Impact of Internet Accessibility on Earnings among Uneducated Population
An Analysis of Current Population Survey Microdata, 2017

Shriyanshi Shikha

Depauw University
December 11th, 2020

Abstract

In the world we live in today, it is impossible to imagine a world without internet access. From educational to entertainment purposes, the internet enables us to access high quality information and opens up more opportunities in life. This paper will focus on the uneducated population in the United States with 13 or less years of education in order to find whether the internet is helping an individual in career development and overcoming the lack of education. The empirical analysis concludes that having internet access at home leads to 0.4-7.3% increase in income, holding other variables constant. In 1994, there was a 10-15% increase in income with internet access at home. This signifies that the internet accessibility does not have much of an economic significance anymore.

1. Introduction

We all know how the Covid-19 pandemic has impacted the world in different ways. Now more than ever we need internet access to overcome the obstacles the pandemic has created and to continue with the normal life. Having access to the internet and owning a smartphone has made it easier to shift to a virtual world. Online courses are becoming more and more popular and now the underprivileged people who could not attend the universities for various reasons now can take courses from the best universities in the world. However, this is not accessible to everyone because this platform requires internet access. How do people who do not have access to the internet make use of such opportunities? Even though there is more than 85% of America's population that has access to the internet, there are a significant number of population with less than 13 years of education in America who do not. According to Current Population Survey data from 2011, 76.2% of non hispanic white households and 87.2% of Asian households reported Internet use at home compared to 58.3 percent of Hispanic households and 56.9 percent of black households (File 2013).

I am interested in knowing whether having accessibility to the internet has any significant effect on the earnings among the underrepresented population with less than 13 years of education. There are so many resources available online for example, social media like Facebook and Instagram has a huge platform of Digital Marketing. People who already own businesses use social media platforms to promote and expand their business which is only possible because of the internet accessibility. People can also use the resources online like online classes, certification and web marketing to compensate for their lower education. For example, the internet entrepreneur Evan Williams CEO and co-founder of Twitter. He dropped out of college to pursue his career in technology and the internet provided him an opportunity to start his own company. Even though this is an extreme example, Internet accessibility could positively affect the income of the underrepresented population.

2. Literature Review

Krueger (1993) conducted a study using the Current Population Survey (CPS) data from 1984 to 1989 to examine how much computer usage affects the earnings of a U.S worker. The study showed that the workers who used computers in their jobs earn 10 to 15 percent higher wages compared to workers who don't use computers. The cost of computing fell down dramatically during the 1980s in turn there was a tremendous increase in the use of computers at work. Since during that time, when computers and the internet were not that popular, more highly educated people tend to use computers "can account for between one-third to one-half of increase in the rate of return of education observed between 1984 and 1989 (Krueger 1993). The estimates made in the paper by Krueger suggested that the evolution of wage structure is tied to the future development of technology. Now in 2020, almost after 26 years Krueger's suggestion still seems true. Although, the evidence from prior research does not exactly answer my question of "How does the internet accessibility affect the income among the uneducated population?" but the empirical evidence gives me more insight into the topic for further research.

DiMaggio and Bonikowski (2008) used the micro data from the CPS from 2000 and 2001 and the OLS regression to show the association between internet usage and income. The study showed that there are "robustly significant positive associations between Web use and earnings growth (DiMaggio and Bonikowski, 2008). This paper partially answers my research question although the research is performed on the entire population of US workers and did not filter according to the education level, I can use the empirical analysis of the paper to answer my question. During that time women were more likely to use the internet (74 percent compared to 67 percent). The differences reflect women's advantage in at-work internet use (48 vs 41 percent) and suggests that increased workplace technology use was responsible for eliminating a gender gap that gave men an advantage during the internet's early years (Ono and Zavondy 2003). I find this data interesting and want to see how is the income affected by the internet use among the

uneducated population especially among the female. A dummy variable was created in the paper to see whether a person used internet in 2000 and 2001 or not by controlling the

