INCREASING BROADBAND ACCESS' IMPACT ON THE ECONOMY AT A LOCAL LEVEL

Randall Naylar, University of Central Arkansas

Kaye McKinzie, University of Central Arkansas

ABSTRACT

As with all technology, broadband is ever-advancing. Not only are internet speeds increasing, but internet access is becoming more readily available in both businesses and homes. Where once those living in rural areas had limited to no access to high-speed internet connections, now they have choices including cable, DSL, satellite, and low-orbit satellite. During the recent COVID lockdowns, our society not only became more reliant on the internet but also became more comfortable with using the internet in ways they had not previously. A major change was working in non-traditional business environments that are often composed of one's home. The US Federal government has recognized the importance of internet access with the Infrastructure Investment and Jobs Act which provided $65 billion for broadband. Recently there have been many stories about how broadband has impacted economic growth. (Why Broadband, n.d.) This directly drove our research. If the US is investing this much money into broadband expansion, where is the proof that it has a direct impact on the economy? We could not find any recent study at the granular level in the US. This study focused on broadband availability and unemployment from 2010 to 2018 at the county level. Some unexpected, interesting insights are provided.

Keywords: broadband, inflation, unemployment, Divisions

INTRODUCTION

There is limited publicly available data on broadband availability and its impact before 2010. This is probably because 2G was introduced in 1991, 3G in 2001, and 4G did not become available until 2006. Even with the introduction of 4G, in 2011 most of the world was still reliant on 2G & 3G. (The World, 2011) In November 2021, the US President signed the Infrastructure Investment and Jobs Act which provided $65 billion for broadband. (NTIA, 2021) This program has provided many opportunities to expand broadband, but what impact will it have? Does broadband impact unemployment?

Broadband has proven to be a breakthrough technology for many users in urban areas and rural areas (Isley & Low, 2022). With a broadband connection available, employment opportunities are often easier to find, especially with the push to a more digital market type (Hasbi, 2017). Widespread broadband availability for consumers can result in many communication opportunities and overall household income impact (Lobo et al., 2019). Broadband is seen as a contributing variable to many economic developments across the United States and the world (Lobo et al., 2019). In recent years, there has been an overall increase in consumers using broadband which contributes to many benefits to the overall economy (Lobo et al., 2019).

Unemployment rates are just one of the many economic factors that broadband can affect. Broadband allows users to connect with businesses and other users in a more efficient way (Hasbi, 2017). Literature from (Lobo et al., 2019) states that geography is no longer a barrier to entry into the employment market because broadband can allow users to apply for career opportunities from the comfort of their homes. The unemployment rate can be a main economic driver when looking at real estate, considering a new business, or even overall well-being in the area (Lobo et al., 2019). Our research takes unemployment rates and applies them to broadband availability to see just how much of an impact broadband has on unemployment rates across the United States.

Early literature has looked at broadband speed and penetration data at a less granular level. For our research, we were able to take data from the census bureau and labor statistics bureau to measure the impact of an available broadband connection on the unemployment rate at the county level. We applied granularity for this study as we dive into forty states across the United States and 1,268 counties within those states. In our research, we wanted to look at how much of an impact broadband availability has on unemployment rates at the county level. Our research focused on the period of 2012 to 2018 to give the literature an update on the most recent census bureau study on broadband availability. Our research took a similar path to how Lobo et al. (2019) conducted their research but instead of looking at broadband speed, we examined the impact of having broadband available.

This research is an extension of other work from similar studies concerning broadband and the economy conducted in many countries. We mirrored the independent variables (IVs) they used in their studies. We also mirrored their approaches using linear regression. Our paper follows a traditional format with the literature review followed by a section discussing the methodological approaches. The results describe these approaches in detail and the paper finishes with our conclusions and recommendations for future work.

LITERATURE REVIEW

Broadband

Since 2010 there have been several studies discussing how higher broadband speeds contribute to a lower unemployment rate in high-speed counties. Much of this early research has focused on broadband speed and not the impact of broadband availability by county. One such study (Lobo et al., 2019) was of the most interest to us due to its similarity, however, the data was from 2016 and prior – thus in need of an update. In this study, focusing on one specific state, Lobo found that in state-funded areas broadband adoption was higher versus other states with no funding (Lobo et al., 2019). One of the key contributions of this work was the addition of geographical variables such as state and counties when analyzing broadband's impact on unemployment.

This research by (Lobo et al., 2019) also helped provide other variables for us to consider. They found that educated counties resulted in a decline in the unemployment rate. They also found that where population density (and amount) increased there was an increase in the unemployment rate. The counties that had higher population amounts resulted in a higher unemployment rate mainly because there was more data represented in that county than in others (Lobo et al., 2019). Working age was a significant variable used in research because it gives a way to look at the geographical information on the county. Counties with an average middle age demographic had lower unemployment rates. The article showed that counties with higher income levels had a lower unemployment rate and this variable had one of the highest significance rates in their findings. All in all, this study helped form the basis for our study.

Some other research that has been performed focuses on the relationship between broadband and the expansion of economic growth (Kolko, 2010). The study suggested that broadband might offer more benefits for rural areas helping businesses connect with larger markets outside of their service area. This paper used broadband data found on the FCC public data website and data from the US census bureau the time they studied was between 1999 and 2006. The variables that this study used in their FCC data were the average number of broadband providers, the average number of broadband providers employment weighted, percent of ZIP code tabulation areas (ZCTAs) with one or more providers, percent of ZCTAs with one or more providers employment weighted, average broadband provider count, average broadband provider count employment weighted, population growth, changes in employment, average pay, household income, change in the type of work location, employment growth area, percentage change with an increase in broadband availability, the highest share of technology inputs, the highest share in computer occupations, county level education attainment. The dependent variable (DV) they used was employment growth, but inflation had the most significant impact on their regression model. This pointed us to use the variable inflation as an impactful variable in our regression modeling.

