

Bridging the Rural – Urban Digital Divide in Residential Internet Access

Brian E. Whitacre

Dissertation submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Economics

Dr. Bradford Mills, Chair
Dr. Jeffrey Alwang
Dr. Richard Ashley
Dr. Everett Peterson
Dr. Daniel Taylor

September 1, 2005
Blacksburg, Virginia

Key words: Digital Divide, Internet, Diffusion, Rural, Logit Decomposition

Copyright 2005, Brian E. Whitacre

Bridging the Rural – Urban Digital Divide in Residential Internet Access

by

Brian E. Whitacre

Dr. Bradford F. Mills, Chair

(ABSTRACT)

This dissertation explores the persistent gap between rural and urban areas in the percentage of households that access the Internet at home (a discrepancy commonly known as the "digital divide"). The theoretical framework underlying a household's Internet adoption decision is examined, with emphasis on the roles that household characteristics, network externalities, and digital communication technology (DCT) infrastructure potentially play. This framework is transferred into a statistical model of household Internet access, where non-linear decomposition techniques are employed to estimate the contributions of these variables to the digital divide in a given year. Differences in Internet access rates between years are also analyzed to understand the importance of temporal resistance to the continuing digital divide.

The increasing prevalence of "high-speed" or broadband access is also taken into account by modeling a decision process where households that choose to have Internet access must decide between dial-up and high-speed access. This nested process is also decomposed in order to estimate the contributions of household characteristics, network externalities, DCT infrastructure, and temporal resistance to the high-speed digital divide.

The results suggest that public policies designed to alleviate digital divides in both general and high-speed access should focus more on the broader income and education inequities between rural and urban areas. The results also imply that the current policy environment of encouraging DCT infrastructure investment in rural areas may not be the most effective way to close the digital divide in both general and high-speed Internet access.

Acknowledgements

First and foremost, I would like to thank Bradford Mills for going above and beyond the call of duty in his role as my committee chair. It was Brad who first introduced this topic to me upon my arrival at Virginia Tech, and I cannot overestimate the contribution of his guidance and feedback. Brad read and edited every version of this dissertation from its infancy, a feat made all the more impressive by my own less-than-distinguished writing ability.

I would also like to thank the remaining members of my committee, Drs. Jeffrey Alwang, Daniel Taylor, Everett Peterson, and Richard Ashley, for their comments and suggestions regarding this work.

I am extremely grateful for the opportunity to pursue a Ph.D. in the Agricultural and Applied Economics at Virginia Tech. My experience with the department and school has been nothing but positive. I owe many thanks to the faculty, students, and staff for making my time as a graduate student such a rewarding experience.

Finally, I would like to thank my fiancée Jill, who, along with my parents John and Theresa, provided all the support and understanding I needed to complete this process. I love you all.

Table of Contents

Table of Contents	iv
List of Figures.....	v
List of Tables	vi
Chapter 1: Introduction	1
1.1 - Problem Statement.....	1
1.2 - Objectives	6
1.3 – Study Structure.....	7
Chapter 2: Conceptual Framework	9
2.1 - Internet Access vs. Internet Use	9
2.2 - Structure of the Internet.....	10
2.3 - The Big Picture.....	18
2.4 - Diffusion Theory	20
2.5 - Adoption Theory.....	23
2.6 - Utility Theory	25
2.7 - A Unified Theory of Residential Internet Access	28
2.8 - Review of the Literature	31
Chapter 3: Empirical Framework.....	53
3.1 - Data.....	53
3.2 - Empirical Specification	58
3.3 - Model Distribution Assumption	62
3.4 - Differentiating Dial-up and High-speed Access in the Adoption Decision	66
3.5 - Assessing the Importance of the Four Factors.....	70
Chapter 4: Results.....	82
4.1 - General Logit Model Results.....	82
4.2 - Decomposition of the General Digital Divide.....	91
4.3 - Inter-temporal Decomposition of the General Digital Divide.....	97
4.4 - Nested Logit Model Results	109
4.5 - Decomposition of the Nested Logit Model	124
4.6 - Decomposition of the Inter-Temporal Nested Logit Model.....	133
Chapter 5: Policy Implications, Limitations, and Conclusions	145
5.1 – Policy Implications for General Access.....	145
5.2 – Policy Implications for High-speed Access.....	150
5.3 – Limitations and Areas for Future Research	153
5.4 – Concluding Remarks.....	155
References.....	157
Appendix A	166
Appendix B	168
Appendix C.....	169
Appendix D.....	170
Appendix E	171
Appendix F	175

List of Figures

Figure 1. Residential Internet Access and the Rural - Urban Digital Divide	2
Figure 2. Internet Structure for Residential Access	10
Figure 3. Overlay of Competing Fiber Optic Networks in the U.S., 2002	11
Figure 4. Middle Mile Infrastructure in South Dakota	12
Figure 5. Use of Last-mile Technologies by Households with Internet Access	14
Figure 6. Technology Diffusion Cycle	19
Figure 7. S-shaped Curve Representing the Rate of Adoption over Time	21
Figure 8. Technology Adopter Categories	24
Figure 9. A More Detailed Technology Diffusion Cycle	30
Figure 10. Nine Regions of the United States	43
Figure 11. Rate of Adoption for Two Types of Innovations	46
Figure 12. Temporal and Geographic Resistance to Hybrid Corn Adoption	50
Figure 13. S-Curves for Various Technologies	51
Figure 14. Multinomial and Nested Decision Processes	67
Figure 15. Nested Logit Tree Structure	68
Figure 16. Age Parameter Values from 1997 and 2003 Regressions	101
Figure 17. Age Profile of Household Heads with Internet Access	102

List of Tables

Table 1. Residential Broadband Market Structure and Consumer Costs, 2003.....	15
Table 2. Technology Adoption Categories and Typical Characteristics	24
Table 3. Household Characteristics by Internet Access.....	34
Table 4. Household Characteristics by Rural / Urban Area.....	35
Table 5. Percent of U.S. Rural / Urban Population Living in Counties with DCT Infrastructure.....	39
Table 6. Residential Dial-up and High-Speed Access Rates by Region.....	43
Table 7. Income and Education levels for Internet and High-speed Adopters	49
Table 8. CPS Household Summary Data.....	54
Table 9. Percent of U.S. Rural / Urban Population Living in Counties with DCT Infrastructure (Broken out by Region).....	55
Table 10. Key Components of Alternative Residential Internet Adoption Models.....	60
Table 11. Variable Summary (Preliminary Regressions)	63
Table 12. LPM, Logit, and Probit Estimates for Internet Access in 2003.....	64
Table 13. Comparison of Partial Effects for LPM, Logit, and Probit.....	65
Table 14. Decomposition of Nested Logit Specification.....	78
Table 15. Inter-temporal Decomposition of Nested Logit Specification.....	80
Table 16. Logit Results for General Internet Access (2000 - 2003).....	84
Table 17. Logit Results for General Internet Access (1997 – 1998)	87
Table 18. Logit Regression for Urban – Rural Internet Access (2003).....	89
Table 19. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access, 1997 - 2003.....	92
Table 20. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access, 1997 – 2003 (Reverse Ordering.....	94
Table 21. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access (No Network Externality Term), 1997 - 2003.....	97
Table 22. Summary of Inter-temporal Decomposition for General Internet Access, 1997 - 2003.....	98
Table 23. Individual Contributions of Characteristics and Parameters to the Inter- temporal Decomposition in General Internet Access, 1997 - 2003	99
Table 24. Summary of Inter-temporal Decomposition for General Internet Access, 2000 - 2003.....	103
Table 25. Individual Contributions of Characteristics and Parameters to the Inter- temporal Decomposition in General Internet Access, 2000 - 2003	104
Table 26. Individual Contributions of Characteristics and Parameters to the Inter- temporal Decomposition in General Internet Access, 1997 – 2003, Order Reversed	106
Table 27. Individual Contributions of Characteristics and Parameters to the Inter- temporal Decomposition in General Internet Access, 2000 – 2003, Order Reversed	106
Table 28. Nested Logit Results for Education (2003)	110
Table 29. Nested and Multinomial Logit Model Comparison - I (2003).....	112
Table 30. Nested Logit Results for Education and Income (2003).....	113
Table 31. Nested and Multinomial Logit Model Comparison - II (2003)	115

Table 32. Nested Logit Results for Education, Income, and Other Household Characteristics (2003)	116
Table 33. Nested and Multinomial Logit Model Comparison - III (2003).....	118
Table 34. Nested Logit Results for Education, Income, Other Household Characteristics, and Network Externalities (2003).....	119
Table 35. Nested and Multinomial Logit Model Comparison - IV (2003).....	120
Table 36. Nested Logit Results for Education, Income, Other Household Characteristics, Network Externalities, and DCT Infrastructure (2003)	122
Table 37. Nested and Multinomial Logit Model Comparison - V (2003).....	123
Table 38. Nested and Multinomial Logit Model Comparison - Summary (2003)	124
Table 39. Nested Logit Decomposition Results	125
Table 40. Nested Logit Decomposition Results (Order Reversed).....	128
Table 41. Nested Logit Decomposition Results (Single Explanatory Variables).....	130
Table 42. Nested Logit Results for Education, Income, Other Household Characteristics, and DCT Infrastructure (No Network Externalities) (2003).....	132
Table 43. Nested Logit Decomposition Results (Comparison with Model Excluding Network Externalities)	133
Table 44. Inter-temporal Nested Logit Decomposition Results	134
Table 45. Inter-temporal Nested Logit Decomposition Results (Order Reversed)	137
Table 46. Inter-temporal Nested Logit Decomposition - Contributions of Parameter Shifts	139
Table 47. Inter-temporal Nested Logit Decomposition - Contributions of Parameter Shifts (Order Reversed)	143

Chapter 1: Introduction

"The future is already here, it's just unevenly distributed."

-William Gibson (science fiction author)

1.1 - Problem Statement

The Internet is arguably the most significant innovation to enter U.S. households since the television. Access to the Internet provides households with an array of previously unavailable opportunities for commerce, education, entertainment, and civic engagement. While more and more households became 'digitally connected' during the 1990s and early 2000s, disparities in residential access to the Internet emerged among various segments of the population. Recent survey results find that Whites show higher rates of access to the Internet than Blacks (Compaine, 2001; NTIA, 2002). Non-Hispanics show higher rates of access than Hispanics. Internet access is also found to increase with household education and income levels (NTIA, 2002).

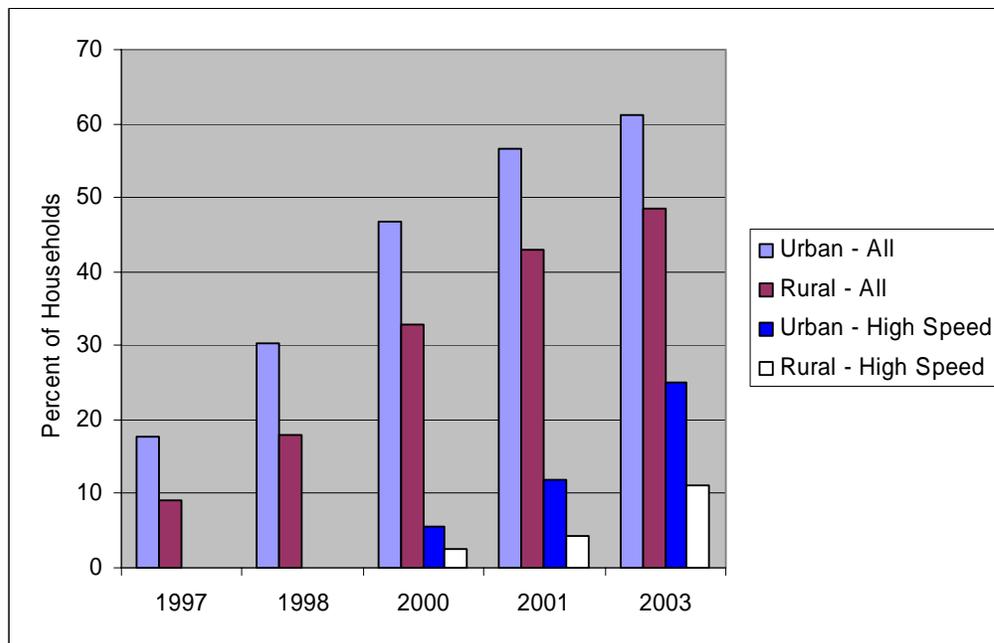
Regional variations in rates of residential Internet access are also found, with perhaps the most notable difference being a 13 percentage points higher rate of residential Internet access among metropolitan area households than among non-metropolitan area households in 2003.¹ This inequality in residential Internet access is generically referred to as the rural – urban digital divide. Current Population Survey Computer and Internet Use Supplemental Survey (CPS) data reveals a dramatic increase in residential Internet access over the period 1997 to 2003, along with a persistent rural - urban digital divide (Figure 1).² A new rural – urban digital divide has also rapidly emerged in high-speed Internet access, with the rate of high-speed residential Internet

¹ This paper uses the 1993 U.S. Census designations of non-metropolitan and metropolitan counties to compare rural - urban area differences in home Internet use. Metropolitan counties generally have populations greater than 100,000 (75,000 in New England) or a town or city of at least 50,000 and are referred to as urban areas. Non-metropolitan counties are those counties not classified as metropolitan and are referred to as rural areas.

² All estimates, unless noted, are based on author's calculations.

access being more than two times higher in urban counties than rural counties in 2003.³ As Malecki (2003) notes, high-speed access is becoming an essential dimension of the Internet due to the increasing prevalence of graphics, audio, and video on the web. These media are important to both recreational and business-oriented users. As the percentage of users with high-speed access climbs, so will the waiting time for dial-up users as web pages insert more multimedia elements in an effort to offer a more impressive on-line experience.

Figure 1. Residential Internet Access and the Rural - Urban Digital Divide



Sources: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, and 2003.

The rural - urban digital divide has drawn attention from government agencies at the local, state, and national level. Significant differences of opinion exist regarding the best way to close the divide, but virtually all agencies agree that the divide needs to be closely monitored.⁴ Two justifications are commonly given for why closing the rural –

³ High-speed access, also called Broadband or advanced service, is defined as 200 Kilobits per second (Kbps) (or 200,000 bits per second) of data throughput. This is about 4 times faster than a 56Kbps dial-up modem, and about 8 times faster than most people’s actual download speeds, since many ISPs’ modems offer a maximum of 28.8 (Strover, 2001)

⁴ The Department of Commerce under the Clinton Administration identified the existence of several digital divides (NTIA 1999), while the Bush Administration instead focused on the increasing rates of usage

urban digital divide is important. First, the digital divide may exacerbate existing inequalities in rural and urban household economic well-being (Drabenstott, 2001; Forestier, 2002). The benefits associated with residential Internet use (including education and income opportunities) can only accrue to those who have access to the technology. Although the provision of Internet access at public places such as libraries does allow those without access in their own home to use the Internet, this "away from home" access falls well below residential access in the intimacy of the online experience. For example, the top three reasons given for Internet use were gathering information for personal needs, entertainment, and education (Georgia Institute of Technology, 1998). However, quickly finding driving directions, looking up the latest baseball scores, and taking a class on-line are all much easier and more convenient to perform from the comfort of home. One of the largest benefits of Internet access is the ability to perform such tasks on a second's notice. This benefit disappears if obtaining such access requires leaving home and making a trip to the nearest Internet accessibility point.

