The Role of Economics, Demographics, and State Policy in Broadband Availability

Kenneth Flamm

February 7, 2005
Revised September, 2005

Preliminary
Comments Only Please

Preliminary Draft Presented at the PURC/London Business School Conference on
"The Future of Broadband: Wired and Wireless, 2005"
Gainesville, Florida, February 24, 2005

Kenneth Flamm
Lyndon B. Johnson School of Public Affairs
The University of Texas at Austin
Box Y
Austin, Texas 78731
kflamm@mail.utexas.edu

The Role of Economics, Demographics, and State Policy in Broadband Availability

Kenneth Flamm

Abstract

This paper constructs a framework for modeling the determinants of broadband penetration in the United States, and applies it to a zip code-level database of economic, demographic, and policy variables constructed by the author. Statistically significant state effects exist, suggesting that, after controlling for a long list of economic and demographic factors, state-level policies appear to have played a role in accelerating or retarding broadband deployment. I find that geophysical variables are important determinants of differences in broadband deployment across zip codes, presumably because of their effects on broadband deployment costs. My results here suggest that omitting these variables from supply side analyses is likely to result in significant estimation biases in empirical econometric work. Geophysical variables are also excellent candidates for use as instruments in future work on broadband demand, where dealing with unmeasured variation in quality of service chosen by consumers is clearly an important problem for researchers, and availability of instrumental variables to deal with this issue has been problematic. Interesting results on the impact of a variety of socioeconomic variables on broadband deployment are available. Income and wealth variables unsurprisingly seem to be among the most important determinants of broadband penetration, but population density's role is much more problematic. When a full set of explanatory variables is used, the results suggest that absolute market size, not some measure of density relative to physical area, is the key determinant. I note that the "digital divide" reducing African American use of broadband, cet. par., seems to have disappeared after 2001, a statistically significant continuing lag in broadband deployment among Native Americans persists, and a new "digital divide" affecting native Hawaiian islanders opened up after broadband deployment began to accelerate in Hawaii in 2002 and 2003. A statistically significant impact of the eRate (Universal Service Fund) program linking schools and libraries was evident in 2002, with a positive impact on the availability of broadband service to consumers in affected zip codes. The eRate funding was distributed among zip codes in a way that made little practical difference for broadband availability, however.

The Role of Economics, Demographics, and State Policy in Broadband Availability

Kenneth Flamm[1]

Given the increasing emphasis among analysts on the role, actual and potential, of information technology in productivity growth,[2] it is not surprising that policies to accelerate deployment of broadband Internet services have sparked considerable discussion recently. Debates about broadband policy often start from the premise that federal, state, and local policies have played an important role in accelerating or retarding deployment of broadband technology. The primary purpose of my analysis is to examine whether substantial variation across states in state-specific factors—including regulatory policies, subsidies, and financial incentives—may have had a detectable impact on broadband deployment. My approach will be to utilize detailed public use data available on broadband deployment at the individual zip code level from the FCC, add to it economic and demographic data from the 2000 population census and 1997 economic census, data on "eRate" and rural health care Universal Service Fund grants, hydrological and terrain data from geophysical data bases, then use this data to estimate the parameters of an economic model describing entry into broadband service markets.

Four major findings emerge from my analysis. First, statistically significant state effects exist, suggesting that, after controlling for a long list of economic and demographic factors, state-level policies appear to have played a role in accelerating or retarding broadband deployment. Second, I find that geophysical variables are important determinants of differences in broadband deployment across zip codes, presumably because of their effects on broadband deployment costs. My results here suggest that omitting these variables from supply side analyses is likely to result in significant estimation biases in empirical econometric work. Geophysical variables are also excellent candidates for use as instruments in future work on broadband demand, where dealing with unmeasured variation in quality of service chosen by consumers is clearly an important problem for researchers, and availability of instrumental variables to deal with this issue has been problematic.

Third, income and wealth variables unsurprisingly seem to be among the most important determinants of broadband penetration. Population density's role is much more problematic. When a full set of explanatory variables is used, the results suggest that absolute market size, not some measure of density relative to physical area, is the key determinant. Finally, interesting results on the impact of a variety of socioeconomic

---

[1] I thank the Policy Research Institute of the LBJ School of Public Affairs, University of Texas at Austin, for financial support for this research. Without implicating them in my errors, I am also grateful to Anindya Chaudhuri for his expert assistance in assembling portions of the data, and to Gerry Faulhaber, Mark Jamison, Jim Prieger, Bob Wilson, Shane Greenstein, Colin Cameron, Chandler Stolp, and participants at the 2005 PURC/London School Conference on the "Future of Broadband," the CIDE-LBJ School Conference on "Internet Use in the Americas," the AEI/Brookings SIEPR 2005 Telecommunications Conference, and seminars at the University of California at Davis and the University of Texas at Austin, for helpful comments on previous versions of this paper.

[2] Influential studies suggesting links between IT deployment and aggregate productivity growth include Oliner and Sichel (2000), Jorgenson (2001), U.S. President, Council of Economic Advisors (2001). A more skeptical view can be found in Gordon (2000).

variables on broadband deployment are available. As examples, I note that the "digital divide" reducing African American use of broadband, cet. par., seems to have disappeared after 2001, a statistically significant continuing lag in broadband deployment among Native Americans persists, and a new "digital divide" affecting native Hawaiian islanders opened up after broadband deployment began to accelerate in Hawaii in 2002 and 2003. A statistically significant impact of the eRate (Universal Service Fund) program linking schools and libraries was evident in 2002, with a positive impact on the availability of broadband service to consumers in affected zip codes.

The FCC View of US Broadband Deployment

The Federal Communications Commission has been gathering data on U.S. broadband service deployment since 1999. The FCC defines a **high-speed** ["broadband"] **line** to be one with a speed exceeding 200 kilobits per second (kbps) in at least one direction, while an **advanced services line** is a high speed line with a 200kbps rate in both directions. There are basically two types of information that are gathered. First, providers of a least 250 high-speed connections within a single state are required to provide state-level data on numbers of lines in service. Providers of less than 250 lines may also voluntarily provide the FCC the same information, but apparently rarely do.[3]

Second, each service provider is required to identify each zip code in which it supplies at least one high-speed line. Obviously, service providers do not supply information for zip codes in which no high-speed service is offered by any provider, and the FCC must estimate these numbers. In doing so, the FCC uses procedures that have a significant effect on the numbers of "zero" service provider zip codes implicitly estimated within its statistics.[4]

Zip codes are not designed as geographic descriptors, but rather as an organizing mechanism for mail delivery routes. Roughly speaking, there are two broad classes of zip codes: "point" zip codes that route mail to a single point (typically a post office with post office boxes or general delivery service, or a large organization), and "geographic" zip codes that funnel mail to a carrier delivery route covering some geographic area. The FCC (as do many commercial zip code data vendors) takes point zip codes and reassigns people living in (or telecomm vendors serving) a mailing address associated with that zip to the closest "geographic" zip code.

Thus, only geographic zip codes show up in the universe of zip codes that the FCC uses in its reports. It is likely that significant amounts of sparsely populated territory with no regular mail carrier service are not included within the boundaries of the geographic zip codes that are caught in this net. Any people or services associated with these "point"

---

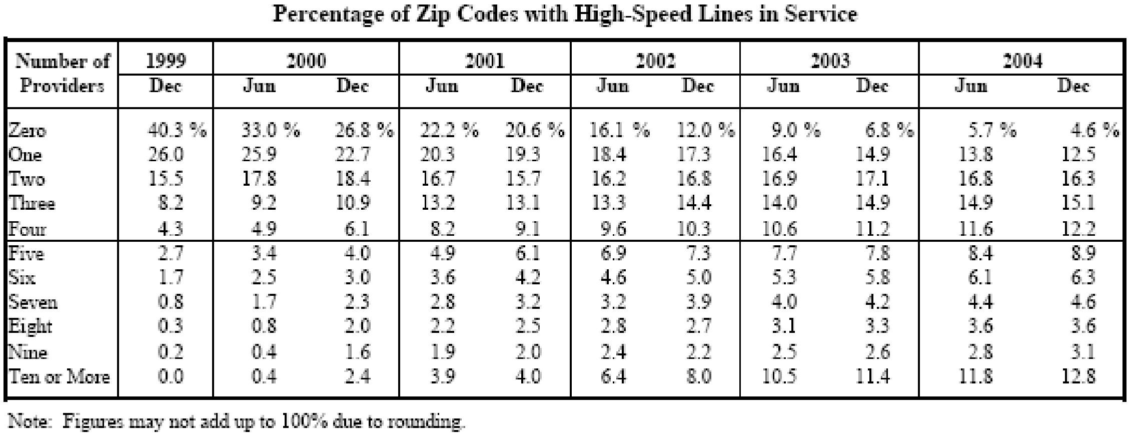
[3] Such voluntarily reported lines accounted for less than .05% of high-speed lines in recent submissions. See FCC, Industry Analysis and Technology Division, Wireline Competiton Bureau, **High-Speed Services for Internet Access: Status as of December 31, 2003**, June 2004, p. 2, available at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/IAD/hspd0604.pdf.

[4] This description is based on my understanding of FCC procedures, based in turn on a teleconference with Roger Wouck, Craig Stroup, Jim Eisner, and Ken Lynch of the FCC on January 27, 2005.

zips are reassigned to the nearest "geographic" zip, whether or not they actually live or operate within the boundaries of the mail delivery area defined by the geographic zip.

If a zip code does not show up in the FCC zip code data bases as associated with any telecommunications service providers, then, it does not necessarily mean that no service is provided to individuals within that zip if it is a point zip. It is possible that service is indeed being provided to an address making use of that point zip, but credited instead to the closest geographic zip. It is also possible that no one receiving mail at that point zip is being provided the service—there is simply no way to tell without accessing the FCC's confidential database. This mapping of telecomm consumers is quite different from the manner in which the Census maps zip codes to physical regions lacking normal mail carrier service, discussed below.

Table 1 shows aggregate U.S. data on "geographic" zip codes in which differing numbers of broadband service providers were available. Note that in December of 1999, over 40% of U.S. zip codes had no providers of high-speed lines; in December 2004, less than 5% of U.S. zip codes had no reporting high-speed line providers.[5]



Percentage of Zip Codes with High-Speed Lines in Service

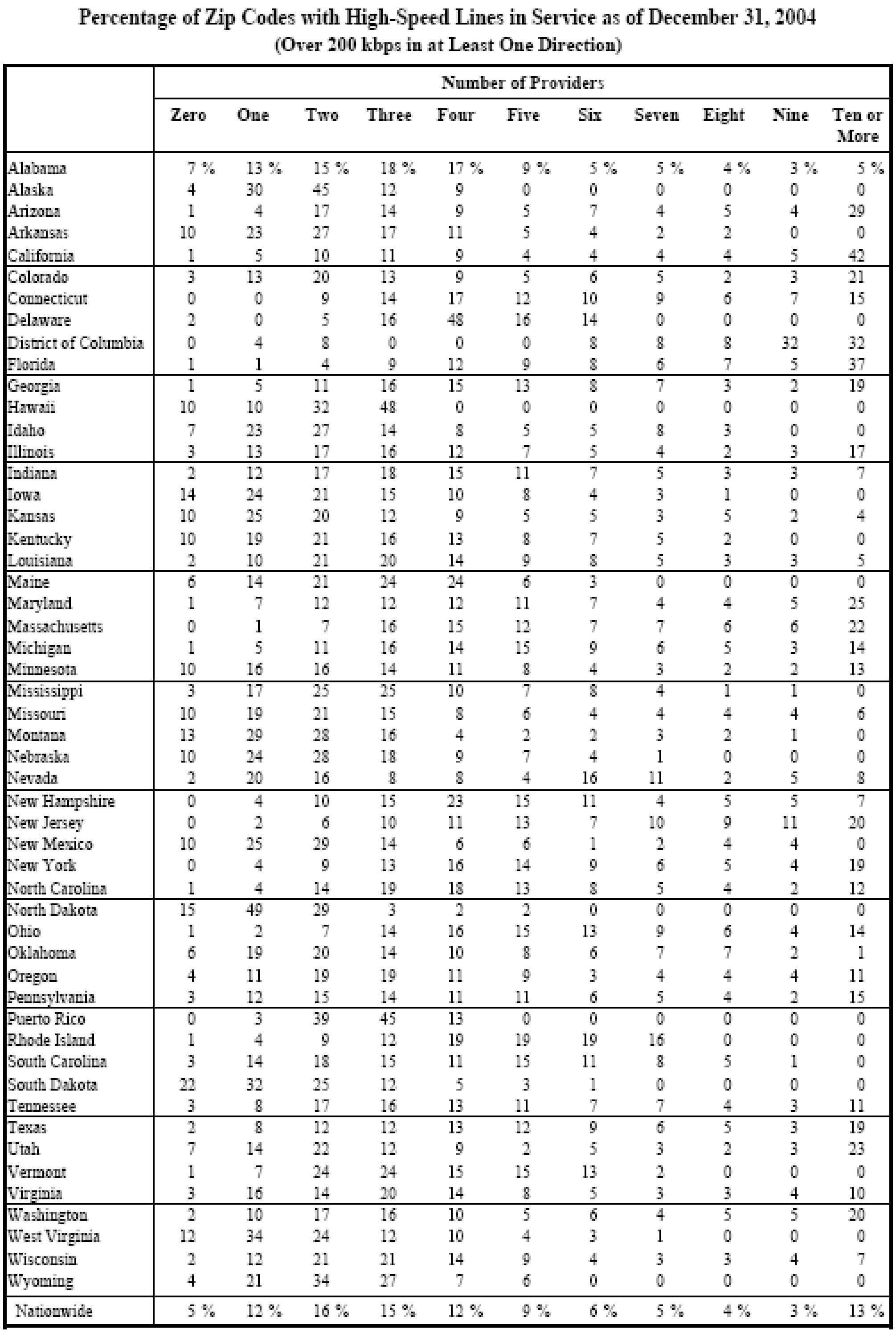
Number of Providers	1999	2000		2001		2002		2003		2004	
	Dec	Jun	Dec	Jun	Dec	Jun	Dec	Jun	Dec	Jun	Dec
Zero	40.3 %	33.0 %	26.8 %	22.2 %	20.6 %	16.1 %	12.0 %	9.0 %	6.8 %	5.7 %	4.6 %
One	26.0	25.9	22.7	20.3	19.3	18.4	17.3	16.4	14.9	13.8	12.5
Two	15.5	17.8	18.4	16.7	15.7	16.2	16.8	16.9	17.1	16.8	16.3
Three	8.2	9.2	10.9	13.2	13.1	13.3	14.4	14.0	14.9	14.9	15.1
Four	4.3	4.9	6.1	8.2	9.1	9.6	10.3	10.6	11.2	11.6	12.2
Five	2.7	3.4	4.0	4.9	6.1	6.9	7.3	7.7	7.8	8.4	8.9
Six	1.7	2.5	3.0	3.6	4.2	4.6	5.0	5.3	5.8	6.1	6.3
Seven	0.8	1.7	2.3	2.8	3.2	3.2	3.9	4.0	4.2	4.4	4.6
Eight	0.3	0.8	2.0	2.2	2.5	2.8	2.7	3.1	3.3	3.6	3.6
Nine	0.2	0.4	1.6	1.9	2.0	2.4	2.2	2.5	2.6	2.8	3.1
Ten or More	0.0	0.4	2.4	3.9	4.0	6.4	8.0	10.5	11.4	11.8	12.8

Note: Figures may not add up to 100% due to rounding.