following variables: age, gender, race, ethnic background, educational attainment, marital status, region and metropolitan residence, union membership, occupational category, occupation level job skills demands and industry. The study found out that "the median earner who used the internet in 2000 and 2001 was paid $.96 per hour more than a comparable nonuser" (DiMaggio and Bonikowski, 2008). This shows that there is a positive correlation between internet accessibility and earnings. However, the data was collected from 2000 to 2001 so the result might not resonate with my empirical study. It's almost been 20 years since the study was conducted so I am excited to see how the results differ and I am also very curious to see how the internet accessibility has changed the status of women in the workplace during the recent times and if there is any effect in the earnings at all.

Houngbonon and Liang (2017) took advantage of a unique town-level data to investigate the effects of broadband on the distribution of income using data from france. Using the OLS regression the estimates suggested that the broadband adoption raised the mean income and lowered the income inequality and the effects came out to be economically significant. Between 2009 and 2013 in France, there was an increase of 34% of average income by having broadband internet access and 80% of fall in the Gini index of income inequality. The study also found out that "the positive effects of broadband Internet on income distribution depends on education, measured by the number of years of schooling" (Houngbonon and Liang, 2017). The last study was the most interesting to me and I wanted to find out how the income distribution looks among the uneducated population in the United States and if there is any increase in the income or not by having the internet accessibility.

3. Theoretical Analysis

There are three different theories in how internet use can increase one's earnings: Human Capital Theory, Social Capital theory, and cultural resource and Signal Theory

(DiMaggio and Bonikowski, 2008). The result of the empirical analysis will be explained under the light of these theories.

According to Human capital theory, internet use affects both technological knowledge as well as human capital per worker as there are several obvious advantages in using the internet like faster communication, high quality information and exposure to learning opportunities. Firms may also invest in human capital when they implement new technologies by training their employees (Fernandez, 2011). The Internet has changed the way some firms produce the goods and services. For example, instead of calling the restaurant and placing the order for food delivery now you can place the order online and even track the arrival of the food using the internet. Since the extent to which one's job uses the internet depends on the job as well as the industry, we will have to control occupation and industry in the empirical analysis.

We can also explain the positive correlation between the internet use and the earnings from looking at internet use as a source of social capital. There are three kinds of social capital enhancement. First, workers who have access to the internet can use it as a means of searching for jobs. Online job search is becoming more and more popular and now you can even know the estimate of salary without even talking to the employer by looking up websites like glassdoor.com. Among those who are unemployed and looking for work, 76.3% of them utilized the internet to search for jobs in 2011 in contrast to 25.5% in 2000 (Faberman and Kudlyak, 2016). Second, workers who use the internet may be exposed to more networking opportunities. Nowadays, linkedin.com has become one of the most popular social networking sites in terms of job search. It enables people to connect with the potential employers and expand their professional network. The last enhancement explains that "employees with large, accessible professional networks may use technology to employ these networks in ways that benefit their employers: for example, getting useful information, contacting clients, or setting up collaborative ventures" (DiMaggio and Bonikowski, 2008).

The last theory talks about how cultural resources and signals can explain Internet use yields higher income. Especially a few years after the internet became accessible to

regular Americans, in early 2000's many of them "regarded the internet as the transformative force that would ignite explosive economic growth" (DiMaggio and Bonikowski, 2008). There is also evidence of employers filtering out the low-quality applicants on the internet job posting just because the people who used the internet back then were considered better for the job. The cultural resource theory describes that the familiarity with high-status activities like the internet puts a person into a favorable position in professional settings because the person can form a relationship with other high status easily compared to those who don't use the internet. However, this theory might not be true anymore since the internet is not a high-status activity anymore in 2020.

