individual
educ = number of years of formal education completed by the individual
a is the intercept
β is the slope parameter associated with educ.
u = disturbance (error term)

By the performance of simple linear regression analysis, we obtain the simple linear regression model

$$\widehat{wage} = -0.777 + 1.626educ + u$$

Specific data about the regression are shown below:



SUMMARY OUTPUT					
Regression Statistics					
Multiple R	0.86425832				
R Square	0.746942443				
Adjusted R Square	0.741441192				
Standard Error	2.283583026				
Observations	48				
ANOVA					
	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	708.0424084	708.0424084	135.776828	2.5269E-15
Residual	46	239.878566	5.214751436		
Total	47	947.9209745			
	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*
Intercept	-0.776968118	2.354229552	-0.330030739	0.74287507	-5.5157864
Years of formal education	1.625660187	0.139513727	11.65233144	2.5269E-15	1.34483362

figure 1

5.2 Multiple Regression

$$wage = a + \beta_1 educ + \beta_2 exper + \beta_3 age + \beta_4 smokr + u$$

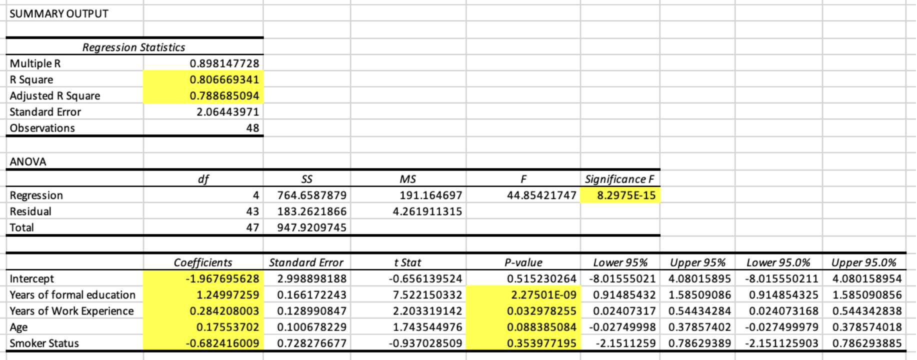
where,

wage = hourly wage of an individual
educ = number of years of formal education completed by the individual
exper = years of work experience
age = age of the individual
smokr = dummy variable for smoker status (1 for smoker, 0 for nonsmoker)
a is the intercept
β_1 is the slope parameter associated with educ.
β_2 is the slope parameter associated with exper,
β_3 is the slope parameter associated with age,
β_4 is the slope parameter associated with smokr,
u = disturbance (error term)

By the performance of multiple linear regression analysis, we obtain the multiple linear regression model

$$\widehat{wage} = -1.977 + 1.250 educ + 0.284 exper + 0.175 age - 0.682 smokr + u$$

Specific data about the regression are shown below:



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.898147728
R Square	0.806669341
Adjusted R Square	0.788685094
Standard Error	2.06443971
Observations	48

ANOVA

	df	SS	MS	F	Significance F
Regression	4	764.6587879	191.164697	44.85421747	8.2975E-15
Residual	43	183.2621866	4.261911315		
Total	47	947.9209745			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.967695628	2.998898188	-0.656139524	0.515230264	-8.01555021	4.08015895	-8.015550211	4.080158954
Years of formal education	1.24997259	0.166172243	7.522150332	2.27501E-09	0.91485432	1.58509086	0.914854325	1.585090856
Years of Work Experience	0.284208003	0.128990847	2.203319142	0.032978255	0.02407317	0.54434284	0.024073168	0.544342838
Age	0.17553702	0.100678229	1.743544976	0.088385084	-0.02749998	0.37857402	-0.027499979	0.378574018
Smoker Status	-0.682416009	0.728276677	-0.937028509	0.353977195	-2.1511259	0.78629389	-2.151125903	0.786293885