Another study (Hatef, et al., 2021) was performed recently but they looked at residential fixed internet access subscriptions per 1,000 households. The variables included income, education, employment status, and housing quality. This was based on December 2016 subscription data from the FCC which doesn't address multiple years of change in the data. They also used data from Aunt Bertha's Social Service (now called Findhelp) which included variables such as types of available service by an organization, and the demand data had information on individual levels and healthcare workers. The most significant variable in their SLR model was employment status which told us how important employment status is when looking at broadband access (Hatef, et al., 2021).

Broadband's role in healthcare has become an important topic according to (Quinton et al., 2021). (Quinton et al., 2021) addressed the national priority for broadband to improve access to healthcare in rural communities. They took medical record data and measured broadband availability and its impact on income and educational achievements and whether there was a primary care physician in their zip code. The time they used was from March 2019 to March 2021. Since this study was conducted, we realized using broadband availability would be the most impact variable to unemployment. Their sensitive analysis performed in this study showed us that the variable broadband availability was an impactful variable to use when studying county-level data (Quinton et al., 2021).

The (Aliyev, 2021) study looked at several different descriptive models and their outcomes. It showed that the largest sample of unhappy people are in the category of jobless and they are mostly male. This gave us the idea to look at certain population groups that were impactful to unemployment. They took several different data points such as the age of males, the age of females, and if they are unemployment or unemployed. The variables ran in the descriptive model were median age, gender, educational attainment, and marital status. The variable median age was highly significant in their regression model. (Aliyev, 2021) told us that it would be relevant to study how unemployment is affected by median age.

The study (Dahliah & Nirwana, 2021) ran several statistical tests for unemployment and whether they had a positive effect on the poverty level. It showed that when unemployment is low the poverty level is important to look at over a time series plot to discover data findings. When certain business sectors have a high employment rate it showed that the poverty level was unaffected, but income levels were lower (Dahliah & Nirwana, 2021). The data they used was a sample from a public research site that had data from East Luwu from 2010 to 2020. This data was monthly data used in regression analysis and ran over ten years. Their multiple linear regression model tested poverty levels and how they affect unemployment. Since their adjusted R-squared told me how strong of a model they conducted it drove us to use a poverty variable in our study.

The study from (Barrero et al., 2021) pointed us to look at median income levels. The variables they used came from survey data they collected in May 2020. The variables collected were demographics, employment status, working, arrangements, earnings, commuting, internet access, age, gender, educational attainment, and income. The benefit of broadband can be a tough ask but that is what this research article was written to do. The article measured overall well-being when you have internet access during the pandemic. During the pandemic, mental health issues jumped significantly due to being confined to locations for long periods. The best SLR model (Barrero et al., 2021) used the variable median household income which is why we wanted to include this in our research.

Still, other research has focused on the impact of broadband speed on economic measurements such as GDP per capita, population density, price, urban population, labor force growth, telecommunication revenue growth, population growth, and the average achieved downlink speed. Rohman & Bohlin (2012) used the dependent variable GDP and ran separate regression models which measured the impact of broadband speed on economic growth in the 38 Organization for Economic Cooperation and Development (OECD) countries over the period 2008-2010. They researched the gap in broadband speed in certain of these countries around the world and how it affects developing nations. They found that broadband plays a vital part in developing nations across the world and doubling the speed will contribute to 0.3% growth.

Geographical

In their article, Whitacre & Gallardo (2020) used data from the census block and the FCC pointing us to the importance of looking at county-level detail to see rural and urban impacts. The period for the data was 2014-2018 and they were able to aggregate the data using census blocks. The variables they used were primarily the percentages of rural vs urban broadband fiber availability, percentage of the population in rural vs urban areas, broadband policies, percentage of state funds, population, income, educational attainment, and poverty rate variables.

Our Focus

There have been a variety of studies concerning broadband and the economy, but not with recent data. Rohman & Bohlin (2012) used data from 33 countries but only had data from 2008-2010. Whereas Hasbi (2017) had more current data (2010-2015), but only from France. And Whitacre & Gallardo (2020) had the most current data (2012-2018) but were only looking at the impact of state policies on broadband availability. What is needed is a more detailed and current look at a more granular level to see if broadband availability directly impacts unemployment.

This research updates prior research in that it uses relatively current data at the US County level of specificity. Studies that were done prior only focused on a certain state within the United States or international countries. Whereas prior research has mainly focused on broadband speed and how speed impacts economic variables, this study looks at the availability of broadband. This study also takes a more narrowed focus rather than looking at the economy or the GDP, it focuses on whether broadband availability has an impact on unemployment rates throughout the US with county-level data.

In our research, we used population below the poverty line, inflation, total amount of African American population, labor force participation, and gross domestic product. These variables were shown to be significant in prior research (Isley & Low, 2022), (Bitetti, 2018), (Bohlin & Rohman, 2012), in each of the articles GDP was shown to have a positive correlation with unemployment rates. The article (Isley & Low, 2022) demonstrated that high GDP was positively correlated with lower unemployment rates. The article (Isley & Low, 2022) focused on demographics in their research as well to show the impact on unemployment. The demographic that had the most impact on unemployment was the population of African Americans. Since (Isley & Low, 2022) showed this demographic having a strong impact we included it in our model. (Adelowokan et al., 2019) saw that the population below the poverty line had an impact on unemployment. Research conducted in the article (Adelowokan et al., 2019) found that poverty caused by unemployment had a significant impact on the growth of the economy and showed a positive relationship between the two variables. All these economic and socioeconomic variables are expected to impact unemployment. In our research, we will be modeling these variables to see if we see the same impact on our dependent variable (unemployment). From Lobo's work (2019), we would hypothesize that our study would also show an inverse relationship between broadband availability and unemployment.

