The Role of Economics, Demographics, and State Policy in Broadband Competition: An Exploratory Study

Kenneth Flamm

October 1, 2004

Revised Preliminary Draft
Comments Only Please

Presented at the Telecommunications Policy and Research Conference
Arlington, Virginia
October 2, 2004

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 Competition: An Exploratory Study

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. Preliminary analysis of the factors affecting penetration of broadband into underserved areas has uncovered a significant role for factors that are not much discussed in the current political debate. At the top of the list is local telephone competition, which seems to be the most important single factor associated with greater broadband availability. A preliminary analysis also suggests that state policies may play an important role. The two influences most often correlated with broadband penetration, income and population density, paradoxically seem to be among the least important determinants of broadband penetration. Income effects, on balance, are positive, but quite small. Industrial activity seems to have a significant impact on broadband use. Several measures of infrastructure quality have significant positive impacts on broadband penetration. Finally, "digital divide" type ethnic, racial and personal variables show up as small, but statistically perceptible factors.

The Role of Economics, Demographics, and State Policy in Broadband Competition: An Exploratory Study

Although the United States was the undisputed leader in development and deployment of the Internet and its underlying technologies, the United States has most definitely not been the global leader in the deployment of ubiquitous high speed broadband.[1] Official International Telecommunications Union statistics listed the United States as number 11 in broadband penetration in 2002, with 6.5 broadband subscribers per 100 inhabitants—about 18% of all Internet subscribers—and about 19% of all households with Internet connectivity making use of broadband.[2] Only 10% of all households had a broadband connection in 2002.

By contrast, the leader in these rankings Korea, had a broadband subscription rate equal to 21.3 percent of its population, and 94 percent of its Internet subscribers had a broadband connection. Some 83% of Korean households with an internet connection made use of broadband, as did 43 percent of all Korean households. Our northern neighbor Canada was number 3 on this list, had more than double the U.S. broadband penetration rate, with half of all its Internet subscribers using a broadband link, and also had roughly double the rates seen in the United States for broadband penetration among both Internet and all households.

Given the increasing emphasis among analysts on the role, actual and potential, of information technology in productivity growth,[3] it is not surprising that policies to accelerate deployment of broadband Internet communications have been a topic for political discussion in recent years. Presidential candidate Al Gore called for building an "information superhighway" for all Americans during the 2000 presidential campaign. Most recently, in February 2004, presidential candidate John Kerry blasted the Bush administration for failing to create a national broadband policy. In March, President Bush embraced the goal of universal broadband access[4], and later publicized a series of policy

[1] I thank Anindya Chaudhuri for his invaluable research assistance, and the Policy Research Institute of the Lyndon B. Johnson School of Public Affairs for its generous financial support for some of the work going into this paper. Without implicating them in my errors, I thank Anindya Chaudhuri, Chandler Stolp, Sharon Gillett, Bill Lehr, James Prieger, and Gerald Faulhaber for helpful comments as this paper was being written.

[2] These rankings are available at http://www.itu.int/ITU-D/ict/statistics/at_glance/top15_broad.html.

[3] 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).

[4] In a speech given in New Mexico on March 26, 2004, MSNBC News reported that President Bush urged that "affordable high-speed Internet access be available to all Americans by 2007, saying it was essential to the nation's economic growth". MSNBC, "Bush calls for universal broadband by 2007," available at http://www.msnbc.msn.com/id/4609864/. CNet News quoted Bush as saying "We ought to have universal, affordable access to broadband technology by the year 2007…And then, we ought to make sure that as soon as possible thereafter, consumers have plenty of choices." See John Borland, "Kerry's broadband policy plans emerging," available at http://zdnet.com.com/2100-1103_2-5197218.html?tag=nl. Infoworld reported that based on official transcripts, "Bush made a call to keep broadband prices low and added that

initiatives intended to promote broadband development and deployment.[5] Concrete measures announced to date in support of this goal include increased depreciation subsidies to capital investment, a proposed extension of a moratorium on Internet access taxes, a federally-funded R&D initiative, FCC regulatory relief for fiber optic connections, a streamlined process for granting broadband providers access to federal land, a government-sponsored standards program for digitization and dissemination of medical information, and a proposed broadening and extension of tax credits for R&D.[6] Challenger John Kerry in June 2004 rolled out proposals to define broadband as a new form of universal service, including a 10 percent tax credit for investments in current broadband technology used in rural and inner city areas, a 20 percent tax credit for investments in next generation high speed investment, over the next 5 years, universal broadband for "first providers" (firefighters and police) by 2006, turning over unused portions of the spectrum for wireless broadband services by the public and first providers, and unspecified measures to promote privates sector investment and increased competition in broadband.[7] A similar proposal to make more spectrum available for wireless broadband was also unveiled by President Bush in a speech on the very same day, June 24, 2004.[8]

Common threads running through all these proposals include a belief that broadband pricing is a significant barrier to greater broadband use, and that insufficient investments in broadband technology by broadband service providers have been a major impediment to wider deployment of broadband. This paper does not aspire to address the first question,[9] but does muster evidence that has some relevance to the second point. Interestingly, the regulatory framework shaping broadband diffusion has been given less emphasis by both candidates, perhaps because it is a patchwork hybrid of both federal and state policy. My analysis will also examine whether substantial variation across states in state-specific factors, including regulatory policies, may be having an 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, and use this data to estimate the parameters of an economic model of entry into broadband service markets.

'Congress must not tax access to broadband technology if we want to spread it around.'". The President clearly pointed to pricing as the key to increasing deployment rates. "'The more the price goes down, the more users there will be,' he said." See http://www.infoworld.com/article/04/03/29/HNbushbroadband_1.html.

[5] See http://www.whitehouse.gov/infocus/technology/tech2.html; http://zdnet.com.com/2100-1104_2-5200196.html.

[6] The R&D program is known as the Networking and Information Technology Research and Development (NITRD) initiative.

[7] See "John Kerry's Plan to Create Millions of High-Wage Jobs in the Industries of the Future," available at http://www.johnkerry.com/pdf/pr_2004_0624b.pdf.

[8] Josh Long, "Bush, Kerry Promise to Expand Broadband over Airwaves," available at http://www.x-changemag.com/articles/481front1.html.

[9] On the first issue, see Anindya Chaudhuri, Kenneth Flamm, and John Horrigan, "An Analysis of the Determinants of Internet Access," also to be presented at TPRC 2004.

FCC Data on Broadband Deployment

The Federal Communications Commission has been gathering data on broadband service deployment since 2000. 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.[10]

Second, each service provider is required to identify each zip code in which it supplies at least one high-speed line. Obviously, the 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. Table 1 shows aggregate U.S. data on 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 2003, less than 7% of U.S. zip codes had no reporting high-speed line providers.[11]

As we shall see below, the FCC appears to be significantly underestimating the number of zero-provider zip codes in its statistics and published reports. Since absolute numbers of uncounted zero provider zip codes have almost certainly declined over time, the rate of increase in broadband penetration across zip codes has been even greater than suggested by these tables.

Percentage of Zip Codes with High-Speed Lines in Service

Number of Providers	1999	2000		2001		2002		2003	
	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 %
One	26.0	25.9	22.7	20.3	19.3	18.4	17.3	16.4	14.9
Two	15.5	17.8	18.4	16.7	15.7	16.2	16.8	16.9	17.1
Three	8.2	9.2	10.9	13.2	13.1	13.3	14.4	14.0	14.9
Four	4.3	4.9	6.1	8.2	9.1	9.6	10.3	10.6	11.2
Five	2.7	3.4	4.0	4.9	6.1	6.9	7.3	7.7	7.8
Six	1.7	2.5	3.0	3.6	4.2	4.6	5.0	5.3	5.8
Seven	0.8	1.7	2.3	2.8	3.2	3.2	3.9	4.0	4.2
Eight	0.3	0.8	2.0	2.2	2.5	2.8	2.7	3.1	3.3
Nine	0.2	0.4	1.6	1.9	2.0	2.4	2.2	2.5	2.6
Ten or More	0.0	0.4	2.4	3.9	4.0	6.4	8.0	10.5	11.4

[10] 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.

[11] Note that these recently published numbers differ from the FCC's original published reports for these years. Problems in the FCC numbers are discussed in footnote 17 below. Based on information detailed in that footnote, the numbers of "zero service" zip codes in this table appear to continue to be undercounted even with the revisions.

Table 1

Percentage of Zip Codes with High-Speed Lines in Service as of December 31, 2003
(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	9 %	14 %	15 %	17 %	18 %	12 %	7 %	4 %	3 %	1 %	0 %
Alaska	3	32	55	9	1	0	0	0	0	0	0
Arizona	2	5	15	17	7	9	4	5	4	4	28
Arkansas	16	25	25	16	8	3	3	2	1	1	0
California	2	6	11	11	8	4	5	4	5	5	39
Colorado	4	13	18	16	10	4	5	3	4	3	20
Connecticut	0	2	10	13	13	9	8	7	7	7	22
Delaware	0	0	4	16	33	32	16	0	0	0	0
District of Columbia	0	0	7	7	0	4	0	7	7	4	63
Florida	1	2	6	11	13	10	9	8	7	7	27
Georgia	2	7	10	15	14	14	8	4	5	2	18
Hawaii	13	43	37	7	0	0	0	0	0	0	0
Idaho	11	23	27	15	9	13	2	0	0	0	0
Illinois	7	14	19	15	9	6	4	4	2	2	18
Indiana	3	16	20	16	13	8	6	4	3	2	8
Iowa	18	27	19	14	10	7	4	2	1	0	0
Kansas	7	19	21	20	10	7	5	3	4	2	1
Kentucky	15	22	18	15	10	8	7	3	1	0	0
Louisiana	5	16	20	19	16	11	8	3	1	0	0
Maine	12	25	27	17	14	3	2	0	0	0	0
Maryland	2	7	11	11	12	8	11	5	4	3	27
Massachusetts	0	2	13	12	14	11	8	7	6	3	24
Michigan	2	10	17	18	11	9	6	5	3	3	17
Minnesota	14	20	16	11	10	5	4	3	3	2	12
Mississippi	5	14	23	24	16	9	5	4	1	0	0
Missouri	14	23	20	13	6	5	4	3	4	4	3
Montana	19	28	26	15	3	2	2	3	1	0	0
Nebraska	17	25	23	15	11	5	3	1	0	0	0
Nevada	4	28	14	9	14	9	15	5	3	0	0
New Hampshire	1	5	13	16	14	16	10	6	5	4	9
New Jersey	0	3	7	9	13	12	10	8	10	13	15
New Mexico	14	27	26	8	11	4	1	4	4	0	0
New York	2	10	12	16	13	11	7	6	4	3	16
North Carolina	2	8	15	20	19	12	8	5	2	2	9
North Dakota	17	54	23	5	2	1	0	0	0	0	0
Ohio	1	7	14	18	17	12	9	6	3	3	11
Oklahoma	7	22	17	17	9	7	6	7	5	1	0
Oregon	5	10	18	18	17	7	5	2	4	4	11
Pennsylvania	7	14	15	14	11	8	6	6	3	3	13
Puerto Rico	0	8	50	34	8	0	0	0	0	0	0
Rhode Island	0	6	3	6	17	28	17	25	0	0	0
South Carolina	4	13	20	17	14	12	9	7	3	1	0
South Dakota	25	32	25	11	3	3	0	0	0	0	0
Tennessee	3	10	16	16	14	12	8	6	4	3	9
Texas	4	10	14	13	10	8	7	6	5	4	19
Utah	10	21	18	12	5	3	2	1	2	2	24
Vermont	4	14	23	22	14	13	9	1	0	0	0
Virginia	8	15	18	17	13	6	4	4	2	2	12
Washington	5	9	18	16	8	5	5	5	6	4	18
West Virginia	21	32	19	13	8	5	1	0	0	0	0
Wisconsin	3	14	21	18	14	8	8	8	4	2	0
Wyoming	8	25	21	28	5	11	1	0	0	0	0
Nationwide	7 %	15 %	17 %	15 %	11 %	8 %	6 %	4 %	3 %	3 %	11 %

Table 2

Similarly, 26% of U.S. zip codes had only one high-speed provider in December 1999, contrasted with only 15% in December 2003. The District of Columbia leads with the largest share of its zip codes with 10 or more high-speed providers (63% in December 2003), trailed by California (39%), Florida and Maryland (27%), and Utah (24%). (See

Table 2.) The least serviced zip codes were South Dakota (25% of zip codes with no providers, 32% with a single provider), West Virginia (21% with no provider, 32% with just one), Montana (19% and 28%), Nebraska (17% and 25%), Iowa (18% and 27%), and Arkansas (16% and 25%). My home, Texas, is somewhere in the middle of the pack, with 19% of its area codes reporting 10 or more providers, and 4% and 10% of its zip codes, respectively, having zero or one provider.

