Evaluation of USDA Broadband Loan Programs: Modeling Broadband Diffusion Across 1999 to 2008*

Robert Dinterman†

November 23, 2015

Abstract

This paper utilizes ZIP code level data on the loan programs in order to accomplish two main goals: 1) finding a policy instrument to further help determine the value of broadband usage and 2) determining how effective the US government is in implementing public investment. The first motivation for this research determines the gains to a region for investing in broadband infrastructure that are unrelated to the natural diffusion of broadband infrastructure. Increases in broadband availability for regions suffers from causality issues without a policy instrument that controls for the natural rate of broadband diffusion. One needs to identify whether broadband increases are due to lower costs in a particular area or higher anticipated revenue streams for a particular reason and the broadband loan programs serve as this instrument. The second motivation evaluates the effectiveness of the broadband loan program, which describes government spending and allow us to identify increases in broadband availability that is separate from market forces that obfuscate the causal impact of broadband for a region. Results indicate that ZIP codes receiving broadband loans experienced approximately 0.092 additional broadband providers annually and that these benefits accrued more towards rural areas than urban areas.

JEL Codes: H81, C22, O33, R11, P25
Keywords: broadband, diffusion, rural, loan evaluation

*__Working paper. Please do not distribute or cite without permission.__

†Email: rdinter@ncsu.edu. Department of Agricultural and Resource Economics, North Carolina State University.

1 Motivation

Economists have long understood the relationship between technology and human capital whereby advances in technology generally lead to increases in the level of human capital for a region. The crux of this relationship is to identify whether new technologies signify an advancement in the current technological level and to measure the gains in productivity from the subsequent increases in human capital. For new technologies such as broadband, it is imperative to determine if this new technological advancement has led to an increase in human capital for its users and attempt to measure this increase in productivity for a region. Previous research on the productivity gains associated with broadband access and use has suffered from two important issues: 1) separating broadband deployment from general increases in economic conditions for a region and 2) identifying user adoption of broadband connections in measuring increases in productivity for a region. This paper aims to address the first issue while the subsequent paper will expand upon this analysis to address the second issue.

Broadband is seen as an information and communications technology (ICT) that is made up of infrastructure which delivers services to end users. As a technology, previous research has indicated that emerging technologies follow a diffusion process to the population it is intended to serve that follows an S-Curve (Attewell, 1992; Geroski, 2000). Previous research indicates the natural diffusion path of broadband availability is determined through economic forces that induce broadband providers to serve the areas with highest expected profit (high revenues and/or low costs) first (Whitacre and Mills, 2007; Whitacre, 2010; Czernich et al., 2011). If broadband diffusion only follows these natural economic forces, then evaluation of its impact on a region cannot disentangle whether broadband increased economic well-being or if expected increases in economic well-being caused broadband providers to better serve the area. The dynamic relationship can be unfurled through an exogenous policy instrument which is related to broadband deployment but unrelated to economic conditions that determine broadband's deployment and effectiveness.

This paper utilizes data on two broadband loan programs administered by the United States Department of Agriculture (USDA) in order to accomplish two main goals: 1) finding a policy instrument to further help determine the value of broadband usage and 2) determining how effective the US government is in implementing public investment. The first motivation for this research determines the gains to a region for investing in broadband infrastructure that are unrelated to the natural diffusion of broadband infrastructure. Increases in broadband availability for regions suffers from causality issues without a policy instrument that controls for the natural rate of broadband diffusion. One needs to identify whether broadband increases are due to lower costs in a particular area or higher anticipated revenue streams for a particular reason and the broadband loan programs serve as this instrument. The second motivation evaluates the effectiveness of the broadband loan program, which describes government spending and allow us to identify increases in broadband availability that is separate from market forces that obfuscate the causal impact of broadband for a region.

The econometric techniques in this paper involve the use of discrete choice methods and count panel estimators (and still trying to figure out the spatial component). Discrete choice method involves use of a logit estimator in order to determine the factors that drove the disbursement of the broadband loans across the United States from 2002 to 2006. The count panel methods are used to model the diffusion of broadband across time and to attribute the increase in broadband availability caused by the broadband loan program. By using ZIP code level Federal Communications Commission (FCC) data from December 1999 to June 2008, the changes in broadband availability due to the broadband loan program is identified. Results indicate that ZIP codes receiving broadband loans experienced approximately 0.092 additional broadband providers annually and that these benefits accrued more towards rural areas than urban areas.

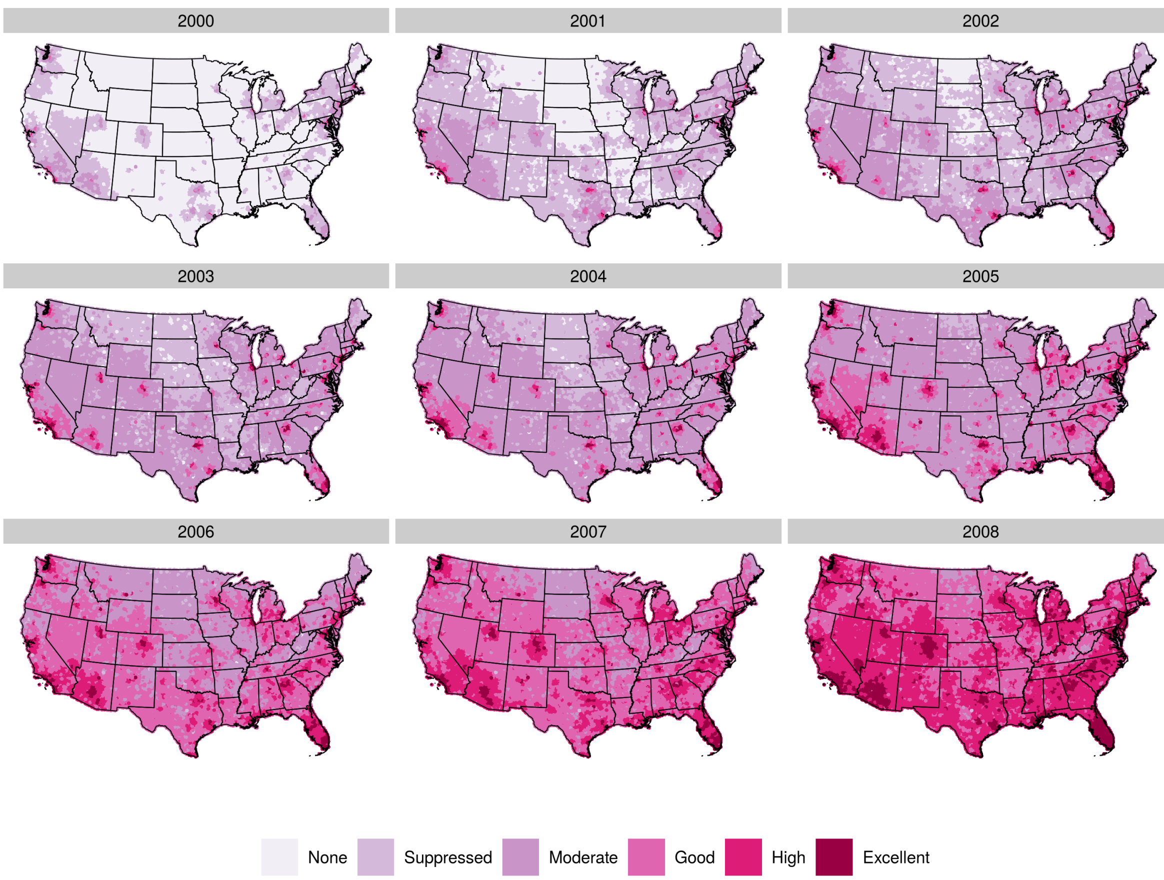
The paper continues as follows. I describe the broadband loan program in section 2. Next, the data are described in section 3. Then methods and results for evaluating disbursement of the loans are described in section 4. Section 5 estimates the impact of the loan programs on number of providers by use of basic panel methods as well as count panel methods. In section 6, I assess the validity of using spatial methods for estimating the impact of the loan program. And finally, the paper concludes with a short discussion in section 7.

2 Broadband Loan Program

The initial roll-out phase of broadband technologies across North America began around 1996 when Rogers Communications introduced the first cable modem service in Canada (FCC, 2005). The

United States quickly followed although the diffusion of broadband technologies was not instantaneous. As seen in figure 1, around 2000 there were many areas within the U.S. which lacked access to broadband. After 2008, the only areas in the United States lacking access to broadband technology were extreme rural areas with limited economic activity. The early adopters of broadband across the United States were areas with favorable economic conditions: densely populated metro areas with high levels of median income.

Figure 1: US Broadband Availability from 2000 to 2008

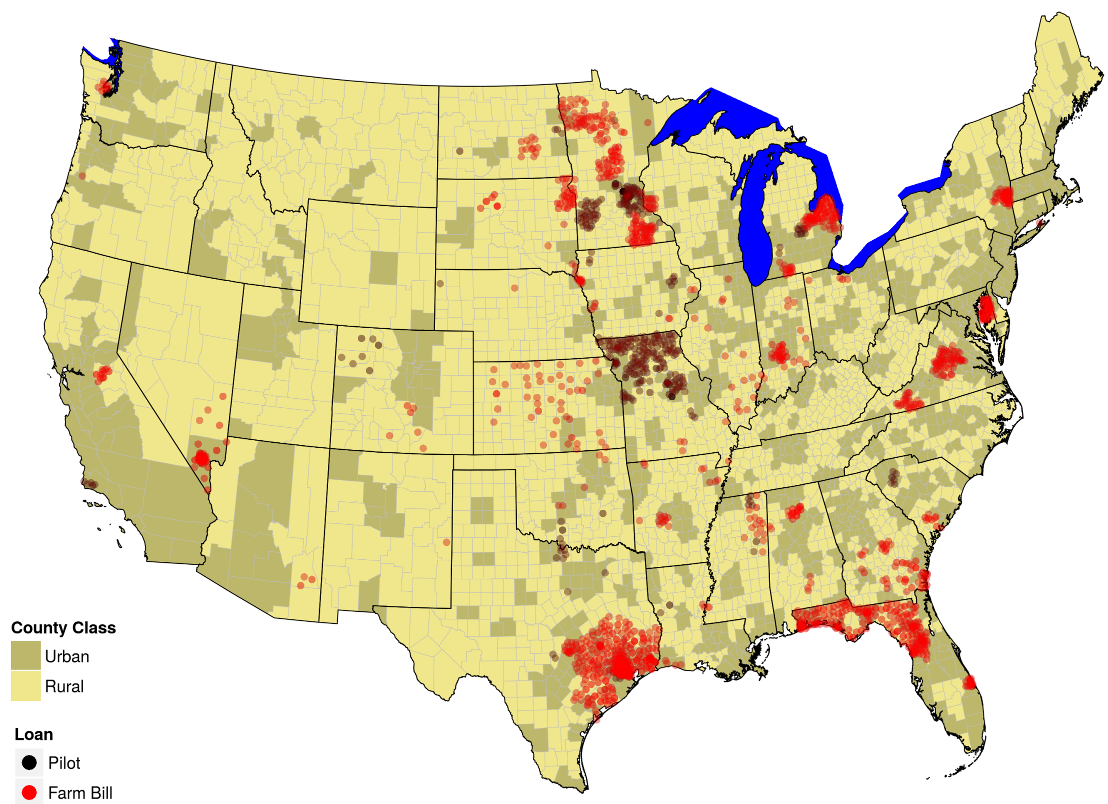


A simple kriging procedure was used using ZIP code level FCC data on number of providers with the suppressed values coded as 2 (Cressie, 1993). The predicted values were then translated to None (0 to 2), Suppressed (2 to 3), Moderate (3 to 5.5), Good (5.5 to 7.5), High (7.5 to 10) and Excellent (more than 10).