Second, the unique nature of the Internet may be particularly suited to solve the age-old rural location problem. The Internet has the potential to reduce the 'rural penalty' associated with high costs of economic transactions stemming from lower market density and greater distance between businesses and other economic agents (Hite, 1997; Malecki, 2003). However, the rural penalty also inhibits telecommunication infrastructure investments that support residential Internet access, particularly high-speed access. Stover (2003) notes that population density influences the development of competitive markets and infrastructure investments. In the case of digital telecommunication infrastructure, the number of area high-speed providers typically depends on a combination of population density and per capita income. Higher numbers of competitive providers in an area market, in turn, spurs high-speed deployment. Thus, in terms of Internet access, the "rural penalty" can be reconceptualized as a "remote penalty," with the most remote towns least likely to enjoy the fruits of the communication revolution (Nicholas, 2002). Given the potential repercussions of the digital divide for economic growth in rural areas relative to urban areas, it is not surprising that many rural

among traditionally underserved groups (NTIA 2002). However, the disparity in usage among various groups was still acknowledged in the Bush Administration report.

coalitions have become active in pushing for a solution. In particular, the Southern Rural Development Center lists support for state research and extension efforts to close the digital divide in the rural south as one of its five priority efforts, while the Rural Utilities Service provided \$1.4B in loans for high-speed access in FY2003 for communities of up to 20,000 people (RUS, 2004).

Most rural development agencies feel that more empirical information on the underlying causes of the rural – urban digital divide is needed to create a policy environment that ensures that residency in rural areas does not raise continued differential barriers to home Internet access. Similarly, the increasing importance of high-speed access for many Internet applications indicates a need for analysis of the household decision between no access, dial-up access, and high-speed access. Historically, the primary course of action of the federal, state, and local governments to address the digital divide has been to provide subsidies for digital communication technology (DCT) infrastructure investments in low-density regions. Such investments are often made without well-defined policies for technology use (Grimes, 1992), and are often designed to suit the suppliers of equipment rather than the potential users (Grimes, 2000). However, technology differences are only one of several possible causes of the rural - urban digital divide, and identification of the most important causes is vital for the creation of cost effective policies to bridge the current divide. A firm understanding of the various contributions of different factors to the current general divide in residential Internet access and the emerging divide in high-speed access is essential in order to create cost effective policies to close the gap in rural – urban residential Internet access.

This dissertation tests hypotheses regarding the roles of various factors in the existence, and persistence, of the digital divide. The relationships of these factors to the household decision on Internet access are then used to propose the most effective set of public policies for decreasing the gap in residential Internet access (both high-speed and dial-up) between rural and urban areas. The following factors influencing Internet access are directly addressed as part of this study:

Household characteristics

Household characteristics play a significant role in the decision on whether or not to adopt Internet access at home and may also be responsible for part of observed regional variations in rates of access. In particular, households with higher income and education levels (such as those found in urban areas) tend to have higher rates of Internet access (McConnaughey and Lader, 1998; Cooper and Kimmelman, 1999). Other household characteristics, such as race / ethnicity, age, and family structure may also affect Internet access rates – perhaps because Internet content may be less suited to the interests of particular population groups.

Temporal Resistance

Early adoption of an innovation is typically associated with regions that have higher income or education levels than the general population. As innovations become more widely diffused throughout the society, the associated benefits often become known with more certainty and the perceived costs of adoption decline (Brown, 1981). Thus, initially higher adoption propensities among high income and education households may dissipate through the natural process of diffusion from early adopters (who are more prominently located in urban areas) to the general population.

Telecommunications Infrastructure

Internet access requires use of, at the minimum, basic telephone service. While universal access to such service exists across all rural and urban areas in the U.S., not all households can access an Internet Service Provider (ISP) by a local call. The cost associated with having to make a long-distance call to connect to the Internet may play only a minor role in the adoption decision, as data from the 2001 CPS indicate that 5.1 percent of rural households with Internet access paid a long distance fee, compared to 3.5 percent of urban households. Thus, telecommunications infrastructure differences are unlikely to be a significant component of regional differences in plain-old telephone service (POTS) Internet use.

On the other hand, important differences do exist between rural and urban areas in the presence of digital communication technology infrastructure that allows for high-

speed Internet access. Data from a Federal Communications Commission survey in 2000 indicates that rural areas lag in high-speed Internet access even after controlling for demographics, with the 70 percent of ZIP code areas with broadband access containing 95 percent of the US population (Prieger, 2003). As noted, such high-speed access is necessary for households to fully benefit from increasingly common audio and video Internet content. Thus, DCT infrastructure differences may be becoming an increasingly important component of the rural – urban digital divide.

Social Networks

Network externalities may also play an important role in determining the magnitude of the benefits associated with residential Internet access. Given the strong local user base of many on-line communities (Horrigan et al. 2001), the value of the Internet to a household in a region may increase with the share of other households in the region that are connected. Inequalities stemming from household attribute differences may also be intensified by network externalities (Graham and Aurigi, 1997). For example, low-income households tend to be geographically clustered. A household in a low-income area is therefore likely to receive fewer benefits from home Internet access than a similar household in a high-income area because a lower proportion of other households in the same geographic cluster are using the Internet.

There is a crucial need to disentangle the roles of these various factors in order to generate and employ policies that directly deal with the most important causes of the divide. This need is addressed through the following objectives of this dissertation.

1.2 - Objectives

1. Determine the magnitude of the rural - urban digital divide in residential Internet access and detail how the gap has evolved over time.
2. Compare the emerging pattern of diffusion for high-speed residential Internet access to the initial pattern for diffusion of dial-up Internet access.
3. Determine the nature of the no access / dial-up / high-speed decision. In particular, does the household decide directly between these three alternatives? Or, is the

decision between high-speed and dial-up made conditional on the decision to access the Internet?

4. Determine the major factors underlying the current digital divide in general Internet access and the emerging divide in high-speed access. In particular, the impacts of the following four factors will be addressed:
 - a. Household characteristics. Differing characteristics between rural and urban households, particularly education and income levels, are likely to account for some of the differences in Internet access between regions.
 - b. Temporal resistance to adoption. Households in urban areas either have characteristics that make them more likely to be early adopters or are more rapidly exposed to the Internet as it diffuses from core to peripheral areas.
 - c. DCT infrastructure differences. High-speed infrastructure such as Digital Subscriber Lines (DSL) or broadband cable can potentially give rise to less expensive or higher quality access in urban areas than in rural areas.
 - d. Social Networks. Benefits associated with greater local content may arise from higher local rates of access for urban households relative to rural households, and may increase the relative value of Internet access in urban areas.
5. Given the underlying causes, generate policies to close the current digital divide and ensure regional equality in access to emerging digital information technologies.

1.3 – Study Structure

The remainder of this dissertation is structured as follows. Chapter 2 distinguishes the concepts of Internet access and Internet use, and discusses why the notion of access is more amenable to empirical analysis. The structure of the Internet is laid out, with particular attention paid to the availability of various technologies to rural households. The chapter then develops a conceptual framework for researching the rural – urban digital divide, building upon existing models of technology diffusion, adoption theory, and utility theory. This framework identifies four factors as probable causes of the divide, and the potential costs and benefits associated with each of the four factors is discussed through a review of the associated literature. Policy prescriptions that might be

appropriate for each factor are then presented to highlight the inherently different policy implications. Chapter 3 outlines the empirical analysis to be undertaken by detailing the data used and laying out the empirical framework. Chapter 4 presents the results of the empirical analysis, and Chapter 5 discusses the implications of these results for defining appropriate policies to deal with the digital divide.

Chapter 2: Conceptual Framework

This chapter first distinguishes between Internet access and Internet use. The structure of the Internet is described, focusing on the importance of its various components to the rural – urban digital divide. The remainder of the chapter uses three well-developed theoretical concepts (diffusion, adoption, and utility theory) to develop a framework with which to examine how residential Internet access is affected by the location of a household. This framework identifies four primary factors that play a role in the diffusion of the Internet among residential areas. Previous research on these factors is also reviewed.

2.1 - Internet Access vs. Internet Use

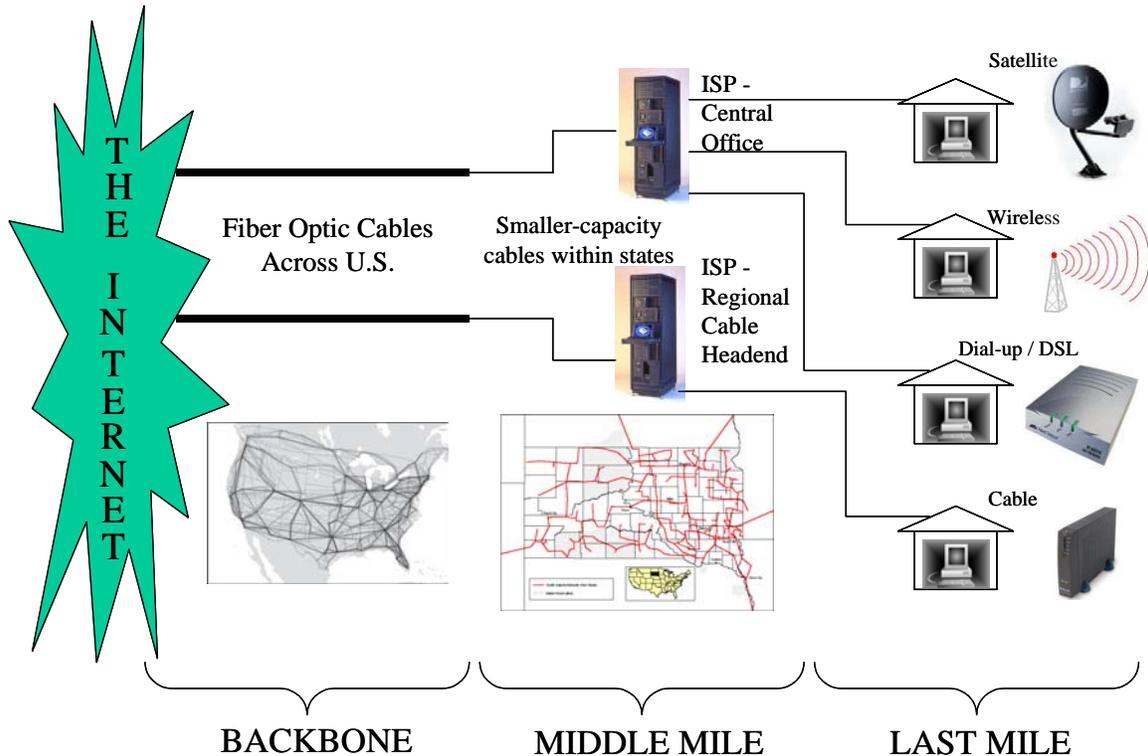
Before developing a theoretical framework for this study, it is important to distinguish between Internet access and Internet use. Residential access is typically defined as having a machine that is connected to the Internet in one's home. Use, on the other hand, refers to what people do with the medium once they have access to it (Hargittai, 2003). Some research has attempted to identify ways of distinguishing different types of Internet use. Warschauer (2002) suggests that conditions such as content, language, literacy, education, and institutional structures must be taken into account when assessing Internet use. The PEW Internet project has conducted surveys asking individuals what type of activities they are involved in on-line (Madden, 2003). However, determining the existence or magnitude of a digital divide in Internet *use* is not a trivial process. Specifically, quantifying the intensity of use is an arduous task. Surveys may be able to tell us the number of hours spent on-line for different types of activities, but given the wide range of computer-related abilities among the general population, this does not necessarily tell us how much use they got out of it. For example, two people may obtain the same amount of information after performing a search on a given topic, but one person might take four times longer to conduct their search. Given these difficulties and the fact that access is a necessary condition for use, comparing Internet access is a more pragmatic way of assessing the connectivity of households. Hence, this study will be concerned with determining the causes of

differences in Internet access in rural versus urban areas, as opposed to looking at differences in how those online are using the Internet.

2.2 - Structure of the Internet

The telecommunications infrastructure that comprises the Internet is made up of three distinct categories: the backbone, the middle mile, and the last mile. The backbone facilities are essentially the main arteries of the Internet, connecting major metropolitan areas at extremely high data rates using fiber optic cables.⁵ The middle mile connects the long-distance channels of the backbone to various Internet Service Providers (ISPs) located throughout the country. The last mile uses different types of technology (such as a dial-up modem, cable modem, DSL, satellite or wireless modem) to connect an ISP to the end user. The components of all three categories are depicted in Figure 2 below.

Figure 2. Internet Structure for Residential Access

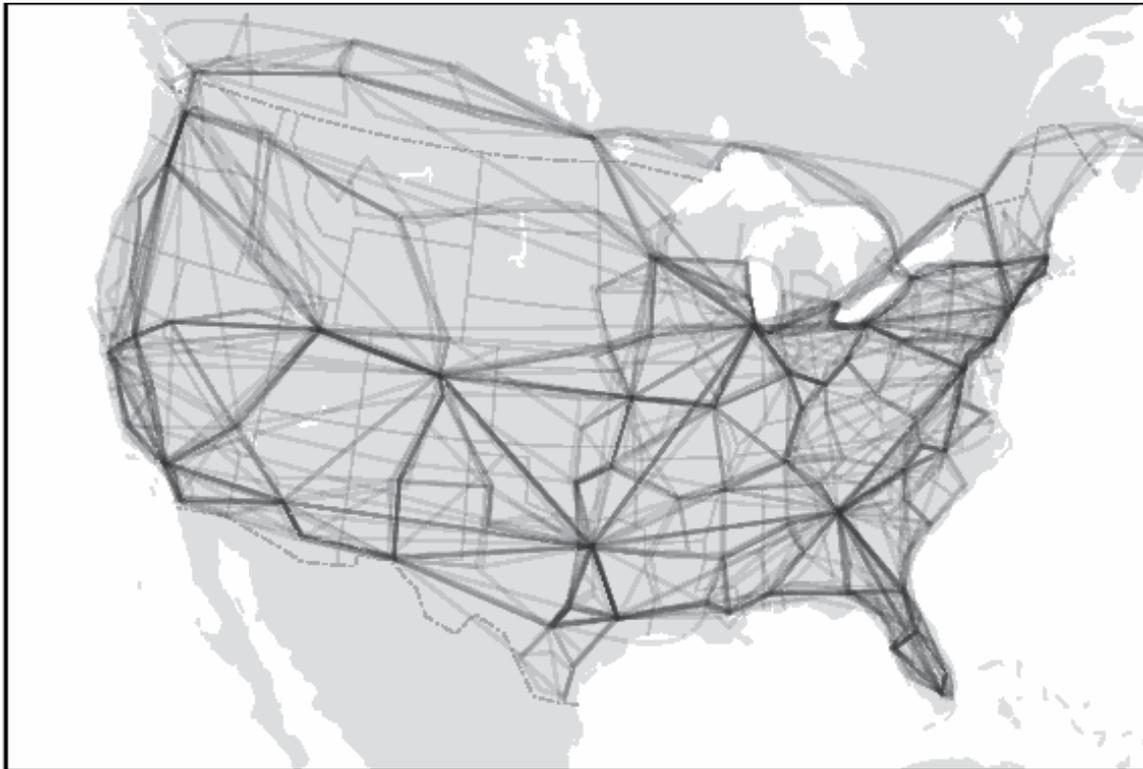


⁵ The NTIA and RUS (2000) note that a single fiber optic cable can carry 400 gigabits / second, which is equivalent to two million broadband signals at 200 kilobits / second.