Table 1

---

[5] Note that these recently published numbers differ from the FCC's original published reports for the years 1999 and 2000. Perceived problems in the early FCC numbers led to an effort to "clean up" data for these years and create the revised numbers. (Per my conversations with FCC officials.).



Percentage of Zip Codes with High-Speed Lines in Service as of December 31, 2004
(Over 200 kbps in at Least One Direction)

	Number of Providers										
	Zero	One	Two	Three	Four	Five	Six	Seven	Eight	Nine	Ten or More
Alabama	7 %	13 %	15 %	18 %	17 %	9 %	5 %	5 %	4 %	3 %	5 %
Alaska	4	30	45	12	9	0	0	0	0	0	0
Arizona	1	4	17	14	9	5	7	4	5	4	29
Arkansas	10	23	27	17	11	5	4	2	2	0	0
California	1	5	10	11	9	4	4	4	4	5	42
Colorado	3	13	20	13	9	5	6	5	2	3	21
Connecticut	0	0	9	14	17	12	10	9	6	7	15
Delaware	2	0	5	16	48	16	14	0	0	0	0
District of Columbia	0	4	8	0	0	0	8	8	8	32	32
Florida	1	1	4	9	12	9	8	6	7	5	37
Georgia	1	5	11	16	15	13	8	7	3	2	19
Hawaii	10	10	32	48	0	0	0	0	0	0	0
Idaho	7	23	27	14	8	5	5	8	3	0	0
Illinois	3	13	17	16	12	7	5	4	2	3	17
Indiana	2	12	17	18	15	11	7	5	3	3	7
Iowa	14	24	21	15	10	8	4	3	1	0	0
Kansas	10	25	20	12	9	5	5	3	5	2	4
Kentucky	10	19	21	16	13	8	7	5	2	0	0
Louisiana	2	10	21	20	14	9	8	5	3	3	5
Maine	6	14	21	24	24	6	3	0	0	0	0
Maryland	1	7	12	12	12	11	7	4	4	5	25
Massachusetts	0	1	7	16	15	12	7	7	6	6	22
Michigan	1	5	11	16	14	15	9	6	5	3	14
Minnesota	10	16	16	14	11	8	4	3	2	2	13
Mississippi	3	17	25	25	10	7	8	4	1	1	0
Missouri	10	19	21	15	8	6	4	4	4	4	6
Montana	13	29	28	16	4	2	2	3	2	1	0
Nebraska	10	24	28	18	9	7	4	1	0	0	0
Nevada	2	20	16	8	8	4	16	11	2	5	8
New Hampshire	0	4	10	15	23	15	11	4	5	5	7
New Jersey	0	2	6	10	11	13	7	10	9	11	20
New Mexico	10	25	29	14	6	6	1	2	4	4	0
New York	0	4	9	13	16	14	9	6	5	4	19
North Carolina	1	4	14	19	18	13	8	5	4	2	12
North Dakota	15	49	29	3	2	2	0	0	0	0	0
Ohio	1	2	7	14	16	15	13	9	6	4	14
Oklahoma	6	19	20	14	10	8	6	7	7	2	1
Oregon	4	11	19	19	11	9	3	4	4	4	11
Pennsylvania	3	12	15	14	11	11	6	5	4	2	15
Puerto Rico	0	3	39	45	13	0	0	0	0	0	0
Rhode Island	1	4	9	12	19	19	19	16	0	0	0
South Carolina	3	14	18	15	11	15	11	8	5	1	0
South Dakota	22	32	25	12	5	3	1	0	0	0	0
Tennessee	3	8	17	16	13	11	7	7	4	3	11
Texas	2	8	12	12	13	12	9	6	5	3	19
Utah	7	14	22	12	9	2	5	3	2	3	23
Vermont	1	7	24	24	15	15	13	2	0	0	0
Virginia	3	16	14	20	14	8	5	3	3	4	10
Washington	2	10	17	16	10	5	6	4	5	5	20
West Virginia	12	34	24	12	10	4	3	1	0	0	0
Wisconsin	2	12	21	21	14	9	4	3	3	4	7
Wyoming	4	21	34	27	7	6	0	0	0	0	0
Nationwide	5 %	12 %	16 %	15 %	12 %	9 %	6 %	5 %	4 %	3 %	13 %

Table 2

Similarly, 26% of U.S. geographic zip codes had only one high-speed provider in December 1999, contrasted with only 12.5% in December 2004. California currently leads in broadband competition, with the largest share of its zip codes with 10 or more high-speed providers (42% in December 2004), trailed by Florida (37%), the District of Columbia (32%), Maryland (25%), and Utah (23%). (See Table 2.) Interestingly, there has been a notable decline in the share of zip codes with this highest level of competition in some of the above states since 2003. In December of 2003, the District of Columbia led with the largest share of its zip codes with 10 or more high-speed providers (63%), trailed by California (39%), Florida and Maryland (27%), and Utah (24%).

The least serviced zip codes at the end of 2004 were South Dakota (22% of zip codes with no providers, 32% with a single provider), North Dakota (15% of zip codes with no providers, 49% with one), West Virginia (12% with no provider, 34% with just one), Montana (13% and 29%), Nebraska (10% and 24%), and Iowa (14% and 24%). My home, Texas, is somewhere toward the front of the pack, with 19% of its area codes reporting 10 or more providers, and 2% and 8% of its zip codes, respectively, having zero or one provider.

As the FCC notes in its reports, high speed line provision seems to be correlated with population density (presumably because the cost of providing individual users such service declines with population density) and median household income (presumably because willingness to pay the higher prices associated with this service increases with income).[6] To what extent each of these factors is causally related to provision of high speed lines, and to what extent it is related to other, as yet unmentioned, factors, is an important question which I address in my analysis. I briefly preview these empirical results below by noting that they suggest that is absolute numbers of consumers per zip code, and not their density, which drives deployment; my results also confirm an important impact of per capita income on broadband diffusion.

Also, note that data where one to three providers have supplied lines are aggregated together in the public use FCC zip code data base, to protect company-sensitive information. This has some consequences when I build a statistical framework to model this data, as described below.

The Census View of Household Economics and Demographics

The most recent U.S. Census Bureau data on population and demographics released at the zip code level are the 2000 Census of Population and Housing figures, which are available for "zip code tabulation areas" (ZCTAs).[7] A very important point to make is

---

[6] See FCC, Industry Analysis and Technology Division, Wireline Competiton Bureau, **High-Speed Services for Internet Access: Status as of December 31, 2003**, June 2004, pp. 4-5, p. 21; **High-Speed Services for Internet Access: Status as of December 31, 2004**, July 2005, p. 5, p. 21;

[7] ZCTA-based Census data are approximations corresponding to actual zip codes. Their construction is explained at http://www.census.gov/geo/ZCTA/zcta_brch_prnt.pdf, and http://www.census.gov/geo/ZCTA/zcta.html. I have discarded "artificial" ZCTAs (unclassified areas, or areas consisting of bodies of water) which do not have a corresponding "real" zip code in the analysis that follows. The census data correspond to the estimates in the Census SF-3 (long form) data base, and were taken from the "Gazeteer" ZCTA file available at http://www.census.gov/geo/www/gazetteer/places2k.html, and from the version of the Census SF-3 database as extracted and made accessible at the University of Missouri's Missouri Census Data Center through http://mcdc2.missouri.edu/cgi-bin/uexplore?/pub/data/sf32000x.

that unlike the FCC, Census procedures map out the areas where households not served by regular mail service routes receive their mail through "point" zip codes, and assign these physical areas to their point zip codes.[8] Approximately 10 percent of zip codes reported in the 2000 census (3245 ZCTAs) are point zips that have been mapped to geographic areas by the Census Bureau. These are mainly sparsely populated rural areas which have simply been annexed by more densely populated neighboring zips in the FCC data.

The least problematic way to link FCC broadband availability by zip, to Census household data for zip codes (ZCTAs), is to restrict the analysis to "geographic" zip codes showing up in both the Census and FCC zip code pools. The FCC's practice of attributing point zip code service to nearby geographic zip is unlikely to create many "false positives" for any broadband availability at all in geographic zips (since it is probably pretty rare—but certainly not impossible—for a sparsely settled rural area with no mail delivery to be served with broadband while a nearby more populated area is not). It seems more likely that the FCC data might be overestimating the number of providers of broadband for ZCTAs corresponding to "geographic" zip codes to which "point" zip codes have been assigned. Restricting our attention to only those Census ZCTAs that are also geographic zip codes, then, is probably a decent approximation to which of these has any service at all. I recognize, however, that I am dropping the 10% of mainly rural ZCTAs that are most likely to not have broadband available.

A limited amount of data (principally establishment numbers, by two digit NAICS code) from the 1997 economic census is also available at the zip code level.[9] I have constructed a data set linking data from the 2000 population and 1997 economic censuses to the FCC "high speed" provider data just described. Every ZCTA corresponding to a geographic zip code in 2000,[10] has been "looked up" in the FCC public use broadband data zip code data files, and the corresponding number of high-speed line providers linked to data from the population census for 2000, and the economic census for 1997. All analysis that follows is based on the database I have constructed using this methodology.

Other Data

Additionally, I have gathered other data on additional, potentially relevant activities available at the zip code level. The Universal Service Administrative Company makes publicly available information on individual eRate (schools and libraries) and rural health care grants funded out of the Universal Service Fund, by zip code of the organization receiving the grant.[11] Data on committed funds (through fall 2004) for each of the

---

[8] For example, the Census ZCTAs for 2000 include 3245 "point" zip codes assigned to rural areas with post office box and general delivery service only. See http://www.census.gov/geo/ZCTA/zcta_tech_doc.pdf

[9] The economic census uses actual zip codes reported by businesses or their administrative units. The only figures available without substantial suppressed or missing detail at the zip code level are establishment numbers by 2-digit NAICS industries, which may be accessed at http://www.census.gov/epcd/ec97zip/downlzip.htm.

[10] This is available at http://www.census.gov/geo/www/tiger/zip1999.html.

[11] The data and other information are available at www.universalservice.org. A small number (94) of miscoded zip code entry errors in the more than 600,000 funded projects were corrected, and the data aggregated up to a funding year sum for an entire zip code.

funding years from 1998 to 2004 for the eRate and rural health care programs were collected, as were authorized disbursements submitted to USAC for the eRate program only.

Physical topographic, land cover, and meteorological data available from the International Satellite Land-Surface Climatology Project, Initiative II, for half degree squares covering the earth's surface was downloaded and matched to this same set of zip codes.[12] Every zip code was assigned values for the half degree cell (roughly 50 km by 50km) in which the latitude and longitude coordinates of its centroid were located. Data collected included mean and standard deviation of slope (maximum change in elevations between every 1km square cell and its eight neighbors within the half degree square area), mean and standard deviation of the Compound Topographic Index (often referred to as the "wetness" index), range between minimum and maximum elevations with the half degree cell. For even more finely spatially disaggregated quarter-degree-by-quarter-degree cells, the distribution of MODIS[13] land cover types was available and used.

In addition, certain additional data thought to be relevant to telephone network cost are available within the databases distributed with the FCC's hybrid cost proxy model:[14] rock hardness, water table level, and minimum and maximum slopes. These data are available, however, at the 1990 census block group level. Using the Missouri Census Data Center's MABLE/Geocorr geographic correspondence engine,[15] these census block group level data were aggregated up to the zip code level using 1990 population weights. Available zip code-level data constructed from this source were weighted mean water table level, population-weighted share of census block groups within a zip code with "normal," soft, and hard rock types, and maximum, minimum, and population-weighted means of maximum and minimum slopes within the census block groups making up a zip code.

Caveats

Before scrutinizing this data, I must note some limitations that come with it. First, by identifying an FCC-defined "high speed line" as "broadband" I am ratifying a definition that glosses over some very real differences among "high speed" lines. Cable broadband connections in the U.S. routinely exceed one-way download speeds of 2-3 megabits per second in many areas, a full order of magnitude greater than the FCC threshold for a high-speed line. Differences in download speeds within the "high-speed" line category are likely to be as great as or greater than magnitudes of differences in download speed between high-speed and low-speed lines established by this definition of broadband!

Second, actual *provision* of a high speed line (what is shown in the FCC data) is different from *availability* of a high speed line. In most instances, it may in fact be I speculate that availability in a zip code-sized area may reasonably be expected to lead to at least one person in that area purchasing the service, if "availability" is also taken to mean some minimal effort within a geographic area to sell the product. But we cannot exclude *a*

---

[12] The data may be found at http://islscp2.sesda.com/ISLSCP2_1/html_pages/data_scale.html.

[13] For Moderate Resolution Imaging Spectroradiometer (MODIS) satellite, launched in 1999.

[14] See http://www.fcc.gov/wcb/tapd/hcpm/.

[15] See http://mcdc2.missouri.edu/websas/geocorr90.shtml.

priori the possibility that more providers are offering the product, and are simply failing in competing for customers.

Operationally, this issue is probably most important in the market for satellite-based broadband services. Satellite-based service is available throughout the United States, in the sense that it is technically possible to put a satellite receiver virtually anywhere in the 50 United States and connect to a satellite-based service provider providing downloads at a speed exceeding 200 kbps. It is, however, prohibitively expensive compared with broadband services delivered through a terrestrial provider, when the latter is available. In addition, a satellite-based service typically requires on-the-ground service and support. Satellite-based service providers are cognizant of this when they attempt to market and sell their services. If we take "availability" to mean investment in a sales and support effort in a specific geographic region, it would seem unlikely that the overlap between provision and technical availability is vastly different than with other modes of service delivery. In any event, satellite and wireless-based broadband remains a tiny segment of the market overall (though potentially important for isolated rural areas), with 549 thousand high-speed lines out of a national total of 37.9 million in December 2004.[16]

Third, zip codes are relatively large chunks of geography. Just because one provider offers service to one customer in one portion of a zip code does not mean that the service is available uniformly throughout the zip code. For this reason, it is reasonable to suppose that our count of high speed line providers within a zip code may err on the overly generous side from the perspective of the totality of residents within that zip code. Nonetheless, without a detailed census of service availability at an even lower level of geographic detail, there is no practical alternative to using a definition like this in assessing broadband competition.