As the FCC notes in its reports, high speed line provision clearly is 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).[12] 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.

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

Census Data

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).[13] 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.[14] 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 an actual zip code in 2000 has been "looked up" in the FCC public use data zip broadband 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

---

[12] 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.

[13] 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.

[14] 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.

the database I have constructed using this methodology, which consisted of 32038 observations on ZCTAs.[15]

Caveats

Before scrutinizing this data, I must note some limitations that come with it. First, by identifying a 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 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 with this gross definition of broadband!

Second, actual *provision* of a high speed line is different from *availability* of a high speed line. In most instances, it may in fact be true that availability to 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 at least some minimal effort within a geographic area to sell the product. But we cannot exclude *a 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 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 availability is vastly different from other modes of service delivery. In any event, satellite and wireless-based broadband remains a tiny segment of the market overall (though potentially important in isolated rural areas), with 367 thousand high-speed lines out of a national total of 28.2 million in December 2003.[16]

[15] Zip codes for Puerto Rico were dropped from the sample because no establishment data from the economic census was available for Puerto Rican zip codes. Another 42 zip code areas were dropped because they are shown in the 2000 population census as belonging to more than one state. In addition, another 44 population census ZCTAs were dropped because the 1997 economic census establishment data showed those zip codes associated with business addresses in more than one state, even though the population census had assigned that zip code to a single state.

[16] See FCC, 2004, op. cit., Tables 1and 2, p. 6. Less than a fifth of the satellite and wireless lines had a high-speed return, compared with 72 percent of high-speed lines overall. The share of satellite and wireless has been declining steadily—from 1.8 percent of high speed lines in December,1999, to 1.3 percent in December, 2003.

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 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.

An Overview of Broadband Competition

Figure 1 displays our tabulation of high-speed line providers in the zip codes in our ZCTA-based database, for 2000 and 2003. As noted earlier, zip codes with from one to three high-speed connection providers have been lumped into a single category.

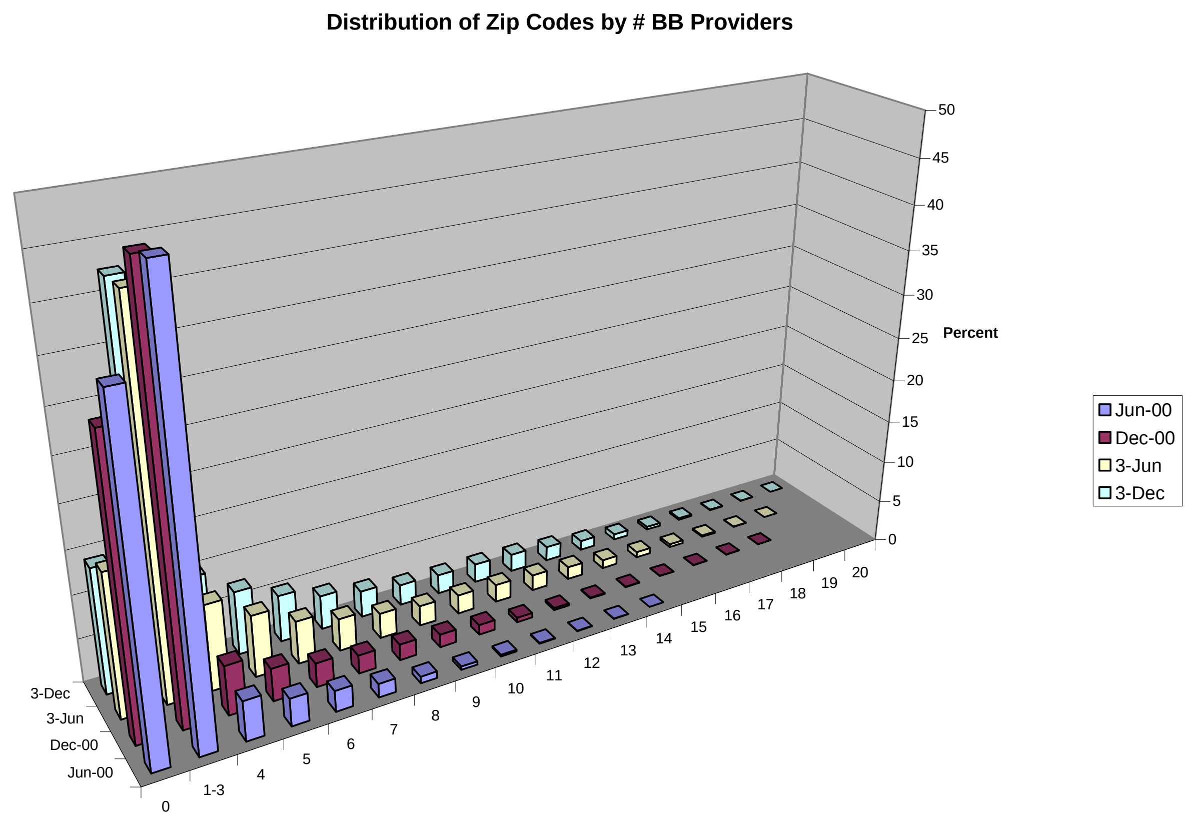

Figure 1

The data show the number of zip codes (identified in the 2000 census) with no high speed service providers declining from 38 percent of the total (for the continental U.S.) in June 2000, to 14 percent of these zip codes in December 2003.[17] Zip codes with 1-3 providers

---

[17] Note that my estimates differ from those appearing in the FCC broadband reports. The FCC now appears to be using a universe of zip codes (i.e., including zip codes in which service providers have **not** reported serving) of under 30,000 zip codes that is substantially smaller than the 32,038 ZCTA "regular" (i.e.,

declined from 48 percent in mid-2000 to 42 percent in late 2003. Zip codes with 10 or more providers accounted for 10.7 percent of the total in 2003, up from .4 percent in mid-2000. Overall, the picture that emerges is one where zip codes "unwired" for broadband have dropped from close to 40 percent at the turn of the century, to about one-third of that figure at the end of 2003. If broadband is the new face of universal service, then the underserved declined greatly, but still account for a visible fraction of the communications landscape.

A rather different picture emerges if we weight the zip codes by the year 2000 population living within, as is done in figure 2. Even in 2000, the almost 40% of zip codes with no high-speed providers accounted for only 6% of the nation's population. Today, those zip codes account for less than one percent of the population. In 2000, zip codes with 1-3 providers had 51 percent of the population, while today less than 12 percent of the population lives in zip codes with such limited competition. To the extent that we are concerned with the availability of broadband, per se, rather than the reasons people may or may not choose to purchase this service, the human dimension of the problem seems substantially smaller, qualitatively, than mere counts of zip codes would appear to indicate. This raises the question of whether programs to make service available, where it does not now exist, would be more effectively and efficiently targeted with rifle-like precision, rather than receiving broad general subsidies.

It is important to remember, though, that we may be missing a significant "quality of service" issue when we frame the discussion in this way. It may well be that our "low quality" definition of broadband (i.e., >200 kbps) is minimizing the real problem, as a rising tide of cheap and increasingly ubiquitous technology raises all boats. 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.

Figures 1 and 2 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

---

excluding artificial codes for unassigned areas and bodies of water) zip codes in which the census estimated people to be resident in 2000 in the 50 states.

For example, the FCC's October 2000 broadband report shows 30.1% of June 2000 zip codes with zero high-speed service providers. See FCC, Industry Analysis Division, Common Carrier Bureau, **High-Speed Services for Internet Access: Subscribership as of June 30, 2000**, October 2000, Table 6. [Note that the FCC later revised this up to 33%; see table 1 from 2003.]

Simply running the census ZCTA 2000 list against the FCC's own June 2000 survey public use file shows 38% of the Census ZCTA codes containing people without high-speed service. My conclusion is that *the FCC's broadband reports substantially undercounted the number of "zero service provider" zip codes in 2000*.

As the number of "true" zero service provider zip codes dropped over time, however, the impact of this bias most likely shrank. The FCC estimates of the fraction of unserved zip codes clearly must be too low, nevertheless, just as their estimates of the share of zip codes being served must be too high. Thus, a significant problem in FCC statistical reports on this subject exists, related to inadequate methods of estimating "zero service" zip codes.

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, pricing), and not availability, may be the real issue in broadband services.

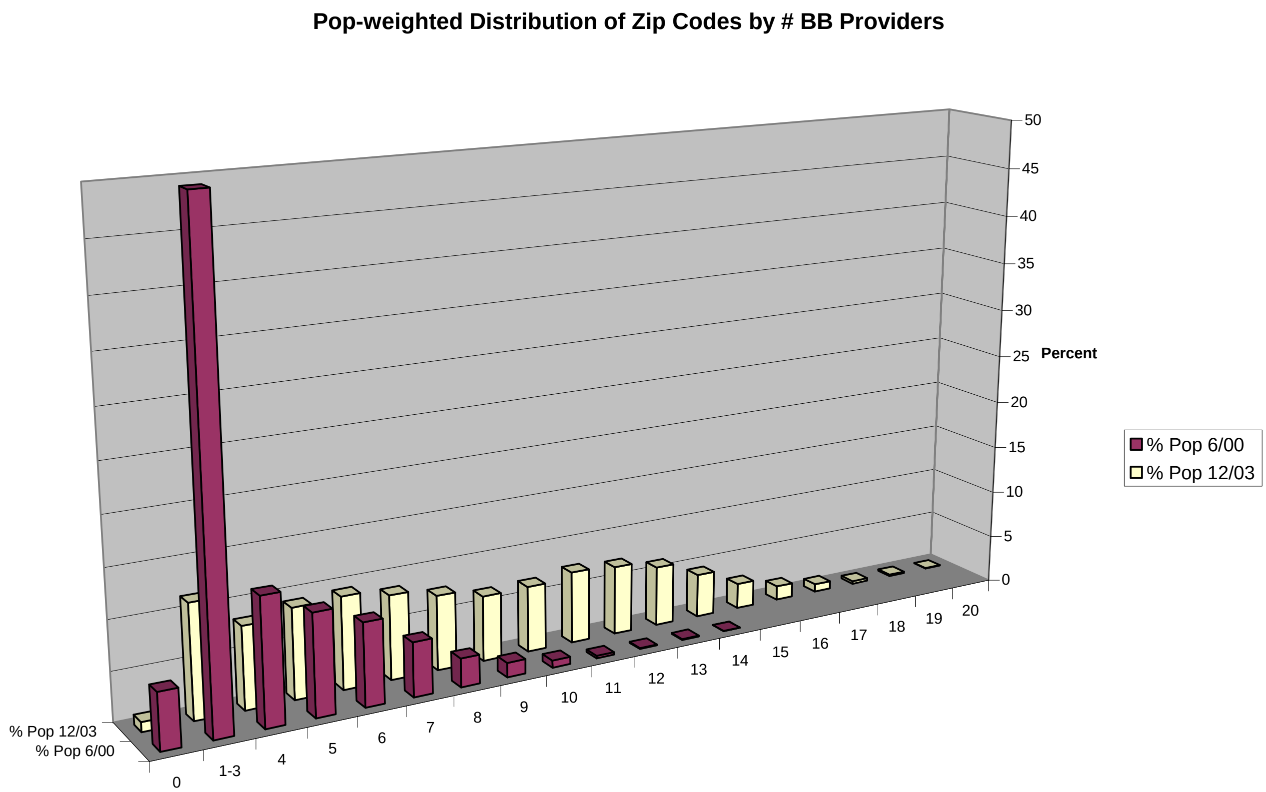

Figure 2

Revaluating the FCC High-Speed Service Area Estimates

In the course of preparing this paper, and constructing the year 2000 census database referenced above, it became clear that the FCC has underestimated the total number of zip codes corresponding to physical residential areas in the United States. As a consequence, the number of zip codes estimated by the FCC to be without service significantly understates the true figure.[18] The calculations underlying this statement are reproduced in the following table:

---

[18] This most likely continued to be the case after 2000, though the calculation is less clear cut, since new zip codes (slowly) get added and old ones dropped over time. But the turnover in "real place" zip codes seems to be quite minimal. If we compare the Census year 2000 residential zip code list (i.e., zip codes listed in the Census 2000 ZCTA list containing data for population and housing in "real" geographic places, excluding "artifical" zip codes for post office boxes and organizations), we find that only 67 such zip codes disappeared between 2000 and 2004. Similarly, if we compare a 2004 zip code list purchased from a private provider with the census 2000 and 1999 lists, and count those that might be real places (i.e., have a FIPS code assigned to them on the 2004 zip code list) and do not show up on either the 1999 or 2000 Census lists, we count 301 new zip codes that are likely to be real places. Thus, there was most likely a net increase of 234 "real place" zip codes over the period 2000-2004.

Revisions to the FCC Estimates of Zip Codes with Zero High-Speed Lines

	Jun-00	Dec-00	1-Jun	1-Dec	2-Jun	2-Dec	3-Jun	3-Dec
Zip Codes w/high speed per FCC (revised	20085	21935	23312	23779	25139	26360	27259	27917
% of zip codes w/zero providers (original)	30.1%	25.0%	22.2%	20.6%	16.1%	12.0%	9.0%	6.8%
% of zip codes w/zero providers (revised)	33.0%	26.8%	22.2%	20.6%	16.1%	12.0%	9.0%	6.8%
Total zips FCC estimate revised (imputed)	29978	29966	29964	29948	29963	29955	29955	29954
Total Census zips (Census/Flamm)	32038	32038	32077	32116	32155	32194	32233	32272
(includes Puerto Rico)								
% zips with zero providers (estimated)	37.3%	31.5%	27.3%	26.0%	21.8%	18.1%	15.4%	13.5%

Source: Author's calculations, as detailed in text.

Table 3

At first glance, it would appear that the FCC's "High Speed Services" reports have significantly *overestimated* both the penetration of high-speed service, geographically, by zip code areas, and the rate of decline in the share of zip codes with no high speed service. The most likely reasons for this are that (1) the FCC has been using a zip code universe based on outdated information from the 1990 Census (the 2000 Census ZCTA estimates were unavailable until relatively recently), and (2) in constructing ZCTAs based on zip codes found in its master address file, the Census Bureau assigned largely rural populations lacking conventional mail service (i.e., receiving mail only through post office boxes and general delivery at post offices) to the zip codes of the post offices through which their mail was accessible. Thus, the Census assigned 3,245 "post office-only" zip codes to "real" populations and places without conventional mail service.[19]

In any event, the operational importance of this overestimate of zip code penetration has little consequence for the U.S. population overall. The share of the *population* living in zip codes with no high speed service providers is a vastly smaller and much more rapidly declining number over this period, so the ideas about trends presented by these zip code area figures is less flawed than might seem at first glance.

Modeling Competition and Entry in Broadband Services

Our main interest is in trying to understand why different numbers of service providers, or no providers at all, provide high-speed services in different zip codes. This is clearly the outcome 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.

---

[19] See U.S. Census Bureau, **Census 2000 ZCTAs ZIP Code Tabulation Areas Technical Documentation**, (Washington: U.S. Census Bureau), 2000, p. 18, available at http://www.census.gov/geo/ZCTA/zcta_tech_doc.pdf. These "PO Box only" zip codes are flagged as such in zip code lists, and may erroneously be discarded with thousands of other "PO Box only" zip codes (for areas with conventional mail service) when zip code lists are being pruned to represent "real" places.

In constructing my model, I have in mind a long-run story about how firms enter the high-speed service 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. In most markets, third party broadband service providers can either compel, through regulatory procedures, or have reached voluntary agreements with, the local cable and telephone monopolies to allow them to invest in interconnects to the incumbents' networks and offer high speed services over these networks after paying a suitable price. 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 think of there being 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 player, 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 an easy question, in theory, given these assumptions. Firm number 1, with the top spot in the profit pecking order, should look at 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, 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 will enter the marketplace. Conversely, if $\prod^*$ is negative, no one will enter and there will be no providers of broadband services.

Conceptually, $\prod^*$ can be thought of as a "reduced form", where profit-maximizing price and quantity have been solved for, and these values then inserted into the expression for profit. $\prod^*$ 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 of 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, "hypothetical maximum profit of the most profitable firm were it to be a monopolist," $\prod^*$, 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 $\prod^* \geq 0$. $\prod^*$ 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;

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

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 models are easier and more intuitive to interpret, however, and we will focus on presenting the logit results, even though we also estimate results from estimation of a probit model of the same reduced form expression for monopoly profit.

How Many Entrants? If we are willing to make some additional assumptions, we can extend this framework to consider how many firms are likely to enter any given market for broadband services. To do so, we must make assumptions about the nature of oligopolistic competition in regional markets for broadband services.

I start by assuming a very simple cost structure, with total cost function for firm i in market j, TC_{ij}, a function of an index of its place in the potential profit line, i, its output, q_i, and a vector of cost variables specific to market j, Z_j, given by

(3) $$TC_{ij}(i,q_i, Z_j)= F(i,Z_j) + v(i,Z_j)\, q_i$$

with $F(i,Z_i)$ its fixed cost to enter, and $v(i,Z_j)$ its marginal unit cost. Note that the fixed costs create economies of scale. As before, I note that the ordering of different types of firms and technologies in terms of costs and potential profitability can itself be dependent on the variables that shift costs and demands in that market.

We start by assuming that the previous question about any entry at all has been answered in the affirmative, and continue by assuming that firms will continue to enter this market as long as the last entrant remains profitable after entry. If firm 1 were to enter, as the monopolist, then profit maximization means it sets a price (suppressing all region subscripts j, since we are considering only a single geographic region) corresponding to the usual markup rule,

(4) $(p-v(1,Z))/p = -1/\eta(p, X)$,

where η is the market price elasticity of demand, and X is a vector of variables that shift demand within a region. Having determined the profit-maximizing price p* and quantity q* as a function of X and Z by solving (2), we can then substitute these into an expression for total profit

(5) $\prod^*(1,Z,X) = [p^*(1,X,Z)-v(1,Z)]\, q^*(1,X,Z) - F(1,Z)$

If $\prod^*$ is greater than zero, the firm should enter the market, otherwise it should not.

Expression (5) is just the reduced form for monopoly profit discussed above, and does not require data on either price or quantity (which we do not have). Using potential monopoly profit (5) as an indicator, or latent variable, for a binary decision to enter or not enter a market leads us very naturally to a logit or probit model of broadband penetration, which we present below. To move on and look at the numbers of firms present in the market, given that entry has occurred, requires further assumptions.

Given that it is profitable for at least most profitable firm to enter the market, and there are profitable opportunities for additional firms to enter the market, how can we model when entry stops? With multiple service providers in a market, we have an oligopoly, and must make additional assumptions about how the oligopolists interact. If we are in a stable free entry equilibrium, moreover, an additional firm will be unprofitable if it chooses to enter the market.

Assume for the moment we have the first N firms in our profit queue operating profitably, and the N+1st firm decides to enter. In the context of broadband, since we have more than one firm, it is probably useful to think about these firms as offering differentiated products, with each firm i offering its own differentiated version of a broadband service product.[20] If it chooses to enter, and Cournot (quantity-taking) assumptions hold, profit maximization means that it should choose a price, and level of output, such that

---

[20] Since prices and quality characteristics of different broadband services typically vary substantially within a given market, it would be unrealistic to posit otherwise.

(6a) $(p_i - v(N+1,Z))/p_i = -\Omega(p_i, q_{-i} X)$,

at a new equilibrium, where q_{-i} is a vector of quantities produced by other firms, which this firm takes as fixed when it makes its own production decisions. Function Ω is the elasticity of inverse demand.[21] Alternatively, it may be more realistic to assume Bertrand (price-taking) assumptions, so that

(6b) $(p_i - v(N+1,Z))/p_i = -1/\eta(p_i, p_{-i} X)$,

at a new equilibrium, where p_{-i} is a vector of prices set by other firms, which this firm takes as fixed when it makes its own production decisions.[22] Given either assumptions (6a) or (6b), we have a system of N+1 equations in N+1 unknowns, and can solve for the p_i's and q_i's as a function of cost shifters Z, demand shifters X, and N+1, the number of firms in the new equilibrium.[23]

Whether this new equilibrium is viable in the long run depends on whether or not the least profitable firm (which we have assumed to be the last and most recent entrant, given our ordering assumptions on entry) makes a profit or not. Let q*(N+1, X, Z) and p*(N+1,X,Z) be the new equilibrium quantity and price for firm N+1 in its new equilibrium. Inserting these into an expression for equilibrium profit, like (5), the new N+1-firm equilibrium will be viable and N+1 firms will remain in the industry if $\prod^*(N+1,Z,X) \geq 0$, and non-negative profits are earned by the last entrant. On the other hand, if $\prod^*(N+1,Z,X)$ is negative, the equilibrium is not sustainable, and a firm will ultimately exit.

Thus, for a long-run equilibrium in which no more than N firms can profitably operate, it must be true that

---

[21] I..e., $\Omega = (q_i / p_i)(\partial P_i / \partial q_i)$, where P_i is the inverse demand curve for firm i's product. For the firm facing a given inverse demand curve, choosing p_i is equivalent to choosing q_i.

[22] I.e., $\eta = (p_i / q_i)/(\partial Q_i / \partial p_i)$, where Q_i is the demand curve for firm i's product. Generally, $\Omega \leq (1/\eta)$, unless all producer's goods are homogeneous, in which case equality holds. See Xavier Vives, **Oligopoly Pricing**, (Cambridge: MIT Press), 1999, pp. 154-160 for a detailed discussion of the Cournot and Bertrand equilibria assumptions.