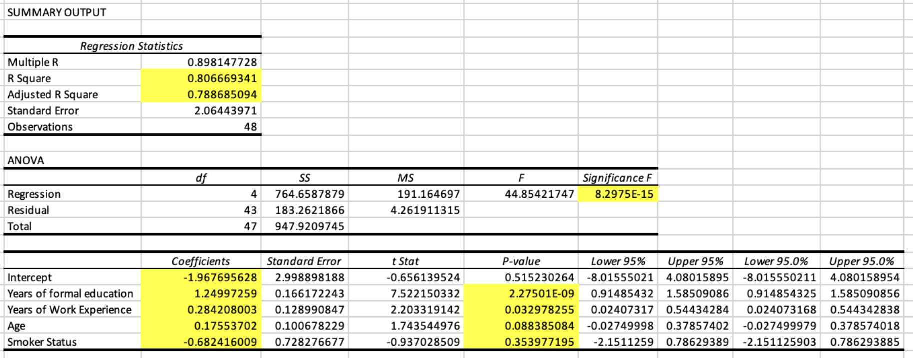
figure 2

5.3 Empirical Methodology

Results from the wage regressions on the many independent variables we picked out speak volumes about their statistical association with wages. The coefficient of linear correlation between wage and educ approximates to be near positive 1. The coefficient of linear correlation between wage and expert approximates to a relatively high positive degree. The coefficient of linear correlation between wage and age approximates to a moderate positive degree. Lastly and not too surprisingly, when we regress wage on smokr, a minute inverse correlation, namely negative coefficient r, suggests two variables move in opposite directions from one another. Additionally, to avoid highly correlated independent variables with R^2 exceeding roughly 0.6, we test to see if all the independent variables should remain untouched in the model, which, to our satisfaction, their respective R^2 stay below 0.45. Since the preliminary steps are taken to ensure a more credible and accurate predictive model, we can now move on to the analysis stage where we analyze the metrics that are most relevant to our study objectives.

5.4 Analysis

We begin our analysis by examining the summary statistics in our Excel sheet.



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.898147728
R Square	0.806669341
Adjusted R Square	0.788685094
Standard Error	2.06443971
Observations	48

ANOVA

	df	SS	MS	F	Significance F
Regression	4	764.6587879	191.164697	44.85421747	8.2975E-15
Residual	43	183.2621866	4.261911315		
Total	47	947.9209745			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.967695628	2.998898188	-0.656139524	0.515230264	-8.01555021	4.08015895	-8.015550211	4.080158954
Years of formal education	1.24997259	0.166172243	7.522150332	2.27501E-09	0.91485432	1.58509086	0.914854325	1.585090856
Years of Work Experience	0.284208003	0.128990847	2.203319142	0.032978255	0.02407317	0.54434284	0.024073168	0.544342838
Age	0.17553702	0.100678229	1.743544976	0.088385084	-0.02749998	0.37857402	-0.027499979	0.378574018
Smoker Status	-0.682416009	0.728276677	-0.937028509	0.353977195	-2.1511259	0.78629389	-2.151125903	0.786293885

figure 3

First, we look at a statistical measurement called the coefficient of determination which indicates how differences or variations in the dependent variable can be explained by the difference in the independent variable(s) when predicting the outcome of a given event. However, R^2 can only go up and never come down with the increased number of regressors. A similar but not identical measurement called "adjusted r squared" is a better statistical number for us to reply on in terms of the overall fitness of the model. In our case, both R^2 and adjust R^2 turn out to be 0.8 and 0.78, high enough for us to trust the explanatory power of the independent variables in our multiple linear regression model.

Secondly, the F-statistic along with the corresponding p-value for that F-statistic is of great importance as it tells us that our model is statistically significant for predicting wages. (let's take a look again at our multiple regression model)

$$\widehat{wage} = -1.977 + 1.250educ + 0.284exper + 0.175age - 0.682smokr + u$$

$$n = 48, R^2 = 0.8067$$

In other words, our model is very likely to get the results right most of the time. With such a small probability of the F-statistic, we rest assured that our model is well constructed and good for result predictions.

Next up, we determine the relationship between the dependent variable and each independent variable by looking at the P-value. some might prefer looking at the t-test which essentially works the same way for our purpose. For instance, we first examine the education variable by laying out the statistics we need for the t-test.

$$t_{age} = \frac{1.2499}{0.1661} \approx 7.522$$

$$df = n - k - 1 = 48 - 4 - 1 = 43$$

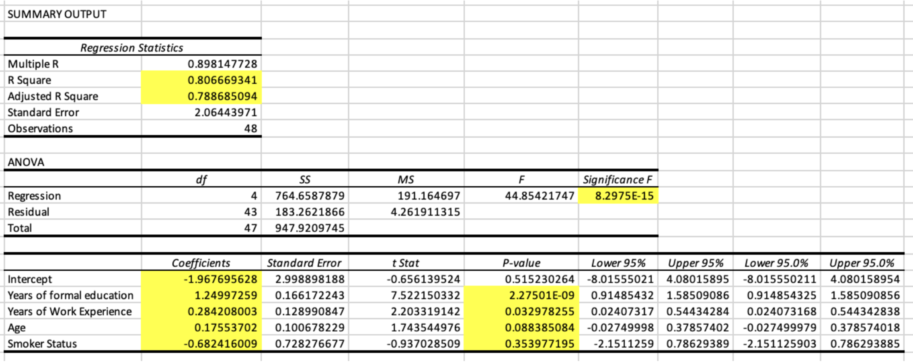
$$c_{0.05} = \pm 2.0167$$

$$c_{0.01} = \pm 2.6951$$

The effect of years of formal education on hourly wage is statistically greater than zero at the 5% (and even at the 1%) significance level.