Finally, it is important to remember that we may be missing a significant "quality of service" issue when we frame the broadband discussion using the FCC definitions. It may well be that this "low quality" definition of broadband (i.e., >200 kbps) is glossing over a new issue. Even if there is relatively wide availability of low grade broadband, there may be substantially greater unevenness in access to high quality, high data rate services that could come to define a new "digital divide". This may be even truer for advanced broadband services that will define new levels of functionality in the near future.

The FCC data seem to indicate that, on the one hand, availability of some (at least "low") level of broadband services seems to be a rapidly diminishing issue for most of the U.S. population. On the other hand, these same data seem to suggest that geographic variance in the degree of competition (as measured by number of service providers in zip codes) has greatly increased. Increasingly, the degree of competition (and presumably, impacts on pricing), and not any availability, may be the real issue in broadband services.

---

[16] See FCC, 2005, op. cit., Tables 1and 2, p. 6. Less than a fifth of the satellite and wireless lines had a high-speed return, compared with 76 percent of high-speed lines overall. The share of satellite and wireless has been declining—from 1.8 percent of high speed lines in December,1999, to 1.5 percent in December, 2004.

Modeling Availability of Broadband Services

Our main present interest is in trying to understand why broadband service is (or is not) being provided in different kinds of zip codes. This is clearly the result of a series of economic decisions, and I next outline a simple and parsimonious economic framework for modeling these decisions, that makes use of available and relatively sparse data.

In constructing my model, I keep in mind a long-run story about how firms enter the high-speed Internet services business. In most markets, there are incumbent cable and local telephone service providers who can use their existing cable and wireline networks to deliver broadband services at a lower cost than *de novo* network builders. Historically, in most markets, third party broadband service providers could either compel their interconnection through regulatory procedures, or had reached voluntary agreements with local cable and telephone monopolies, which permitted them to connect to the incumbents' networks and offer high speed services over these networks after paying a suitable price. (It is suggestive that recent declines in numbers of zip codes with the most "competition"—10 or more providers—in many states coincide with easing of regulatory requirements that forced network owners to make their facilities available to third party service providers.) There are also growing numbers of "wi-fi"-type wireless service providers available in some U.S. markets, and much more expensive satellite-based services are theoretically available in virtually every part of the U.S.

Given economic conditions in every local market, we can conceive of an order of potential returns to providing broadband services. Let us order the potential entrants into a given market by their potential economic gains from entering the broadband service market, with index number 1 assigned to the player that receives the highest return from entering the market, number 2 assigned to the next most profitable entrant, etc. The order of different classes of providers on that list, by technology, will vary with supply-side cost factors, and demand-side consumer socioeconomic demographics, from market to market.

One way to think of this is as a line of M potential entrants to the broadband market in every zip code, with the type of company and technology with the highest potential profits holding number 1, and the lowest profit potential entrant holding number M.

Entry vs. No Entry. Will any firm at all enter the market? This a relatively easy question, in theory, given these assumptions, and the question this paper focuses on. Firm number 1, with the top spot in the profit pecking order, should consider what would happen if it entered the market as the sole provider of broadband services. If it couldn't make money as the local broadband monopolist, then no one else further down the line is going to be able to make money either. If on the other hand it can make money as a monopoly provider, it should go ahead and enter.

Thus, if there is any profit to be made by the most profitable potential broadband monopolist, at least one firm should enter the market. If $\prod^*$ is the maximum monopoly profit to be made by the potential entrant with the most to gain, the rule for any entry at all to come about is that if $\prod^*$ exceeds zero, some provider should enter the marketplace in the long run. Conversely, if ∏* is negative, no one will enter and there will be no providers of broadband services.

Conceptually, ∏* can be thought of as a "reduced form", where profit-maximizing price and quantity for a broadband monopolist have been solved for, and these values then inserted into the expression for monopoly profit. ∏* will be a function of variables that shift costs, and variables that shift demand. This is very convenient, since some of the variables we will be considering might conceivably shift either demand or cost, and this means that we do not have to worry unduly about identification or simultaneity issues. The down side is that when we observe the net impact of some given factor on entry into a market, we don't know whether that is working through the demand side, or the cost side, or both.

This framework is by nature long-term, since it relies on firms entering or exiting markets in accordance with their long-run profits. At any given moment in time, we can think of a large number of observations over individual regional markets as being "perturbed" by random factors from their long-run equilibria. In addition, in an industry subject to rapid technological change, like broadband, it is reasonable to suppose that the equilibrium number of providers for a market will change over time as technological change alters costs. In essence, we will be assuming that across regions (zip codes), entry (or lack thereof) reflects some deterministic calculation of profit given a static snapshot of costs at some time, plus disturbances that are distributed randomly across regions.

The natural structure for analyzing this problem is that of a logit or probit-type model. That is, there is an underlying "latent" variable," the hypothetical profit of the most potentially profitable firm were it a monopolist, ∏*, which we do not observe, but whose value determines a binary "entry" variable E which takes on value 0 if $\prod^* < 0$, value 1 if $\prod^* \geq 0$. ∏* is, however, a function of a vector of cost shifters Z, and demand shifters X, which we do observe. Then, our model is given by

(1) $\prod^* = X\,b + Z\,c + \varepsilon$, where ε is a random disturbance term;[17]

(2) and $E=1$ if $\prod^* \geq 0$,
$E=0$ if $\prod^* < 0$.

Given observed data on X, Z, and the entry decisions of firms, we can estimate the function $X\,b + Z\,c$ and use our coefficient estimates to evaluate the impact of changes in the X and Z variables on the probability that a firm will enter into a market. If we assume ε follows a logistic distribution, we have the logit model; if ε is distributed normally, we have the probit model. The logistic and normal distributions are very similar, and in practice, logit and probit models typically yield very similar results. Coefficients in logit

---

[17] The error term in (1) may also be thought of as containing within it a zip code-specific, mean zero "random effect" component reflecting the dispersion of unmeasured, time-constant factors across zip codes.

models are easier and more intuitive to interpret, however, and I will focus on presenting the logit results.[18]

Data Issues

I have constructed a unique database that joins together seven different data sources describing market-related cost and demand variables at the individual zip code level. The components, some previously described, of this database include:

1. FCC data on the number of firms providing at least one high-speed line to a geographic region, at the zip code level, discussed above. Recall that because of aggregation related to confidentiality concerns, data for zip codes with 1 to 3 providers have been aggregated together in the public data set.[19]
2. Detailed data for individual ZCTAs, also discussed earlier, from the 2000 U.S. Population and Housing Census. Detailed population and housing characteristics, including education, race and ethnicity, labor force status, industry and type of employment, income, housing characteristics, etc., are aggregated and available at the ZCTA level in the Census SF3 data set.[20] A short summary of these data are also available as a downloadable "2000 U.S. Gazetteer" file.[21] The 2000 Census data have been made available in a very user-friendly standard zip code level extract by the Missouri Census Data Center, and we have used their cleaned-up and processed version of this data.[22]
3. Data on numbers of establishments in ZCTAs at the two digit NAICs industry level, from the 1997 U.S. Economic Census.[23]

---

[18] This framework can be extended to consider how many firms are likely to enter any given market for broadband services if appropriate additional assumptions about the nature of oligopolistic competition in regional markets for broadband services are made. See Kenneth Flamm, "The Determinants of Broadband Competition: Economics, Demographics and State Policy," presented at the Telecommunications Policy Research Conference, Arlington, Virginia, October 2, 2004, and the earlier version of this paper presented at the PURC/London Business School Conference on "The Future of Broadband: Wired and Wireless, 2005" Gainesville, Florida, February 24, 2005.

[19] FCC data are also available on numbers of CLECs providing competition for incumbent telephone companies, by zip code. Competition in local voice telephone services might be expected to affect the costs of providing DSL-based high speed data services, and therefore the costs of broadband provision. In the long-run, many, if not all, of the same demand and supply shifters affecting broadband service provider numbers might be expected to affect local telephone competition.. Furthermore, state policy may also be an important factor in determining the extent of local telephone competition. We do not pursue the topic of estimating a "partially" reduced form in which local telephone competition is used as an explanatory variable for broadband deployment within a zip code (which would have the effect of removing indirect impacts of state policy, via telephone competition, from the state policy effects on broadband that we estimate) in this paper. This approach would bring with it some additional problems of endogeneity and identification which are discussed in Flamm (2004) and Flamm (2005), cited in the last footnote. We proceed for the remainder of this paper by using the standard "completely" reduced form.

[20] See http://www.census.gov/support/SF3ASCII.html for links to extensive documentation on this data set.

[21] See http://www.census.gov/geo/www/gazetteer/places2k.html. This is helpful for an overview of the structure of the ZCTAs and zip code-related issues that are addressed below.

[22] Available at http://mcdc2.missouri.edu/cgi-bin/broker?_PROGRAM=websas.uex2dex.sas&_SERVICE=appdev9&path=/pub/data/sf32000x&dset=uszips&view=0.

[23] These data may be found at http://www.census.gov/epcd/ec97zip/downlzip.htm.

4. Data on zip codes in use in November 1999 (a data file published by the Census),[24] 2000 (from the Population and Housing Census) distinguishing between point and geographic zips, and an electronic listing of current census and FIPs codes purchased from zipwise.com in February 2004.[25]
5. Data on commitments of funds by the Universal Service Fund to funded grants to schools and libraries to support communications and Internet connections, and to rural health care service providers, for the years 1999-2004, also discussed above. Fund "commitments" are the stage prior to disbursement, so these data represent likely spending on Internet connections for schools and libraries in the several years after their commitment. Data were also available on authorized disbursements for the schools and libraries ("eRate") program, but not for the rural health care program.[26]
6. The ISLSCPI II hydrological, topographic, and land cover data discussed earlier.
7. Topographic data from the FCC Hybrid Cost Model, also described above, aggregated up from the census block group level to the zip code level.

An extensive effort went into "cleaning" these data and making them consistent across sources. All ZCTAs/zip codes where the Census showed no population living were dropped (typically, most of these cases were zip codes that spanned more than a single state). In addition, I dropped all ZCTAs not linked to "geographic" zip codes included by the FCC in its universe (discussed above), as well as zip codes listed in the 1997 economic census that show businesses with addresses in multiple states (even when the population census showed the corresponding ZCTA as spanning only a single state). I dropped Puerto Rico and other US territories from the sample (establishment data from the economic census was unavailable), and purged all ZCTAs where any of the economic, social, and demographic variables I wished to use were missing. The resulting data set was combined with available geophysical data for 1990 zip codes derived from the FCC Hybrid Cost model. In the resulting "cleaned" data for zip codes containing valid observations for all variables, eight states (Alaska, Connecticut, District of Colombia, Delaware, Massachusetts, Nevada, New Jersey, and Rhode Island) had no included zip codes with zero providers in one of the years 2001-2003. This would mean that dummy variables for state effects in these areas would not be identified, and would lead to "quasi-complete separation" of the data (inability to compute a maximum likelihood estimator for an intercept term) were they to be included in the sample. Zip codes in these states were then dropped, resulting in a sample of 25,216 zip codes spread across 43 states.[27]

---

[24] See http://www.census.gov/geo/www/tiger/zip1999.html.

[25] See http://www.zipwise.com.

[26] Some early public use data file may also be found at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/neca.html. A description of the program may be found at http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/Monitor/mr03-4.pdf.

[27] The original sample of 32,081 zip codes listed in the 2000 population census ZCTA data (32,038 after removing duplicates of 42 multi-state zip codes listed for more than one state). After merging this with 29,915 "geographic" zips compiled by the census in November 1999, and some preliminary cleaning, we had 27,774 observations. After further cleaning the data, removing zip codes located in non-state US territories (like Puerto Rico and Saipan), merging with the FCC Hybrid Cost Proxy Model data based on

Estimating a Model of Entry

Our initial effort was to estimate the model described by equation (2) above, using both logit and probit assumptions about the error distribution term. We assume a linear approximation to the profit function described by (1), and estimate an equation of the form

(3) Prob(E=1) = F(X b +Z c), derived from (1) and (2) above,

where F is assumed to be the cumulative density function for either the logistic (logit) or normal (probit) distribution, depending on the assumption about the error term in (1). Probit results differed minimally from logit results, and we henceforth discuss only the logit equation, where coefficients have more intuitive interpretations.

The published empirical econometric literature on the subject of what variables are important in determining either broadband supply or costs is relatively small.[28] The FCC "high-speed" line reports, referenced above, typically provide simple tables showing that greater broadband penetration in zip codes seems correlated positively with both per capita income and population density.

The classes of variables I first included in my analysis were (C / D notation indicates whether they likely affect costs or demand):

--Population density, measures of the percent of the population in rural areas, percent living on farms (C or D)
--Geographic location (latitude and longitude) (C or D) [a preliminary analysis suggested that both might be significant; I also constructed a "heartland" variable measuring absolute distance in degrees from latitude -95]

1990 zip codes, and dropping our 8 states with no "cleaned" zip codes without broadband available, we wound up with a sample with 25,216 zip codes in it.

[28] Earlier studies of this subject include T. Grubesic, "The geodemographic correlates of broadband access and availability in the United States," **Telematics and Informatics,** 21, 2004, pp. 335-358; J. Prieger, "The Supply Side of the Digital Divide: Is There Equal Availability in the Broadband Internet Access Market?" **Economic Inquiry**, vol. 41, no. 2, 2003, pp. 346-363; D. Gabel & F. Kwan, 2001, "Accessibility of Broadband Telecommunication Services by Various Segments of the American Population," in B. Compaine and S. Greenstein, eds., **Communications Policy in Transition: The Internet and Beyond**, MIT Press, 2001, pp. 295-320; S. Gillett & W. Lehr, "Availability of Broadband Internet Access: Empirical Evidence," Presented at Telecommunications Policy Research Conference, September 25-27, 1999, Alexandria VA, http://itc.mit.edu/itel/docs/MISC/LehrGillettTPRC99_0523.doc; D. Gabel, and G.L. Huang, "Promoting Innovation: Impact of Local Competition and Regulation on Deployment of Advanced Telecommunications Services for Businesses," 2003, http://itc.mit.edu/itel/docs/2003/promo_innov.pdf; and J.A. Hausman, J.G. Sidak, and H.J. Singer, "Cable Modems and DSL: Broadband Internet Access for Residential Customers," **American Economic Review**, vol. 19, May 2001. The Prieger study is most similar to the current paper, but uses 1990 Census data, early (unrevised) data from the FCC, and a sparser set of explanatory variables to estimate a probit equation describing broadband entry. The one econometric study of broadband price I have seen (Hausman, Sidak, and Singer) uses a very small sample of prices and basically finds that only a dummy for Roadrunner (a quality indicator?) is statistically significant. No included household income and age variables, dialup access price, or population density carries either a large or statistically significant coefficient. Note that price drops out of the reduced form I am estimating.

