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

Kenneth Flamm

February 7, 2005

Preliminary Draft
Comments Only Please

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

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

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

Kenneth Flamm

Abstract

This paper constructs a framework for modeling the determinants of broadband penetration in the United States, and applies it to a zip code-level database of economic, demographic, and policy variables constructed by the author. My analysis suggests that state policies may play an important role, and that statistical methods are useful in assessing this role. The ranking of state effects produced by my model seems to correlate with casual impressions of the effectiveness of state policies as portrayed in the press and trade journals. Terrain effects (presumably increasing or decreasing the cost of installing and maintaining a network) seem to be significant in some parts of the country. Two factors often associated with broadband penetration, income and population density, unsurprisingly seem to be among the most important determinants of broadband penetration. The much maligned eRate program does not appear to play a statistically significant role in encouraging broadband use. Industrial activity seems to have a significant impact on local broadband availability. Professional and technical service establishments seem to have the largest such impact. Common perceptions of the effects of gender, education, and rural location on broadband penetration seem to be supported by a causal analysis that attempts to control for confounding factors. Age effects as estimated in this paper do not support the conventional wisdom. Finally, "digital divide" type ethnic, racial and personal variables show up as small, but statistically perceptible effects. There were reduced odds of broadband provision in zip codes with larger Afro-American and Native American populations in 2001, but the gap seems to be closing for Afro-Americans.

.

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.

Common threads running through discussions of broadband policy 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,[4] but does muster evidence that has some relevance to the second point. My analysis will also examine whether substantial variation across states in state-specific factors, including regulatory policies, may be having an impact on broadband

---

[1] I thank Anindya Chaudhuri and the members of my Policy Research Project for their invaluable research assistance in gathering the Universal Service Fund data used in this analysis, 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] 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.

deployment. My approach will be to utilize detailed public use data available on broadband deployment at the individual zip code level from the FCC, add to it economic and demographic data from the 2000 population census and 1997 economic census, data on "erate" and rural health care Universal Service Fund grants, hydrological and terrain data from geophysical data bases, then use this data to estimate the parameters of a reduced form reduced form +economic model of entry into broadband service markets.

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.[5]

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. In doing so, the FCC makes certain assumptions that have a significant effect on the numbers of "zero" service zip codes implicitly estimated to exist within its statistics.[6]

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

Thus, only geographic zip codes show up in the universe of zip codes that the FCC uses in its reports—it is likely that significant amounts of sparsely populated territory with no regular mail carrier service are not included within the boundaries of the geographic zip codes that are caught in this net. Any people or services associated with these "point" zips are reassigned to the nearest "geographic" zip, whether or not they actually live or operate within the boundaries of the mail delivery area defining the geographic zip. If a

---

[5] 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.

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

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

Table 1 shows aggregate U.S. data on "geographic" zip codes in which differing numbers of broadband service providers were available. Note that in December of 1999, over 40% of U.S. zip codes had no providers of high-speed lines; in December 2003, less than 7% of U.S. zip codes had no reporting high-speed line providers.[7]

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

Table 1

[7] 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.

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. geographic 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).[8] 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).[9] A very important point to make is that unlike the FCC, Census procedures map out the areas not served by regular mail service routes that receive their mail through "point" zip codes, and assign these physical areas to their point zip codes.[10]

Thus, the least problematic way to link FCC broadband availability by zip, to Census data for zip codes (ZCTAs), is to restrict the universe to "geographic" zip codes showing up in both the Census and FCC zip code pools. The FCC's practice of attributing point zip code service to nearby geographic zip is unlikely to create many "false positives" for any broadband availability at all (since it is probably pretty rare—but certainly not impossible—for a sparsely settled rural area with no mail delivery to be served with

---

[8] 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.

[9] 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.

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

broadband while a nearby more populated area is not). On the other hand, the FCC data may well overestimate the number of providers of broadband for ZCTAs corresponding to "geographic" zip codes to which "point" zip codes have been assigned. Restricting our attention to only those Census ZCTAs linked to geographic zip codes, then, we are likely to come up with a decent approximation to which of these have any service at all, but our figures on number of providers per geographic zip are more likely to be contaminated with noise from point zip broadband addresses that are assigned to their closest neighboring geographic zip.

A limited amount of data (principally establishment numbers, by two digit NAICS code) from the 1997 economic census are also available at the zip code level.[11] 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, less those ZCTAs making use of zip codes shown as "point" zip codes in the Census's November 1999 master list of postal zip codes,[12] 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 the database I have constructed using this methodology.

Other Data

Additionally, I have gathered data on additional, potentially relevant variables available at the zip code level. From the Universal Service Administrative Company, I have collected information on individual eRate (schools and libraries) and rural health care grants funded out of the Universal Service Fund, by zip code of the organization receiving the grant.[13] A small number (94) of zip code entry errors in the more than 600,000 funded projects were corrected, and the data aggregated up to a funding year sum for an entire zip code. Committed funds (through fall 2004) for each of the funding years from 1998 to 2004 for the eRate and rural health care programs, as were authorized disbursements submitted to USAC for the eRate program only.

Physical topographic, land cover, and meteorological data available from the International Satellite Land-Surface Climatology Project, Initiative II, for half degree squares covering the earth's surface was downloaded and matched to zip codes.[14] Every zip code was assigned values for the half degree cell (roughly 50 km by 50 km) in which the latitude and longitude coordinates of its centroid were located. Data collected included mean and standard deviation of slope (maximum change in elevations between every 1km square cell and its eight neighbors within the half degree cell), mean and standard deviation of the Compound Topographic Index (often referred to as the

[11] 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.
[12] This is available at http://www.census.gov/geo/www/tiger/zip1999.html.
[13] The data and other information are available at www.universalservice.org.
[14] The data may be found at http://islscp2.sesda.com/ISLSCP2_1/html_pages/data_scale.html.

"wetness" index), range between minimum and maximum elevations with the half degree cell, and the predominant MODIS[15] land cover type code for every half degree cell.

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]

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

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

[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.

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 FCC high-speed line providers in both geographic and point 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.

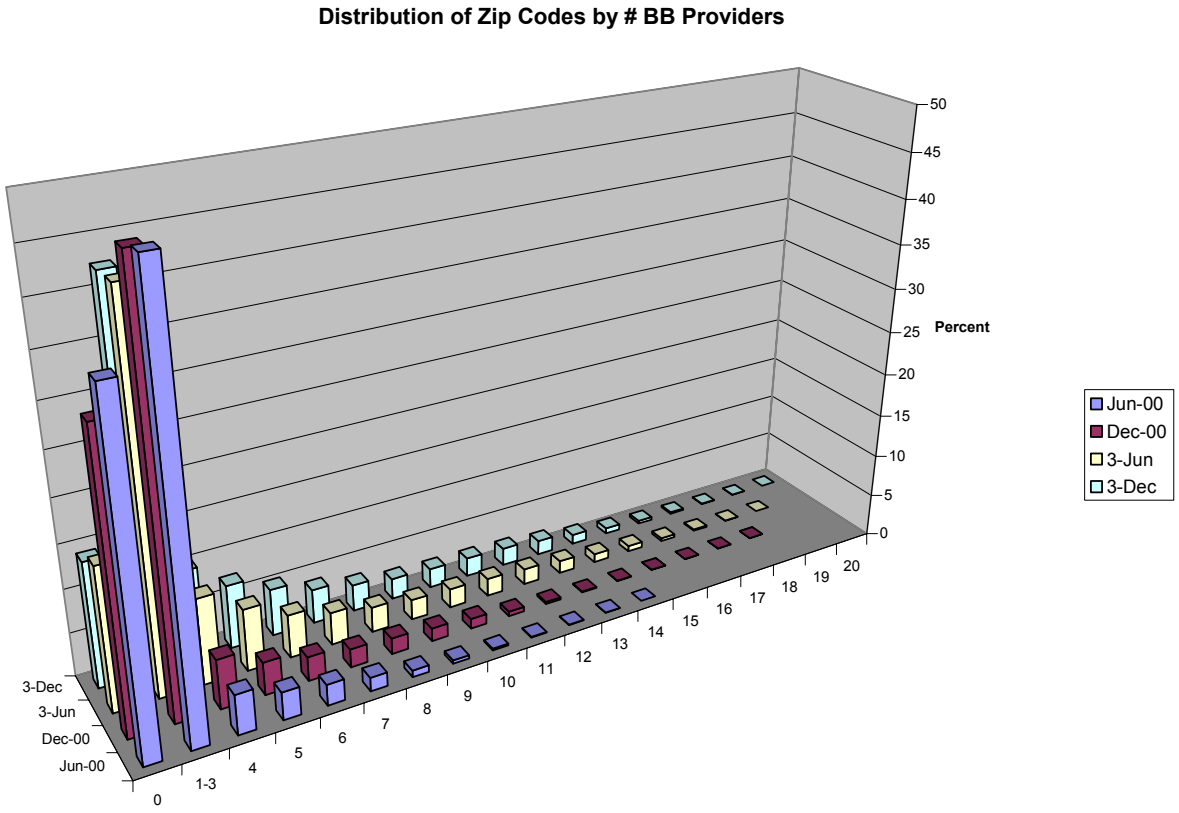

Figure 1

The data show the number of zip codes (ZCTAs 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] Note that this figure should be considered an upper bound on "zero service" ZCTAS, since, as just remarked, broadband providers serving ZCTAs corresponding to point zip codes may have been assigned by the FCC to the closest geographic zip code. Zip codes with 1-3 providers 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

[17] This calculation is discussed in Kenneth Flamm," The Determinants of Broadband Competition: Economics, Demographics, and State Policy," presented at TPRC, Washington, D.C., September 2004.

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 Census zip codes not showing up in the FCC broadband-provisioned zip code list 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 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.

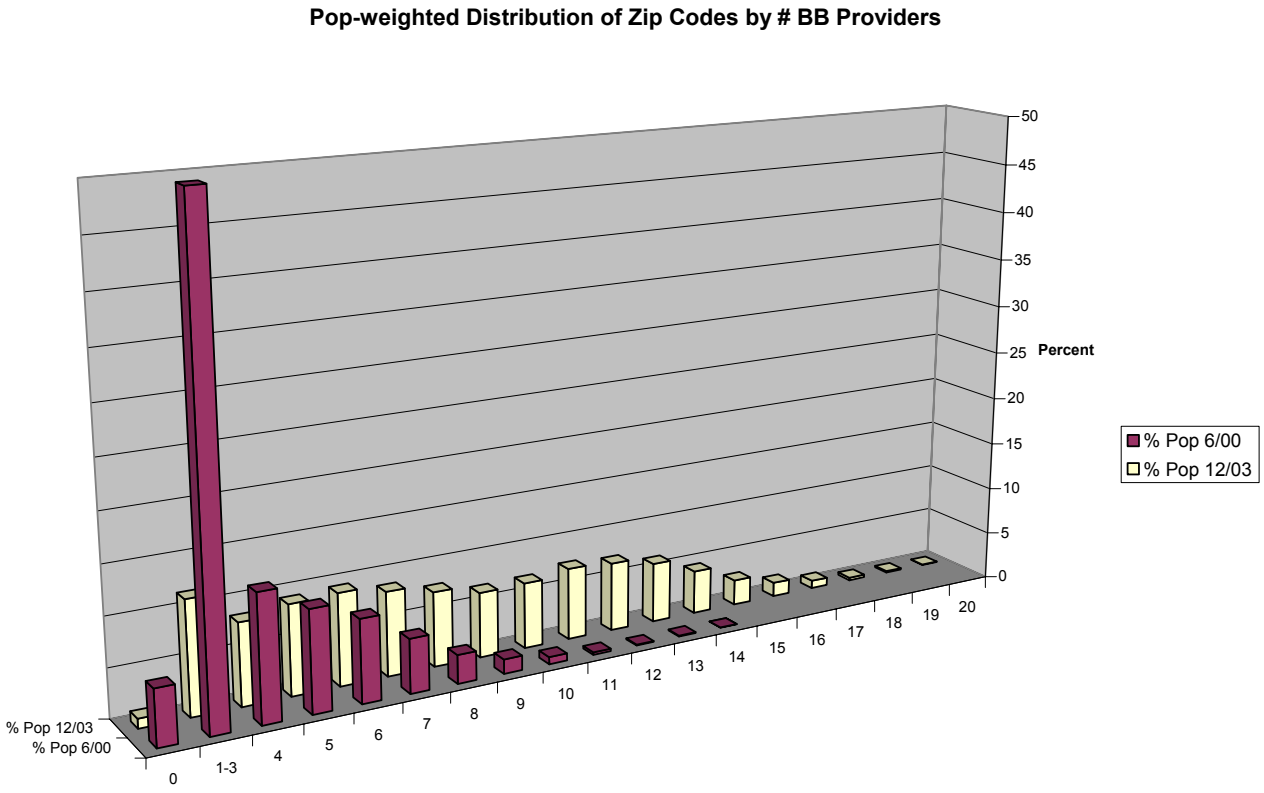

Figure 2

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.

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 $E=1$ if $\prod^* \geq 0$,
$E=0$ if $\prod^* < 0$.

Given observed data on X, Z, and the entry decisions of firms, we can estimate the function X b + Z c and use our coefficient estimates to evaluate the impact of changes in the X and Z variables on the probability that a firm will enter into a market. If we assume ε follows a logistic distribution, we have the logit model; if ε is distributed normally, we have the probit model. The logistic and normal distributions are very similar, and in practice, logit and probit models typically yield very similar results. Coefficients in logit 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.[18] 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

(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.[19] Alternatively, it may be more realistic to assume Bertrand (price-taking) assumptions, so that

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

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

[19] 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.

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.[20] 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.[21]

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

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

Thus, we can calculate $\prod^*$, 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 $\prod^*$ 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) $N \leq g(Z,X) < N+1.$

---

[20] 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.

[21] 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.

