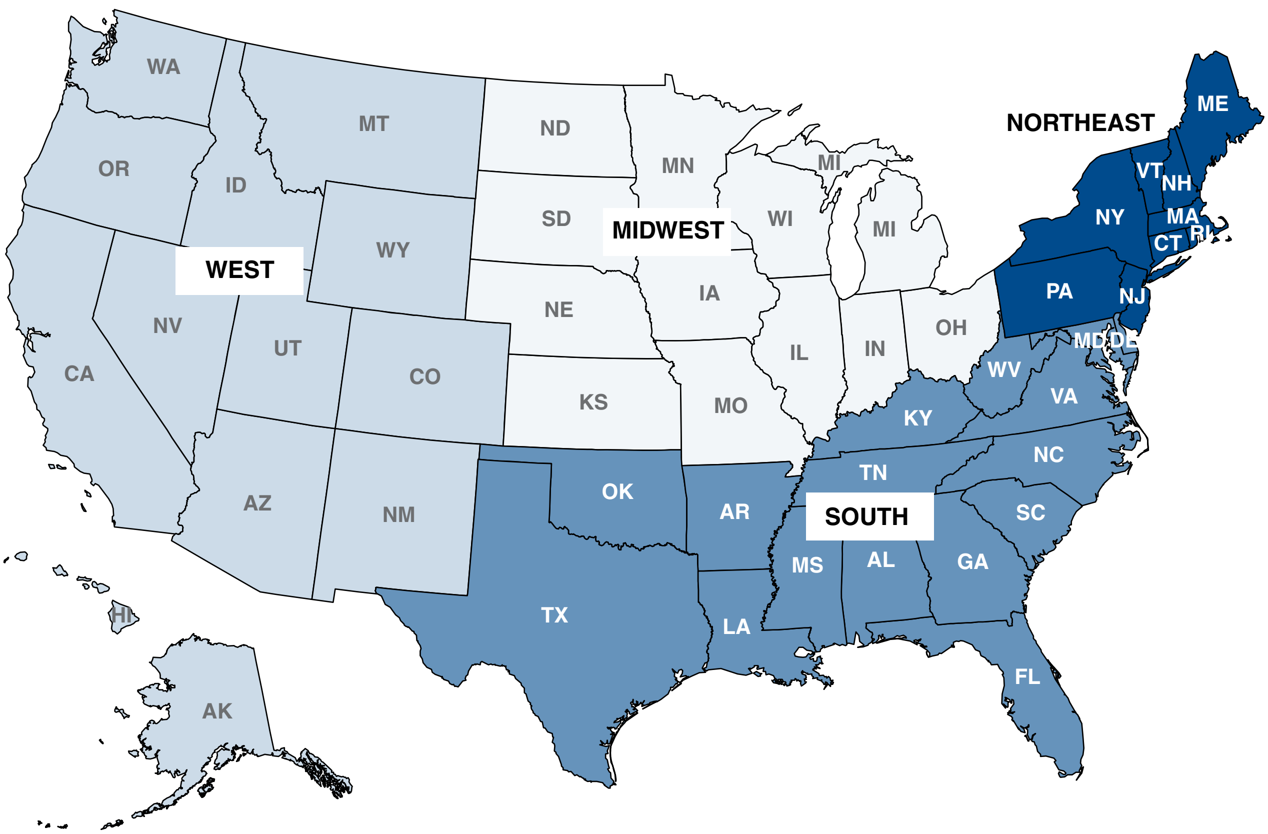
Appendix A: Bureau of the Census Regions

Figure A-1

Bureau of the Census Regions



Appendix B: Data Used in the Study

Agricultural Resources Management Survey

Most of the farm-level data used in the analysis are from the 2007 Agricultural Resources Management Survey (ARMS), conducted annually by USDA's Economic Research Service and National Agricultural Statistics Service. The survey collects information needed to measure the financial condition (farm income, expenses, assets, and debts) and operating characteristics of farm businesses, the cost of producing agricultural commodities, and the well-being of farm operator households. The target population of the survey is operators of farm businesses representing agricultural production in the 48 contiguous States. A farm is defined as an establishment that sold or normally would have sold at least $1,000 of agricultural products during the year. Farms can be organized as proprietorships, partnerships, family corporations, nonfamily corporations, or cooperatives. Data are collected from one operator per farm, the senior farm operator. A senior farm operator is the operator who makes most of the day-to-day management decisions. For this study, operator households organized as nonfamily corporations or cooperatives and farms run by hired managers are excluded.

June Agricultural Survey Data

The June Agricultural Survey is conducted by USDA's National Agricultural Statistics Service every year to provide estimates of farm numbers and land in farms, crop acres planted, grains and oilseeds in storage, livestock inventories, and land values. In 1997, 1999, 2001, 2003, 2005, and 2007, the surveys included questions about Internet access. In 2005 and 2007, the questions addressed broadband Internet access. The 2007 computer usage estimates are based on responses from over 31,400 agricultural operations and represent all sizes and types of farms.

Federal Communications Commission Form-477 Data

The Federal Communications Commission collects data from Internet service providers twice annually: in June and December. The data have been collected since December 1999 through what is called the FCC Form-477. The data collected include number of lines, various company characteristics, and most importantly where broadband is currently provided. Initially, very small providers were not required to file the form, but eventually all providers were required to return the form to the FCC every 6 months. On the form, a company specifies each ZIP Code in which it has customers. Recently, the FCC has started to collect more detailed information on type of broadband service and location of broadband service provision.

PEW Surveys

PEW data are collected by Princeton Survey Research Associates. Sample sizes for each survey are around 2,500, with total sampling error of +/-2 percent and sampling error of +/-3 percent for Internet users. Data were collected from various surveys and reports available from PEW Internet & American Life Project's website (http://www.pewinternet.org/index.asp). Urban, as defined by the U.S. Census Bureau, includes any population concentration of 2,500 or more people.