METHODOLOGY

We began our research on unemployment (the dependent variable) by first exploring the data, then conducting normality assumptions and other statistical tests before beginning our regression analysis. We ran multiple regressions with different groups of categorical variables before the final regression with two different sets of geographic variables and numeric variables. The addition of geographic data led us to also explore geographical patterns relating to both unemployment and our regression models.

Numerical Variables

This research begins with just considering the numerical variables to determine if our new variable (broadband) has a statistical impact on unemployment. To do this we begin with correlation analysis, followed by single linear regression (SLR), and finally multiple linear regression (MLR). SAS Studio was used for all statistical analyses in this study.

Exploratory Data Analysis

Our data came from multiple publicly available sources. (FRED, n.d; The Center, n.d.) To begin our analysis we started with exploring the data for patterns and trends as well as conducting the required normality assumptions. The main concern in this research focused on the relationship and strength of relationships between variables. This was modeled through SLR, MLR, and finally geographical mapping. Tableau was used for all mapping in this study.

Modeling

Our dependent variable (DV) was a single numerical variable (unemployment). But, we had many independent variables (IV) that were both numerical and categorical. That made this a very large problem to solve. After consulting the Advanced Research Computing website (UCLA, n.d.), we chose to analyze in steps to reduce the problem size. We began by establishing baseline models of just numerical variables using linear regression. Then we reduced the categorical variables first due to sample size and then used one-way ANOVA before building multiple linear regression (MLR) models. Our last modeling included two MLRs to see if our reduction methods were worthwhile. One MLR had all the categorical variables that survived the sample size test and one had only those categorical variables that were in each of the final MLR models for Division, State, and/or County.

One-way ANOVA

We began our analysis of the geographical categorical variables of US Census Bureau Divisions US States, and US Counties. The One-way ANOVA test let us know if any of these three variables were significant at the variable level. It also provided sample sizes and Box-Whiskers plots. This allowed us to not only eliminate any categories that failed to meet the sample size condition but also analyze the Box-Whiskers charts for differences. Once we identified categorical variables that failed the sample size tests, we recorded them as “other” and attempted a second one-way ANOVA.

Linear Regression

A multitude of regression tests was done to determine if our addition of broadband and/or geographic categorical variables helped predict unemployment and if the reduction of these variables before adding them to the final model impacted the final model. We began with just the numerical variables starting with building SLRs and then an MLR to establish our baseline models.

The next set of MLRs focused on the geographic variables. Once the geographic variables were shown to have an impact on unemployment via the one-way ANOVA, they were converted into binary variables. Those that passed the sample size text ($N \geq 30$), were put into regression models. We ran MLRs focusing on each variable independently (Division, State, County) and then combining them into one model. This last modeling was done with two models. One began with all the geographic variables that were in the models we just built. The other had more variables because it started with all the variables that passed the sample size test.

Lastly, we merged models by including the numeric variables and the geographic variables.

Pattern

The final analysis we conducted was a visual analysis of the geographical patterns. We considered Division, State, and County independently. First, we looked at the strength and direction of the correlation of these geographical impacts. Then we looked at each variable and if it was in or not in a particular MLR. Then we compared the results of the two.

RESULTS

Including all 9 of the US Census Bureau Divisions, 40 US States, and 1,268 US Counties along with our other independent variables made for a very large set of independent variables. This data covered 2012 – 2018 and was obtained from publicly available websites.

We began with modeling our numerical variables to establish a baseline. Then we reduced the categorical variables, built geographic models, and finally built models with the geographic variables and the numerical variables to determine if the addition of geographical variables helped. SAS Studio was used for all the statistical analyses.

Once the statistical analysis was complete, we mapped the modeling results to determine patterns. This was done using Tableau software.

Variables

Our new variables being added to the data from the economic study came from The Center (n.d.): Year, State, County, and percent of broadband available per household. The economic data came from FRED (n.d.): Unemployment Rate %, Percent of Population Below the Poverty Line, Median Household Income, % of People by County Population, Inflation, Population of African Americans, Total Population Per County, Labor Force Amount Per County, Median Age by County, and GDP By County. We then merged these into one data set before importing them into SAS Studio.

Our variables were tested using Bartlett's Kolmogorov-Smirnov (SAS PROC SPECTRA) test which showed small p-values (< 0.001) which drove us to reject white noise for all numerical variables. As a double-check for cointegration concerns, we also used Johanson and the Vector Error Correction model and the Dickey-Fuller tests (SAS VARMAX) showing Pr<Rho (0.001) and Pr < Tau (<0.001) for all variables.

Categorical

The categorical variables used were: Year (mode 2017), State (mode Georgia), County, and Region (mode South Atlantic).

Numerical

Most of our variables were numerical. Summary statistics are supplied in Table 1.

Table 1. Summary Statistics

	Mean	Median	Standard Deviation
Year	2016	2017	1.8
% Unemployment Rate	0.05	0.047	0.0206
% Broadband Available Per Household	0.84	0.84	0.08
Labor Force Amount Per County	92,377	33,933	241,508
Employment Amount Per County	87,175	32,090	226,225
Unemployment Amount Per County	5,202	1,700	16,418
Median Household Income	52,277	50,300	12,651
Median Age by County	40.34	40.3	5.24
% People by County Pop with Bach Degree +	23.31	21.1	9.66
% Population Below the Poverty Level by County	15.49	15	5.54
Population of African Americans Per County	22,278	2,980	69,250
Total Population Per County	184,801	73,156	478,316
Inflation	1.93	1.8	0.17
GDP By County	9,849,093	2,547,646	31,176,891

Linear Regression (LR) with just numerical variables

This research focused on adding the influence of geographical factors at the county level and broadband availability on unemployment. We begin with determining which of these 11 numerical variables (10 from prior research cited in the literature review and broadband availability) had the most influence on unemployment. We follow this with an MLR of all of these numerical variables.

