Examining the Effects of Internet Access on Earnings in Alabama:

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, among other things, helps us have more efficient, informed and interconnected lives. It is interesting to note that just a few decades ago the internet was considered a luxury but now it is increasingly becoming an essential component of our everyday existence, changing everything from how we bank to how we communicate with others. 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[1], 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 policies in the state of Alabama 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. Access to the internet can give these disadvantaged groups the tools necessary to search for and apply for better paying jobs as well as learn new skills and earn certifications in growing industries. In states like Alabama where the median household income is among the lowest in the country at only US$45,182[2] and where a history of racial inequality has resulted in an income gap of approximately 10.2% it is of the utmost importance that policy makers pursue not only good intentioned but well-informed programs. Due to limited resources, federal and state governments cannot afford to undertake every initiative that seems reasonable, so studies that examine the real impact of these proposals are very important. Current programs like the FCC's Lifeline Program[3] 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.

[1]Federal Communications Commission. 2018. Eighth Broadband Progress Report. https://www.fcc.gov/reports-research/reports/broadband-progress-reports/eighth-broadband-progress-report

[2]U.S. Census Bureau. 2018. Quick Facts: Alabama. https://www.census.gov/quickfacts/al

[3]Federal Communications Commission. 2018. Lifeline Program for Low-Income Consumers. https://www.fcc.gov/general/lifeline-program-low-income-consumers

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 can show that there is evidence of a positive relationship between access to the internet and earnings especially for Black individuals. These findings will be the basis for offering a clearer perspective on public policy and in order to make appropriate recommendations.

Data

We are using data from the 2016 American Community Survey (ACS) from the state of Alabama to examine if access to the internet has an effect on earnings and if that effect varies across races. The 2016 ACS Alabama data sets, provided by the United States Census Bureau, contains 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, earnings and educational attainment. Access to the internet was contained within the household level data, which was then aggregated to the individual level in order to account for each individual's access level. In order to better understand the relationship between earnings and having access to the internet we removed individuals who paid for internet access; limiting our data to individuals who either received free internet or had no access at all. Individuals who paid for access were removed in order to avoid endogenous variable bias, due the fact that the decision to purchase internet would be strongly correlated to the dependent variable earnings. One important implication for our analysis is that it provides a 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 to the internet on earnings, we chose to use yearly earnings for full-time workers. 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, individuals between $14 - 62$ years old. Lastly, due to a limitation of the Alabama ACS data, small sample sizes in some race categories restricted our sample selection to White, Black and Hispanic individuals.

Table 1 and 2 display descriptive statistics for the continuous variables in our data set. Table 1 displays the statistics for the 372 individuals with free access and Table 2 for the 1268 individuals with no internet access. 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 construct four multiple regression analyses using OLS estimators with 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 are easier to interpret. To address the first research question regarding the effect of access to the internet on earnings, the variable of the most interest was included as the dummy variable Free Access, with the reference category of No Access. The other factors impacting people's earning were race, gender, age, and educational attainment, to account for them we included the categorical variables of race as Black and Hispanic with White as the reference category, Female with Male as reference category, and High School Diploma or GED, Some College, Undergraduate Degree or Higher with no High School Diploma as the reference category. In regards to age, we also included ageˆ2 as a variable to account for the nonlinearity of the relationship between earnings and age. The resulting model was of the form:

$$\log(earnings) = \beta_0 + \beta_1 FreeAccess + \beta_2 Hispanic + \beta_3 Black + \beta_4 Female + \beta_5 Age(Years) +$$

$$\beta_6 Age^2 + \beta_7 HighSchoolDiplomaorGED + \beta_8 SomeCollege + \beta_9 UndergraduateDegreeorHigher + \mu$$

In order to examine if the relationship between access to the internet and earnings varied by race we stratified the data by race consequently removing the categorical variables of race from the model. The comparison between the three resulting models helps us determine if the premium or penalty for having free internet access varies across the three selected ethnic groups in Alabama. Three models where constructed for each race in the following form:

$$\log(earnings)(race) = \beta_0 + \beta_1 FreeAccess + \beta_4 Female + \beta_5 Age(Years)$$

$$+\beta_6 Age^2 + \beta_7 HighSchoolDiplomaorGED + \beta_8 SomeCollege + \beta_9 UndergraduateDegreeorHigher + \mu$$

It is worth noting that a Breusch-Pagan-Godfrey test was performed on all four models to test for homoscedasticity. The general and White models both fail to reject the null hypothesis of homoscedasticity. However, the Hispanic and Black models both reject the null hypothesis of homoscedasticity at 1% and 10% respectively. Since we already did a logarithmic transformation on earnings the robust standard errors were calculated and included in Black and Hispanic models along with the adjusted F-statistic in Table 4.