A troubling issue at the turn of the milenia centered around the pending digital-divide within the United States which existed as differences broadband access between rural and urban areas

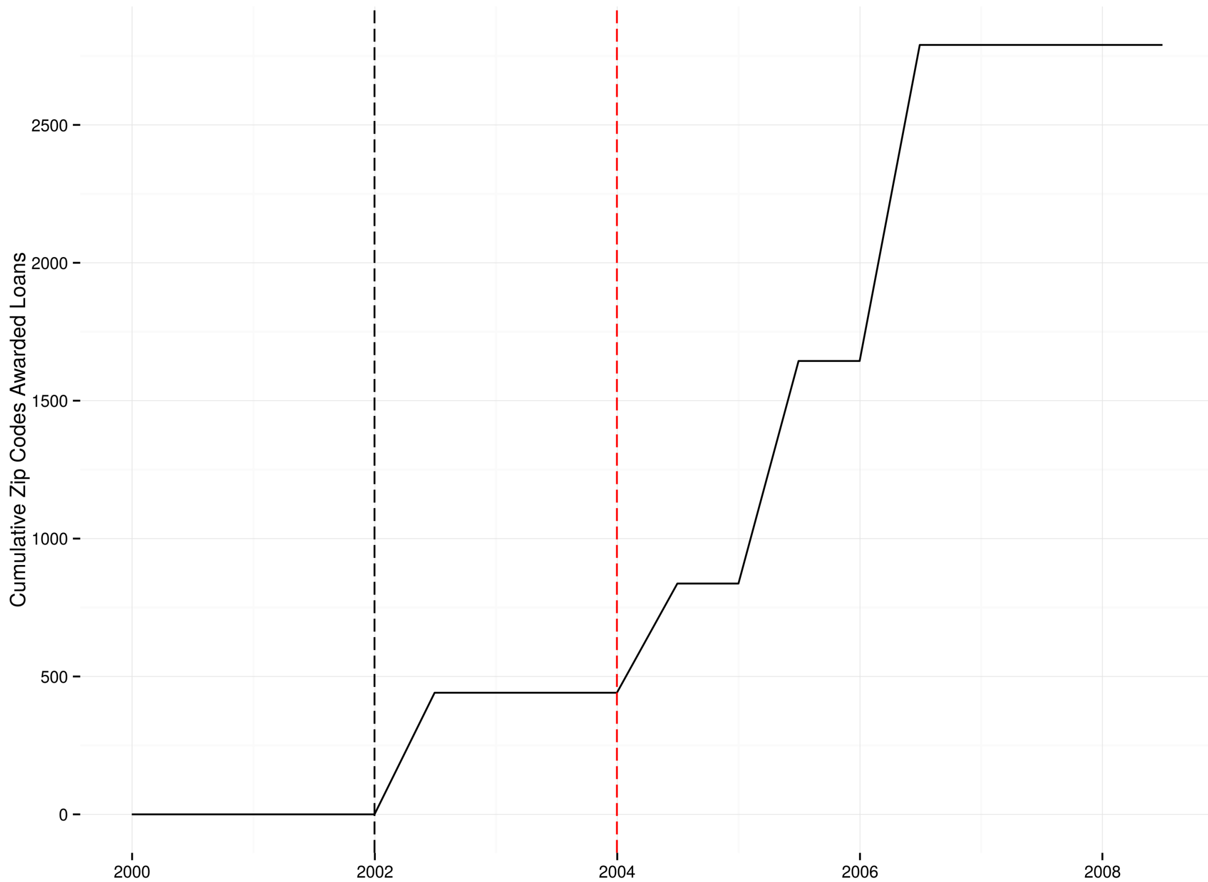
(Norris, 2001; DiMaggio et al., 2001; Wade, 2002; Antonelli, 2003). Rural areas consistently lagged urban areas in terms access to high-speed internet, which was thought to be a critical component to economic development in the changing US economy. This growing concern spurred government policies aimed at reducing the difference between regional broadband access. In December 2000, Congress authorized a broadband pilot program (Pilot) to help expand broadband access in underserved rural communities through subsidized loans. Program eligibility criteria included having a population of 20,000 or fewer, having no prior access to broadband, and providing a minimum matching contribution of 15 percent by recipients of the loan. Loans were extended mainly to small telecommunications services firms at varying (subsidized) interest rates; most participating communities qualified for a "hardship rate" of 4 percent (Cowan, 2010).

Figure 2: Zip Codes Receiving Broadband Loans



The Pilot program was administered by the United States Department of Agriculture's (USDA)

Figure 3: USDA Loans by Zip Code Across Time



Black dotted line indicates the Pilot Program while Red dotted line indicates the Current Program.

Rural Utilities Service (RUS) as a way to gauge interest and effectiveness for a larger scale program. In the first year, there were 12 loans worth $100 million disbursed. The Agricultural Appropriations Act of 2002 allowed for a second round of funding for 16 loans of $80 million to complete the total funding of the program. Overall, the Pilot program totaled $180 million in loans to 98 communities located in 13 states. After the Pilot program, the 2002 Farm Bill established the USDA Rural Development Broadband Program that expanded upon the scope of the Pilot program. By 2007, the program had approved 70 loans in 40 states, totaling over $1.22 billion serving 1,263 communities and 582,000 households (Development, 2007).

Although loans were explicitly targeted for rural areas, some of the loans went to urban areas. Presumably, the urban loans were either because of geographical constraints of existing broadband infrastructure or because of a lack of shovel-ready projects in rural areas. Data on loans were obtained through RUS at the ZIP code level and the year the loans were disbursed. Table 1 displays the distribution of ZIP codes across the rural-urban divide. Compared to the rest of the United States, both the Pilot and Farm Bill programs disbursed a higher number of loans to rural ZIP codes, although there were still more than 30% of the loan recipients located in urban ZIP codes. Data are not available for the locations of rejected loan applicants, although it is known for the Farm Bill that there were 76 applicants in 2003, 39 in 2004, 37 in 2005, and 33 in 2006.

Table 1: County Classification by Loan Type

Loan	Metro	Rural Adjacent	Rural Non-Adjacent
Pilot	35.4 %	26.8 %	37.8 %
Farm Bill	48.5 %	37.1 %	14.4 %
None	54.2 %	26.1 %	19.7 %

A concern for both the Pilot and Farm Bill programs relates to an overly broad definition of what constitutes a "rural" community. For example, a 2005 audit by the USDA's Inspector General chided RUS for having extended nearly 12 percent of total loan funding to suburban communities located near large cities. A follow-up audit found that this situation was not remedied, noting that between 2005 and 2008 broadband loans were extended to 148 communities within 30 miles of cities with populations greater than 200,000 - including Chicago and Las Vegas (Kruger, 2013). [1]

The primary intent of the USDA broadband loan program is to extend loans to broadband providers in order to increase broadband availability that will affect households and firms only after the appropriate channels flow to the end consumer. These loans could result in an increase in a providers coverage, the number of providers, or an increase in the quality of broadband, which may not be completely measurable with the available data. Within the context of evaluating the success of the USDA broadband loan program, the only available data on increases in broadband availability is through the number of providers that serve a ZIP code. The following section details the data used in this paper.

[1] Our data do not contain any loans disbursed around Chicago. This could be an error in our data or from the previous report as Houston appears to be a large city with loans disbursed around the area.

3 Data

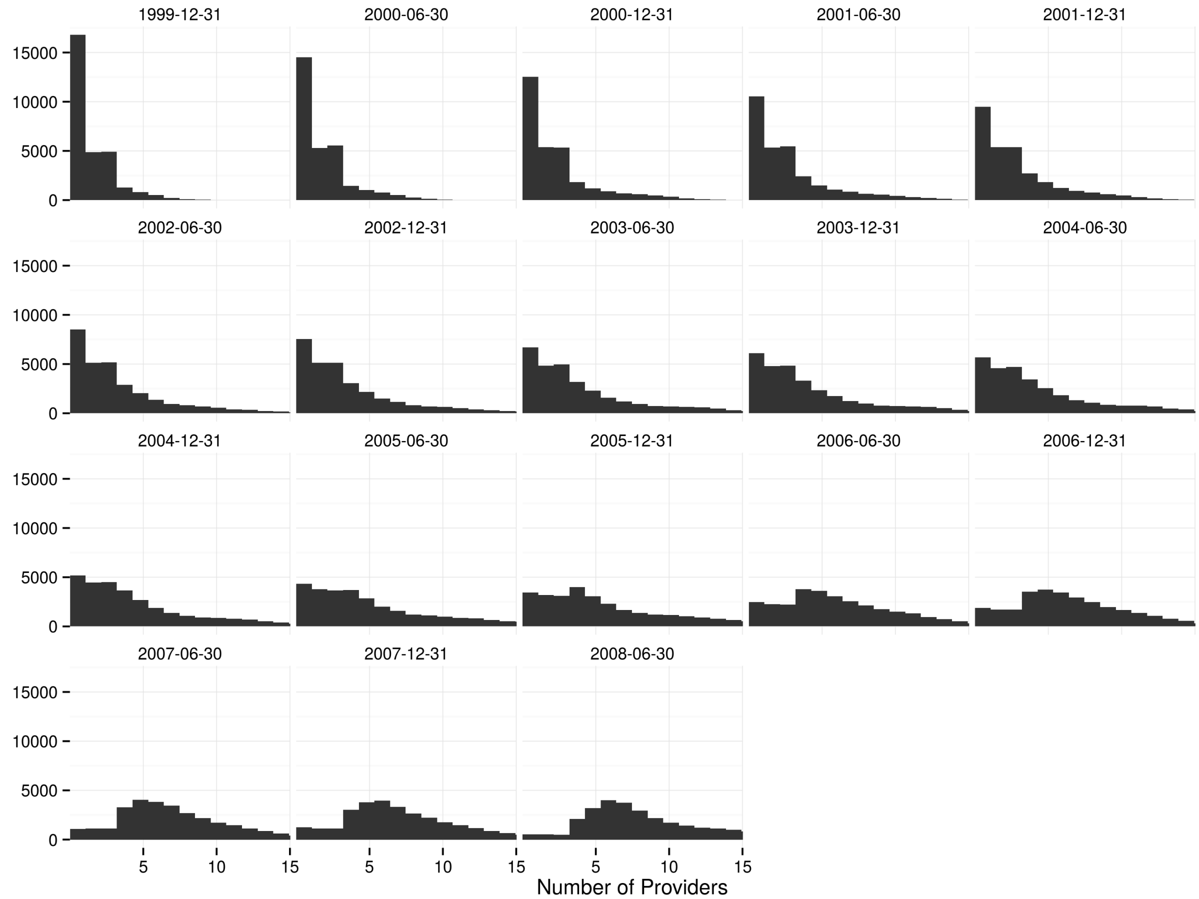
3.1 FCC Form 477 Data

The Federal Communications Commission (FCC) provides data on number of broadband providers at the ZIP code level from 1999 through 2008 from telecommunication providers that are required to file Form 477 twice a year. Broadband providers that serve more than 250 lines are required to submit Form 477, which reports the ZIP codes that they provide service. The FCC definition of broadband across this time period is providers offering fixed-location Internet access connections faster than 200 kilo-bits per second in either download or upload speeds. Subsequently, since 2008 the FCC has progressively increased the minimum internet connection speed to define broadband and changed the geographical unit of analysis from the ZIP code level to the census tract level. These changes limits the time-frame for analysis of broadband diffusion. This serves as our basis for measuring the level of broadband available to a ZIP code. The data do not give insight into the total households served within a ZIP code or the quality of services, so the measure is imprecise. Kolko (2010) demonstrates that a monotonic relationship exists between number of providers and coverage of broadband within a ZIP code. He does this by combining FCC ZIP code level data and Forrester Research's December 2005 Technographic benchmark survey of 60,000 households of broadband availability at the ZIP code level. This serves as justification to use number of broadband providers as a proxy of broadband availability as well as quality.

The FCC data take on a count value of 0, 1-3*, 4, 5, 6, . . . 31. The value 1-3* is a suppressed value of broadband providers for confidentiality purposes. Within the literature on broadband, the most common way to deal with the suppression issue is to code the suppressed values as 2. I choose to maintain this distinction in the analysis, although results where I shift all of the unsuppressed count values down by two (0, 1, 2, . . . 29) are not qualitatively different. The nature of count data implies that the data are not normally distributed and this can be confirmed via figure 4 as across each time period. The non-normal nature of the data is expanded upon in section 5.2.

As measured in December of 1999, FCC Form 477 data indicate that approximately 40% of the 30,000 ZIP codes lacked broadband access. This value rapidly declined as seen in figure 5 until around 2006 where only remote ZIP codes with sparse population did not have access. By the end of the available FCC data, there were effectively no ZIP codes without access although Grubesic

Figure 4: Broadband Providers by Zip Code Across Time



Note: data are suppressed if there are between 1 and 3 providers for a zip code. For clarity of picture, a random uniform draw between 1 and 3 replaced the suppressed data.

(2008) notes that this is not necessarily the same as universal access as a ZIP code may not fully serve its population.

ZIP codes, in a sense, can be described as broadband adopters, although this particular measure is of availability and not necessarily use of broadband. The FCC did not track the percentage of subscribers to broadband in a geographical area until the second half of 2008. At that time, the FCC also changed the reporting unit from ZIP code level data to census tract level data because the latter is a consistent polygon while the former may change across time. Because of the change in reported data, this paper only uses data from the FCC through June of 2008.