Backbone

There are currently approximately 50 Internet backbone providers in the continental U.S., made up of cable systems, electric utilities, and municipalities. (TeleGeography, 2003). Figure 3 shows an overlay of competing U.S. fiber optic networks as of 2002. While it is true that the majority of these lines primarily connect urban centers, access to the backbone does not appear to be a problem for rural areas (NTIA and RUS, 2000; CBC, 1999, FCC 2000). This is because gaining access to the backbone can be accomplished in several different ways, with the most common being through local telephone providers. Given the near-universal status of phone service (NTIA, 1999) and the significant number of existing backbone facilities, accessibility to the backbone does not appear to be a major source of the digital divide.⁶

Figure 3. Overlay of Competing Fiber Optic Networks in the U.S., 2002



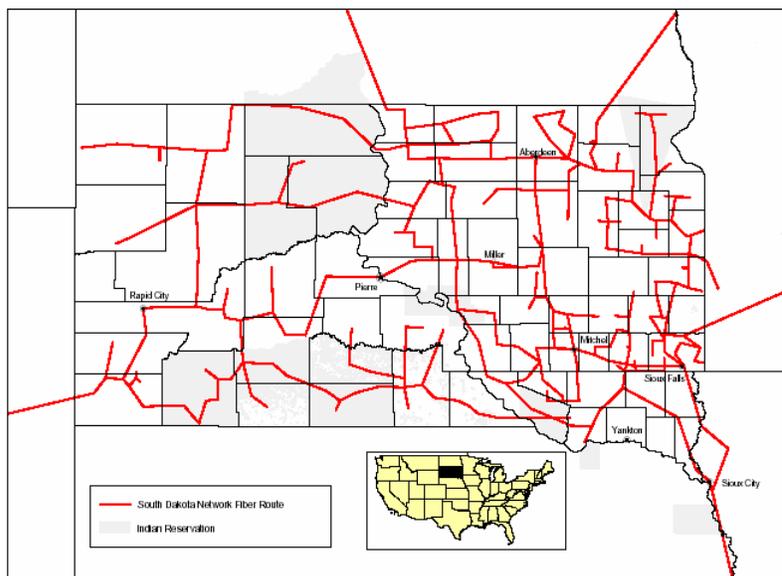
Source: TeleGeography Incorporated: U.S. Internet Geography 2003

⁶ The NTIA estimated that 94.1 percent of all U.S. households had phone service in 1999.

Middle Mile

The middle mile transports Internet traffic from the backbone to an ISP. ISPs are typically run through "central offices," which are telecommunications offices that are centralized in a specific locality to handle the telephone service for that locality, or through regional cable headends, which is the cable provider's version of the central office (Prieger, 2003). These offices typically have smaller capacity fiber optic cables running from the backbone to their offices in the area that they serve. Many middle mile facilities were originally built for ordinary phone and cable operations by incumbent telephone and cable providers (FCC 2002). Additionally, some states have invested in state-wide fiber optic networks, such as the one shown for South Dakota in Figure 4. While some organizations have claimed that the large distances between some rural areas and the backbone are problematic in terms of broadband access (NECA 2001), the FCC (2000) and NTIA (2000) indicate that the available middle mile facilities are adequate for providing both dial-up and high-speed access for rural areas. This is because extensive facilities for middle mile transports already exist, and the existing transports continue to have expanded capacity due to innovative techniques to compress and modulate the signals being carried. Hence, middle mile infrastructure also appears to be adequate across the U.S., and is not a prominent factor in the rural – urban digital divide.

Figure 4. Middle Mile Infrastructure in South Dakota



Source: FCC (2000)

Last Mile

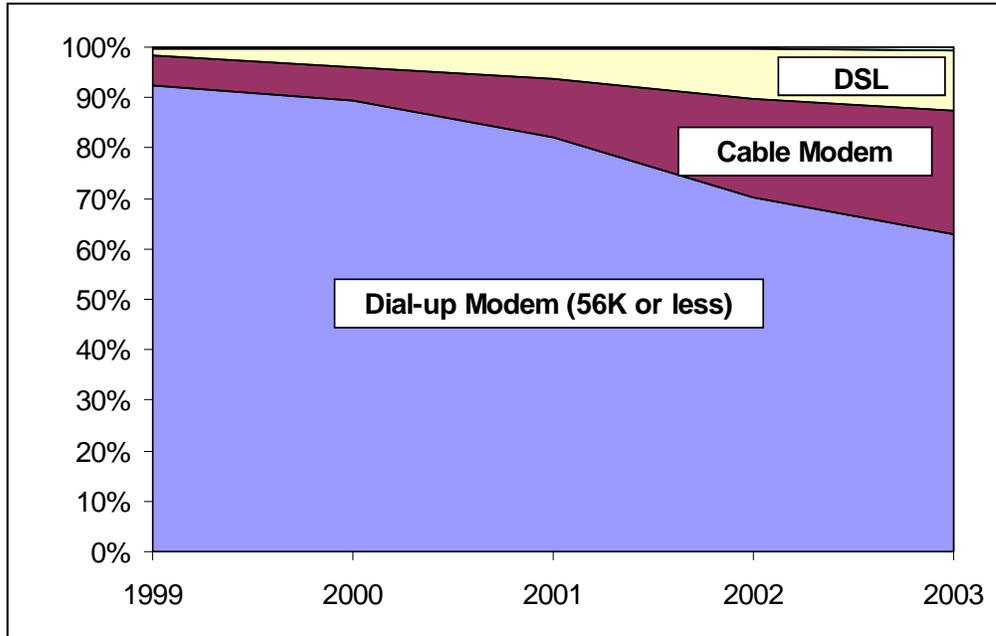
The infrastructure that connects ISPs to their customers is known as the last mile. This connection can be made over copper phone lines (such as with dial-up modems or DSL), over coaxial or fiber optic cable television lines, or via satellite / wireless communication systems. Some of these technologies are more conducive to urban areas, while some are more effective in rural areas. Understanding how each of these technologies work and recognizing the patterns in last mile infrastructure investment are important parts in determining the potential contribution of last mile infrastructure differences to the rural – urban digital divide.

In the early days of the Internet, residential access was limited to a dial-up modem that connected directly over the household's phone line, with maximum speeds reaching 28.8K in 1994 and 56K in 1996 (Encyclopedia Britannica, 2004). ISPs were sparsely located in these days, and some rural locations wishing to have access needed to place long-distance calls to reach the nearest ISP. As the Internet grew in popularity, ISPs became more commonplace, in part because of market forces equating supply with demand, but also because of various programs enacted to ensure "universal access" in the dial-up market. For example, the Rural Internet Access Authority in North Carolina stated its goal for the year 2000 as "the provision of local dial-up Internet access from every telephone exchange" in the state. By the late 1990s, the vast majority of households in both rural and urban areas of the U.S. were able to connect to the Internet via a local call.⁷ Thus, telecommunications infrastructure differences are unlikely to be a significant component of regional differences in plain-old telephone service (POTS) Internet access. While dial-up access was becoming nearly universal, demand for high-speed residential access was on the rise, perhaps due to the existence of high-speed access at places of work and the increasing frustration associated with long wait times and missed phone calls encountered when using dial-up modems. Cable companies, phone companies (through DSL), and satellite / wireless companies all entered the high-speed marketplace, where residential customers have increased dramatically since the

⁷ CPS data from 2000 indicate that 4.0 percent of urban Internet users paid a long-distance fee to access the Internet, compared to 4.7 percent of rural Internet users.

turn of the century. Figure 5 shows the type of last mile connections that Internet households have used over the period 1999 – 2003.

Figure 5. Use of Last-mile Technologies by Households with Internet Access



Source: Nielson Media Research, FCC Form 477

While the majority of households that access the Internet were still using dial-up modems in 2003, broadband technologies comprise a rapidly increasing share.⁸ Understanding how these high-speed options operate is critical to identify their potential to affect the rural – urban digital divide. Table 1 summarizes the residential broadband market structure and typical costs as of 2003. The start-up cost and the monthly fee associated with these services will impact both the adoption (yes /no) decision of the household and the type of service selected.

⁸ Recall that the terms "broadband" and "high-speed" are used interchangeably in this paper. DSL and cable modems are the dominant types of residential broadband access, as discussed later in this section.

Table 1. Residential Broadband Market Structure and Consumer Costs, 2003

	DSL	Cable	Satellite	Wireless
Market Share	32%	66%	<1%	<1%
Start-up Cost	\$0 - \$200	\$0 - \$200	\$400 - \$700	\$300 - \$500
Monthly Cost	\$30 - \$45	\$30 - \$50	\$50 - \$70	\$50 - \$80

Source: FCC Form 477, www.broadbandbuyer.com

Cable

Cable companies were the first to jump on board the broadband wagon, since most of the original lines laid back in the 1950's and 1960's were made of coaxial cable. The typical coaxial cable used for television services is a natural candidate for providing broadband data services, due to its large bandwidth (NTIA and RUS, 2000). However, most cable networks had to be upgraded to allow for two-way transmission, since these networks were originally built to provide video programming in only one direction. These upgrades consist of the installation of routers, switches, and a cable modem termination system to allow both upstream (from end user to ISP) and downstream (from ISP to end user) transmission (FCC, 2002).⁹ Most of these upgrades were performed during the late 1990's (NCTA, 2004). Upgrading a cable system for broadband service is not cheap (over \$65 billion was spent between 1996 and 2002 (NCTA, 2004)), and cable companies often need to estimate the high-speed subscription rates before investing in an area upgrade. Similarly, new cable system builds will not necessarily provide broadband to all customers – if the number of homes served is too low, cable companies cannot justify the additional expense of two-way service (NTIA and RUS, 2000). This has obvious implications for rural areas, since the density of homes in these areas is almost by definition lower than urban areas.

Cable television service has been estimated to be available to 81 – 86 percent of the U.S. population, with rural locations making up most of the unavailable areas (RUS 2000). Additionally, as noted above, cable systems located in rural areas are less likely to

⁹ It is worth noting that the downstream rate of data transfer is typically much higher than the upstream rate for most technologies (cable, asymmetrical DSL, wireless, and satellite). This is because the majority of end users are more interested in receiving information such as large pictures or video files than they are in sending these types of files. While symmetrical DSL does exist, it makes up less than 1% of the residential market (FCC Form 477, December 2003).

have broadband capability. Hence, the availability and performance capabilities of cable infrastructure for rural areas will likely have a significant impact on the high-speed digital divide, due in part to the dominance displayed by cable modems in the residential broadband market.

DSL

Digital Subscriber Lines are the second most widely used residential broadband service, and are gaining ground on cable. The vast majority (over 99%) of these lines are asymmetrical, meaning that the downstream rate of data transfer is faster than the upstream rate. A PEW Internet survey in early 2004 found that DSL makes up 42 percent of the home broadband market, compared with 31 percent a year earlier (Horrigan, 2004). When the first broadband subscription statistics came out in 1999, DSL made up only 16 percent of the market. Hence, DSL use is becoming much more commonplace in the residential broadband market. While cable systems require extensive upgrades to provide high-speed data services, most (but not all) of the copper loops in the existing telephone system can provide some form of DSL service by simply adding equipment at both ends of the connection (NITA and RUS, 2000). However, most of the loops that DSL cannot operate on are found in rural areas. This is primarily because of the length restriction required to support DSL services. In general, if a line extends more than 18,000 feet (approximately 3.5 miles) from the central office, the line is not DSL capable. These longer lines must incorporate inductors capable of maintaining voice quality, which block the higher frequencies needed for DSL broadband (FCC, 2002). While a large portion of incumbent local exchange carriers (ILECs) are investing in DSL technology deployment, this investment is primarily occurring in urban areas (FCC, 2000; Prieger 2003).¹⁰ The lower density associated with rural areas ensures that there will be fewer people within the 3.5 mile radius, dampening investment incentives for these areas. Some products extending the availability of DSL have been developed, but they are far from widespread deployment. For example, remote DSL

¹⁰ The Telecommunications Act of 1996 resulted in a dramatic increase in the number of competitive local exchange carriers (CLECs) that compete with the ILECs. However, CLECs do not represent a large portion of the market for DSL as of 2003. Data from the FCC Industry Analysis and Technology Division (2003) indicates that only 5% of all DSL lines are provided by non-ILECs.

Access Multipliers (DSLAMs) and loop extender technologies have the potential to reach more rural locations, but a survey of rural service providers in 2002 found that many were concerned about the cost of these items, coupled with historically low demand in their service areas (FCC, 2002).

The distance restriction associated with DSL broadband and the lack of significant deployment of DSL extenders supports the idea that the capability of the area central office (which determines whether or not DSL is possible) will have a significant impact on the emerging digital divide in high-speed access.

Satellite

While satellite Internet access has accounted for approximately one percent of all residential broadband use, it has the potential to play a large role in connecting primarily rural areas. A satellite dish installed at the end user's house connects with company satellites orbiting above the equator, rendering concerns about wire distances obsolete. As of 2004 there were two dominant satellite Internet providers in the market (DirecWay and Starband), both of who claim to serve all fifty states and can assure users of a clear satellite signal as long as they have a clear line of sight to the southern sky.¹¹ Prior to 2000, satellite Internet provided only downstream capability, with users resorting to dial-up modems to upload data. However, the technology has recently matured to allow for two-way capability via satellite, although upload speeds are still too slow to officially classify satellite access as "high-speed."¹² (FCC, 2002). Since the customer must pay for the installation of the satellite dish, this type of Internet access is inherently more expensive than either cable or DSL. Hence, it is unlikely that urban households will choose satellite access when they have the option to choose cable or DSL. But for those households without cable or DSL options, satellite access provides an opportunity for high-speed access, albeit at a higher price.

¹¹ DirecWay's internet site (www.DirecWay.com) indicates that 90% of U.S. households have this capability.

¹² While "high-speed" access is officially defined as 200 Kbps of data throughput (see footnote 3), satellite upload speeds are typically only around 50 – 100Kbps (although higher speeds are available for a price).