Current Population Survey

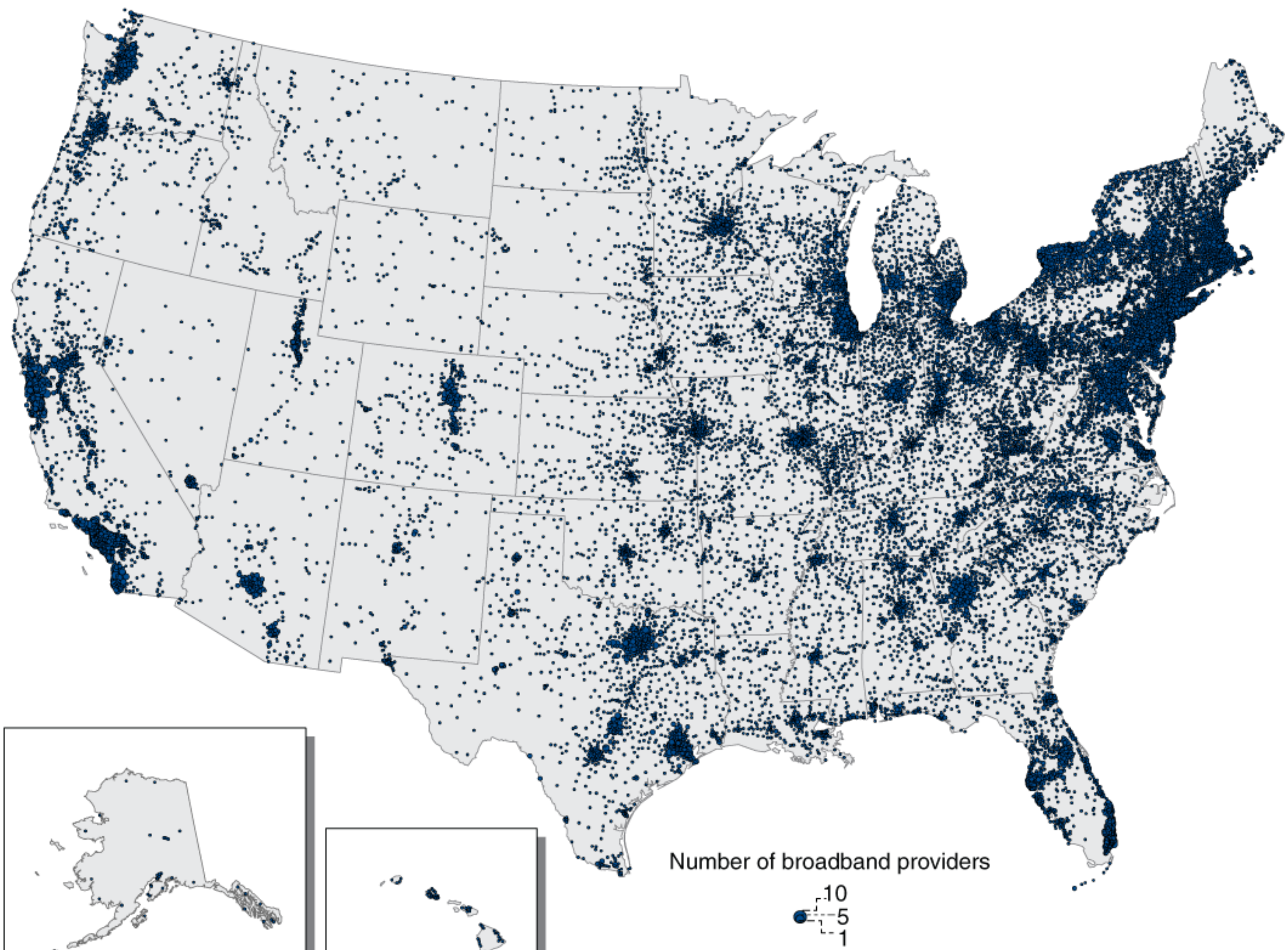
The U.S. Census Bureau conducts monthly Current Population Surveys. The surveys constitute 57,000 households containing 134,000 persons. On an irregular basis, the Bureau's monthly survey includes questions on Internet and related technologies. These special surveys provide broad-based and statistically reliable information on the ways that information technologies, in general, and broadband more specifically are transforming the way we live, work, and learn. As of this writing, the last survey on information technology took place in October 2007, but only asked four Internet-related questions, whereas an October 2003 survey asked detailed questions on what, where, and how the Internet was used. Estimation conducted in this report had a sampling error no greater than +/-2 percent. Urban, as defined by the U.S. Census Bureau, includes any population concentration of 2,500 or more people.

