Transcript of Oral Presentation, email is: zmilnes@mail.umw.edu

Hi, my name is Zachary Milnes, I am with the University of Mary Washington and today I'll be discussing the project I did, which is The Effect of Internet Download Speeds on Income in a County.

The research question is, like I said before, it is what, if any effect would an increase in internet download speeds in an area have on per capita income in a county.

We already know that the introduction of the internet increased wages, especially those of white-collar workers. But I predict that there's also a beneficial effect caused by increased efficiency through faster speeds which would lead to further wage increases.

What motivated me to do this project? The Internet is becoming ever more important in our lives and it's accelerating faster and faster, you know 10 years ago the internet was nothing like it is now and the introduction of the internet has done many things for the people on it. It's made marketing and sales easier and it made Amazon possible. Logistics is much faster and easier for everybody to do. It's made communication easier with the introduction of social media and it's easier to learn about a variety of subjects. Some of them are behind paywalls but a lot of it is very free and you can look at it anytime you want from the comfort of your own home.

The literature on this subject mostly focuses on the gap in broadband availability between rural and urban areas. Recently the gap has narrowed but it still persists because rural areas really lack the necessary funding and infrastructure to get broadband really into their areas, and there's also oversubscription problems. You know, maybe they do have broadband, but they don't have enough so people use it and it's not as efficient as say an urban area which has lots and lots of it invested. In one paper from 1991 which was most relevant to my research, it suggested that people who used computers at work compared with those that did not, made 10 to 15% higher wages, and this paper was written in 1991 using data from 1989, so there's definite room for growth in what they found.

So the economic theory which drove my model is really just… it's the basic of the basic. It's the Cobb-Douglas production function. The internet in my mind is a measure of total factor productivity, it is new technology which has made it easier for labor and capital to produce. So an increase in Broadband speeds would cause the steepening of these curves causing more production for less. A more in-depth explanation of this, what you see here are really quick mock-ups of what a Cobb-Douglass Production function would look like. You have Labor and Capital on those bottom axes and then production on the Y, directly across from the labor axis is the labor line. And it can take a little bit to really understand. This is deceptively simple looking so don't worry about it. But if you look at the difference between these two, there's, I mean, it might be hard to see but there's a definite difference in the curves. The one on the left we're going to call pre-internet, you know, this is just a representation of what it might have been like for labor and capital productivity. On the right hand side we introduced the internet, no the curve

is steeper and you can put in less labor or capital to get higher production. Of course there's diminishing marginal returns to everything but it boosts production across the board and that's theoretically what should happen, and from what I saw, that is what happened.

On my equation I used a two-stage least squares regression. It's notoriously difficult to do and I used a panel data set so that's even worse. My dependent variable of course is income and I decided to take the natural log of it so that I could find the change in percentage as I feel like that is more valuable to look at then the random numbers you would get from not doing that. Other explanatory variables, I had median age and age squared together. Age is known to be a determinant of income and to correct for, you know, bias, you have to actually include an age squared variable as well so that's what I did, that's why I had the age squared variable. And the Unemployment rate which of course is related to income. If you have higher unemployment people aren't making money so per capita income goes down. Then I also have percentage of people with broadband that's, you know, the most iffy variable, it still belongs there, but you know especially in urban areas where like 98% of people have broadband it's not necessarily really applicable, but because of that rural gap, it still needs to be in there to correct for it. Next we have Bachelor's and Above. This is an education variable. So, education is pretty directly correlated with income, if you have a bachelor's degree or above rather than a high school diploma, we can expect you to make more money overall over lifetime than someone with a high school diploma. And then of course the main independent variable would be the median download speeds. That's what I'm really looking at. Of course the system of equations, I had to reverse the income and median download speeds because it's entirely possible that there's simultaneity bias, and that basically means that, sure, median download speed might effect income but income might also effect median download speed at the same time, for instance if you're in a wealthy area, maybe you have better resources to pay for that broadband infrastructure which would cause higher download speeds which causes higher productivity which causes higher income, therefore it's simultaneous. And that's really the entire reason that I had to do two-stage least squares. I instrumented the median download speeds with the percentage of white-collar jobs in a county and I did that because to instrument a variable it has to be really heavily correlated with the independent variable (i.e. income) and it can't be correlated with any error terms, so anything not included in the equation. Percentage of white-collar jobs is probably very correlated with bachelor's degrees, but that's in the equation so it should be fine and that's why I did that.

So how did I get my data? I got my data from, at least for the Broadband download speeds, I got it from this independent research place called the Measurement Lab. I found them on sort of a whim but in my opinion it's much better than getting the data from the government because the way the government gets their ISP download speed data is they asked them to self-report and if you've ever signed a contract with an ISP to get like say.. I don't know one gigabyte of download speed, that's a lot, but say you sign that contract and then you check on any website and you'll have like a hundred and 50 megabytes, so ISP's lie and that is why I had to go out and get this independent data. They had the data from 2014 to 2018, so that's over the course of 5 years, and then I included 135 counties across the continental United States just to expand my data set for a total of 675 observations. Most of the other explanatory variables like

age and bachelor's degrees came from the United States Census Bureau's American Community Survey. So there's really no chance of that being inaccurate unless the government is inaccurate in which case that's a whole other problem.

So what did my regression show? Yeah, it looks like median download speed does in fact have an effect on per capita income, for instance looking at this graph I can tell you that it's predicted that a 10 megabyte per second increase in average download speed in a county should increase per capita income by 7%. That's pretty dang good. Now there's reasons to doubt this and I will get to that in a second.

Two-Stage least squares, aside from being notoriously difficult, the coefficients that are calculated are not completely accurate because the math that goes into it is very… It's kind of like almost cheating to get what you want, but it's better than letting simultaneity into your equation which would screw up your coefficients even more, so it's almost like a necessary evil. So that is a definite issue, and then finally for future research. I'd like to see someone use data from the current time, the quarantine, and compare it with other recessions or recession like events in the past. This is a natural experiment on how much we can really get done at home and how much is essential and I'm kind of kicking myself because I didn't do this, you know, because I started this before coronavirus was even a thing. So it would be interesting to see that in the future.

That is it for my presentation, if you have any questions, feel free to email me. I will try to respond in a timely manner and thank you for watching