**Appendix B. Technical Methodology**

**Data Sources**

This study combines detailed data on broadband deployment and fixed connections from the Federal Communications Commission (FCC), and demographics from the American Community Survey (ACS) to determine the variation in accessibility and subscription of broadband across the United States.

***FCC Form 477 Broadband Data***

The two primary datasets for broadband availability and subscription come from FCC’s Internet Access Service Reports as of December 2015.[[1]](#footnote-1) The FCC compiles these reports from Form 477 data that all facilities-based broadband providers are required to file twice a year.[[2]](#footnote-2)

***Fixed Broadband Deployment Data***

For the analysis on broadband availability, the Fixed Broadband Deployment Data from Form 477 was used. This data is available at the census block level, which is the smallest geographic entity for which the U.S. Census Bureau tabulates decennial census data.[[3]](#footnote-3) Each broadband provider submits lists of census blocks where they offer Internet access service to at least one location at speeds exceeding 200 kbps in at least one direction, with additional information about the technology and bandwidth provided. Each row in the dataset compiled from this data corresponds to a provider within a census block. In the raw data, there can thus be multiple rows for any given census block, based on the number of providers offering broadband service in that block. Note that a provider that reports deployment of a particular technology and bandwidth in a census block may not necessarily offer that particular service everywhere in the census block. This dataset thus provides details on whether or not each census block has any service provision, and about the number of providers for each speed category in a block.

***Residential Fixed Connections Data***

For the analysis on broadband subscription, the FCC’s Residential Fixed Connections dataset was used. This dataset provides the number of residential fixed Internet access service connections of at least 10 Mbps downstream and 1 Mbps upstream per 1,000 households by Census Tract based on Form 477 fixed broadband subscribership data.[[4]](#footnote-4) In other words, this data tells us the proportion of the population within any given census tract that has ‘adopted’ broadband by subscribing to a fixed connection with at least 10Mbps downstream speeds. The FCC reports this data in quintiles, plus a sixth “zero percent” category.

***American Community Survey Demographics data***

The US Census Bureau’s American Community Survey (ACS) was the source for demographic data. To assign demographic data to all census tracts, the 2011-2015 five-year ACS estimates were used, which are the most recent vintage as of publication.[[5]](#footnote-5) Primary variables in this dataset include population, poverty, race, age, education, income, nationality, and housing characteristics of residential tracts. All data are downloaded through the National Historical Geographic Information System (NHGIS).[[6]](#footnote-6) The ACS 5-year estimates represent data collected between January 1, 2011 and December 31, 2015. It is the most geographically-precise dataset offered by the ACS, but the data represents a 5-year average in all of the reported data categories.

***Decennial Census data***

The 2010 decennial census was the source for aggregate population data within census blocks. This is the only public data source providing universal population counts at this geographic level. The data was used in the aggregation process described below.

**Constructing a merged broadband-demographics dataset**

***Aggregation of availability data from block to tract level***

Broadband availability data is at a more detailed census block level, while subscription and demographic data is at the census tract level. This necessitates the aggregation of availability data from block to tract level in order to make it compatible with the other datasets and create a merged model-ready dataset. To achieve this aggregation, the availability data was first aggregated by census block, wherein each row specifies whether or not there is service provision by different speed categories for a specific census block based on the number of providers in the raw FCC data. The speed categories are chosen based on the following characteristics:

|  |  |
| --- | --- |
| **Speed Category** | **Justification** |
| 3 Mbps and above | Broadband speed permitting high-quality, non-video Internet browsing and some video streaming |
| 10 Mbps and above | Broadband speed standard in some developed countries |
| 25 Mbps and above | Formal FCC definition of broadband speed |
| 1,000 Mbps (or 1 Gbps) and above | A prominent speed tier that is advertised and sold, often in the context of Fiber Optic cable |

In aggregating from block to tract, we first calculate the total population in all blocks within the tract that have no service provision. If a majority (greater than 50 percent) of population in all blocks in a given tract have no provider in a certain speed category, we qualify the tract as having no service provision. If exactly 50 percent of population in the tract has no provider, then that tract is given the benefit of the doubt and categorized as having service provision at the highest-qualifying speed tier. If a majority (greater than 50 percent) of population in all blocks in a given tract have one or more providers in a certain speed category, it then means that the tract has service provision.