Appendix C: Using the FCC Data

Broadband Availability Over Time

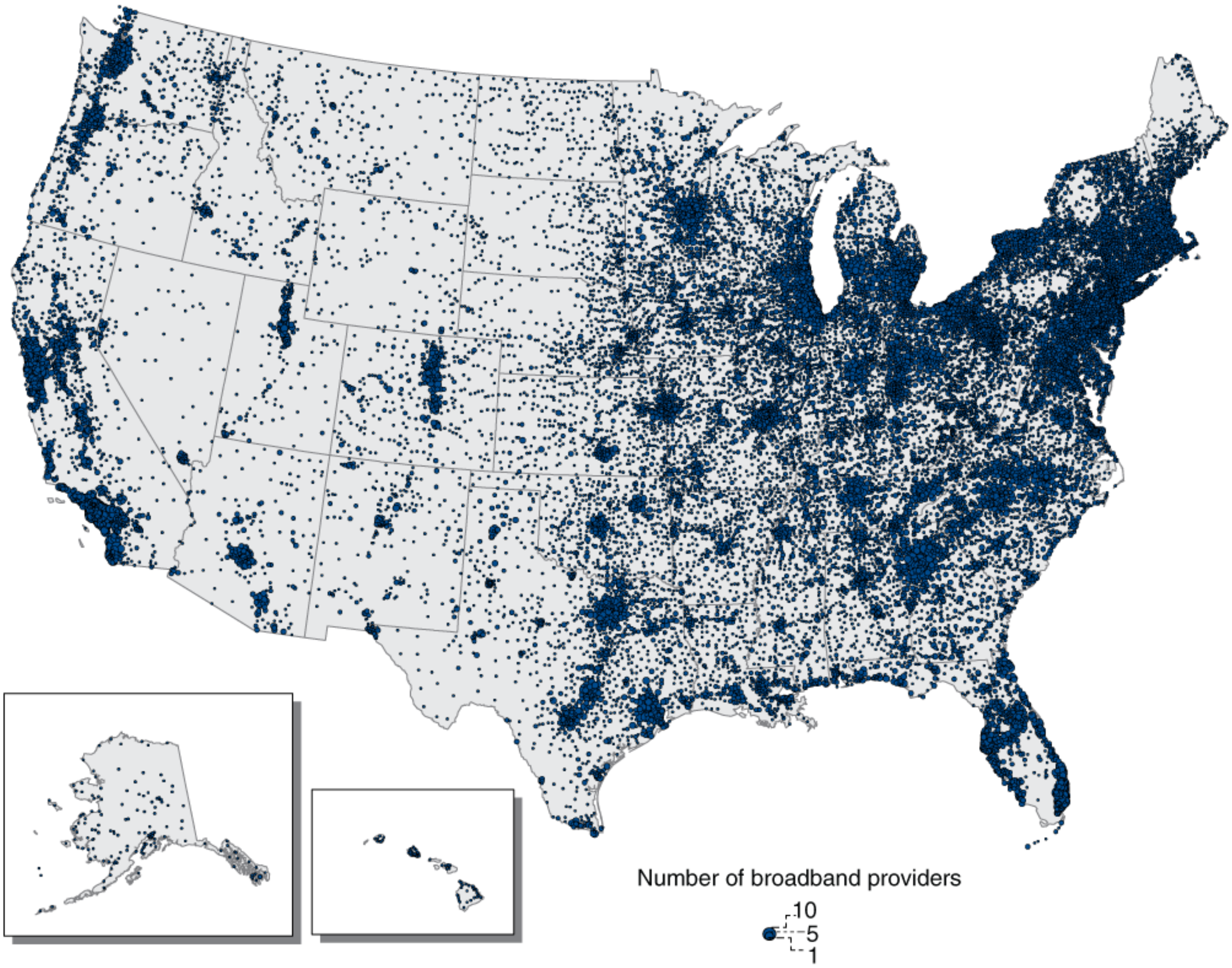
The growth in broadband availability has been rapid. This can be seen in the following maps (fig. C-1 - C-4) developed from the FCC data that show broadband availability from December 2000 through December 2006. Each dot in the maps represents a ZIP Code area that has at least one provider. The dots are located at the population center for their corresponding ZIP Codes. Dots are used to represent broadband availability as rural broadband service most commonly radiates out from towns into the surrounding countryside. These maps give a contrasting view of broadband availability from the isopleth map (fig. 6, p. 16) shown in the main body of this report. Some persistent wilderness areas and the development of broadband service along arterial roadways become apparent in the dot maps over time (from 2000 to 2006).

Figure C-1
Number of high-speed service providers by ZIP Code, 2000



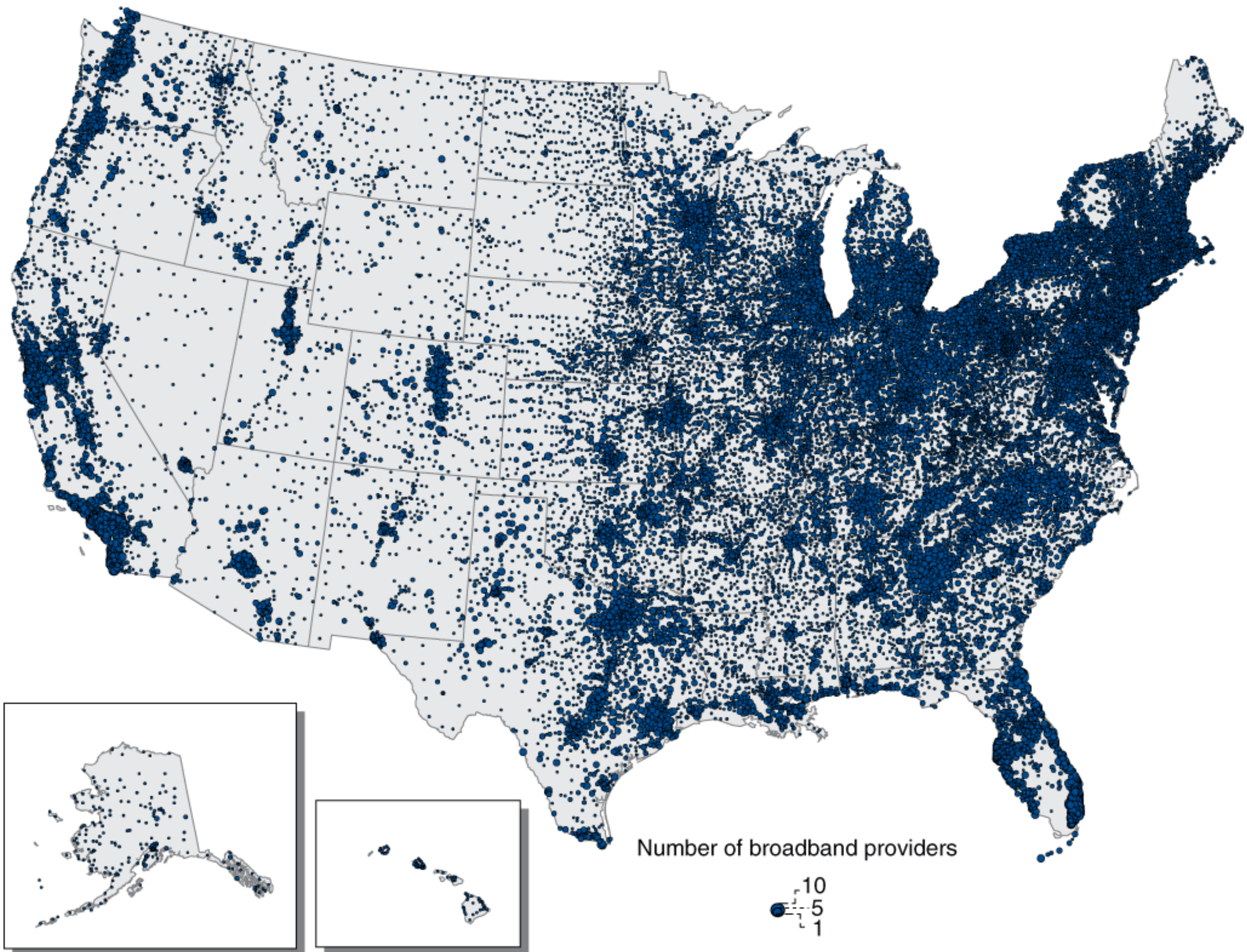
Source: ERS, using FCC data.