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 $\prod^*$ 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 seven 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.[22] The CLECs may have their own physical local networks, or may be reselling access to the ILEC's network.[23] 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 the ZCTA level in the Census SF3 data set.[24] A short summary of these data are also available as a downloadable "2000 U.S. Gazetteer" file.[25]
4. Data on numbers of establishments in ZCTAs at the two digit NAICs industry level, from the 1997 U.S. Economic Census.[26]
5. Data on zip codes in use in 1999 (a data file published by the Census),[27] 2000 (from the Population and Housing Census, and an electronic listing of current census and FIPs codes purchased from zipwise.com in February 2004.[28]
6. Data on commitments of funds by the Universal Service Fund to funded grants to schools and libraries to support communications and Internet connections, and to

---

[22] 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.

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

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

[25] 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.

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

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

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

rural health care service providers, for the years 1999-2004, as discussed above. Fund "commitments" are the stage prior to disbursement, so these data represent likely spending on Internet connections for schools and libraries in the several years after their commitment. Data were also available on authorized disbursements for the schools and libraries ("erate") program, but not for the rural health care program.[29]

7. The ISLSCPI II hydrological, topographic, and land cover data discussed above.

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 all ZCTAs not linked to "geographic" zip codes included by the FCC in its universe (discussed above);
- 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 and Delaware, 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 dummy variable for these areas would not be identified, and would lead to "quasi-complete separation" of the data (inability to compute a maximum likelihood estimator for an intercept term for DC and Delaware) were they 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 to 30,279 The

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

intersection of this set with the Census November 1999 list of "geographic" zips, and dropping DC and Delaware, gives us 27,739 data points.

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

where F is assumed to be the cumulative density function for either the logistic (logit) or normal (probit) distribution, depending on the assumption about the error term in (1).

The received empirical econometric literature on the subject of what variables are important in determining either broadband supply or costs is relatively small.[30] The FCC "high-speed" line reports, referenced above, typically provide simple tables showing that greater broadband penetration in zip codes seems correlated positively with both per capita income and population density. The 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)

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

- 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)
- 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" and rural health care grant value committed to a zip code by the Universal Service Fund for years 1999-2004, (C or D)
- Geophysical and hydrological data-- mean and standard deviation of slope, range of elevation, and composite topographic ("wetness") index for ½ degree square areas containing the latitude and longitude coordinates of the centroid of a Census ZCTA; MODIS land cover classification (type of vegetation/physical land cover). (C)

Specification Issues

Before turning to actual empirical results, three further issues related to the specification of the empirical model need be discussed. The first of these is my use of geophysical and hydrological variables , the second is my assumptions about functional form, the third is a discussion of possible use of data measuring the extent of local telephone competition.

Geophysical and hydrological data. A very preliminary version of this paper, working with data for December 2000, based on casual observation of geographic patterns of Internet use in earlier research on broadband use, experimented with use of nonlinear

functions of longitude as explanatory variables in estimating the reduced form described above. Much to my chagrin, these variables were statistically significant, no matter how many additional variables were added to the equation.

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

Accordingly, geophysical and hydrological data corresponding to a ½ degree grid of the earth were compiled and used to assign values for terrain variables to zip codes, based on the ½ degree square in which a zip code's centroid coordinates were located (see discussion above).

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

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

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

Local Telephone Competition. FCC data is also available on numbers of CLECs providing competition for incumbent telephone companies, by zip code. Competition in local voice telephone services might be expected to affect the costs of providing DSL-based high speed data services, and therefore the costs of broadband provision. In the long-run, 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

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

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.

One might wish to ignore potential endogeneity of the CLEC competition variables in an econometric analysis of broadband entry, and estimate the "partial" reduced form model. To do so without bias would 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 assuming this kind of causal structure might have seemed an almost reasonable approximation back in 2000, it grew increasingly tenuous over time. More recently, voice-over-IP (VOIP) voice communication services delivered by broadband service providers have shown substantial growth, and are beginning to have some serious impact on local telephone services markets. Broadband clearly is now affecting voice services competition!

We proceed for the remainder of this paper by using the standard "completely" reduced form. If one is willing to assume that the recursive causal structure suggested here (CLECs affect broadband, but not the converse) is reasonable, or if one is blessed with relevant exogenous variables available as instruments, then comparison of the coefficients from the two models might give us information about the likely impact of the

explanatory variables on CLEC competition.[32] We should note, however, that even if CLEC entry were 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 a CLEC variable to be correlated with the error term (producing bias) in our broadband equation if we have omitted variables in the broadband equation that also affect CLEC entry.[33]

Initial Estimates

I initially estimated a full (all available variables) version of the binary logit model based on equation (9) (all variables, logarithmic functional form for continuous variables), estimated separately for December 2000, December 2001, December 2002, and December 2003. The probability of any broadband provision at all was modeled.

State dummy variables were denoted as Sx, where x is the numeric FIPs code for that state. Texas (S48, FIPs=48) is the excluded state dummy incorporated into the intercept term. MODIS land cover variable Mx takes on value 1 if MODIS land cover type x is the predominant vegetation type in the half degree square in which a zip code's centroid is located. (M0, for example, denotes MODIS land cover type 0—water bodies, including oceans, seas, lakes, reservoirs, and rivers.) Note that the "eRate" and rural health care grant variable used in these models is cumulative commitments for the *prior* grant years, i.e., through grant year 1999 in December 2000, through 2000 in December 2001, etc.

The logit equations have considerable explanatory power: a generalized R-square measure (max re-scaled R-square) is .60 in 2000, .56 in 2000, .48 in 2002, .42 in 2003. In a model with just an intercept and state dummy variables, the equivalent generalized R-square measures are .23, .19, .15, and .11, respectively.

---

[32] Indeed, it should be possible (though I do not pursue the idea in this paper) to estimate a two-equation 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 direct and indirect 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 a real concern.

[33] 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.

From these preliminary logit results, a list that include all variables that were at least marginally statistically significant (i.e., reject the hypothesis of equality with zero at the 10% confidence level) in *any* of these years, plus all state dummy variables, was then compiled. This common variable list was then used to define regressors in separate binary logits re-estimated for December 2001, 2002, and 2003. From each of these separate logit equations, in turn, a list of 20 variables (including state dummies), that were not marginally significant in any of these three years was compiled. A formal Wald test of the hypothesis that these coefficients were zero could not be rejected in any of these 3 years.[34]

Imposing these restrictions produced the binary (broadband-no broadband) logits for these 3 years shown in appendix A. This same variable set was used to estimate ordered logit models corresponding to equation (8) above. As noted earlier, substantially more demanding theoretical assumptions must be made to justify this model; furthermore, the FCC data are probably a substantially more noisy measure of the unobserved left-hand-side outcome (because of the FCC practice of assigning "point" zip code competitors to the nearest "geographic" zip). In all cases, the score test for the proportional odds assumption[35] (which must be maintained in an ordered logit model) leads us to reject that assumption.[36] To model numbers of competitors properly, I conclude, we must move to a more complex model—partial proportional odds, or continuation ratio models are two attractive alternatives, but we defer that effort to another day.

Instead, I will try to take advantage of the fact that we have repeated observations over time on individual zip codes to better model possible within-zip code effects that cannot be modeled in a single cross section of zip codes. My approach is to estimate a so-called marginal model, i.e., one that estimates the population mean response conditional on a set of independent variables. The mean probability of high speed availability in zip code j, hj, conditional on observed vector of covariates X will be assumed to be given by the logit function

$$(9) \qquad \log (hj/(1 - hj)) = X'\beta,$$

with the variance of broadband availability around conditional mean hj some known function of this mean, and within-zip-codes association of broadband availability over time depends on some fixed set of association parameters and the mean. Equation (9) also follows from equations (1) and (2) above; in that case, however, precise distributional assumptions about an error term were made, and maximum likelihood methods used to produce an estimator whose properties depended on those statistical assumptions.

Unlike the standard logit model defined by equations (1) and (2) above, my marginal model avoids precise distributional assumptions about the broadband availability

[34] The Wald chi-square statistics (with 20 degrees of freedom) were 16.1 in 2001, 14.0 in 2002, and 18.6 in 2003, so the hypothesis could not be rejected. (The smallest p value for the relevant chi-square was .55.)
[35] This is just the assumption that coefficients are constant from one cut point to the next. In a probit model, the very same assumption is generally labeled the "parallel lines" assumption.
[36] The chi-square statistics for these score tests are 2406 for 2001, 2665 for 2002, and 2716 for 2003, all with 218 degrees of freedom. The p-values for these tests are all less than .0001, and we decisively reject proportional odds.

variable. I employ the method of generalized estimating equations (GEE) to estimate the parameters of (9). To the extent that I correctly specify the variance and within-zip-codes association structure over time, my estimates will be efficient. But even if I approximate the variance and association structure poorly, the GEE estimator will be consistent, and a robust "sandwich" covariance estimator can be constructed that will provide asymptotically correct standard errors. I trade off efficiency for robustness, and the ability to make use of within-zip-code information to better model my longitudinal panel data.

Zip code data for 2001 and 2002 were combined in a single data set, and GEE used to estimate (9) using this two year panel. There is of course a real possibility that coefficients may change over time—as do technology and policy in this arena—so my first step was to specify a totally general model that permitted all coefficients to change from one period to the next. Starting with the specification shown in appendix B, I estimated logit model (9) using GEE.[37] The results were:

Table 3: Inclusive specification

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-104.763	28.8772	-161.362	-48.1651	-3.63	0.0003
lpopden	0.8054	0.0765	0.6556	0.9553	10.53	<.0001
lland	0.9325	0.0711	0.7931	1.0718	13.11	<.0001
lpophous	-0.3450	0.2656	-0.8656	0.1756	-1.30	0.1940
long	0.0086	0.0063	-0.0038	0.0210	1.36	0.1739
hland	0.0057	0.0080	-0.0100	0.0215	0.71	0.4761
lslopemn	-0.0895	0.0865	-0.2590	0.0799	-1.04	0.3004
lslopesd	0.1123	0.1103	-0.1039	0.3286	1.02	0.3087
lelevrang	0.0406	0.0884	-0.1327	0.2139	0.46	0.6459
lctimn	0.1614	0.1013	-0.0371	0.3599	1.59	0.1110
M1	-0.2887	0.1330	-0.5493	-0.0280	-2.17	0.0300
M4	-0.1574	0.0873	-0.3286	0.0137	-1.80	0.0714
M5	-0.1581	0.1317	-0.4162	0.0999	-1.20	0.2297
M7	-0.2527	0.4200	-1.0759	0.5706	-0.60	0.5475
M11	-5.0239	0.7100	-6.4155	-3.6323	-7.08	<.0001
e31	0.0672	0.0396	-0.0104	0.1449	1.70	0.0895
e44	0.0282	0.0130	0.0028	0.0536	2.18	0.0296
e54	0.0986	0.0383	0.0234	0.1737	2.57	0.0101
e56	0.0503	0.0720	-0.0908	0.1914	0.70	0.4846
e72	0.0744	0.0243	0.0269	0.1220	3.07	0.0021
e81	0.0459	0.0277	-0.0084	0.1003	1.66	0.0974
S1	0.0373	0.2125	-0.3792	0.4538	0.18	0.8607
S5	-0.0430	0.1663	-0.3691	0.2830	-0.26	0.7959
S6	0.7130	0.3333	0.0597	1.3662	2.14	0.0324
S8	0.6086	0.2902	0.0398	1.1775	2.10	0.0360
S9	0.3462	0.7124	-1.0501	1.7425	0.49	0.6270
S12	1.3488	0.5628	0.2457	2.4520	2.40	0.0166
S15	2.2078	2.0021	-1.7162	6.1318	1.10	0.2701
S16	-0.3940	0.2903	-0.9630	0.1750	-1.36	0.1747
S17	-0.3175	0.1415	-0.5947	-0.0402	-2.24	0.0248
S18	-0.6746	0.1927	-1.0523	-0.2970	-3.50	0.0005

[37] I assumed a diagonal (independent), within-zip working correlation structure over time, which undoubtedly is not correct, and therefore sacrificed some efficiency for simplicity in producing my GEE estimates, which nonetheless are consistent and have correct standard errors. The software I used (Proc GENMOD in SAS is quite fragile and not very robust in estimating large, complex models (convergence is often an issue).