--Establishment counts for two-digit NAICs industries (D)
--Dummy variables to account for state policies and programs that might affect either broadband cost or demand (Texas normalized as baseline) (C or D)
--Percent of the population in very detailed age groups, with a disability, living in households with married couples (D)
--Racial composition of population (percent of population single race Black, Indian, Asian, Hawaiian, or Other, multi-race, single race white as baseline) (D)
--Percent of population of Hispanic ethnicity
--Percent of population in detailed educational status categories (D)
--English-speaking abilities of population (D)
--Average commute time to work, in minutes (D)
--Percent of population with Disabled status (D)
--Participation in labor force or armed forces, employment status (D)
--Broad categories of industry of employment, profession (D)
--Average household and family incomes, per capita income (D, possibly C)
--Percent of population poor, female, living in group quarters, institutionalized, living in married couple households (D)
--Occupied housing density, percent houses occupied, percent in crowded housing (D or C)
--Percent of homes with no telephone, no car, no indoor plumbing, percent of housing structures by number of units in structure, percent in mobile homes, percent of housing units in 2-4, 5-19, 20 and up housing unit configurations (D or C)
--Average rent and home value (D)
--Cumulative "eRate" and rural health care grant value committed to a zip code by the Universal Service Fund for years 1999-2004, (C or D)
--Geophysical and hydrological data-- mean and standard deviation of slope, range of elevation, and composite topographic ("wetness") index for ½ degree square areas containing the latitude and longitude coordinates of the centroid of a Census ZCTA; MODIS land cover classification (type of vegetation/physical land cover) for quarter degree map grid squares containing zip code centroid; FCC data on rock hardness, slope, water table level aggregated to 1990 zip code from 1990 census block groups. (C)

Specification Issues

Before turning to actual empirical results, two further issues related to the specification of the empirical model need be discussed. The first of these is my use of geophysical and hydrological variables, the second is my assumptions about functional form.[29]

Geophysical and hydrological data. A very preliminary version of this paper, working with data for December 2000, based on casual observation of geographic patterns of Internet use in earlier research on broadband use, experimented with use of nonlinear functions of longitude as explanatory variables in estimating the reduced form described above. Much to the author's chagrin, these variables were statistically significant, no matter how many additional variables were included in the equation's specification.

---

[29] A possible third issue, not included here, is a discussion of possible use of data measuring the extent of local telephone competition to estimate a "partially reduced form" equation. See Flamm (2004, 2005) for further discussion of this issue.

Observing that the changing pattern of effect on broadband availability (highest in the "heartland" at the center of the country, lower as one moved east or west) roughly coincided with mountain ranges (the Rockies and the Alleghenies), terrain effects immediately came to mind as a possible explanation. This was my original motivation for exploring what if any impact terrain variables might have on broadband economics.

Interestingly, the original longitude-based terrain proxies did not have the same strikingly significant effects in preliminary regressions for December of 2001, 2002, and 2003. The FCC's original high speed data for 1999 and 2000 had significant quality problems, and the data was later revised and reissued by the FCC after following up with selected respondents. Taking this as evidence that the data for 1999 and 2000 is likely to be of lower quality, with greater noise, than in following years, I have focused my analysis on the years after 2000.

Nonetheless, the original stimulus to experiment with geophysical determinants of broadband availability (working through costs, most likely) turned out to be a highly productive impulse. As we shall see, geophysical determinants are statistically significant, and their omission appears to create bias in estimates for at least some of the other factors affecting broadband deployment.

Functional Form. A very preliminary version of this paper experimented with a variety of functional forms for continuous variables (linear, logarithmic, square roots). Generally, natural logs produced marginally better results (measured by log likelihood or Akaike Information Criterion), and in some cases we rejected a linear specification in favor of logs when both were nested within a common specification. A limited amount of experimentation with logs and linear forms for continuous variables was undertaken in this paper, with similar results. In all cases the log form yielded superior fits, and there was virtually no impact on the signs and relative magnitudes of effects.[30]

I report results using a logarithmic functional form for continuous variables below. With a logit model, the coefficient of the log of an independent variable can also be interpreted as the elasticity of the odds ratio with respect with respect to the independent variable, which is quite convenient in interpreting coefficients.[31]

Initial Estimates

I initially estimated a full (150 variables) version of the binary logit model based on equation (4) below (all variables, logarithmic functional form for continuous variables), estimated separately for December 2001, December 2002, and December 2003. The probability of any broadband provision at all was modeled.

---

[30] For continuous geophysical data from the ISLSCPI II database (slope, elevation, CTI index, etc.) a small number of zero values were replaced with an artificial value of .001 prior to taking logs)

[31] That is, when log(odds ratio) = a + b log(x), b equals *d*log(odds ratio)/*d*log(x), which is the elasticity of log odds ratio with respect to x.

State dummy variables were denoted as Sx, where x is the numeric Federal Information Processing Standard (FIPS) code for that state. Texas (S48, FIPS=48) is the excluded state dummy incorporated into the intercept term. Race variables take “white” as the omitted exhaustive category, education takes “less than a high school education” as the baseline, occupation uses “production, transportation, and material moving” as the omitted type, industry takes “other industries” as the base, rock hardness takes “normal” rock hardness as the norm, and region takes “urban” as the excluded type. MODIS land cover variable Mx takes gives the percentage of land cover type x in the quarter degree square in which a zip code’s centroid is located. (M0, for example, denotes MODIS land cover type 0—water bodies, including oceans, seas, lakes, reservoirs, and rivers; there are 18 MODIS land cover classifications) Urban land cover (M13) is the omitted baseline. Note that the “eRate” and rural health care grant variable used in these models is cumulative commitments for the *prior* grant years, i.e., through grant year 2000 in December 2001, through 2001 in December 2002, etc.

The set of preliminary logit equations have considerable explanatory power: a generalized R-square measure (max re-scaled R-square) is .56 in 2001, .48 in 2002, and .43 in 2003.[32] In a model with just an intercept and state dummy variables, the equivalent generalized R-square measures are .18, .14, and .12, respectively.

A total of 150 explanatory variables (including the intercept) were used in these initial logit estimates.[33] We can, however, take advantage of the fact that we have repeated observations over time on individual zip codes to estimate coefficients more precisely, and to allow for possible neglected heterogeneity across zip codes, and correlation over time. My approach is to estimate a so-called marginal model, i.e., one that estimates the mean broadband availability over all zip codes conditional on some given set of independent variables. (Heuristically, we can think of an effect estimated this way as measuring the hypothetical impact on broadband availability averaged across all zip codes were they to have some given set of measured characteristics, relative to mean broadband availability in all zip codes were they to have the same set of other measured characteristics without the effect.)

The mean probability of high speed availability in zip code j, hj, conditional on observed vector of covariates Xj in that zip code, is assumed to be given by the logit function

$$(4)\ \log (hj/(1 - hj)) = Xj'\beta,$$

assuming the variance of broadband availability around conditional mean hj is some known function of this mean, and where the within-zip-codes association of broadband

---

[32] Percent concordant observed responses was 91.9 in 2001, 91.9 in 2002, and 92.5 in 2003.

[33] This same variable set was used to estimate an ordered logit model describing the number of competitors selling broadband services in a zip code. As noted earlier, substantially more demanding theoretical assumptions must be made to justify this model; furthermore, the FCC data are probably a substantially more noisy measure of the unobserved left-hand-side outcome. In all cases, the score test for the proportional odds assumption (which must be maintained in an ordered logit model) leads us to reject that assumption. To model numbers of competitors properly, I conclude, we must move to a more complex model—partial proportional odds, or continuation ratio models are two attractive alternatives.

availability over time is assumed to depend on some fixed set of association parameters and the mean. Equation (4) also follows from equations (1) and (2) above; in that case, however, precise distributional assumptions about an error term were made, and maximum likelihood methods used to produce an estimator whose properties hinged on those statistical assumptions.

Unlike the standard logit model defined by equations (1) and (2) above, this "population-averaged" marginal model avoids precise distributional assumptions about the broadband availability variable, and therefore enjoys a certain degree of robustness with respect to possible heterogeneity over zip code and correlation over time in error terms.[34] I employ "pooled logit," a variant of the method of generalized estimating equations (GEE) that makes the fewest assumptions about the distribution of hj in (4), in order to estimate the parameters of (4).[35] This estimator will be consistent, and a robust "sandwich" covariance estimator can be constructed that will provide asymptotically correct standard errors. I trade off efficiency for robustness, and the ability to make use of within-zip-code information to better model my longitudinal panel data. GEE estimators typically provide reasonably efficient estimates, even when association structures over time used to construct estimators are not correct.

Zip code data for December, 2001 through December, 2003 were combined in a single data set, and pooled logit used to estimate (4) using this three year panel. There is of course a real possibility that coefficients may change over time—as do technology and policy in this arena—so my first step was to specify a totally general model that permitted all coefficients to change from one period to the next. With time period-specific interaction effects for all coefficients (including shifts in the intercept over time), there were 450 different effects estimated using 75,648 observations (3 years x 25,216 observations).

---

[34] Another alternative would be to estimate a so-called "random effects" logit model, where additional assumptions are added to the statistical model: an additive random intercept effect from a known common distribution is drawn independently and added on to the equation describing broadband availability in each zip code, and strict exogeneity of covariates conditional on these unobserved effects is assumed. This approach estimates the impact of a variable conditional on the realized random effect specific to an individual zip code. It is well known that, adopting the statistical assumptions of the random effects model, the population-averaged effect is smaller in absolute value (attenuated) compared to the zip-specific effect conditional on being in any specific zip code. With the random effects model, as the variance of the random effect across zip codes gets smaller and approaches zero, this difference between the population-averaged effect and an individual effect will also shrink toward zero. See Jeffrey M. Woolridge, **Econometric Analysis of Cross Section and Panel Data**, (Cambridge, MIT Press), 2002, pp. 470-472, 484-485.

[35] I assume a diagonal (independent), within-zip working correlation structure over time, and therefore sacrifice some potential efficiency for simplicity and robustness in producing my GEE estimates, which nonetheless are consistent and have correct standard errors. The GEE estimator with an independence error correlation structure is also called the "pooled" logit estimator and makes minimal assumptions about exogeneity of covariates and dynamics (lagged values of dependent and independent variables are permitted as conditioning variables) in order to produce consistent and asymptotically normal parameter estimates of marginal population parameters. See Woolridge, 2002, pp. 401-405, 482-483. In addition, convergence can be an issue in GEE models which assume a more complex working correlation structure, when there are large numbers of variables and large data sets, as was the case.

A large number of the effects in this initial analysis seemed small and close to zero, or seemed to change little over time. My first step was to assemble a group of attractive candidate variables or changes over time that seemed small in value and statistically not significant, and then test whether we would be justified in dropping these variables or time interactions in further analyses. My initial criteria were a small value for the coefficient and an empirical standard error that did not compel us to reject the hypothesis that the coefficient was zero at even relatively loose standards for statistical significance (the 10% level).

One broad group of restrictions involved certain variables that seemed to have no effect in all years. Variables in this group included certain geophysical variables[36] and a number of census-derived variables.[37] Then there was another group of variables that seem constant over time. These include state level effects for Arizona, Arkansas, California, Florida, Georgia, Indiana, Kentucky, Louisiana, Minnesota, Mississippi, Nebraska, New Mexico, New York, North Carolina, Ohio, Oklahoma, Oregon, Pennsylvania, South Carolina, Utah, Virginia, Wisconsin, and Wyoming. They also included certain geophysical variables[38] and census variables.[39] The individual coefficients for all these variables were not significant, nor did groups and subsets of these variables, tested jointly, lead us to reject the zero null hypothesis using standard Wald tests.

A total of 232 coefficients for this set of variables and changes over time could thus be eliminated by restricting their value to zero. A joint Wald test on all these constraints produced a chi-squared value of 257.7 with 232 degrees of freedom, which does not lead us to reject the zero null at the 10 percent significance level.[40] A score (Lagrange multiplier) test for the same hypotheses produces a chi-squared statistic of 240.7, again

---

[36] Longitude, distance from a central US longitude, population-weighted mean water table depth, percent of rock in a zip code that is a "soft" type, weighted mean of maximum slope, minimum slope within a zip code's census block groups, maximum slope within the zip code, mean CTI within the half degree square containing the zip code centroid.

[37] Percent Hispanic ethnicity, percent of population unemployed, percent of population in various detailed age categories except for 20 to 24, and 85 and over, percent of population of a multiracial background, percent of population of a single racial background other than black, American Indian, or Hawaiian Islander Polynesian, percent speaking no English, percent in the armed forces, percent in the civilian labor force, percent not in the labor force, percent employed in retail trade, percent in service occupations, percent in sales and office occupations, percent in construction occupations, average commute time, percent poor, occupied housing per land area, percent homes lacking a phone, percent living in crowded housing.

[38] Standard deviation of slope, range from minimum to maximum elevation within the half degree square containing the centroid of a zip code, population-weighted mean of minimum slope within a zip code, MODIS land cover for types 0, 1, 3, 5, 6, 7, 8, 10, 11, 12, 14, 16, 17—water bodies, evergreen needle leaf forests, deciduous needle leaf forests, mixed forests, closed shrub lands, open shrub lands, woody savannas, grasslands, permanent wetlands, croplands, cropland/natural vegetation mosaic, barren, and unclassified, respectively.

[39] Percent age 85 or over, percent with a disability, percent with limited English, percent with some high school, percent with some college, percent with bachelor's degree, percent with graduate or professional degree, percent working in manufacturing, percent female, percent living in institutions, percent living in group quarters, percent in housing without complete plumbing, percent working in education, percent living in housing with 5 to 19 units, percent living in mobile homes, percent living on farms, per capita income, log of average home value.