Figure C-2
Number of high-speed service providers by ZIP Code, 2002



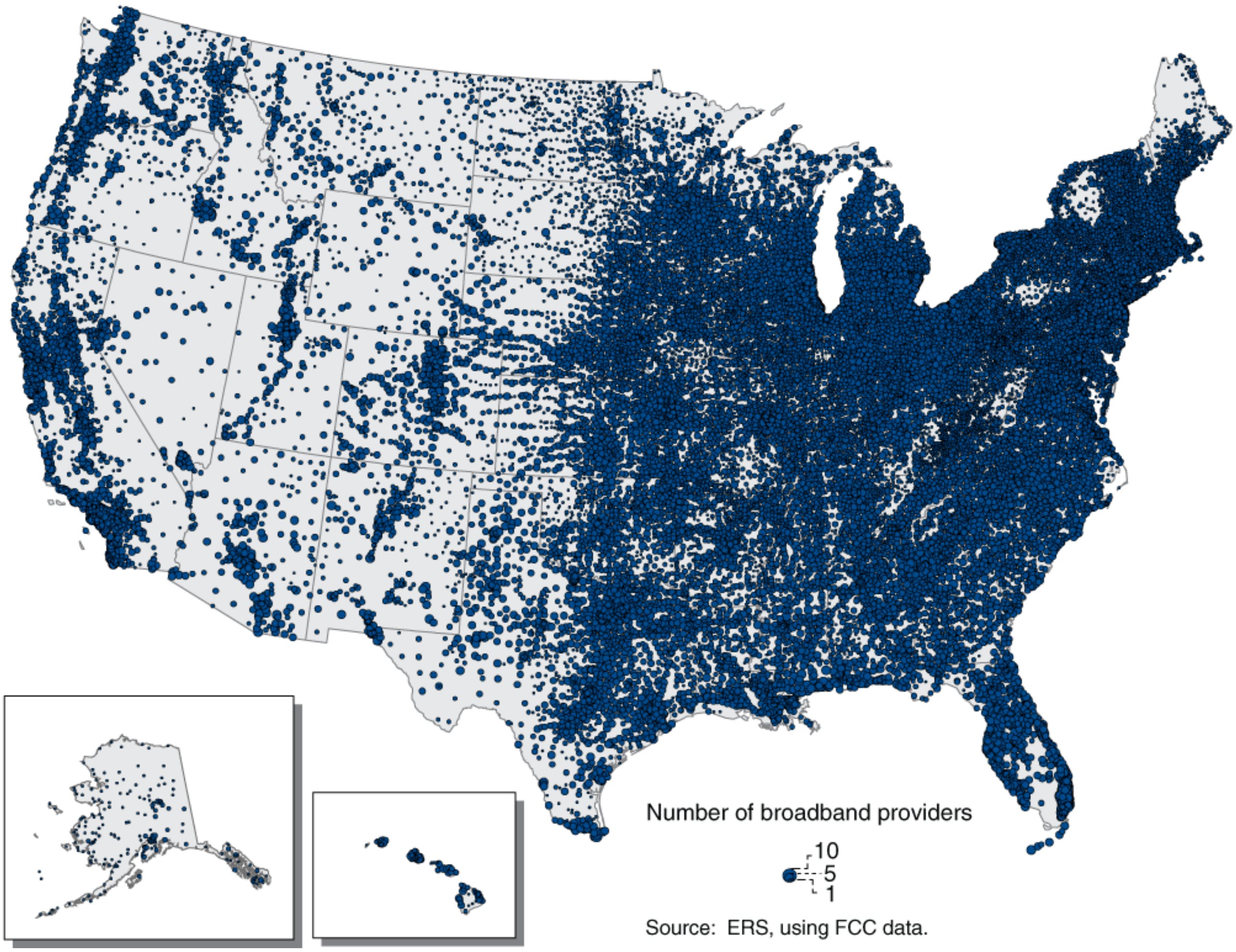
Source: ERS, using FCC data.

Figure C-3
Number of high-speed service providers by ZIP Code, 2004



Source: ERS, using FCC data.

Figure C-4
Number of high-speed service providers by ZIP Code, 2006



Source: ERS, using FCC data.