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
S19		-0.9422	0.1478	-1.2318	-0.6526	-6.38	<.0001
S20		-0.4449	0.1589	-0.7563	-0.1334	-2.80	0.0051
S21		-0.2226	0.1741	-0.5639	0.1186	-1.28	0.2010
S22		0.2259	0.2227	-0.2106	0.6624	1.01	0.3104
S23		0.0541	0.2622	-0.4598	0.5680	0.21	0.8367
S24		1.0392	0.3936	0.2678	1.8107	2.64	0.0083
S25		1.8742	0.8794	0.1506	3.5978	2.13	0.0331
S26		0.0527	0.2153	-0.3692	0.4746	0.24	0.8066
S27		-1.0322	0.1530	-1.3321	-0.7323	-6.75	<.0001
S28		-0.2207	0.2146	-0.6413	0.1999	-1.03	0.3037
S29		-0.2528	0.1386	-0.5246	0.0189	-1.82	0.0682
S30		0.0440	0.2349	-0.4164	0.5043	0.19	0.8515
S31		-0.5959	0.1620	-0.9134	-0.2783	-3.68	0.0002
S32		-1.8257	0.4996	-2.8050	-0.8465	-3.65	0.0003
S35		-0.1742	0.2686	-0.7007	0.3523	-0.65	0.5166
S36		0.8456	0.2179	0.4186	1.2726	3.88	0.0001
S37		0.4882	0.2393	0.0191	0.9572	2.04	0.0414
S38		-0.6803	0.1779	-1.0289	-0.3317	-3.82	0.0001
S39		0.2576	0.2301	-0.1934	0.7085	1.12	0.2630
S41		0.6329	0.3534	-0.0597	1.3255	1.79	0.0733
S42		-0.8289	0.1663	-1.1548	-0.5030	-4.98	<.0001
S46		-0.3148	0.1808	-0.6691	0.0394	-1.74	0.0816
S47		0.8536	0.2622	0.3397	1.3674	3.26	0.0011
S49		-0.7742	0.3203	-1.4020	-0.1463	-2.42	0.0157
S51		-0.4127	0.1799	-0.7654	-0.0601	-2.29	0.0218
S53		0.2272	0.3371	-0.4336	0.8880	0.67	0.5004
S54		0.0798	0.1825	-0.2780	0.4375	0.44	0.6621
S55		-0.6795	0.1882	-1.0484	-0.3105	-3.61	0.0003
PctUrban		0.0011	0.0020	-0.0028	0.0050	0.57	0.5680
PctOnFarms		-0.0088	0.0036	-0.0158	-0.0018	-2.45	0.0142
cerate		0.0259	0.0409	-0.0542	0.1060	0.63	0.5264
crhc		0.0272	0.7957	-1.5323	1.5867	0.03	0.9728
pctage55_74		0.0143	0.0054	0.0038	0.0248	2.68	0.0075
pctover75		-0.0094	0.0083	-0.0257	0.0070	-1.12	0.2619
PctBlack1		-0.0015	0.0024	-0.0062	0.0033	-0.60	0.5480
PctIndian1		-0.0063	0.0035	-0.0131	0.0005	-1.82	0.0684
PctAsian1		0.0061	0.0296	-0.0519	0.0642	0.21	0.8359
PctHawnPI1		-0.1479	0.0622	-0.2698	-0.0260	-2.38	0.0174
PctEnglis2		0.0166	0.0056	0.0057	0.0276	2.97	0.0029
PctSomeHig		0.0056	0.0071	-0.0084	0.0195	0.78	0.4347
PctHighSch		0.0174	0.0055	0.0065	0.0282	3.14	0.0017
PctSomeCol		0.0192	0.0057	0.0080	0.0304	3.35	0.0008
PctBachelo		0.0161	0.0077	0.0010	0.0312	2.10	0.0361
PctGradPro		0.0150	0.0103	-0.0052	0.0351	1.45	0.1464
PctArmedFo		0.9017	0.2841	0.3447	1.4586	3.17	0.0015
PctDisable		-0.0027	0.0047	-0.0118	0.0064	-0.58	0.5595
PctCivLabF		0.9175	0.2879	0.3532	1.4818	3.19	0.0014
PctNotInLF		0.9106	0.2879	0.3463	1.4750	3.16	0.0016
PctManufac		-0.0044	0.0042	-0.0127	0.0039	-1.03	0.3036
PctEducati		-0.0094	0.0052	-0.0196	0.0007	-1.82	0.0695
PctHealthS		-0.0004	0.0050	-0.0102	0.0094	-0.07	0.9403
PctTransOc		-0.0002	0.0049	-0.0097	0.0094	-0.04	0.9707
PctConsOcc		0.0000	0.0047	-0.0092	0.0093	0.01	0.9936
lPCI		0.4127	0.1516	0.1156	0.7098	2.72	0.0065
PctFemale		-0.0049	0.0071	-0.0189	0.0091	-0.69	0.4903
occhdn		-0.0001	0.0000	-0.0002	-0.0001	-4.26	<.0001
PctOccupie		0.0126	0.0045	0.0037	0.0215	2.77	0.0057
PctPlumbin		0.0182	0.0100	-0.0015	0.0379	1.81	0.0699
PctNoCars		0.0035	0.0062	-0.0087	0.0157	0.57	0.5711
PctAgeUnit		0.0130	0.0062	0.0009	0.0251	2.11	0.0346
PctAgeUn15		-0.0199	0.0064	-0.0324	-0.0073	-3.11	0.0019
PctBuiltBe		0.0084	0.0057	-0.0028	0.0197	1.47	0.1429
lavgage		1.0793	0.3537	0.3861	1.7725	3.05	0.0023
lAvgrent		0.1251	0.1011	-0.0731	0.3233	1.24	0.2161
lAvghval		0.3644	0.0851	0.1976	0.5311	4.28	<.0001
time	1201	12.3330	22.3115	-31.3967	56.0626	0.55	0.5804

time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lpopden*time	1201	0.1340	0.0770	-0.0169	0.2850	1.74	0.0819
lpopden*time	1202	0.0000	0.0000	0.0000	0.0000	.	.

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
lland*time	1201	0.0266	0.0700	-0.1106	0.1639	0.38	0.7037
lland*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lpophous*time	1201	0.1000	0.2836	-0.4559	0.6558	0.35	0.7244
lpophous*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
long*time	1201	-0.0033	0.0063	-0.0157	0.0091	-0.52	0.6026
long*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
hland*time	1201	-0.0112	0.0081	-0.0270	0.0047	-1.38	0.1673
hland*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lslopemn*time	1201	0.0825	0.0904	-0.0946	0.2597	0.91	0.3613
lslopemn*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lslopesd*time	1201	-0.2088	0.1148	-0.4339	0.0163	-1.82	0.0690
lslopesd*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lelevrang*time	1201	0.0450	0.0914	-0.1340	0.2241	0.49	0.6220
lelevrang*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lctimn*time	1201	-0.1529	0.1130	-0.3744	0.0685	-1.35	0.1759
lctimn*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
M1*time	1201	0.1373	0.1350	-0.1273	0.4020	1.02	0.3091
M1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
M4*time	1201	0.2296	0.0887	0.0558	0.4035	2.59	0.0096
M4*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
M5*time	1201	0.2269	0.1288	-0.0256	0.4794	1.76	0.0782
M5*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
M7*time	1201	-0.4629	0.4086	-1.2638	0.3380	-1.13	0.2573
M7*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
M11*time	1201	2.6069	0.4213	1.7812	3.4326	6.19	<.0001
M11*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e31*time	1201	0.0111	0.0303	-0.0483	0.0706	0.37	0.7139
e31*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e44*time	1201	-0.0247	0.0150	-0.0542	0.0048	-1.64	0.1002
e44*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e54*time	1201	0.0111	0.0403	-0.0679	0.0901	0.28	0.7831
e54*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e56*time	1201	-0.1049	0.0933	-0.2877	0.0779	-1.12	0.2608
e56*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e72*time	1201	-0.0591	0.0260	-0.1100	-0.0082	-2.27	0.0230
e72*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e81*time	1201	0.0159	0.0286	-0.0401	0.0719	0.56	0.5779
e81*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S1*time	1201	-0.0090	0.2411	-0.4815	0.4636	-0.04	0.9704
S1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S5*time	1201	0.2182	0.1655	-0.1062	0.5426	1.32	0.1874
S5*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S6*time	1201	-0.3312	0.3226	-0.9635	0.3010	-1.03	0.3045
S6*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S8*time	1201	-0.0022	0.2612	-0.5142	0.5098	-0.01	0.9933
S8*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S9*time	1201	1.8954	0.8812	0.1683	3.6225	2.15	0.0315
S9*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S12*time	1201	-0.9718	0.5823	-2.1132	0.1696	-1.67	0.0952
S12*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S15*time	1201	-7.5279	2.0368	-11.5200	-3.5358	-3.70	0.0002
S15*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S16*time	1201	-0.3858	0.2934	-0.9609	0.1893	-1.31	0.1886
S16*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S17*time	1201	-0.3687	0.1428	-0.6485	-0.0888	-2.58	0.0098
S17*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S18*time	1201	-0.1841	0.1823	-0.5415	0.1733	-1.01	0.3126
S18*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S19*time	1201	-0.6828	0.1542	-0.9850	-0.3806	-4.43	<.0001
S19*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S20*time	1201	-0.3531	0.1573	-0.6615	-0.0447	-2.24	0.0248
S20*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S21*time	1201	0.0363	0.1763	-0.3093	0.3819	0.21	0.8369
S21*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S22*time	1201	-0.1076	0.2415	-0.5809	0.3658	-0.45	0.6561

S22*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S23*time	1201	0.6619	0.2325	0.2062	1.1175	2.85	0.0044

S23*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S24*time	1201	-1.4275	0.4044	-2.2202	-0.6349	-3.53	0.0004

Parameter		Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
S24*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S25*time	1201	-0.2387	0.8987	-2.0001	1.5226	-0.27	0.7905
S25*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S26*time	1201	-0.4658	0.2275	-0.9117	-0.0199	-2.05	0.0406
S26*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S27*time	1201	-0.1822	0.1596	-0.4951	0.1306	-1.14	0.2536
S27*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S28*time	1201	-0.2867	0.2205	-0.7188	0.1454	-1.30	0.1934
S28*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S29*time	1201	0.0424	0.1332	-0.2186	0.3035	0.32	0.7501
S29*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S30*time	1201	-0.2987	0.2361	-0.7614	0.1640	-1.27	0.2058
S30*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S31*time	1201	-0.0380	0.1609	-0.3534	0.2773	-0.24	0.8131
S31*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S32*time	1201	0.6136	0.4412	-0.2512	1.4784	1.39	0.1644
S32*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S35*time	1201	-0.1630	0.2337	-0.6210	0.2950	-0.70	0.4854
S35*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S36*time	1201	0.0722	0.2089	-0.3371	0.4816	0.35	0.7294
S36*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S37*time	1201	-0.3651	0.2313	-0.8185	0.0883	-1.58	0.1145
S37*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S38*time	1201	0.1497	0.2052	-0.2525	0.5520	0.73	0.4656
S38*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S39*time	1201	-0.2159	0.2186	-0.6443	0.2125	-0.99	0.3233
S39*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S41*time	1201	0.0242	0.3557	-0.6730	0.7214	0.07	0.9458
S41*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S42*time	1201	-0.0652	0.1711	-0.4005	0.2702	-0.38	0.7033
S42*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S46*time	1201	-0.3365	0.1932	-0.7152	0.0421	-1.74	0.0815
S46*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S47*time	1201	-1.0725	0.3004	-1.6613	-0.4837	-3.57	0.0004
S47*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S49*time	1201	-0.2182	0.3026	-0.8113	0.3749	-0.72	0.4709
S49*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S51*time	1201	-0.1899	0.1732	-0.5294	0.1495	-1.10	0.2728
S51*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S53*time	1201	0.1129	0.3457	-0.5646	0.7905	0.33	0.7439
S53*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S54*time	1201	-0.7070	0.1894	-1.0783	-0.3357	-3.73	0.0002
S54*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S55*time	1201	0.3393	0.1903	-0.0336	0.7122	1.78	0.0745
S55*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctUrban*time	1201	0.0013	0.0018	-0.0022	0.0048	0.72	0.4704
PctUrban*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctOnFarms*time	1201	0.0017	0.0038	-0.0057	0.0091	0.44	0.6583
PctOnFarms*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
cerate*time	1201	0.0125	0.0366	-0.0591	0.0842	0.34	0.7320
cerate*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
crhc*time	1201	-0.3023	1.1433	-2.5431	1.9384	-0.26	0.7914
crhc*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
pctage55_74*time	1201	-0.0060	0.0054	-0.0165	0.0046	-1.10	0.2703
pctage55_74*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
pctover75*time	1201	0.0054	0.0086	-0.0114	0.0222	0.63	0.5302
pctover75*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctBlack1*time	1201	-0.0047	0.0025	-0.0095	0.0002	-1.87	0.0612
PctBlack1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.

PctIndian1*time	1201	-0.0029	0.0033	-0.0095	0.0036	-0.87	0.3832
PctIndian1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctAsian1*time	1201	-0.0304	0.0251	-0.0796	0.0188	-1.21	0.2255
PctAsian1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctHawnPI1*time	1201	0.1444	0.0688	0.0095	0.2794	2.10	0.0359
PctHawnPI1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctEnglis2*time	1201	0.0019	0.0058	-0.0095	0.0134	0.33	0.7408
PctEnglis2*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctSomeHig*time	1201	0.0096	0.0074	-0.0050	0.0241	1.29	0.1965

Parameter		**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
PctSomeHig*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctHighSch*time	1201	0.0059	0.0059	-0.0057	0.0176	1.00	0.3190
PctHighSch*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctSomeCol*time	1201	0.0107	0.0061	-0.0012	0.0225	1.76	0.0784
PctSomeCol*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctBachelo*time	1201	0.0156	0.0081	-0.0002	0.0314	1.94	0.0528
PctBachelo*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctGradPro*time	1201	0.0015	0.0100	-0.0182	0.0212	0.15	0.8790
PctGradPro*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctArmedFo*time	1201	-0.1223	0.2180	-0.5495	0.3050	-0.56	0.5749
PctArmedFo*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctDisable*time	1201	0.0020	0.0049	-0.0076	0.0117	0.41	0.6803
PctDisable*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctCivLabF*time	1201	-0.1192	0.2214	-0.5532	0.3148	-0.54	0.5903
PctCivLabF*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctNotInLF*time	1201	-0.1149	0.2213	-0.5488	0.3189	-0.52	0.6035
PctNotInLF*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctManufac*time	1201	-0.0037	0.0044	-0.0123	0.0049	-0.83	0.4038
PctManufac*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctEducati*time	1201	0.0010	0.0055	-0.0098	0.0118	0.19	0.8524
PctEducati*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctHealthS*time	1201	-0.0085	0.0053	-0.0190	0.0019	-1.61	0.1077
PctHealthS*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctTransOc*time	1201	-0.0008	0.0050	-0.0106	0.0090	-0.16	0.8699
PctTransOc*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctConsOcc*time	1201	-0.0051	0.0050	-0.0149	0.0047	-1.02	0.3066
PctConsOcc*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lPCI*time	1201	-0.2222	0.1498	-0.5158	0.0714	-1.48	0.1380
lPCI*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctFemale*time	1201	-0.0128	0.0071	-0.0267	0.0011	-1.80	0.0720
PctFemale*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctGQPop*time	1201	-0.0060	0.0058	-0.0174	0.0054	-1.04	0.3000
PctGQPop*time	1202	0.0017	0.0065	-0.0110	0.0144	0.26	0.7959
occhdn*time	1201	0.0000	0.0000	-0.0001	0.0001	0.66	0.5067
occhdn*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctOccupie*time	1201	-0.0028	0.0047	-0.0120	0.0065	-0.58	0.5588
PctOccupie*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctPlumbin*time	1201	0.0189	0.0103	-0.0013	0.0391	1.83	0.0669
PctPlumbin*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctNoCars*time	1201	0.0016	0.0069	-0.0118	0.0151	0.24	0.8117
PctNoCars*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctAgeUnit*time	1201	-0.0130	0.0060	-0.0249	-0.0012	-2.16	0.0309
PctAgeUnit*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctAgeUn15*time	1201	0.0166	0.0063	0.0043	0.0289	2.65	0.0081
PctAgeUn15*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctBuiltBe*time	1201	-0.0104	0.0059	-0.0218	0.0011	-1.77	0.0767
PctBuiltBe*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lavgage*time	1201	-0.6254	0.3023	-1.2179	-0.0328	-2.07	0.0386
lavgage*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lAvgrent*time	1201	0.0255	0.1064	-0.1831	0.2341	0.24	0.8109
lAvgrent*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lAvghval*time	1201	0.2775	0.0845	0.1119	0.4431	3.28	0.0010
lAvghval*time	1202	0.0000	0.0000	0.0000	0.0000	.	.