[40] The p-value is .12.

with 232 degrees of freedom, which again does not lead us to reject the null hypothesis that the zero constraints hold true at the 10 percent significance level.[41] The two tests are asymptotically equivalent, but as is typical, the Wald test seems the more conservative one in considering whether to reject the null in this finite sample.

Imposing these constraints, we can now estimate a more parsimonious model with a mere 218 coefficients, which we report in Table 3 below.

Table 3: Parsimonious Model, Broadband Availability 2001-03

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Note: Statistically significant coefficients at 5% level have been bolded.
Coefficient names with suffix 01 and 02 refer to incremental changes for 2003 values in 2001 or 2002 for most variables (exceptions are rock hardness, HARD03, and racial categories, e.g., Black03, where year suffixes indicate entire effect for that year.

Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	2.9174	3.649	-4.2345	10.0694	0.8	0.424
t01	-0.402	1.8406	-4.0094	3.2055	-0.22	0.8271
t02	0.2848	1.5717	-2.7958	3.3653	0.18	0.8562
USF Funding:						
erate03	0	0.0001	-0.0002	0.0003	0.28	0.783
erate01	0.0004	0.0004	-0.0003	0.0012	1.13	0.2579
erate02	**0.0008**	**0.0004**	**0.0001**	**0.0015**	**2.14**	**0.0325**
rhc	-0.0056	0.006	-0.0174	0.0061	-0.94	0.3487
Industry effects:						
e31	-0.016	0.0203	-0.0559	0.0238	-0.79	0.4296
e3101	**0.0728**	**0.0171**	**0.0393**	**0.1063**	**4.26**	**<.0001**
e3102	**0.0546**	**0.0271**	**0.0014**	**0.1078**	**2.01**	**0.0442**
e44	**0.0624**	**0.0152**	**0.0325**	**0.0922**	**4.09**	**<.0001**
e4401	**-0.0545**	**0.0184**	**-0.0906**	**-0.0184**	**-2.96**	**0.0031**
e4402	-0.0253	0.0176	-0.0597	0.0091	-1.44	0.1488
e54	-0.003	0.0189	-0.0399	0.034	-0.16	0.8741
e5401	**0.0993**	**0.0306**	**0.0394**	**0.1593**	**3.25**	**0.0012**
e5402	0.0588	0.0512	-0.0414	0.1591	1.15	0.2501
e56	-0.0338	0.0872	-0.2048	0.1371	-0.39	0.6979
e5601	-0.0628	0.0774	-0.2146	0.0889	-0.81	0.4169
e5602	-0.0055	0.0482	-0.1	0.0889	-0.12	0.9084
e62	**-0.023**	**0.0043**	**-0.0313**	**-0.0146**	**-5.38**	**<.0001**
e6201	0.0242	0.0246	-0.0241	0.0724	0.98	0.3259
e6202	-0.0096	0.0056	-0.0205	0.0013	-1.72	0.085
e72	0.0506	0.0259	-0.0003	0.1014	1.95	0.0511
e7201	-0.0332	0.0278	-0.0876	0.0212	-1.2	0.2314
e7202	0.0305	0.0258	-0.0202	0.0811	1.18	0.2383
e81	-0.0189	0.0391	-0.0955	0.0577	-0.48	0.629
e8101	0.058	0.0401	-0.0207	0.1366	1.44	0.1487

---

[41] The p-value is .33 this time.

e8102	0.0261	0.0357	-0.0438	0.0961	0.73	0.4639
Geophysical:						
lelevrang	**-0.2066**	**0.069**	**-0.3419**	**-0.0713**	**-2.99**	**0.0028**
HARD01	**0.0038**	**0.0009**	**0.002**	**0.0056**	**4.1**	**<.0001**
HARD02	0.0009	0.0011	-0.0012	0.003	0.82	0.4126
HARD03	-0.0011	0.0013	-0.0037	0.0014	-0.88	0.3814
lslopesd	**0.1789**	**0.0641**	**0.0533**	**0.3045**	**2.79**	**0.0053**
MINSLOPE_WGTMEAN	0.0052	0.0061	-0.0068	0.0172	0.84	0.3987
modisqc00	**-0.0672**	**0.0275**	**-0.1212**	**-0.0132**	**-2.44**	**0.0147**
modisqc01	**-0.0701**	**0.0276**	**-0.1241**	**-0.016**	**-2.54**	**0.011**
modisqc02	**-0.0736**	**0.0375**	**-0.1471**	**-0.0002**	**-1.96**	**0.0495**
modisqc0201	-0.0096	0.0278	-0.0641	0.045	-0.34	0.7309
modisqc0202	0.03	0.0191	-0.0074	0.0674	1.57	0.1159
modisqc03	-0.0142	0.0676	-0.1466	0.1183	-0.21	0.834
modisqc04	**-0.0702**	**0.0276**	**-0.1242**	**-0.0162**	**-2.55**	**0.0109**
modisqc0401	0.0023	0.0017	-0.001	0.0056	1.36	0.1734
modisqc0402	0.0028	0.0015	-0.0002	0.0057	1.84	0.0652
modisqc05	**-0.0624**	**0.0276**	**-0.1166**	**-0.0083**	**-2.26**	**0.0239**
modisqc06	-0.0333	0.0684	-0.1673	0.1007	-0.49	0.6263
modisqc07	**-0.0665**	**0.0277**	**-0.1209**	**-0.0122**	**-2.4**	**0.0165**
modisqc08	**-0.0677**	**0.0277**	**-0.1219**	**-0.0134**	**-2.44**	**0.0145**
modisqc09	0.024	0.0614	-0.0964	0.1444	0.39	0.6964
modisqc0901	-0.0895	0.0555	-0.1982	0.0193	-1.61	0.107
modisqc0902	-0.0503	0.049	-0.1463	0.0458	-1.03	0.3048
modisqc10	**-0.0659**	**0.0276**	**-0.1199**	**-0.0118**	**-2.39**	**0.017**
modisqc11	**-0.1252**	**0.0583**	**-0.2395**	**-0.011**	**-2.15**	**0.0317**
modisqc12	**-0.0653**	**0.0276**	**-0.1195**	**-0.0112**	**-2.36**	**0.018**
modisqc14	**-0.0642**	**0.0276**	**-0.1183**	**-0.01**	**-2.32**	**0.0201**
modisqc15	0.0178	1.2852	-2.5012	2.5367	0.01	0.989
modisqc1501	0.44	0.6592	-0.852	1.7321	0.67	0.5045
modisqc1502	0.1551	0.6557	-1.1301	1.4403	0.24	0.8131
modisqc16	**-0.0922**	**0.031**	**-0.153**	**-0.0314**	**-2.97**	**0.0029**
modisqc17	-0.0401	0.0334	-0.1055	0.0253	-1.2	0.2292
Population Characteristics:						
lpopden	**1.221**	**0.0781**	**1.0678**	**1.3742**	**15.62**	**<.0001**
lpopden01	**-0.2084**	**0.0711**	**-0.3478**	**-0.069**	**-2.93**	**0.0034**
lpopden02	**-0.2707**	**0.057**	**-0.3824**	**-0.1589**	**-4.75**	**<.0001**
lland	**1.1753**	**0.07**	**1.0382**	**1.3125**	**16.8**	**<.0001**
lland01	**-0.1907**	**0.0626**	**-0.3135**	**-0.068**	**-3.05**	**0.0023**
lland02	**-0.2101**	**0.0511**	**-0.3103**	**-0.1099**	**-4.11**	**<.0001**
lPCI	**0.3338**	**0.1208**	**0.0971**	**0.5705**	**2.76**	**0.0057**
PctAge20_24	0.0197	0.0133	-0.0063	0.0458	1.48	0.1381
PctAge20_2401	**-0.0419**	**0.0149**	**-0.071**	**-0.0127**	**-2.81**	**0.0049**
PctAge20_2402	**-0.0328**	**0.0125**	**-0.0573**	**-0.0084**	**-2.63**	**0.0086**
PctDisabled	-0.0054	0.0039	-0.013	0.0021	-1.41	0.1583
PctEnglishLimited	**0.0188**	**0.007**	**0.0052**	**0.0325**	**2.71**	**0.0068**
PctFemale	**-0.0133**	**0.0064**	**-0.0259**	**-0.0007**	**-2.08**	**0.0379**
PctOver85	-0.0249	0.0151	-0.0545	0.0048	-1.64	0.1008
PctMarriedCouples	**-0.0127**	**0.0041**	**-0.0208**	**-0.0046**	**-3.07**	**0.0022**
PctMarriedCouples01	0.0086	0.0045	-0.0003	0.0175	1.89	0.0581
PctMarriedCouples02	0.0035	0.0037	-0.0039	0.0108	0.92	0.3551
PctOnFarms	-0.0057	0.0031	-0.0119	0.0004	-1.83	0.0678
PctRural	**-0.0099**	**0.0039**	**-0.0175**	**-0.0023**	**-2.55**	**0.0107**
PctRural01	**0.0076**	**0.0038**	**0.0001**	**0.015**	**1.98**	**0.0473**
PctRural02	0.0018	0.0032	-0.0046	0.0081	0.54	0.5864

Ethnicity:						
PctBlack101	**-0.0059**	**0.0025**	**-0.0109**	**-0.0009**	**-2.33**	**0.0197**
PctBlack102	-0.0011	0.0029	-0.0067	0.0046	-0.37	0.713
PctBlack103	-0.0034	0.0036	-0.0105	0.0037	-0.93	0.3537
PctHawnPI101	0.0226	0.0465	-0.0685	0.1137	0.49	0.6266
PctHawnPI102	**-0.1471**	**0.053**	**-0.2509**	**-0.0433**	**-2.78**	**0.0055**
PctHawnPI103	**-0.1543**	**0.0446**	**-0.2416**	**-0.0669**	**-3.46**	**0.0005**
PctIndian101	**-0.0133**	**0.0034**	**-0.02**	**-0.0065**	**-3.85**	**0.0001**
PctIndian102	**-0.0137**	**0.0038**	**-0.0212**	**-0.0061**	**-3.56**	**0.0004**
PctIndian103	**-0.014**	**0.0046**	**-0.023**	**-0.005**	**-3.05**	**0.0023**
Education:						
PctSomeHighSchool	0.0086	0.0062	-0.0036	0.0207	1.39	0.1658
PctHighSchool	**0.0144**	**0.0058**	**0.0031**	**0.0256**	**2.5**	**0.0126**
PctHighSchool01	0.0035	0.0046	-0.0055	0.0125	0.77	0.444
PctHighSchool02	0.0037	0.0037	-0.0036	0.0109	1	0.3188
PctSomeCollege	**0.0213**	**0.005**	**0.0114**	**0.0311**	**4.24**	**<.0001**
PctBachelors	**0.0223**	**0.007**	**0.0086**	**0.0359**	**3.2**	**0.0014**
PctGradProf	**0.0205**	**0.0095**	**0.0018**	**0.0392**	**2.15**	**0.0314**
Industry of Employment:						
PctEducation	**-0.0107**	**0.0048**	**-0.0201**	**-0.0013**	**-2.24**	**0.0253**
PctManufacturing	**-0.0055**	**0.0027**	**-0.0107**	**-0.0002**	**-2.05**	**0.04**
Occupation:						
PctFarmFishOccs	-0.0138	0.0078	-0.0291	0.0015	-1.76	0.0781
PctFarmFishOccs01	0.0152	0.0084	-0.0013	0.0317	1.81	0.0702
PctFarmFishOccs02	0.0004	0.0065	-0.0124	0.0132	0.06	0.9531
PctHealthSA	0.0017	0.0064	-0.0107	0.0142	0.27	0.7853
PctHealthSA01	-0.0132	0.007	-0.0269	0.0006	-1.87	0.0614
PctHealthSA02	-0.0072	0.006	-0.019	0.0047	-1.19	0.2354
PctManProfOccs	0.0035	0.0044	-0.0052	0.0121	0.78	0.4333
PctManProfOccs01	-0.0014	0.0043	-0.0099	0.0071	-0.32	0.7497
PctManProfOccs02	0.0003	0.0035	-0.0065	0.0071	0.08	0.9378
Housing:						
lavgage	0.1877	0.4286	-0.6522	1.0277	0.44	0.6613
lavgage01	**-0.9788**	**0.439**	**-1.8391**	**-0.1184**	**-2.23**	**0.0258**
lavgage02	**-0.7811**	**0.3786**	**-1.5232**	**-0.039**	**-2.06**	**0.0391**
lAvghval	**0.3739**	**0.0744**	**0.228**	**0.5198**	**5.02**	**<.0001**
lAvgrent	0.1656	0.1231	-0.0756	0.4069	1.35	0.1785
lavgrent01	0.0433	0.1391	-0.2293	0.3158	0.31	0.7555
lavgrent02	0.0286	0.1116	-0.1901	0.2474	0.26	0.7975
PctAgeUnit5	0.0134	0.0083	-0.0028	0.0296	1.62	0.106
PctAgeUnit501	-0.0136	0.0089	-0.0312	0.0039	-1.53	0.1269
PctAgeUnit502	-0.0032	0.0075	-0.0179	0.0115	-0.43	0.6697
PctBuiltBefore1940	-0.0018	0.0055	-0.0126	0.009	-0.33	0.7426
PctBuiltBefore194001	0.0071	0.0058	-0.0042	0.0184	1.23	0.2204
PctBuiltBefore194002	0.0079	0.005	-0.0018	0.0176	1.6	0.1106
PctGQPop	**0.0208**	**0.0088**	**0.0035**	**0.038**	**2.36**	**0.0181**
PctInInstitutions	**-0.0325**	**0.0097**	**-0.0516**	**-0.0135**	**-3.35**	**0.0008**
PctNoCars	**-0.0187**	**0.0076**	**-0.0336**	**-0.0037**	**-2.45**	**0.0143**
PctNoCars01	**0.0191**	**0.0086**	**0.0022**	**0.036**	**2.21**	**0.027**
PctNoCars02	**0.0167**	**0.0066**	**0.0038**	**0.0296**	**2.53**	**0.0114**
PctPlumbing	**0.0316**	**0.009**	**0.0139**	**0.0492**	**3.5**	**0.0005**
PctMobileHomes	**-0.0084**	**0.0027**	**-0.0136**	**-0.0031**	**-3.11**	**0.0019**