Correlation

Based on the literature review we conducted, we expected to see a strong correlation between unemployment and inflation, but we found that the strongest correlation with unemployment was with the year variable. Unemployment is decreasing from 2012 to 2018 creating a negative correlation. As seen in Table 2, even the percentage of broadband per household had a stronger correlation with unemployment than inflation. As the unemployment rate is increasing, broadband and inflation are decreasing - both of which were expected. What we did not expect was to see that year was a stronger determinant than both broadband and inflation.

Table 2. Correlations

	Unemployment Rate %	p-value
Unemployment Rate %	1	
Year	-0.57478	0.0001
Percent of Population Below the Poverty Line	0.38735	0.0001
Median Household Income	-0.32587	0.0001
% Broadband Available Per Household	-0.31428	0.0001
% of People by County Population	-0.19628	0.0001
Inflation	-0.18926	0.0001
Population of African Americans	0.12855	0.0001
Total Population Per County	0.12418	0.0001
Labor Force Amount Per County	0.10876	0.0001
Median Age by County	-0.09847	0.0001
GDP By County	0.07301	0.0001

Regression

We begin with conducting SLR and then MLR. Within each table, the coefficient "b" is noted in the first numeric column indicating the regression equation. We also list the standard regression weight (beta), the squared semi-partial correlation, as well as the VIF as indicators of the models.

3 SLRs.

If we were to have run just one SLR, we would have chosen the year IV as it had the strongest correlation with the DV (unemployment). And although we did this, we felt it prudent to also analyze broadband and inflation to compare the impacts of our focus variables.

Year. This IV had the strongest correlation with the DV so it should also produce the strongest SLR. As seen in Table 3, the result was a moderate strength model as the Adj R^2 was only 0.3303.

Table 3. Year SLR

Predictor	b	95% Confidence Limits LL	UL	Pr > |t|	*beta*	sr^2	VIF
Intercept	1,332	1,286	1,377	<0.0001	0		0
Year	-0.66	-0.681	-0.635	<0.0001	-0.575	0.330	1

Fit statistic: Adj R^2 = 0.3303

Note: beta is the standard regression weight. sr2 is the squared semi-partial correlation Type I.

Broadband. Our focus variable did not have quite as strong of a relationship with the DV, but it did produce a viable model as noted in Table 4. The model strength was not as strong as the Year SLR model, but it was also a moderate model with an Adjusted R^2 of 0.0986.

Table 4. Broadband SLR

Predictor	b	95% Confidence Limits LL	UL	Pr > |t|	*beta*	*sr^2*	VIF
Intercept	11.8870	11.3863	12.3877	<0.0001	0	0	0
Broadband	-8.1659	-8.7609	-7.5710	<0.0001	-0.3143	0.0988	1

Fit statistic: Adj $R^2 = 0.0986$
Note: beta is the standard regression weight. sr2 is the squared semi-partial correlation Type I.

Inflation. When we completed our literature review we believed that this IV would be our strongest predictor and it was not. It was worse than our new IV (broadband). When looking at the results of the SLR in Table 5, it is also evident that the model strength is weak with an Adj R^2 of 0.0357.

Table 5. Inflation SLR

Predictor	b	95% Confidence Limits LL	UL	Pr > |t|	*beta*	*sr^2*	VIF
Intercept	9.4676	8.912	10.0232	<0.0001	0	0	0
Inflation	-2.2921	-2.5789	-2.0052	<0.0001	-0.1893	0.0358	1

Fit statistic: Adj $R^2 = 0.0357$
Note: beta is the standard regression weight. sr^2 is the squared semi-partial correlation Type I.

MLR.

Our final numerical MLR included 9 of the 11 variables that we began the model with. We used backward elimination to remove variables with an alpha of 0.05. The only two variables removed were the total population per county and the percent of broadband availability as noted in Table 6. So where we were able to create a strong model with an Adj R^2 of 0.6631, our newly introduced variable (broadband) was not a contributing IV to the model. Broadband was removed due to a large p-value, yet inflation stayed in the model with several other IVs that had a weaker correlation to the DV.

Table 6. Numerical MLR

Predictor	b	95% Confidence Limits LL	95% Confidence Limits UL	Pr > |t|	*beta*	*sr^2*	VIF
Intercept	1644.043	1598.924	1689.162	<.0001	0		0
Year	-8.17E-01	-0.8393	-0.7943	<.0001	-0.714	0.33	1.5
Labor Force Amount	1.56E-06	0	0	<.0001	0.1826	1E-04	19
Median HH Inc	4.07E-05	0	0	<.0001	0.25	0.114	4.5
Median Age County	5.62E-02	0.0486	0.0637	<.0001	0.1429	3E-04	1.4
% of People County	-6.66E-02	-0.0719	-0.0613	<.0001	-0.312	0.027	2.4
% of Population	1.80E-01	0.1694	0.1899	<.0001	0.4827	0.077	2.9
Population of AA	1.22E-06	0	0	0.0015	0.0409	6E-04	2.5
Inflation	1.16E+00	0.9457	1.3755	<.0001	0.0958	0.008	1.2
GDP By County	-1.06E-08	0	0	<.0001	-0.16	0.001	18

Fit statistic: Adj R^2 = 0.6631

Note: beta is the standard regression weight. sr^2 is the squared semi-partial correlation Type I.

At this point, we had four viable models. The strongest of these models (Table 7) was the MLR with nine of the numerical IVs. To try and improve on this, we attempted to add geographical categorical variables.