The total number of ZIP codes in the sample is 29,588. Of these, 339 ZIP codes received Pilot

Table 2: ZIP Codes Without Access

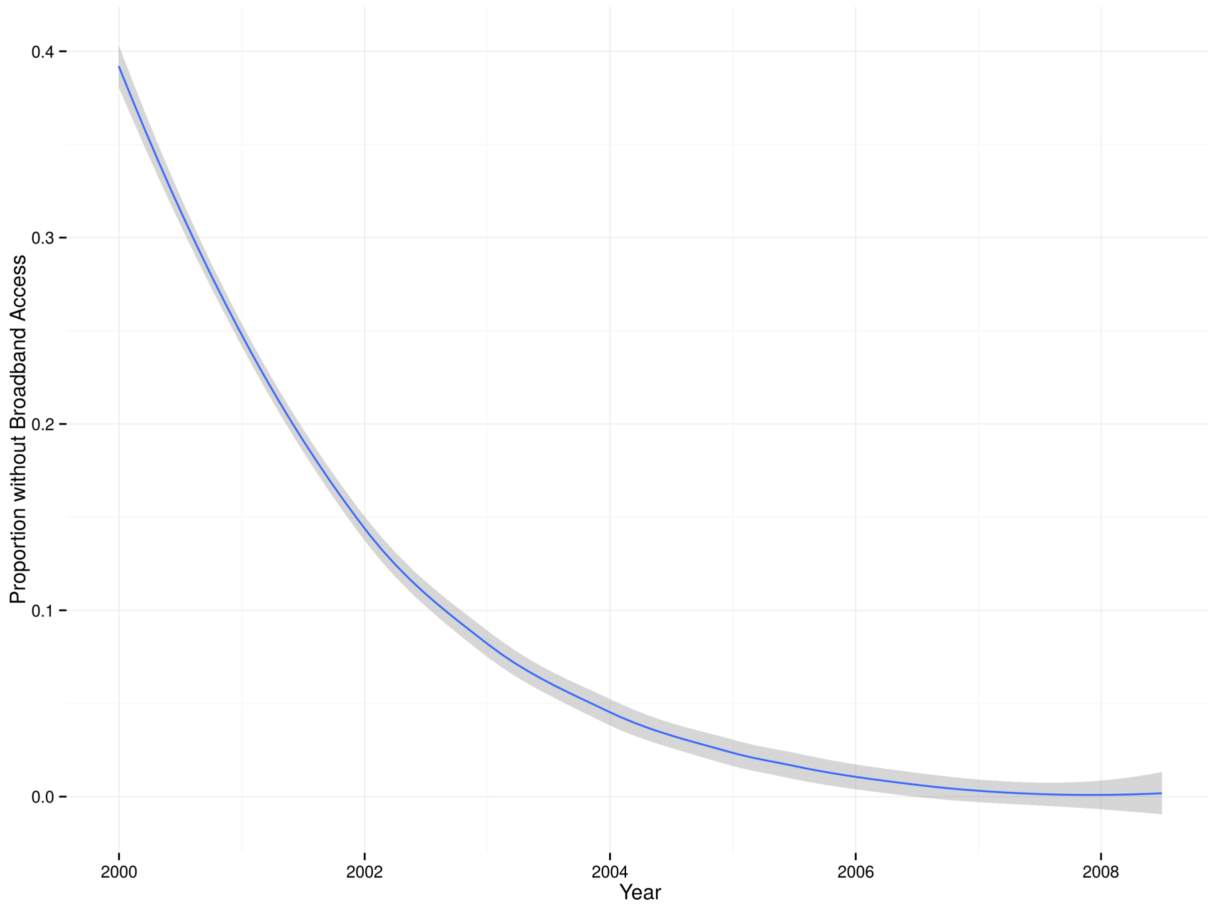
time	No Loan	Pilot	Farm Bill
1999-12-31	11,318 (40.6%)	174 (51.3%)	447 (32.9%)
2000-06-30	8,645 (31.0%)	165 (48.7%)	350 (25.8%)
2000-12-31	6,793 (24.4%)	118 (34.8%)	252 (18.5%)
2001-06-30	4,853 (17.4%)	104 (30.7%)	178 (13.1%)
2001-12-31	4,031 (14.5%)	93 (27.4%)	141 (10.4%)
2002-06-30	3,125 (11.2%)	76 (22.4%)	114 (8.4%)
2002-12-31	2,286 (8.2%)	67 (19.8%)	84 (6.2%)
2003-06-30	1,675 (6.0%)	56 (16.5%)	59 (4.3%)
2003-12-31	1,323 (4.7%)	44 (13.0%)	50 (3.7%)
2004-06-30	847 (3.0%)	37 (10.9%)	20 (1.5%)
2004-12-31	633 (2.3%)	33 (9.7%)	15 (1.1%)
2005-06-30	562 (2.0%)	8 (2.4%)	16 (1.2%)
2005-12-31	283 (1.0%)	5 (1.5%)	6 (0.4%)
2006-06-30	201 (0.7%)	2 (0.6%)	5 (0.4%)
2006-12-31	130 (0.5%)	0 (0.0%)	3 (0.2%)
2007-06-30	34 (0.1%)	0 (0.0%)	0 (0.0%)
2007-12-31	69 (0.2%)	1 (0.3%)	0 (0.0%)
2008-06-30	14 (0.1%)	0 (0.0%)	0 (0.0%)

loans and 1,359 received the Farm Bill loans. At the end of 1999, there were 41% of ZIP codes without broadband access while across ZIP codes that would eventually receive a Pilot or Farm Bill loans this values was 51% and 33%, respectively. In table 2, we can see that the Pilot ZIP codes lagged the result of the United States until approximately 2006 while the Farm Bill ZIP codes consistently had a higher percentage of broadband access, This casual observation in differences for access to broadband across time piques interest in the disbursement and evaluation of the two loan programs. Further inspection requires more rigorous statistical analysis to verify robustness of the resulting increase in broadband access across Pilot and Farm Bill ZIP codes as well as quantifying this increase in access. I now turn to the other data sources used in order to investigate the broadband loan programs.

3.2 Zip Code Level Data

The unit of observation for this study is at the ZIP code level. ZIP codes are defined via the United States Postal Service (USPS) and represent postal routes. These routes change over time, new ones created each year, some are deleted each year, and do not necessarily represent an enclosed area. To

Figure 5: Zip Codes without Broadband Access



deal with the enclosed area problem, I use a ZIP Code Tabulated Area (ZCTA) shapefile provided by the US Census Bureau which approximates each ZIP code to an enclosed area which covers the land mass of the United states using Geographic Information Systems (GIS) technology. This particular shapefile also includes interpolated values for the 2003 population of a ZCTA via Census block populations.

I choose to use the 2004 ZCTA shapefile to deal with the potential issue of changing ZIP code boundaries and those entering/leaving the USPS system as 2004 represents the midpoint of the years of interest. The changing ZIP codes could potentially be a problem if the changing boundaries directly, or indirectly through a covariate, affect the number of providers in a ZIP code. This potential bias cannot signed due to the unknown nature of the changing ZIP code boundaries. It is not known if the boundary changes were to positively or negatively influence the provider count.

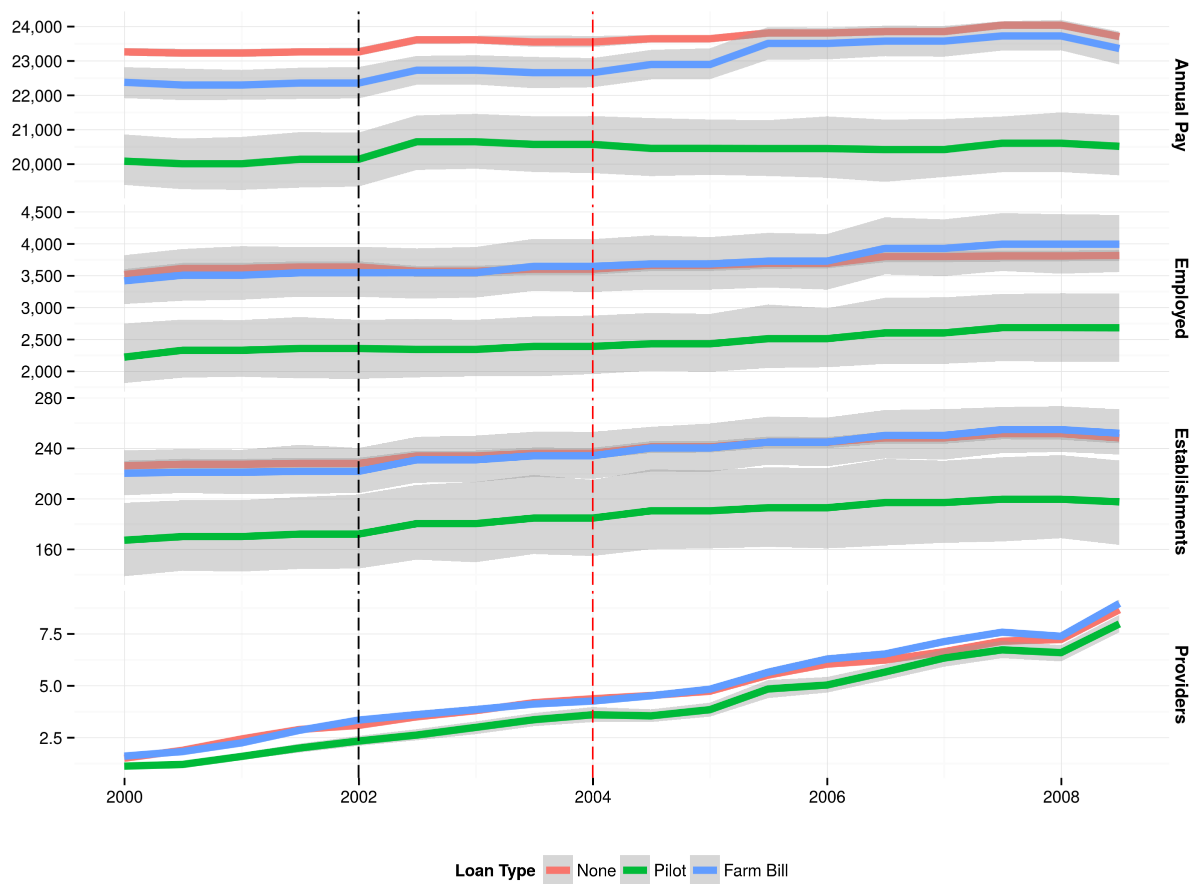
Within the framework of this study, it is not feasible to test for whether or not changing ZIP code boundaries are influencing the results because of the lack of available ZCTA shapefiles. It is noted that changing ZIP codes may influence results but I choose to ignore this issue because this is not testable and the direction of the bias is unknown. All of the ZIP code level data can be seen in table 3.

Another potential issue with ZCTAs is the endogenous nature of their size. As ZIP codes are based upon postal routes, this implies that the time to service a postal route should be roughly equal across ZIP codes. It is not the case that ZIP codes are randomly drawn, they are designed by the postal service and thus their goal is to efficiently serve the United States. The size of the ZIP code therefore needs to be controlled for by calculating the area of the ZCTA. The total square miles of each ZCTA is calculated from the ZCTA shapefile through GIS software in R. Holding all else constant, it should be the case that larger ZIP codes have more broadband providers because they span a greater area which allows a different broadband provider to potentially serve the area.

The ZCTA shapefile also allows for the calculation of stands for the topographical features of the area which partially determine the costs of deploying broadband in an area. In order to control for this, I use the Terrain Ruggedness Index (TRI) for each ZCTA in the 2004 shapefile. TRI is a measurement developed by Riley, et al. (1999) to express the amount of elevation difference between adjacent cells of a digital elevation grid. This is calculated through elevation shapefiles from `DIVA-GIS.org` and the R function `terrain()` from the raster package. The values range from 0 to 100 and Kolko (2012) demonstrates that costs to deploy broadband are correlated with the terrain characteristics.

The US Census Bureau produces the County Business Patterns (CBP) dataset which provides information on number of establishments, employment during the week of March 12, first quarter payroll, and annual payroll across 6-digit North American Industry Classification System (NAICS) and various geographies. The annual data are extracted from the Business Register (BR), a database of all known single and multi-establishment employer companies maintained and updated by the US Census Bureau. Data on number of employees and payroll values at the ZIP code level suffer from suppression issues due to the sensitivity of being able to identify particular businesses if there are a limited number operating in a ZIP code. Suppression affects approximately 2% of ZIP codes, however the total number of establishments is not suppressed for any ZIP codes. I use annual

Figure 6: Zip Code Data Trends



Black dotted line indicates the Pilot Program while Red dotted line indicates the Current Program.

data from 1999 to 2008 on total number of establishments[2] as a way to proxy for the economic activity in a ZIP code that should be positively related to the number of broadband providers as their potential customer base grows. In addition, I include annual payroll divided by number of employees[3] to proxy for wages in the ZIP code. Average annual pay is deflated to 1999 dollars via the Consumer Price Index (CPI).

Figure 6 presents time trends of the relevant economic conditions for all ZIP codes and further

[2] The values for number of employees and establishments track each other fairly well across time which, along with suppression issues, drives my choice of using number of establishments as opposed to number of employees.