It’s critical to qualify the potential assignment errors inherent to these techniques. First, the FCC Form 477 data qualifies an entire census block as being serviced by a given provider if any address is serviced. This naturally leads to potential overestimation of coverage within a given census block. Second, because the demographic and income data we use in this analysis are not available at the block level, we aggregate blocks to census tracts, which required us to either qualify an entire tract as served or not served by broadband. However, since broadband may only serve some of the census blocks within a tract, this method will miscategorize a certain share of the population. Our methodology causes these mischaracterizations to skew in a positive direction—meaning, on net we overestimate the number of people served by broadband. This margin of error does not meaningfully change the overall findings, but it does mean individual tract data, in particular, should be read with caution.

Finally, note that there are two types of speed data in the Form 477 dataset, advertised speed and typical speed. Based on comments from an expert roundtable held by Brookings in the summer of 2016, advertised speed was chosen as the speed variable used in the analysis.

***Geographic typologies***

We assign all census tracts to one of the following, mutually exclusive categories: “large metro,” “small metro,” and “nonmetropolitan” areas. Of the 381 metropolitan statistical areas (MSA) defined by the

Office of Management and Budget in 2013, the 100 with the largest total population---per the 2010 decennial census---are considered large metro areas. Tracts falling within any MSA not among the 100 largest are considered to be in a small metro area. All others are considered rural. For tracts that fall within the 100 largest metropolitan areas, we further categorize them as either “cities” or “suburbs.” City tracts are those that fall within the first named city in the MSA’s official title, as well as all other cities in the MSA title with population totals greater than 100,000 (also per the 2010 decennial census). Tracts that fall within the 100 largest metropolitan areas but outside of these primary cities are considered suburban.

**Model**

The brief uses an Ordinary Least Squares (OLS) multiple regression to investigate the association between demographic variables and broadband subscription rates at the census tract level. The model is based on the same FCC and ACS data vintages described in this appendix. The model did not include female or male share of tract population because past research work did not find rates to vary by gender.[[7]](#footnote-7)

We predict broadband subscription rates based on the following formula:

Where *t* designates tract and *p* designates the ten independent variables.

The ten independent variables are: Poverty Rate; Population Density per Mile; Broadband Availability at 25 Mbps (Boolean); Share of Population over Age 25 with No More than a High School Diploma; Share of Population Aged 65 or Older; Foreign Born Share of Total Population; Black Share of Total Population; Hispanic Share of Total Population; Rural Neighborhood (Boolean); and Families Share of Total Population.

Because the FCC reports tract-level subscription data in categorical quintiles, we converted each quintile into the median value of the given quintile to create a continuous dependent variable. For example, the “0 to 20 percent” category became a value of 10 percent. To check for validity, we also tested the categorical data as the upper bound of the quintile and the model reported the exact same results.

**Table A1: Regression of Variables on Metropolitan Broadband Subscription Rates, 2014 Single Year⁺**



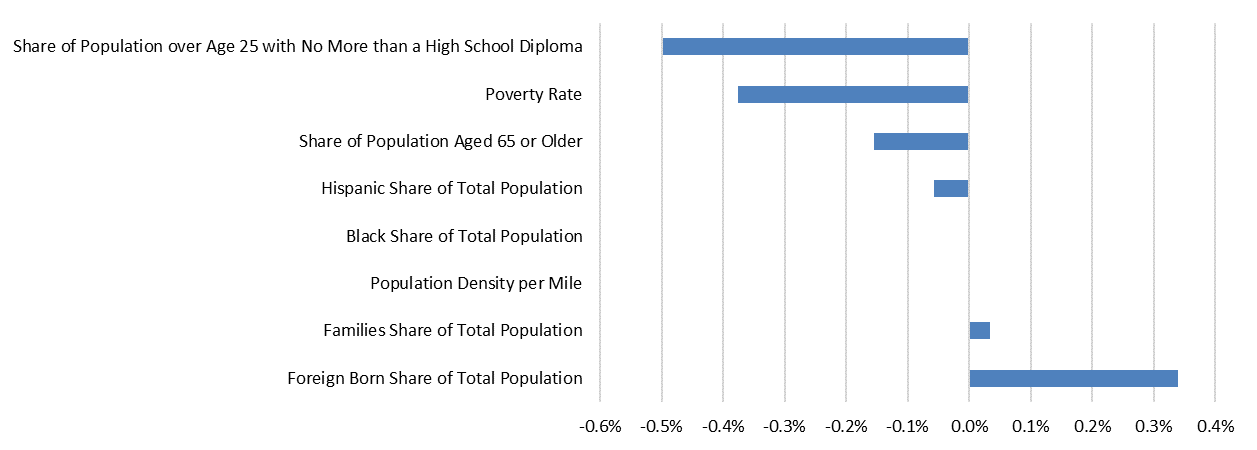
*Source: Brookings Institution analysis of 2011-15 American Community Survey and FCC data*