4. Empirical Analysis

i. The Data

Table 1. Summary Data Table

Variable	Obs	Mean	Std. Dev.	Min	Max
earnweek	12,618	702.0469	469.5563	1	2885
female	108,128	.4920927	.4999398	0	1
male	108,128	.5079073	.4999398	0	1
exp	108,128	22.26929	14.45173	-1	58
educYears	108,128	11.81765	1.713959	0	13
compUse	43,360	.7183579	.4498049	0	1
white	108,128	.7698468	.4209327	0	1
black	108,128	.1373372	.3442046	0	1
asian	108,128	.0451687	.2076749	0	1
hispanic	108,128	.2340929	.4234325	0	1
married	108,128	.4709788	.4991594	0	1
femaleMarr~d	108,128	.2311797	.421589	0	1
metropolitan	108,128	.7903411	.4070671	0	1
compJob	108,128	.0099974	.0994865	0	1
compInd	108,128	.0049386	.0701017	0	1

Source: Current Population Survey Computer and Internet Use Supplement Nov 2017

Before comparing the different regression results that analyze internet accessibility and income, some independent variables need to be described along with the summary data in *Table 1.* All the data were collected from the Current Population Survey (CPS), a monthly household survey fielded continually by the Bureau of Census and based on

stratified probability samples of the non-institutionalized U.S. population" (DiMaggio and Bonikowski, 2008). Since the data generation process (DGP) does not meet the requirements of Classical Econometric Model (simple random sampling), PWLS

(Probability Weighted Least Squares) regression was also used other than the OLS (Ordinary Least Squares) in this analysis. The data was derived from the latest available CPS Computer and Internet Use supplement from Nov 2017. This paper has a similar way of selecting variables as DiMaggio and Bonikowski (2008) and Krueger (1993). Only samples with age 18 to 64 were selected because this is the typical labor force in the United States. This study also omits the sample with more than 13 years of education because that is the threshold for the uneducated population. The reason for choosing the threshold will be explained further in the next section. The samples that were categorized as “Not in Universe" for internet accessibility were dropped because the data collected by CPS had an inclusion criteria that defined niu as not being part of the population. So it’s better to drop it for more accuracy in the result.