[3] To correct for suppression issues, I proceed in three steps. First, if a suppressed value has an unsuppressed value in the previous year. If the previous year is suppressed, then I replace the suppressed value with the value of the following year. If both of these values are suppressed, then I replace the suppressed value with the average across all years. Because the variable of interest is a ratio between annual payroll and number of employees, there should be no issue with the differing size of a ZIP code as it relates to suppression.

breaks down ZIP codes by those which received either the Pilot or Farm Bill broadband loans. Each series is plots the sample mean and a 90% confidence interval as calculated through bootstrapped standard errors. The first (black) dotted line indicates when Pilot loans were disbursed and the second (red) dotted line is for the Farm Bill. Throughout the entire period, the Pilot loan ZIP codes lags behind all for employment statistics as well as number of broadband providers. For the Farm Bill loans, it appears that the ZIP codes track the rest of the US across the years fairly well for all variables except for average annual pay. For average annual pay, the Farm Bill ZIP codes lag behind the rest of the US and this is a statistically significant difference until there is a visible uptick in annual pay for Farm Bill ZIP codes that occurs around 2005. After 2005, the Farm Bill ZIP codes still lag behind the rest of the US but the difference is not statistically significant like before 2005. Aside from the annual pay uptick, It is difficult to determine visually if the Pilot and/or Farm Bill loans made a discernible impact in increasing the number of providers, which continues to motivate the need for this study.

A potentially confounding issue in ZIP code level data is that the area and number of establishments are negatively correlated (value of -0.0729). This stems from the endogenous nature of postal route size as more dense regions will have smaller ZIP codes on average in order to serve all of the establishments/people. Further, sparsely populated areas are more likely to have larger ZIP codes as it would be inefficient for remote areas to have more postal routes as there are fixed costs associated with postal routes. Because of this, one needs to further control for the density of firms in a ZIP code as an arguably more important proxy to customer base that broadband providers view as potential revenue streams.

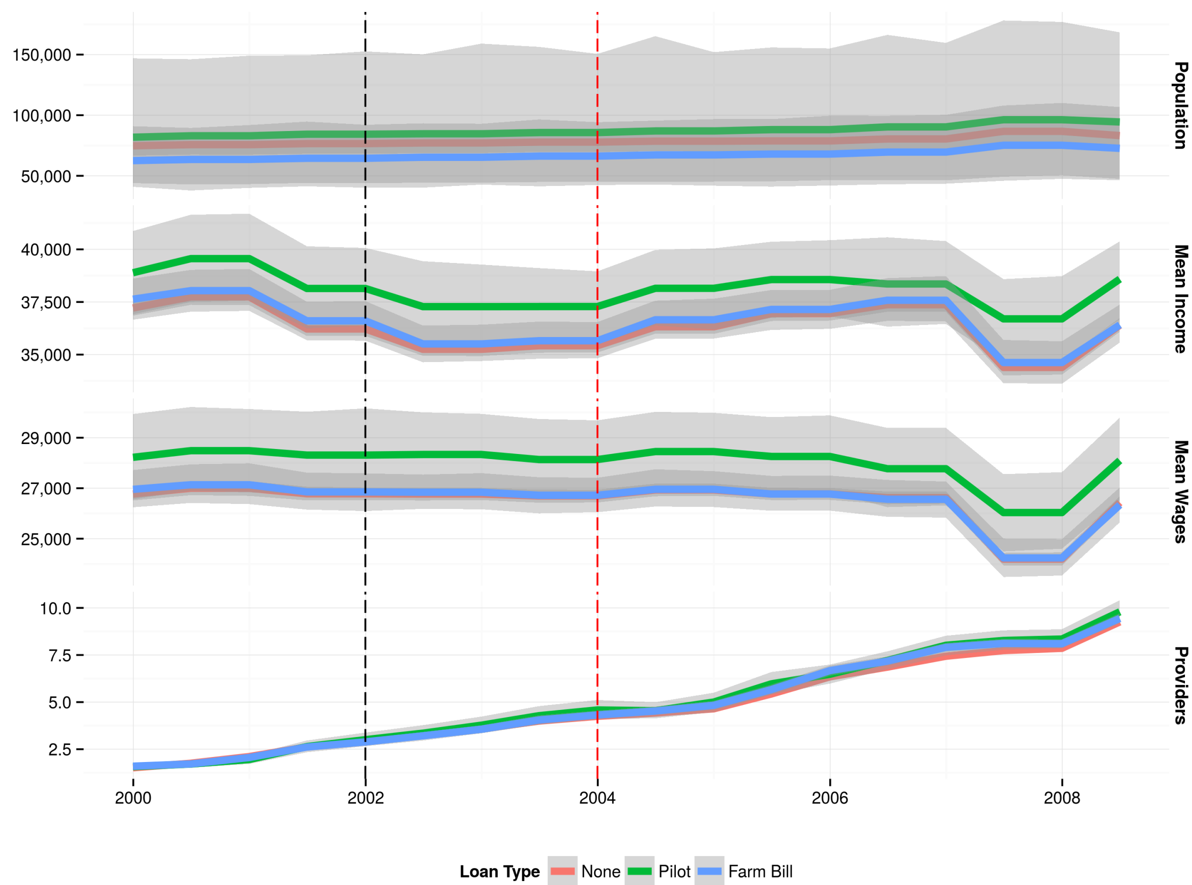
Table 3: Zip Code Level Data

Year	Category	n	Providers	Establishments	Employed	Pay	TRI	AREA
1999	All Zips	29,588	1.51 (1.58)	226 (377)	3,520 (7,391)	23,267 (9,922)	15.6 (20.2)	99.1 (245)
1999	Farm Bill	1,359	1.6 (1.46)	220 (335)	3,420 (7,454)	22,380 (8,407)	7.98 (12.7)	118 (218)
1999	Pilot	346	1.13 (1.32)	167 (266)	2,223 (4,317)	20,087 (7,474)	7.75 (9.3)	116 (181)
2000	All Zips	29,588	1.88 (1.84)	228 (378)	3,616 (7,596)	23,233 (10,135)	15.6 (20.2)	99.1 (245)
2000	Farm Bill	1,359	1.82 (1.57)	221 (334)	3,509 (7,738)	22,301 (8,504)	7.98 (12.7)	118 (218)
2000	Pilot	346	1.19 (1.31)	170 (271)	2,331 (4,511)	20,011 (7,306)	7.75 (9.3)	116 (181)
2001	All Zips	29,588	2.9 (2.7)	228 (377)	3,639 (7,691)	23,266 (9,861)	15.6 (20.2)	99.1 (245)
2001	Farm Bill	1,359	2.85 (2.55)	222 (334)	3,547 (7,748)	22,358 (8,280)	7.98 (12.7)	118 (218)
2001	Pilot	346	2.01 (1.95)	172 (273)	2,359 (4,548)	20,141 (7,655)	7.75 (9.3)	116 (181)
2002	All Zips	29,588	3.51 (3.08)	234 (383)	3,573 (7,395)	23,618 (10,200)	15.6 (20.2)	99.1 (245)
2002	Farm Bill	1,359	3.61 (3.31)	231 (344)	3,547 (7,661)	22,731 (8,375)	7.98 (12.7)	118 (218)
2002	Pilot	346	2.62 (2.54)	180 (285)	2,345 (4,447)	20,650 (7,707)	7.75 (9.3)	116 (181)
2003	All Zips	29,588	4.18 (3.48)	236 (386)	3,604 (7,411)	23,555 (14,152)	15.6 (20.2)	99.1 (245)
2003	Farm Bill	1,359	4.11 (3.18)	234 (348)	3,648 (7,836)	22,660 (8,190)	7.98 (12.7)	118 (218)
2003	Pilot	346	3.33 (3.06)	185 (292)	2,392 (4,466)	20,575 (7,876)	7.75 (9.3)	116 (181)
2004	All Zips	29,588	4.54 (3.53)	241 (394)	3,665 (7,461)	23,649 (9,772)	15.6 (20.2)	99.1 (245)
2004	Farm Bill	1,359	4.51 (3.13)	240 (355)	3,686 (7,922)	22,898 (8,259)	7.98 (12.7)	118 (218)
2004	Pilot	346	3.49 (2.86)	191 (300)	2,433 (4,488)	20,458 (8,019)	7.75 (9.3)	116 (181)
2005	All Zips	29,588	5.52 (4.2)	245 (399)	3,695 (7,476)	23,817 (9,853)	15.6 (20.2)	99.1 (245)
2005	Farm Bill	1,359	5.65 (3.64)	245 (360)	3,730 (8,042)	23,511 (8,884)	7.98 (12.7)	118 (218)
2005	Pilot	346	4.82 (3.84)	193 (305)	2,513 (4,635)	20,454 (8,114)	7.75 (9.3)	116 (181)
2006	All Zips	29,588	6.24 (3.64)	248 (404)	3,803 (7,679)	23,858 (9,878)	15.6 (20.2)	99.1 (245)
2006	Farm Bill	1,359	6.55 (3.24)	250 (367)	3,930 (8,588)	23,578 (8,562)	7.98 (12.7)	118 (218)
2006	Pilot	346	5.63 (3.57)	197 (312)	2,603 (4,821)	20,424 (8,363)	7.75 (9.3)	116 (181)
2007	All Zips	29,588	7.15 (3.51)	252 (407)	3,812 (7,672)	24,045 (9,920)	15.6 (20.2)	99.1 (245)
2007	Farm Bill	1,359	7.59 (3)	255 (366)	3,996 (8,581)	23,728 (8,630)	7.98 (12.7)	118 (218)
2007	Pilot	346	6.75 (3.77)	200 (319)	2,686 (5,098)	20,608 (8,266)	7.75 (9.3)	116 (181)
2008	All Zips	29,588	8.67 (4.36)	248 (402)	3,821 (7,705)	23,712 (9,868)	15.6 (20.2)	99.1 (245)
2008	Farm Bill	1,359	8.97 (3.69)	252 (361)	3,996 (8,464)	23,362 (8,853)	7.98 (12.7)	118 (218)
2008	Pilot	346	7.98 (4.05)	198 (316)	2,684 (5,002)	20,518 (8,550)	7.75 (9.3)	116 (181)

3.3 County Level Data

ZIP code level data is not always available at an annual rate, which means that to further analyse the diffusion of broadband I need to augment county level data with the previous ZIP code data. Using the ZCTA shapefile from 2004, I overlay each ZCTA on a county level shapefile for the US to determine the county that each ZCTA lies within. For ZCTAs that span multiple counties, I assign the ZCTA to the county which has the majority of the overlapped area. I also utilize the county shapefile to calculate the total square miles of each county to help control for density issues related to population.

Figure 7: County Data Trends



Black dotted line indicates the Pilot Program while Red dotted line indicates the Current Program.

The main source of county level statistics comes from the Internal Revenue Service (IRS) County

Income dataset. This comes from the IRS Statistics of Income division which bases its county data on administrative records of individual income tax returns (Forms 1040) filed during the 12-month period of January 1 to December 31. The IRS provides data at the county level from 1989 to 2013 at the county level on number of tax returns filed (proxies households), number of exemptions filed (proxies population), Adjusted Gross Income (AGI, which proxies income), Wages and Salaries income, Dividends before exclusions, and Interest received. Because not all individuals within the United States are required to file tax returns, this data only represent the tax filling portion. I utilize total number of exemptions to proxy for the population of a county in a given year, which should be a more accurate reflection of population as opposed to the Census values because the IRS data can more accurately capture migration movements within the United States. Population of a county is expected to be positively related to number of broadband providers as it represent the potential customer base that a broadband provider would see. I further proxy for population density by dividing the number of exemptions by the area of a county. Population density should lower the costs of deploying broadband for an area as a broadband providers does not need to lay as much cable to connect customers to the backbone of their system.

In addition, I utilize AGI divided by number of tax returns filed and Wages and Salaries income divided by number of tax returns filed to proxy the mean level of income and wages for a county. Both of these are deflated to 1999 dollars through CPI. These are used as a cross-check for the ZIP code level average annual pay, which the mean level of the ZIP codes should be similar to those of counties. The ZIP code level is preferred because of the finer granularity of ZIP code level over a county level statistic. Both of these measures are expected to be positively related to number of broadband providers as this represents quality of the customer base that a potential broadband provider would see.