[23] We assume that a new Nash equilibrium exists and is unique in what follows. Steven T. Berry, "Estimation of a Model of Entry in the Airline Industry," **Econometrica**, vol. 60, no. 4, July 1992, shows that one set of sufficient assumptions for this to be the case include (1) that firm profits decline as more rivals enter, (2) that the profitability ranking does not change if the set of potential entering firms changes, and (3) that differences across firms affect only their fixed costs, and that variable profits therefore are identical across firms. As Berry notes, the last assumption has the effect of making the post-entry equilibrium among firms symmetric.
Alternatively, one could simply assume that firms take turns in deciding whether or not to enter the industry, in order of profitability, and add an explicitly sequential element to the game. In the context of telecommunications and broadband markets, one could make the argument that this latter assumption, in lieu of (3), is a rough description of the historical advantages of incumbency in the construction of telecommunication networks. Both Berry, above, and T. Bresnahan and P. Reiss, "Empirical Models of Discrete Games," **Journal of Econometrics**, vol. 48, 1991, note that using profitability as the order of entry can define a unique equilibrium in models of this sort.

(7) $\prod^*(N,Z,X) \geq 0, \qquad \prod^*(N+1,Z,X) < 0.$

Thus, we can calculate ∏*, the profitability of the last firm to enter the market, for successive values of N, and use this function to determine how many firms, N, can profitably enter any given market. Assuming that function ∏* is continuous and decreasing in N over the relevant empirical ranges for the variables in (7), we can solve for the N* that just sets long run profit equal to zero, as N*=g(Z,X). We can then rewrite the conditions for N being the equilibrium number of firms, (7), as

(8) $N \leq g(Z,X) < N+1.$

In effect, function g gives the value of an unobserved latent variable, which in turn determines the number of firms that can profitably enter a regional market.

The "natural" way to model entry into regional broadband markets, then, will be to use an ordered logit or probit model, where bounds on the value of latent variable g determine how many firms enter a market. This approach does have some down sides relative to the much simpler model described earlier of the binary decision to enter, however. For one thing, function g described by 8 is likely to be highly nonlinear. In addition, we are adding many additional assumptions about the nature of equilibrium in an imperfectly competitive market in order to derive (8).

<u>An Illustration.</u> The easiest way to understand the approach just outlined is to give a simple example of the underlying principles. Suppose equilibrium in a regional market can be described as a symmetric (all firms identical), Cournot (quantity-taking) equilibrium. These assumptions are adopted merely to illustrate the logic and method described by equations (3) through (8) in a particularly simple case, so explicit and relatively simple algebraic expressions could be derived. The underlying framework described above and used in this paper does NOT assume identical competitors.

Assume a differentiated product is produced by each firm, and symmetric demands are described by linear inverse demand functions like

$$P_i = \alpha - \beta q_i - \gamma \sum_{j \neq i} q_j \quad .$$

Let $\sigma = \frac{\gamma}{\beta}$, which is equal to 1 if different firms' products are perfect substitutes, and 0 if they are not substitutable and do not affect each other's market. Costs for firm i are also assumed to be linear, and given by

$$TC_i = F + cq_i .$$

After working through some tedious algebra, it is possible to show that with N firms in the industry, a symmetric equilibrium is characterized by an equilibrium profit ∏* given by

$$\Pi^* = \frac{(\alpha - c)^2}{\beta(2+(N-1)\sigma)^2} - F .$$

Solving for the N* that sets this value to zero, we have

$$N^* = \frac{1}{\sigma}\left(\sqrt{\frac{(\alpha-c)^2}{F\beta}} - 2\right) + 1.$$

The function on the right-hand side of this last equation gives a value for the latent variable that can be used to determine optimal N in this example.

For the moment, however, I shall only worry about the binary enter/don't enter decision, and the simpler logit model. I will briefly return to my more complex model of the number of entrants at the end of this paper.

Data Issues

I have constructed a unique database that joins together six different data sources describing market-related cost and demand variables at the individual zip code level. The components 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. We have discussed this data 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.
2. FCC data on the number of CLECs (competitive local telephone service providers) selling telephone service in competition with the incumbent ILEC, also available at the zip code level.[24] The CLECs may have their own physical local networks, or may be reselling access to the ILEC's network.[25] Since telephone line-based DSL channels are one major form of broadband, the extent to which alternative telephone service providers are available and compete to provide access lines to potential Internet service providers may be expected to increase as the measured number of CLECs increases.
3. Detailed data for individual ZCTAs, 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

---

[24] See http://www.fcc.gov/Bureaus/Common_Carrier/Reports/FCC-State_Link/IAD/lcom0604.pdf for a more extensive discussion of these data and their limitations. As with the high-speed lines survey, at least one end user must receive service for the CLEC to be counted as serving that zip code. The data can be found at http://www.fcc.gov/wcb/iatd/comp.html.

[25] In December 2003, about 23% of the switched access lines provided by CLECs were over their own local loop facilities.

the ZCTA level in the Census SF3 data set.[26] A short summary of these data are also available as a downloadable "2000 U.S. Gazetteer" file.[27]

4. Data on numbers of establishments in ZCTAs at the two digit NAICs industry level, from the 1997 U.S. Economic Census.[28]
5. Data on zip codes in use in 1999 (a data file published by the Census),[29] 2000 (from the Population and Housing Census, and an electronic listing of current census and FIPs codes purchased from zipwise.com in February 2004.[30]
6. Data on commitments of funds by the Universal Service Fund to grants to schools and libraries to support communications and Internet connections, for the years 1999-2001. 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. Grantees' zip codes are available in a public use file, and these data were aggregated by the author to provide totals for every zip code and year.[31]

An extensive effort went into "cleaning" these data and making them consistent across sources. I note that the cleaning process included

- dropping all ZCTAs/zip codes where the Census showed no population living. Typically, most of these cases were zip codes that spanned more than a single state, and the Census apparently chose not to attempt to allocate population in these zip codes across states, although housing often was;
- dropping zip codes listed in the 1997 economic census as business addresses that do not correspond to residential census zip codes listed by the 2000 population census, and therefore may not correspond to a "real" physical, geographic addresses;
- dropping zip codes listed in the 1997 economic census that show businesses with addresses in multiple states, even though the population census may show that same ZCTA as spanning only a single state;
- dropping Puerto Rico from the sample (establishment data from the economic census was unavailable);
- dropping ZCTAs where per capita income was missing, or where median rent, housing value, household income, or family income were zero or missing;
- dropping zip codes from the District of Columbia, where all zip codes remaining after the above cleaning had access to high speed lines, and there were no zip codes without high speed access. This would mean that a dummy variable for the District of Columbia would not be identified, and

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

[27] 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.

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

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

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

[31] The public use data file may 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.

would lead to "quasi-complete separation" of the data (inability to compute a maximum likelihood estimator for an intercept term for DC) were DC to be included in the sample.

From an original sample of 32,081 "real" unique 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), some 30,306 remained after the above cleaning. Other missing variables reduced the number of observations available for model estimation in 2000 to 30,279.

Estimating a Model of Entry

Our initial effort is 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

(9) $\text{Prob}(Ei=0) = F(- X b - Z c)$, 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). Note that our coefficients are *reversed* in sign (something that improves the profitability of broadband for an entrant reduces the probability of no broadband in a region), since we are estimating the probability of *not* having a high speed line, given observed values for X and Z.[32]

The received empirical econometric literature on the subject of what variables are important in determining either broadband supply or costs is relatively small.[33] The FCC

---

[32] This is to preserve greatest comparability to the ordered categorical model we intend to estimate later. All coefficients describe determinants of the probability of a category that does not change (i.e., the probability of no broadband), used as our base, when we proceed to subdivide the broadband category into finer boxes later.

[33] 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

"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 analysis I present below shows that it would be a mistake to assume causality from this evident correlation.

The classes of variables I will include in my logit analysis are (C / D notation indicates whether they likely affect costs or demand):

- Population density, measures of the percent of the population in urban 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]
- 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)
- Numbers of CLECs providing competition for incumbent telephone companies (constructing categorical variables representing 1-3 CLECs, 4-9 CLECS, 10 or more CLECs; no CLECs is baseline) (C)
- Percent of the population in very detailed age groups (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 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 poor, female, living in group quarters, institutionalized (D)
- Occupied housing density, percent houses occupied, percent in crowded housing (D)
- Percent of homes with no telephone (D, possibly C)
- Percent of households with no car, indoor plumbing (D, possibly C as proxy for infrastructure quality)
- Average rent and home value (D)
- Cumulative "eRate" grant value committed to a zip code by the Universal Service Fund for years 1999-2000 (C or D)

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.

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 assumptions about the relationship between local telephone competition and broadband competition; the second is my assumptions about functional form.

Local Telephone Competition. A brief further discussion of the "local telephone competition" variable described above is useful. In the long-run, the number of CLECs entering local markets to compete with ILECs is likely to have a reduced form very similar to the reduced form derived above for broadband service providers, with many, if not all, of the same demand and supply shifters that show up in the reduced form for broadband service provider numbers. Furthermore, state policy may also be an important factor in determining the extent of local telephone competition.

Thus, we can think of two variants of our reduced form equation for number of broadband service providers. In the first, we include variables describing the number of local telephone competitors. In this specification, the state dummy variables **exclude** the impact of any policies affecting local telephone competition. The CLEC competition variable reflects the outcome of a separate subsystem of cost and demand equations. If we assume that broadband does not appear in the CLEC supply and demand equations, then we can take CLEC as predetermined (exogenous) from the standpoint of broadband markets, and we are estimating a "partial" reduced form conditional on the number of CLEC competitors.

In the second variant, we substitute an expression for the number of local telephone competitors, similar to that based on (8) above, to form a "completely" reduced form equation for broadband entrants, a function of all demand and supply shifters appearing in both sets of equations (broadband and CLEC entry).The coefficients of the supply and demand shifters in this completely reduced form reflect both their *direct* impact on broadband profitability, and their *indirect* impact on broadband via local telephone competition. The state dummies now ***include*** the impact of state policies affecting local telephone competition on broadband profitability. Both specifications are valid, but different effects are being identified in coefficients for variables other than those describing numbers of CLECs. In the "completely" reduced form model without CLEC numbers, all variables include their net impact on broadband after factoring in both direct and indirect (through local exchange competition) effects. In the "partial" reduced form model variant where CLEC numbers explicitly control for local exchange competition, other coefficients exclude any indirect impact on CLEC competitor numbers.

In constructing and labeling both of the above two models as "reduced forms," and in ignoring estimation issues related to potential endogeneity of the CLEC competition variables in my statistical analysis of broadband entry, I implicitly assume an asymmetry—that local telephone competition affects broadband provision costs, but that broadband provision has no effect on local telephone voice services competition.. While this may seem like reasonable approximation to reality for the year 2000, it grows increasingly tenuous over time. Currently, voice-over-IP (VOIP) voice communication services delivered by broadband service providers have shown substantial growth, and are beginning to have some impact on local telephone services markets. This was not the case until recently, however, and we can hope that we are justified in making this assumption for the year 2000 in the results discussed below.

Nonetheless, if one is concerned about bias or endogeneity issues related to the potential effect of broadband availability on CLEC voice services competition, the "completely" reduced form is the preferred model variant. The price paid for its use is that both direct and indirect (working through local exchange competition) effects of explanatory variables on broadband competition are being combined in estimated coefficients.