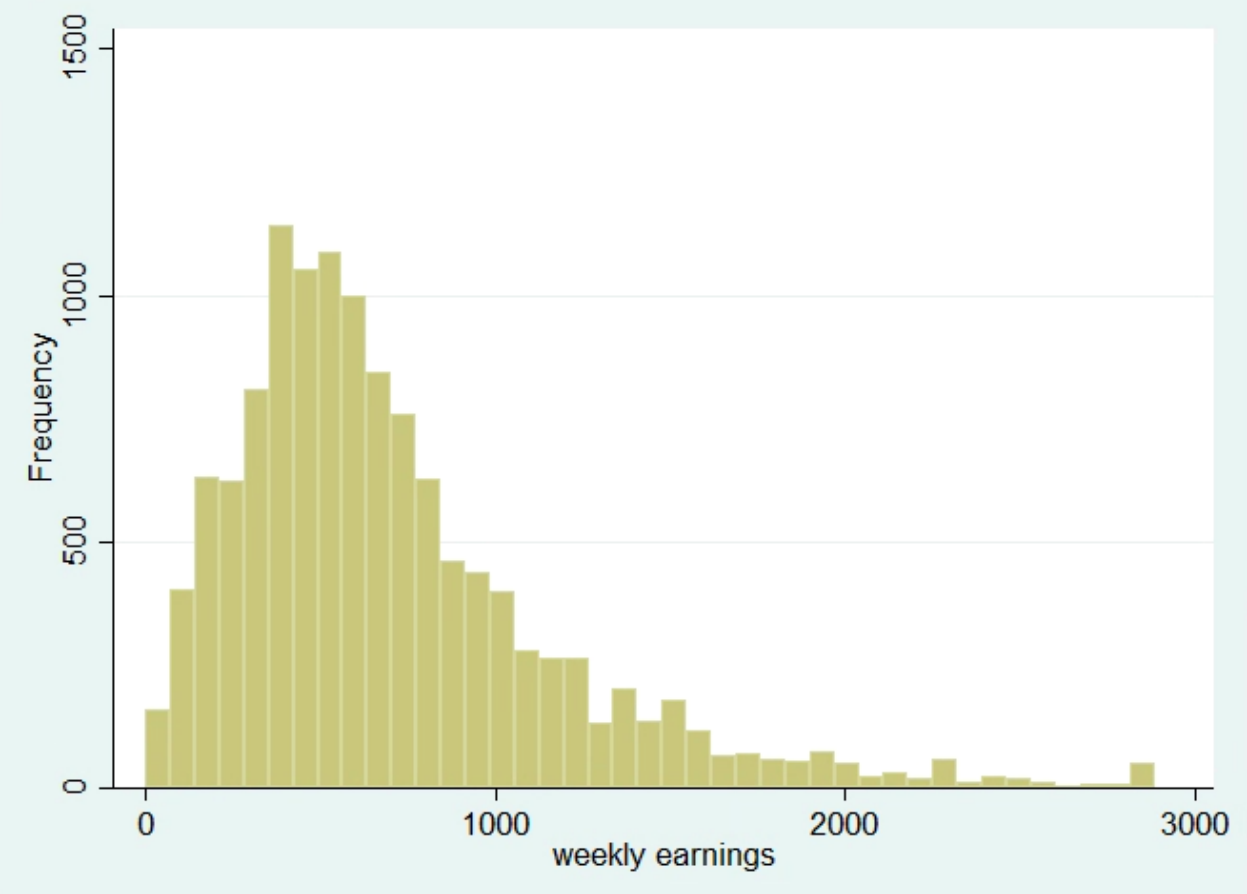
Table 1 summarizes the variable used in the regression analysis. Occupation and industry are controlled as explained above and are encoded for this table to compJob and compInd to show how much of the sample have computer or Internet related jobs or industry. By encoding compJob and compInd it was found that there only a small portion of the sample that has computer or Internet related jobs. This is reflected in the percentage of internet accessibility (compUse variable), we can see that only 71.8% of the sample used the internet at home according to the data collected in Nov 2017. As shown in *Figure 2*, the most skewed data is the years of education. The majority of the sampled individuals have a high school diploma or one year of college experience. Therefore, we can conclude that the study for this paper is heavily weighted towards high school graduates and college dropouts.

The dependent variable in this regression is Income. The natural log is applied to the income as the exponential is always while doing the growth calculation. It also reduces the heteroskedasticity in the data. Using the statistical software package STATA, the raw data was encoded by creating dummy variables that could correlate with the most important independent variable: Internet Use (compUse). The variables that are controlled are: experience, experience squared, years of education, female, race, hispanic, marital status, metro (people who live in the city), the interaction term called femaleMarried for married female, occupation and industry. The experience variable was

created by age minus schooling minus 6 (the typical age to begin school in the U.S.) and the quadratic term (experience squared) was also included (Mincer, 1958). As shown in *Figure 1*, the income vs experience relationship is curved since too little experience would yield lower income but being old or having too much experience would reduce one's earnings. The years of education also need to be controlled in the regression because even though the empirical study is focusing on the uneducated population the education still has the standard deviation greater than 1.

One of the biggest limitations in the sampled data is that the Current Population Survey for Computer and Internet Use Supplement does have any traditional annual income variable. The closest variable related to the annual income is weekly earnings called earnweek. There are two limitations of using weekly earnings instead of the annual income. One that affects the result in this study is that the weekly earnings compresses all the sampled individuals who make more than $2,885 into one category. To clear that issue, the value was replaced by (.), which means "missing value" in STATA. Dropping them is a good idea because it would allow the entire row to fit into the category in the income column. This essentially means that even if income data is not used in the regression analysis, some data won't be available because the individual earns more than $2,884 in a particular week. This is the biggest source of omitted variable bias in this data generation process. Another limitation is also an important factor and it affects my study as the weekly earnings excludes the self-employed and the non-wage/salaried employees. I wanted to see how people who form their own career without college education are affected by the internet accessibility like I mentioned above CEO of twitter, Evan Williams. But, innovative people like that are not included in the data and this inclusion might affect the significance of the paper.