The breakdown of how population and income at the county level change across time for counties receiving Pilot or Farm Bill loans as well as the rest of the United States can be seen in figure 7. All of the series are within the 90% confidence interval of each other as calculated through bootstrapped standard errors. On average, counties receiving at least one Pilot loan have larger populations and have higher wages and incomes than the rest of the United States. Further, the Farm Bill counties are smaller than the rest of the United States in population and appear to fluctuate around the rest of the United States in terms of income. As seen in table 4, the Wages and Salaries are slightly

Table 4: County Level Data

Year	Category	n	Providers	Population	Income	Wages
1999	All	3,089	1.52 (1.02)	75,700 (239,426)	37,242 (10,250)	26,795 (7,558)
1999	Farm Bill	286	1.6 (0.871)	62,516 (203,311)	37,633 (8,515)	26,955 (6,525)
1999	Pilot	105	1.57 (0.98)	81,860 (292,645)	38,889 (10,519)	28,223 (8,493)
2000	All	3,089	1.94 (1.21)	76,643 (242,692)	37,771 (10,839)	27,026 (7,810)
2000	Farm Bill	286	1.89 (0.997)	63,507 (207,043)	38,038 (8,621)	27,136 (6,562)
2000	Pilot	105	1.83 (1.06)	83,032 (296,372)	39,560 (10,856)	28,487 (8,735)
2001	All	3,089	2.76 (1.66)	77,547 (246,354)	36,248 (9,490)	26,807 (7,479)
2001	Farm Bill	286	2.75 (1.52)	64,475 (212,296)	36,605 (7,918)	26,854 (6,328)
2001	Pilot	105	2.82 (1.68)	84,336 (303,094)	38,142 (9,929)	28,312 (8,575)
2002	All	3,090	3.41 (1.98)	78,110 (248,066)	35,279 (9,050)	26,795 (7,175)
2002	Farm Bill	286	3.39 (1.73)	65,308 (215,686)	35,509 (7,534)	26,842 (6,138)
2002	Pilot	105	3.56 (2.17)	84,689 (306,482)	37,279 (9,741)	28,336 (8,294)
2003	All	3,090	4.13 (2.31)	78,827 (250,192)	35,449 (8,965)	26,718 (7,088)
2003	Farm Bill	286	4.19 (1.92)	66,277 (220,103)	35,659 (7,608)	26,721 (6,094)
2003	Pilot	105	4.43 (2.54)	85,699 (311,352)	37,283 (9,472)	28,136 (8,059)
2004	All	3,090	4.54 (2.41)	79,601 (251,664)	36,350 (9,423)	26,958 (7,183)
2004	Farm Bill	286	4.68 (1.96)	67,239 (225,307)	36,654 (8,361)	26,958 (6,206)
2004	Pilot	105	4.77 (2.45)	87,016 (318,103)	38,150 (9,779)	28,452 (8,121)
2005	All	3,090	5.9 (2.9)	79,865 (252,250)	36,978 (9,468)	26,796 (7,052)
2005	Farm Bill	286	6.17 (2.27)	67,940 (229,920)	37,146 (8,364)	26,770 (6,031)
2005	Pilot	105	6.22 (2.85)	88,104 (324,217)	38,565 (9,865)	28,258 (7,954)
2006	All	3,090	7.15 (2.52)	81,472 (256,096)	37,400 (9,881)	26,652 (7,109)
2006	Farm Bill	286	7.54 (2.09)	69,587 (235,284)	37,583 (8,957)	26,562 (6,185)
2006	Pilot	105	7.61 (2.63)	90,342 (330,947)	38,355 (10,192)	27,772 (7,993)
2007	All	3,090	7.8 (2.48)	87,846 (274,491)	34,420 (10,306)	24,217 (7,331)
2007	Farm Bill	286	8.11 (2)	75,206 (253,252)	34,623 (8,886)	24,238 (6,264)
2007	Pilot	105	8.31 (2.64)	96,276 (354,242)	36,698 (10,534)	26,035 (8,267)
2008	All	3,090	9.26 (3.01)	84,336 (266,711)	36,425 (9,067)	26,460 (6,887)
2008	Farm Bill	286	9.44 (2.4)	72,675 (251,376)	36,414 (7,987)	26,339 (5,995)
2008	Pilot	105	9.79 (2.94)	94,422 (352,693)	38,588 (9,675)	28,099 (7,898)

higher than the ZIP code values although they track each others trends fairly well except the visual uptick around 2005 in annual pay for Farm Bill ZIP codes is not reflected at the county level. This would indicate that use of the ZIP code level data improves precision of estimates. Further visual inspection tells a similar story to the ZIP code graph of attributes, it is difficult to see a discernible impact across time that one may ascribe to either loan program.

The last county level variable utilized in this study is of policy implication for the loan programs and that is the classification of the county as metro, rural but adjacent to a metro county (Rural Adjacent), and rural but non-adjacent to a metro county (Rural Non-Adjacent) which is updated in years that end in 3 (1993, 2003, 2013). I choose to use the values for 2003 as this would be a little bit before the halfway point in the analysis. These classifications come from the rural-urban continuum code as calculated by the USDA Economic Research Service (ERS) which utilizes these classifications in order to distinguish metropolitan counties by the population size of their metro area, and non-metropolitan counties by degree of urbanization and adjacency to a metro area. The classification is important to the loan programs as rural counties were the targeted group to receive these loans and thus it is expected that rural counties to be more affected by the programs.

4 Determinants of Loan Receipt

The stated objectives from RUS for disbursement of loans were to target under-served rural communities with fewer than 20,000 inhabitants, as determined by the Census Bureau, and not located in a Standard Metropolitan Statistical Area. In table 1, it is evident that some of these qualifications were not met as loans were disbursed to metro areas. This section aims to be more statistically rigorous in examination of whether or not the ZIP codes receiving these loans were more likely to qualify under the stated goals of the programs. In order to do this, I turn to discrete choice models.

Suppose that the probability of receiving a loan takes the form:

$$Pr\left(Loan_z = 1\right) = f\left(Rural_z, Pop_z < 20{,}000, Prov_z < 4, X_z\right) \tag{1}$$

where the X matrix includes ZIP code and county characteristics that are believed to affect the probability of receiving a loan. For evaluating the Pilot, Farm Bill, and both loan programs the key variables of interest are their stated goals, all of which are taken to be dummy variables: rural areas, with fewer than 20,000 inhabitants, and low levels of broadband.

The Pilot program began in December of 2000, in order to stay consistent with the known information at the time the values of all variables (with exception of ZCTA population) are taken to be from December 2000. While these three variables are the key in evaluation, it is also important to control for other factors that may have attributed to the probability of receiving a loan during this time. These factors include the total number of broadband providers for a ZIP code, total population, total establishments, average income, and TRI. Number of broadband providers is suspected to be negatively related to probability of receiving a loan as more providers in an area would make servicing an area less attractive. All other variables (except TRI) are expected to be positively associated with the probability of receiving a loan as they indicate higher potential revenues that would make it attractive for a broadband provider to attempt to service an area and thus apply for a loan. The TRI variable is ambiguous *a priori*. An area with a high TRI is more likely to not be serviced by broadband providers due to the high costs which is supposedly targeted via these programs. However, if the terrain is prohibitively rugged, then the subsidized rate of the loan may not be enough to offset the high costs and therefore be negatively related to receiving these loans. These are all empirical questions that I address.

Equation 1 is typically be estimated through maximum likelihood methods involving either a logistic or probit regression. The logistic regression takes the form of:

$$Pr\left(Loan_z = 1\right) = \frac{1}{1 + e^{-(\beta_1 Rural_z + \beta_2 Pop_z < 20{,}000 + \beta_3 Prov_z < 4 + \beta_4 X_z)}} \tag{2}$$

where the coefficients to be estimated have the interpretation of log-odds ratio. The probit regression takes the form of:

$$Pr\left(Loan_z = 1\right) = \Phi\left(\beta_1 Rural_z + \beta_2 Pop_z < 20,000 + \beta_3 Prov_z < 4 + \beta_4 X_z\right) \tag{3}$$

where $\Phi(\cdot)$ is the normal cumulative density function. In practice, the fitted values and marginal effects between a logit and probit specification should be similar. Therefore, I present results of five models that are estimated as both a logit, in table 5, and probit, table 6 as a robustness check across specifications.

The first model is intentionally simple in that the only covariates involved are the three criteria

Table 5: Probability of Recveiving Loan (Logit)

	Dependent variable:				
	iloans		ipilot	ibip1234	
	(1)	(2)	(3)	(4)	(5)
I(Prov_num <3)	0.593***	0.647***	0.013	0.791***	0.319***
	(0.089)	(0.128)	(0.293)	(0.140)	(0.099)
I(SUMBLKPOP <20000)	−0.331***	0.145	−0.043	0.204**	0.413***
	(0.083)	(0.092)	(0.220)	(0.099)	(0.101)
rucadj	0.354***	0.485***	0.245	0.535***	0.619***
	(0.061)	(0.063)	(0.151)	(0.069)	(0.071)
rucnonadj	0.025	0.161**	0.781***	−0.108	−0.044
	(0.072)	(0.077)	(0.148)	(0.092)	(0.094)
Prov_num		−0.020	−0.163***	0.010	0.0004
		(0.024)	(0.057)	(0.025)	(0.016)
log(SUMBLKPOP + 1)		0.053	−0.037	0.089**	0.074*
		(0.037)	(0.070)	(0.043)	(0.043)
log(est)		0.162***	0.143**	0.161***	0.205***
		(0.035)	(0.070)	(0.040)	(0.041)
logAPay_R2		−0.372***	−0.557***	−0.288***	−0.404***
		(0.077)	(0.147)	(0.088)	(0.089)
tri		−0.047***	−0.047***	−0.046***	−0.046***
		(0.003)	(0.007)	(0.003)	(0.003)
Constant	−3.122***	−0.489	1.377	−2.072**	−0.715
	(0.072)	(0.799)	(1.514)	(0.918)	(0.920)
Observations	29,588	29,588	29,588	29,588	29,588
Log Likelihood	−6,444.219	−6,163.733	−1,783.288	−5,221.596	−5,242.630
Akaike Inf. Crit.	12,898.440	12,347.470	3,586.577	10,463.190	10,505.260
Note:					*p<0.1; **p<0.05; ***p<0.01

Table 6: Probability of Recveiving Loan (Probit)

	Dependent variable:				
	iloans		ipilot	ibip1234	
	(1)	(2)	(3)	(4)	(5)
I(Prov_num <3)	0.273***	0.309***	−0.016	0.371***	0.153***
	(0.040)	(0.059)	(0.109)	(0.063)	(0.046)
I(SUMBLKPOP <20000)	−0.155***	0.052	−0.012	0.072	0.174***
	(0.039)	(0.044)	(0.083)	(0.046)	(0.047)
rucadj	0.169***	0.232***	0.088	0.249***	0.287***
	(0.029)	(0.031)	(0.058)	(0.033)	(0.033)
rucnonadj	0.011	0.069*	0.319***	−0.066	−0.036
	(0.034)	(0.037)	(0.058)	(0.042)	(0.042)
Prov_num		−0.013	−0.069***	0.001	−0.002
		(0.011)	(0.022)	(0.011)	(0.007)
log(SUMBLKPOP + 1)		0.021	−0.017	0.033*	0.027
		(0.017)	(0.027)	(0.019)	(0.019)
log(est)		0.079***	0.063**	0.080***	0.099***
		(0.016)	(0.027)	(0.018)	(0.018)
logAPay_R2		−0.166***	−0.216***	−0.124***	−0.178***
		(0.036)	(0.059)	(0.040)	(0.040)
tri		−0.018***	−0.017***	−0.017***	−0.017***
		(0.001)	(0.002)	(0.001)	(0.001)
Constant	−1.723***	−0.565	−0.004	−1.290***	−0.662
	(0.033)	(0.376)	(0.610)	(0.415)	(0.417)
Observations	29,588	29,588	29,588	29,588	29,588
Log Likelihood	−6,443.988	−6,180.506	−1,782.742	−5,235.689	−5,258.723
Akaike Inf. Crit.	12,897.980	12,381.010	3,585.483	10,491.380	10,537.450

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

stated by the broadband loan programs: dummies for rural adjacent, rural non-adjacent, fewer than 4 broadband providers, and fewer than 20,000 inhabitants. For both the logit and probit specifications, ZIP codes with fewer than 4 broadband providers as well as rural adjacent areas were more likely to receive a broadband loan, which is consistent with *a priori* expectations of the loan program. However, ZIP codes with fewer than 20,000 inhabitants were less likely to receive a broadband loan, which is a result that would indicate that the loan programs failed in disbursing loans to areas with fewer than 20,000 inhabitants. This result may be due to omitted variables in the model such as number of establishments which is negatively related to populations of under 20,000 and potentially positively related to the probability of receiving a loan. Because of this, I turn to extending the model.

The second model includes a number of variables taken from December 2000 in addition to the first model. These variables are number of broadband providers, the log of population, the log of number of establishments, the log of average pay, and TRI. For program evaluation, the results generally change to the expected outcome for an effective broadband program. Places with fewer than 4 broadband providers and rural areas (both adjacent and non-adjacent) are significant and positively related to the probability of receiving a loan. The previous perplexing result of places with fewer than 20,000 inhabitants being significantly less likely to receive loans is no longer present in the model. The model now has the predicted interpretation that a ZIP code with fewer than 20,000 inhabitants is more likely to receive a broadband loan, although this is not a statistically significant result in either specification. The additional regressors in the second model are each significant and of the expected sign (except TRI, which did not have an expected sign).

The first and second models aggregate both the Pilot and Farm Bill loans, which may not be a reasonable assumption. To verify if the two loans programs yield qualitatively similar results, I split the loan programs up. For the third model, the dependent variable takes on a value of 1 only if a ZIP code received a Pilot loan where before the dependent variable would have taken a value of 1 if a ZIP code received either the Pilot or Farm Bill loan. With the Pilot loan, the logit and probit gives qualitatively similar results and there is not a statistically significant result which indicates that the Pilot was disbursed in a manner that opposes its stated goals. The Pilot loans were disbursed to rural areas (adjacent and non-adjacent) with a higher probability than metro areas. Further, although results do not display a significantly positive relationship with under-served areas there is a significantly negative relationship with the number of providers for a zip code. The USDA definition of an under-served area may not be consistent with the chosen metric of fewer than 4 broadband providers, so finding a significant and negative relationship with number of providers may be consistent with what the USDA finds to be under-served areas.