Table 7. Numerical regression results

		Adj R^2
SLR	Year	0.3303
	Broadband	0.0986
	Inflation	0.0357
MLR	9 Ivs	0.6631

Geographic Variables

We collected data from all the US Census Bureau's 9 Divisions, from 40 US states, and 1,268 counties for all 6,608 entries. Some states did not report at the county level, so we could not use that data. We first tested our data to comply with linear regression requirements. We then tested the divisions, states, and counties individually in three models before combining them.

One-way ANOVA

We began by analyzing the Division, State, and County separately to see if they statistically impact our DV. Each of these models showed that the three variables were statistically significant to predict unemployment with $p <= 0.001$. But the F-test does not tell us which variables in those categories are significant. Had any not been significant, then we could conclude that there was no impact, and then regression would not be needed.

Our next approach to reduce the sample size was to eliminate categories with each of the variables (Division, State, County) based on a sample size of less than 30 (for normality assumptions). For those categories with sample sizes greater than 30, we used the SAS results from the One-Way ANOVA to see if the box-whisker plots revealed any categories that were statistically significantly different than others. This test was not conclusive, but it did reduce the number of states and counties to consider as noted in Table 8 (N≥30) due to the sample sizes.

Table 8. Sample size remaining

	N >=30
Division	all 9
State	32 of 40
County	20 of 1268

Second One-Way ANOVA

We did not do this for Divisions because all of them passed the sample size tests.

We revisited the One-Way ANOVA with two approaches. The first approach recoded the State and County so that any State or County that had a small sample size was recorded at Other. The second approach consisted of making two data sub-sets where the data was reduced by eliminating (deleting rows) any State (first data sub-set) and then County (second data sub-set) where that category had a small sample size.

From this, we conducted four One-Way ANOVAs to test if State and/or County were still significant when predicting unemployment (including the other variable or deleting the other variable). They were significant in all cases with $p <= 0.001$. Indicating that regardless of how we parsed it, there were some states and counties that might influence our DV.

MLR for variable reduction

Before conducting these MLRs, we created new binary variables for each of the categories still in our model. We then ran separate MLRs for Division, State, and County. By doing this we were able to eliminate many more IVs as noted in Table 8's individual columns.

Wanting to see if we could further reduce the number of geographic variables, we ran one more MLR with the Division, State, and County IVs and were able to further reduce our IVs as noted in Table 9's last two columns. Of note is that after we eliminated for multicollinearity using the sign, VIF, and proportion of variation, we had a stronger model with the states by themselves, but when combined with the divisions and counties, two of those states were eliminated.

Table 9. Geographical MLR results.

	Individual Geographic Models		Combined Geographic Model	
	#	Adj R2	#	Adj R2
Divisions	6	0.1758	2	
States	25	0.3207	23	0.3197
Counties	5	0.0470	1	

MLR with geographic and numerical variables

Having established our baseline model with an Adj R^2 of 0.6631 and a combined geographic model with an Adj R^2 of 0.3197 we looked at combining these variables in an attempt to get a stronger model where geographic variables could better help to predict unemployment.

MLR with all geographical variables that survived the sample size test

Our next regression was to build a model with the geographic variables that passed the sample size test along with all eleven of our numerical variables. As seen in Table 10, we ended up with a strong model. The categorical variables were a subset of those that survived the categorical regression reduction with no Divisions, 13 States, and no Counties. Whereas we thought Inflation would be in our model as it was noted in prior works, it was not, but your new numerical variable (broadband) was in the model.

Table 10. Final MLR with categorical variables N>30

Predictor	b	95% Confidence Limits LL	UL	Pr > |t|	*beta*	*sr²*	VIF
Intercept	1,192	1,155	1,231	<.0001	0	.	0
Year	-0.582	-0.601	-0.564	<.0001	-0.509	0.209	1.241
% Broadband Avail	-1.700	-2.126	-1.275	<.0001	-0.065	0.003	1.242
Labor Force Amount	-1.37E-06	-1.58E-06	-1.15E-06	<.0001	-0.160	0.009	2.896
Percent of Population	0.128	0.122	0.134	<.0001	0.345	0.096	1.239
Population of African	2.98E-06	2.28E-06	3.69E-06	<.0001	0.100	0.004	2.609
Arizona	2.982	3.271	2.693	<.0001	-0.155	0.023	1.053
California	2.819	2.980	2.659	<.0001	-0.285	0.067	1.219
Connecticut	1.392	1.753	1.032	<.0001	-0.057	0.003	1.023
Illinois	1.424	1.576	1.273	<.0001	-0.140	0.019	1.038
Iowa	-0.747	-0.582	-0.911	<.0001	0.068	0.004	1.052
Louisiana	0.518	0.699	0.338	<.0001	-0.043	0.002	1.049
Michigan	1.193	1.341	1.044	<.0001	-0.120	0.014	1.032
Nebraska	-0.613	-0.432	-0.795	<.0001	0.051	0.002	1.058
Ohio	0.838	0.977	0.699	<.0001	-0.090	0.008	1.040
Oregon	0.910	1.125	0.695	<.0001	-0.063	0.004	1.022
Pennsylvania	1.124	1.273	0.976	<.0001	-0.115	0.012	1.062
Washington	1.761	1.960	1.562	<.0001	-0.132	0.017	1.032
West Virginia	1.196	1.410	0.982	<.0001	-0.083	0.007	1.024

Fit statistic: Adj R2 = 0.6292 Note: beta is the standard regression weights. sr^2 is the squared semi-partial correlation Type I.

MLR with all geographical variables that survived the MLR reduction modeling.

Since the previous model only included categorical variables that survived the MLR reduction, we chose to run a model that began with only these IVs and the numerical IVs and not all the geographical IVs that survived the MLR reduction.

As seen in Table 11, this MLR included fewer of our numerical variables, and it included our new variable (broadband) but not inflation. It also included an additional two States, but no Divisions or Counties. Compared to the previous model, this model was also not as strong as the previous model indicating that reducing the Divisions, States, and Counties before running this model was not a good idea (Table 12).