We proceed for the remainder of this paper by using the "partially" reduced form. If one is willing to assume that the recursive causal structure suggested here (CLECs affect broadband, but not the converse) is reasonable, then comparison of the coefficients from the two models gives us information about the likely impact of the explanatory variables on CLEC competition.[34] We should note, however, that even if CLEC entry is predetermined from the standpoint of broadband entry (i.e., broadband entry does not appear as an argument in the equations determining CLEC entry), it is possible for our CLEC variable to be correlated with the error term in our broadband equation if we have omitted variables in the broadband equation that also affect CLEC entry.[35]

Functional Form. In a study of demand for Internet service making use of data on individual households, Chaudhuri, Flamm, and Horrigan (2004) found some evidence that the impact of income increases on Internet demand, while remaining positive, declined at higher levels of income. Initial estimates of the logit and probit regression models based on equations (1) and (2) above use a functional form that included 3 income variables, average rent and house value, average housing age, zip code land area,

---

[34] Indeed, it should be possible (though I do not pursue the idea in this paper) to estimate a two-equation recursive system: one equation giving broadband competitors as a function of a set of variables plus local telephone competition, the other equation giving local telephone competitors as a function of a subset of the same variables (and possibly, broadband competition). This would allow more precise estimates of the separate impacts of all these variables on both CLEC competition and broadband competition, along with estimated standard errors for both sets of effects.

Prieger (2001, see above) takes such an approach in estimating a bivariate probit, binary choice model of entry including both a broadband and CLEC equation. He constructs a test for correlation between the CLEC variable and the error in the broadband equation, and interprets it as indicating that the CLEC variable is endogenous. His results are not completely comparable to mine (putting aside the large differences in the data sets used to estimate these relationships), since his specification excludes a number of statistically significant variables included in my specification, which could lead to apparent correlation between the CLEC variable and the broadband error term. His exogeneity tests also rely on the assumptions that unbundled network element prices (which he in effect uses as an instrument) are exogenous, and do not show up as arguments in the broadband equation, both of which could potentially be questioned. But the underlying issue raised by Prieger's analysis—that CLEC entry may well be an endogenous variable—is certainly a real concern.

[35] This would make CLEC entry look like an endogenous variable (i.e., it would be correlated with the error term in the broadband reduced form). But any such omitted variables would create bias issues for our estimated broadband coefficients quite independently of their possible effect in creating dependencies between the residual error term in this equation and the CLEC entry variable.

population density, number of households, average household size, commute time, occupied housing density, and average population per housing unit as simple linear continuous variables. An alternative version of the model with the natural logarithms of the first 11 of these variables (the last two were maintained in their original untransformed form) substituted for the original untransformed linear values was estimated. Yet another version of the model with the square roots of the 11 variables substituted for the original untransformed linear values was also estimated. The log transformation of these variables clearly did the best.

Figure 3 shows the relative change in response of these transformations to changes in income over a representative range for income variables. The odds of broadband connectivity increase linearly, of course, with a linear term included. The logarithmic transformation substantially flattens the response as income increases beyond a certain minimal level, and the square root transformation produces an outcome intermediate between the linear and logarithmic cases.

The results of estimating the model with the most inclusive set of variables (all of those discussed above) to explain broadband penetration rates among U.S. zip codes in December 2000, and the logarithmic functional form for the continuous variables just discussed, are shown in appendix A. In interpreting coefficients, please remember that it is the probability of **not** having high-speed lines that is being modeled. A probit model was also estimated and (as is usually the case) produces very similar results, but is not shown, since logit coefficients are quite similar, and are much easier to interpret.

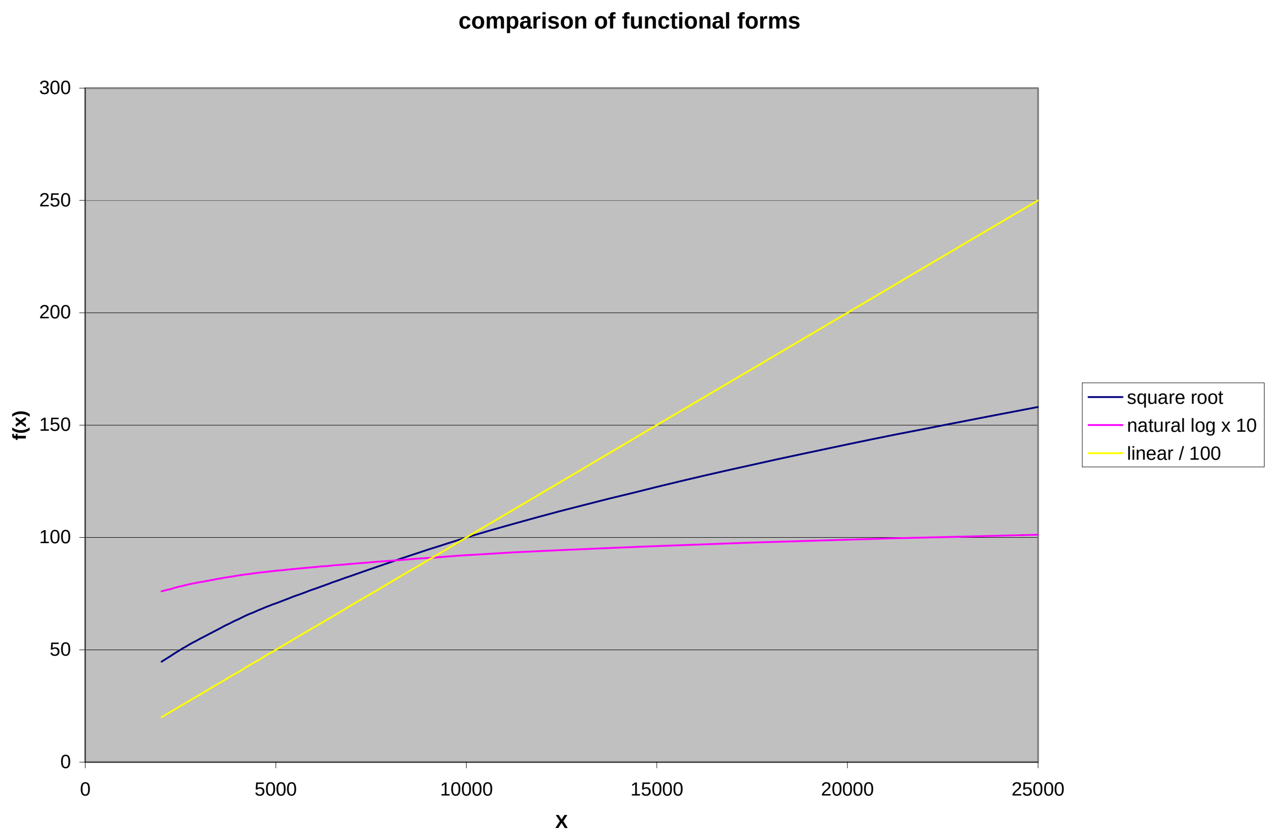

Figure 3

A comparison of various model fit criteria for the most inclusive (all variable) versions of the model with the various functional forms is shown in Table 4. Both the square root and logarithmic transformations of these variables are superior to the simple linear model. By all criteria (Akaike Information Criterion, Schwarz (Bayesian Information) Criterion, simple log likelihood, various pseudo R^2 measures), the logarithmic transformation gives the superior fit. All regressions reported in the remainder of this paper will use this specification.

It should be noted, moreover, that choice of functional form did effect the sign and/or magnitude of some of the estimated coefficients. I shall note where this was the case in my discussion of the results.

Variables other than the state intercepts were responsible for most of the explanatory power of this analysis. When only state dummy variables were used as independent variables in the model, note that the percent of concordant observations plummeted from over 90 percent to about 70 percent, discordant observations jumped from 7 to 26 percent, and pseudo R-squared measures dropped to between .1 and .2, compared to .45 to .65 with a full set of explanatory variables.

Model fit with different transformations of selected continuous variables

Binary logit model	Linear	Square Root	Logarithm
AIC	19755	19156	18998
SC	20878	20279	20121
-2 Log L	19484.8	18886.0	18727.8
-2 Log L with intercept only	37245.5	37245.5	37245.5
McFadden's pseudo r^2	0.4769	0.4929	0.4972
Cox & Snell pseudo r^2	0.4438	0.4547	0.4575
Nagelkerke max-rescaled pseudo r^2	0.627	0.6424	0.6464
Number of Observations	30279	30279	30279
Zip codes without broadband	9235	9235	9235
Number of Variables (including Intercept)	135	135	135
Percent Concordant Observations	92.3	92.6	92.6
Percent Discordant Observations	7.6	7.2	7.1
Percent Tied Observations	0.1	0.2	0.3

Table 4

The fit of this model was quite good. With any of the functional forms considered, over 90 percent of the 30,279 observed outcomes were correctly predicted (with fewer than 8 percent discordant, less than 1 percent tied). A generalized R^2 measure was equal to .63 or higher, a very respectable outcome in a large cross-section.[36]

A formal statistical test of the linear versus the logarithmic forms for this group of ten continuous variables can be constructed by constructing an "artificial model" containing both linear and logarithmic terms, then testing whether separately restricting all linear terms, and all logarithmic terms, respectively, to equal zero can be rejected. We cannot reject the restrictions that imply the logarithmic form is the correct version of the model, but do reject the restrictions that imply the linear form is correct version of the model, at the 5 percent significance level.[37]

---

[36] The generalized R^2 measure calculated here is Nagelkerke's (1991) "Max-rescaled R^2". -2 Log likelihood was reduced to about 53 percent of its intercept-only value by using the dependent variables in addition to an intercept term, and the model overall is highly significant (a chi-squared test based on likelihood ratio, score, or Wald tests rejects the hypothesis that the non-intercept terms are zero at a significance level below .01 percent).

[37] The test results were as follows,using a model containing both linear and logarithmic terms:

	Wald Chi-squared	DF	Pr>Chi-squared
Hypothesis:			
Logarithmic terms =0	739.0500	11	<.0001
Linear terms =0	18.051	11	0.0804

Table 5 shows the complete coefficient estimates from a full version of the model (all variables, logarithmic functional form for 11 of the 13 continuous variables):