The fourth and fifth models inspect the Farm Bill loans but at different time periods. The fourth model is consistent with the time period selected for models 1 through 3 (December 2000). Because the Farm Bill loans were disbursed later than the Pilot, the fifth model utilizes variables from December 2002 to control for potential changes in economic conditions between 2000 and 2002 that may affect results. For the Farm Bill loans, the results appear to indicate a stronger relationship between the stated goals of the program and the probability of receiving a loan. The measure for under-served areas and fewer than 20,000 inhabitants are of the expected sign and significant across the probit and logit specification. Rural areas that are adjacent to metro areas are more likely to receive loans, which is also consistent with the stated goals. However, rural areas which are not adjacent to metro areas have the incorrect sign although they are not statistically different from 0.

In all, it appears that the loan program did serve areas which were consistent to its stated goals. Further, it appears that there is evidence that the Farm Bill loans improved upon the Pilot loans. This result is to be expected as the Pilot loan program was limited in scope, which would reduce the power of detecting an increase in probability of receiving a loan, and those within the USDA did not have previous experience with reviewing broadband loan applications. While the Inspector General may have found specific evidence and incidences of misappropriation of loans, I find results consistent with the stated goals. If there was widespread misappropriation of loans, then I would have found statistically significant effects in the opposite direction as predicted. Since I do not find this result and instead see statistically significant results in the predicted direction, I conclude that, on the whole, both the Pilot and Farm Bill loans were directed at their intended population of interest. The next step is to determine if the targeted areas saw a benefit in broadband access through the Pilot and Farm Bill loan programs.

5 Loan Program Effectiveness

The targeted goal of both the Pilot and Farm Bill loan programs was to increase broadband availability to under-served areas. This could be through increasing the broadband infrastructure in place, adding last mile telecommunications providers, or providing programs that advocate broadband at public places such as libraries. The only available data are number of broadband providers for a ZIP code as collected by the FCC, which can capture some of these increases but not all. The imprecision of broadband availability lessens the power to identify any effect that the broadband loan programs may have had on broadband availability. From a program evaluation standpoint, failure to find significant results does not necessarily mean that the program was not effective. However, if statistically significant results are found using an imprecisely measured dependent variable, then it is fairly certain that the loan programs did have an effect on broadband availability and credibly assess an estimate of this value. The next subsection utilizes standard panel econometric techniques to address the programs impact while the following subsection uses the preferred Poisson regression to account for the count nature of number of providers.

5.1 Basic Panel Methods

The typical starting point for analyzing a dataset which spans across both time and cross-sections is to implement panel estimators. These estimators typically utilize what are termed fixed and random effects that are either at the individual level, time level, or both. A general framework for

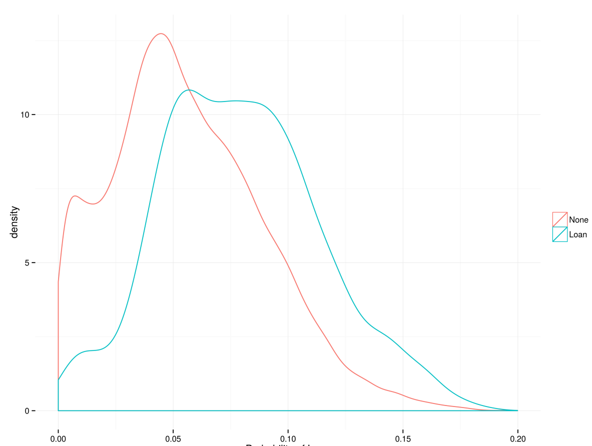


Figure 8: *Logit Density Plot*

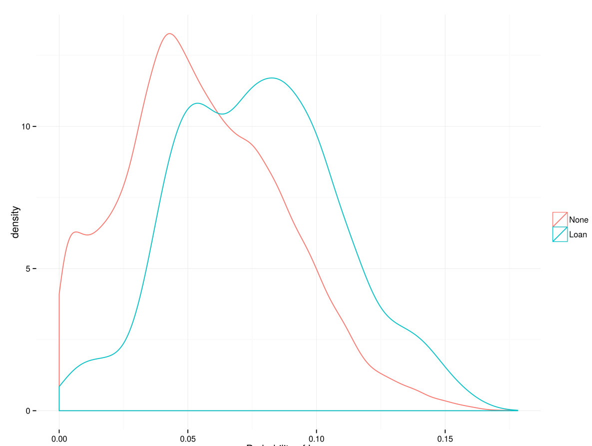


Figure 9: *Progit Density Plot*

this analysis can be formulated as such:

$$Prov_{z,t} = \mu_z + \tau_t + \beta_1 Loan_{z,t} + \beta_2 X_{z,t} + \varepsilon_{z,t} + \alpha_z + \alpha_t \tag{4}$$

where $Prov_{z,t}$ represents the number of broadband providers in ZCTA z at time t; the particular variable of interest is $Loan_{z,t}$ which is a dummy variable indicating if a zip code has been awarded (and assumed deployed) a subsidized loan from the USDA at time t or before; and $X_{z,t}$ represents a set of other economic and demographic variables at the zip code level that determine the level of broadband provision. The variables chosen for $X_{z,t}$ are log of number of establishments, log of population, and log of average annual pay to control for potential revenue streams for broadband providers; TRI and a second order polynomial for area of ZCTA to proxy for potential costs to a broadband provider of servicing a ZCTA; population density and employment density to control for potential confounding effects related to ZCTA size; and dummy variables for rural adjacent and non-adjacent ZCTAs as a way to evaluate whether or not a digital-divide existed during this time frame. As a reminder, the data are biannual from December 1999 to June 2008 across 29,588 ZCTAs.

The μ_z parameter is a zip code level fixed effect that may or may not be present in the particular model; τ_t is a time fixed effect that may or may not be present; α_z is a zip code random effect that may or may not be present in the model; and α_t is a time random effect that may or may not be present in the model.

Four models are presented in table 7 which are variants of a fixed effects estimator. The first is a pooled estimator which does not control for any time or ZCTA effects, effectively assuming that each observation in the model is taken from the same distribution. The second is a ZCTA fixed effect estimator, which implies that μ_z is estimated and this case with a dummy variable for each ZCTA in the sample. The third estimator is a time fixed effect estimator, τ_t, which should account for structural temporal changes associated with broadband diffusion. The fourth model combines individual and time fixed effects, μ_z and τ_t. Models two and four include individual fixed effects which does not allow for the estimation of time-invariant characteristics and limits the scope of analysis as the digital-divide cannot be evaluated.

Across the four models in table 7, the estimate on the broadband loan program are all positive

Table 7: Fixed Effects

	Dependent variable:			
	Prov_num			
	(1)	(2)	(3)	(4)
iloans	1.751***	2.376***	0.042**	0.094***
	(0.025)	(0.028)	(0.018)	(0.019)
log(est)	1.013***	1.480***	1.015***	0.661***
	(0.003)	(0.018)	(0.002)	(0.012)
log(Pop_IRS)	0.515***	14.550***	0.486***	2.962***
	(0.004)	(0.053)	(0.003)	(0.042)
logAPay_R2	0.395***	0.408***	0.283***	−0.050***
	(0.011)	(0.020)	(0.008)	(0.013)
rucadj	−0.023**		−0.087***	
	(0.011)		(0.008)	
rucnonadj	0.308***		0.215***	
	(0.014)		(0.010)	
ZCTA Fixed Effects	No	Yes	No	Yes
Time Fixed Effects	No	No	Yes	Yes
Observations	532,584	532,584	532,584	532,584
R^2	0.489	0.198	0.630	0.019
Adjusted R^2	0.489	0.187	0.630	0.018
Note:				*p<0.1; **p<0.05; ***p<0.01

and statistically significant at the 5% level. The coefficients range from 0.042 to 2.376, although the interpretation for each model is different as to how the broadband loan programs affect broadband providers. I interpret the positive coefficients across the broadband loan programs as suggestive evidence that the broadband loan programs had an impact on number of broadband providers. As number of broadband providers is imprecisely measured, finding even small effects is a positive sign for the effectiveness of the loan program. The other covariates in the fixed effects models are all of the predicted sign except for the coefficient associated with average annual pay for the ZCTA and time fixed effects model which produces a counter-intuitive result that average pay is inversely related with broadband availability. While this result is puzzling, I believe it is symptomatic of a larger issue that these may not be appropriate models which I expand upon at the end of this section. I now turn to the random effects models for further inspection of the loan programs.

Table 8 displays the random effects estimators which are an alternative approach to the assumption of fixed effects. The three random effects estimators are considered: ZCTA random effects, time random effects, and ZCTA and time random effects. The ZCTA random effects implies an error component associated with each ZCTA that is uncorrelated with the other regressors, α_z is present. The time random effects implies an error component in the model that, α_t is present. The last contains both of the previous error components which implies both α_z and α_t are present and uncorrelated with other regressors.

Across all of the random effects models a similar story to the fixed effects estimators is seen for the broadband loan program coefficients, all of them are positive and statistically significant. I see this as suggestive evidence that the broadband loan programs had a positive impact, but I do not conclude the broadband loan program was effective through this result. As an additional cross check, the other covariates do maintain the predicted sign across the models[4].

Although it is tempting to conclude that inference from both the fixed and random effects models would indicate that the broadband loan programs had a significant impact for improving broadband access across ZIP codes, further inspection of regression diagnostics indicate that this conclusion is not necessarily appropriate. For proper inference, both a fixed and random effects estimator in this setting necessitate that the error term approximately follows a normal distribution. An effective

[4] The coefficients associated with TRI, ZIP code area, population density, and establishment density are not displayed because of limited table space. The results available upon request.

Table 8: Random Effects

	Dependent variable:		
	Prov_num		
	(1)	(2)	(3)
iloans	2.983***	0.045**	0.118***
	(0.028)	(0.018)	(0.018)
log(est)	1.040***	1.015***	0.964***
	(0.006)	(0.002)	(0.005)
log(Pop_IRS)	0.722***	0.486***	0.595***
	(0.008)	(0.003)	(0.008)
logAPay_R2	0.425***	0.283***	0.033***
	(0.017)	(0.008)	(0.012)
rucadj	0.451***	−0.087***	0.001
	(0.026)	(0.008)	(0.024)
rucnonadj	0.922***	0.215***	0.327***
	(0.031)	(0.010)	(0.029)
ZCTA Random Effect	Yes	No	Yes
Time Random Effects	No	Yes	Yes
Observations	532,584	532,584	532,584
R^2	0.186	0.630	0.171
Adjusted R^2	0.186	0.630	0.171
Note:	*p<0.1; **p<0.05; ***p<0.01		

check for this assumption is through a Quantile-Quantile plot which plots the theoretical distribution of residuals across the sample distribution given a normal distribution. In figures 10 and 11, the Q-Q plots for the time fixed effect and time random effect models are shown. If the assumption of a normally distributed error is satisfied, then the points for each plot should roughly lie around the red dotted line. Significant disturbance from this assumption can be visually seen which indicates that these models are not appropriate for modeling the number of broadband providers across the US for 1999 to 2008. This divergence is likely due to the fact that number of broadband providers is a count statistic which takes on a Poisson-like distribution. Because of this, I turn to a more appropriate model which takes into account the count nature of the number of broadband providers and allows for more appropriate inference on the broadband loan programs.