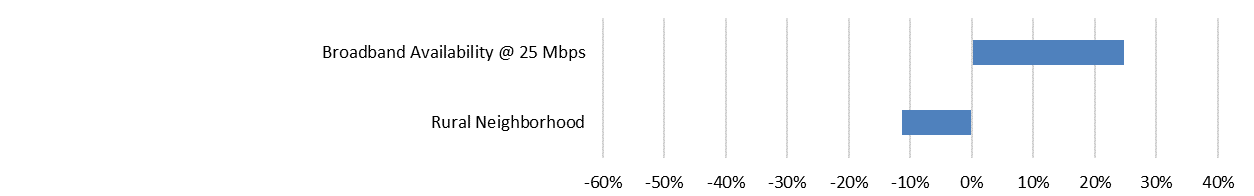
The results of this regression, plus seven alternative models, are shown in Table A1. All told in Model 8—the preferred model—nine variables were found to have significance at the 1 percent level. Overall, this model explains about 55 percent of the variation in broadband subscription for the 72,222 tracts that had data for all fields.

**Model Analysis**

Overall, the model demonstrates the strong, significant relationships that exist between a number of neighborhood characteristics and broadband subscription levels. The preferred model includes ten variables, nine of which were significant (Figure A1). It is worth stressing these results are not suggesting causality or directionality to the results; rather, they present relationships between subscription rates and the given independent variables. Similarly, it’s important to recognize that the model cannot specify many variables likely to impact subscription demand, including digital skills.

**Figure A1: Estimated change in neighborhood broadband subscription rate per 1 percentage point increase in the variable, 2015 \***





*\* adjusted R² = 0.547; n = 72,222; F = 8,705.26*

*Source: Brookings Institution analysis of 2011-15 American Community Survey and FCC data*

Education levels and poverty are associated the largest effects, and both relate to broadband subscription in ways one would expect.[[8]](#footnote-8) For education, every 1 percent increase in the share of residents with no more than a high school diploma suggests broadband subscription rates will drop by 0.49 percent. This may seem like a small effect, but it is far from it. While adults with no more than a high school diploma represent 41 percent of the median neighborhood’s population, there are over 1,700 neighborhoods where this population share jumps above 75 percent. Based on the model’s findings, it is not surprising that 71 percent of these low-educated neighborhoods fall into the low-subscription category.

Poverty rates function in a similar way: every 1 percent increase in the share of residents living in poverty suggests broadband subscription rates will drop by 0.38 percent. Here again, the median neighborhood poverty rate of 13 percent says little about the 4,379 neighborhoods where over 40 percent of people live below the poverty line. For those neighborhoods of concentrated poverty, a lack of broadband subscriptions is yet another barrier to opportunity for its residents. These risks only grow when neighborhoods house people with both lower education levels and higher poverty rates.

Age also has a negative relationship with subscription rates, although the effects are smaller than education and income. As expected, neighborhoods with larger shares of retirement-age individuals demonstrate lower subscription levels. Specifically, every 1 percent increase in the share of residents over the age of 65 suggests broadband subscription rates will drop by 0.15 percent. These impacts will be especially acute in the hundreds of neighborhoods where retirees can represent over half of the population.

In contrast, the share of foreign-born residents carries the largest positive effect among the demographic variables. The model finds that every 1 percent increase in the foreign-born population leads to a 0.34 percent increase in broadband subscription. One possibility is that foreign-born populations’ enhanced global connectivity—such as family, social, and even business connections outside the country—could increase their interest in broadband subscriptions. However, the level of relationship warrants further study.