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	7.8914	31.0898	0.0644	0.7996
lpopden	1	-0.6018	0.2481	5.8809	0.0153
lland	1	-0.7707	0.2476	9.6883	0.0019
lhhs	1	-0.1155	0.2465	0.2194	0.6395
pophous	1	-0.0398	0.0388	1.0517	0.3051
lhhsiz	1	-0.1284	0.4271	0.0903	0.7638
lat	1	-0.00478	0.0166	0.0825	0.7740
long	1	0.0511	0.0109	21.9646	<.0001
hland	1	0.0616	0.0112	30.4227	<.0001
e31	1	-0.0235	0.00865	7.3874	0.0066
e44	1	-0.00309	0.00474	0.4244	0.5148
e54	1	-0.0195	0.0116	2.8367	0.0921
e56	1	-0.0852	0.0197	18.7906	<.0001
e61	1	0.0316	0.0709	0.1981	0.6563
e62	1	-0.0520	0.0118	19.3081	<.0001
e71	1	-0.0204	0.0236	0.7452	0.3880
e72	1	-0.00058	0.00703	0.0068	0.9341
e81	1	-0.0143	0.0120	1.4229	0.2329
S1	1	-1.1555	0.2172	28.2971	<.0001
S2	1	1.3625	0.8712	2.4460	0.1178
S4	1	-0.4703	0.2890	2.6481	0.1037
S5	1	-0.0211	0.1804	0.0137	0.9067
S6	1	0.0321	0.3576	0.0081	0.9284
S8	1	0.1606	0.2297	0.4890	0.4844
S9	1	-3.6106	0.5754	39.3735	<.0001
S10	1	-4.2113	0.6774	38.6523	<.0001
S12	1	-2.1875	0.2939	55.3845	<.0001
S13	1	-0.6248	0.2454	6.4851	0.0109
S15	1	1.3117	1.1483	1.3047	0.2533
S16	1	-0.1054	0.3499	0.0907	0.7633
S17	1	-1.0200	0.2175	21.9893	<.0001
S18	1	-0.9249	0.2553	13.1248	0.0003
S19	1	0.5458	0.2267	5.7964	0.0161
S20	1	-0.9409	0.1821	26.6890	<.0001
S21	1	-1.0965	0.2543	18.5909	<.0001
S22	1	0.0219	0.1834	0.0143	0.9048
S23	1	-5.0143	0.5172	94.0047	<.0001
S24	1	-4.7794	0.4244	126.8106	<.0001
S25	1	-4.8381	0.5379	80.9045	<.0001
S26	1	-1.5090	0.3058	24.3488	<.0001
S27	1	0.3980	0.2710	2.1563	0.1420
S28	1	-0.1163	0.2055	0.3202	0.5715
S29	1	-0.2800	0.1864	2.2553	0.1332
S30	1	0.0623	0.3310	0.0355	0.8506
S31	1	0.2601	0.2176	1.4290	0.2319
S32	1	1.0121	0.4063	6.2045	0.0127
S33	1	-5.6126	0.5582	101.1156	<.0001
S34	1	-3.4300	0.4571	56.3011	<.0001
S35	1	-0.3405	0.2369	2.0657	0.1506
S36	1	-4.3135	0.4096	110.9242	<.0001
S37	1	-1.9215	0.3145	37.3375	<.0001
S38	1	-0.3556	0.3043	1.3652	0.2426
S39	1	-2.3119	0.2977	60.3080	<.0001
S40	1	0.1941	0.1677	1.3403	0.2470
S41	1	-1.6789	0.4192	16.0430	<.0001
S42	1	-2.6642	0.3523	57.1957	<.0001
S44	1	-4.9535	0.9172	29.1653	<.0001
S45	1	-1.8672	0.3079	36.7705	<.0001
S46	1	-0.0384	0.2674	0.0207	0.8857
S47	1	-1.0591	0.2345	20.4006	<.0001
S49	1	0.0410	0.3108	0.0174	0.8950
S50	1	-5.3320	0.4870	119.8591	<.0001
S51	1	-3.0572	0.3286	86.5376	<.0001

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S53	1	-0.0418	0.4134	0.0102	0.9195
S54	1	-2.9981	0.3067	95.5362	<.0001
S55	1	-0.5928	0.2746	4.6605	0.0309
S56	1	-0.1716	0.3245	0.2796	0.5970
C1	1	-0.8856	0.0498	315.8213	<.0001
c4_9	1	-1.4224	0.1516	88.0368	<.0001
c10plus	1	-2.2334	1.0263	4.7356	0.0295
PctUrban	1	-0.00097	0.000973	0.9879	0.3202
PctOnFarms	1	0.00137	0.00358	0.1458	0.7026
erate00	1	-0.00024	0.000225	1.1246	0.2889
PctAge5_9	1	-0.0143	0.0202	0.5064	0.4767
PctAge10_1	1	-0.0148	0.0192	0.5924	0.4415
PctAge15_1	1	-0.00622	0.0202	0.0944	0.7586
PctAge18_1	1	-0.0305	0.0220	1.9268	0.1651
PctAge20_2	1	-0.0171	0.0202	0.7119	0.3988
PctAge25_3	1	-0.0178	0.0199	0.8060	0.3693
PctAge35_4	1	-0.0122	0.0190	0.4150	0.5194
PctAge45_5	1	-0.0166	0.0183	0.8262	0.3634
PctAge55_5	1	0.000413	0.0197	0.0004	0.9833
PctAge60_6	1	-0.0120	0.0199	0.3621	0.5473
PctAge65_7	1	-0.00307	0.0192	0.0255	0.8732
PctAge75_8	1	-0.0243	0.0201	1.4576	0.2273
PctOver85	1	0.0140	0.0237	0.3482	0.5551
PctMultRac	1	-0.0235	0.0115	4.2029	0.0404
PctBlack1	1	0.00545	0.00198	7.5703	0.0059
PctIndian1	1	0.0102	0.00333	9.2874	0.0023
PctAsian1	1	-0.0206	0.0138	2.2190	0.1363
PctHawnPI1	1	-0.00764	0.0264	0.0837	0.7723
PctOther1	1	-0.0152	0.00764	3.9564	0.0467
PctHispPop	1	0.00505	0.00524	0.9303	0.3348
PctEnglish	1	-0.00681	0.00613	1.2330	0.2668
PctEnglis2	1	-0.0196	0.0121	2.6113	0.1061
PctNoEngli	1	0.0362	0.0215	2.8223	0.0930
lcommut	1	-0.1137	0.0809	1.9740	0.1600
PctSomeHig	1	-0.00508	0.00625	0.6605	0.4164
PctHighSch	1	-0.00353	0.00500	0.4965	0.4811
PctSomeCol	1	-0.00954	0.00530	3.2359	0.0720
PctBachelo	1	-0.00353	0.00693	0.2599	0.6102
PctGradPro	1	-0.00098	0.00848	0.0135	0.9075
PctArmedFo	1	0.1324	0.3027	0.1913	0.6618
PctDisable	1	0.00374	0.00415	0.8128	0.3673
PctCivLabF	1	0.1532	0.3092	0.2454	0.6203
PctUnemplo	1	0.00829	0.00507	2.6779	0.1018
PctNotInLF	1	0.1473	0.3091	0.2273	0.6335
PctManufac	1	0.00238	0.00348	0.4694	0.4932
PctRetailT	1	0.0109	0.00493	4.9184	0.0266
PctEducati	1	0.00619	0.00475	1.7014	0.1921
PctHealthS	1	-0.00170	0.00453	0.1405	0.7078
PctManProf	1	-0.00309	0.00698	0.1961	0.6579
PctService	1	-0.0107	0.00673	2.5061	0.1134
PctSalesOf	1	-0.00709	0.00689	1.0582	0.3036
PctTransOc	1	-0.00269	0.00707	0.1446	0.7037
PctConsOcc	1	-0.00903	0.00675	1.7918	0.1807
lhhinc	1	-0.0120	0.4867	0.0006	0.9803
lfaminc	1	0.3146	0.2696	1.3612	0.2433
lPCI	1	-0.7733	0.4289	3.2501	0.0714
PctPoor	1	0.00162	0.00394	0.1693	0.6807
PctFemale	1	0.00150	0.00627	0.0570	0.8114
PctGQPop	1	0.00277	0.00831	0.1113	0.7387
PctInInsti	1	0.00942	0.00936	1.0124	0.3143
occhdn	1	0.000117	0.000049	5.7873	0.0161
PctOccupie	1	-0.0121	0.00210	33.4154	<.0001
PctCrowded	1	0.0225	0.0121	3.4283	0.0641
PctPlumbin	1	-0.00034	0.00809	0.0018	0.9665
PctNoPhone	1	0.000069	0.00568	0.0001	0.9903
PctNoCars	1	-0.0163	0.00486	11.2228	0.0008
PctAgeUnit	1	-0.0128	0.00465	7.5943	0.0059
PctAgeUn15	1	-0.00355	0.00515	0.4736	0.4913
PctBuiltBe	1	0.00985	0.00467	4.4508	0.0349

lavgage	1	-0.3905	0.2380	2.6928	0.1008
lAvgrent	1	-0.1325	0.0922	2.0637	0.1508
lAvghval	1	-0.5182	0.0737	49.4695	<.0001

Table 5

There are two hypotheses that it is useful to test immediately. The first is that our geographic metrics—latitude, longitude, heartland [absolute degrees from longitude -95 degrees Greenwich]—have zero effect. The second is that our local voice telecommunications competition measures have zero effect.

Strictly speaking, the last restriction is not necessarily a hypothesis; imposing it (and **including** any exogenous variables that only affect the cost and demand equations for voice telephone service provision) would identify a different model (a "fully" reduced form in which a function determining numbers of CLEC providers in terms of the other variables has been substituted for actual numbers of local exchange competitors) even if CLEC competition was a significant determinant of broadband provision cost. However, **not** rejecting the latter hypothesis when CLEC variables **are** included would put into doubt any link between voice telephone line and broadband competition (assuming away any complications having to do with endogeneity of CLEC competition).[38]

Table 6 shows the outcome of undertaking these tests, and estimating a model in which these two distinct sets of constraints are imposed. We resoundingly reject both sets of constraints, at significance levels far below .01%. Both geographic location measures and local telephone competition seem to be statistically significant determinants of broadband competition (again, the validity of this inference for local telephone competition depends crucially on our assumption that CLEC competition is predetermined from the standpoint of broadband entry). We interpret the geographic location variables as capturing "terrain" effects; see the discussion below.

[38] Previous caveats about the potential endogeneity of CLEC competition and its potential to bias our results if CLEC entry is not predetermined from the standpoint of broadband entry apply; the "fully" reduced form is to be preferred if there is doubt on this issue.

Model fit with				
Binary logit model	No Geo Variables		No Telco Competition Variables	
AIC	19039		19345	
SC	20137		20443	
-2 Log L	18775.1		19081.3	
-2 Log L with intercept only	37245.5		37245.5	
McFadden's pseudo r^2	0.4959		0.4877	
Cox & Snell pseudo r^2	0.4567		0.4511	
Nagelkerke max-rescaled pseudo r^2	0.6452		0.6374	
Number of Variables (including Intercept)	132		132	
		Pr > ChiSq w/ 3 df		Pr > ChiSq w/ 3 df
Likelihood Ratio Test Statistic for Constraints	47.3	3.00652E-10	353.5	2.56505E-76
Wald Test for Constraints	46.9017	3.64703E-10	343.7649	3.33987E-74
Percent Concordant Observations	92.5		92.3	
Percent Discordant Observations	7.2		7.4	
Percent Tied Observations	0.3		0.3	

Table 6

Based on these preliminary results, I estimated a more parsimonious model by eliminating variables that were both not statistically significant at the 5 or 10% levels, and had small point estimates of impacts on broadband penetration based on estimated odds ratios. Variables dropped included latitude (longitude and "heartland" remained significant), Hawaiian ethnicity, "Other" ethnicity, Hispanic ethnicity, most variables measuring English-speaking aptitude, all educational level variables, most labor force status variables, most occupational variables, shares of population in poverty, female, disabled, institutionalized, and living in group quarters, and percentages of the population living in housing with no plumbing or phone.

The age variables, which were highly collinear in their original disaggregated form, were reformulated into three groups: percentage of the population in ages 18 to 24, ages 25 to 44, and ages 45 and up. They did not become statistically significant, however. The percentage of population with some college education was the only educational variable that proved significant.

Table 7 shows the results of fitting this more parsimonious model. Not all retained variables were statistically significant (in particular, average household income and family income had modest effects and large standard errors, and many state dummy variables could not be distinguished from zero),

The fit of the model remained quite good. Some 92.5 percent of the 30,000+ observed outcomes continued to be correctly predicted (with 7.2 percent discordant, and .3 percent tied). The generalized R^2 measure clocked in at .645.