It is interesting to note that in these preliminary results, none of the longitude-related variables (long, hland[38]) are significant. The only "terrain" variables that are statistically significant are several of the MODIS land cover codings with a negative impact (the baseline in the intercept is code 13, urban building cover), and log of mean CTI (the "wetness" index) is on the fringes of marginal statistical significance. A point estimate of the effect of the latter is quite small, however, but positive. An increase of 10% in wetness increases the odds of broadband by 1.6%. One is hard-pressed not to be tempted to interpret this positive effect as related to increased ease of digging in the ground and installing cable!

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. Differences between coefficients in 2001 and 2002 were first estimated as offsets from 2002 values. In cases where the offset for 2001 was statistically significant, but the base coefficient for 2002 was not, the variable was reformulated as two separate effects—one for 2001 and one for 2002, in order to make it easier to drop variables that were not significant.

Table 4 shows the results of fitting this more parsimonious model, and dropping all coefficients not at least marginally signficiant (i.e., at the 10% level). The results are broadly consistent with all results shown for 2001-2003 in the appendixes, and the preliminary GEE estimates. A detailed discussion of the results is worthwhile.

Table 4: Parsimonious Model

Analysis Of GEE Parameter Estimates
Empirical Standard Error Estimates

Parameter	Estimate	Standard Error	95% Confidence Limits		Z	Pr > |Z|
Intercept	-101.757	28.0273	-156.690	-46.8249	-3.63	0.0003
lpopden	0.8533	0.0571	0.7414	0.9651	14.95	<.0001
lland	0.9325	0.0524	0.8298	1.0352	17.80	<.0001
M1	-0.2243	0.0860	-0.3929	-0.0557	-2.61	0.0091
M11	-3.4139	0.7183	-4.8217	-2.0061	-4.75	<.0001
e31	0.0708	0.0255	0.0208	0.1208	2.77	0.0056
e44	0.0259	0.0116	0.0032	0.0486	2.23	0.0254
e54	0.1120	0.0260	0.0610	0.1630	4.30	<.0001
e72	0.0748	0.0230	0.0297	0.1200	3.25	0.0011
e81	0.0580	0.0192	0.0204	0.0957	3.02	0.0025
S6	0.4926	0.1561	0.1866	0.7986	3.16	0.0016
S8	0.7503	0.2132	0.3324	1.1681	3.52	0.0004
s901	2.3830	1.0775	0.2712	4.4949	2.21	0.0270
S12	0.8356	0.2576	0.3307	1.3404	3.24	0.0012
s1501	-6.4801	1.5136	-9.4468	-3.5135	-4.28	<.0001
S17	-0.3133	0.1164	-0.5413	-0.0852	-2.69	0.0071
S18	-0.8014	0.1276	-1.0516	-0.5513	-6.28	<.0001
S19	-1.0333	0.1162	-1.2609	-0.8056	-8.90	<.0001
S20	-0.4762	0.1362	-0.7432	-0.2092	-3.50	0.0005
s2301	0.8121	0.1770	0.4651	1.1590	4.59	<.0001
S24	1.1461	0.3702	0.4205	1.8718	3.10	0.0020
S25	1.8213	0.5180	0.8060	2.8366	3.52	0.0004
s2601	-0.3633	0.1405	-0.6387	-0.0880	-2.59	0.0097
S27	-1.1628	0.1020	-1.3628	-0.9629	-11.40	<.0001
S29	-0.1849	0.0953	-0.3718	0.0020	-1.94	0.0525

[38] H[eart]land is defined as the absolute value of (longitude –(- 95) degrees).

S31		-0.5993	0.1176	-0.8299	-0.3688	-5.09	<.0001
S32		-1.4960	0.4362	-2.3509	-0.6411	-3.43	0.0006
S36		1.0128	0.1294	0.7592	1.2665	7.83	<.0001
S37		0.2882	0.1478	-0.0014	0.5778	1.95	0.0511
S38		-0.6031	0.1243	-0.8468	-0.3595	-4.85	<.0001
S41		0.5958	0.1929	0.2177	0.9739	3.09	0.0020
S42		-0.8358	0.0959	-1.0236	-0.6479	-8.72	<.0001
S46		-0.4763	0.1291	-0.7294	-0.2233	-3.69	0.0002
S47		0.7798	0.2275	0.3340	1.2257	3.43	0.0006
S49		-0.8518	0.2230	-1.2888	-0.4147	-3.82	0.0001
S51		-0.4647	0.1106	-0.6814	-0.2480	-4.20	<.0001
s5401		-0.5056	0.1133	-0.7277	-0.2835	-4.46	<.0001
S55		-0.7116	0.1652	-1.0353	-0.3879	-4.31	<.0001
PctOnFarms		-0.0094	0.0028	-0.0149	-0.0039	-3.33	0.0009
Parameter		**Estimate**	**Standard Error**	**95% Confidence Limits**		**Z**	**Pr > |Z|**
pctage55_74		0.0129	0.0042	0.0047	0.0211	3.07	0.0021
pctblack101		-0.0052	0.0015	-0.0082	-0.0022	-3.41	0.0006
PctIndian1		-0.0066	0.0026	-0.0117	-0.0015	-2.53	0.0113
PctHawnPI1		-0.0899	0.0269	-0.1427	-0.0371	-3.34	0.0008
PctEnglis2		0.0165	0.0039	0.0087	0.0242	4.18	<.0001
PctHighSch		0.0160	0.0031	0.0099	0.0221	5.12	<.0001
PctSomeCol		0.0190	0.0033	0.0125	0.0254	5.76	<.0001
PctBachelo		0.0223	0.0053	0.0119	0.0326	4.22	<.0001
PctArmedFo		0.8759	0.2751	0.3366	1.4151	3.18	0.0015
PctCivLabF		0.8960	0.2797	0.3478	1.4443	3.20	0.0014
PctNotInLF		0.8887	0.2797	0.3405	1.4370	3.18	0.0015
PctEducati		-0.0066	0.0040	-0.0144	0.0012	-1.66	0.0973
lPCI		0.4133	0.1118	0.1942	0.6323	3.70	0.0002
pctfemale01		-0.0133	0.0055	-0.0240	-0.0026	-2.43	0.0153
occhdn		-0.0001	0.0001	-0.0002	0.0000	-1.76	0.0778
PctOccupie		0.0046	0.0019	0.0009	0.0084	2.45	0.0143
PctPlumbin		0.0291	0.0081	0.0133	0.0449	3.61	0.0003
PctAgeUn15		-0.0177	0.0059	-0.0292	-0.0062	-3.02	0.0025
pctbuiltbe02		0.0107	0.0053	0.0003	0.0210	2.02	0.0430
lavgage		0.7129	0.2308	0.2607	1.1652	3.09	0.0020
lAvghval		0.4734	0.0670	0.3421	0.6047	7.07	<.0001
lpopden*time	1201	0.0788	0.0170	0.0455	0.1121	4.64	<.0001
lpopden*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e44*time	1201	-0.0215	0.0105	-0.0421	-0.0008	-2.04	0.0413
e44*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
e72*time	1201	-0.0581	0.0219	-0.1010	-0.0152	-2.66	0.0079
e72*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S17*time	1201	-0.3091	0.1115	-0.5276	-0.0905	-2.77	0.0056
S17*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S19*time	1201	-0.5617	0.1135	-0.7841	-0.3393	-4.95	<.0001
S19*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S20*time	1201	-0.2752	0.1263	-0.5227	-0.0278	-2.18	0.0293
S20*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S24*time	1201	-1.4064	0.3808	-2.1528	-0.6599	-3.69	0.0002
S24*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S47*time	1201	-1.0183	0.2656	-1.5388	-0.4977	-3.83	0.0001
S47*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
S55*time	1201	0.4249	0.1635	0.1043	0.7454	2.60	0.0094
S55*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctHawnPI1*time	1201	0.1052	0.0634	-0.0191	0.2296	1.66	0.0971
PctHawnPI1*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
PctAgeUn15*time	1201	0.0133	0.0055	0.0025	0.0241	2.41	0.0161
PctAgeUn15*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lavgage*time	1201	-0.3092	0.1298	-0.5636	-0.0548	-2.38	0.0172
lavgage*time	1202	0.0000	0.0000	0.0000	0.0000	.	.
lAvghval*time	1201	0.1018	0.0344	0.0345	0.1692	2.96	0.0030
lAvghval*time	1202	0.0000	0.0000	0.0000	0.0000	.	.

Population density is statistically significant in all my estimated models, has the expected sign, and has a relatively large impact. We can interpret the coefficient of its log as an elasticity: a 10% increase in population density, cet. par., results in an 8.5%

increase in the odds ratio for broadband within a zip code in 2000. It seems to have been declining over time.

Zip code land area is very close in size and sign to population density. The coefficient of the log of land area is also interpretable as an elasticity, and suggests that a 10% increase in land area, for given population density and all else, results in a 9.3% increase in the odds of broadband. Taken together with the result on population density, the two coefficients are similar enough in size to suggest that broadband odds increase a little less than proportionately as the population of any zip code increases.

Terrain variables M1 (Evergreen needleleaf forests) and M11 (permanent wetlands, a permanent mixture of surface water and herbaceous or woody vegetation) both make the final cut. Point estimates suggest that a pine forest canopy land cover cuts the odds ratio for broadband by about 20 percent, while a permanent swamp pretty much kills hopes for broadband, lowering the odds to about 3 percent of the odds with all else equal in an urban setting.

Industry presence—at least if it is the right industry—seems to significantly increase the odds of broadband in a zip code. An additional establishment in manufacturing (NAICS code 31, e31 is number of establishments in this sector), or accommodation and food services (NAICS 72) raise the odds of broadband in 2002 by about 7 percent. An establishment in other services (NAICS 81) clocks in at 6 percent, while NAICS 44 (retail trade) yields a 2.6% improvement in the odds. Another professional, scientific, and professional services establishment (NAICS 54) raises the odds by a whopping 11%. The effects in NAICS 44 and 72 seem to have grown significantly over time, and were substantially lower in 2001.

Farm settings don't seem to help the odds of broadband entry, but also don't hurt it a lot. A one percentage point increase in the percent of the population living on farms lowers the odds of broadband by about one percent.

Older people, or at least the middle-aged, do not hurt the odds of broadband entry into a zip code. A one percentage point increase in the share of the population in the 55 to 74 age bracket increases the odds very slightly, by about one percent, all else being equal. Similarly, the impact of an increase in the share of the population in the civilian labor force, or the armed forces, or even being over 16 and not in the labor force—all of which are associated with being over 16—is to substantially increase the odds of broadband.

Ethnicity has measurable impacts on the odds of broadband in a zip code. A one percentage point increase in the percentage of the population identified as black reduced broadband odds by a half a percent in 2001. Interestingly, this digital divide seems to be closing—this differential was not statistically significant in 2001. A one percentage point increase in the native Hawaiian population share reduced the odds of broadband by almost 9% in 2002, on the other hand, and this digital divide seems to be growing compared to 2001. A one percentage point increase in Native Americans' share of the population is associated with a 2/3 percentage point reduction in broadband odds, and this

seems to have been static over 2001-02. A one percentage point increase in the share of speakers of English as a second language is associated with a 1.7 percent increase in the broadband odds ratio. This, perhaps, reflects the superior educational attainment of those who master two languages.

Education has strong effects on greater broadband availability. A one percentage point increase in the share of the population with a high school through college BA degree increases the odds of broadband by from 1.6 to 2.2 percent. The effect increases with educational attainment.

Working in education (or educational services), on the other hand, seems to reduce the broadband odds slightly. An increase of one percentage point in the share of the population working in this sector lowers broadband odds by about 7/10 of a percent. This is most easily explained by access to high speed internet connections at work substituting for a high speed connection in the home.

Gender has the stereotypical effect on broadband use. A one percentage point increase in the female share of the population is associated with a 1.3 percent decline in the odds of broadband.

Per capita income has a strong effect on broadband use. A 10% increase in income increase the odds of broadband use by about 4 percent.

