APPENDIX A: THE DATA

TO

THE RESIDENTIAL AND COMMERCIAL BENEFITS OF RURAL BROADBAND: EVIDENCE FROM CENTRAL APPALACHIA

FINAL REPORT

JULY 2005

MARK L. BURTON

MICHAEL J. HICKS

Introduction

This Appendix describes the data collected as part of the preceding study. Since much of the data is publicly available we provide only selected summary statistics. Our intent is to highlight strengths and weaknesses of the data, and explain the potential impact of on our study of varying definitions of the data. We begin with a general description of the study, temporally and regionally. We then describe the data collected and analyzed as part of the analysis of consumers, followed by a similar description for the commercial analysis. Much of the data was evaluated in both studies, and so as to avoid redundancies we will only note how the data was employed in the later studies, providing the bulk of the discussion for the first application of the data. We also include descriptions of how calculations for each transformation of data or aggregation scheme were performed.

The Study Design

This study evaluates consumer and commercial benefits to broadband, through analysis of a variety of data from Kentucky, Ohio, Pennsylvania and West Virginia. The region of particular interest is West Virginia, thus the surrounding areas have been chosen to provide greater robustness to a number of the hypotheses examined in the course of the preceding study. The time period we analyzed stretched from 1990 through 2005, but was obviously different for data of varying vintage. In each case we were constrained by the availability and definition of data (or cost of acquiring better data). This is an unavoidable dilemma in research of this nature.

The study itself had a number of analytical components from maps, to advanced econometric models. We gathered data from the *Federal Communications Commission* (FCC), the *Bureau of Labor Statistics* (BLS), the *Census* and the *Bureau of Economic Analysis* (BEA). We compared these data to other sources such as the West Virginia Public Service Commission and a number of advocacy groups, industry data and those collected and evaluated from other sources. We also collected our own data through a telephonic survey of West Virginia households.

The data we have employed is the best available. We are familiar with both the shortcomings and strengths of the data, and the manner in which the geographic definitions have been made. In the final analysis, if these data prove insufficient to the reader in providing a basis for evaluating the questions we have attempted to answer (as distinct from our ability to answer these questions) then no economic, environmental, political or social study will be acceptable.

The Consumer Study

We begin our consumer study in terms of regional access to broadband telecommunications. We employ the regions created by state legislators (counties) and the subsets created by the U.S. Postal Service (zip codes). We note that zip codes include population, infrastructure and geological considerations. The most important of these is

population as zip codes are designed to better manage mail traffic to households and businesses. Zip codes change (though rarely) and the largest impact in recent changes to zip codes have been in the dense suburban areas of Ohio and Pennsylvania. Zip codes are used to provide regional analysis of broadband coverage, and also provide Census estimates of key demographic and economic information.

Broadband Access

The Federal Communications Commission has collected data since 1998 on broadband access by zip code in the United States. These data have been criticized for both the technical and geographic content. The FCC defines broadband as data transfer, in one direction at 200 kbs. One Commissioner challenges this definition as "so 1997." And, while we admire this air of technological savoir faire, the alternative definitions do no better in defining what is "state of the art" nor what is likely to be extant eight years hence. What is clear is that this definition marked a point of departure from traditional dial-up internet connections to something different. It still does, and so in whatever failures this definition provides, it at least allows a clear demarcation between those without any access to technology better than dial-up.[1] Among the better informed analysis of these problems (Odlyzko, 2003) notes that the Postal Service meets the current (and virtually all proposed) definitions of broadband as a box of CD's mailed overnight potentially meet any speed and data based definition. The many new or alternative technological definitions which rely on symmetric data transfer or greater speeds are subject to the same criticisms. The studies are not better, merely different.

The FCC defines regional access at the zip code level. Thus, access of a 200 kbs asymmetric data transfer by a single household, or by every household, meets the FCC definition of access. This definition is clearly adequate for most of the study region. However, for the very rare zip codes with highly spatially unbalanced populations, this

[1] One advocacy group noted that this definition was contrived by Telecom friendly members of Congress to protect regional monopolies by distorting the reality of broadband. They further argued that 200kbs could be achieved by enhanced dial-up. We wonder, if this is the case, why then the ILEC's don't simply offer the enhanced dial-up to the remaining unserved zip codes (at a trivial cost) and thus report to the FCC that everyone in the US now has broadband access.

definition permits the misinterpretation that all residents enjoy broadband access. We discourage such a lazy interpretation, but caution the critic as well, these data are substantially correct.

As researchers with considerable experience with data we are satisfied that these data represent the phenomenon of internet access as adequately as most other regional or technological data. By comparison we feel compelled to note that all data has errors. Indeed, the breathtakingly high levels of the shadow economy (10-27% of West Virginia's Gross Domestic Product), which distorts the poverty measures; the arbitrariness of the rural/urban divide; and the bias generated by poor literacy in reporting Census data are examples of data concerns.