Figure 1. Weekly Income Frequency Distribution among sample with 13 or less years of education

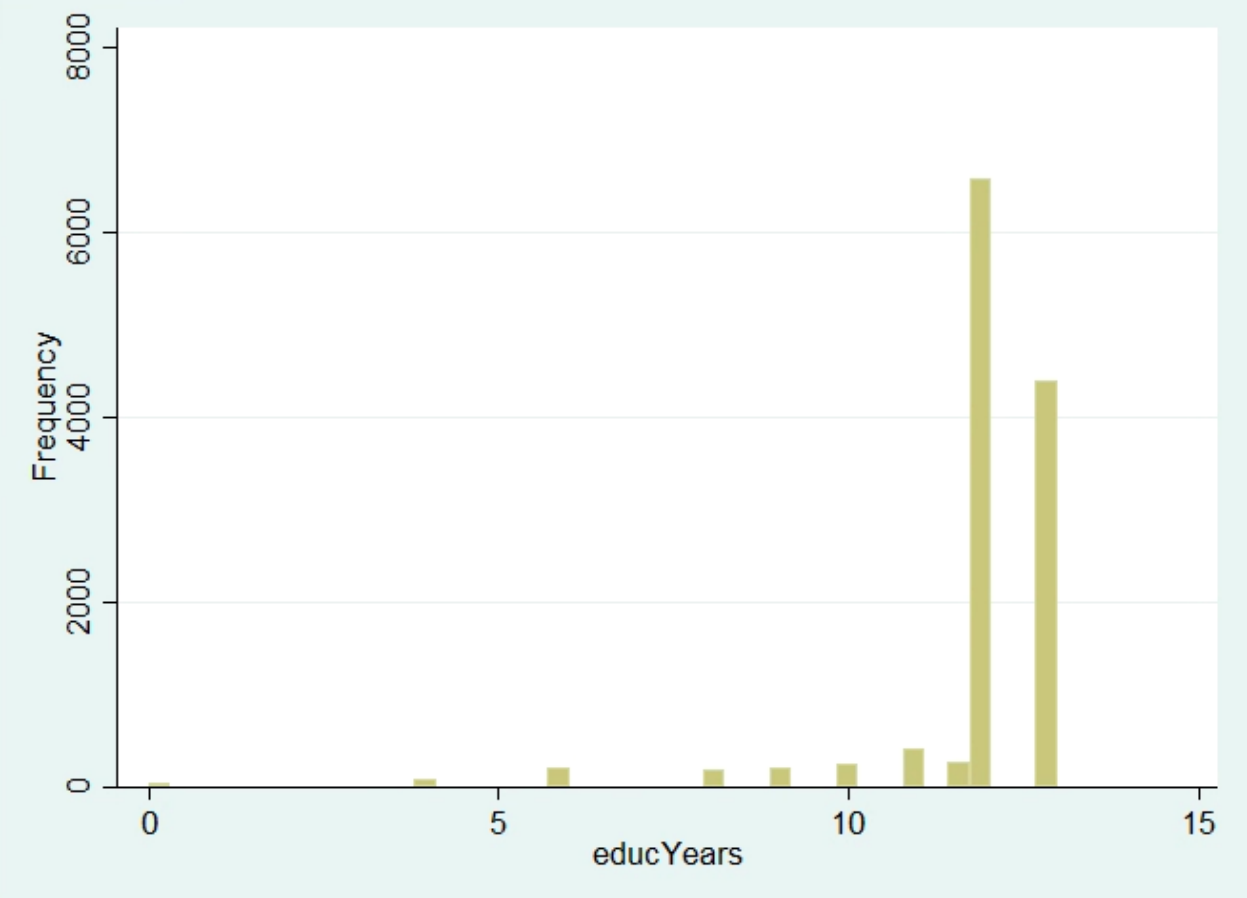


Source : Current Population Survey Computer and Internet Use Supplement Nov 2017, ages 18 to 64