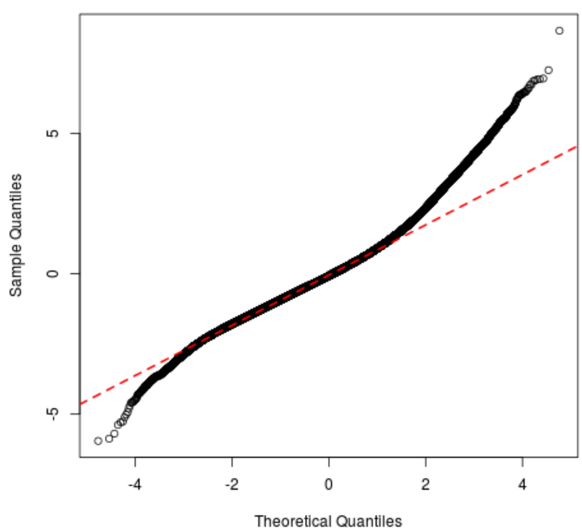


Figure 10: *Time Fixed Effects Q-Q Plot*

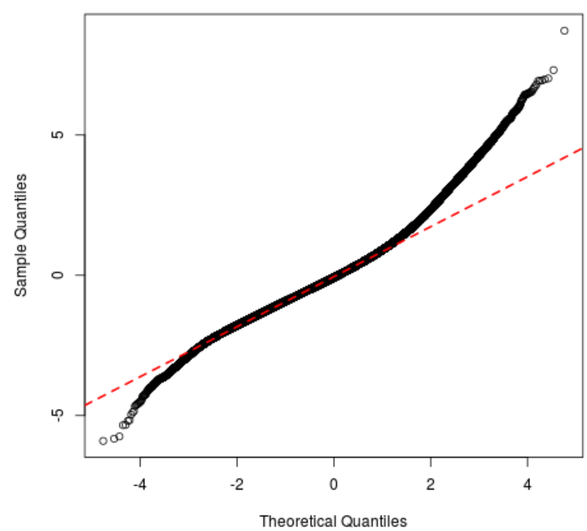


Figure 11: *Time Random Effects Q-Q Plot*

5.2 Count Panel Methods

I start with by making use of the count nature of the broadband providers variable by assuming it follows a Poisson distribution:

$$
\begin{aligned}
Prov_{z,t}|X_{z,t},\beta \sim Pois\left(\lambda_{z,t}\right) \\
log\left(\lambda_{z,t}\right) = \beta_0 + \beta_1 Loan_{z,t} + \beta_2 X_{z,t} + \tau_t + \varepsilon_{z,t}
\end{aligned} \tag{5}
$$

This nonlinear model is solved via maximum likelihood estimation (MLE). In the context of a regression of this form, the coefficients associated with the model can be interpreted as the effect of a one-unit change in regressors on the conditional mean. For the main variable of interest, receipt of a broadband loan, this can be seen as a biannual percentage increase in the number of broadband providers for a ZIP code (the sample average across all years is approximately 5). Another way to express the effects of a change in a predictor is to present the average response across the entire sample. This is done by taking the average across the fitted values multiplied by the coefficient of interest. For the broadband loan programs, subsetting the data using only the ZIP codes which received the broadband loans is an appropriate candidate of presenting the marginal effects of the broadband loan program's impact effectiveness (Cameron and Trivedi, 2005).

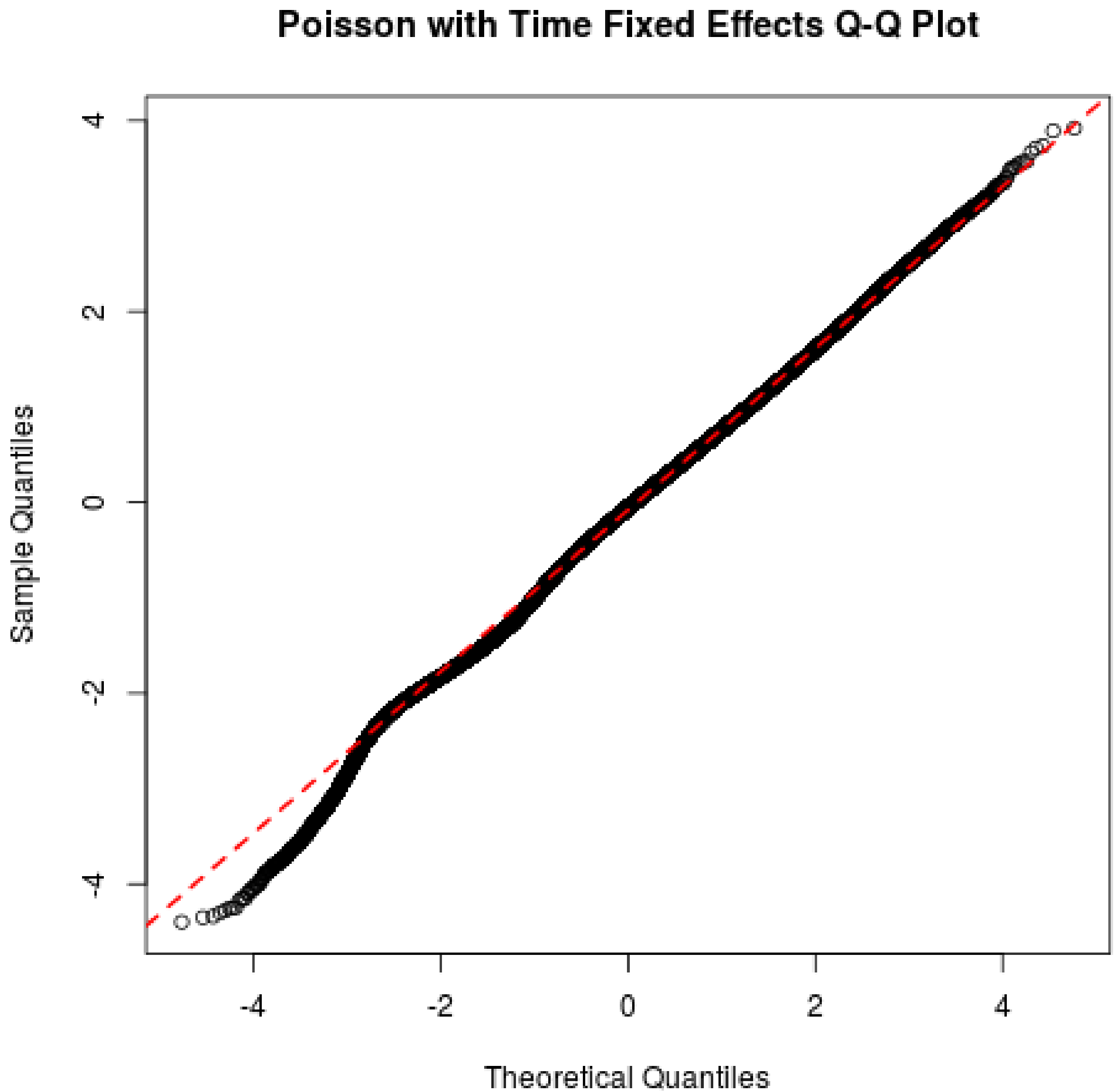
One potential issue with a Poisson regression is that a Poisson distribution has a limiting property that the conditional mean and the conditional variance are equal, typically termed equi-dispersion. In the case that the distribution of the fitted values does not reflect equi-dispersion, either resulting from over-dispersion or under-dispersion, the standard errors associated with the coefficients of the regressors may be incorrect. Two solutions are presented to account for this potential problem in the form of a quasi-MLE estimator (Quasi-Poisson) and use of a negative binomial distribution for the dependent variable.

Table 9 presents results of the preferred model specification of modeling the number of broadband providers across 1999 to 2008. The first model is the Poisson regression, the second is a Quasi-Poisson regression, and the third is a Negative Binomial Regression. In addition, figure 12 displays the Q-Q plot for the Poisson regression which visually indicates a better fit than previous estimators. The Poisson and Quasi-Poisson produce identical estimates for the coefficients but the standard errors differ due to relaxing the assumption that the conditional mean and variance are identical. The standard errors are larger for the Quasi-Poisson model and the dispersion parameter associated is less than one (0.68), indicating under-dispersion in the predictions which can also be seen at

Table 9: Poisson Regressions

	Dependent variable:		
	Prov_num		
	Poisson	*glm: quasipoisson* *link = log*	*glm: Negative Binomial(1)* *link = log*
	(1)	(2)	(3)
iloans	0.010***	0.010***	0.007
	(0.004)	(0.003)	(0.004)
log(est)	0.246***	0.246***	0.247***
	(0.001)	(0.0004)	(0.0005)
log(Pop_IRS)	0.077***	0.077***	0.090***
	(0.001)	(0.0005)	(0.001)
logAPay_R2	0.093***	0.093***	0.065***
	(0.002)	(0.002)	(0.002)
rucadj	−0.035***	−0.035***	−0.037***
	(0.002)	(0.002)	(0.002)
rucnonadj	−0.050***	−0.050***	−0.055***
	(0.003)	(0.002)	(0.002)
Time Fixed Effects	Yes	Yes	Yes
Observations	532,584	532,584	532,584
Log Likelihood	−998,786.700		−1,289,931.000
Akaike Inf. Crit.	1,997,631.000		2,579,920.000
Note:			*p<0.1; **p<0.05; ***p<0.01

Figure 12: Poisson Q-Q Plot



the lower tail of the Q-Q plot. The Negative Binomial regression also indicates under-dispersion through its associated dispersion parameter (0.17), although inspection across coefficient estimates between the Negative Binomial and Poisson regressions show similar results. In particular, the broadband loan coefficient for the Poisson regressions indicates an increase of the conditional mean of number of broadband providers of 1% attributed to the programs for each ZIP code biannually while the Negative Binomial indicates a 0.7% increase. This significant result gives evidence that the broadband loan programs, taken as a whole, were effective to some degree.

Calculating the average marginal effect for the Poisson regression indicates 0.046 additional broadband providers biannually per ZIP code receiving a loan. This interpretation is for the average ZIP code, which is not an accurate representation of the broadband loan programs because they were specifically targeted, and disbursed, to under-served rural communities as seen in tables 5 and 6. Further, there were two separate loan programs which may have varying degrees of effectiveness due to the scope of projects that were funded. I turn to stratifying the Poisson regression across the Pilot and Farm Bill loan programs as well as an interaction term across rural adjacent and rural non-adjacent counties to better capture the effectiveness of the loan programs.

Table 10 displays four models, the first is the first model from table 9 repeated for ease of comparability. The second model interacts the broadband loan variable across rural adjacent and rural non-adjacent to inspect whether or not the broadband loan programs were, on average, more effective for its targeted rural areas as opposed to the metro areas. To interpret the results for an average rural adjacent ZIP code that received a broadband loan, this is a linear combination of three coefficients (loan, Rural Adjacent, and the interaction of loan and Rural Adjacent) which combines to be 0.044, or an effect of 4.4% increase in conditional mean of providers. Constructing a Chi-square distributed linear hypothesis test of this combined effect equaling 0 is significant with a p-value of less than 0.001. The associated average marginal effect for ZIP codes in a rural adjacent area receiving a loan is an increase in 0.20 broadband providers biannually. A similar story results for the rural non-adjacent ZIP codes with a combined effect of 0.023 which rejects the null of a zero effect with a p-value of 0.007. The associated average marginal effect is an increase of 0.11 broadband providers biannually. The takeaway from the second model is that when broadband loans were disbursed to rural areas, the associated effects were larger than in metro areas as was the intended goal of the program.

The third model separates the broadband loan programs instead of combining the two. The results indicate that the Pilot loan program had a significant impact of increasing broadband providers by 4% of the conditional mean while the Farm Bill loan program did not have a significant impact. The average marginal effect is approximately 0.19 additional broadband providers biannually per ZIP code receiving a Pilot loan. The resulting difference can be either because of an associated lag of implementation of loans or that the Pilot loans were more effective. Previous results indicated

Table 10: Poisson Auxiliary Regressions

	Dependent variable:			
	Prov_num			
	(1)	(2)	(3)	(4)
iloans	0.010***	−0.043***		
	(0.004)	(0.005)		
ipilot			0.040***	0.049***
			(0.007)	(0.009)
icur			−0.003	−0.074***
			(0.004)	(0.005)
rucadj	−0.035***	−0.040***	−0.035***	−0.040***
	(0.002)	(0.002)	(0.002)	(0.002)
rucnonadj	−0.050***	−0.055***	−0.050***	−0.055***
	(0.003)	(0.003)	(0.003)	(0.003)
iloans:rucadj		0.127***		
		(0.008)		
iloans:rucnonadj		0.122***		
		(0.010)		
ipilot:rucadj				−0.027
				(0.017)
ipilot:rucnonadj				−0.003
				(0.017)
icur:rucadj				0.174***
				(0.009)
icur:rucnonadj				0.173***
				(0.012)
Time Fixed Effects	Yes	Yes	Yes	Yes
Observations	532,584	532,584	532,584	532,584
Log Likelihood	−998,786.700	−998,634.000	−998,772.900	−998,551.700
Akaike Inf. Crit.	1,997,631.000	1,997,330.000	1,997,606.000	1,997,171.000
Note:				*p<0.1; **p<0.05; ***p<0.01

that the Farm Bill loan program was better at disbursing loans to its targeted area, which may have actually been a deterrent in spurring broadband providers because the targeted areas were predicted to have low broadband access to begin with. In order to evaluate the effectiveness further, the Pilot and Farm Bill loan programs are broken down further.

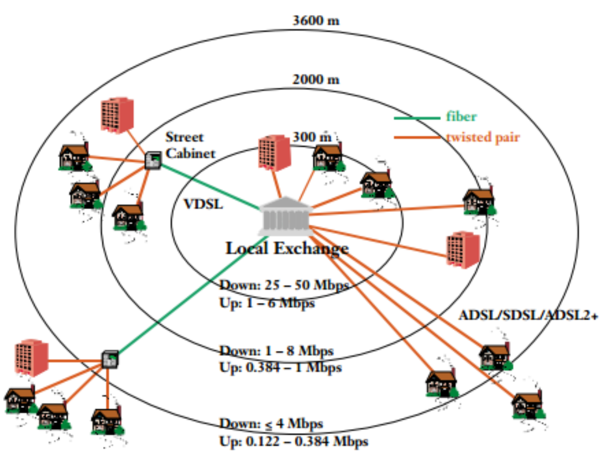
The fourth and final model separates the broadband loan programs and interacts each with the rural adjacent and rural non-adjacent dummies. This specification combines the previous two models for greater detail in each loan program's effectiveness in expanding broadband availability, but the effects are a combination of parameters. For the Pilot loans, across both the rural adjacent (-0.018) and non-adjacent (-0.008), the linear combination of coefficients is not statistically different from zero at respective p-values of 0.22 and 0.55. Since these values are measured relative to the baseline of a metro ZIP code, the Pilot loans effectiveness in increasing broadband availability stemmed from its limited loans that were disbursed to metro areas which has an estimated average marginal effect of 0.231 broadband providers biannually per ZIP code.