Furthermore, the P-value represents the mathematical probability, upon hypothesis testing, that tells us whether not or to reject the null hypothesis $H_0 : \beta_j = 0$ stating each independent variable has zero correlation with the dependent variable. Changes in regressors will not be associated with changes in regressand. Simply put, when P-value is smaller than the significance level, usually set at 5% in one tail test and 2.5% in a two-tail test, then we favor the $H_a : \beta_j \neq 0$, alternative hypothesis and decide that there is sufficient evidence in the sample to conclude that a non-zero correlation exists between the 2 variables. Since we choose a 5% significant level, years of formal education, and years of work experience both pass the hypothesis testing and are worthy of staying in the model.

The last word on the variable: age. Although corresponding p-values of age and smoker status do not make the cuts of 5% significance level and therefore are deemed statistically insignificant and should be left out of the model. We want to bring to your attention that it might still be better to keep the "age" variable in the model for the sake of alleviating bias in model parameters even at the expense of increased likelihood that the coefficient estimates are further from the correct population value. Let me explain.



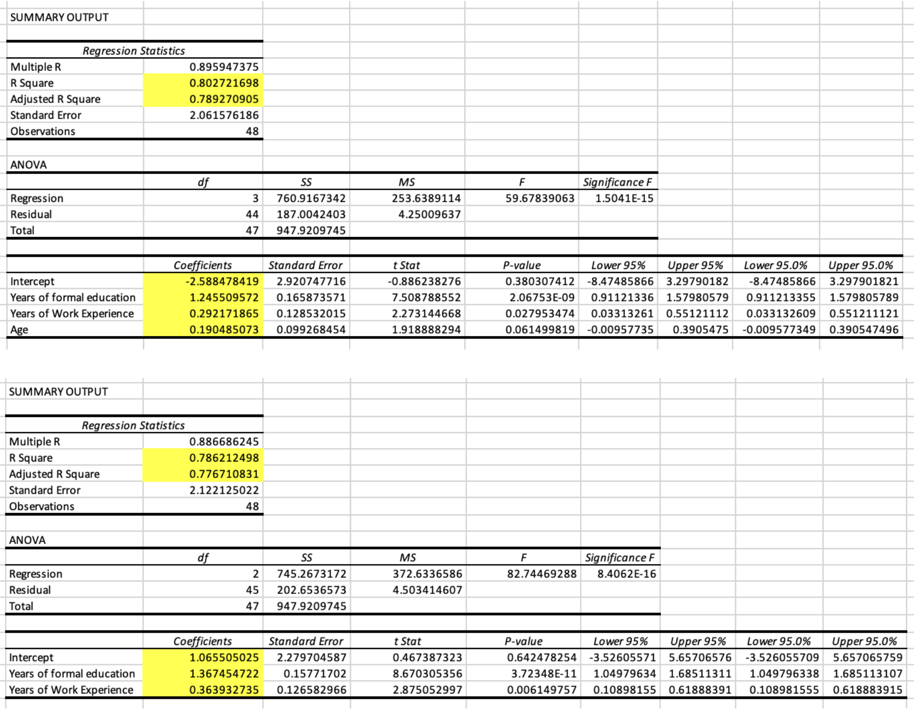
SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.898147728
R Square	0.806669341
Adjusted R Square	0.788685094
Standard Error	2.06443971
Observations	48

ANOVA

	df	SS	MS	F	Significance F
Regression	4	764.6587879	191.164697	44.85421747	8.2975E-15
Residual	43	183.2621866	4.261911315		
Total	47	947.9209745			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-1.967695628	2.998898188	-0.656139524	0.515230264	-8.01555021	4.08015895	-8.015550211	4.080158954
Years of formal education	1.24997259	0.166172243	7.522150332	2.27501E-09	0.91485432	1.58509086	0.914854325	1.585090856
Years of Work Experience	0.284208003	0.128990847	2.203319142	0.032978255	0.02407317	0.54434284	0.024073168	0.544342838
Age	0.17553702	0.100678229	1.743544976	0.088385084	-0.02749998	0.37857402	-0.027499979	0.378574018
Smoker Status	-0.682416009	0.728276677	-0.937028509	0.353977195	-2.1511259	0.78629389	-2.151125903	0.786293885



SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.895947375
R Square	0.802721698
Adjusted R Square	0.789270905
Standard Error	2.061576186
Observations	48

ANOVA

	df	SS	MS	F	Significance F
Regression	3	760.9167342	253.6389114	59.67839063	1.5041E-15
Residual	44	187.0042403	4.25009637		
Total	47	947.9209745			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	-2.588478419	2.920747716	-0.886238276	0.380307412	-8.47485866	3.29790182	-8.47485866	3.297901821
Years of formal education	1.245509572	0.165873571	7.508788552	2.06753E-09	0.91121336	1.57980579	0.911213355	1.579805789
Years of Work Experience	0.292171865	0.128532015	2.273144668	0.027953474	0.03313261	0.55121112	0.033132609	0.551211121
Age	0.190485073	0.099268454	1.918888294	0.061499819	-0.00957735	0.3905475	-0.009577349	0.390547496

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.886686245
R Square	0.786212498
Adjusted R Square	0.776710831
Standard Error	2.122125022
Observations	48

ANOVA

	df	SS	MS	F	Significance F
Regression	2	745.2673172	372.6336586	82.74469288	8.4062E-16
Residual	45	202.6536573	4.503414607		
Total	47	947.9209745			

	Coefficients	Standard Error	t Stat	P-value	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	1.065505025	2.279704587	0.467387323	0.642478254	-3.52605571	5.65706576	-3.526055709	5.657065759
Years of formal education	1.367454722	0.15771702	8.670305356	3.72348E-11	1.04979634	1.68511311	1.049796338	1.685113107
Years of Work Experience	0.363932735	0.126582966	2.875052997	0.006149757	0.10898155	0.61888391	0.108981555	0.618883915

figure 4

By comparison of the three summary tables in *figure 4* above, the only difference is that each table below has one independent variable omitted. Meanwhile, we should keep an eye on the slope parameters, and coefficients of correlation r for each remaining regressor. As the categorical variable is left out, the coefficient of educ and exper remain more or less the same due to the noncorrelation between educ, exper, and smokr. A different phenomenon happens when we omit age variable. This time each coefficient of educ and expert jumps high than when age is omitted, which shows the error term is now more correlated with the remaining variables and their coefficients contain upward bias due to the high correlation between educ, exper, and age variables.

Omitted variables are a common cause of Endogeneity, a sticky problem the inclusion of the age variable as a proxy variable can help lessen. Despite the variance of the predicted y being narrowed because of the smaller number of independent variables, it is not worth the extra bias and inconsistency in all of our overestimated OLS estimators skewing the results further in our study ($Bias \Uparrow$ **versus** $Variance \downarrow$). This type of tradeoff reminds us of the importance of weighing the pros and cons in the process of model construction. Without one size fit all solution, we can only aspire to get closer to the best solution. It reflects the principle of opportunity cost in the field of Economics.

6 Results

In our collected survey of 48 cross-sectional observations of New York State residents, we found significant relationships between hourly wages and years of formal education, and years of work experience ($p < 0.05$ for each). Specifically, we found a ceteris paribus effect of 1.25$ increase (± 0.166) in hourly wage for every 1-year increase in years of formal education and a ceteris paribus effect of 0.284$ increase (± 0.129) in hourly wage for every 1-year increase in years of work experience.

7 Conclusion

Investing in education is one of the most crucial and rewarding decisions an individual, a society, or a nation can make. It yields a wide range of benefits that extend far beyond wage growth, positively impacting communities and the overall progress of a nation. In our study, we had a chance to closely examine a small cohort of New York State residents with the aim of studying the effect of education attainment, workers' experience in the workforce, age, and smoker status. We conclude that education and work experience are indeed positively correlated with wages which is more or less in line with common sense in today's society. Although age and smoking history do not associate with wages in our regression analysis, they are still meaningful factors to be considered in policymaking as well as more advanced econometric studies.

References

[Baum et al.(2006)Baum, Ford, and Hopper] Charles L Baum, William F Ford, and Jeffrey D Hopper. The obese smoker's wage penalty. *Social science quarterly*, 87(4):863–881, 2006.

[Dickson(2013)] Matt Dickson. The causal effect of education on wages revisited. *Oxford Bulletin of Economics and Statistics*, 75(4):477–498, 2013.

[Kosters(1990)] Marvin H Kosters. Schooling, work experience, and wage trends. *The American Economic Review*, 80(2):308–312, 1990.

[Weiss(1995)] Andrew Weiss. Human capital vs. signalling explanations of wages. *Journal of Economic perspectives*, 9(4):133–154, 1995.