Table 11. MLR with categorical variables N>30 and survived categorical MLR

Predictor	b	95% Confidence Limits LL	95% Confidence Limits UL	Pr > |t|	*beta*	sr^2	VIF
Intercept	1069	1028	1110	<.0001	0	.	0
Year	-0.513	-0.534	-0.493	<.0001	-0.449	0.330	1.196
% Broadband Available	-4.324	-4.779	-3.869	<.0001	-0.166	0.018	1.156
Alabama	1.177	-1.546	-1.272	<.0001	-0.175	0.005	1.080
Arizona	3.755	-1.403	-1.102	<.0001	-0.140	0.002	1.068
California	3.103	-1.360	-1.049	<.0001	-0.130	0.001	1.068
Connecticut	1.083	-1.330	-1.003	<.0001	-0.120	0.001	1.076
Florida	1.166	-1.331	-1.002	<.0001	-0.119	0.001	1.066
Georgia	1.409	-3.271	-2.936	<.0001	-0.314	0.059	1.082
Illinois	1.708	-1.858	-1.524	<.0001	-0.170	0.013	1.056
Louisiana	1.733	-1.877	-1.538	<.0001	-0.168	0.015	1.051
Michigan	1.691	-1.369	-0.986	<.0001	-0.102	0.004	1.042
North Carolina	1.252	-1.933	-1.534	<.0001	-0.144	0.013	1.036
Ohio	1.204	-2.263	-1.821	<.0001	-0.153	0.017	1.035
Oregon	1.452	-2.293	-1.818	<.0001	-0.143	0.017	1.026
Pennsylvania	1.166	-1.692	-1.213	<.0001	-0.100	0.008	1.027
Washington	2.042	-4.071	-3.439	<.0001	-0.196	0.037	1.022
West Virginia	2.056	-1.481	-0.685	<.0001	-0.045	0.002	1.016

Fit statistic: Adj R2 = 0.5441 Note: beta is the standard regression weights. sr^2 is the squared semi-partial correlation Type I.

Table 12. Categoricals in ALL MLR w/ multicollinearity

	MLR N≥30	MLR Geog Reduced
Division	0	0
State	13	15
County	0	0
Numerical	5	2
Adj R^2	0.6292	0.5441

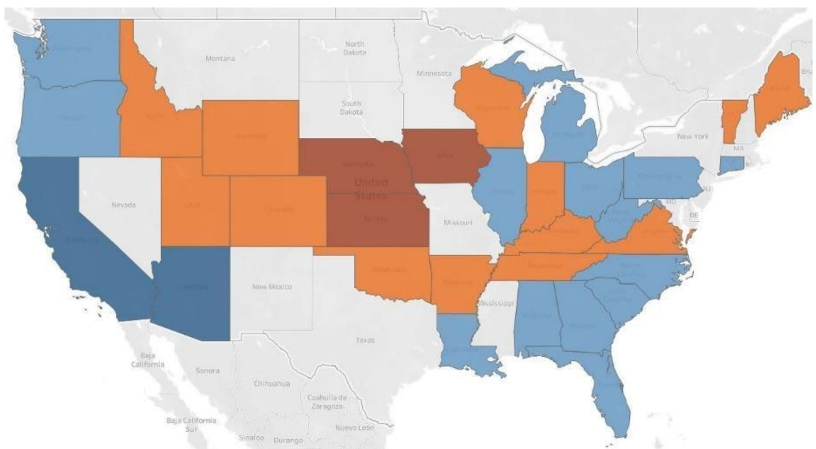
Geographic pattern analysis

Once the geographic variable categories were converted into binary variables, we were able to look at their influence on our DV by using a correlation matrix. We mapped this using the strength of the correlation as well as looking at what variables survived the different regressions. The motivation of this research was to see if broadband impacted unemployment when measuring it at different geographic levels: Division, State, and County. In addition to analyzing these geographic variables in regression with our other variables, here we took a look at geographical differences using Tableau's mapping features without other IVs.

Positive or Negative Influencers.

We could not determine a distinct pattern when looking at divisions' or counties' correlation with unemployment, but when we looked at the states, we saw that there was a stronger negative correlation (red) with states in the center of the US and generally a stronger positive (blue) correlation with states on the West or Southeastern coasts (Figure 1). A positive correlation means that those states were more likely to have more unemployment. Since data is recorded at the county level, this correlation although depicted for each State is measuring correlations for each county within the state with the DV.

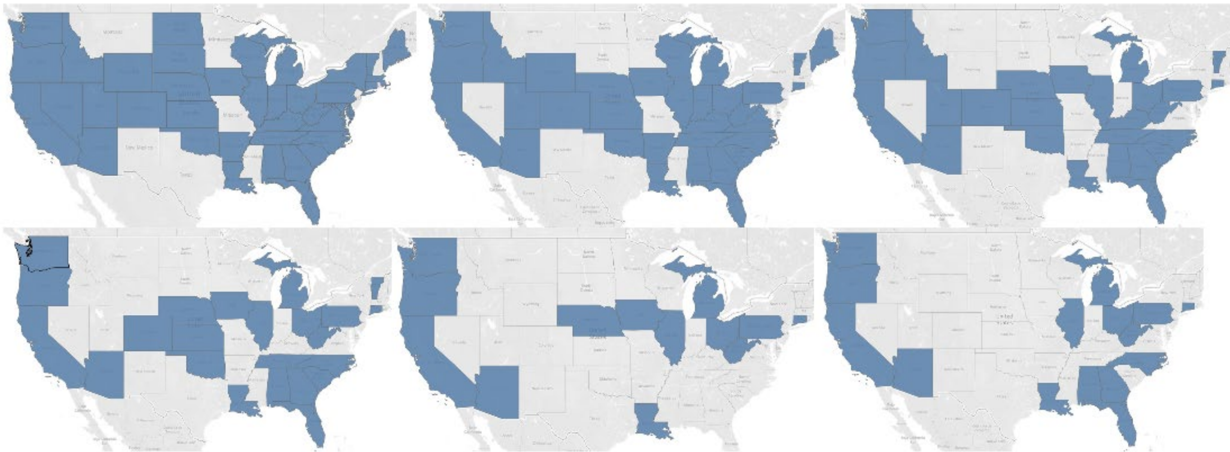
Figure 1. State correlation with unemployment



Geography and Regression.