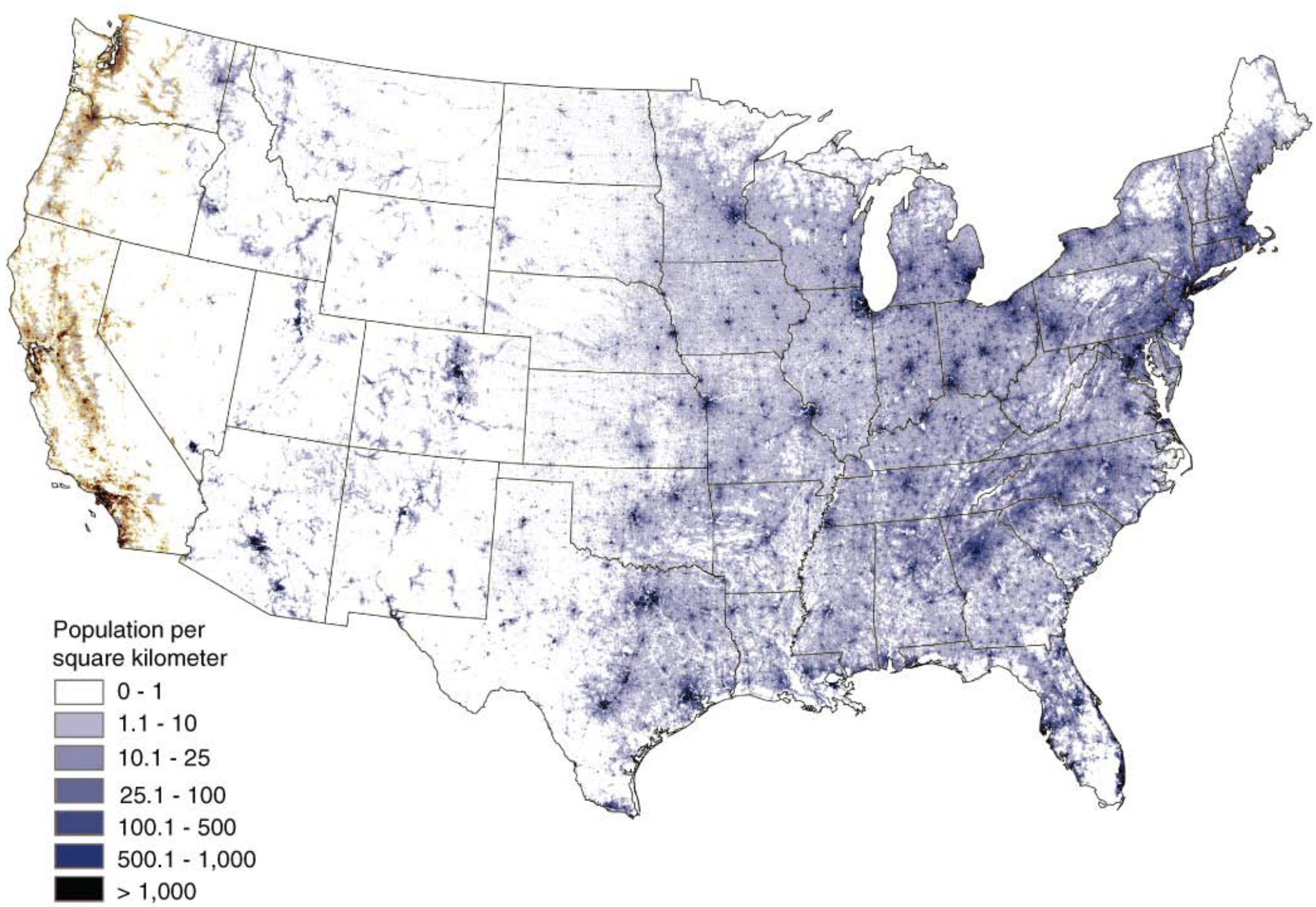
Broadband deployment has generally followed population density. This can be seen from the population density map (fig. C-5).

Enhancing the FCC Data

Population and adjoining area affect broadband availability. These facts underlie our enhancement of the FCC broadband availability data. From the FCC data we developed broadband availability density maps that constitute our most basic measure for a number of our research applications. This basic broadband database is composed of equally sized 2-km sub-ZIP Code zonal building blocks. The databases are further refined and adapted to each line of research in our broader rural broadband Internet study.

Essentially, these 2-km grid cells show the likelihood of having broadband available at any location within the lower 48 States at different points of time. To generate these broadband probability surfaces, we first locate the number of providers at the population centroid of the ZIP Code, where providers are most likely to maximize potential customers. Next, we pass a kernel density function over the provider locations using an 8-km search radius.

Figure C-5
Population density across the United States, 2000



Source: ERS using Bureau of the Census, Census of the Population, 2000.

The resulting surface provides estimates of provider access with the highest likelihood at the population centroid of ZIP Codes and decreasing probability toward the edge of the search radius. This search radius is larger than the typical limitation of DSL Internet service of 15,000 feet (4.5 km); due to technical reasons, DSL service cannot go beyond a certain distance from its signal's point of origin without additional equipment along the telephone line. The 8-km search radius helps to balance our assumption that all providers within a ZIP Code are centrally within the ZIP Code. Likelihood of service increases with more providers within a ZIP Code. Overlapping provision areas increase the likelihood of service to any location within the overlap, so high provision in adjoining zonal areas further increases the likelihood of broadband availability.

Our density map was tested against June Agricultural Survey data of farm broadband use. The June Agricultural Survey (JAS) data are a geographic-based survey of farms in the lower 48 States. Internet use data have been collected since 1997. The JAS Internet data give geographic- and time-specific use and non-use of broadband Internet. (See Appendix B for further discussion of the JAS data.)

The density map matched very well with the JAS data in all areas except what is essentially the Great Plains region. The challenge here is the large geographic size of some ZIP Code areas, suggesting that the population centroid indicates less well the broadband Internet service area. The location of schools is used to further define the likelihood of broadband Internet service in an area; schools are useful because of their widespread use of broadband Internet. With the additional data, the surface map was adjusted to include additional provision areas. The resulting broadband density is essentially a likelihood measure—the probability of broadband Internet access for any given point in geographic space. Likelihood of broadband Internet access is centered in urban areas and radiates out from these urban centers (fig. C-6). The FCC data and the various selected indices that we developed here, one of which is shown in figure C-7, form the basis for much of the analysis in this report.