Housing stock characteristics are associated with a variety of effects. An increase in the occupied housing density, all else being equal, has a very slight but negative effect on broadband provision. A one percentage point increase in the share of housing units older than 50 years reduces broadband odds by about 1.8 percent, and this effect seems to have been growing over time. An increase in the share of homes occupied (not vacant), and the share of home with indoor plumbing, both possible signals of home quality, are associated with greater broadband penetration. Similarly, an increase of 10% in mean home value is associated with a 5% increase in broadband odds. More homes built before 1940 seems to translate into slightly greater broadband odds, an effect that again may be associated with home quality. An increase in mean housing age, on the other hand, translates into more broadband—this may in fact be a proxy for distance from a central office, and DSL availability. The complex effects associated with measured housing characteristics suggests an array of diverse factors at work.

State effects—which it is natural to associate with differences in policies, across states, given the array of other variables for which we control in this analysis—can be parsed into groups. My baseline is Texas (FIPS 48), which had a modest but sustained Internet and broadband subsidy program in place from 1996 to 2004 (TIF, the Telecommunications Infrastructure Fund, which distributed about $1.5 billion over this period), and a relatively competitive regulatory environment. For a state not assigned to one of these groups, I could not reject the hypothesis that state effects were the same as Texas in overall impact on broadband odds.

Greater encouragement of broadband use in 2002: California (6), Colorado(8), Florida(12), Maryland(24), Massachusetts(25), New York(36), North Carolina(37), Oregon(41), Tennessee(47). Maryland, Tennessee increasing relative to 2001.

Less encouragement of broadband use in 2002: Illinois(17), Indiana(18), Iowa(19), Kansas(20), Minnesota(27), Missouri(29), Nebraska(31), Nevada(32), North Dakota(38), Pennsylvania(42), South Dakota(46), Utah(49), Virginia(51), Wisconsin(55). Illinois, Iowa, Kansas increasing in 2002 relative to 2001. Wisconsin decreasing relative to 2001.

Greater in 2001, parity in 2002: Connecticut(9), Maine(23).

Less in 2001, parity in 2002: Hawaii(15), Michigan(26), West Virginia(54).

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 (with the exceptions sketched out above) 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.

Our eRate and rural health care grant variables did not prove to have either substantial or statistically significant effects on broadband penetration. These programs were not designed to promote general use of broadband by homes and businesses. They do not seem to be having any incidental impacts.

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).

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.

Finally, it will be possible to apply these methods to model broadband penetration by zip code in other years. A more sophisticated random effects approach to estimating a model utilizing a panel of zip code data 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 we have seen is currently 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) of the population now lives in areas where broadband services are not currently being provided. Nonetheless, our attempt to dissect the relative contributions of different factors to the recent historical dynamics of broadband penetration is potentially useful in understanding the possible influence of many of these same forces in accelerating or retarding the next generation of advanced information services.

My analysis suggests that state policies may play an important role, and that statistical methods are useful in assessing this role. The ranking of state effects produced by my model seems to correlate with casual impressions of the effectiveness of state policies as portrayed in the press and trade journals.

To some extent, at least, geography is destiny. Terrain effects (presumably increasing or decreasing the cost of installing and maintaining a network) seem to be significant in some parts of the country.

Two factors often associated with broadband penetration, income and population density, unsurprisingly seem to be among the most important determinants of broadband penetration. 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 local broadband availability. Professional and technical service establishments seem to have the largest such impact.

Common perceptions of the effects of gender, education, and rural location on broadband penetration seem to be supported by a causal analysis that attempts to control for confounding factors. Age effects estimated in this paper do not support the conventional wisdom. A higher share of the population in the 55-75 age group increases the odds of broadband.

Finally, “digital divide” type ethnic, racial and personal variables show up as small, but statistically perceptible effects. There were reduced odds of broadband provision in zip

codes with larger Afro-American and Native American populations in 2001, but the gap seems to be closing for Afro-Americans.

Appendix A

Broadband Logit Equations Using "Significant" Variables

12/2000 - 12/2003

The LOGISTIC Procedure

Model Information

Data Set	WORK.PROBSTATB
Response Variable	bpen1200
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	27812
Number of Observations Used	27811

Response Profile

Ordered Value	bpen1200	Total Frequency
1	1	20879
2	0	6932

Probability modeled is bpen1200=1.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	31234.530	17189.815
SC	31242.764	18095.466
-2 Log L	31232.530	16969.815

R-Square 0.4012 Max-rescaled R-Square 0.5946

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	14262.7150	109	<.0001
Score	11504.2241	109	<.0001
Wald	4484.4051	109	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-6.8775	30.2914	0.0515	0.8204
lpopden	1	0.7173	0.0399	322.4688	<.0001
lland	1	0.6428	0.0326	388.0563	<.0001
lpophous	1	-0.0224	0.2142	0.0109	0.9167
long	1	0.00499	0.00436	1.3091	0.2525
hland	1	-0.0272	0.00631	18.5594	<.0001
lslopesd	1	-0.2585	0.0680	14.4560	0.0001
lelevrang	1	0.3013	0.0689	19.1216	<.0001
lctimn	1	0.0985	0.0851	1.3391	0.2472
M0	1	0.2276	0.1149	3.9276	0.0475
M1	1	-0.0845	0.0975	0.7503	0.3864

M4	1	0.1426	0.0689	4.2809	0.0385
M7	1	0.0332	0.2795	0.0141	0.9055
M8	1	0.1390	0.2075	0.4491	0.5027
M11	1	-1.0378	1.3842	0.5621	0.4534
e31	1	0.0240	0.00909	6.9606	0.0083
e44	1	0.00815	0.00528	2.3866	0.1224
e54	1	0.0666	0.0150	19.7234	<.0001
e56	1	0.1107	0.0214	26.8033	<.0001
e61	1	-0.00395	0.0802	0.0024	0.9607
e62	1	0.0383	0.0130	8.6469	0.0033
e72	1	0.00933	0.00810	1.3251	0.2497
e81	1	0.00311	0.0128	0.0594	0.8074
S1	1	0.2004	0.1448	1.9148	0.1664
S4	1	1.0173	0.3211	10.0354	0.0015
S5	1	-0.7769	0.1432	29.4174	<.0001
S6	1	0.0735	0.1756	0.1749	0.6758
S8	1	-0.0936	0.1824	0.2635	0.6078
S9	1	0.5306	0.4094	1.6798	0.1949
S12	1	1.2731	0.2576	24.4181	<.0001
S15	1	0.8835	0.9384	0.8863	0.3465
S16	1	-0.1560	0.2073	0.5665	0.4516
S17	1	-0.1870	0.1135	2.7129	0.0995
S18	1	-0.3407	0.1379	6.1010	0.0135
S19	1	-1.2851	0.1336	92.5161	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S20	1	0.3523	0.1392	6.4081	0.0114
S22	1	-0.3552	0.1571	5.1143	0.0237
S23	1	1.5303	0.2330	43.1459	<.0001
S24	1	3.6443	0.6071	36.0307	<.0001
S26	1	0.2561	0.1433	3.1927	0.0740
S27	1	-0.9029	0.1367	43.6117	<.0001
S28	1	-0.0343	0.1649	0.0432	0.8353
S29	1	-0.6037	0.1192	25.6467	<.0001
S31	1	-0.8848	0.1521	33.8550	<.0001
S32	1	-0.9327	0.3678	6.4296	0.0112
S33	1	2.8896	0.4933	34.3171	<.0001
S34	1	2.0159	0.6162	10.7009	0.0011
S35	1	-0.0196	0.2297	0.0073	0.9320
S36	1	2.2272	0.1888	139.1011	<.0001
S37	1	0.0242	0.1566	0.0240	0.8770
S38	1	-0.00239	0.1709	0.0002	0.9889
S39	1	0.5522	0.1480	13.9274	0.0002
S40	1	-0.7060	0.1433	24.2904	<.0001
S41	1	2.0617	0.2343	77.4458	<.0001
S42	1	0.3630	0.1434	6.4110	0.0113
S44	1	1.1885	0.8146	2.1287	0.1446
S45	1	0.1377	0.1919	0.5151	0.4730
S46	1	-0.3830	0.1770	4.6796	0.0305
S47	1	-0.1275	0.1494	0.7282	0.3935
S49	1	0.0457	0.2601	0.0309	0.8606
S50	1	2.3451	0.2829	68.7181	<.0001
S51	1	0.9762	0.1515	41.5327	<.0001
S54	1	1.3769	0.1479	86.6172	<.0001
S55	1	-0.5529	0.1375	16.1729	<.0001
S56	1	0.0319	0.2732	0.0136	0.9071
PctUrban	1	0.00121	0.00115	1.1023	0.2938
PctOnFarms	1	-0.00385	0.00365	1.1121	0.2916
cerate99	1	-0.00041	0.0210	0.0004	0.9842
crhc99	1	0.0649	1.0696	0.0037	0.9516
PctAge55_5	1	-0.0105	0.00972	1.1624	0.2810
PctAge60_6	1	-0.0134	0.0106	1.6002	0.2059
PctAge65_7	1	-0.00639	0.00797	0.6432	0.4225
PctAge75_8	1	0.0104	0.00999	1.0905	0.2964
PctOver85	1	-0.0288	0.0164	3.0905	0.0788
PctBlack1	1	-0.00641	0.00195	10.8314	0.0010
PctIndian1	1	-0.00560	0.00336	2.7714	0.0960
PctHawnPI1	1	-0.0279	0.0403	0.4812	0.4879
PctOther1	1	0.0153	0.00630	5.8880	0.0152
PctEnglis2	1	0.0125	0.00609	4.2431	0.0394

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
PctNoEngli	1	-0.0462	0.0214	4.6716	0.0307
PctSomeHig	1	0.0135	0.00662	4.1676	0.0412
PctHighSch	1	0.00935	0.00517	3.2626	0.0709
PctSomeCol	1	0.0171	0.00543	9.9533	0.0016
PctBachelo	1	0.0184	0.00740	6.2125	0.0127
PctGradPro	1	0.0195	0.00940	4.3019	0.0381
PctArmedFo	1	-0.1177	0.2965	0.1576	0.6914
PctCivLabF	1	-0.1262	0.3026	0.1741	0.6765
PctNotInLF	1	-0.1242	0.3025	0.1686	0.6814
PctManufac	1	-0.00555	0.00374	2.1959	0.1384
PctRetailT	1	-0.00865	0.00544	2.5260	0.1120
PctEducati	1	-0.0102	0.00513	3.9457	0.0470
PctHealthS	1	-0.00145	0.00492	0.0868	0.7682
PctService	1	0.0160	0.00444	12.9711	0.0003
PctSalesOf	1	0.00858	0.00475	3.2535	0.0713
PctTransOc	1	0.00558	0.00498	1.2535	0.2629
PctConsOcc	1	0.00617	0.00480	1.6487	0.1991
lPCI	1	0.6451	0.1397	21.3233	<.0001
PctFemale	1	-0.00448	0.00697	0.4138	0.5201
PctInInsti	1	-0.00771	0.00656	1.3817	0.2398
occhdn	1	-0.00008	0.000095	0.6958	0.4042
PctOccupie	1	0.0107	0.00385	7.7251	0.0054
PctPlumbin	1	-0.00668	0.00993	0.4529	0.5010
PctNoPhone	1	0.00105	0.00643	0.0265	0.8707
PctNoCars	1	0.0122	0.00602	4.0931	0.0431
PctAgeUnit	1	0.0112	0.00511	4.7708	0.0289
PctAgeUn15	1	0.00642	0.00579	1.2293	0.2675
PctBuiltBe	1	-0.0121	0.00525	5.2885	0.0215
lavgage	1	0.6454	0.2641	5.9723	0.0145
lAvgrent	1	0.2642	0.0985	7.1870	0.0073
lAvghval	1	0.5168	0.0783	43.5854	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lpopden	2.049	1.895	2.216
lland	1.902	1.784	2.027
lpophous	0.978	0.643	1.488
long	1.005	0.996	1.014
hland	0.973	0.961	0.985
lslopesd	0.772	0.676	0.882

statsig binary model, 1200, lpophous 197

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lelevrang	1.352	1.181	1.547
lctimn	1.103	0.934	1.304
M0	1.256	1.003	1.573
M1	0.919	0.759	1.113
M4	1.153	1.008	1.320
M7	1.034	0.598	1.788
M8	1.149	0.765	1.726
M11	0.354	0.023	5.340
e31	1.024	1.006	1.043
e44	1.008	0.998	1.019
e54	1.069	1.038	1.101
e56	1.117	1.071	1.165
e61	0.996	0.851	1.166
e62	1.039	1.013	1.066
e72	1.009	0.993	1.026
e81	1.003	0.978	1.029
S1	1.222	0.920	1.623
S4	2.766	1.474	5.190
S5	0.460	0.347	0.609
S6	1.076	0.763	1.518

S8	0.911	0.637	1.302
S9	1.700	0.762	3.793
S12	3.572	2.156	5.918
S15	2.419	0.384	15.222
S16	0.856	0.570	1.284
S17	0.829	0.664	1.036
S18	0.711	0.543	0.932
S19	0.277	0.213	0.359
S20	1.422	1.083	1.868
S22	0.701	0.515	0.954
S23	4.620	2.926	7.293
S24	38.257	11.639	125.747
S26	1.292	0.975	1.711
S27	0.405	0.310	0.530
S28	0.966	0.699	1.335
S29	0.547	0.433	0.691
S31	0.413	0.306	0.556
S32	0.393	0.191	0.809
S33	17.985	6.840	47.291
S34	7.507	2.244	25.121
S35	0.981	0.625	1.538
S36	9.274	6.405	13.428
S37	1.025	0.754	1.393
S38	0.998	0.714	1.395