Wireless

Wireless broadband Internet access is similar to satellite in that it makes up a very small percentage of all broadband use, but could be important for rural households. Wireless service is split into two distinct categories, multipoint multichannel distribution systems (MMDS) and local multipoint distribution systems (LMDS). LMDS offers higher data rates than MMDS, but has a much shorter range (3 to 4 miles versus 35 miles for MMDS). A residential wireless customer typically purchases a radio transmitter / receiver that communicates with the provider's central antenna site, which is similar to the central office for DSL or the headend for the cable company (FCC, 2000). Wireless service providers can enhance their networks (via additional radio towers) with significantly less expense than that required to update cable or DSL infrastructure, primarily because they do not incur the sizeable costs of installing and maintaining wires running to the end user. Hence, in some instances it is much more feasible to install a wireless network in a remote area than it is a cable or DSL system. However, the density issue still remains for wireless service, as installing additional radio towers is an expense that wireless service providers will only undertake if it is profitable.

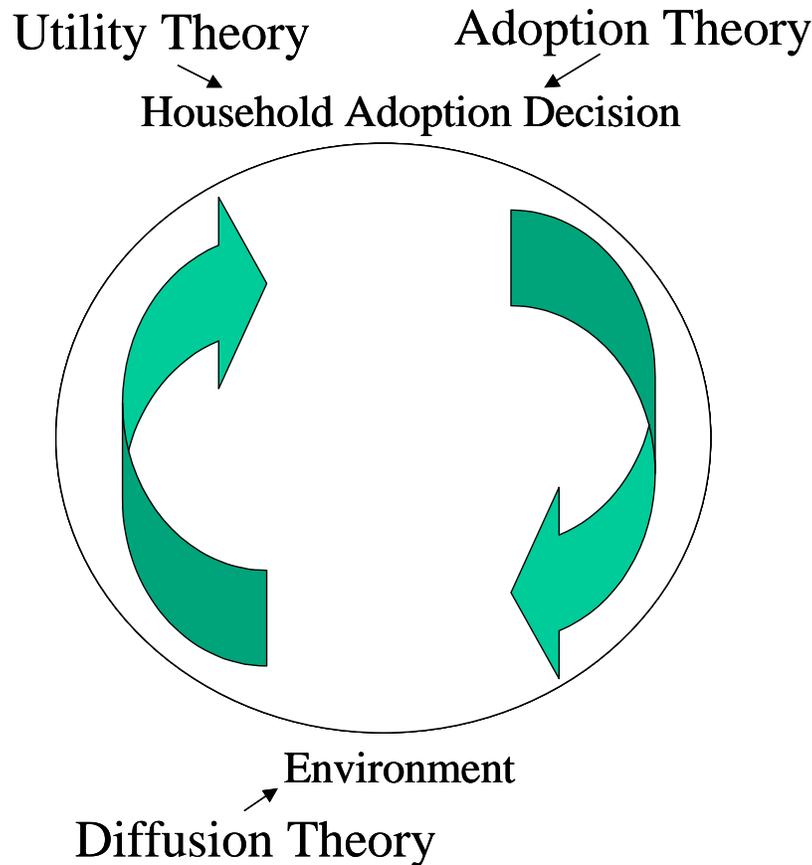
As Table 1 indicated, cable and DSL dominated the residential broadband market as of 2003, with satellite and wireless each accounting for less than one percent of all high-speed subscribers. Additionally, when they are available, cable and DSL technologies are cheaper for households to install and use, lowering the costs of access. However, as noted above, the wired nature of these networks generate disproportionately high costs for rural areas. Therefore, investment decisions with respect to cable and DSL networks in a given area may play an important role in the high-speed digital divide.

2.3 - The Big Picture

Studying residential Internet access requires an understanding of the relationship between households and their environment. Each individual adoption decision stems from the interaction of a household with its environment. Figure 6 summarizes the nature of this relationship in the cycle of technology diffusion. The socio-economic theories of adoption usually focus on either the household or the environment component. For

example, diffusion theory focuses on how the flow of information and household adoption decisions over time affect the broader environment. On the other hand, adoption theory isolates factors that influence the household decision making process, while utility theory models how the household makes the adoption decision.

Figure 6. Technology Diffusion Cycle



The household adoption decision is made at a single point in time. The theoretical structure for assessing this decision is adoption theory, which emphasizes the role of individual characteristics in the choice to adopt or not adopt the innovation. Adoption theory, in turn, is inextricably linked with utility theory, where the perceived costs and benefits of the innovation are compared for that particular household. These perceived costs and benefits are affected by the four factors discussed in chapter 1: household characteristics, temporal resistance to the innovation, social networks to which

the household belongs, and the technology environment (i.e. the available telecommunication infrastructure) where the household is located.

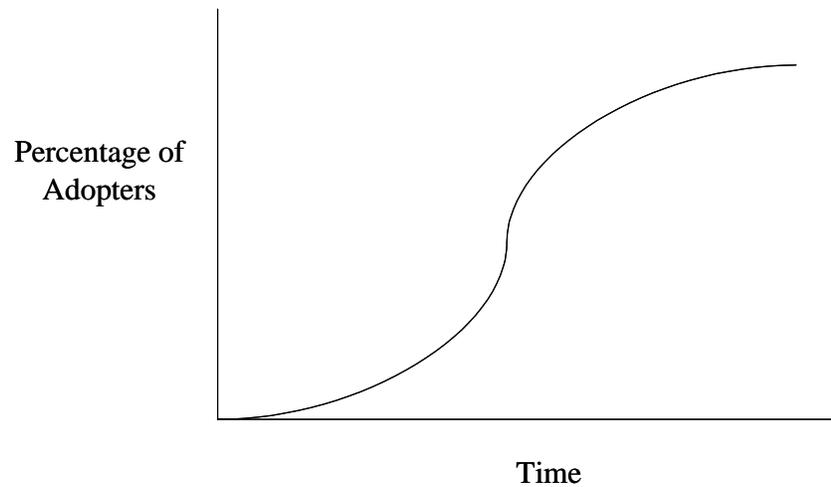
While the individual household adoption decision is affected by its environment, the environment is likewise affected by individual adoption decisions. In particular, aggregate demand stemming from such decisions determines the feasibility of investments, like digital communication technology (DCT) infrastructure, that underlie the type of Internet access available to households (Malecki and Boush, 2003). Individual household characteristics, social networks, and temporal resistance thus indirectly affect the technology environment and influence the adoption decisions of other households. Hence, even as infrastructure is playing a role in the adoption decision of a household, it is also being affected by this decision, since higher adoption levels may lead to future additional infrastructure investment. Diffusion theory incorporates this temporal dimension by allowing for feedback from the individual adoption decision into the environment.

The next three sections look at diffusion, adoption, and utility theory from a historical perspective, and show how these theories contribute to the analysis of residential Internet access.

2.4 - Diffusion Theory

The origin of diffusion theory can be traced back more than a century to a French sociologist named Gabriel Tarde. In 1903, Tarde came up with the notion of an S-shaped adoption curve (Figure 7), while implying that individuals learned about an innovation by copying someone else's adoption behavior. However, the study of the diffusion of innovations did not obtain academic notice until 1943, when sociologists Ryan and Gross published a study dealing with a new type of corn to be planted in Iowa fields. This study spurred additional research, culminating in Rogers' (1960) effort to synthesize the most significant findings and compelling theories. Rogers' work is generally accepted as the authoritative view on innovation diffusion, going through five editions as additional theories developed and more examples emerged over time.

Figure 7. S-shaped Curve Representing the Rate of Adoption over Time



The most recent edition of Rogers' Diffusion of Innovations (2003) defines diffusion as the interaction of four main elements:

- (1) An innovation
- (2) Communication channels
- (3) Time
- (4) A social system

For the purposes of this study, these four main elements are relatively well defined. The following section relates the most important concepts in diffusion theory to residential Internet access.

Innovation

This study is concerned with "residential Internet access," which can be defined as whether anyone in a household has accessed the Internet from that residence for a given year. Hence, the adoption decision occurs at the household level, and diffusion occurs as potential adopters make the decision to obtain Internet access.

Communication channels

Communication channels are simply means of conveying a message from one individual to another, in this case regarding the perceived benefits and costs of residential Internet access. These channels can vary from small-scale social networks (such as

discussions with friends or family members) to large-scale media in the form of television, radio, or newspaper. Given the potentially intimidating technological nature of the Internet, personal channels may be more successful than mass media messages in encouraging residential adoption, especially for first-time Internet users (Goolsbee and Klenow, 2002).

Time

Given the rapid diffusion of residential Internet access compared to other household-level innovations such as the automobile or telephone¹³, understanding the temporal dimension of the diffusion of Internet access is likely very important in determining how the rural – urban digital divide will evolve. Individual characteristics certainly play a role in determining when a household decides to adopt¹⁴, but place-based resistance could also be important. Temporal resistance in the flow of information over space, usually from a core environment (e.g. urban areas with a prevalent high-tech sector) to peripheral (remote) locations has been well documented, with the classic example dealing with hybrid corn in the Midwest (Griliches, 1957). Cultural and economic characteristics, along with social networks and the availability of relevant technology, are factors that influence temporal diffusion. These factors often differ by rural / urban location. For example, high-speed infrastructure is more prevalent in urban areas due to higher population density and per-capita income. Hence, the temporal resistance to Internet diffusion could be dramatically different in urban and rural areas, given the potential for technology adopter categories to vary across locations and the influence that infrastructure location could hold.

Social System

Rogers defines a social system as a "set of interrelated units engaged in joint problem solving to accomplish a common goal." In the context of residential Internet access, the "common goal" could be defined as achieving affordable access for all

¹³ Rogers (2003) notes that while the automobile and telephone required 40 and 50 years, respectively, to reach 10 percent of U.S. households, the Internet reached this threshold in only 5 years.

¹⁴ Section 2.5 on adoption theory discusses how these individual characteristics affect the adoption decision (see Table 2).

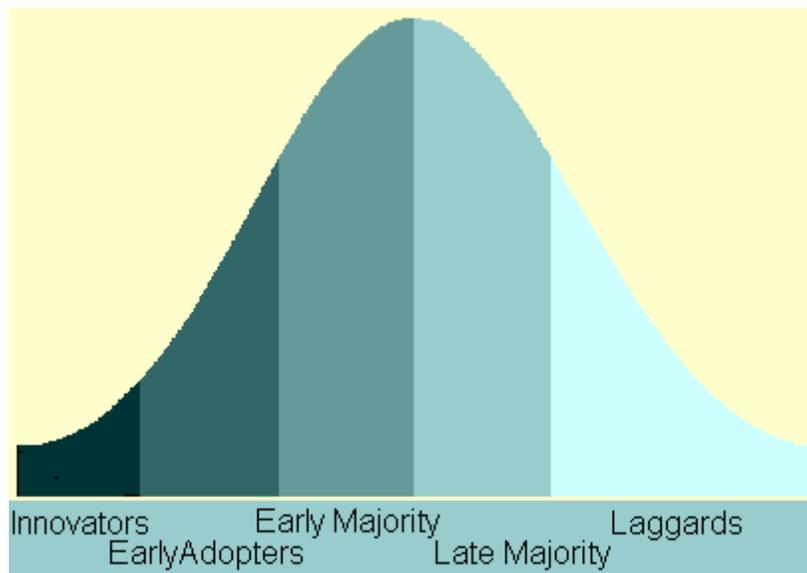
households who have determined that they want it. Hence, the social system is made up of all households in the United States, since this is the set of units that could potentially adopt the innovation. It is worth noting that the social system could have sub-sets of units who may be more enthusiastic about accomplishing the goal (for instance, a state pushing for broadband access for all of its households). Additionally, local social networks may impact adoption propensities to the extent that "peer pressure" or amount of local content influences the perceived costs and benefits for a household.

As described above, the theory of innovation diffusion is certainly applicable to residential Internet access, but by itself is not an all-encompassing story. The broad depiction of the factors involved is often not amenable to empirical research. The next section delves into adoption theory and why individual households choose to adopt innovations.

2.5 - Adoption Theory

While diffusion theory focuses on the temporal aspect of the innovation's dispersion and how the environment responds to this dispersion, an interrelated yet distinct theoretical structure (adoption theory) determines whether or not each individual agent decides to implement the innovation. Adoption theory emphasizes the role of individual characteristics in determining whether or not adoption occurs on a case-by-case basis. The primary difference between diffusion theory and adoption theory is that diffusion occurs among units of a social system, while adoption takes place in the mind of an individual (Rogers and Shoemaker, 1971). Brown (1981) praised the work of Rogers (1960) for including both theories, indicating that its focus on the adoption process makes it the most parsimonious theoretical framework available for innovation diffusion. Rogers hypothesized that potential adopters were normally distributed in terms of their innovativeness, and that five separate categories of potential adopters existed. These categories range from innovators – the first to adopt an innovation – to laggards, who are set in their ways and suspicious of change. Figure 8 displays this distribution and the categories that theoretically make up all potential adopters, while Table 2 lists some characteristics that are associated with each of the categories.

Figure 8. Technology Adopter Categories



Source: Rogers, 2003

Table 2. Technology Adoption Categories and Typical Characteristics

Adopter Categories	Typical Characteristics
Innovators	Eager to try new ideas. More years of formal education, higher income. Higher social status. Risk takers.
Early Adopters	Role models for other members of social system. Upward social mobility, able to lead opinions.
Early Majority	Interact frequently with peers. Deliberate before adopting new ideas.
Late Majority	Respond to pressure from peers. Approach innovation with caution, unwillingness to risk scarce resources
Laggards	Resistant to innovation. Suspicious of change, hold on to traditional values. Isolated.

Source: Rogers, 2003

As noted previously, time is one of the four main elements of diffusion. The categories of adopters listed in Table 2 certainly play a role in the diffusion process by influencing the time frame within which individuals choose to adopt. Hence, the

adoption decision is affected by the characteristics of the household, along with numerous other factors that dictate the perceived costs and benefits of the innovation. To formally discuss how this adoption decision is made, we turn to utility theory.

2.6 - Utility Theory

Adoption theory uses individual characteristics as the primary determinant of whether or not an innovation will be implemented (Rogers, 2003). For example, the characteristics described in Table 2 suggest that some individuals are more likely than others to adopt innovations quickly. However, in making the decision to adopt, the individual must weigh the perceived costs and benefits of the innovation. Many of these costs and benefits are non-monetary and must be inferred using utility theory. Due to the discrete nature of the adoption decision, proposed models conform to the theory of random utility, where a choice is observed as what is assumed to be the result of optimizing behavior (Ben-Akiva and Lerman, 1985).