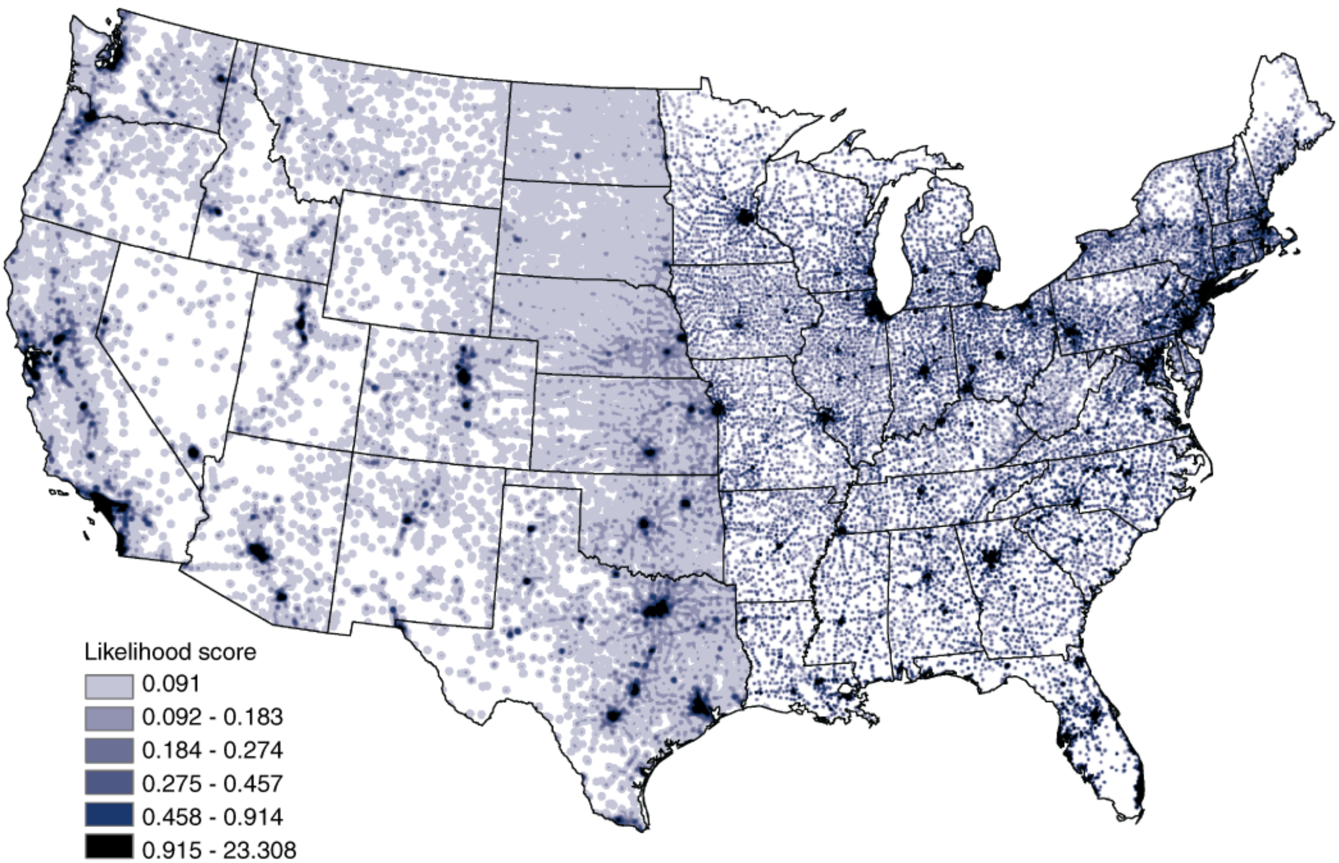
Our data were left as geo-specific or reconstituted into either county or ZIP Code areas for our analysis. Reconstitution into ZIP Code or county-level estimates required using population weighting of the provider likelihood estimate. A county aggregation of the data can be seen in figure C-7.

Estimation of Expected Number of Broadband Providers Using Population Measurements

To estimate the expected number of broadband providers, we began with the FCC's measure of the number of providers in a ZIP Code area in an unmodified form. To avoid disclosure of information on individual companies, areas that recorded 1-3 providers were combined into one category and assigned a value of 2. We used a population-based regression analysis to measure the relationship between the number of broadband providers and three components of U.S. population distribution: the overall population in the ZIP Code area in 2005, the percent of the ZIP Code's population living in urban areas as defined by the U.S. Census Bureau in 2000 (the most recent data on urban-

Figure C-6

Broadband availability based on 2005 FCC ZIP reporting and distance to schools in the West



Note: White areas would be expected not to have broadband Internet access. The darker the area, the more likely the area will have broadband.

Source: ERS using Bureau of the Census, Census of the Population, 2000.