As our postscript suggests, we believe there is need for data collection and analysis, and that this will become more urgent as technology continues to change the benefits derived by consumers and businesses for different types of information technology and the services generated by these difference. However, the evidence to date strongly suggests that the FCC broadband data is very effective in its effort to portray the "dial-up" versus "high speed" divide both technologically and geographically.

Census Data

The 2000 Census represented the most ambitious national Census in history, and the data obtained from this effort is remarkably descriptive. These data are collected at the Census Tract level and aggregated through multi-state regions. The data are intended to be "as of" March, 2000, but are released over several a considerable period of time. Direct data on residential broadband access or internet use is limited to the Current Population Survey. We have reviewed studies of these data for the 2000 Census. We chose not to directly model these data for two reasons. First, the interesting questions surrounding these data have largely been answered and second, there are no geographical links to the data so we cannot isolate any proximal study region.

The data from the 2000 Census we employ in this study are some of the typical characteristics, which we divide by internet access (see Table 1, pg 12 of this report). While these data are almost universally viewed as the gold standard in data quality, conditions such as literacy levels may bias some of the reported data.

We report only limited ranges of these data, since simple access was not among the direct study questions in this report. We refer the reader to a number of potential studies of geographic and demographic features of broadband access noted in the text. Among these are Grubesic [2002]; Grubesic and Murray [2003] and Chaudhuri and Flamm [2005].

The Survey Data

The survey instrument employed in this analysis was an extensive questionnaire regarding internet knowledge and use, matched to individual (and regional) data. The survey consisted of 600 random calls to households, with 200 each collected in each of West Virginia's three Congressional districts. The choice to separate the survey in this way was motivated by the Steering Committee's desire to observe regional variation in the data collection process. The survey itself enjoyed a response rate in excess of 30 percent, which is unusually good for a questionnaire of this length. The responses provide a sample of the State's population which is enjoys significance better than the 2.5 percent level. The questions employed in our analysis are reported in the text. Table A-2 provides summary statistics to these data.

Table A-2 Summary Statistics of Selected Survey Data

Variable	Mean	Median	Maximum	Minimum	Std. Dev.
Number of Adults in Family	1.87	2	5	1	0.730078
Number of Kids in Family	0.48	0	6	0	0.931837
Age of Youngest Family Member	8.23	12	12	1	4.350707
Median Age	39.05	39.6	46.3	22	3.261337
Percent Population > 25 with BA degree	0.091	0.087	0.279612	0	0.048993
Median HH income	30,201.14	29,112	62,445	0	7,454.883
Have Broadband at Work	1.226667	1	3	0	1.159395
Broadband Price (actual, but not necessarily known)	35.34531	35	50	10	8.130141

Summary

The data employed in the residential demand estimates represent the best available for this type of analysis. These data form the bulwark of research regarding broadband (and many other questions) and are well accepted by researchers and informed policymakers. However, these data are not perfect, and in an attempt to improve upon known shortcomings we collected our own primary data on internet usage, price, and household demographics. We also recommend more attention to data collection occur (see Section 8 of the report).

Firm Productivity Study

Three separate estimation strategies were employed to evaluate firm productivity impacts of broadband. The first of these was a firm level estimate of broadband's productivity impact employing microdata from West Virginia and Kentucky. The second analysis consisted of a series of estimates of panel data consisting of county level data on all counties in Kentucky, Ohio, Pennsylvania and West Virginia from 1990 through 2003. This model permitted the estimate of broadband impacts on industry productivity as measured through the impact on worker wages in these industries. We review the data to support these modeling efforts in turn. The third process involved a simple comparison of firms located in the Corridor G region based upon the presence of broadband access. The goal of this analysis was to evaluate the relative productivity of firms based solely on their location decision with respect to residential broadband.

Firm Level Productivity Estimates of Broadband

The first of these matched the FCC's 2000 broadband access data with an existing microdata set of firms employed in an earlier study of firm productivity and highway access along the Appalachian Development Corridor G (RTE 119).[2] This study matched

[2] Hicks, Michael J. The Impact of Appalachian Corridors on Small Business, Transportation Research Board National Academy of Sciences, Proceedings, Transportation and Economic Development 2001. See

firm and regional specific data with GIS estimates of actual route and Euclidean distance to Corridor G.

By matching the FCC data to the firm specific and regional data we were able to estimate the impact of broadband on firm productivity (the average product of labor, which is defined as the Revenues per worker in an establishment). See summary statistics in Table A-2.