PctOccupied	**0.0108**	**0.0041**	**0.0027**	**0.0188**	**2.63**	**0.0085**
PctOccupied01	-0.0048	0.0036	-0.0119	0.0022	-1.35	0.1755
PctOccupied02	-0.0023	0.003	-0.0082	0.0036	-0.78	0.4355
PctUnits2	-0.0059	0.0136	-0.0326	0.0208	-0.44	0.6633
PctUnits201	**0.0341**	**0.0138**	**0.007**	**0.0612**	**2.47**	**0.0135**
PctUnits202	0.0136	0.0112	-0.0084	0.0356	1.21	0.2255
PctUnits20up	**-0.0392**	**0.0174**	**-0.0734**	**-0.0051**	**-2.25**	**0.0243**
PctUnits20up01	0.0336	0.0225	-0.0105	0.0776	1.49	0.135
PctUnits20Up02	**0.0939**	**0.0306**	**0.0339**	**0.1539**	**3.07**	**0.0022**
PctUnits5	0.0087	0.0097	-0.0104	0.0277	0.89	0.3745
pophous	**-0.2752**	**0.1203**	**-0.5111**	**-0.0394**	**-2.29**	**0.0222**
pophous01	**0.2195**	**0.0722**	**0.0779**	**0.361**	**3.04**	**0.0024**
pophous02	**0.1642**	**0.0525**	**0.0613**	**0.2671**	**3.13**	**0.0018**
State Effects						
S1	**-0.9735**	**0.2783**	**-1.519**	**-0.428**	**-3.5**	**0.0005**
s101	**0.8596**	**0.277**	**0.3168**	**1.4024**	**3.1**	**0.0019**
s102	**0.6674**	**0.2295**	**0.2175**	**1.1173**	**2.91**	**0.0036**
S12	0.4053	0.3292	-0.2399	1.0504	1.23	0.2182
S13	-0.1684	0.1774	-0.5161	0.1793	-0.95	0.3425
S15	-0.5796	1.4641	-3.4492	2.29	-0.4	0.6922
s1501	**-6.6529**	**1.679**	**-9.9437**	**-3.3621**	**-3.96**	**<.0001**
s1502	0.3065	1.3788	-2.3959	3.0089	0.22	0.8241
S16	-0.1685	0.3682	-0.8902	0.5532	-0.46	0.6472
s1601	-0.6185	0.3478	-1.3001	0.0631	-1.78	0.0753
s1602	-0.2087	0.3074	-0.8112	0.3937	-0.68	0.4971
S17	**-0.7676**	**0.2002**	**-1.16**	**-0.3752**	**-3.83**	**0.0001**
s1701	0.0581	0.1756	-0.2862	0.4023	0.33	0.741
s1702	**0.3378**	**0.1521**	**0.0397**	**0.636**	**2.22**	**0.0264**
S18	**-0.8435**	**0.1732**	**-1.1829**	**-0.504**	**-4.87**	**<.0001**
S19	**-1.528**	**0.1862**	**-1.8929**	**-1.1631**	**-8.21**	**<.0001**
s1901	-0.1705	0.1591	-0.4823	0.1413	-1.07	0.2838
s1902	**0.3838**	**0.1325**	**0.1241**	**0.6436**	**2.9**	**0.0038**
S20	**0.521**	**0.2368**	**0.0569**	**0.9852**	**2.2**	**0.0278**
s2001	**-1.4166**	**0.2255**	**-1.8585**	**-0.9747**	**-6.28**	**<.0001**
s2002	**-1.1239**	**0.2065**	**-1.5287**	**-0.7191**	**-5.44**	**<.0001**
S21	**-0.6237**	**0.1825**	**-0.9813**	**-0.266**	**-3.42**	**0.0006**
S22	0.1613	0.1908	-0.2126	0.5352	0.85	0.3979
S23	**-1.0815**	**0.2714**	**-1.6134**	**-0.5496**	**-3.99**	**<.0001**
s2301	**1.7757**	**0.269**	**1.2484**	**2.3029**	**6.6**	**<.0001**
s2302	**1.1598**	**0.2557**	**0.6587**	**1.6609**	**4.54**	**<.0001**
S24	-0.2249	0.4481	-1.1032	0.6534	-0.5	0.6157
s2401	-0.3729	0.4247	-1.2054	0.4596	-0.88	0.38
s2402	**1.1396**	**0.2914**	**0.5685**	**1.7108**	**3.91**	**<.0001**
S26	**-0.4413**	**0.1788**	**-0.7917**	**-0.091**	**-2.47**	**0.0136**
S27	**-1.6144**	**0.195**	**-1.9966**	**-1.2321**	**-8.28**	**<.0001**
s2701	0.28	0.172	-0.057	0.617	1.63	0.1035
s2702	0.2402	0.1526	-0.0588	0.5392	1.57	0.1154
S28	**-0.503**	**0.1943**	**-0.8839**	**-0.1222**	**-2.59**	**0.0096**
S29	**-0.9311**	**0.183**	**-1.2898**	**-0.5724**	**-5.09**	**<.0001**
s2901	**0.5862**	**0.1448**	**0.3025**	**0.8699**	**4.05**	**<.0001**
s2902	**0.5094**	**0.1245**	**0.2654**	**0.7534**	**4.09**	**<.0001**
S30	-0.0485	0.1984	-0.4373	0.3403	-0.24	0.8068
S31	**-0.7112**	**0.1578**	**-1.0205**	**-0.402**	**-4.51**	**<.0001**
S33	1.121	1.0265	-0.891	3.1329	1.09	0.2748
s3301	0.0082	0.9527	-1.859	1.8754	0.01	0.9931
s3302	-1.3039	0.9702	-3.2055	0.5977	-1.34	0.179

S35	-0.2371	0.2756	-0.7773	0.3032	-0.86	0.3897
S36	**0.7008**	**0.2095**	**0.2903**	**1.1114**	**3.35**	**0.0008**
S37	0.3211	0.2218	-0.1136	0.7558	1.45	0.1477
S38	0.1488	0.232	-0.3058	0.6035	0.64	0.5211
s3801	**-0.766**	**0.2396**	**-1.2355**	**-0.2964**	**-3.2**	**0.0014**
s3802	**-0.8978**	**0.2001**	**-1.29**	**-0.5055**	**-4.49**	**<.0001**
S39	-0.0082	0.199	-0.3981	0.3818	-0.04	0.9673
S4	0.2912	0.4757	-0.6411	1.2235	0.61	0.5404
S40	0.0383	0.1637	-0.2826	0.3591	0.23	0.815
S41	**0.6986**	**0.267**	**0.1752**	**1.2219**	**2.62**	**0.0089**
S42	**-1.0568**	**0.1955**	**-1.44**	**-0.6735**	**-5.4**	**<.0001**
S45	-0.3881	0.2373	-0.8533	0.077	-1.64	0.1019
S46	**-0.662**	**0.2144**	**-1.0822**	**-0.2418**	**-3.09**	**0.002**
s4601	-0.0978	0.206	-0.5016	0.306	-0.47	0.6349
s4602	0.2486	0.1665	-0.0777	0.5749	1.49	0.1353
S47	0.4516	0.3869	-0.3066	1.2099	1.17	0.243
s4701	**-1.0317**	**0.3857**	**-1.7877**	**-0.2756**	**-2.67**	**0.0075**
s4702	-0.0681	0.3759	-0.8048	0.6686	-0.18	0.8562
S49	-0.3654	0.4758	-1.298	0.5672	-0.77	0.4425
s4901	-0.6379	0.4571	-1.5339	0.2581	-1.4	0.1629
s4902	-0.4806	0.4148	-1.2937	0.3324	-1.16	0.2466
S5	-0.0741	0.1673	-0.4021	0.2538	-0.44	0.6578
S50	**0.994**	**0.5357**	**-0.056**	**2.044**	**1.86**	**0.0635**
s5001	**-1.3798**	**0.5108**	**-2.381**	**-0.3785**	**-2.7**	**0.0069**
s5002	**-1.0111**	**0.4706**	**-1.9335**	**-0.0887**	**-2.15**	**0.0317**
S51	**-0.8159**	**0.1916**	**-1.1915**	**-0.4403**	**-4.26**	**<.0001**
S53	**-0.7455**	**0.3137**	**-1.3604**	**-0.1306**	**-2.38**	**0.0175**
s5301	**0.968**	**0.279**	**0.4212**	**1.5148**	**3.47**	**0.0005**
s5302	**0.8868**	**0.2232**	**0.4494**	**1.3241**	**3.97**	**<.0001**
S54	-0.394	0.2521	-0.8881	0.1001	-1.56	0.1181
s5401	-0.1879	0.2096	-0.5986	0.2229	-0.9	0.37
s5402	0.3171	0.1697	-0.0154	0.6496	1.87	0.0616
S55	**-0.8159**	**0.1924**	**-1.193**	**-0.4388**	**-4.24**	**<.0001**
S56	0.4659	0.2995	-0.1211	1.053	1.56	0.1198
S6	**0.4917**	**0.2531**	**-0.0044**	**0.9877**	**1.94**	**0.052**
S8	**0.7187**	**0.2496**	**0.2296**	**1.2079**	**2.88**	**0.004**

General Observations

Note that for most variables, changes in coefficients over time (when not constrained to equal zero in this “parsimonious” model are parameterized as an incremental shift from a base coefficient value for December 2003. The exceptions to this rule are coefficients for race and ethnicity (black, American Indian, etc.) and share of rock classified as “hard,” which are specified as individual standalone effects for 2001 through 2003. I will confine most of my discussion of these results to coefficients for which we reject the zero null hypothesis at the 5% significance level. I will include marginally significant coefficients (i.e., reject equality with zero at 10% level but not at 5% level) on occasion when useful. While I will generally only identify time-specific coefficients for years in which they were significant, inability to reject the hypothesis that they are zero in other years may simply mean that estimated standard errors were larger, and not that the coefficient was not important in other years.

Although not necessarily apparent in Table 3, coefficient values and estimated standard errors were sensitive to omission of variables. That is, variations of the above equation estimated without a full set of regressors initially led to different “parsimonious” models and conclusions about what effects were important. Since this is a reduced form equation, one must conclude that using incomplete sets of regressors (including geophysical and housing stock variables that prove statistically significant in the results above, but seem to be rarely used in empirical studies of telecommunications markets) introduces a serious risk of biased coefficient estimates and flawed statistical inference in modeling both supply and demand.

Note that most zip codes actually had access to broadband in all years in my sample. Thus, in my working sample of 25,216 zip codes, 20,367 had broadband available in 2001, 22,603 in 2002, and 23,895 in 2003. As a consequence, even changes in variables that have significant effects on the odds of broadband availability in many zip codes may still end up having only slight impacts on the overall availability of broadband. Increases (or decreases) in the odds of broadband will have little aggregate impact if these odds are mainly very high already in most areas.

The Usual Suspects: Population Density

One truism of discussion of broadband supply is the assertion that population density, working through provisioning costs, is a major factor in explaining why broadband supply varies greatly across regions. This assertion, illustrated with a simple cross-tabulation of population density versus number of zip codes with differing degrees of apparent competition among active service providers is a staple of FCC reports on its Form 477 data.

At first glance, this seems to hold up in my empirical results. (See Table 4—time interaction effects have been added to base 2003 effects and empirical standard errors recalculated when necessary to produce estimates of total effect for a given year.) The log of population density is statistically significant at the 5% level in my estimated models, has the expected sign, and has a relatively large impact. We can interpret the coefficient of its log as an elasticity: a 10% increase in population density, cet. par., results in an 22% increase in the odds ratio for broadband on average in 2003. It seems to have been increasing over time.

Table 4 Population Density-related Effects

Variable	Estimate	Std. Error	95% Conf. Interval		Chi-Square
lpopden 2001	1.0126	0.0535	0.9078	1.1175	358.28
lpopden 2002	0.9503	0.0758	0.8018	1.0988	157.32
lpopden 2003	1.221	0.0781	1.0678	1.3742	244.11
lland 2001	0.9846	0.0502	0.8862	1.0829	385.05
lland 2002	0.9652	0.071	0.8262	1.1043	185.06
lland 2003	1.1753	0.07	1.0382	1.3125	282.16
Exp(pophous 2003)	0.7594	0.0914	0.5998	0.9614	
Exp(PctPlumbing)	1.0321	0.0093	1.014	1.0505	

However, the coefficient of the log of population density (i.e., population / land area) is remarkably similar to the coefficient of the log of land area in all three years. Since both variables are measured as natural logs, equality between the two coefficients would mean that population alone, a measure of geographic market size presumably—and not density—is what is increasing the odds of broadband.[42] A formal statistical test of equality between these coefficients in all three years, jointly, does lead us to reject the hypothesis that is the effects of absolute population size, and not population density, that the density variable is capturing.[43] The implication is that the frequently observed correlation between population density and broadband availability is really capturing the relationship between absolute market size and broadband availability, since denser zip codes are also zip codes with larger populations.

Two other variables related to population size and density yield statistically significant coefficients. All else equal, increased population per housing unit actually *reduced* the odds of finding broadband available in a zip code in 2003: an average increase of one person per housing unit (almost a 50% increase, since the mean in my sample was 2.31 persons per unit) reduced the relative odds of broadband by almost a quarter. (Note that I have exponentiated my estimated coefficients to translate these into an impact on odds ratios, and calculated empirical standard errors for the transformed coefficients; I will generally do this in evaluating the impact of variables where I have not used a logarithmic specification.)[44]

The percent of housing units lacking complete plumbing (hot and cold running water, flush toilet, and bath or shower) had a positive, statistically significant impact on broadband availability. I interpret this as a measure of the absolute size of the urban area an urban zip code is located in. Detailed study of responses for this variable in the 2000 census shows that this variable is higher in rural areas, and in the largest metropolitan areas with older residential housing cores. In fact, there is a clear positive correlation between percent of housing units with incomplete plumbing and absolute size of urban area in which the housing unit is located.[45] I control separately for the percent of a zip code's population which is rural or living on a farm (see analysis below), but this is the only variable that is directly associated with the absolute size of an urban area in which a zip code may be located. I therefore interpret this variable as a proxy related to size of urban markets.[46]

---

[42] Since a ln(pop/land) + b ln(land) = a ln(pop) + (b-a) ln (land), and (b-a)=0 with equal coefficients.

[43] The Wald test for the joint hypothesis of equality between the two coefficients in all three years is 3.75, with 3 degrees of freedom. The probability of a chi-square statistic of this or greater magnitude under the null of equality is .29, so we are not even close to rejecting the null at a 5% significance level. A score test produces an identical result, with a chi-square of 3.73 and 3 degrees of freedom.

[44] Note also that despite the fact that differences from 2003 for population/housing unit are statistically significant for 2001 and 2002, we cannot reject the hypothesis that the entire effect (i.e., after adding the increment onto the 2003 base) is zero.