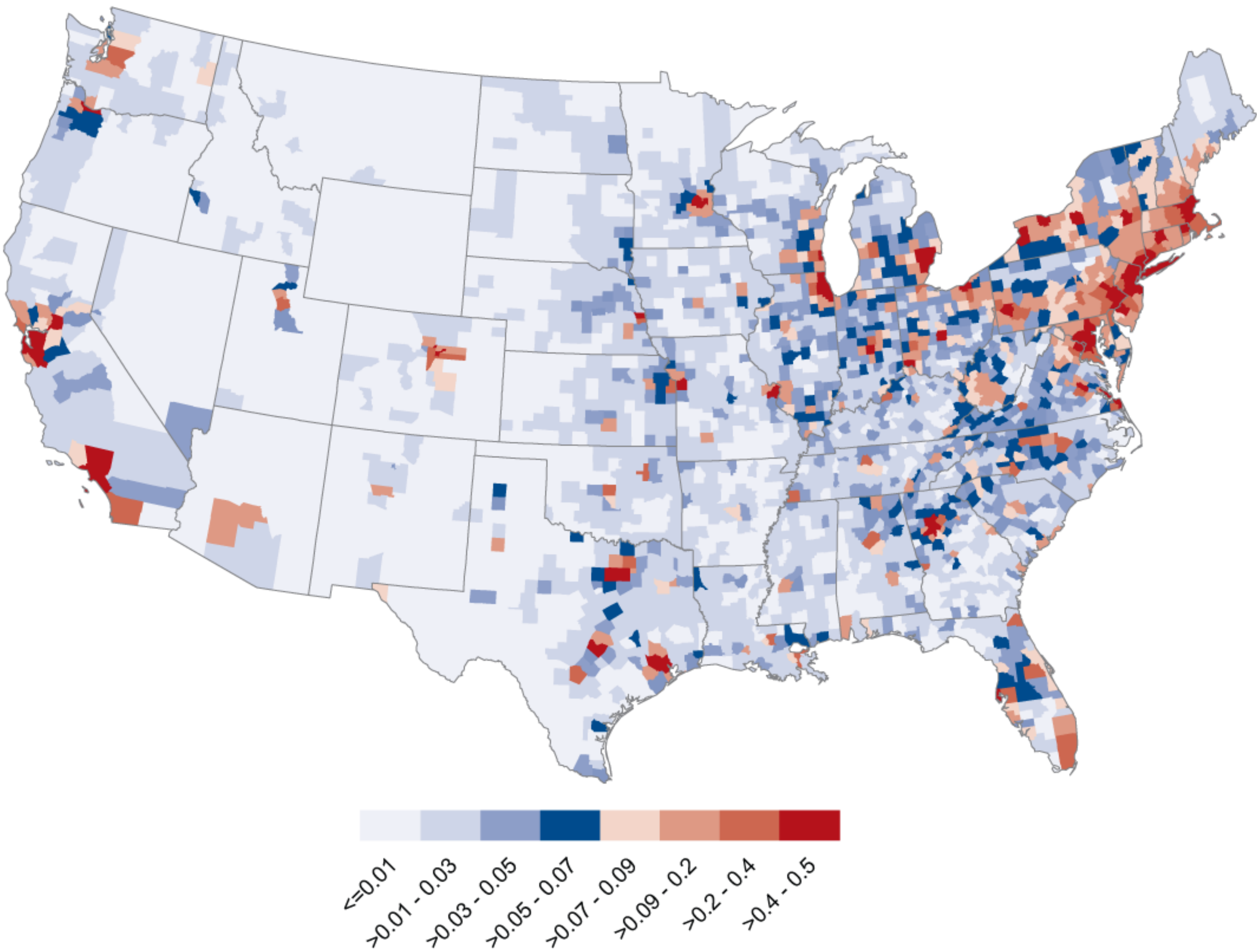
rural population), and a measure of the ZIP Code's accessibility to nearby populations.

ZIP Code area accessibility is used to distinguish areas that might be similarly sized but are differently positioned relative to large population centers. All things being equal, a nonmetro ZIP Code area adjacent to a metro area will likely have more providers than a more isolated area. To calculate accessibility for each ZIP Code area *a*, we identified all other ZIP Code areas within 200 miles, divided the population of each of these areas by the distance (squared) to area *a*, and summed the results. Values range from 0.0002 in Prudhoe Bay, Alaska to 150 million in Los Angeles (table C-1).

The independent variables were logged for the regression analysis to meet the assumption of linearity with the dependent variable. Parameter estimates represent the change in the number of broadband providers that occurs with a 1-percent increase in the independent variable. For example, a 1-percent increase in population size will increase the number of providers by 1.07. The standardized parameter estimates indicate the relative strength of the influence of each independent variable. Population size has twice the effect

Figure C-7

County representation of average broadband provision per square kilometer, 2000



Source: ERS using Bureau of the Census, Census of the Population, 2000.

Table C1

Descriptive statistics and linear regression results measuring the effect of population size, percent urban, and population accessibility on number of broadband providers, 2006

	Mean	Standard deviation	Minimum	Maximum	Parameter estimate	Standardized estimate
Dependent variable: Number of broadband providers, 2006	6.59	3.58	0	21	n/a	n/a
Independent variables: Population size, 2005	9,922	13,878	0	114,726	1.08	0.55
Percent urban, 2000	39.93	43.90	0	100	2.54	0.22
Population accessibility, 2005	20,944.09	1,045,331.36	0	150,709,763	0.21	0.13

Note: Descriptive statistics for the independent variables are shown for unlogged values; logged values were used in the regression analysis. All parameter estimates are significant at the .01 level. The proportion of variation in broadband provision explained by the regression (adjusted R-square) equals .63.

Source: USDA, Economic Research Service, using data from the FCC and U.S. Census Bureau.

on number of broadband providers compared with percent urban, which in turn has a stronger effect than accessibility.

Parameter estimates may be used to calculate the expected number of broadband providers in a ZIP Code area. Residuals from the regression show the difference between the actual and expected number of providers (fig. 9, p. 18). For example, Grafton, West Virginia's ZIP Code area had 10,316 people, was 59 percent urban, and had an accessibility measure of 385. Taking the natural logs of these values (4, 0.2, and 2.6, respectively), multiplying them by their parameter estimates, and summing the results show a predicted value for Grafton of 5.4 providers. The number of broadband providers in this ZIP Code area, as reported by the FCC in 2006, was 2, so the model indicates that Grafton has roughly 3 fewer providers than predicted.

Appendix D—Modeling Broadband Use on the Farm

Discrete choice models are interpreted in terms of an underlying behavioral model, the so-called random utility maximization (RUM) model. The decisionmaker chooses the alternative with the highest utility. Let x_{ij} be an attribute vector of alternatives *j* that individual *i* faces; let b be the impacts of the changes of the attributes; let e_{ij} be a random component. The random utility function of alternative *j* for individual *i* can then be written as $U_{ij} = b'x_{ij} + e_{ij}$. Suppose that alternative *j* is chosen and that alternative *k* is not chosen. Individual *i* will choose *j* to maximize the random utility function, if and only if $U_{ij} > U_{ik}$ for any $k \neq i$. Since e_{ij} is a random component of the individual utility function, the probability that individual *i* actually chooses alternative *j* is written as $P\ (U_{ij}>U_{ik})$ for any $k \neq i$. The true utilities of the alternatives are considered random variables, so the probability that the alternative is chosen is defined as the probability that it has the greatest utility among the available alternatives.