Table 7: Parsimonious Model

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	17.6888	1.6308	117.6443	<.0001
lpopden	1	-0.7365	0.0331	493.8782	<.0001
lland	1	-0.8863	0.0275	1038.9818	<.0001
long	1	0.0511	0.0106	23.1833	<.0001
hland	1	0.0605	0.0109	30.9485	<.0001
e31	1	-0.0220	0.00836	6.9449	0.0084
e54	1	-0.0216	0.0113	3.6416	0.0564
e56	1	-0.0922	0.0193	22.8725	<.0001
e62	1	-0.0511	0.0111	21.3559	<.0001
e81	1	-0.0201	0.0113	3.1761	0.0747
S1	1	-1.1554	0.2085	30.7061	<.0001
S2	1	1.1623	0.7851	2.1917	0.1388
S4	1	-0.4419	0.2802	2.4872	0.1148
S5	1	-0.0171	0.1628	0.0110	0.9166
S6	1	-0.00354	0.3449	0.0001	0.9918
S8	1	0.0677	0.2041	0.1099	0.7403
S9	1	-3.6650	0.5429	45.5777	<.0001
S10	1	-4.2425	0.6592	41.4242	<.0001
S12	1	-2.0918	0.2882	52.6661	<.0001
S13	1	-0.6367	0.2381	7.1511	0.0075
S15	1	1.5001	0.9683	2.3999	0.1213
S16	1	-0.1782	0.3025	0.3470	0.5558
S17	1	-1.0488	0.1646	40.5830	<.0001
S18	1	-0.9641	0.2089	21.2935	<.0001
S19	1	0.5422	0.1479	13.4476	0.0002
S20	1	-0.9314	0.1463	40.5277	<.0001
S21	1	-1.1261	0.2210	25.9718	<.0001
S22	1	0.00136	0.1760	0.0001	0.9938
S23	1	-5.0463	0.4620	119.3207	<.0001
S24	1	-4.8658	0.4009	147.3275	<.0001
S25	1	-4.8609	0.5024	93.6215	<.0001
S26	1	-1.5743	0.2206	50.9388	<.0001
S27	1	0.3753	0.1456	6.6455	0.0099
S28	1	-0.0950	0.1949	0.2377	0.6259
S29	1	-0.2876	0.1427	4.0603	0.0439
S30	1	0.00467	0.2528	0.0003	0.9852
S31	1	0.2775	0.1588	3.0539	0.0805
S32	1	1.0185	0.3949	6.6528	0.0099
S33	1	-5.6311	0.5165	118.8707	<.0001
S34	1	-3.5240	0.4296	67.3036	<.0001
S35	1	-0.3748	0.2185	2.9412	0.0863
S36	1	-4.3611	0.3598	146.9352	<.0001
S37	1	-1.9459	0.3006	41.9101	<.0001
S38	1	-0.3599	0.1751	4.2251	0.0398
S39	1	-2.3706	0.2526	88.0943	<.0001
S40	1	0.1724	0.1523	1.2810	0.2577
S41	1	-1.7050	0.3862	19.4942	<.0001
S42	1	-2.7077	0.3100	76.2866	<.0001
S44	1	-4.8957	0.8959	29.8600	<.0001
S45	1	-1.8730	0.2983	39.4217	<.0001
S46	1	-0.0408	0.1798	0.0515	0.8205
S47	1	-1.0544	0.2139	24.3043	<.0001
S49	1	-0.0481	0.2861	0.0282	0.8666

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S50	1	-5.3464	0.4358	150.4819	<.0001
S51	1	-3.0840	0.3065	101.2580	<.0001
S53	1	-0.1107	0.3581	0.0957	0.7571
S54	1	-3.0955	0.2734	128.1802	<.0001
S55	1	-0.6199	0.1728	12.8638	0.0003
S56	1	-0.2510	0.2822	0.7911	0.3738
C1	1	-0.8910	0.0496	322.7211	<.0001
c4_9	1	-1.4270	0.1505	89.8615	<.0001
c10plus	1	-2.2497	1.0261	4.8068	0.0283
PctUrban	1	-0.00127	0.000953	1.7620	0.1844
erate00	1	-0.00020	0.000221	0.7858	0.3754
PctMultRac	1	-0.0273	0.0111	6.0812	0.0137
PctBlack1	1	0.00604	0.00183	10.9098	0.0010
PctIndian1	1	0.0114	0.00273	17.4781	<.0001
PctAsian1	1	-0.0209	0.0123	2.8642	0.0906
PctNoEngli	1	0.0188	0.0152	1.5273	0.2165
lcommut	1	-0.1197	0.0747	2.5657	0.1092
PctSomeCol	1	-0.00903	0.00322	7.8812	0.0050
PctDisable	1	0.00419	0.00375	1.2471	0.2641
PctUnemplo	1	0.00627	0.00473	1.7566	0.1851
PctRetailT	1	0.00785	0.00430	3.3342	0.0679
PctEducati	1	0.00679	0.00398	2.9122	0.0879
PctService	1	-0.00829	0.00342	5.8701	0.0154
lhhinc	1	-0.3338	0.2918	1.3084	0.2527
lfaminc	1	0.3878	0.2583	2.2552	0.1332
lPCI	1	-0.4752	0.1986	5.7246	0.0167
occhdn	1	0.000114	0.000048	5.7871	0.0161
PctOccupie	1	-0.0132	0.00169	61.7179	<.0001
PctCrowded	1	0.0155	0.0109	2.0170	0.1555
PctNoCars	1	-0.0164	0.00451	13.1696	0.0003
PctAgeUnit	1	-0.00646	0.00322	4.0271	0.0448
PctBuiltBe	1	0.00277	0.00178	2.4278	0.1192
lAvghval	1	-0.5124	0.0668	58.8167	<.0001

Using the logit model allows us to interpret these coefficients in terms of odds ratios, in this case, the impact of a one unit change in a variable on the odds of NOT having broadband available (i.e., the probability of no broadband / the probability of broadband). Odds ratios based on these estimated coefficients are shown in Table8.

Population density is statistically significant in all my estimated models, and has the expected sign, but has a relatively small impact. (Note that with the linear functional form for the 11 continuous variables, population density has the *opposite* sign from that suggested by the simple table presented in the FCC high-speed reports- this is one of the variables with an effect that is sensitive to functional form.)

A meaningful estimate of the impact of a change in the population density variable on broadband use can be had by simulating the impact of a 10 percent increase in population density in every zip code in my sample. That is, I increased the population density in every zip code by 10%, then recalculated the probability of having no high-speed provider based on the model coefficients. Summing over zip codes, I thus estimate the expected number of zip codes with no broadband. The 10 percent increase in density results in a -2.3% decline in no-broadband zip codes. The effect is not huge, but it is definitely negative, and statistically significant.

Conditional on population density, zip code size (in square miles) appears to be the best measure of market size. It had a large and statistically significant impact on the probability of no broadband. The simulated impact of a ten percent increase in the physical area of every U.S. zip code would be an almost 3 percent decline in the fraction of zip codes with broadband.

Geographic location is another statistically significant and somewhat surprising effect. Only longitude (long, hland) seems to be significant. Longitude is a negative number in these data (measured in degrees west of Greenwich), and gets more negative as one moves west. The odds ratio estimate for this variable suggests a 5 percent increase in the odds of not having broadband with every degree of eastward (positive) movement in a zip code. My constructed "heartland" measure [absolute value of (longitude less -95 degrees, the approximate longitude of Leavenworth, Kansas, and St. Joseph, Missouri)] is statistically significant, and one degree of movement away from the heartland is associated with a 6 percent increase in the odds of not having broadband.

The net effect is that as one moves west, or toward the "heartland" in the center of the country, the probability of having broadband increases, then decreases as one continues moving west toward the Rockies, then increases as one gets closer to the west coast. The most compelling candidate explanation is that our crude geographic location measures are picking up "terrain effects" here, notably, the Alleghanies and Appalachia in the east, then the Rockies as we move west.[39] Construction of a more direct measure of "ruggedness" of terrain within a zip code would seem worthwhile, and a more refined "terrain" variable is high on the priority list as we continue to analyze these data.

Industry establishment counts (e31, e54, etc.) are statistically significant at the 5 percent level for industries 31-33 (manufacturing), 54 (retail trade), and technical services), 56 (Administrative services and waste management), and 62 (health care and social assistance). NAICS industry 54 (professional, scientific and technical services) and 81 (other services) are marginally significant (at the 10% but not the 5% level). A point estimate of the impact of another establishment in one of these industries within a zip code ranges from a 9 percent reduction in the odds of not having broadband, in administrative services (e56), to a roughly 2% reduction in the odds of not having broadband, in other services (e81). Industry activity within a zip code clearly promotes broadband penetration.

Another dimension of industry presence is in the share of the labor force in a zip code that works in different industries. Although marginally significant (not quite at the 5% level, but significant at the 10% level), two of these coefficients (PctRetailT, PctEduc,) had similar values, increasing the probability of **not** having broadband service. The retail and trade work force may be less skilled than workers in other sectors, and the educational work force may simply have access to broadband at work and be less prone to pay for it at home. These effects are quite small: a point estimate of the impact of a one

[39] Anecdotal stories of entrepreneurs setting up directional 802.11b wireless relays to beam broadband service from rural peaks down to hamlets nestled in isolated valleys below are a staple of broadband chit-chat at telecommunications meetings!

percentage point increase in the portion of the labor force employed in either of these sectors is a .7 to .8 percent increase in the odds of not having a broadband provider.

Service occupations, on the other hand, are a slight plus for broadband. A one percent increase in the share of the labor force in a service occupation translates into a .8 percent decline in the odds of not having broadband.

Table 8 Parsimonious Model

Odds Ratio Estimates

Effect	**Point Estimate**	**95% Wald Confidence Limits**	
lpopden	0.479	0.449	0.511
lland	0.412	0.391	0.435
long	1.052	1.031	1.075
hland	1.062	1.040	1.085
e31	0.978	0.962	0.994
e54	0.979	0.957	1.001
e56	0.912	0.878	0.947
e62	0.950	0.930	0.971
e81	0.980	0.959	1.002
C1	0.410	0.372	0.452
c4_9	0.240	0.179	0.322
c10plus	0.105	0.014	0.788
PctUrban	0.999	0.997	1.001
erate00	1.000	0.999	1.000
PctMultRac	0.973	0.952	0.994
PctBlack1	1.006	1.002	1.010
PctIndian1	1.011	1.006	1.017
PctAsian1	0.979	0.956	1.003
PctNoEngli	1.019	0.989	1.050
lcommut	0.887	0.766	1.027
PctSomeCol	0.991	0.985	0.997
PctDisable	1.004	0.997	1.012
PctUnemplo	1.006	0.997	1.016
PctRetailT	1.008	0.999	1.016
PctEducati	1.007	0.999	1.015
PctService	0.992	0.985	0.998
lhhinc	0.716	0.404	1.269
lfaminc	1.474	0.888	2.445
lPCI	0.622	0.421	0.918
occhdn	1.000	1.000	1.000
PctOccupie	0.987	0.984	0.990
PctCrowded	1.016	0.994	1.037
PctNoCars	0.984	0.975	0.992
PctAgeUnit	0.994	0.987	1.000
PctBuiltBe	1.003	0.999	1.006
lAvghval	0.599	0.526	0.683

Local telephone exchange competition: our CLEC local exchange provider variables are by far the most important single determinant of broadband availability across the U.S. The estimated effects are large and highly significant. Having 1 to 3 CLECs active in a zip code reduces the odds of not seeing broadband provided in a zip code by 59 percent! Estimated standard errors are very small (a 95% confidence interval for this number is 55 to 63 percent). Having 4 to 9 CLECS active reduces the no broadband odds by 76 percent. Having 10 or more CLECs reduces the no broadband odds by 89% (a point estimate, the confidence interval is much wider in this case—1 to 79%).

