Examining the Effects of Internet Access on Earnings:

Can access to the internet help close the racial income gap?

Mark Russeff, Renato Albolea, Shuai Ma

Introduction

In today's society, reliable access to the internet has become something that we all take for granted; an integral part of our daily lives. We are constantly connected via phone, laptop or tablet to a vast world of information that is always just a click away. The internet helps us live more efficient, well informed and interconnected lives. Something that was once considered a luxury a few decades ago is increasingly becoming an essential component of our everyday existence. Unfortunately, for some this life changing technology is still out of reach. According to a 2018 Federal Communications Commission's (FCC) report nearly 19 million Americans, or 6% of the population, still lack access to high speed internet. Many argue that this lack of access places these already vulnerable groups at a sizable social and economic disadvantage. In response, policy makers worldwide have begun to recognize unrestricted internet access as a modern human right and have moved to provide free access to all of their citizens. While these policies are rooted in social justice, free speech and equality, the economic component is also very important. Closing the "information gap" and bridging the chasm between those who are exploiting the vast potential of technology and those who are not should lead to increased economic opportunity for all. On the other hand, free and unrestricted access to the internet may not have any effect on one's earnings and resources may be more effectively allocated to other programs. The purpose of this paper is to examine whether the policies providing free internet access can have a positive economic effect on a person's earnings, specifically for already disadvantaged minority groups. First, we will ask: does access to the internet have any effect on a person's earnings? Then, does that premium or penalty vary by race?

As more policy makers begin to embrace the idea of free unrestricted internet access as a modern human right, it is becoming increasingly important that they understand the effects of such policies both socially and economically. If providing free access to the internet can increase an individual's earnings and in turn raise their economic wellbeing then such policies should engender broad support. However, more importantly, if they can also have a positive impact on racial groups historically disadvantaged by systemic racism and economic oppression then they could be a significant tool in the efforts to close the racial income gap. So as policy makers earmark funds for more jobs programs, retraining initiatives, and other strategies to combat income inequality it is important that they are aware of all possible factors effecting the earnings of their citizens. Current programs like the FCC's Lifeline Program which subsidizes broadband services for the poor are woefully underfunded and offer a mere $9.25 per month for internet service, if studies can show a substantial economic benefit from investing in such programs then perhaps they will expand and be more effective moving forward.

In order to address our research questions in the context of these policy challenges we will be estimating four different earnings models. All of the models will be estimated by preforming a multiple regression analysis using ordinary least squares (OLS) estimators. The first model will be addressing whether or not access to the internet effects earnings for our whole sample. While the other three models will be modeled on stratified data using the categorical covariate race in order to identify interactions between our categorical covariate and the other variables, particularly access to the internet. Then by interpreting our results we hope to garner some evidence of a relationship between access to the internet and earnings so that we can offer a clearer perspective on public policy and the right to internet access.

Data

We are using data from the 2016 American Community Survey (ACS) to examine if access to the internet has an effect on earnings and if that effect varies across race using data from the state of Alabama. The 2016 ACS Alabama data sets provided by the United States Census Bureau contained both individual and household level data which we joined to create one data set containing a representative sample of 47,659 individual level observations across 17 variables relating to various demographic characteristics like age, race, sex and educational attainment. Access to the internet was assessed via the ACCESS variable contained within the household level data, which was then aggregated to the individual level in order to determine each individual's access level. In order to better address the effect access to the internet has on earnings we removed respondents who reported that they had "paid for access". This limited our data to individuals who either received free internet or had no access at all. The implication for our analysis is that this provides a more contextual result for policy makers who could be evaluating the possible economic benefits of providing free internet to those with no access. Since we are measuring the effect of access on earnings we chose to use yearly earnings for full-time workers so any individuals with earnings of 0 or less were removed from the sample. In addition, we also chose to limit our sample to people of working age, so individuals between 14 – 62 years old. Lastly, due to small sample sizes in other race categories our sample selection was limited to White, Black and Hispanic individuals. This is a limitation of the Alabama ACS data; future analysis may require the use of a representative national sample with more observations.

Table 1 and 2 display the descriptive statistics for the continuous variables in our data set. Table 1 displays the 372 observations with free access and Table 2 for the 1268 observations with no access. Immediately we can observe that the average earnings for an individual with free access is about 23.5% more than for someone with no access at US$40,218.28 and US$32,564.90 respectively. Both tables have the same max earnings at US$356,000 which is likely the earnings cap of the data but min earnings for no access is only US$100 compared to US$1,600 for free access. While the range may be large for both tables the interquartile range is far more compact. It is interesting to note as well that the average age for both free access and no access is quite close, within two years, this goes against the intuition that older people may be indifferent to whether or not they have access to the internet.