[45] See S. Gasteyer and R.T. Vaswani, "Still Living Without the Basics in the 21st Century: Documenting the Extremes of the Water Crisis in the United States," (Washington: Rural Community Assistance Partnership), 2004, Table 3, p. 13, available at http://www.rcap.org/resources/basics.html.

[46] The effect, while highly statistically significant (a score test of the null hypothesis that these effects are zero yields a chi-square of 12.48, with 3 degrees of freedom), is small. An increase of one percentage point in share of units lacking complete plumbing (which is better than a doubling, since the mean in my sample

Taken together, all of the above evidence suggests that it is absolute size of market, and not population density, that influences broadband availability. Nonetheless, there are some indications that denser housing types within zip codes are associated with differences in broadband availability (see the discussion of housing below).

The Usual Suspects: Income and Wealth

Variables related to household income and wealth have the expected positive impacts on broadband availability. The effects are small, however. The elasticity of the broadband odds ratio with respect to per capita income is about 1/3. Because most zip codes have high odds already for broadband availability, the aggregate empirical elasticity of broadband availability with respect to a hypothetical ten percent across the board rise in per capita income in all zip codes is much smaller, and falls over time as broadband becomes increasingly available. A simulated ten percent increase in per capita income throughout my sample leads to point estimates of a mere .35% increase in the mean number of zips with broadband availability in 2001, .22% in 2002, and only a .12% increase in 2003. A ten percent decline in per capita income everywhere reduces broadband zip code numbers by a slightly greater .39% in 2001, .25% in 2002, and .14% in 2003.

A slightly higher elasticity with respect home value, the principal store of much consumer wealth, is measured—about .37. Another major store of consumer wealth, an automobile, also appears to be statistically significant: a one percentage point increase in the share of households *without* autos reduces the relative odds of broadband availability in a zip code by almost 2 percent.[47]

Table 5 Income and Wealth-related Effects

Variable	Estimate	Std. Error	95% Conf. Interval		Chi-Square
lpci	0.3338	0.1208	0.0971	0.5705	7.64
lavghval	0.3739	0.0744	0.228	0.5198	25.23
Exp(PctNoCars 2003)	0.9815	0.0075	0.9669	0.9963	

Rural vs. Urban Status

Greater rural character in a zip code clearly seems to reduce broadband availability. (Table 6) With a one percentage point increase in the share of rural population, odds of broadband being available in a zip code were reduced by one percent. The percent of the population living on farms had a marginally significant impact (i.e., reject a zero effect at the 10% level, but not at the 5% level), but an even smaller effect. If we were to not reject

---

was .95%) leads to a three percent increase in broadband availability. Nationwide, the percent of units without complete plumbing rose from .41 percent in urban areas with from 50,000 to 99,999, to .80 percent in urbanized areas with more than 5 million inhabitants. See Gasteyer and Vaswani, op. cit.

[47] This effect is opposite in sign from the effect one would expect to see if broadband use were causing consumers to give up their cars—i.e., a positive association between no cars and broadband availability. In any event, the data on car use is for early 2000, and the data on broadband for late 2001 and after, so it would be difficult to argue that the latter was causing the former. Reading availability of a car as an indicator for household wealth would seem more sensible.

this effect as equal to zero, a one percentage point increase in the share of farm-dwelling population leads to a .6 percent reduction in the broadband odds.

Table 6 Rural vs. Urban Effects

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(Pctrural 2002)	0.9919	0.0029	0.9863	0.9976
Exp(Pctrural 2003)	0.9902	0.0038	0.9827	0.9977

Industry Presence

The right sort of industry seems to significantly increase the odds of broadband in a zip code—particularly professional scientific, and technical services—while the wrong sorts of establishments—particularly administrative, support, waste management, remediation services—may have reduced it substantially. (Table 7) An additional establishment in manufacturing (NAICS code 31, e3101 is number of establishments in this sector in 2001) raised the odds of broadband by about 6 percent in 2001. An additional NAICS 44 (retail trade) establishment yielded a 3.8% improvement in the broadband odds in 2002, and a 6.4% increase in 2003. Another professional, scientific, and technical services establishment (NAICS 54) raised the odds in 2001 by a whopping 10%. An accommodation and food services (NAICS 72) establishment increased the odds of broadband in 2002 by about 8.4 percent. An establishment in other services (NAICS 81) clocked in at a 4 percent rise in 2001. An additional NAICS 56 (Administrative, support, waste management, remediation services) in 2001 or NAICS 62 (Health care and social assistance) in 2002 seemed to actually reduce the probability of broadband being offered, by 9 and 3 percent, respectively.

Table 7 Industry Presence

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(e3101)	1.0584	0.0235	1.0132	1.1055
Exp(e4402)	1.0377	0.0131	1.0123	1.0638
Exp(e4403)	1.0643	0.0162	1.033	1.0966
Exp(e5401)	1.1012	0.023	1.0569	1.1472
Exp(e5601)	0.9078	0.033	0.8455	0.9748
Exp(e6202)	0.968	0.0059	0.9565	0.9796
Exp(e7202)	1.0844	0.0193	1.0473	1.1229
Exp(e8101)	1.0398	0.0187	1.0038	1.0772

Household Demographics

In 2001, zip codes with larger shares of college-age students (20 to 24 years old) seemed to have somewhat lower odds of broadband availability. This seems to have disappeared over time, however, and one may speculate that availability of higher speed Internet connection on campus discouraged this group from purchasing home broadband connections until more competitive pricing and availability was available. A higher share of married couples seems to discourage broadband availability—perhaps because broadband competes for time with other household social activities.

A one percentage point increase in the share of the population with limited English-speaking ability increases broadband odds by almost 2 percent. A reasonable interpretation of this is that, cet. par., immigrants and migrant workers are more likely to use higher quality Internet service to maintain contact with their homes and families overseas.

A larger female population seemed less likely to encourage broadband availability. A one percentage point increase in the share of women in the population reduced broadband odds by about 1.3 percent. This mirrors similar findings using entirely different data and models to estimate a broadband demand function empirically.[48]

Table 8 Household Demographics

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(PctAge20_24 2001)	0.9781	0.0105	0.9577	0.9989
Exp(PctMarriedCouples2002)	0.9909	0.0036	0.9838	0.9981
Exp(PctMarriedCouples2003)	0.9867	0.0041	0.9787	0.9948
Exp(PctEnglishLimited)	1.019	0.0071	1.0052	1.033
Exp(PctFemale)	0.9868	0.0063	0.9744	0.9993

The Digital Divide: Ethnicity

A population's ethnicity or race had perceptible impacts on the odds of broadband in a zip code though our reduced form gives us no clue as to whether this was being expressed through demand for, or supply of, broadband service. (Table 9) A one percentage point increase in the percentage of the population identified as black reduced broadband odds by .6 percent in 2001. Interestingly, this digital divide seems to be closing—this differential was not statistically significant after 2001.

A one percentage point increase in the native Hawaiian/Polynesian islander population share, all else equal, reduced the odds of broadband by about 14% in 2002 and 2003. Interestingly, this digital divide seems to have opened up since 2001. In our cleaned sample of 52 Hawaiian zip codes, roughly 40 percent had no broadband available in 2001. In 2002 and 2003, zips with no broadband service dropped to under 4 percent of zip codes in my Hawaiian sample. A major expansion in broadband availability in Hawaii seems to have occurred throughout Hawaii between late 2001 and late 2002. That expansion also seems to have created a "digital divide" of sorts where none previously existed, leaving native Hawaiian communities behind in disproportionate numbers.

Native American communities seem to suffer from a smaller, but very persistent lag in broadband availability. A one percentage point increase in Native Americans' share of

---

[48] See K. Flamm and A. Chaudhuri, "An Analysis of the Determinants of Broadband Access," August, 2005, to be presented at the Telecommunications Policy Research Conference, September 24, 2004, accessible at http://www.tprc.org/TPRC05/Sat830Sess05.htm#NewEcon .

the population was associated with a 1.3 to 1.4 percentage point reduction in broadband odds, and this seems to have been relatively static over 2001-03.

Share in population of Asian and multiracial groups, as well as self-identification as Hispanic, had no statistically significant effect on broadband availability, and was dropped from the parsimonious model after preliminary statistical tests.

Table 9 Ethnicity and the Digital Divide

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(PctBlack101)	0.9941	0.0025	0.9892	0.9991
Exp(PctHawnPI102)	0.8632	0.0457	0.7781	0.9576
Exp(PctHawnPI103)	0.857	0.0382	0.7853	0.9353
Exp(PctIndian101)	0.9868	0.0034	0.9802	0.9935
Exp(PctIndian102)	0.9864	0.0038	0.979	0.9939
Exp(PctIndian103)	0.9861	0.0045	0.9772	0.995

Education

Greater educational attainment within a zip code had relatively strong effects on greater broadband availability. A one percentage point increase in the share of the population with a high school degree increased broadband odds by about 1.8 percent. A percentage point more of the population with some college, a BA, or a graduate or professional degree increased the odds of broadband by a little over 2 percent. The effect generally increased with educational attainment, mirroring results obtained from estimates of broadband demand using entirely different models and data sets.[49]

Table 10 Education

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(PctHighSchool 2001)	1.018	0.0053	1.0078	1.0284
Exp(PctHighSchool 2002)	1.0182	0.0054	1.0077	1.0289
Exp(PctHighSchool 2003)	1.0145	0.0058	1.0031	1.026
Exp(PctSomeCollege)	1.0215	0.0051	1.0115	1.0316
Exp(PctBachelors)	1.0225	0.0071	1.0086	1.0366
Exp(PctGradProf)	1.0207	0.0097	1.0018	1.0399

Employment by Industry

The share of the population working in two broad industry groups seemed to reduce the broadband odds slightly. An increase of one percentage point in the share of the population working in manufacturing lowered broadband odds by about half of a percent. It would seem that manufacturing employment encourages broadband use-enhancing skills less than employment in other sectors.

---

[49] See Flamm and Chaudhuri, 2005.

Perhaps more surprisingly, a one percentage point increase in the share of the population working in education services reduced the odds of broadband by about 1 percent. This is most easily explained by access to high speed internet connections at work—schools and colleges—being substituted for a high speed connection that might otherwise be purchased in the home.

Table 11 Employment by Industry

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(PctEducation)	0.9893	0.0047	0.9801	0.9987
Exp(PctManufacturing)	0.9945	0.0027	0.9894	0.9998

Housing Stock Characteristics

Characteristics of housing units turn out to have a surprisingly varied and strong set of effects on the odds of broadband availability. They are summarized in Table 12. Log of average age of housing units (which is most likely inversely related to quality of housing stock and ease of connection and distribution of a high speed Internet connection) has a pronounced negative impact on broadband. A 10 percent increase in average housing age is associated with an 8 percent decline in the odds of broadband being available.

Increased density in housing units in some cases seems to be associated with greater broadband availability. An increase of one percentage point in the share of population living in group quarters (rooming houses, group homes, dormitories, military quarters, hotels, etc.) led to a two percent increase in broadband odds. A one percentage point increase in the share of housing units occupied (as opposed to being vacant), all else being equal, increased broadband odds by a percentage point in 2002 and 2003. An increase in the share of housing units in buildings with 2 to 4 units bumped broadband odds up by almost 3 percent in 2001.

In other cases, denser living situations reduced broadband availability. A one percentage point increase in the share of institutionalized housing (requiring a pass or escort to enter or leave, e.g., prisons, hospital wards) reduced broadband odds by 3.2 percent. Very large scale housing complexes (20 or more units), all else equal, also seem to depress broadband availability, with a one percentage point increase in the share of these housing units reducing the odds by almost 4 percent. One possible explanation is that it is simply more expensive to retrofit large complexes or high rises for broadband distribution.

Table 12 Housing Stock Characteristics

Variable	Estimate	Std. Error	95% Conf. Interval		Chi-Square
lavgage 2001	-0.791	0.305	-1.3887	-0.1933	6.73
Exp(PctGQPop)	1.021	0.009	1.0036	1.0388	
Exp(PctInInstitutions	0.968	0.0094	0.9497	0.9866	
Exp(PctMobileHomes)	0.9917	0.0027	0.9865	0.9969	
Exp(PctOccupied 2002)	1.0085	0.0036	1.0014	1.0156	
Exp(PctOccupied 2003)	1.0108	0.0041	1.0028	1.019	

Exp(PctUnits2 2001)	1.0286	0.0098	1.0095	1.048
Exp(PctUnits20Up 2003)	0.9615	0.0167	0.9293	0.9949

The eRate Program

The largest program supporting Internet and broadband access subsidized by government is the so-called eRate program, which takes Universal Service Fund revenues and makes grants to schools and libraries. Though not primarily designed to promote broadband access outside of schools and libraries, it is possible that extending broadband access to these institutions makes “spillover” to nearby residences more economic. It is also possible that school and library broadband connections could stimulate demand for greater home access by users of these facilities.

The data seem to suggest a statistically significant effect of eRate funding on broadband availability in zip codes containing institutions receiving that funding. For every ten thousand dollars of cumulative eRate funding authorizations received through 2001, broadband odds in a zip in 2002 increased by .8 percent on average. This is a relatively large effect (mean cumulative eRate funding authorizations in our sample of zip codes through 2001 was $318,000). On the other hand, this was at the early stages of the takeoff in US broadband penetration, and there is no evidence of a similarly strong effect a year later, in 2003, as US broadband use accelerated.

In addition, the distribution of eRate funds among zip codes was such that the aggregate impact of eRate funding in working to make broadband available where it otherwise might not have been was quite small. Using the point estimates of eRate impacts shown in Table 3, I simulated the expected impact of zeroing out eRate funding in my sample of 25,000+ zip codes. It turns out that eRate funding, on average, made a difference in making broadband available in about 21 of these zip codes in 2001 and 2002, and just 1 in 2003. While the eRate funding had a reasonably significant impact on the odds of broadband availability, it was distributed (perhaps unsurprisingly, since it was not designed for this purpose) in a way that did little to change broadband availability to otherwise unserved regions.

No impact was detected for the rural health care program administered by USF.

Table 13 eRate Impact

Variable	Estimate	Std. Error	95% Conf. Interval	
Exp(erate 2002)	1.0081	0.0040	1.0002	1.0160

Geography and Terrain

One of the most interesting results of my analysis is the large number of statistically significant geography and terrain variables, and the sensitivity of empirical results to their use. These variables are almost certainly working through service provision cost. Table 14 displays a summary of these results.