This effect most likely operates through widening of the availability of telephone (and some cable) access at competitive prices to potential broadband ISPs. Another cut at understanding the impact of competition can be constructed by simulating the impact of having 1 to 3 CLECs operating in all zip codes where there are currently none. Our simulation, based on probabilities computed from this logit model, suggest the impact of this change would be a 21 percent reduction in the share of zip codes in our sample without broadband lines, from 30 to 24 percent of these zip codes.

The signs and magnitudes of other coefficients in this model are relatively insensitive to dropping these Telco competition variables from the model, suggesting that their effect is quite independent of other factors included in the model. Note once again that we have assumed that broadband entry did not stimulate local voice competition in the year 2000; while this assumption may have been reasonable then, it clearly no longer is.

The eRate program: The USF-based eRate program for schools and libraries has been subject to considerable criticism in recent years, principally over accounting and efficiency issues. Our point estimate of its impact is slightly negative, suggesting it encourages broadband penetration, but was not statistically significant. (Note that the program had a very slight but statistically significant impact in improving broadband connectivity with a linear functional form for the continuous variables.) A thousand dollar eRate grant, according to Table 6 (coefficient for erate00, cumulative eRate commitments over 1999-2000), lowers the odds of not having a broadband provider by less than .1 percent.

To get a better handle on this effect, I simulated the effect of giving every zip code in my sample a $100,000 eRate grant. This roughly $3 billion dollar "virtual" program (30,000 zip codes x $100,000) would decrease the expected number of zip codes without broadband by about .6 percent, not a terribly large impact. On the other hand, expanding broadband connectivity to the general public is not the primary objective of this program, and it is perhaps somewhat unfair to judge it by this standard.

Racial, ethnic, and personal characteristics: My results here track other studies of the determinants of Internet demand.[40] Zip codes with greater Afro-American and American Indian populations have statistically significant higher odds of not having access to broadband, while zip codes with larger shares of Asian populations have marginally significant lower odds of not having access to broadband. Zip codes with higher percentages of disabled persons and non-English speakers have coefficients suggesting a slight increase in the probability of not having broadband, but none of these effects are statistically significant.

Cars and commute time: Again tracking the findings of other recent studies of Internet demand,[41] average commute time to work appears to have a positive impact on broadband penetration. The effect is slight however, and not statistically significant. (Here again, a linear functional form produces a slight effect that is statistically significant.) To better understand these magnitudes, I simulated the impact of a 10 percent increase in average commute time in every zip code in my sample. The result was a .4 percent decrease in zip codes with no broadband providers. Car ownership on the other hand has a statistically significant and reasonably large effect: a one percentage

[40] For example, Chaudhuri, Flamm, and Horrigan, 2004.
[41] Ibid.

point increase in the share of households with no car decreases the odds of no broadband by 1.6 percent.

Income, Poverty, and Unemployment: Neither poverty nor unemployment have statistically significant impacts on broadband penetration, though coefficient signs are as might be expected. (Here, once more, things are a little different with a linear functional form: both an increase in the percentage of people below the poverty line and an increase in the share of the unemployed have slight but statistically significant effects in raising the odds of not having broadband.).

The effects of income are statistically significant, but somewhat more complex. I included three income variables: average household income (AvgHHInc), average family income(AvgFamInc), and per capita income(PCI). Only one (PCI) is statistically significant at the one percent level, but the other two have relatively large point estimates (with large standard errors). Household income, like PCI, increases broadband penetration, while family income has the opposite effect. I speculate that per capita and household income are more relevant to demand, while cet. par., greater family income is associated with higher labor compensation costs on the cost side.

In any event, to understand the impact of these variables on broadband penetration, I simulated the effects of a 10 percent increase in all three variables in every zip code in my sample. The result was a 1.3% decrease in the share of zip codes without broadband. The effects of income, on balance, seem to be relatively small, though statistically significant.

Infrastructure Quality. Variables associated with quality of infrastructure seemed to have an important impact on broadband service provision. The share of housing units less than 5 years old (PctAgeUnit) has a significant effect in reducing the odds of no broadband , while the share of units built before 1940 (PctBuiltBe) increased the odds of no broadband. Higher average home values (lAvgHval) also had a significant effect in reducing the odds of no broadband.

State Policy: Our last category of variable is a state dummy, assumed to pick up the effects of state policy on both the costs and demand for broadband. In reality this dummy will pick up any influence unique to an entire state that is not captured by one of my other variables. I posit that state policies are likely to be the most important of these factors.

I have retained all of these dummies in Table 7, and now consider a list of statistically significant ones based on Table 9. Recall that the base case is Texas, so the odds ratios show the impact of that state's policies on the odds of not having broadband relative to the zip code were it instead annexed to Texas (and all other variables kept constant). They are divided into two groups: those where state effects increase the odds of not having broadband, and those where state policies reduce the odds of not having broadband, always relative to the Texas base case.

Statistically significant states where policies are more encouraging to broadband are denoted with a +, states with negative impacts with a -. The numbering system for the dummy variables is based on the state's FIPS code.

Table 9. State Policies

Odds Ratios Relative to Texas

	Effect	Point Estimate	95% Wald Confidence Limits	
+ AL	S1	0.315	0.209	0.474
AK	S2	3.197	0.686	14.895
AZ	S4	0.643	0.371	1.113
AR	S5	0.983	0.715	1.353
CA	S6	0.996	0.507	1.959
CO	S8	1.070	0.717	1.596
+ CT	S9	0.026	0.009	0.074
+ DE	S10	0.014	0.004	0.052
+ FL	S12	0.123	0.070	0.217
+ GA	S13	0.529	0.332	0.844
HI	S15	4.482	0.672	29.906
ID	S16	0.837	0.463	1.514
+ IL	S17	0.350	0.254	0.484
+ 18	S18	0.381	0.253	0.574
- IA	S19	1.720	1.287	2.298
+ KS	S20	0.394	0.296	0.525
+ KY	S21	0.324	0.210	0.500
LA	S22	1.001	0.709	1.414
+ ME	S23	0.006	0.003	0.016
+ MD	S24	0.008	0.004	0.017
+ MA	S25	0.008	0.003	0.021
+ MI	S26	0.207	0.134	0.319
- MN	S27	1.455	1.094	1.936
MS	S28	0.909	0.621	1.332
+ MO	S29	0.750	0.567	0.992
MT	S30	1.005	0.612	1.649
NE	S31	1.320	0.967	1.802
- NV	S32	2.769	1.277	6.004
+ NH	S33	0.004	0.001	0.010
+ NJ	S34	0.029	0.013	0.068
NM	S35	0.687	0.448	1.055
+ NY	S36	0.013	0.006	0.026
+ NC	S37	0.143	0.079	0.257
+ ND	S38	0.698	0.495	0.983
+ OH	S39	0.093	0.057	0.153
OK	S40	1.188	0.881	1.602
+ OR	S41	0.182	0.085	0.387
+ PA	S42	0.067	0.036	0.122
+ RI	S44	0.007	0.001	0.043
+ SC	S45	0.154	0.086	0.276
SD	S46	0.960	0.675	1.366
+ TN	S47	0.348	0.229	0.530
UT	S49	0.953	0.544	1.670
+ VT	S50	0.005	0.002	0.011
+ VA	S51	0.046	0.025	0.083
WA	S53	0.895	0.444	1.806
+ WV	S54	0.045	0.026	0.077
+ WI	S55	0.538	0.383	0.755
WY	S56	0.778	0.447	1.353

Table 9

The results suggest that there is a large group of states where broadband penetration, cet. par., is substantially more pervasive than in Texas. Interestingly, these states are largely on the East Coast. While one might suppose that these results are sensitive to our geography variables, which associate eastward movement with higher odds of no broadband, they are not, surprisingly. Pretty much the same sorts of odds ratios are observed with the geography variables dropped from the model.

One is struck by the fairly large number of states in the south and center of this nation who seem to be doing better than might otherwise be predicted in broadband. Pennsylvania, which is known for an activist state policy designed to promote broadband use, shows up as a much better than average promoter of broadband. Nevada, Minnesota, and Iowa appear to have significantly higher odds of no broadband, compared to Texas, given other characteristics of those states' zip codes. And there is a large group, including California, which appears to be statistically indistinguishable from Texas.

If we are indeed measuring the impacts of state policies with these dummy variables, the magnitudes are striking. Reductions in the odds of no broadband service of 90% or more in zip codes are present in the states at the head of this list.

Finally, it is probably worth noting that policies that the degree of competition in local telephone service markets is a complex mix of federal and state policies. Since we are measuring that quite independently of our state "policy" dummy, we may be stripping one of the most potent dimensions of state policy relevant to broadband use from our "policy" variable.

The Dogs That Did Not Bark

It is worth mentioning that a few factors that are sometimes mentioned as significant in the context of broadband markets did not show up with large or statistically significant effects. Numbers of households, population per housing unit, household size, and age structure had little discernable impact on broadband penetration. For given population density, physical size of zip code appears to be the scale variable relevant to entry.

Next Steps

This very preliminary initial analysis of a rich data set on broadband penetration has yielded some intriguing first results. Some obvious additional directions clearly need to be explored.

An immediate next step would simply be to estimate the ordered logit and probit models (based on equation 8, above), and take advantage of the additional information available on different numbers of providers operating in different zip codes. Indeed, I have already taken a first pass at doing this. Unfortunately, however, the immediate generalization of these models to ordered logit models fail the so-called constant proportional odds test, and probit models fail their conceptual equivalent, the so-called "equal slopes" or "parallel lines" test, and by quite a lot (i.e., the hypothesis of homogeneity of coefficients are rejected at extremely small significance levels). These tests assess whether the coefficients of an ordered logit model change as one moves from one cutpoint to the next, and our first pass at an ordered categorical model suggest they do.

Possible alternatives to be explored in the near future include a generalized ordered logit model, the continuation ratio model, and a partial proportional odds model, all of which relax the assumption that coefficients of the equation determining the value of the latent variable are constant from one cutpoint to the next.

Another issue is the sensitivity of these results to the functional form of the equation. A high priority is to further relax the linearity assumption, and consider interactions between variables as well.

Finally, it will be possible to apply these methods to model broadband penetration by zip code in other years. A random effects approach to estimating a model utilizing a panel of zip code data, rather than a single cross section, should also be possible. Indeed, the most interesting policy questions concern what determines the level of competition, not whether there is any service at all—that question that we have seen is already answered in the affirmative for 99% of the U.S. population.

Preliminary Conclusions

To circle back to the debate over broadband policy with which this paper started, it is now clear that only a small fraction (most likely under one percent) lives in areas where broadband services are not currently being provided. Our attempt to assess some of the factors affecting penetration of broadband into some of these underserved areas has uncovered a significant role for factors that are not much discussed in the current political debate.

At the top of the list is local telephone competition, which seems to be the most important single factor associated with greater broadband availability. Much effort needs to be devoted to disentangling the evolving causal relationships between local voice competition and broadband competition. Our preliminary analysis also suggests that state policies may play an important role.

The two factors most often correlated with broadband penetration, income and population density, paradoxically seem to be among the least important determinants of broadband penetration. Income effects, on balance, are positive, but relatively small. Population density increases broadband penetration, but again, has a relatively small effect, when other variables are taken into account.

The much maligned eRate program does not appear to play a statistically significant role in encouraging broadband use. On the other hand, it was not intended to be a solution to a more general broadband access problem.

Industrial activity seems to have a significant impact on broadband use. Interestingly, car ownership has a significant effect on broadband penetration.

Finally, "digital divide" type ethnic, racial and personal variables show up as small, but statistically perceptible effects. There are reduced odds of broadband provision in zip codes with larger Afro-American and Native American populations.