The Farm Bill loans differ from the Pilot loans across rural areas. Both the rural adjacent (0.06) and rural non-adjacent (0.045) linear combinations for the Farm Bill loans were statistically significant at p-values less than 0.001. The associated average marginal effects are respectively 0.28 and 0.21 broadband providers biannually per ZIP code. This result bodes well for evaluation of the targeted areas for broadband improvement with the Farm Bill loan programs. As the Farm Bill began disbursing loans around 2005, an average ZIP code in a rural adjacent area would expect to have seen an increase of over 1 broadband provider by June 2008, the end of the available data.

6 Spatial Inspection

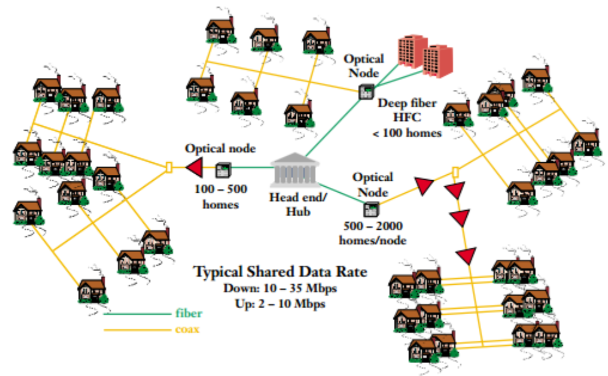
One concern for the process of broadband diffusion is that there can potentially be spatial correlation which poses a problem for inference. It is known that expanding the infrastructure of broadband technologies depends upon the existing infrastructure. Broadband expands in a spatial pattern from a hub outward to nodes and then to the end consumer. Each of these connections, from hub to node and node to consumer, have physical limitations as the broadband signal transmitted through wires degrades over increased distances which will limit capacity for internet access. Dependent upon the actual technology for broadband, a node is able to expand between 3 to 5 km of maximum bit rate to consumers to satisfy speed requirements for broadband[5] (Corning, 2005). When one ZIP code has access to broadband then there is a higher probability of broadband access for all ZIP codes within 3 to 5 km. The actual deployment of broadband depends upon the technology used to transmit data and the location of the deployed lines. As seen in figure 13 for DSL, the location of the local exchange and street cabinets determine the potential coverage of DSL for households and firms. In figure 14, we see that the process is similar for cable in that the location of the hub and optical nodes determine the potential coverage. The locations of local exchanges and hubs are not observed in the available data, which lessens the precision of estimates for broadband availability. However, if this diffusion process of broadband deployment exhibits spatial dependence that affects the estimated impacts of the broadband loan programs on broadband availability then this should be detectable.

Figure 13: Deployment Process of DSL



Network architectures for various forms of xDSL, note the xDSL bandwidth is dependent on distance from the local exchange/central office or the remote street cabinet, from Corning (2005)

Figure 14: Deployment Process of Cable



Cable TV, Hybrid Fiver Coax (HFC) Architectures, from Corning (2005)

I utilize geospatial methods in attempt to detect if spatial dependence poses a problem for the inference of the preferred count models. If spatial dependence is a problem, then inspection of the residuals for the preferred model should exhibit a significant level of spatial correlation. While failure to detect spatial correlation does not indicate that there is no spatial dependence, it can alleviate concerns or objections that the results from the count models are driven by spatial dependence.

[5] FCC definition for broadband is internet access of at least 200 kilobits per second downstream or upstream.

A more sophisticated way to evaluate the potential for spatial autocorrelation is through the use of an empirical variogram. A variogram is defined as the variance of the difference between field values at two locations across realizations of the field citecressie1993statistics. There are theoretical variograms that describe a spatial process via parameters which can be estimated, however before getting to that step it is best to look at the empirical variogram which is defined as:

$$\gamma\left(h\right)=Var\left(Y\left(s+h\right)-Y\left(s\right)\right) \tag{6}$$

where $Y()$ indicates the variable of interest as a function of the location, s denotes the location of a spatial process (usually with longitude and latitude) and h is the difference between two locations. This is the analogue to a partial autocorrelation function in the time series methods, except a variogram is the inverse of correlation. A typical empirical variogram of a spatial process will start with a low value on the variogram (indicating high correlation), then increase up to a particular value given a distance, and ultimately appear to be flat from there on. If this happens, then it indicates that closer observations are highly correlated while further away values are not as related. In the case of inspecting spatial correlation for broadband diffusion, the process is first modeled aspatially and then the residuals are inspected for spatial correlation.

I use the residuals from the baseline Poisson regression and map the points via the longitude and latitude of a ZCTA's centroid. The first empirical variogram that I evaluated was a pooled variogram, taking all of the 18 time periods and combining them into one. The ZCTAs are then binned up by distances so that all ZCTAs that are less than 2,500 meters apart are one grouping, the next is between 2,500 and 5,000 meters and so forth. The pooled variogram is seen in figure 15.

The pooled variogram does not appear to exhibit much of a pattern in the residuals based upon distance between ZCTAs. The semivariance appears to be fairly constant across all distances, which would indicate that a spatial model may not be necessary because of the lack of spatial correlation. Although visual inspection does not appear to indicate spatial correlation, I attempted to fit a functional variogram to the pooled residuals in R in the event that spatial correlation exists but is not visible. Attempts to fit a matern and exponential variograms to the residuals did not yield results because of convergence issues. Failure to fit a functional variogram is taken to be an indication that spatial methods are not appropriate for the model.

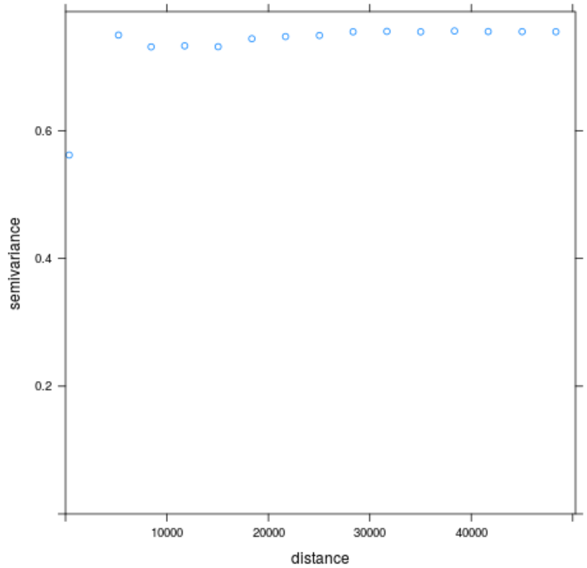


Figure 15: *Pooled Variogram*

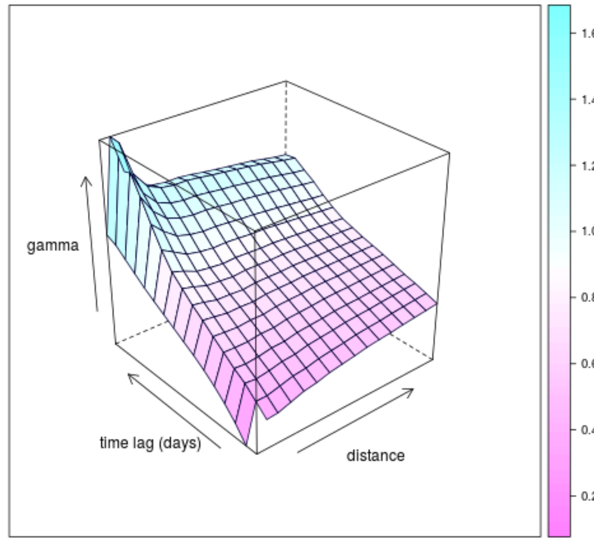


Figure 16: *Spatio-Temporal Variogram*

As a further corss check for spatial correlation, I turned to a spatio-temporal variogram displayed in figure 16. The results are as follows and a similar result as before is established. Visually, there does not appear to be a pattern in the residuals and further convergence issues to fit spatio-temporal variograms are encountered. While this lack of determining spatial correlation through estimation of functional variograms does not establish that spatial correlation does not exist, I take this result to mean that the modeling tools available to me cannot determine a spatial structure for estimating broadband diffusion. Because of this, I cannot establish a justification for use of spatial methods.

7 Conclusion

This paper's immediate goals are to evaluate the USDA broadband loan programs that disbursed loans to under-served communities from 2002 to 2007. The evaluation is two-fold by first examining if the programs were disbursed to their targeted population and then in determining if the areas that received the broadband loans experienced an increase in broadband availability. While there is anecdotal evidence from the Inspector General that both the Pilot and Farm Bill loan programs had misappropriated loans, on the whole both programs appeared to reach their intended target of under-served rural areas with fewer than 20,000 people. Further, there appears to be a detectable effect of an increase of approximately 0.092 additional broadband providers annually across ZIP codes receiving broadband loans. The associated increase in broadband availability indicates the Pilot loan program accrued more additional providers than the Farm Bill and that, in general, loans to rural areas had a larger impact than to metro areas.

The results largely dictate that the broadband loans had an effect of increasing broadband availability to their targeted areas. Results are further interpreted to indicate that the effect of increasing broadband availability has an associated lag after the disbursement of loans as seen that the marginal effects associated with the Pilot loan program are larger than the Farm Bill. It is also interpreted that the Farm Bill was more effective in reaching its targeted population as seen in the discrete choice models. I would expect for the increases in broadband availability to the Farm Bill areas to be more detectable beyond 2008. Unfortunately, this is not directly testable as FCC data beyond 2008 are compatible with the FCC data used in this analysis.

While this paper has determined an increase in broadband availability, it is unclear what the economic impact of an additional broadband provider on a ZIP code is. This stems from the inability to translate an additional provider into a measure of consumer surplus as publicly available data on broadband usage is not readily available. Future research aims address the economic impact of broadband availability through the use of USDA Agricultural Resource Management Survey, confidential farm level data.

References

Antonelli, C. (2003). The digital divide: understanding the economics of new information and communication technology in the global economy. *Information Economics and Policy 15*(2), 173–199.

Attewell, P. (1992). Technology diffusion and organizational learning: The case of business computing. *Organization Science 3*(1), 1–19.

Cameron, A. C. and P. K. Trivedi (2005). *Microeconometrics: methods and applications*. Cambridge university press.

Corning, I. (2005, June). Broadband technology overview. Technical Report WP6321.

Cowan, T. (2010). *An overview of USDA rural development programs*, Volume 20. Washington, DC.

Cressie, N. (1993). Statistics for spatial data: Wiley series in probability and statistics.

Czernich, N., O. Falck, T. Kretschmer, and L. Woessmann (2011). Broadband infrastructure and economic growth*. *The Economic Journal 121*(552), 505–532.

Development, U. S. R. (2007). *USDA Rural Development: bringing broadband to rural America*. US Dept. of Agriculture, Rural Development.

DiMaggio, P., E. Hargittai, et al. (2001). From the 'digital divide' to 'digital inequality': Studying internet use as penetration increases. *Princeton: Center for Arts and Cultural Policy Studies, Woodrow Wilson School, Princeton University 4*(1), 4–2.

FCC (2005). The internet: A short history of getting connected. https://transition.fcc.gov/omd/history/internet/. Accessed: 2014-12-01.

Geroski, P. A. (2000). Models of technology diffusion. *Research policy 29*(4), 603–625.

Grubesic, T. H. (2008). Spatial data constraints: Implications for measuring broadband. *Telecommunications Policy 32*(7), 490–502.

Kolko, J. (2010). A new measure of us residential broadband availability. *Telecommunications Policy 34*(3), 132–143.

Kolko, J. (2012). Broadband and local growth. *Journal of Urban Economics 71*(1), 100–113.

Kruger, L. G. (2013). Broadband loan and grant programs in the usda's rural utilities service. *Broadband Internet: Access, Regulation and Policy*, 95.

Norris, P. (2001). *Digital divide: Civic engagement, information poverty, and the Internet worldwide.* Cambridge University Press.

Wade, R. H. (2002). Bridging the digital divide: new route to development or new form of dependency? *Global governance*, 443–466.

Whitacre, B. E. (2010). The diffusion of internet technologies to rural communities: a portrait of broadband supply and demand. *American behavioral scientist*.

Whitacre, B. E. and B. F. Mills (2007). Infrastructure and the rural—urban divide in high-speed residential internet access. *International Regional Science Review 30*(3), 249–273.