We began with 40 US States as seen in the far left of Figure 2. The next image is of the states that met the sample size criteria (32 states). As we see on the top right, only 25 of these states were in the regression model with just states and unemployment. The bottom row depicts the 23 states that were in the regression model with Divisions, States, and Counties. The middle bottom image is the model with MLR when the 2 Divisions, 23 States, and 1 county were modeled with the numerical IVs. The final map shows the MLR when all geographic regions that passed the sample size criteria were modeled with the numerical IVs. In general, we saw that the states that had the most impact on unemployment were those on the west coast, the Great Lakes region, and the Gulf/southeast of the US. When comparing these states with the correlation map in ***Figure 1***, those were also the ones with the most positive (blue) correlation with unemployment as shown in Figure 1. This means those states with the most influence on unemployment are also the ones with a stronger positive correlation to unemployment.

Figure 2. States and Regression



A summary of the geographic variable reduction from each of the modeling is summarized in Table 13. They are also the same maps reflected in the State maps in Figure 2, but here you can also see how in each of these runs which ones also had Divisions and Counties in the models.

Table 13. Geographic variables

	N >=30	Individual Geographic Models	Combined Geographic Model	MLR Geog Reduced	MLR N≥30
Division	all 9	6	2	0	0
State	32 of 40	25	23	15	13
County	20 of 1268	5	1	0	0

CONCLUSION

The advancement of broadband availability and speed are improving every day and prior research data is obsolete as soon as it is completed due to this continued advancement. Thus this research and the impact of broadband is an ongoing research area of interest. Prior research did not look at the granularity of broadband's impact at the county level. This research updates past research with newer data, data at the US County level, and adds broadband availability in modeling unemployment.

Insights

There are many reasons to promote more access to broadband internet such as for entertainment, social connection, or decreasing unemployment. Some of the motivation for this research was due to the potential of broadband internet use from both businesses and homes impacts on the economy and ties to impacts on inflation. This drove our focus on independent variables of year (over time), broadband availability, and inflation as predictors of unemployment. When running an SLR, the year was the best predictor of unemployment, and broadband availability was a better predictor than inflation. But when running an MLR, both year and inflation remained in the model with other variables, but broadband did not. The implication was that time itself was a better predictor of unemployment than either broadband availability or inflation. We note however that over time (as time increased) broadband availability also increased where as inflation fairly stagnated during this time.

One major addition of this research was the granularity of our data at the county level. We note that unemployment takes on a graphical nature and in different areas (counties) there are varying unemployment rates as well as different levels of broadband access. When we introduced our geographic variables into the regression models, the models did not improve, but our new variable of broadband did remain in the model eliminating all the other numeric variables other than year. Thus, when attempting to predict unemployment some geographic variables did remain in the model as predictors and both year and broadband access did as well. This led us to believe that geography and broadband are influencers, but an SLR model with just year is a simpler and more accurate model.

When considering the impact of geography on unemployment we first attempted to use the county-level data but quickly found that the granularity was not practical. Using States, Divisions, and Regions (as defined by the US Census Bureau) did have an impact on the model. Taking a deeper look at these geographic variables we found a striking result. The geographic variables that did impact the regression model were also the geographic variables that had positive correlations with unemployment (more unemployment). Those that had negative correlations with unemployment (more employment) did not. Looking back on the previous paragraph, noting that the numerical variables remaining to predict unemployment were year and broadband access and combining it with the understanding that the geographic features remaining were those predicting more unemployment leads us to want to conduct more research on those areas with the most unemployment over time. They are the areas in the greatest need of job growth.

Overall, what we saw was that broadband availability does impact unemployment when considering county-level data at the State level. If this data is collected again in the future, we should see a direct impact of the recent US investment in broad expansion.

Future Work

This study is limited by only considering 2012 to 2018 data. Obtaining data before 2010 as well as obtaining more current data could strengthen this research into the impacts of broadband on unemployment. During this period, the inflation rate has been relatively small, however, before 2010 as well as in 2021 the inflation rate has shown significant changes. Expanding the time frame of the study could show a larger impact on inflation. But this also has to be weighed with the changing nature of broadband with 4G being introduced in 2011. If studies included data before 2011, they might also need to classify the nature of the broadband availability as 2G, 3G, 4G, or in the future 5G.

An expansion of the time frame would also show significant differences in broadband access. Although access to the Internet by general US citizens began in the 1980s (A Brief, n.d.), access to broadband in US homes did not begin until 2000. (Internet, 2021) This study only covers half the time when we have had access to broadband in US homes. Expanding this study back to 2000 could also shed insight into the impacts of broadband on unemployment as there was more variability in unemployment rates before 2012. This study ends in 2018 which was when the US government began investing in expanding broadband (USDA, 2018). Obtaining more current data might also show the impacts of the more recent explosion of broadband investment and the impact on unemployment. However, this needs to be carefully analyzed with the impacts of many people working from home during the COVID pandemic.

Lastly, future work should take a specific focus on those areas with the highest unemployment rates as this research showed that those geographic regions had more influence on predicting unemployment than those areas with smaller unemployment rates. The proliferation of broadband may have more impacts on these areas than those that already have low unemployment rates.