Figure 1. above shows the weekly income frequency distribution in the sample. The data is skewed to the right even though the right tail is cut off at $2,884 because of the limitations in the data generation process. This is also a really important information since the income distribution is generally skewed to the right, the sample used in the regression analysis represents the population well. CPS is sometimes oversampled in some categories so this is a valid concern.

Figure 2. below shows the frequency distribution is highest for the high school graduates and college dropouts. Since the years of education still varies (SD>1), education is going to be controlled in the analysis.

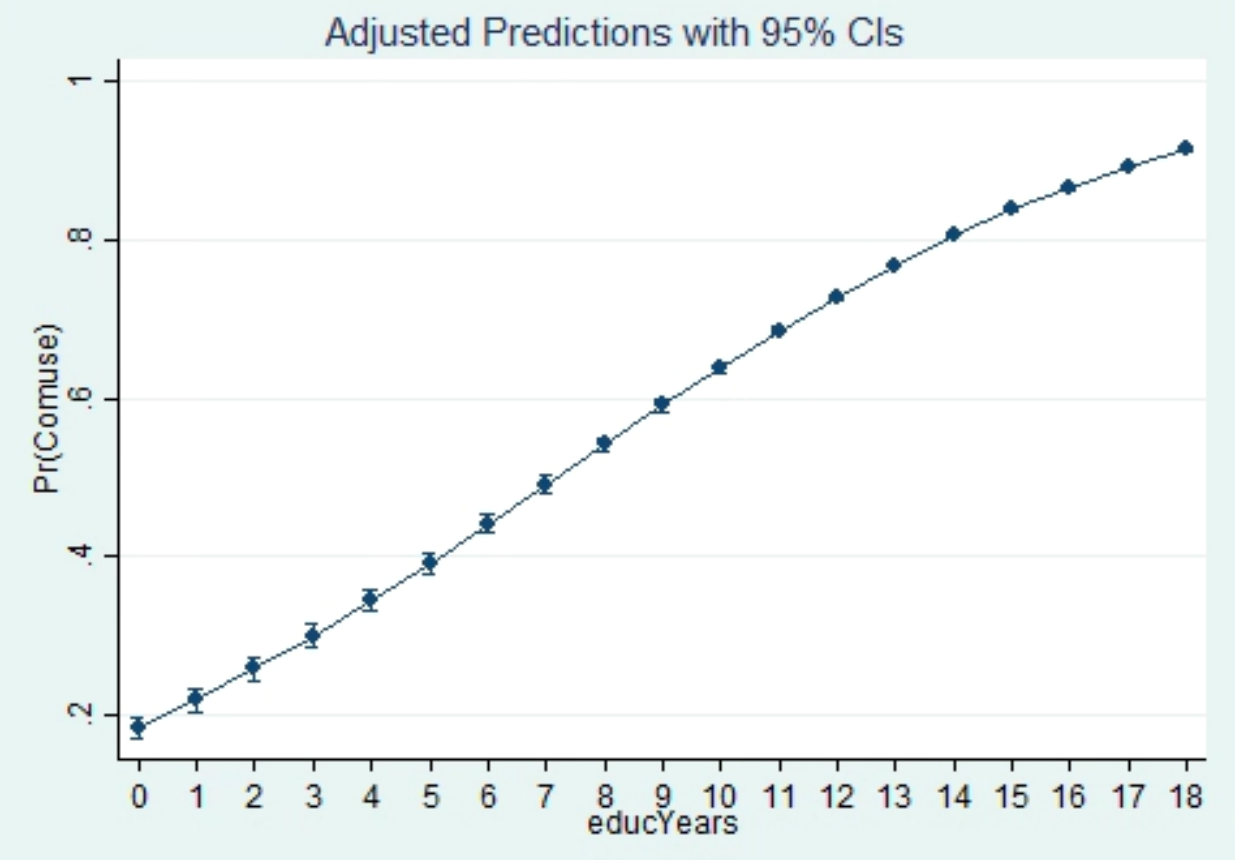
Figure 2. Frequency distribution of Years of education