Table of Descriptive Statistics for Free Access

Table 1: Free Access

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Age (years)	372	42.08	12.54	17	31	53	62
Earnings (dollars)	372	40,218.28	35,526.30	1,600	18,500	50,250	356,000

Table of Descriptive Statistics for No Internet Access

Table 2: No Access

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
Age (years)	1,268	43.84	12.82	16	32	55	62
Earnings (dollars)	1,268	32,564.91	30,005.72	100	15,900	40,000	356,000

Modeling Approach

For our analysis we have chosen to perform a series of multiple regression analyses using OLS estimators. For this analysis we used a logarithmic transformation of the dependent variable, earnings. Even though there was no evidence of heteroscedasticity we chose to use log(earnings) because conceptually percentage changes across our independent variables makes more sense. Next, we addressed variable selection. To address the first research question that the effect of internet access to person's earning, the most interested variable which was included as the dummy variable was Free Access. Since we removed the observations of "Paid Access" and "N/A" from the sample. The reference category for Free Access was No Access in our model. The other factors which we believed would impact on people's earning were race, gender, age, and education. So, we included the categorical variables of race as Black and Hispanic with White as the reference category, Female with Male as reference category, High School Diploma or GED, Some College, Undergraduate Degree or Higher with no High School Diploma as the reference category. The variables for age and ageˆ2 were included, ageˆ2 was included to account for the no linearity of the relationship between earnings and age.

To address our second research question of if the penalty or premium from internet assess vary by race, we remove the categorical variable of Hispanic and Black and stratified the data set by race - Black, Hispanic, White. In the second model, we compare the change of premium/penalty from Free Internet Access on difference races.

This constructed a model of the form:

Model for research question 1:

$$\log(wage) = \beta_0 + \beta_1 FreeAccess + \beta_2 Hispanic + \beta_3 Black + \beta_4 Female + \beta_5 Age(Years) + \beta_6 Age^2 + \mu$$

Model for research question 2:

$$\log(wage) = \beta_0 + \beta_1 FreeAccess + \beta_4 Female + \beta_5 Age(Years) + \beta_6 Age^2 + \mu$$

Empirical Results

Table 3 presents the estimate for our log earnings equation for our whole sample. Our estimates suggest that earnings are positively associated with educational attainment and age, both of which is what we would expect. Earnings are also negatively associated with being female or black. Controlling for all other variables our estimate suggests that a black person will make 13.15% less than a white person on average. This is both a statistically significant finding at the 1% level and is economically significant because it illustrates the problem of racial inequality in earnings. If we look at the free access coefficient we can see that controlling for all other variables free internet access leads to a 10.5% premium in earnings on average over having no access. This result is statistically significant at the 5% level and is also highly economically significant. We can say with some confidence that based on our estimated log earnings equation, earnings are positively associated with having free access to the internet. This is an important finding for our research and leads us into the second part of our analysis.

Table 4 presents the three estimated log earnings equations on the tratified data, each model representing a different race. When we examine the Free Access coefficient we can see that for Black people controlling for all other variable there is a 19% earnings premium for free access on average over having no access at all. This result is statistically significant at the 5% level and is extremely economically significant, more than making up for the 13.15% penalty in earnings for being black compared to being white. These estimates suggest a possible avenue to closing the racial earnings gap. For white people controlling for all other variables there is only a 5.65% premium on average for having free access. This is not a statistically significant estimate nor is it particularly economically significant.

Table 3: Results

	Dependent variable:
	ln_earnings
Free Access	0.100^{**}
	(0.042)
Hispanic	0.013
	(0.082)
Black	-0.141^{***}
	(0.036)
Female	-0.373^{***}
	(0.035)
Age (Years)	0.069^{***}
	(0.010)
Age Squared	-0.001^{***}
	(0.0001)
High School Diploma or GED	0.184^{***}
	(0.048)
Some College	0.318^{***}
	(0.051)
Undergraduate Degree or Higher	0.659^{***}
	(0.065)
Constant	8.436^{***}
	(0.202)
Observations	1,640
R^2	0.212
Adjusted R^2	0.207
Residual Std. Error	0.681 (df = 1630)

Table 4: Results

	Dependent variable:		
	ln_earnings		
	White	Black	Hispanic
	(1)	(2)	(3)
Free Access	0.055	0.177**	0.304
	(0.051)	(0.076)	(0.249)
Female	−0.472***	−0.236***	−0.128
	(0.043)	(0.060)	(0.158)
Age (Years)	0.075***	0.074***	−0.022
	(0.013)	(0.018)	(0.051)
Age Squared	−0.001***	−0.001***	0.0004
	(0.0002)	(0.0002)	(0.001)
High School Diploma or GED	0.212***	0.117	0.147
	(0.062)	(0.083)	(0.180)
Some College	0.336***	0.287***	0.165
	(0.067)	(0.088)	(0.233)
Undergraduate Degree or Higher	0.739***	0.468***	0.631**
	(0.081)	(0.121)	(0.291)
Constant	8.340***	8.170***	10.192***
	(0.254)	(0.358)	(1.032)
Observations	976	586	78
R^2	0.244	0.153	0.210
Adjusted R^2	0.238	0.143	0.131
Residual Std. Error	0.662 (df = 968)	0.713 (df = 578)	0.619 (df = 70)
F Statistic	44.623*** (df = 7; 968)	14.968*** (df = 7; 578)	2.652** (df = 7; 70)

Note: *p<0.1; **p<0.05; ***p<0.01

Conclusion