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
S39	1.737	1.300	2.322
S40	0.494	0.373	0.654
S41	7.859	4.966	12.440
S42	1.438	1.085	1.904
S44	3.282	0.665	16.200
S45	1.148	0.788	1.672
S46	0.682	0.482	0.965
S47	0.880	0.657	1.180
S49	1.047	0.629	1.743
S50	10.434	5.993	18.166
S51	2.654	1.972	3.572
S54	3.963	2.965	5.296
S55	0.575	0.439	0.753
S56	1.032	0.604	1.764
PctUrban	1.001	0.999	1.003
PctOnFarms	0.996	0.989	1.003
cerate99	1.000	0.959	1.042
crhc99	1.067	0.131	8.683
PctAge55_5	0.990	0.971	1.009
PctAge60_6	0.987	0.967	1.007
PctAge65_7	0.994	0.978	1.009
PctAge75_8	1.010	0.991	1.030
PctOver85	0.972	0.941	1.003
PctBlack1	0.994	0.990	0.997
PctIndian1	0.994	0.988	1.001
PctHawnPI1	0.972	0.899	1.052
PctOther1	1.015	1.003	1.028
PctEnglis2	1.013	1.001	1.025
PctNoEngli	0.955	0.916	0.996
PctSomeHig	1.014	1.001	1.027
PctHighSch	1.009	0.999	1.020
PctSomeCol	1.017	1.007	1.028
PctBachelo	1.019	1.004	1.033
PctGradPro	1.020	1.001	1.039
PctArmedFo	0.889	0.497	1.589
PctCivLabF	0.881	0.487	1.595
PctNotInLF	0.883	0.488	1.598
PctManufac	0.994	0.987	1.002
PctRetailT	0.991	0.981	1.002
PctEducati	0.990	0.980	1.000
PctHealthS	0.999	0.989	1.008
PctService	1.016	1.007	1.025
PctSalesOf	1.009	0.999	1.018
PctTransOc	1.006	0.996	1.015

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PctConsOcc	1.006	0.997	1.016
lPCI	1.906	1.450	2.506
PctFemale	0.996	0.982	1.009
PctInInsti	0.992	0.980	1.005
occhdn	1.000	1.000	1.000
PctOccupie	1.011	1.003	1.018
PctPlumbin	0.993	0.974	1.013
PctNoPhone	1.001	0.989	1.014
PctNoCars	1.012	1.000	1.024
PctAgeUnit	1.011	1.001	1.021
PctAgeUn15	1.006	0.995	1.018
PctBuiltBe	0.988	0.978	0.998
lavgage	1.907	1.136	3.199
lAvgrent	1.302	1.074	1.580
lAvghval	1.677	1.438	1.955

Association of Predicted Probabilities and Observed Responses

Percent Concordant	91.9	Somers' D	0.838
Percent Discordant	8.0	Gamma	0.839
Percent Tied	0.1	Tau-a	0.314
Pairs	144733228	c	0.919

statsig binary model, 1201, lpophous 200

The LOGISTIC Procedure

Model Information

Data Set	WORK.PROBSTATB
Response Variable	b1201
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	27812
Number of Observations Used	27811

Response Profile

Ordered Value	b1201	Total Frequency
1	1	22623
2	0	5188

Probability modeled is b1201=1.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	26765.863	15094.386
SC	26774.097	16000.037
-2 Log L	26763.863	14874.386

R-Square 0.3479 Max-rescaled R-Square 0.5629

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	11889.4770	109	<.0001
Score	10256.0975	109	<.0001
Wald	3890.7504	109	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-98.1298	34.2296	8.2186	0.0041
lpopden	1	0.9423	0.0437	464.0052	<.0001
lland	1	0.9620	0.0361	711.4234	<.0001
lpophous	1	-0.2427	0.2168	1.2530	0.2630
long	1	0.00775	0.00519	2.2279	0.1355
hland	1	0.000318	0.00732	0.0019	0.9653
lslopesd	1	-0.1255	0.0734	2.9227	0.0873
lelevrang	1	0.1363	0.0744	3.3578	0.0669
lctimn	1	-0.00173	0.0951	0.0003	0.9855
M0	1	0.0627	0.1229	0.2602	0.6100
M1	1	-0.1961	0.1096	3.2005	0.0736
M4	1	0.0167	0.0735	0.0513	0.8208
M7	1	-0.5989	0.2939	4.1519	0.0416
M8	1	0.0273	0.2226	0.0150	0.9024
M11	1	-2.3995	1.5913	2.2738	0.1316
e31	1	0.0785	0.0128	37.3619	<.0001
e44	1	0.00300	0.00663	0.2050	0.6507
e54	1	0.1054	0.0201	27.6279	<.0001
e56	1	-0.0584	0.0250	5.4753	0.0193
e61	1	-0.0445	0.1003	0.1968	0.6573
e62	1	0.0152	0.0156	0.9584	0.3276
e72	1	0.0136	0.0106	1.6408	0.2002
e81	1	0.0582	0.0168	11.9249	0.0006
S1	1	0.0859	0.1718	0.2498	0.6172
S4	1	0.1507	0.3718	0.1644	0.6851
S5	1	0.2507	0.1539	2.6519	0.1034
S6	1	0.2880	0.2113	1.8591	0.1727
S8	1	0.6549	0.2234	8.5909	0.0034
S9	1	2.0999	1.0560	3.9544	0.0467
S12	1	0.3887	0.2888	1.8120	0.1783
S15	1	-6.1201	0.9252	43.7616	<.0001
S16	1	-0.7928	0.2193	13.0748	0.0003
S17	1	-0.6486	0.1238	27.4632	<.0001
S18	1	-0.8424	0.1553	29.4299	<.0001
S19	1	-1.5463	0.1390	123.6919	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S20	1	-0.7428	0.1470	25.5270	<.0001
S22	1	0.2105	0.1869	1.2678	0.2602
S23	1	0.6052	0.2838	4.5488	0.0329
S24	1	-0.4644	0.2515	3.4083	0.0649
S26	1	-0.4150	0.1648	6.3369	0.0118
S27	1	-1.1206	0.1453	59.4528	<.0001
S28	1	-0.4468	0.1809	6.0997	0.0135
S29	1	-0.1538	0.1284	1.4348	0.2310
S31	1	-0.5506	0.1503	13.4273	0.0002
S32	1	-1.2438	0.3990	9.7181	0.0018
S33	1	0.5231	0.3898	1.8006	0.1796
S34	1	0.5491	0.5157	1.1336	0.2870
S35	1	-0.4081	0.2431	2.8193	0.0931
S36	1	0.8192	0.2055	15.8886	<.0001
S37	1	0.0946	0.1995	0.2248	0.6354

S38	1	-0.4302	0.1697	6.4240	0.0113
S39	1	0.0426	0.1762	0.0585	0.8089
S40	1	0.0243	0.1492	0.0266	0.8704
S41	1	0.5779	0.2398	5.8078	0.0160
S42	1	-0.9299	0.1655	31.5591	<.0001
S44	1	0.2109	0.8603	0.0601	0.8063
S45	1	-0.0426	0.2294	0.0346	0.8525
S46	1	-0.5555	0.1740	10.1911	0.0014
S47	1	-0.1229	0.1731	0.5042	0.4776
S49	1	-0.9846	0.2760	12.7293	0.0004
S50	1	-0.3122	0.2588	1.4558	0.2276
S51	1	-0.6256	0.1626	14.7949	0.0001
S54	1	-0.5760	0.1550	13.8046	0.0002
S55	1	-0.2783	0.1655	2.8286	0.0926
S56	1	0.2604	0.2897	0.8080	0.3687
PctUrban	1	0.00210	0.00138	2.3266	0.1272
PctOnFarms	1	-0.00743	0.00354	4.4087	0.0358
cerate00	1	0.0375	0.0271	1.9179	0.1661
crhc00	1	-0.3461	0.7568	0.2091	0.6474
PctAge55_5	1	0.0144	0.00940	2.3468	0.1255
PctAge60_6	1	-0.00251	0.0103	0.0590	0.8081
PctAge65_7	1	0.0115	0.00785	2.1628	0.1414
PctAge75_8	1	-0.00580	0.00993	0.3415	0.5589
PctOver85	1	0.00250	0.0161	0.0240	0.8768
PctBlack1	1	-0.00579	0.00209	7.6333	0.0057
PctIndian1	1	-0.00934	0.00342	7.4856	0.0062
PctHawnPI1	1	0.00687	0.0488	0.0198	0.8881
PctOther1	1	0.00721	0.00689	1.0961	0.2951
PctEnglis2	1	0.0156	0.00624	6.2126	0.0127

statsig binary model, 1201, lpophous 203

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
PctNoEngli	1	0.0161	0.0239	0.4564	0.4993
PctSomeHig	1	0.0193	0.00671	8.2899	0.0040
PctHighSch	1	0.0274	0.00526	27.1632	<.0001
PctSomeCol	1	0.0342	0.00555	37.9300	<.0001
PctBachelo	1	0.0356	0.00742	23.0899	<.0001
PctGradPro	1	0.0220	0.00946	5.4067	0.0201
PctArmedFo	1	0.8227	0.3365	5.9769	0.0145
PctCivLabF	1	0.8471	0.3418	6.1426	0.0132
PctNotInLF	1	0.8452	0.3418	6.1167	0.0134
PctManufac	1	-0.00793	0.00388	4.1739	0.0411
PctRetailT	1	-0.00200	0.00543	0.1354	0.7129
PctEducati	1	-0.00862	0.00514	2.8130	0.0935
PctHealthS	1	-0.00800	0.00499	2.5755	0.1085
PctService	1	-0.00021	0.00449	0.0023	0.9619
PctSalesOf	1	0.00347	0.00475	0.5329	0.4654
PctTransOc	1	-0.00010	0.00500	0.0004	0.9848
PctConsOcc	1	-0.00484	0.00483	1.0045	0.3162
lPCI	1	0.2118	0.1383	2.3469	0.1255
PctFemale	1	-0.0194	0.00694	7.7761	0.0053
PctInInsti	1	-0.0108	0.00671	2.6058	0.1065
occhdn	1	-0.00012	0.000072	2.6310	0.1048
PctOccupie	1	0.00967	0.00391	6.1371	0.0132
PctPlumbin	1	0.0354	0.00967	13.4338	0.0002
PctNoPhone	1	0.00512	0.00642	0.6366	0.4250
PctNoCars	1	0.00297	0.00619	0.2303	0.6313
PctAgeUnit	1	-0.00010	0.00540	0.0003	0.9853
PctAgeUn15	1	-0.00444	0.00592	0.5623	0.4533
PctBuiltBe	1	-0.00107	0.00530	0.0410	0.8396
lavgage	1	0.5045	0.2890	3.0477	0.0809
lAvgrent	1	0.1643	0.0971	2.8640	0.0906
lAvghval	1	0.6280	0.0782	64.5710	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lpopden	2.566	2.355	2.796
lland	2.617	2.438	2.809
lpophous	0.784	0.513	1.200

long	1.008	0.998	1.018
hland	1.000	0.986	1.015
lslopesd	0.882	0.764	1.019

statsig binary model, 1201, lpophous 204

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lelevrang	1.146	0.991	1.326
lctimn	0.998	0.828	1.203
M0	1.065	0.837	1.355
M1	0.822	0.663	1.019
M4	1.017	0.880	1.174
M7	0.549	0.309	0.977
M8	1.028	0.664	1.590
M11	0.091	0.004	2.053
e31	1.082	1.055	1.109
e44	1.003	0.990	1.016
e54	1.111	1.068	1.156
e56	0.943	0.898	0.991
e61	0.956	0.786	1.164
e62	1.015	0.985	1.047
e72	1.014	0.993	1.035
e81	1.060	1.025	1.095
S1	1.090	0.778	1.526
S4	1.163	0.561	2.409
S5	1.285	0.950	1.737
S6	1.334	0.882	2.018
S8	1.925	1.242	2.983
S9	8.166	1.031	64.694
S12	1.475	0.838	2.598
S15	0.002	<0.001	0.013
S16	0.453	0.294	0.696
S17	0.523	0.410	0.666
S18	0.431	0.318	0.584
S19	0.213	0.162	0.280
S20	0.476	0.357	0.635
S22	1.234	0.856	1.780
S23	1.832	1.050	3.195
S24	0.629	0.384	1.029
S26	0.660	0.478	0.912
S27	0.326	0.245	0.434
S28	0.640	0.449	0.912
S29	0.857	0.667	1.103
S31	0.577	0.430	0.774
S32	0.288	0.132	0.630
S33	1.687	0.786	3.622
S34	1.732	0.630	4.758
S35	0.665	0.413	1.071
S36	2.269	1.516	3.394
S37	1.099	0.743	1.625
S38	0.650	0.466	0.907

statsig binary model, 1201, lpophous 205

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
S39	1.044	0.739	1.474
S40	1.025	0.765	1.373
S41	1.782	1.114	2.852
S42	0.395	0.285	0.546
S44	1.235	0.229	6.666
S45	0.958	0.611	1.502
S46	0.574	0.408	0.807
S47	0.884	0.630	1.242
S49	0.374	0.218	0.642
S50	0.732	0.441	1.215
S51	0.535	0.389	0.736
S54	0.562	0.415	0.762
S55	0.757	0.547	1.047

S56	1.297	0.735	2.289
PctUrban	1.002	0.999	1.005
PctOnFarms	0.993	0.986	1.000
cerate00	1.038	0.985	1.095
crhc00	0.707	0.161	3.118
PctAge55_5	1.015	0.996	1.033
PctAge60_6	0.997	0.978	1.018
PctAge65_7	1.012	0.996	1.027
PctAge75_8	0.994	0.975	1.014
PctOver85	1.002	0.971	1.035
PctBlack1	0.994	0.990	0.998
PctIndian1	0.991	0.984	0.997
PctHawnPI1	1.007	0.915	1.108
PctOther1	1.007	0.994	1.021
PctEnglis2	1.016	1.003	1.028
PctNoEngli	1.016	0.970	1.065
PctSomeHig	1.019	1.006	1.033
PctHighSch	1.028	1.017	1.038
PctSomeCol	1.035	1.024	1.046
PctBachelo	1.036	1.021	1.051
PctGradPro	1.022	1.003	1.041
PctArmedFo	2.277	1.177	4.403
PctCivLabF	2.333	1.194	4.558
PctNotInLF	2.329	1.192	4.550
PctManufac	0.992	0.985	1.000
PctRetailT	0.998	0.987	1.009
PctEducati	0.991	0.981	1.001
PctHealthS	0.992	0.982	1.002
PctService	1.000	0.991	1.009
PctSalesOf	1.003	0.994	1.013
PctTransOc	1.000	0.990	1.010

statsig binary model, 1201, lpophous 206

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PctConsOcc	0.995	0.986	1.005
lPCI	1.236	0.943	1.621
PctFemale	0.981	0.968	0.994
PctInInsti	0.989	0.976	1.002
occhdn	1.000	1.000	1.000
PctOccupie	1.010	1.002	1.017
PctPlumbin	1.036	1.017	1.056
PctNoPhone	1.005	0.993	1.018
PctNoCars	1.003	0.991	1.015
PctAgeUnit	1.000	0.989	1.011
PctAgeUn15	0.996	0.984	1.007
PctBuiltBe	0.999	0.989	1.009
lavgage	1.656	0.940	2.918
lAvgrent	1.179	0.974	1.426
lAvghval	1.874	1.608	2.184