Utility theory has been present since the early days of economics, when Adam Smith noted that value in use is not necessarily the same as value in exchange. Jeremy Bentham is credited in the late 18th century with using the term "utility" to refer to amounts of pleasure, and was also the first to associate utility with numerical magnitude. The next century saw numerous additions to utility theory, which became generally accepted in the study of economics by the 1870s (Stigler, 1950). After much research on the existence and measurability of utility, most papers turned to utility maximization and its relationship with demand.¹⁵ With utility maximization accepted as a common goal in economic theory, empirical studies of adoption in the twentieth century incorporated this concept into their hypothesized models. A common theme among these studies was that the likelihood of adoption was modeled as a function of the perceived costs and benefits of the innovation, which in turn were influenced by individual characteristics, such as education and income levels (Feder, Just, and Zilberman, 1985; Mahajan and Peterson, 1985).

The standard household adoption model, then, can be specified via a utility function aggregating the perceived costs and benefits of the innovation. In particular, a

¹⁵ See, for example, Jevons (1911) and Marshall (1925).

household will adopt the innovation if the perceived utility of the benefits outweighs the perceived disutility of the costs. Thus, the decision of the household will be determined by the following equation:

$$y_i^* = U(B_i) - U(C_i) \quad (1)$$

where $U(B_i)$ and $U(C_i)$ are measures of the perceived benefits and costs of the innovation for household i , respectively. y_i^* is a latent measure of the aggregate utility associated with the innovation, and the household will be observed to adopt if $y_i^* > 0$. Empirically, incorporating measures of utility into a model is not feasible (particularly when dealing with cross-section data), and most analysis tends to link the costs and benefits to variables that can be observed (Finke and Huston, 2003; Wellington 1993). Hence, a typical adoption model will take the form:

$$y_i^* = U(B_i) - U(C_i) = \beta' X_i + \varepsilon_i \quad (2)$$

where X_i is a vector of observable characteristics that influence perceived benefits and costs, β' is the associated parameter vector, and ε_i is the associated error term.¹⁶ While y_i^* is a latent variable, it is observed that $y_i = 1$ (meaning the innovation is adopted) if $y_i^* > 0$, and $y_i = 0$ otherwise. Given the binomial nature of this choice, the model is typically specified via a discrete choice framework, such as logit, probit, or linear probability models. The choice among which type of model to use is made based on the inherent restrictions and the ease of parameter interpretation associated with each model.¹⁷

Adoption and utility theory are clearly applicable to residential Internet access. A household is expected to adopt Internet access if the benefits of this access outweigh its costs. The associated costs and benefits, in turn, can be linked to observable household characteristics and conditions. The main question, then, is what characteristics and conditions are related to the utility associated with Internet access, and what is the nature of their relationship? Drawing on diffusion, adoption, and utility theory, three main categories of such characteristics (household characteristics, DCT infrastructure, and social networks) are discussed below. It is important to note that temporal resistance is

¹⁶ For simplicity, it is assumed that the measures of perceived benefits and costs enter the household decision process linearly. Given this relationship between Internet access and observable characteristics, the discussion that follows is framed around the perceived costs and benefits of particular characteristics.

¹⁷ See section 3.3 for a further discussion of these tradeoffs among the commonly used discrete choice models.

not included as a unique factor because the adoption decision analyzed in this dissertation comes from a cross-section of households at a unique point in time. However, the relative importance of these characteristics may change over time, and adding a temporal dimension to the analysis will capture this type of effect.

Household Characteristics

Household characteristics influence both the costs and benefits associated with residential Internet access. Many of the household characteristics shown for the five categories of potential adopters (Table 2) are difficult to observe or assign a value to – for example, the ability to lead opinions or the level of interaction with peers. Some, however, are easily observable and quantifiable, such as income and education level. Other household characteristics may play a role in determining the utility of Internet access, particularly with respect to the amount of content online catering to households with the characteristic. Examples of such characteristics are race / ethnicity, age, family structure (especially the number of school-age children), and gender. For example, an Indian household may derive less benefit from Internet access than a White household due to smaller amounts of content tailored to their ethnic background. Similarly, a household with several school-aged children may see more of an advantage in residential Internet access than a household with no children because of the educational opportunities for children presented by the Internet. Hence, a number of observable household characteristics can influence the costs and benefits associated with Internet access.

Digital Communication Technology Infrastructure

DCT infrastructure may also increase the benefits associated with residential Internet access. As noted in section 2.2 above, different types of high-speed access technologies vary by the way they carry the broadband signal and by their costs to consumers. The presence (or lack thereof) of various types of infrastructure will have an impact on the cost to, and hence utility of, residential households considering what type of Internet access to adopt. Thus, the availability of telecommunications infrastructure where a household is located certainly plays a role in determining the level of Internet

access adopted. However, the role played by this infrastructure is likely to be more important for high-speed access than it is for dial-up access. In the early days of the Internet, some households had to pay for a long-distance call to access the Internet, which certainly increased the costs associated with such access. Such long-distance fees have become scarce, implying that telecommunications infrastructure differences are unlikely to be a significant component of regional differences in plain-old telephone service (POTS) Internet access. However, access to high-speed infrastructure is highly unequal between rural and urban areas. This is because population density is the primary influence for infrastructure investments (Strover, 2003). Hence, households located in less dense areas may not be able to obtain high-speed access via the same infrastructure as households located in more populated areas.

Social Networks

While individual characteristics are certainly important to the adoption decision, the social networks that a potential adopter is connected to may also play an important role in determining the magnitude of the benefits associated with residential Internet access. Recent research suggests that many on-line communities consist mostly of local users (Horrihan et al. 2001). Given this strong local user base, the value of the Internet to a household in a region may increase as the share of other households in the region that are connected increases (note that this is essentially the definition of a network externality). Inequalities stemming from household attribute differences may also be intensified by network externalities (Graham and Aurigi, 1997). For example, low-income households tend to be geographically clustered. A household in a low-income area is therefore likely to receive fewer benefits from home Internet access than a similar household in a high-income area because a lower proportion of other households in the same geographic cluster are using the Internet.

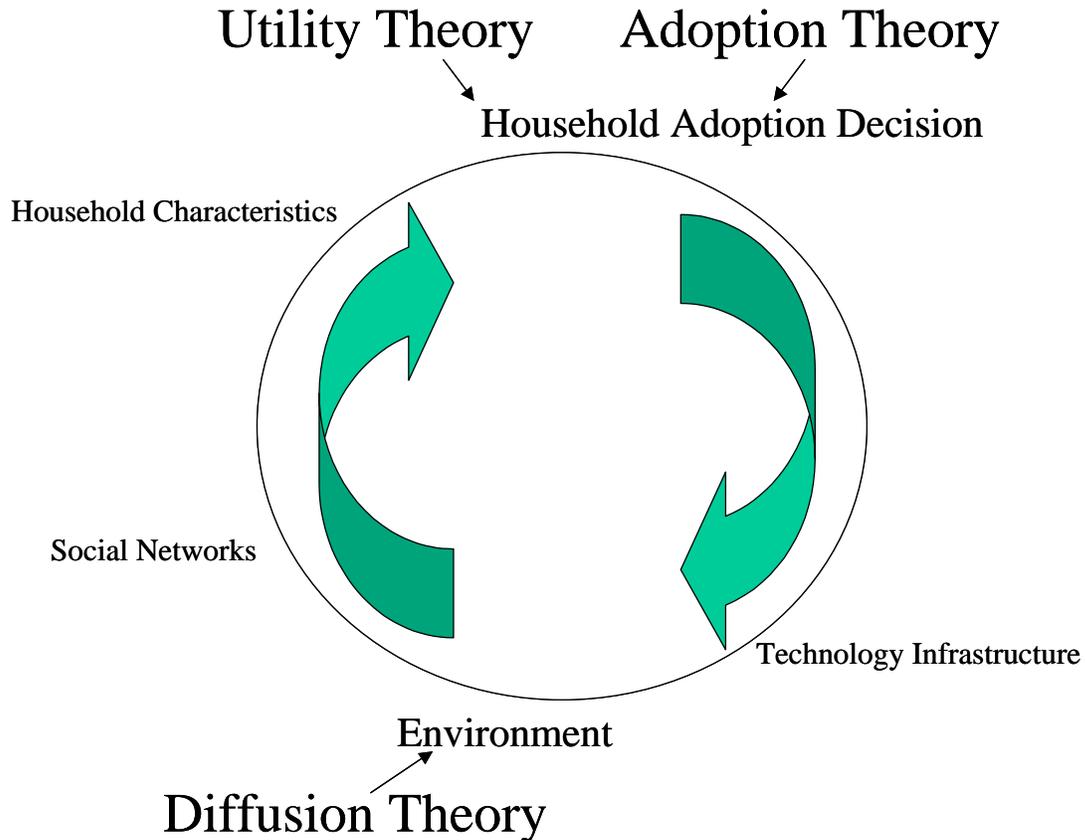
2.7 - A Unified Theory of Residential Internet Access

The previous three sub-sections have described the structures of diffusion, adoption, and utility theory and shown how they are applicable to studying residential

Internet access. In light of these underlying structures, we now return to the big picture and compose a unified framework for studying residential Internet access.

Figure 6 showed the technology diffusion cycle, which dealt with the interaction between the household adoption decision and the environment. In order to empirically analyze a decision on whether or not a household will implement the innovation at a given point in time, adoption and utility theory are used to outline factors influencing the perceived costs and benefits associated with the innovation. As discussed in section 2.5, there are three primary factors that affect the perceived costs and benefits of residential Internet access at any point in time – household characteristics, telecommunications infrastructure, and social networks. Hence, the household adoption model is specified in terms of these factors. Diffusion theory adds a temporal dimension to the analysis, and determines how the environment (DCT infrastructure) responds to the aggregation of individual household adoption decisions. Figure 9 shows how these factors fit into the technology diffusion cycle.

Figure 9. A More Detailed Technology Diffusion Cycle



While diffusion theory determines how the environment evolves with changes in household adoption decisions, the temporal dimension also allows the environment to continuously affect adoption decisions. At any point in time, the household makes a decision based on its current characteristics, social networks, and infrastructure availability. As time moves on, temporal resistance changes with additional information about the innovation and the potential for investments in infrastructure to occur. As more information becomes available on the innovation, it may become more appealing to households with "late adopter" characteristics, such as the late majority or laggard categories in Table 2. Hence, a household could decide not to adopt at time t , but decide to adopt at time $t+1$, without any change in characteristics or infrastructure levels. These higher rates of adoption could spur investments in high-speed infrastructure, demonstrating the cyclical nature of the relationship.

The next section reviews the relevant literature associated with each of the four factors in an effort to accomplish three goals:

- (1) To determine the potential costs and benefits of residential Internet access associated with each factor.
- (2) To provide descriptive statistics on each factor. The data is primarily taken from the 1997, 1998, 2000, 2001, and 2003 CPS supplemental surveys dealing with residential Internet access. Supplemental data sources are used for statistics regarding DCT infrastructure.
- (3) To discuss policy prescriptions that may play a role in bridging the portion of the urban – rural digital divide stemming from that factor.

2.8 - Review of the Literature

Internet access has been a hot topic in several disciplines since the mid-1990's. Sociologists explore its implications for human interaction, communication experts delve into how it has affected the transmission of information, and economists are interested in how people decide to adopt and use the innovation as well as its implications for economic interactions. This section reviews literature on these questions, focusing on the potential costs and benefits from adoption of Internet access associated with the four factors identified in the previous sections.

Household Characteristics

Household characteristics, particularly education and income, influence the relative costs and benefits of Internet access and, thus, the overall level of utility that the household receives from such access. Household income and education levels have been found to play a consistently important role in determining residential Internet access propensities (e.g. McConnaughey, Nila, and Sloan, 1995; McConnaughey and Lader, 1998; Cooper and Kimmelman, 1999). Table 3 provides descriptive statistics from the CPS data on U.S. households that access the Internet and do not access the Internet from home in 1997, 1998, 2000, 2001, and 2003. For all years, households that access the Internet from home have, on average, higher levels of income and education. Moss and Mitra (1998) determine that monetary costs associated with Internet access are likely to

be a greater deterrent to access for low-income households, since most costs are fixed - such as the purchase of a computer. They also find that Internet access is not an essential household good, implying that households with higher incomes will have relatively low costs of access, since a smaller percentage of their income will be spent on essential goods. Similarly, more educated household heads may have greater exposure to digital technologies, which increases the perceived benefits and lowers the perceived costs of Internet access (Mills and Whitacre, 2003). Differential propensities to access the Internet by income and education levels may also be intensified by the relative abundance of content targeted to high income and well- educated groups. Evidence suggests that there is less content on the Internet catering to the ‘underserved’ population - those without high rates of household access (Children’s Partnership Report, 2000; Greenman, 2000). For example, information on entry-level jobs, low rent housing, and neighborhood assistance programs that is of particular use to low-income households is rarely posted on the Internet. As a result, the perceived benefits of home Internet access by low-income and low educational asset households may be low relative to the perceived benefits for higher income and educational asset households.

Other household characteristics shown in Table 3 are also likely to influence residential Internet access. In 1995, McConnaughey, Nila, and Sloan found that households that access the Internet are disproportionately likely to have White than non-White and non-Hispanic than Hispanic household heads. These racial – ethnic differences may stem, in part, from underlying differences in income and education levels, but Internet content may be less suited to the interests of households headed by Blacks, other non-White racial groups, and Hispanics, thereby generating less utility associated with accessing the Internet from home, *ceteris paribus*. Differences in residential Internet access among racial and ethnic groups continued to exist in the late 1990s and into the turn of the century (McConnaughey and Lader, 1998; NTIA, 2000). The age of the household head may also influence the propensity to access the Internet from home. Rose (2003) found that younger household heads are more likely to have been exposed to digital technologies in school and, therefore, more comfortable using the Internet from home. Hence, the perceived benefits from Internet access are potentially higher for households with younger heads. A household’s propensity to access the

Internet is also influenced by the composition of the household. Children likely have both positive and negative effects on a household's propensity to access the Internet. Children are often exposed to computers and the Internet at school (NTIA, 2002), thus increasing the household propensity to access the Internet from home. Further, with an additional child the benefits of Internet use are spread over an additional household member, while the cost of home Internet access is usually fixed. On the other hand, an additional child lowers the disposable income of the household, which decreases the household's propensity to access the Internet. Similarly, the propensity to access the Internet may be higher in households headed by married couples than for single households, as the costs of access are split between two adults. Early studies of Internet access also found females to have a lower propensity to use the Internet (Bimber, 2000; NTIA 1999). However, more recent findings indicate the Internet gender gap has dissipated (NTIA, 2002). Whether or not household members are employed can also have an impact on the likelihood of residential Internet access. Forestier, Grace, and Kenney (2002) find that households with an employed member may also be more likely to access the Internet at home, due to the potential benefits associated with performing work-related tasks at home. The utility of Internet access for retired household heads is debatable: these individuals tend to be older, and hence are less likely to use the Internet (Bell, Ready, and Rainie, 2004); however, the additional time spent at home may increase the benefits associated with having Internet access available there.