Source : Current Population Survey and Internet Use Supplement, Nov 2017, ages 18 to 64

A probit (probability unit) model was utilized on probability of having internet access at home in order to determine who should be considered as "uneducated" in this study. The fitted line is slightly curved in the *Figure 3* below. As shown in the figure, until 13 years of education an additional year of education increases the probability of having internet access at home whereas after 13 years, the rate of increase in probability declines as the line is curved after the threshold. The installation of the internet is more prominent in the sampled individuals who are less educated compared to the people with higher education. Therefore, to analyses the results of having internet access at home, it is reasonable to draw a line between educated and uneducated at 13 years of education.

Figure 3. The impact of years of education on the probability of having Internet access at Home



Source: Current Population Survey Computer and Internet Use Supplement Nov 2017, ages 18 to 64

ii. OLS and PWLS Regression Analysis

In total five regressions were conducted. The model (1) was run to measure the raw differential (regression of log of earnings on the Internet Use). The model (2) includes internet use and years of education. Model (3) adds the core demographic variables like race variables, marital status, sex and living in a city (metro). Model (4) controls variables on occupation and industry and lastly model (5) uses the PWLS regression instead of the OLS regression.

As shown in the bottom row of *Table 2* the model 5 has the highest adjusted R^2. Adjusted R^2 measures the goodness of fit in the regression by penalizing the number of

predictors. If there are more variables in the regression the penalty is also heavier. Using more variables in the regression analysis will increase the goodness-of-fit because more

predictors will be used to measure the dependent variable which is income in this analysis. But there is a problem with R^2, it does not decrease as you increase the number of variables that are predicted. These regressions often over-fit the sample. It is a problem because the model is not flexible enough to predict the trends in the population data. Therefore, misleading high R^2 values are produced with less ability to predict.

Table 2. Regression results for Internet Use at Home

	(1)	(2)	(3)	(4)	(5)
	OLS	OLS	OLS	OLS	PWLS
	ln(Earnings)	ln(Earnings)	ln(Earnings)	ln(Earnings)	ln(Earnings)
Internet access at Home (1 = yes)	0.0636**	0.0496*	0.0741***	**0.0382***	**0.0158**
	(0.0204)	(0.0206)	(0.0182)	(0.0175)	(0.0198)
Years of Education		0.0266***	0.0506***	0.0360***	0.0388***
		(0.00597)	(0.00563)	(0.00564)	(0.00630)
Experience			0.0565***	0.0424***	0.0422***
			(0.00243)	(0.00243)	(0.00281)
Experience Squared			-0.000958***	-0.000707***	-0.000703***
			(0.0000492)	(0.0000488)	(0.0000566)
Female			-0.319***	-0.175***	-0.159***
			(0.0219)	(0.0245)	(0.0298)
Married Female			-0.131***	-0.105***	-0.107**
			(0.0314)	(0.0305)	(0.0344)
_cons	6.293***	5.985***	5.337***	6.004***	5.984***
	(0.0175)	(0.0712)	(0.112)	(0.245)	(0.251)
N	6558	6558	6558	6558	6558
R-sq	0.001	0.005	0.251	0.433	0.451
Standard errors in parentheses					
* p<0.05	** p<0.01	*** p<0.001"			
Variables included as controls					
race, occupation, industry, hispanic, marital status and metropolitan					

As shown in *Table 2,* the model (5) is the best model for predicting the data with the highest adjusted R^2. The regular R^2 value was penalized in the adjusted R^2. It is also shown that model (4) is just as good as model (5) since they have almost similar adjusted R^2. Therefore, in this regression model, the PWLS and OLS seem to perform equally well. The experienced square term is also shown as negative which is how it should be according to the theory.