The model also found racial groups’ share of neighborhood population to have relatively small effects on broadband subscription. Larger Hispanic populations are significant but their negative impacts on subscription are relatively small, while larger black populations do not significantly impact subscription rates. These findings do not suggest these groups are unimportant to track for broadband policy purposes, especially since surveys regularly confirm major gaps by race, and individual-level regression models have found race to be a significant factor on subscription.[[9]](#footnote-9) Instead, this neighborhood-level model confirms further research should investigate racial gaps in broadband subscription and how they relate to other features such as age, income, and education.

Finally, the model did report some critical results from the two dummy variables. When controlling for broadband availability at 25 Mbps—the same measure used in the first two findings and a faster benchmark than the 10 Mbps qualification used in the subscription data—the model found the presence of neighborhood broadband connections increases likely subscription rates by a striking 25 percent. This result suggests that deploying even higher-speed networks can promote greater subscription rates among local populations. The rural flag also was significant and led to an 11 percent drop in subscription rates. Considering the model controls for other factors like availability, income, and education, it is possible that other barriers to subscription do exist specifically for rural residents.

In the future, this model’s explanatory ability should improve with more specific subscription data provided by the FCC. Due to the quintile nature of reported broadband subscription rates, there are unknown limitations to the model’s accuracy.

1. FCC Internet Access Service Reports can be found online at <https://www.fcc.gov/reports-research/reports/internet-access-services-reports/internet-access-services-reports> [accessed May 2017]. [↑](#footnote-ref-1)
2. Four types of entities must file Form 477’s to the FCC: Facilities-based Providers of Broadband Connections to End Users, Providers of Wired or Fixed Wireless Local Exchange Telephone Service, Providers of Interconnected Voice over Internet Protocol (VoIP) Service, and Facilities-based Providers of Mobile Telephony (Mobile Voice) Service. For more information, see the FCC description online at <https://transition.fcc.gov/form477/WhoMustFileForm477.pdf> [accessed May 2017]. [↑](#footnote-ref-2)
3. Blocks are statistical areas bounded by visible features, such as streets and railroad tracks, and by nonvisible boundaries, such as selected property lines, city and county limits. Generally, census blocks are small in area; for example, a block in a city bounded on all sides by streets. Census blocks in suburban and rural areas may be large, irregular, and bounded by a variety of features, such as roads, streams, and transmission lines. Census blocks cover the entire territory of the United States, Puerto Rico, and the Island Areas. As of 2010, there are 11,078,297 census blocks in the 50 states and the District of Columbia, 77,189 blocks in Puerto Rico, and 10,850 blocks in the Island Areas. [↑](#footnote-ref-3)
4. Census Tracts are small, relatively permanent statistical subdivisions of a county or equivalent entity that are updated by local participants prior to each decennial census as part of the Census Bureau's Participant Statistical Areas Program. The primary purpose of census tracts is to provide a stable set of geographic units for the presentation of statistical data. Census tracts generally have a population size between 1,200 and 8,000 people, with an optimum size of 4,000 people. As of 2010, there are 73,056 census tracts in the 50 states and the District of Columbia, 945 tracts in Puerto Rico, and 132 tracts in the Island Areas. [↑](#footnote-ref-4)
5. Full details on the American Community Survey can be found online at <https://www.census.gov/programs-surveys/acs/> [accessed May 2017]. [↑](#footnote-ref-5)
6. The National Historical Geographic Information System website is <https://www.nhgis.org/> [accessed May 2017]. [↑](#footnote-ref-6)
7. Andrew Perrin and Maeve Duggan, “Americans’ Internet Access: 2000-2015” (Washington, DC: Pew Research Center, 2015). [↑](#footnote-ref-7)
8. While prior research findings finding correlations between poverty and education, the model’s variance inflation factors (VIF) did not find multicollinearity among any of the variables. [↑](#footnote-ref-8)
9. For an example of wireline broadband subscription research with an individual-level model, see: S. Derek Turner, “Digital Denied: The Impact of Systemic Racial Discrimination on Home-Internet Adoption,” (Washington: Free Press, 2016). [↑](#footnote-ref-9)