Table 3. Household Characteristics by Internet Access

Household Characteristics	Internet					No Internet				
	1997	1998	2000	2001	2003	1997	1998	2000	2001	2003
Income										
Under \$20K	8.38	7.75	8.06	8.40	9.70	34.14	34.38	36.00	40.01	40.67
\$20K - \$50K	31.37	30.87	31.53	32.15	32.35	40.98	42.52	42.47	42.07	42.29
\$50K +	60.24	61.36	60.40	59.44	57.95	24.87	23.09	21.52	17.91	17.03
Education										
No High School	1.22	2.82	3.69	4.53	4.49	18.41	20.34	23.21	25.23	25.69
High School	11.52	17.26	20.25	21.88	22.02	33.47	34.55	35.31	36.49	37.25
Some College	30.18	29.38	30.93	31.38	30.91	26.35	25.44	24.25	24.17	23.53
College Plus	57.09	50.54	45.13	42.21	42.56	21.77	19.66	17.23	14.11	13.52
Race / Ethnicity										
White	90.92	90.00	88.48	87.22	86.28	83.14	82.02	79.66	79.31	78.01
Black	4.74	5.12	6.52	7.64	7.66	13.25	14.52	16.65	17.43	16.93
Other	4.34	4.87	5.00	5.13	6.06	3.62	3.46	3.70	3.26	5.06
Hispanic	3.50	3.85	4.94	5.70	6.88	8.78	9.64	11.73	12.58	14.93
Household Composition										
Married	68.69	70.11	67.74	66.58	64.82	52.50	49.03	45.01	40.09	39.41
Age of Head	41.26	43.34	43.93	44.34	45.28	48.79	49.41	50.48	51.70	51.32
Headed by Male	73.17	67.91	61.45	59.03	57.55	57.85	55.15	51.17	48.74	48.91
Avg. # of Children	-	-	0.47	0.77	0.73	-	-	0.47	0.47	0.47
Employment										
Employed	87.91	83.11	80.00	78.64	76.42	65.01	62.96	59.86	54.65	52.93
Retired	4.60	6.87	8.59	9.06	10.81	20.75	22.66	24.71	27.18	27.49
Internet at work	42.78	29.09	30.15	43.23	39.30	8.25	11.31	12.20	13.71	11.66

Sources: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, and 2003.

Given the potential for various characteristics to affect the utility associated with residential Internet access, differences in characteristics are likely to account for some of the rural – urban digital divide. For many of the characteristics discussed above, rural areas have distinctly different characteristics from urban areas. Table 4 displays descriptive statistics for rural and urban households. In all years, the average rural household has lower income and education levels than its urban counterpart. As discussed above, lower income and education levels tend to decrease the perceived benefits from Internet access, implying that rural access rates would be lower simply based on these characteristics. Other characteristics also differ between rural and urban areas. Several characteristic differences may tend to decrease rural access rates relative to urban rates, while others may tend to increase them. In particular, rural areas display lower levels of diversity, have older household heads, and are more likely to have dual-headed households than urban areas. Rural households are also more likely to be headed by a male, and have a much larger percentage of retired household heads. Rural households are also consistently less likely to have a computer in the house, and are less likely to access the Internet from work or elsewhere (as well as from home). Some of the characteristics associated with rural areas should increase the likelihood of residential

Internet access (lower levels of racial and ethnic diversity; higher levels of dual-headed households in rural areas relative to urban areas), while others should decrease the likelihood (older household heads and lower rates of access from work).

One point that is somewhat hidden by the Internet statistics in Table 4 is that the dial-up rates for rural and urban areas have converged over the period 2000 – 2003. In 2000, most of the general digital divide (13.8 percentage points) was with regard to dial-up access, as very few households had high-speed connections. However, by 2003 the *high-speed* digital divide (13.8 percentage points) more than made up the entire *general* digital divide (12.8 percentage points)! Hence, dial-up rates have become nearly equal over this time frame, and the digital divide seems to have shifted to one of high-speed access. This trend will be important for the decompositions performed in this dissertation and for policy prescriptions involving the digital divide.

Table 4. Household Characteristics by Rural / Urban Area

	Urban					Rural				
	1997	1998	2000	2001	2003	1997	1998	2000	2001	2003
Family Characteristics										
Income										
Under \$20K	27.72	24.44	21.30	20.60	20.37	39.16	36.78	33.52	32.67	31.25
\$20K - \$50K	38.83	38.57	36.49	35.74	35.43	41.87	41.97	42.41	40.75	40.67
\$50K +	33.45	36.97	42.20	43.62	44.19	18.97	21.24	24.07	26.56	28.09
Education										
No High School	14.17	13.86	13.33	12.86	12.12	21.56	21.67	19.90	19.03	17.87
High School	28.06	27.72	26.60	26.62	26.29	37.49	37.62	37.20	36.75	36.67
Some College	27.50	26.93	27.63	28.35	27.94	24.84	25.03	25.40	26.88	27.54
College Plus	30.26	31.49	32.45	32.17	33.64	16.11	15.69	17.50	17.34	17.92
Race / Ethnicity										
White	83.11	82.92	82.12	82.12	81.48	89.40	89.49	89.25	89.53	88.59
Black	12.83	12.86	13.22	13.19	12.40	8.19	8.13	8.06	7.92	7.70
Other	4.07	4.23	4.66	4.69	6.12	2.41	2.38	2.69	2.55	3.71
Hispanic	9.00	8.99	9.78	9.98	11.25	3.74	4.25	4.56	4.34	5.89
Household Composition										
Married	54.22	54.07	54.25	53.74	53.64	58.51	58.25	58.04	56.91	57.23
Age of Head	47.17	47.22	47.06	47.16	47.22	49.25	49.64	49.81	50.09	50.03
Headed by Male	60.07	58.15	55.26	53.90	53.71	61.17	60.91	57.37	55.84	55.12
Avg. # of Children	-	-	0.47	0.64	0.60	-	-	0.47	0.62	0.63
Employment										
Employed	70.37	70.39	70.29	69.33	68.09	64.77	64.27	64.50	63.03	61.08
Retired	17.40	17.45	16.66	16.48	16.54	21.21	21.42	21.53	21.26	22.48
Internet at Work	15.35	17.92	22.12	31.91	30.13	7.51	9.79	11.90	20.23	18.64
Computer Characteristics										
Computer in House										
Internet	41.75	47.62	57.45	62.53	67.63	30.53	36.17	44.38	50.70	57.03
Home										
Home	17.69	30.44	46.67	56.57	61.23	9.08	18.05	32.90	42.94	48.40
Work										
Work	15.35	17.92	22.12	31.91	30.13	7.51	9.79	11.90	20.23	18.64
Elsewhere										
Elsewhere	16.56	22.40	26.67	39.27	44.08	8.57	14.01	16.54	27.35	29.97
High Speed										
High Speed	-	-	5.47	11.82	25.01	-	-	2.48	4.30	11.22

Sources: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, and 2003.

Using a set of household characteristics similar to the ones outlined above, Mills and Whitacre (2003) find that rural – urban differences in these household characteristics, particularly lower income and education levels in rural areas, account for approximately two-thirds of the rural – urban gap in general residential Internet access in 2001. Thus, from a policy standpoint, closing the rural – urban digital divide may require addressing some of the broader inequities in rural income and education levels. However, as Mills and Whitacre (2003, p.239) note, "market forces have shown little tendency to dissipate such social disparities, and public support for initiatives to address social inequalities that underlie the digital divide is limited."

The characteristics identified above can also play a role in the emerging high-speed digital divide. High-speed access is a more recent innovation than dial-up, only becoming available for households in the late 1990's (compared to the early 1990's for dial-up). Hence, only households with very low resistance to new innovations may currently be likely to adopt this new technology. Rogers (2003) indicates that education and income levels are correlated with the speed of adoption, a notion that is reinforced by Strover (2003) in the context of high-speed Internet access. Additionally, the merits of high-speed access are important when considering the types of households that might adopt. The primary benefit associated with high-speed access is a much quicker connection, which is of importance to experienced users of the Internet, as well as those interested in applications requiring large amounts of information, such as video files or data transfers (Malecki, 2003). Thus, high-speed connections may offer higher benefits to those households with characteristics that might be associated with these applications – for instance, households with children might be interested in on-line gaming, while employed household members might use the connection to perform work from home (Telecommunications Industry Association, 2003). Similarly, retired or older individuals may only be interested in the Internet as a means of communicating with their younger friends or family members via email. E-mail services are very similar between dial-up and high-speed access, and households may not be able to justify the additional expense of high-speed access if e-mail will be the primary use of the service. To date, no empirical analysis has identified the importance of household characteristics in the adoption of high-speed access relative to dial-up access.

DCT Infrastructure

As noted previously, it is unlikely that telecommunications infrastructure differences are a significant component of regional differences in plain-old telephone service (POTS) Internet access. This is because of the near-universal status of phone service, and the spread of ISPs that has allowed for local access calls from almost any location. However, important differences in DCT infrastructure that allows for high-speed Internet access do exist in rural and urban areas. The literature on this topic consistently indicates that urban areas are more likely than rural areas to benefit from higher levels of DCT infrastructure. Currently, there are two primary methods of obtaining high-speed Internet access – Digital Subscriber Lines (DSL) and cable modem.¹⁸ Data from the FCC in 2003 indicates that these two methods made up 99 percent of the residential broadband market, with satellite and wireless access making up the other 1 percent. Given that DSL and cable are wire-intensive technologies, it is no surprise that urban areas have the upper hand when it comes to the presence of this type of infrastructure. Prieger (2003) finds that rural areas lag in high-speed Internet access even after controlling for demographics, as the 70 percent of all ZIP code areas that have broadband access contain 95 percent of the US population. Glasmeier and Wood (2003) obtained similar results, although their analysis was limited to the state of Pennsylvania. Particularly, they found that while 83 percent and 95 percent of urban households had DSL and cable high-speed Internet infrastructure (respectively) available to them, only 21 percent and 63 percent of rural households had similar availability. Further, rural areas continue to be significantly less likely to have broadband DCT infrastructure even after controlling for income differences. Grubestic (2003) performed regression analysis on the availability of DSL in Ohio, and concluded that household density and urban / rural status were significant explanatory variables. The NTIA and RUS (2000) generalized this finding for the entire U.S., noting that far fewer digital subscriber lines exist in low population density areas. Gabe and Abel (2002) showed that the gap in integrated services – digital network (ISDN) lines between rural and urban areas has widened rather

¹⁸ See Chapter 3 for an in-depth discussion on the structure of the Internet, and the roles that DSL and cable modems play.

than closed in recent years. Malecki and Boush (2003) looked at the capability of telecommunication infrastructure in four southern states, and found that the highest capability (meaning DSL availability) was disproportionately found in metropolitan areas. As evidence of how these inequalities affect residential demand, the CPS data on high-speed Internet access in Table 4 shows that only 11.2 percent of rural households used the Internet through a high-speed connection in 2003, compared to 25 percent of urban households. Further, the rural – urban gap in high-speed access widened dramatically between 2000 and 2003. As noted, such high-speed access is necessary for households to fully benefit from increasingly common audio and video Internet content. As an indication of the diffusion of high-speed access, Appendix A provides maps of the U.S. showing the number of high-speed providers by area code in 2000, 2001, and 2003. In 2000 there were large areas with no reporting providers (30 percent of all zip codes), and very few with 7 or more providers. By 2003, however, only 9 percent of zip codes had no reporting high-speed providers, and a much higher percentage of zip codes with 7 or more providers existed. Hence, the diffusion of high-speed infrastructure increased dramatically over these three years, resulting in a change in the DCT infrastructure environment. As a result, DCT infrastructure differences may be becoming an increasingly important component of the high-speed rural – urban digital divide.

As further evidence of these differences, Table 5 below displays descriptive statistics on the percentage of the U.S. population that live in counties with DCT infrastructure.¹⁹ This information is presented for the years 2000, 2001, and 2003. Rural areas lag behind urban areas in terms of both cable and DSL infrastructure for each year of data. Furthermore, between 2003 and 2001, the gap between rural and urban areas has actually increased in terms of cable infrastructure, while it has decreased for DSL infrastructure.

¹⁹ The source of and assumptions behind this infrastructure data are discussed in Section 3.1.

Table 5. Percent of U.S. Rural / Urban Population Living in Counties with DCT Infrastructure

	2000	2001	2003
Cable			
Rural	4.66%	5.47%	44.10%
Urban	25.08%	27.68%	75.75%
DSL			
Rural	3.43%	6.39%	29.55%
Urban	21.61%	32.05%	42.39%

Sources: *Television and Cable Factbook*, NECA Tariff #4 Data

As Table 5 shows, rural areas currently lag urban areas in terms of DCT infrastructure, and some studies indicate that this trend is likely to continue. Greenstein and Lizardo (1999) find that levels of DCT infrastructure are associated with high population density, degree of urbanization, and regional levels of economic activity. Gabe and Abel (2002) determined that demand aggregation plays a vital role in DCT infrastructure investment decisions. Further, many rural areas have made infrastructure investments for high-speed access at public institutions, leaving an even smaller residual group of potential users potentially served by private infrastructure investments. For example, state and local governments in rural areas of several states have set up their own dedicated networks, taking away the potential for these institutions to serve as "anchor tenants" for a new broadband network that could also service residential users (Parker, 2000). Additionally, line densities in many low population rural areas are much too low to support market investments in DCT infrastructure (Cooper, 2002). Evidence also suggests that attempts to increase industry competition through the Telecommunications Act of 1996 have seen limited success in terms of creating incentives to expand infrastructure investments in low-density regions (Cooper and Kimmelman, 1999; Warf, 2001). For example, Greenstein (2000) finds that the factors that lead small ISPs to offer new services, like large demand bases and prior local investments, are disproportionately found in urban areas. Zolnierick et al. (2003) find that facilities-based entry by new local competitors is more likely to occur in large urban telephone markets. Hence, urban areas are more likely than rural areas to benefit from the presence of DCT infrastructure capable of delivering high-speed Internet access.

The general conclusion that rural areas lag urban areas in terms of availability of DCT infrastructure is not a new one. Numerous government organizations have attempted to enact legislation purporting to deal with this problem, including several bills currently under review by the U.S. Congress.²⁰ Most legislation deals with subsidizing DCT infrastructure in rural areas or inducing additional infrastructure investment, perhaps by relaxing competition requirements on the local telephone companies (this idea is a component of the Tauzin-Dingell bill, initially introduced in 1999). This has spurred additional research dealing with the efficacy of such legislation. Grimes (2000) is perhaps the most critical of the majority of these programs, indicating that they are poorly thought out and designed more to suit the objectives of equipment and service suppliers than of potential users. He highlights examples in Scotland where infrastructure was highly underutilized after being built for rural areas. The most important conclusion that Grimes reaches is that simply providing telecommunications infrastructure does *not* guarantee that it will be used. He notes several additional aspects that need to be addressed in order to affect demand for such services, including increasing awareness of its availability, and providing training on how to use it.