Table 14 Geography and Terrain

Variable	Estimate	Std. Error	95% Conf. Interval		Chi Sq.
lelevrang	-0.2066	0.069	-0.3419	-0.0713	8.96
lslopesd	0.1789	0.0641	0.0533	0.3045	7.79
Exp(HARD01)	1.0038	0.0009	1.002	1.0056	
Exp(lelevrang)	0.8134	0.0561	0.7104	0.9312	
Exp(lslopesd)	1.1959	0.0767	1.0547	1.356	
Exp(modisqc00)	0.935	0.0258	0.8859	0.9869	
Exp(modisqc01)	0.9323	0.0257	0.8833	0.9841	
Exp(modisqc02 2001)	0.9201	0.0304	0.8625	0.9817	
Exp(modisqc02 2003)	0.929	0.0348	0.8632	0.9998	
Exp(modisqc04 2001)	0.9344	0.0256	0.8855	0.986	
Exp(modisqc04 2002)	0.9348	0.0257	0.8858	0.9866	
Exp(modisqc04 2003)	0.9322	0.0257	0.8832	0.984	
Exp(modisqc05)	0.9395	0.026	0.8899	0.9918	
Exp(modisqc07)	0.9357	0.0259	0.8862	0.9879	
Exp(modisqc08)	0.9346	0.0259	0.8852	0.9867	
Exp(modisqc09 2001)	0.9366	0.0279	0.8835	0.9929	
Exp(modisqc10)	0.9363	0.0258	0.887	0.9883	
Exp(modisqc11)	0.8823	0.0514	0.787	0.9891	
Exp(modisqc12)	0.9367	0.0259	0.8874	0.9889	
Exp(modisqc14)	0.9378	0.0259	0.8884	0.99	
Exp(modisqc16)	0.9119	0.0283	0.8582	0.9691	

MODIS land cover types for water bodies (MODIS type 0), evergreen needle leaf (1) and broadleaf (2) forests, deciduous broadleaf forests (4), mixed forests (5), open shrub lands (7), woody savannas (8), other savannas (9), grasslands (10), permanent wetlands (11), croplands (12), cropland/natural vegetation mosaic (14), and barren or sparsely vegetated land (16) all seem to lower broadband availability relative to a built-up urban land cover baseline. For most of these land cover types, a one percentage point increase in the land cover type is associated with a 4 to 6 percent decline in broadband odds. Exceptions are permanent wetlands (11) where the penalty is about double—a twelve percent decline in broadband odds—and barren scrublands (16) with a nine percent decline in the odds of broadband.

Other significant terrain variables are the range between maximum and minimum elevations within the half degree map grid containing a zip code centroid, and the standard deviation of slopes within the half degree map grid. The former is a measure of “vertical rise” within an area, while the latter might be considered a measure of “hilliness” within an area. The empirical results suggest a ten percent increase in “vertical rise” is associated with about a 2 percent decline in broadband odds, while a ten percent increase in “hilliness” is associated with a 2 percent increase in the odds ratio. Interestingly, hilliness seems to be more advantageous than flat or smoothly rising or falling terrain, perhaps because it makes wireless distribution technology more economic over larger areas.

Finally, an increase in the share of hard rock over soft and normal rock types seems to have a very slight, but statistically significant effect in 2001. A one percentage point increase in the share of hard rock seems to have been associated with a slight increase of .38 percent in the odds of broadband. One may speculate that there may be some advantages to building foundations for distribution systems and other facilities on hard rock, or this may simply be an artifact of type I error inevitably realized when large numbers of parameters are estimated.

State Effects

Fixed state effects estimated in this paper would naturally be associated with differences in government policies regulating or encouraging broadband use across states, but could reflect any source of variation that is specific to a state, whether or not it is the result of a specific state government's policy. For example, if a private group or foundation conducted a statewide campaign to subsidize broadband, or if an incumbent local exchange carrier operating in a state decide to make that state the focus of a broadband rollout, that "state effect" would be indistinguishable statistically from a policy enacted by an actual state government.

On the other hand, there are few if any studies which actually attempt to assess the efficacy of a state's policies in encouraging or retarding broadband availability. At best, one might compile a list of types of policies pursued by state governments, and then try to relate the selection of policies pursued to the outcomes observed.[50] But such lists are inevitably incomplete, rarely if ever measure success in design and execution of individual policies on the lists, and lack the flexibility to include novel or imaginative policies that have features extending beyond the "cookie cutter" taxonomies used to build such lists.

To begin, I emphatically reject the null hypothesis that all state policy effects are zero with a formal statistical test.[51] Table 15 displays statistically significant state effects. State names are displayed next to variable names, which incorporate FIPS codes.

My baseline is Texas (FIPS 48), which had a modest but sustained Internet and broadband subsidy program in place from 1996 to 2004 (TIF, the Telecommunications Infrastructure Fund, which distributed about $1.5 billion over this period), and a relatively competitive regulatory environment. Overall, my prior would be that Texas had a very encouraging policy environment for broadband use. For a state and time period not listed in Table 15, I could not reject the hypothesis that individual state effects were the same as Texas in overall impact on broadband use.

Table 15 State Policy Effects

---

[50] For a recent example of such an attempt, see S. Wallsten, "Broadband Penetration: An Empirical Analysis of State and Federal Policies," AEI-Brookings Joint Center for Regulatory Studies Working Paper 05-12, June 2005, available at http://www.aei-brookings.org/admin/authorpdfs/page.php?id=1161 .

[51] The Wald test statistic for the joint hypothesis that all 78 state/time effects in my parsimonious model are zero was 754.3 with 78 degrees of freedom; the probability that this statistic, distributed as chi-square under the null, would be greater than or equal to this was less than .0001. A score test for the same hypothesis yielded a test statistic of 735.6, with the same degrees of freedom, and an equally tiny probability that it would be greater or equal to this value under the null.

Variable	Estimate	Std Error	95% Conf. Interval		State
Exp(s1 2003)	0.3778	0.1051	0.2189	0.6518	Alabama
Exp(s15 2001)	0.0007	0.0008	0.0001	0.0069	Hawaii[52]
Exp(s16 2001)	0.4552	0.1063	0.288	0.7194	Idaho
Exp(s17 2001)	0.4919	0.0763	0.3629	0.6666	Illinois
Exp(s17 2002)	0.6507	0.1099	0.4673	0.9061	
Exp(s17 2003)	0.4641	0.0929	0.3135	0.6871	
Exp(s18)	0.4302	0.0745	0.3064	0.6041	Indiana
Exp(s19 2001)	0.183	0.0296	0.1332	0.2512	Iowa
Exp(s19 2002)	0.3185	0.0536	0.229	0.4429	
Exp(s19 2003)	0.217	0.0404	0.1506	0.3125	
Exp(s20 2001)	0.4084	0.0633	0.3014	0.5533	Kansas
Exp(s20 2002)	0.5472	0.0931	0.392	0.7639	
Exp(s20 2003)	1.6838	0.3988	1.0585	2.6784	
Exp(s21)	0.536	0.0978	0.3748	0.7664	Kentucky
Exp(s23 2001)	2.002	0.5031	1.2234	3.2761	Maine
Exp(s23 2003)	0.3391	0.092	0.1992	0.5772	
Exp(s24 2001)	0.55	0.1451	0.328	0.9225	Maryland
Exp(s24 2002)	2.4962	1.1416	1.0185	6.1175	
Exp(s26)	0.6432	0.115	0.4531	0.913	Michigan
Exp(s27)	0.199	0.0388	0.1358	0.2917	Minnesota
Exp(s28)	0.6047	0.1175	0.4132	0.8849	Mississippi
Exp(s29 2001)	0.7083	0.1137	0.5171	0.9701	Missouri
Exp(s29 2002)	0.6559	0.1101	0.4721	0.9114	
Exp(s29 2003)	0.3941	0.0721	0.2753	0.5642	
Exp(s31)	0.491	0.0775	0.3604	0.669	Nebraska
Exp(s33 2001)	3.0931	1.4468	1.2367	7.7363	New Hampshire
Exp(s36)	2.0154	0.4222	1.3368	3.0386	New York
Exp(s38 2001)	0.5395	0.0996	0.3758	0.7746	North Dakota
Exp(s38 2002)	0.4729	0.0911	0.3242	0.6897	
Exp(s41)	2.0109	0.5369	1.1915	3.3936	Oregon
Exp(s42)	0.3476	0.068	0.2369	0.5099	Pennsylvania
Exp(s46 2001)	0.4677	0.0855	0.3269	0.6694	South Dakota
Exp(s46 2002)	0.6614	0.1283	0.4522	0.9674	
Exp(s46 2003)	0.5158	0.1106	0.3389	0.7852	
Exp(s47 2001)	0.5599	0.1166	0.3723	0.842	Tennessee
Exp(s49 2003)	0.6939	0.3302	0.2731	1.7633	Utah
Exp(s49 2001)	0.3667	0.108	0.2059	0.653	
Exp(s51)	0.4422	0.0848	0.3038	0.6439	Virginia
Exp(s53 2003)	0.4745	0.1489	0.2566	0.8775	Washington
Exp(s54 2001)	0.5589	0.1176	0.37	0.8442	West Virginia
Exp(s55)	0.4422	0.0851	0.3033	0.6448	Wisconsin
Exp(s8)	2.0519	0.5121	1.2581	3.3464	Colorado

Groups of states break out relative to Texas as follows:

---

[52] In interpreting this coefficient, recall that 40 percent of my sample's Hawaii zip codes in 2001 had no broadband service available; a year later, that number had changed to 4 percent.

Greater encouragement of broadband use: Kansas, 2003; Maine, 2001; Maryland, 2002; New Hampshire, 2001; New York all years; Oregon, all years; Colorado, all years.

Lesser encouragement of broadband use: Alabama, all years; Hawaii, 2001; Idaho, 2001; Illinois, all years; Indiana, all years; Iowa, all years; Kansas, 2001-02; Kentucky, all years; Maine, 2003; Maryland, 2001; Michigan, all years; Minnesota, all years; Mississippi, all years; Missouri, all years; Nebraska, all years; North Dakota, 2001-02; Pennsylvania, all years; South Dakota, all years; Tennessee, all years; Utah, 2001-03; Virginia, all years; Washington, 2003; West Virginia, 2001; Wisconsin, all years.

No statistically significant differences, any years: California; Florida; Georgia; North Carolina; New Mexico; Ohio; Oklahoma; South Carolina; Vermont; Wyoming.

These rankings are loosely consistent with appreciations of state policies to advance broadband penetration found in the press. They may be useful when combined with other information. For example, one interesting result is that Pennsylvania's broadband promotion policies, including the often discussed "Chapter 30" policy, which traded significant deregulation of telephone rates in exchange for incumbent local exchange carrier commitments to accelerate deployment of broadband, seem to have performed notably worse in spurring broadband availability, cet. par., than, say, policies pursued in Texas.[53]

Other Issues

One possible concern is that causality runs from broadband use to certain of my explanatory variables, particularly use of a car by households, industry establishment variables, and the eRate and rural health care subsidy programs administered by the Universal Service Fund. But the population and housing census variables measure zip code characteristics in early 2000, prior to the widespread availability of broadband, and thus measure the preexisting demographic and economic characteristics of a zip code before broadband availability had any real opportunity to transform the demographics and economics of households. Similarly, the establishment data for industry by NAICS code measures the industry base in a zip code in 1997, in an even earlier and more distant period. The significant impacts of composition of the industrial base (employment by industry and establishments by industry variables) on broadband availability, years after that base is measured, mean there are serious endogeneity and causality issues to be faced if broadband availability is being used to explain contemporaneous industrial growth and composition.

The eRate and rural health care programs are notorious for long lags between applications, approvals, and expenditures. These long lags make it improbable that current broadband availability has played a significant role in generating current expenditures on eRate programs. The functioning of the eRate program, and its

---

[53] See, for example, "Pennsylvania PUC Strikes Broadband Deal With Verizon," **Telecom Policy Report**, July 23, 2003, A.K. Glasmeier, L.E. Wood, A.N. Kleit, and S. Strover, "Broadband Internet Service in Rural and Urban Pennsylvania: A Common Wealth or Digital Divide?", 2002, available at http://www.utexas.edu/research/tipi/reports2/akg%20CRP%20final%20report%20rev3%206-03.pdf.

relationship to supply and demand for broadband service among consumers and businesses, is a subject that could fruitfully be explored in future research.

Finally, a major finding of this paper is the importance and statistical significance of geophysical and terrain variables in a reduced form explaining broadband availability. This opens the door to a possible solution to a major problem for empirical research on broadband demand—the lack of data measuring broadband service quality. Unmeasured and uncontrolled variation in broadband quality is clearly a problem for econometric estimation of broadband demand functions, since quality is likely to be positively correlated with reported service price. Use of geophysical and terrain variables, which shift costs, as instruments, opens the door to a possible solution to these problems.

Conclusion

This initial analysis of a rich new data set on broadband penetration has yielded some intriguing results. My analysis suggests that state policies may play an important role, and that statistical methods are useful in assessing this role. The ranking of state effects produced by my model seems to correlate with casual impressions of the effectiveness of state policies as portrayed in the press and trade journals.

To some extent, at least, geography is destiny. Terrain effects (presumably increasing or decreasing the cost of installing and maintaining a network) seem to be significant and important. Their use as instrumental variables also offers a possible solution to the problem of unmeasured quality variation in estimation of broadband demand functions.

Two factors, income and population density, are often said to be associated with broadband penetration. Income and wealth variables unsurprisingly seem to be among the most statistically significant determinants of broadband penetration. Population density's role is much more problematic. When a full set of explanatory variables is used, the results suggest that absolute market size, not some measure of density relative to physical area, is the key determinant. Density seems much more important in understanding the how the detailed composition of the housing stock affects deployment costs. My results suggest that denser housing types may work to increase or decrease broadband availability, depending on the details.

The Universal Service Fund's eRate program did appear to play a statistically significant role in encouraging broadband use in 2002. On the other hand, it was not intended to be a solution to a more general broadband access problem, and in practice it seems to have had a minimal impact. As the scale of broadband deployment accelerated in 2003, we could not reject the hypothesis that the impact of this program on broadband availability was effectively nil.

Industrial activity seems to have a significant impact on local broadband availability. Professional and technical service establishments seem to have the largest such impact.

Common perceptions of the effects of gender, education, and rural location on broadband penetration seem to be supported by a causal analysis that attempts to control for confounding factors.

Finally, “digital divide” type ethnic, racial and personal variables show up as small, but statistically perceptible effects. There were reduced odds of broadband provision in zip codes with larger Afro-American and Native American populations in 2001. Though the gap seemed to be closing for Afro-Americans in 2002-03, a new gap for native Hawaiian Polynesians opened up, and the Native American deficit continued.