Empirical Results

Table 3 presents the estimate for our log earnings equation for the whole sample. The estimates suggest that earnings are positively associated with educational attainment, both of which is expected, and this relationship is highly economically and statistically significant at the 1% level across each educational variable controlling for all other variables. Age is also positively associated with earnings, until the age of approximately 52 years old when earnings begin to decrease. As expected, 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 internet access. This result is statistically significant at the 5% level and is also highly economically significant. We can argue 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; does this effect vary across races?

Table 4 presents the three estimated log earnings equations on the stratified data, each estimated for a different race. When we examine the Free Access coefficient we can see that, controlling for all other variables, there is on average a 19% earnings premium for Black individuals with free access over having no access at all. This result is statistically significant at the 5% level and is extremely economically significant. For white people controlling for all other variables there is only a 5.65% premium on average for having free access, but this finding is not statistically significant not even at the level of 10% nor is it economically significant. For Hispanics almost all of the coefficients are not statistically significant at any level, more than likely due to the small sample size of 78 observations.

Conclusion

The multiple regression analysis in the first model shows that free access to the internet has an important positive effect on earnings. After controlling for age, race, gender and educational attainment, having free access to the internet resulted in a 10.5% earnings premium over having no access to the internet. In the second set of models that stratified the data by race this relationship holds, however only the result for black individuals remains statistically significant. However, it did show that for black individuals having free access to the internet resulted in an economically significant 18% earnings premium over having no access to the internet.

While the findings of this study may be compelling, it is worth acknowledging that there are some limitations to our analysis. First, although we were able to find a positive relationship between access to the internet and earnings in the state of Alabama, it does not mean that this same relationship holds in all other states particularly in regards to race. Generalizing these results to a national population would violate the Gauss Markov assumption of random sampling because our sample was very specifically drawn from the state of Alabama. This would result in a national model that was biased and not a valid estimator. Second, there may be other variables that affect earnings that were not included in the model. These variables could be things like years of experience in a particular job or the industry the individual works in, by including these variables in future models it could increase their precision and would better satisfy the zero conditional mean assumption, therefore reducing the potential for omitted variable bias. Third, only one of the coefficients for Hispanics in our stratified model was statistically significant, this was likely due to the fact that our sample size contained only 78 observations. A larger sample would be required to better determine the relationship between earnings and access to the internet for Hispanic individuals. Forth, the models for Black and Hispanic individuals both rejected the null hypothesis of homoscedasticity, even after the logarithmic transformation was performed. This demonstrates the problem of heteroscedasticity, to address it we included robust standard errors for those models but in the future a better specified model will be necessary. Finally, there are many reasons why someone may not have internet access, the survey did not differentiate between people who choose not to pay for the internet and those who cannot afford it. If this were addressed in future studies the results would more accurately reflect the relationship between earnings and receiving free internet because you could not afford it. For the purpose of policy, it is our recommendation that more specific and in-depth samples are used moving forward.

In today's economy the internet has become an increasingly important tool in every aspect of the career development process; from finding and applying to new jobs to learning new skills. The positive relationship between access to these tools and an individual's earnings is an important one. Even more important is the implication for income equality, our findings would suggest that free access to the internet has a larger effect for black individuals therefore it could help close the racial income gap in Alabama where it has remained stagnant for far too long. For policy makers, a positive earnings effect can raise tax revenues, boost standards of living and lower the number of people requiring state assistance. This could be achieved by investing a relatively small amount into subsidizing the cost of internet access for those who cannot otherwise afford it. Based on our analysis this could both raise the meager median income in Alabama as well as begin to close the racial earnings gap that has been such a challenge to address. These policies could also be enacted nationally by expanding programs like the FCC's Lifeline program which is already in place. With that being said, future research will be required in order to assess the possible positive earnings effect this could have on a national scale.

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)
F Statistic	48.618^{***} (df = 9; 1630)
Note:	*p<0.1; **p<0.05; ***p<0.01

Table 4: Results

	Dependent variable:		
	ln_earnings		
	White	Black	Hispanic
	(1)	(2)	(3)
Free Access	0.055	0.177**	0.304
	(0.051)	(0.072)	(0.286)
Female	−0.472***	−0.236***	−0.128
	(0.043)	(0.061)	(0.160)
Age (Years)	0.075***	0.074***	−0.022
	(0.013)	(0.016)	(0.049)
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)	(0.089)	(0.182)
Some College	0.336***	0.287***	0.165
	(0.067)	(0.089)	(0.246)
Undergraduate Degree or Higher	0.739***	0.468***	0.631*
	(0.081)	(0.138)	(0.349)
Constant	8.340***	8.170***	10.192***
	(0.254)	(0.321)	(0.942)
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

Contributions

All of us made similar contributions to the project across many areas and worked together on all elements. However, we each individually led for certain sections: Renato for programming/model specification, Shuai for EDA and Mark for final writing.