Association of Predicted Probabilities and Observed Responses

Percent Concordant	92.1	Somers' D	0.843
Percent Discordant	7.8	Gamma	0.844
Percent Tied	0.1	Tau-a	0.256
Pairs	117368124	c	0.921

statsig binary model, 1202, lpophous 207

The LOGISTIC Procedure

Model Information

Data Set	WORK.PROBSTATB
Response Variable	b1202
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	27812
Number of Observations Used	27811

Response Profile

Ordered Value	b1202	Total Frequency
1	1	24893
2	0	2918

Probability modeled is b1202=1.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	18677.984	11421.930
SC	18686.217	12327.581
-2 Log L	18675.984	11201.930

R-Square 0.2357 Max-rescaled R-Square 0.4819

statsig binary model, 1202, lpophous 208

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	7474.0538	109	<.0001
Score	7196.5268	109	<.0001
Wald	2596.2961	109	<.0001

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
Intercept	1	-106.4	37.8244	7.9079	0.0049
lpopden	1	0.8111	0.0510	253.3306	<.0001
lland	1	0.9418	0.0420	501.9759	<.0001
lpophous	1	-0.3189	0.2235	2.0368	0.1535
long	1	0.0101	0.00638	2.4884	0.1147
hland	1	0.0159	0.00895	3.1630	0.0753
lslopesd	1	0.0124	0.0863	0.0207	0.8857
lelevrang	1	0.0663	0.0880	0.5676	0.4512
lctimn	1	0.1727	0.1168	2.1859	0.1393
M0	1	0.2062	0.1519	1.8423	0.1747
M1	1	-0.3374	0.1291	6.8358	0.0089
M4	1	-0.1734	0.0877	3.9104	0.0480
M7	1	-0.2161	0.3974	0.2956	0.5866
M8	1	-0.1622	0.2562	0.4007	0.5267
M11	1	-4.9608	2.0033	6.1323	0.0133
e31	1	0.0678	0.0196	11.9933	0.0005
e44	1	0.0295	0.0115	6.6309	0.0100
e54	1	0.0953	0.0314	9.1934	0.0024
e56	1	0.0508	0.0406	1.5685	0.2104
e61	1	0.0187	0.1527	0.0149	0.9028
e62	1	-0.00857	0.0187	0.2109	0.6461
e72	1	0.0744	0.0180	17.1062	<.0001
e81	1	0.0470	0.0251	3.5172	0.0607
S1	1	0.1024	0.2105	0.2367	0.6266
S4	1	0.0269	0.4549	0.0035	0.9529
S5	1	0.0643	0.1738	0.1370	0.7112
S6	1	0.5909	0.2693	4.8126	0.0283
S8	1	0.6921	0.2753	6.3199	0.0119

S9	1	0.1617	0.6972	0.0538	0.8166
S12	1	1.3524	0.5395	6.2831	0.0122
S15	1	1.9730	2.2853	0.7454	0.3879
S16	1	-0.4044	0.2558	2.5005	0.1138
S17	1	-0.2628	0.1479	3.1562	0.0756
S18	1	-0.6382	0.1917	11.0856	0.0009
S19	1	-0.8049	0.1550	26.9771	<.0001

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S20	1	-0.3363	0.1618	4.3192	0.0377
S22	1	0.3660	0.2307	2.5168	0.1126
S23	1	-0.1387	0.3345	0.1719	0.6784
S24	1	0.9882	0.4331	5.2059	0.0225
S26	1	0.0967	0.2316	0.1745	0.6761
S27	1	-0.9174	0.1634	31.5403	<.0001
S28	1	-0.1566	0.2199	0.5075	0.4762
S29	1	-0.1273	0.1470	0.7500	0.3865
S31	1	-0.4873	0.1652	8.7068	0.0032
S32	1	-1.8705	0.4221	19.6355	<.0001
S33	1	-0.5585	0.4008	1.9413	0.1635
S34	1	1.5453	1.0424	2.1977	0.1382
S35	1	-0.1624	0.2748	0.3491	0.5546
S36	1	0.7432	0.2589	8.2388	0.0041
S37	1	0.4403	0.2746	2.5712	0.1088
S38	1	-0.5497	0.1750	9.8713	0.0017
S39	1	0.2760	0.2391	1.3319	0.2485
S40	1	0.1928	0.1700	1.2869	0.2566
S41	1	0.5135	0.2963	3.0046	0.0830
S42	1	-0.8401	0.1972	18.1446	<.0001
S44	1	-0.5143	0.8658	0.3529	0.5525
S45	1	-0.3421	0.2794	1.4996	0.2207
S46	1	-0.2198	0.1867	1.3851	0.2392
S47	1	0.9980	0.2633	14.3622	0.0002
S49	1	-0.7830	0.2996	6.8290	0.0090
S50	1	-0.0938	0.3288	0.0814	0.7754
S51	1	-0.4323	0.1997	4.6888	0.0304
S54	1	0.0870	0.1783	0.2381	0.6256
S55	1	-0.5832	0.1961	8.8447	0.0029
S56	1	0.3747	0.3372	1.2345	0.2665
PctUrban	1	0.00125	0.00173	0.5193	0.4711
PctOnFarms	1	-0.00895	0.00358	6.2413	0.0125
cerate01	1	0.0250	0.0285	0.7710	0.3799
crhc01	1	-0.00975	0.8742	0.0001	0.9911
PctAge55_5	1	0.0287	0.00955	9.0432	0.0026
PctAge60_6	1	8.017E-6	0.0105	0.0000	0.9994
PctAge65_7	1	0.0186	0.00787	5.5619	0.0184
PctAge75_8	1	-0.00931	0.0101	0.8465	0.3576
PctOver85	1	-0.00369	0.0166	0.0496	0.8238
PctBlack1	1	-0.00053	0.00246	0.0468	0.8286
PctIndian1	1	-0.00747	0.00366	4.1546	0.0415
PctHawnPI1	1	-0.1520	0.0703	4.6736	0.0306
PctOther1	1	0.00304	0.00774	0.1540	0.6948
PctEnglis2	1	0.0188	0.00660	8.1186	0.0044

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
PctNoEngli	1	-0.00128	0.0274	0.0022	0.9629
PctSomeHig	1	0.00744	0.00700	1.1296	0.2879
PctHighSch	1	0.0202	0.00554	13.3011	0.0003
PctSomeCol	1	0.0232	0.00584	15.7520	<.0001
PctBachelo	1	0.0210	0.00776	7.3470	0.0067
PctGradPro	1	0.0195	0.0102	3.6574	0.0558
PctArmedFo	1	0.9067	0.3726	5.9213	0.0150
PctCivLabF	1	0.9239	0.3778	5.9807	0.0145
PctNotInLF	1	0.9156	0.3778	5.8745	0.0154
PctManufac	1	-0.00314	0.00411	0.5825	0.4454

PctRetailT	1	-0.00021	0.00563	0.0014	0.9706
PctEducati	1	-0.00954	0.00531	3.2305	0.0723
PctHealthS	1	8.601E-6	0.00516	0.0000	0.9987
PctService	1	0.000805	0.00466	0.0298	0.8629
PctSalesOf	1	0.000154	0.00487	0.0010	0.9748
PctTransOc	1	-0.00070	0.00516	0.0184	0.8922
PctConsOcc	1	-0.00011	0.00497	0.0005	0.9829
lPCI	1	0.4248	0.1469	8.3627	0.0038
PctFemale	1	-0.00333	0.00712	0.2183	0.6403
PctInInsti	1	0.00328	0.00761	0.1863	0.6660
occhdn	1	-0.00014	0.000080	3.1203	0.0773
PctOccupie	1	0.0120	0.00410	8.5834	0.0034
PctPlumbin	1	0.0139	0.0104	1.7819	0.1819
PctNoPhone	1	0.0103	0.00662	2.4286	0.1191
PctNoCars	1	0.000480	0.00651	0.0054	0.9413
PctAgeUnit	1	0.0127	0.00609	4.3274	0.0375
PctAgeUn15	1	-0.0205	0.00622	10.8281	0.0010
PctBuiltBe	1	0.00947	0.00554	2.9234	0.0873
lavgage	1	1.0855	0.3256	11.1162	0.0009
lAvgrent	1	0.1396	0.0994	1.9700	0.1604
lAvghval	1	0.3694	0.0811	20.7478	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lpopden	2.250	2.036	2.487
lland	2.564	2.362	2.785
lpophous	0.727	0.469	1.126
long	1.010	0.998	1.023
hland	1.016	0.998	1.034
lslopesd	1.012	0.855	1.199

statsig binary model, 1202, lpophous 211

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lelevrang	1.069	0.899	1.270
lctimn	1.189	0.945	1.494
M0	1.229	0.913	1.655
M1	0.714	0.554	0.919
M4	0.841	0.708	0.998
M7	0.806	0.370	1.756
M8	0.850	0.515	1.405
M11	0.007	<0.001	0.355
e31	1.070	1.030	1.112
e44	1.030	1.007	1.053
e54	1.100	1.034	1.170
e56	1.052	0.972	1.139
e61	1.019	0.755	1.374
e62	0.991	0.956	1.028
e72	1.077	1.040	1.116
e81	1.048	0.998	1.101
S1	1.108	0.733	1.674
S4	1.027	0.421	2.505
S5	1.066	0.759	1.499
S6	1.806	1.065	3.061
S8	1.998	1.165	3.427
S9	1.175	0.300	4.610
S12	3.867	1.343	11.132
S15	7.192	0.082	633.973
S16	0.667	0.404	1.102
S17	0.769	0.575	1.027
S18	0.528	0.363	0.769
S19	0.447	0.330	0.606
S20	0.714	0.520	0.981
S22	1.442	0.917	2.266
S23	0.871	0.452	1.677
S24	2.686	1.149	6.278
S26	1.102	0.700	1.734
S27	0.400	0.290	0.550
S28	0.855	0.556	1.316
S29	0.880	0.660	1.175
S31	0.614	0.444	0.849

S32	0.154	0.067	0.352
S33	0.572	0.261	1.255
S34	4.689	0.608	36.174
S35	0.850	0.496	1.457
S36	2.103	1.266	3.493
S37	1.553	0.907	2.660
S38	0.577	0.410	0.813

statsig binary model, 1202, lpophous 212

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
S39	1.318	0.825	2.106
S40	1.213	0.869	1.692
S41	1.671	0.935	2.987
S42	0.432	0.293	0.635
S44	0.598	0.110	3.263
S45	0.710	0.411	1.228
S46	0.803	0.557	1.157
S47	2.713	1.619	4.545
S49	0.457	0.254	0.822
S50	0.910	0.478	1.734
S51	0.649	0.439	0.960
S54	1.091	0.769	1.547
S55	0.558	0.380	0.820
S56	1.455	0.751	2.817
PctUrban	1.001	0.998	1.005
PctOnFarms	0.991	0.984	0.998
cerate01	1.025	0.970	1.084
crhc01	0.990	0.179	5.494
PctAge55_5	1.029	1.010	1.049
PctAge60_6	1.000	0.980	1.021
PctAge65_7	1.019	1.003	1.035
PctAge75_8	0.991	0.971	1.011
PctOver85	0.996	0.964	1.029
PctBlack1	0.999	0.995	1.004
PctIndian1	0.993	0.985	1.000
PctHawnPI1	0.859	0.748	0.986
PctOther1	1.003	0.988	1.018
PctEnglis2	1.019	1.006	1.032
PctNoEngli	0.999	0.947	1.054
PctSomeHig	1.007	0.994	1.021
PctHighSch	1.020	1.009	1.032
PctSomeCol	1.023	1.012	1.035
PctBachelo	1.021	1.006	1.037
PctGradPro	1.020	1.000	1.040
PctArmedFo	2.476	1.193	5.140
PctCivLabF	2.519	1.201	5.282
PctNotInLF	2.498	1.191	5.238
PctManufac	0.997	0.989	1.005
PctRetailT	1.000	0.989	1.011
PctEducati	0.991	0.980	1.001
PctHealthS	1.000	0.990	1.010
PctService	1.001	0.992	1.010
PctSalesOf	1.000	0.991	1.010
PctTransOc	0.999	0.989	1.009

statsig binary model, 1202, lpophous 213

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PctConsOcc	1.000	0.990	1.010
lPCI	1.529	1.147	2.039
PctFemale	0.997	0.983	1.011
PctInInsti	1.003	0.988	1.018
occhdn	1.000	1.000	1.000
PctOccupie	1.012	1.004	1.020
PctPlumbin	1.014	0.994	1.035
PctNoPhone	1.010	0.997	1.024
PctNoCars	1.000	0.988	1.013
PctAgeUnit	1.013	1.001	1.025

PctAgeUn15	0.980	0.968	0.992
PctBuiltBe	1.010	0.999	1.021
lavgage	2.961	1.564	5.604
lAvgrent	1.150	0.946	1.397
lAvghval	1.447	1.234	1.696

Association of Predicted Probabilities and Observed Responses

Percent Concordant	91.9	Somers' D	0.840
Percent Discordant	7.9	Gamma	0.841
Percent Tied	0.1	Tau-a	0.158
Pairs	72637774	c	0.920

The LOGISTIC Procedure

Model Information

Data Set	WORK.PROBSTATB
Response Variable	b1203
Number of Response Levels	2
Model	binary logit
Optimization Technique	Fisher's scoring

Number of Observations Read	27812
Number of Observations Used	27811

Response Profile

Ordered Value	b1203	Total Frequency
1	1	26198
2	0	1613

Probability modeled is b1203=1.