Using the conceptual framework put forth by Flamm and Chaudhuri (2007), we assume that the underlying utility of Internet use is related to economic and demographic attributes (Flamm and Chaudhuri, 2007). The purchase-decision outcome (*j*) consists of one of three choices: no purchase, dial-up, or broadband. The x_{ij} are represented by the explanatory variables described in the farm businesses and broadband section. Under certain restrictive assumptions, the probability that individual *i* chooses alternative *j* can be expressed as $P_{ij} = P(U_{ij}>U_{ik}) = \exp(b'x_{ij})/\sum_{k=1}^{K} \exp(b'x_{ik})$.

The model that results under certain distribution assumptions about e is usually called the conditional logit model. Apart from the assumptions underlying the RUM model, the conditional logit model implies that "the choice probabilities have the property which is called independence of irrelevant alternatives" (McFadden, 1974). This means that the ratio of the probabilities of choosing two alternatives is independent of the characteristics of all other choice possibilities. The conditional logit model is sometimes called the multinomial logit model. Following Greene (1993) and Maddala (1983), this latter term is reserved for models where the probabilities of the individual making a certain choice are functions of the characteristics of the individual, while the term conditional logit model is used when the choice probabilities are functions of characteristics of the choice alternatives. The way the problems are set up is different in these two models. The likelihood functions, however, will be the same.

Appendix E—Quasi-Experimental Design

Quasi-experimental design (QED) is a statistical approach that simulates an ex-post laboratory experiment featuring both a treatment and control group. Selection of control and treatment in QED, unlike a true laboratory experiment, is not perfectly random, hence the term "quasi." Treatment groups are self-selected. Control groups are selected based on their characteristic similarity with the initial, or pre-treatment, characteristics of the treatment groups. The QED approach taken here follows those of Isserman and Rephan (1995). SAS and other software were used in the analysis. A couple of the SAS routines used here were initially developed by Isserman.

The closeness between counties that is used to select the control counties is derived using a discrete measure called Mahalanobis distance. Mahalanobis distance measures the similarity between the treatment county and each county that could potentially be part of a control group. The measure is derived from the differences between the treatment county's and another county's characteristics' measures. The Mahalanobis distance is

$$MAHAL_{bj}=(X_b-X_j)^T\,\Sigma^{-1}(X_b-X_j),$$

where b is the treatment county, j is the potential control group county, X is the vector of variables that measure a county's characteristics, and Σ is the variance-covariance matrix of the variables calculated over all possible control counties. There are a number of ways to compare treatment versus control groups in QED. In the application here, there is one control county for each treatment county. No control county is allowed to appear more than once in the control group. The pairwise counties are the basic unit of analysis. The difference in growth is computed for each pair. The mean and standard deviations of these differences are computed, as are t-statistics between the treatment and control groups.

Robustness checks were made by analyzing prior-period growth rates. A tautology did not exist between the selection of control counties and their post-economic growth measures as the selection of control counties employs a large array of spatial and socioeconomic factors. Control and treatment county growth rates were more similar in the prior period, 1997-2000, than in the treatment period, 2002-2006. Selection criteria for treatment groups were relaxed and strengthened (i.e., cutoff points in broadband likelihood were increased and decreased). No appreciable change in outcomes was found as a result of these changes; the model was not sensitive to minor changes in treatment group inclusiveness.

More analysis on the robustness, however, needs to be completed to further substantiate the results and address more completely the issue of causality. Treatment group selection and control group characteristic variables will be further varied to test sensitivity in selection process.

Appendix F: Economic Research Service Broadband Workshop, September 29-30, 2008

Broadband in the Rural Economy

Keynote: Rural Digital Economy
Edward Malecki, The Ohio State University

Internet and Rural Business Activity

Broadband Deployment and Economic Development in Kentucky and North Carolina
Mitch Renkow, North Carolina State University

Rural Broadband Internet Use and Rural Economy
Peter Stenberg, Economic Research Service, RRED

Comparing Rural Retailer Internet Users and Non-Users: Access Speed, Demographics, Attitudes and Beliefs.
Leslie Stoel and Stan Ernst, Ohio State University

Rural Grocers and Technology Adoption: Attitude Matters. Size Matters More.
Stan Ernst and Leslie Stoel, The Ohio State University

Food and Nonfarm Rural Business

Internet Marketing of Nursery and Greenhouse Products
Enefiok Ekanem and Fisseha Tegegne, Tennessee State University

Positive Examples and Lessons Learned from Rural Small Business Adoption of E Commerce Strategies
David Lamie, David Barkley, Clemson University, and Deborah Markley, University of Missouri

IT and E-Commerce Companies
John Leatherman, Kansas State University, and Hanas Cader, South Carolina State University

Farm and Rural Households

Farm Businesses and Broadband Internet Use
Mitchell Morehart and Peter Stenberg, Economic Research Service

Farming and the Internet: Reasons for Nonuse
Brian Briggeman and Brian Whitacre, Oklahoma State University

What Skills Are at the End of Broadband Cables in Rural America? Do They Match Up with Firms Wishing to Engage Rural Sourcing?
Doug Morris and Lyndon Goodridge, University of New Hampshire

Digital Economy

IT in the Global Economy
Catherine Mann, Peterson Institute for International Economics and Brandeis University

Effects of Broadband Deployment on Output and Employment
Robert Crandall, Brookings Institution

Home Broadband Adoption in the United States: Patterns, Barriers, and Consequences
John Horrigan, Pew Internet & American Life Project

Community Internet Use

The Role of the Internet in Rural Community Participation—Examples from Recent Survey Data
Michael Stern and Alison E. Adams, Oklahoma State University

Rural Distance Education
Janet Poley, University of Nebraska-Lincoln and President of the American Distance Education Consortium

Economic Impact of Rural Telemedicine
Brian Whitacre, Oklahoma State University