Malecki (2000) concurs that DCT investment in rural areas is not a "magic solution," and indicates that rural areas need to demonstrate that they have adequate demand for such technology. Similarly, Fox and Porca (2001) indicate that the promotion of such services in rural areas should be encouraged only when demand will develop, and also imply that this expansion should be funded by user fees when possible. On the other hand, there is significant support for the provision of government subsidies to rural areas. The NTIA and RUS (1999) point out that additional support for high-cost areas will be necessary to provide universal service. While most legislation has dealt with subsidizing infrastructure, other bills (such as Tauzin-Dingell) focus instead on encouraging investment from other sources such as local telephone companies. For DSL access in particular, the question seems to be whether access would be more effectively provided by incumbent local exchange carriers (ILECs) or competitive local exchange

²⁰ See, for example, bills S 2307 (Rural Broadband Enhancement Act), S 877 (Broadband Internet Regulatory Relief Act), and HR 2420 (Internet Freedom and Broadband Deployment Act, or Tauzin-Dingell).

carriers (CLECs).²¹ The Telecommunications Industry Association (2003) details how forcing incumbent local telephone companies to provide competitors with access to broadband facilities will undermine the incumbent companies' incentives to invest in additional facilities. On the other hand, Grubestic (2001) believes that such competition is necessary, and will lead to additional investments from competitors (as opposed to the incumbents). However, since 95 percent of all residential DSL access was provided by incumbent telephone companies in 2003, encouraging investment from competitors could be a lengthy process (FCC Industry Analysis and Technology Division, 2003). Hence, the debate on whether investments in DSL infrastructure will come from incumbents or competitive exchanges is open-ended at best. Crandall, Jackson, and Singer (2003) also explore levels of private investment in high-speed infrastructure. They attempt to predict residential broadband penetration rates using a logistic curve, and come up with a rate of 65% by 2010.²² They then imply that these higher rates will spur investment in cable and DSL infrastructure, up to \$1.6B per year for DSL, and \$4.7B per year for cable. However, it is unclear whether this investment will be for infrastructure in previously unserved areas or to promote competition in areas with existing service.

Differences in DCT infrastructure between rural and urban areas may be an important component of the rural – urban digital divide in high-speed access. If high-speed access is not available via DSL or cable modem in a given area, the likelihood of a residence obtaining high-speed access is greatly diminished. But it is worth noting that technologies (such as satellite or wireless access) do exist with the capacity to allow high-speed access in unwired areas. However, these technologies are still in their infancy, making up only 1 percent of all residential broadband subscriptions in 2003. Public policies, like infrastructure subsidies, may be necessary to ameliorate some spatial discrepancies in rural – urban area DCT infrastructure if infrastructure discrepancies are documented to be an increasingly important factor in the digital divide (Parker, 2000; Youtie 2000). As noted earlier, a number of federal, state and local initiatives have been developed to support telecommunications infrastructure investments in low-density

²¹ ILECs are carriers historically franchised to provide local phone service, while CLECs compete with ILECs by leasing lines or building their own facilities.

²² Due to the nature of the paper (discussing the effects of ubiquitous broadband adoption), these rates are highly optimistic - predicting that 95% of all households will have broadband access by 2021.

regions. For example, the Rural Access Authority in North Carolina was created to provide local dial-up Internet access from every telephone exchange in the state. Other states like Washington and Virginia have also provided grants to rural areas to increase high-speed Internet access. Additionally, the 109th U.S. Congress is in the process of reviewing several bills dealing with the promotion of high-speed access in rural or underserved areas. However, no empirical studies have documented the strength of the link between DCT infrastructure and household Internet access (high-speed or POTS). Therefore, understanding the role that DCT infrastructure plays in the rural – urban digital divide is necessary to determine the relevance and weight that should be given to infrastructure- oriented initiatives. In the adoption framework, DCT infrastructure subsidies lower the costs associated with high-speed access, but the perceived benefits may still not be high enough to achieve adoption. The provision of additional services, such as technology centers showing how to use the equipment, may be necessary to spur penetration rates in rural areas. Such technology centers are not uncommon, as the Community Technology Center's Network (www.ctcnet.org) has over one thousand such centers throughout the U.S., funded by both public (U.S. Department of Health and Human Services) and private (Microsoft, AT&T, AOL) institutions.

Social Networks

CPS data from 2000, 2001, and 2003 document substantial regional variation in Internet access that could give rise to the existence of network externalities. Figure 10 divides the U.S. into nine distinct regions, and Table 6 displays rates of dial-up and high-speed access for rural and urban areas within these regions.²³ Some regions clearly have higher rates of access than others (particularly high-speed). Such variation may imply that some regions have higher benefits from network externalities than others.

²³ The nine regions are based on the breakout used in the Current Population Survey supplement. Analysis of states comprising each region indicates that rates of residential Internet access between states in a region are relatively similar. See Appendix C for state-level Internet access rates.

Figure 10. Nine Regions of the United States

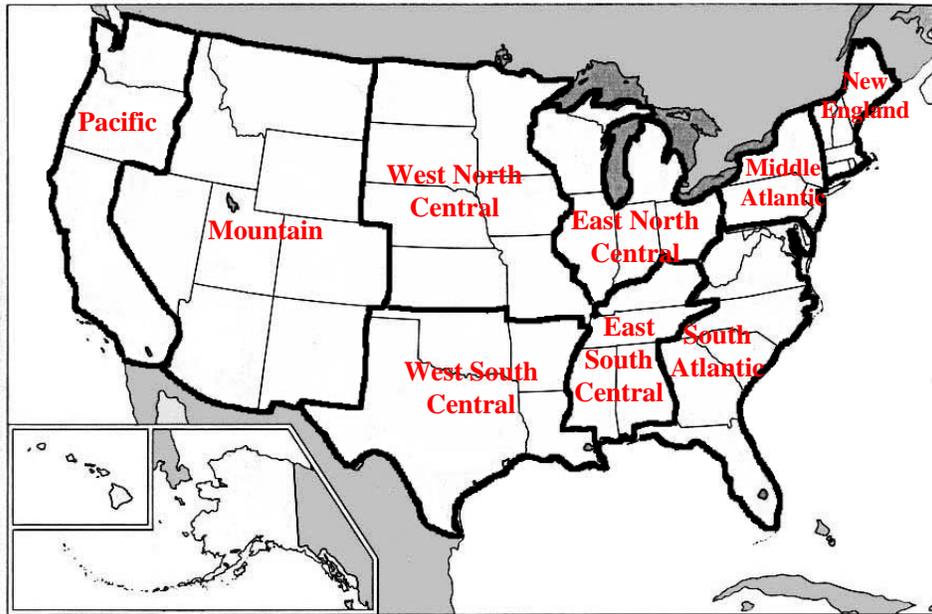


Table 6. Residential Dial-up and High-Speed Access Rates by Region

Region	Dial-up						High-Speed					
	2000		2001		2003		2000		2001		2003	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Pacific	39.31	48.81	44.20	46.18	42.95	36.47	2.88	6.64	6.12	17.89	16.68	29.46
Mountain	39.49	42.21	42.45	47.83	41.38	40.59	1.86	5.09	4.82	9.96	11.33	20.46
West North Central	30.30	42.21	39.42	45.48	37.05	36.68	2.29	5.77	4.43	12.64	13.17	27.32
West South Central	25.54	33.62	32.80	40.89	33.26	34.14	2.28	5.22	2.30	8.94	8.84	17.33
East North Central	33.89	41.28	43.88	46.10	39.68	37.33	2.02	4.20	2.69	8.51	9.15	21.47
East South Central	23.50	38.29	30.17	43.12	32.50	34.69	3.61	4.54	5.98	7.05	9.45	20.43
South Atlantic	28.92	39.83	36.23	45.73	35.64	37.62	2.42	4.83	3.02	9.62	9.75	22.95
Middle Atlantic	37.25	41.57	40.14	44.93	37.85	35.04	3.16	5.14	7.58	12.43	15.40	27.39
New England	48.83	43.96	52.11	45.57	45.98	36.01	2.06	7.46	6.73	15.91	17.90	30.95
Entire U.S.	30.42	41.20	38.64	44.75	37.18	36.22	2.48	5.47	4.30	11.82	11.22	25.01

Source: CPS Computer and Internet Use Supplements, 2000, 2001, and 2003

The role of network externalities is well accepted in the telecommunications industry. In fact, the notion of an externality (the value that a network member obtains increases as more members enter the network) was used as the economic foundation to justify subscribership subsidies in an effort to promote universal phone service (Barnett and Kaserman, 1998). Network externalities may also play an important role in determining the magnitude of the benefits associated with residential Internet access. A survey of Internet users conducted by Horrigan et al. (2001) found that many Americans

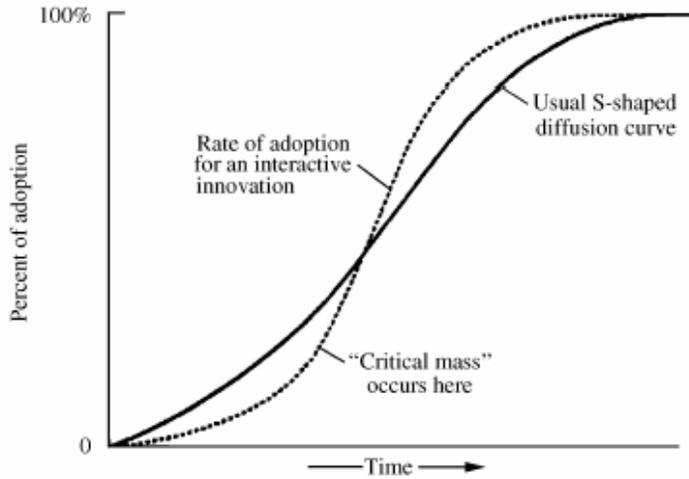
used the Internet to become more connected with their local community. This included setting up meetings, arranging neighborhood gatherings, and checking on local community news. Hence, the value of the Internet to a household in a region may increase as the share of other households in the region that are connected increases. Cremer (2000) composes a simple model to show that as Internet access in a region approaches universal status, new opportunities exist for how the network might be used (such as broadcasting services). Graham and Aurigi (1997) find that inequalities stemming from household attribute differences might also be intensified by network externalities. For example, low-income households tend to be geographically clustered. A household in a low-income area is therefore likely to receive fewer benefits from home Internet access than a similar household in a high-income area because a lower proportion of other households in the same geographic cluster are using the Internet. De Castro and Jensen-Butler (2003) note that for the majority of individuals, the Internet is more suited to interacting with people already known than to establishing new contacts. Similar to Graham and Aurigi, de Castro and Jensen-Butler demonstrate that network externalities can intensify differences in access associated with household attributes, implying that adoption will begin sooner and diffusion will occur faster in more educated and higher income regions when network externalities are present.

The idea that network externalities play a role in the adoption of technologies is supported by Goolsbee and Klenow (2002), who found that people are more likely to buy their first home computer in areas where a high fraction of households already own computers or when a large share of their friends and family own computers. New adopters can potentially benefit from these prior adopters due to the physical proximity of knowledge on setting up and using the computer. This result is strengthened by the fact that these patterns persisted even after controlling for household characteristics such as education and income levels. Additionally, the regression analysis performed in this study indicated that households accessing the Internet appeared to be highly influential in promoting computer ownership in nearby households. Estimates from the study imply that these spillovers are extremely important in the adoption of home computers – potentially doubling the rate of adoption. Goolsbee and Klenow suggest that promoting

users who interact heavily with subgroups that have low rates of adoption would be beneficial in increasing adoption in these areas.

Mahler and Rogers (1999) offer a somewhat different view on network externalities as they apply to the adoption of interactive telecommunications innovations, such as the Internet. They emphasize the concept of a "critical mass," defined as the minimal number of adopters for an innovation, after which the further rate of adoption is self-sustaining. Figure 11 displays the S-shaped diffusion curve of an interactive innovation that is assumed to respond to critical mass, alongside a "usual" S-shaped diffusion curve. The key features associated with this critical mass curve are the lower rate of adoption in the initial phase, and a much higher rate of adoption once the critical mass point has been reached. This concept is more applicable to some innovations, such as fax machines and video conferencing, and less applicable to others, such as cellular phones (which connect to an existing base of phone users). The authors note that, due to the interactive nature of these innovations, network externalities affect the utility of *both* potential adopters and previous adopters. Hence, this backward flow of utility to past adopters may provide additional incentive to spread even more positive perceptions about the innovation. So, according to Mahler and Rogers, network externalities may play an even larger role for interactive innovations such as the Internet than they would for "normal" innovations. In this context, these network externalities may have become increasingly important once critical mass was reached, which the authors speculate to be 20 – 30 percent of all potential adopters (around 1997 – 1998 according to CPS data).

Figure 11. Rate of Adoption for Two Types of Innovations



Source: Mahler and Rogers, 1999

While the majority of the literature focuses on the benefits associated with network externalities, Mason and Valletti (2001) note that perceived costs could also arise as the number of network users increases. In particular, during peak usage times (such as evening hours), negative externalities in the form of congestion can exist when too many users of a single ISP are on-line at one particular time. Although this concept deals more with Internet "use" than it does with "access," the reduced number of ISPs in rural areas may play a role in decreasing the utility associated with Internet access if households believe that congestion will affect their on-line use. Hence, the negative aspects of network externalities must also be considered. Additionally, it is feasible that network externalities might not play a large role in the adoption of the Internet. While some literature (such as Horrigan et al discussed above) believe that communicating with a local neighborhood is an integral part of Internet access, other sources (Faulhaber and Hogendorn, 2000) believe that the primary use of the Internet is to communicate over long distances, and hence demand externalities will be minimal. The weight that network externalities carry in the current Internet environment has yet to be evaluated empirically. As shown in Table 6 above, some states (and regions) are clearly more connected than others, while some display large digital divides between rural and urban areas. These discrepancies may indicate that network externalities do exist, and are substantial; or they may simply be the work of the other three factors in the adoption / diffusion process.

Mills and Whitacre (2003) find a strong positive association between residential Internet access and aggregate regional levels of Internet access after controlling for household characteristics. However, this relationship does not necessarily arise from network externalities, as higher rates of regional access may be due to underlying differences in telecommunications infrastructure or core to periphery diffusion patterns (discussed below) that are not controlled for in their study.

If network externalities are found to play an important role in the current rural – urban digital divide, then there is likely to be considerable reinforcement of initial advantages in local Internet access; with households in areas with high aggregate levels of access showing generally high adoption propensities, while households with similar characteristics in areas with low aggregate levels of access show lower adoption propensities. If the current rural – urban digital divide stems, in part, from network externalities, then concerted efforts to promote widespread access in low access localities through digital-villages or subsidized area user groups may be warranted.