REFERENCES

A Brief History of the Internet. (n.d.). *Online Learning Library.* Retrieved December 5, 2002, from https://www.usg.edu/galileo/skills/unit07/internet07_02.phtml.

Adelowokan, O. A., Maku, O. E., Babasanya, A. O., & Adesoye, A. B. (2019). Unemployment, poverty, and economic growth in Nigeria. *Journal of Economics and Management, 35*, 5–17. https://doi.org/10.22367/jem.2019.35.01

Aliyev, K. (2021). Unemployment and (un)happiness: Life satisfaction approach to enhance policy efficiency for developing countries. *Journal of International Studies, 14*(4), 220–235. https://doi.org/10.14254/2071-8330.2021/14-4/15

Barrero, J. M., Bloom, N., & Davis, S. J. (2021, July 1). Internet Access and its Implications for Productivity, Inequality, and Resilience. *National Bureau of Economic Research*. https://www.nber.org/papers/w29102

Bitetti, R. (2018). Unemployment and entrepreneurship: a bottom-up analysis. *SOCIOLOGIA DEL LAVORO, 149*, 94–117. https://doi.org/10.3280/sl2018-149007

Dahliah, D., & Nirwana Nur, A. (2021). The Influence of Unemployment, Human Development Index and Gross Domestic Product on Poverty Level. *Golden Ratio of Social Science and Education, 1*(2), 95–108. https://doi.org/10.52970/grsse.v1i2.84

FRED Economic Data, n.d. *Economic Research Federal Reserve Bank of St. Louis.* Retrieved September 23, 2022, from https://fred.stlouisfed.org/.

Hasbi, M. (2017, June 24*). Impact of very high-speed broadband on local economic growth: Empirica.* 14th ITS Asia-Pacific Regional Conference, Kyoto 2017: Mapping ICT into Transformation for the Next Information Society. Retrieved September 1, 2022, from https://ideas.repec.org/p/zbw/itsp17/168484.html

Hatef, E., Ma, X., Shaikh, Y., Kharrazi, H., Weiner, J. P., & Gaskin, D. J. (2021). Internet Access, Social Risk Factors, and Web-Based Social Support Seeking Behavior: Assessing Correlates of the "Digital Divide" Across Neighborhoods in The State of Maryland. *Journal of Medical Systems, 45*(11). https://doi.org/10.1007/s10916-021-01769-w

Internet/Broadband Fact Sheet. (April 7, 2021). *The Pew Research Center*. Retrieved December 5, 2022, from https://www.pewresearch.org/internet/fact-sheet/internet-broadband/.

Isley, C., & Low, S. A. (2022). Broadband adoption and availability: Impacts on rural employment during COVID-19. *Telecommunications Policy, 46*(7), 102310. https://doi.org/10.1016/j.telpol.2022.102310

Kolko, J. (2010, January). Does Broadband Boost Local Economic Development? - *ppic.org*. Retrieved September 15, 2022, from https://www.ppic.org/wp-content/uploads/content/pubs/report/R_110JKR.pdf

Lobo, B. J., Alam, M. R., & Whitacre, B. E. (2019, June 4). Broadband speed and unemployment rates: Data and measurement issues. *Telecommunications Policy*. Retrieved from ttps://www.sciencedirect.com/science/article/abs/pii/S0308596118303823

National Telecommunications and Information Administration. *NTIA 2021* https://broadbandusa.ntia.doc.gov/resources/federal/federal-funding

Quinton, J. K., Ong, M. K., Vangala, S., Tetleton-Burns, A., Webb, A., Sarkisian, C., Casillas, A., Kakani, P., Han, M., & Pirtle, C. J. (2021). The Association of Broadband Internet Access and Telemedicine Utilization in Rural Western Tennessee: an observational study. *BMC Health Services Research*, 21(1), NA. https://link.gale.com/apps/doc/A675240763/HRCA?u=anon~b4379ed0&sid=googleScholar&xid=ada51b6f

Rohman, I. K., & Bohlin, E. (2012, April 4). Does Broadband Speed Really Matter for Driving Economic Growth? Investigating OECD Countries. *Papers.ssrn.com.* https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2034284

The Center on Technology, Data, and Society. (n.d.) *Arizona State University.* Retrieved September 23, 2022, from https://techdatasociety.asu.edu/broadband-data-portal/home

"The World in 2011: ITC Facts and Figures" Archived 2012-05-10 at the Wayback Machine, *International Telecommunication Union (ITU*), Geneva, 2011

UCLA Advanced Research Computing Statistical Methods and Data Analysis. (n.d.) Choosing the Correct Statistical Test. Retrieved August 15, 2021, from https://stats.idre.ucla.edu/other/mult-pkg/whatstat/

USDA Launches New Program to Create High-Speed Internet e-Connectivity in Rural America. (December 13, 2018). *USDA*. Retrieved from https://www.usda.gov/media/press-releases/2018/12/13/usda-launches-new-program-create-high-speed-internet-e-connectivity.

Whitacre, B., & Gallardo, R. (2020). State broadband policy: Impacts on availability. Telecommunications Policy, 44(9), 102025. https://doi.org/10.1016/j.telpol.2020.102025

Why Broadband? (n.d.) *The World Bank*. Retrieved December 6, 2023, from https://ddtoolkits.worldbankgroup.org/broadband-strategies/building-broadband/why-broadband#:~:text=Although%20numerous%20studies%20have%20found,1.50%20percent%20(Figure%201.3).

QRBD

Quarterly Review of Business Disciplines

May 2023

Volume 10
Number 1

A Journal of International Academy of Business Disciplines
Sponsored by University of North Florida
ISSN 2334-0169 (print)
ISSN 2329-5163 (online)