It is also to be noted that STATA omitted the industry variables because of the collinearity in model (4) and model (5). The omitted variables had the similar categories with similar features.

Since model (4) and (5) perform equally well the values can be referred to any of the two. According to model (4) the standard error of the estimated coefficient is 0.0175. The exact coefficient estimate is $e^{0.0382}- 1 \cong 0.038$ because the natural log has been taken for the dependent variable. To interpret the result appropriately, anti-log has to be applied. Since 0.0175 * 1.96 = 0.034, 95% of confidence interval constructed as 3.8 +/-3.43% will cover the true parameter value about the impact of internet access at home on weekly earnings. Even if the coefficient estimate is not as good, having internet access at home does increase the income by approximately 7%. The result signifies how important it is to have internet access at home by controlling education, core demographic variables, experience, occupation and industry. We have to also keep in mind that the results exclude the richest groups in the sample because of the limitations mentioned above. Thus, if we have the true data we might get different results.

Breusch-Pagan test was also conducted on model (4) to check for the heteroskedasticity in the data. STATA returned an extremely low P-value which signifies that the null hypothesis was rejected with significant evidence. Heteroskedasticity is considered as a violation of identical distribution of error term in Classical Econometric Model (CEM), the standard errors reported on *Table 2.* should be broken. Therefore, STATA needs to report the robust standard errors. It is also shown in *Table 2* that the reported standard errors are identical. This could be because of taking the natural log on the earnings variable or because of the large sample size (6558).

Breusch-Pagan Test on STATA

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: i.cinethp educYears exp expSq i.race i.hispan i.sex i.marst
                    femaleMarried i.metro i.occ i.ind

         chi2(673)     =  3846.59
         Prob > chi2   =   0.0000

.
.
```

5. Conclusion and Further Research

The purpose of conducting this research and performing the econometric analysis of the Internet Use among the uneducated population was to find out the impact of having internet access on earnings in 2017.

The empirical analysis using the CPS data from 2017 predicts that having internet access at home does have a statistically significant positive impact on earnings. The findings of this study aligns with the studies introduced in the Literature Review. However, according to Krueger's (1993) there was 10-15% increase in income in 1994, Internet Use at home does not have much of an economic impact anymore. This could be because of a lot of reasons. It's been almost 30 years since the research was conducted by Krueger and the world does not look the same anymore. There has been an advent of a lot of new technologies since then especially the advent of smartphones and data plans. The Internet is also not only used for educational purposes but also for recreational activities.

Thus, further research should be conducted with data on possession of smartphones and there should also be a categorization of why the internet is being used. I believe that

this is a part of the omitted variable biases and I did not have any control over it in this paper.

References

DiMaggio, P. and Bonikowski, B. (2008). "Make Money Surfing the Web? The Impact of Internet Use on the Earnings of U.S. Workers." *American Sociological Review*, Vol. 73(2), pp. 227–250.

Krueger, A.B. (1993). "How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989". *The Quarterly Journal of Economics, 108*(1), pp. 33-60.

Houngbonon, G and Liang, G (2017). "Broadband internet and income inequality". hal-01653815.

Brenčič, V. (2012). "Wage posting: Evidence from job ads. *The Canadian Journal of Economics / Revue Canadienne d'Économique." 45*(4), pp.1529-1559.

File, T. (2013). "Computer and Internet use in the United States." Population Characteristics.

Mincer, J. (1958). "Investment in Human Capital and Personal Income Distribution." *The Journal of Political Economy,* Vol. 66(4), pp. 281-302.

Faberman, J. R. and Kudlyak, M. (2016). "What does online job search tell us about the labor market?" *Economic Perspectives,* Federal Reserve Bank of Chicago, No. 1, January 2016.