NOTE: 1 observation was deleted due to missing values for the response or explanatory variables.

Model Convergence Status

Convergence criterion (GCONV=1E-8) satisfied.

Model Fit Statistics

Criterion	Intercept Only	Intercept and Covariates
AIC	12318.091	8028.505
SC	12326.324	8934.155
-2 Log L	12316.091	7808.505

R-Square 0.1496 Max-rescaled R-Square 0.4182

The LOGISTIC Procedure

Testing Global Null Hypothesis: BETA=0

Test	Chi-Square	DF	Pr > ChiSq
Likelihood Ratio	4507.5860	109	<.0001
Score	4730.7469	109	<.0001
Wald	1796.4013	109	<.0001

Analysis of Maximum Likelihood Estimates

Standard Wald

Parameter	DF	Estimate	Error	Chi-Square	Pr > ChiSq
Intercept	1	-74.6829	40.8883	3.3361	0.0678
lpopden	1	0.8978	0.0592	229.9804	<.0001
lland	1	1.0828	0.0487	493.7814	<.0001
lpophous	1	-0.5382	0.2445	4.8481	0.0277
long	1	0.0199	0.00752	7.0203	0.0081
hland	1	0.00715	0.0107	0.4489	0.5029
lslopesd	1	0.2230	0.1070	4.3444	0.0371
lelevrang	1	-0.2107	0.1095	3.7030	0.0543
lctimn	1	0.2769	0.1476	3.5193	0.0607
M0	1	0.2279	0.1952	1.3622	0.2431
M1	1	-0.0368	0.1649	0.0499	0.8233
M4	1	-0.1486	0.1082	1.8851	0.1698
M7	1	-0.0934	0.4693	0.0396	0.8422
M8	1	-0.2563	0.3008	0.7262	0.3941
M11	1	12.0026	1537.3	0.0001	0.9938
e31	1	0.0386	0.0254	2.3164	0.1280
e44	1	0.0533	0.0159	11.2054	0.0008
e54	1	0.0266	0.0374	0.5062	0.4768
e56	1	0.0882	0.0546	2.6102	0.1062
e61	1	0.1426	0.2070	0.4748	0.4908
e62	1	-0.00981	0.0232	0.1794	0.6719
e72	1	0.0436	0.0227	3.6777	0.0551
e81	1	0.0697	0.0346	4.0532	0.0441
S1	1	-0.4561	0.2480	3.3839	0.0658
S4	1	0.1340	0.5153	0.0676	0.7949
S5	1	0.2733	0.2209	1.5302	0.2161
S6	1	0.7581	0.3007	6.3539	0.0117
S8	1	0.4121	0.3274	1.5841	0.2082
S9	1	1.2790	1.1561	1.2240	0.2686
S12	1	1.4229	0.7434	3.6637	0.0556
S15	1	2.6219	2.4690	1.1276	0.2883
S16	1	0.1182	0.3063	0.1489	0.6996
S17	1	-0.2909	0.1835	2.5129	0.1129
S18	1	-0.1779	0.2788	0.4069	0.5235
S19	1	-0.9604	0.1844	27.1402	<.0001

statsig binary model, 1203, lpophous 216

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
S20	1	0.9363	0.2387	15.3904	<.0001
S22	1	0.7874	0.3269	5.8015	0.0160
S23	1	-1.0185	0.3799	7.1880	0.0073
S24	1	0.2969	0.4574	0.4212	0.5163
S26	1	0.0500	0.3103	0.0260	0.8720
S27	1	-0.9144	0.1951	21.9694	<.0001
S28	1	0.3904	0.3214	1.4752	0.2245
S29	1	-0.4084	0.1731	5.5634	0.0183
S31	1	-0.4125	0.1985	4.3179	0.0377
S32	1	-1.1968	0.4638	6.6572	0.0099
S33	1	-0.3136	0.5343	0.3444	0.5573
S34	1	1.1643	1.0582	1.2106	0.2712
S35	1	-0.1507	0.3174	0.2254	0.6349
S36	1	0.7614	0.3246	5.5021	0.0190
S37	1	0.0630	0.3217	0.0384	0.8446
S38	1	0.4245	0.2210	3.6904	0.0547
S39	1	0.6868	0.3791	3.2811	0.0701
S40	1	0.1893	0.2085	0.8238	0.3641
S41	1	1.0833	0.3623	8.9404	0.0028
S42	1	-0.5639	0.2405	5.4972	0.0190
S44	1	-1.0638	0.8793	1.4638	0.2263
S45	1	-0.1482	0.3658	0.1641	0.6854
S46	1	-0.3782	0.2148	3.1005	0.0783
S47	1	0.9510	0.3803	6.2524	0.0124
S49	1	-0.1019	0.3503	0.0846	0.7711
S50	1	0.7197	0.4768	2.2781	0.1312
S51	1	-0.2469	0.2563	0.9282	0.3353
S54	1	-0.1705	0.2089	0.6658	0.4145
S55	1	-0.3143	0.2667	1.3884	0.2387
S56	1	0.6289	0.4283	2.1559	0.1420
PctUrban	1	-0.00331	0.00196	2.8438	0.0917
PctOnFarms	1	-0.00021	0.00411	0.0026	0.9597
cerate02	1	-0.00645	0.00254	6.4611	0.0110

crhc02	1	-0.8239	0.4800	2.9461	0.0861
PctAge55_5	1	0.0304	0.0107	8.0285	0.0046
PctAge60_6	1	0.0141	0.0116	1.4569	0.2274
PctAge65_7	1	0.0168	0.00858	3.8467	0.0498
PctAge75_8	1	0.00352	0.0113	0.0966	0.7560
PctOver85	1	-0.00033	0.0189	0.0003	0.9862
PctBlack1	1	-0.00182	0.00296	0.3766	0.5394
PctIndian1	1	-0.00409	0.00418	0.9586	0.3276
PctHawnPI1	1	-0.1497	0.0765	3.8308	0.0503
PctOther1	1	-0.00239	0.00904	0.0697	0.7917
PctEnglis2	1	0.0239	0.00746	10.2913	0.0013

The LOGISTIC Procedure

Analysis of Maximum Likelihood Estimates

Parameter	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq
PctNoEngli	1	0.0290	0.0355	0.6681	0.4137
PctSomeHig	1	0.0161	0.00793	4.1316	0.0421
PctHighSch	1	0.0195	0.00631	9.5107	0.0020
PctSomeCol	1	0.0243	0.00668	13.2478	0.0003
PctBachelo	1	0.0246	0.00880	7.7930	0.0052
PctGradPro	1	0.0275	0.0116	5.6122	0.0178
PctArmedFo	1	0.5961	0.4027	2.1910	0.1388
PctCivLabF	1	0.6068	0.4081	2.2105	0.1371
PctNotInLF	1	0.5989	0.4082	2.1531	0.1423
PctManufac	1	-0.00518	0.00471	1.2074	0.2718
PctRetailT	1	-0.00860	0.00627	1.8834	0.1700
PctEducati	1	-0.00256	0.00587	0.1907	0.6624
PctHealthS	1	0.0113	0.00581	3.7705	0.0522
PctService	1	0.00169	0.00516	0.1077	0.7428
PctSalesOf	1	0.00383	0.00541	0.5014	0.4789
PctTransOc	1	0.00966	0.00577	2.8014	0.0942
PctConsOcc	1	0.0100	0.00555	3.2574	0.0711
lPCI	1	0.3755	0.1686	4.9604	0.0259
PctFemale	1	-0.00448	0.00791	0.3202	0.5715
PctInInsti	1	0.00456	0.00892	0.2614	0.6092
occhdn	1	-0.00014	0.000069	4.4238	0.0354
PctOccupie	1	0.0160	0.00459	12.2301	0.0005
PctPlumbin	1	0.0220	0.0116	3.6080	0.0575
PctNoPhone	1	0.00679	0.00741	0.8382	0.3599
PctNoCars	1	-0.0154	0.00720	4.5436	0.0330
PctAgeUnit	1	0.0147	0.00699	4.4268	0.0354
PctAgeUn15	1	-0.0132	0.00702	3.5607	0.0592
PctBuiltBe	1	-0.00387	0.00635	0.3718	0.5420
lavgage	1	1.6067	0.3617	19.7312	<.0001
lAvgrent	1	0.1595	0.1133	1.9828	0.1591
lAvghval	1	0.4306	0.0912	22.2659	<.0001

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lpopden	2.454	2.185	2.756
lland	2.953	2.684	3.249
lpophous	0.584	0.362	0.943
long	1.020	1.005	1.035
hland	1.007	0.986	1.028
lslopesd	1.250	1.013	1.542

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
lelevrang	0.810	0.654	1.004
lctimn	1.319	0.988	1.762
M0	1.256	0.857	1.841
M1	0.964	0.698	1.331
M4	0.862	0.697	1.066
M7	0.911	0.363	2.285

M8	0.774	0.429	1.395
M11	>999.999	<0.001	>999.999
e31	1.039	0.989	1.092
e44	1.055	1.022	1.088
e54	1.027	0.954	1.105
e56	1.092	0.981	1.216
e61	1.153	0.769	1.730
e62	0.990	0.946	1.036
e72	1.045	0.999	1.092
e81	1.072	1.002	1.147
S1	0.634	0.390	1.030
S4	1.143	0.416	3.139
S5	1.314	0.852	2.027
S6	2.134	1.184	3.848
S8	1.510	0.795	2.869
S9	3.593	0.373	34.639
S12	4.149	0.966	17.811
S15	13.761	0.109	>999.999
S16	1.125	0.617	2.051
S17	0.748	0.522	1.071
S18	0.837	0.485	1.446
S19	0.383	0.267	0.549
S20	2.551	1.598	4.072
S22	2.198	1.158	4.171
S23	0.361	0.172	0.760
S24	1.346	0.549	3.298
S26	1.051	0.572	1.931
S27	0.401	0.273	0.587
S28	1.478	0.787	2.774
S29	0.665	0.473	0.933
S31	0.662	0.449	0.977
S32	0.302	0.122	0.750
S33	0.731	0.256	2.083
S34	3.204	0.403	25.491
S35	0.860	0.462	1.602
S36	2.141	1.133	4.045
S37	1.065	0.567	2.001
S38	1.529	0.991	2.357

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
S39	1.987	0.945	4.178
S40	1.208	0.803	1.818
S41	2.955	1.452	6.010
S42	0.569	0.355	0.912
S44	0.345	0.062	1.934
S45	0.862	0.421	1.766
S46	0.685	0.450	1.044
S47	2.588	1.228	5.454
S49	0.903	0.454	1.795
S50	2.054	0.807	5.229
S51	0.781	0.473	1.291
S54	0.843	0.560	1.270
S55	0.730	0.433	1.232
S56	1.876	0.810	4.342
PctUrban	0.997	0.993	1.001
PctOnFarms	1.000	0.992	1.008
cerate02	0.994	0.989	0.999
crhc02	0.439	0.171	1.124
PctAge55_5	1.031	1.009	1.053
PctAge60_6	1.014	0.991	1.038
PctAge65_7	1.017	1.000	1.034
PctAge75_8	1.004	0.981	1.026
PctOver85	1.000	0.963	1.037
PctBlack1	0.998	0.992	1.004
PctIndian1	0.996	0.988	1.004
PctHawnPI1	0.861	0.741	1.000
PctOther1	0.998	0.980	1.015
PctEnglis2	1.024	1.009	1.039
PctNoEngli	1.029	0.960	1.104
PctSomeHig	1.016	1.001	1.032
PctHighSch	1.020	1.007	1.032
PctSomeCol	1.025	1.011	1.038
PctBachelo	1.025	1.007	1.043

PctGradPro	1.028	1.005	1.052
PctArmedFo	1.815	0.824	3.996
PctCivLabF	1.835	0.824	4.083
PctNotInLF	1.820	0.818	4.051
PctManufac	0.995	0.986	1.004
PctRetailT	0.991	0.979	1.004
PctEducati	0.997	0.986	1.009
PctHealthS	1.011	1.000	1.023
PctService	1.002	0.992	1.012
PctSalesOf	1.004	0.993	1.015
PctTransOc	1.010	0.998	1.021

statsig binary model, 1203, lpophous 220

The LOGISTIC Procedure

Odds Ratio Estimates

Effect	Point Estimate	95% Wald Confidence Limits	
PctConsOcc	1.010	0.999	1.021
lPCI	1.456	1.046	2.026
PctFemale	0.996	0.980	1.011
PctInInsti	1.005	0.987	1.022
occhdn	1.000	1.000	1.000
PctOccupie	1.016	1.007	1.025
PctPlumbin	1.022	0.999	1.046
PctNoPhone	1.007	0.992	1.022
PctNoCars	0.985	0.971	0.999
PctAgeUnit	1.015	1.001	1.029
PctAgeUn15	0.987	0.973	1.001
PctBuiltBe	0.996	0.984	1.009
lavgage	4.986	2.454	10.131
lAvgrent	1.173	0.939	1.464
lAvghval	1.538	1.286	1.839

Association of Predicted Probabilities and Observed Responses

Percent Concordant	92.0	Somers' D	0.842
Percent Discordant	7.8	Gamma	0.844
Percent Tied	0.3	Tau-a	0.092
Pairs	42257374	c	0.921