Table A-2 Selected Corridor G Data

	Mean	Median	Maximum	Minimum	Std. Dev.
Distance to Corridor G (feet)	28,587.55	19,855.18	138,529.7	330.488	25079.23
BROADBAND	0.972421	1	1	0	0.163775
Age of Firm	18.15707	12	196	0	20.66663
URBAN Zip	0.685366	1	1	0	0.464405
POPULATION	11713.94	10329	25012	114	7890.45
Number of Households	5,044.238	4663	10476	49	3,222.29
West Virginia Dummy	0.809843	1	1	0	0.392455
Per Capita Income	19,105.24	17867	36812	7157	5,575.244

The specific methods of calculating the distance to Corridor G, and the productivity arguments are offered in the original work, and again in the technical appendix on firm productivity appended to this report.

The Broadband Access and Worker Wage Study

Analysis of worker wages leads to an imputation of labor productivity since at the county level labor markets are typically quite competitive. Competitive labor markets will thus generate wage responses due to productivity increases.

In order to estimate this response, we aggregate broadband coverage from the FCC zipcode data to the county level. We used the zip code populations, to estimate the proportion of the county's population (from the 2000 Census) with broadband access for each year from the earliest FCC data through 2003.

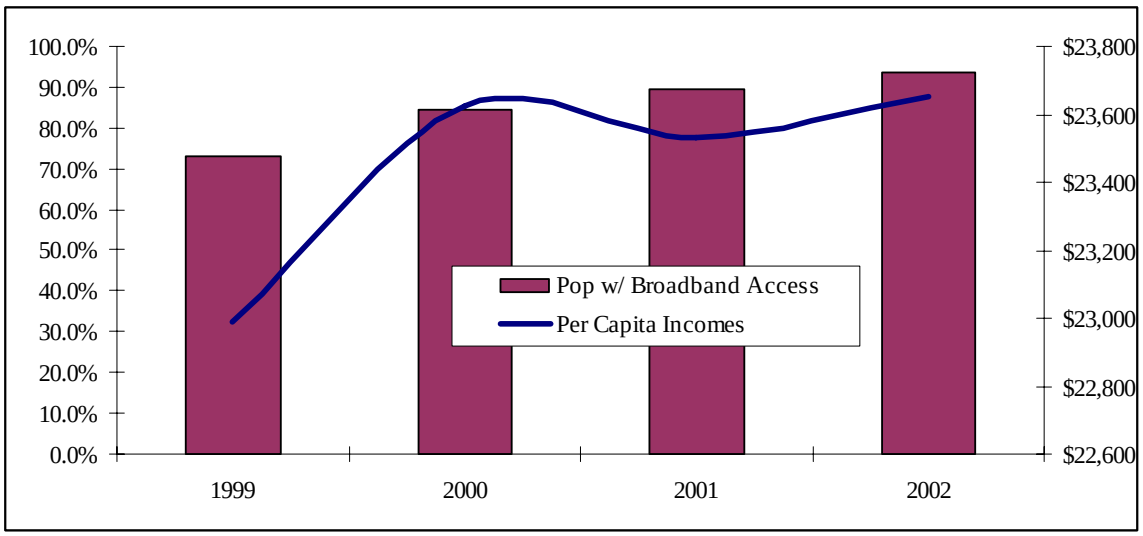
also The Impact of Appalachian Highway Corridors on the Scope of Small Business Activity, ATI 99-15 Rahall Appalachian Transportation Institute.

Further, we created a spatial weighting matrix W, for a series of spatial autocorrelation functions which appear in the estimation. The spatial weight matrix is simply the average of some variable X, in each of the contiguous counties to the county in question. Hence, the spatial weighting matrix is known as a first order contiguity matrix in this setting.

The remaining data on incomes, population, employment were collected from the *Department of Commerce's, Bureau of Economic Analysis, Regional Economic Information Systems*, and placed into 2002 dollars (for nominal values) using the consumer price index, all consumers.

The binary variables in these estimates included a recession dummy, which was coded for the year in which the National Bureau of Economic Research declared a recession in the United States. The 2001 binary variable was included to account for the drop off in investment following the dot.com bust. These data are far too extensive to display in this setting. Selected data appear in Figure A-1.

Figure A-1, Selected Regional Data



The Firm Location Analysis

The final productivity estimate we attempted was to evaluate the role of firm location decisions with respect to broadband. To accomplish this, we manually eliminated all firms with SIC codes which occurred simultaneously in zip codes with either broadband or without broadband access. An excerpt of this data appears in Table 7 (page 27) of the main study document.

Summary

The firm level productivity data and the regional productivity data are compiled from public sources, and are available from the Federal Communications Commission, the Department of Commerce and the National Bureau of Economic Research. These data represent the best available time series data on broadband, income, population, inflation and recessions for the United States.

Contact:

Michael J. Hicks
hicksm@marshall.edu
Michael.hicks@afit.edu
Center for Business and Economic Research
Marshall University
1 John Marshall Drive
Huntington, WV 25755