Temporal Resistance

The last of the four factors affecting the adoption and diffusion of residential Internet access is temporal resistance. There are two separate components of temporal resistance to an innovation. The first is based on the characteristics of potential adopters and how these characteristics affect the time of adoption. Rogers (2003) best defines these characteristics, and his five proposed categories dealing with the temporal aspect of adoption were shown previously in Table 2. The second component deals with the flow of information from core to peripheral areas. Griliches (1957) first reported this phenomenon for his study of hybrid corn in the Midwest. He noted that farmers closer to the developers and producers of the hybrid seed were the first to plant it, while those farmers further away adopted the innovation later as information on the benefits of the seed slowly diffused outward to them. The literature associated with both of these components sheds some light on their potential effects on residential Internet access.

In his discussion on adoption theory, Rogers (2003) indicates that early adopters of an innovation typically have notably higher incomes or education levels than the general population. Additional studies have supported this notion – Feder, Just, and

Zilberman (1985) found that higher income levels may allow households to better bear the risks associated with a new innovation, while Schultz (1975) indicated that higher education levels allow households to access and process information on innovations more efficiently. The diffusion of residential Internet access is no exception to this pattern, as Table 7 indicates that early adopters had relatively higher income and education levels than those who adopted later. This pattern holds true for both general access and high-speed access, although the difference between early and later adopters appears to be larger for education levels than for income levels. One primary reason for the existence of this pattern is that as innovations become more widely diffused throughout the social system, the associated benefits often become known with more certainty and the perceived costs of adoption decline. Early adopters may, however, retain an information seeking and processing skills gap relative to late adopters. Mason and Hacker (2003) note that late adopters are initially more likely to use computers and the Internet for personal rather than economic purposes as communication and information technologies become “domesticated” (Cummings and Kraut 2002). While this type of differential Internet *use* may be of concern for future studies, access comparisons are only concerned with characteristics that affect the adoption decision. Thus, initially higher adoption propensities among high income and education households may dissipate through the natural process of diffusion from early adopters (who are more prominently located in urban areas) to the general population. Major innovations like the telephone, automobile, and TV have all shown such a pattern of diffusion (Rogers, 2003). From a policy perspective, if the current divide is due to the differential presence of early adopters in urban areas, it is likely to dissipate over time without outside intervention. Convergence in the rural – urban rates of residential Internet access can still be increased, however, by subsidizing access among households without early adopter characteristics.

Table 7. Income and Education levels for Internet and High-speed Adopters

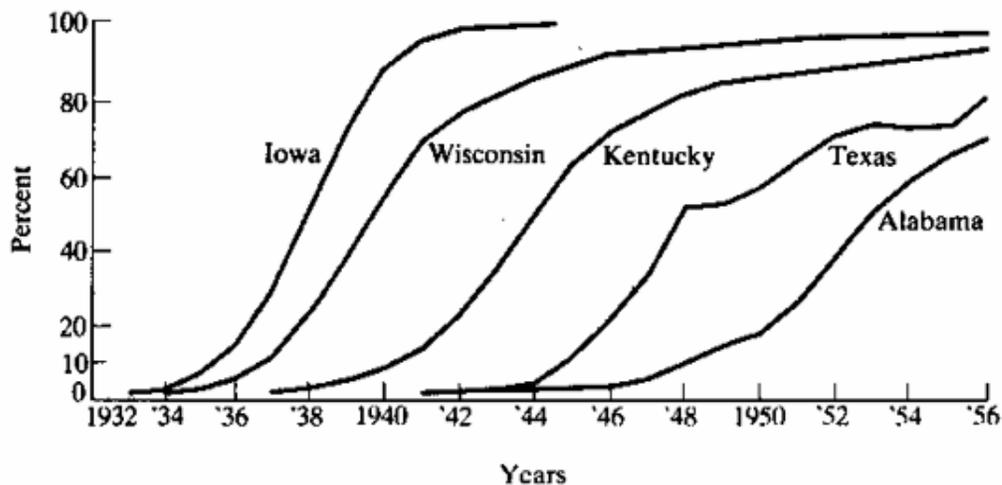
	Internet Access (any type)					High-speed		
	1997	1998	2000	2001	2003	2000	2001	2003
Family Income								
Under \$20K	8.38	7.75	8.06	8.40	9.70	6.12	7.14	7.59
\$20K - \$50K	31.37	30.87	31.53	32.15	32.35	26.45	24.73	25.03
\$50K +	60.24	61.36	60.40	59.44	57.95	67.43	68.13	67.38
Education								
No High School	1.22	2.82	3.69	4.53	4.49	3.42	3.00	3.04
High School	11.52	17.26	20.25	21.88	22.02	15.78	16.96	17.07
Some College	30.18	29.38	30.93	31.38	30.91	29.24	30.38	29.05
College Plus	57.09	50.54	45.13	42.21	42.56	51.54	49.64	50.84

Sources: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, and 2003.

Alternatively, the dynamics of residential Internet access adoption may be directly linked to place through temporal resistance in the flow of information over space, usually from a core environment (e.g. urban areas with a prevalent high-tech sector) to peripheral (remote) locations. Strover (1999, p. 16) notes that “rural citizens often lack the skills or knowledge to realize the importance of digital information and communication and ... to assure digital infrastructure in their areas.” Hence, individuals who reside closer to such technology may be more likely to understand its value earlier than those who are removed from it. This is demonstrated in the original example by Griliches (1957), whose research on hybrid corn documents not only an S-shaped pattern of adoption for an area over time, but also similar patterns occurring later in more remote areas (in this case further away from the developers of the seed, who were located in the Corn Belt near Iowa (Figure 12)). The crucial components of this S-shaped adoption pattern are an increasing adoption phase as more households are exposed to the innovation and a plateau as all households have been exposed and have decided whether or not to adopt it. Additionally, regions that are further away from the "technological center" of the innovation will respond as information flows to them. Looking at this idea from a telecommunications standpoint, Townsend (2001) discusses the rise of network cities in the U.S. These cities are metropolitan areas with high levels of advanced telecommunications technologies that dominate measures of Internet connectivity, such as number of domain names and links to other metropolitan areas. These areas could be considered as the core environment of the Internet, with information diffusing from each of them. Households located in these areas may be more likely to have access earlier

than households that are further away, simply because they are located near an area where the technology is dominant, and are the first to see the potential benefits of such access. As in the early adopter framework, the rural – urban digital divide will lessen over time as information on Internet access diffuses to rural areas. Again, however, the speed of diffusion can be increased through policy interventions. In this case efforts need to focus on increasing knowledge about the Internet and its uses in rural areas. In fact, there have been notable successes at public schools in closing the rural – urban knowledge gap among youths with respect to information technologies (NTIA, 2002). The provision of community technology centers (focusing more on the adult population) would be one way to increase knowledge in rural areas.

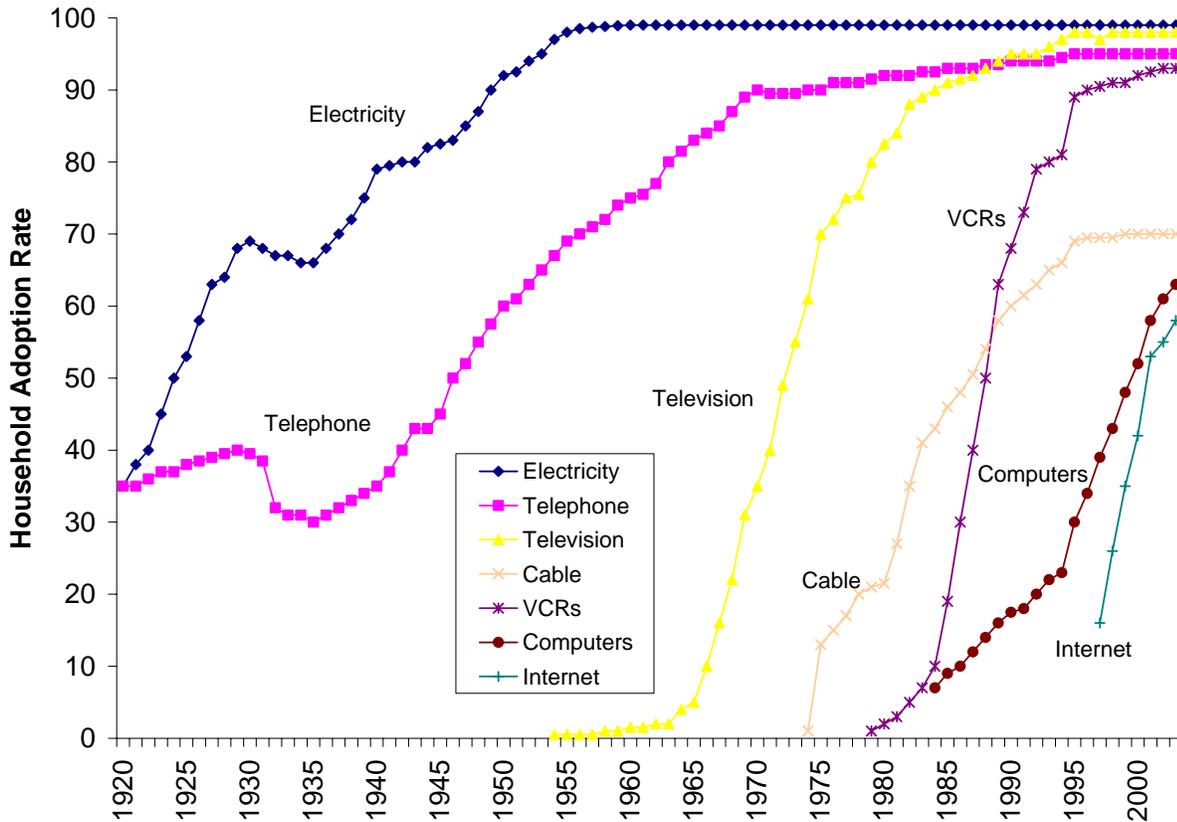
Figure 12. Temporal and Geographic Resistance to Hybrid Corn Adoption



Source: Griliches, 1957

As noted previously, major innovations like the telephone and television have shown temporal resistance to diffusion. The S-curves for a number of various technologies is displayed in Figure 13. These innovations were initially introduced in primarily urban areas, and the first adopters of these innovations typically displayed the "early adopter" characteristics discussed by Rogers.

Figure 13. S-Curves for Various Technologies



Sources: Data on TV and electricity: U.S. Department of Commerce, *Historical Statistics of the United States: Colonial Times to 1970*, p. 783. Cable data: AC Nielsen Co, as reported by National Cable Television Association, www.ncta.com. VCRs, TV data: Consumer Electronics Association, E-Brain, www.ebrain.org. Computer, Internet data: U.S. Department of Commerce www.ntia.doc.gov/ntiahome/, 1997 – 2003 Current Population Surveys

As Figure 13 demonstrates, the S-curves for the earlier technologies (electricity, telephone) have a much lower slope than those for more recent technologies (VCRs, computers). In particular, the Internet has shown an incredible rate of diffusion, reaching more than 50 percent of U.S. households in less than 10 years.²⁴ During this diffusion process, the access rates in rural and urban regions have increased at a similar pace, but rural households have consistently lagged behind urban areas. Given the numerous sources and instant availability of information widely available to today's households such as radio, television, and newspaper, it may be unlikely that restrictions in the flow of

²⁴ Based on the assumption that residential Internet access was not available until 1994 (when privately operated backbone became available), and CPS data from 2003 indicating that more than 50% of U.S. households had Internet access.

information account for a sizeable portion of the rural – urban digital divide. Hence, adopter characteristics may be more important than core-to-periphery information flow in the temporal diffusion of Internet access. However, no empirical studies have effectively tested this hypothesis.

This chapter has developed a conceptual framework dealing with residential Internet access centering around four major factors. These factors have inherently different policy implications for closing the rural - urban digital divide. It is essential to note that the usefulness of each policy will be based on the contribution of the relevant factor to the rural – urban digital divide. Empirical methods to determine the contribution of each factor to the current rural – urban digital divide are addressed in Chapter 3.

Chapter 3: Empirical Framework

The first section of Chapter 3 describes the data to be used. Section 3.2 specifies an empirical model that addresses the four factors introduced in the previous chapter (household characteristics, DCT infrastructure, social networks, and temporal resistance). Two distinct models are then introduced, with section 3.3 focusing on a model of general Internet access, and section 3.4 specifying a more complex model differentiating between dial-up and high-speed access. Section 3.5 then outlines the methodology for assessing the importance of factors underlying the divide for both of these models. This framework allows for inference of the roles that the four factors play in the observed rural – urban digital divide in both general access and in the emerging divide in high-speed access.

3.1 - Data

The empirical data for this dissertation has two distinct components. The household characteristics hypothesized to play a role in the Internet adoption decision are obtained from survey data, while information on DCT infrastructure is aggregated from data on individual cable and DSL providers. This section discusses these data sources in more detail.

The household survey data used in this dissertation comes from Current Population Survey Supplemental Questionnaires on Household Computer and Internet Use. These nationally representative surveys of roughly 50,000 households collect basic demographic and employment information on household members on a monthly basis, while the supplement focuses specifically on residential computer and Internet use in 1997, 1998, 2000, 2001, and 2003.²⁵ The major drawback with this dataset is that the lowest level of geographic information available on a household is rural or urban status within a state.

Residential Internet access is defined by a positive response to the question, "Does anyone in the household connect to the Internet from home?" Unfortunately, questions regarding high-speed access were not included at the time of the 1997 and 1998 surveys, but numbers of households with high-speed access in these years were likely

²⁵ Note that this data has already been used in the descriptive analysis provided in Chapters 1 and 2.

negligible. Hence, only data from 2000, 2001, and 2003 will be used to analyze the high-speed digital divide. Table 8 below provides information on the number of rural and urban households in each year of the data, along with the number of households that had Internet (any type) and high-speed access in those years. Table 8 indicates that a large number of households exist in both the urban and rural samples for all included years.

Table 8. CPS Household Summary Data

	Urban			Rural		
	Total	Internet	Highspeed	Total	Internet	Highspeed
1997	27,676	4,868	-	9,094	907	-
1998	27,220	8,338	-	8,954	1,782	-
2000	26,413	12,368	1,456	8,601	3,045	218
2001	31,006	17,722	3,653	10,605	4,826	520
2003	29,841	18,456	7,524	10,331	5,333	1,300

Sources: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, and 2003.

A major contribution of this dissertation is the construction of a dataset detailing the availability of digital technology infrastructure at the level of rural-urban status within a state. This data is constructed from two separate sources dealing with the dominant types of residential high-speed Internet access (cable and DSL). Information on the availability of cable Internet access is documented in the *Television and Cable Factbook* produced by Warren Publishing each year. This publicly available data source lists all cable systems in the U.S. (over 9,300 in 2003), and includes information on the counties served by each system as well as whether or not the system offers cable Internet access. Thus, the *Television and Cable Factbooks* from 2000, 2001, and 2003 provide comprehensive information on the level of cable Internet access available in the U.S. during those years.²⁶ Similarly, the Tariff #4 dataset available from the National Exchange Carriers Association (NECA) provides information on the DSL capability of every central office switch in the U.S (approximately 38,000 in 2003), along with the city or town served by each central office switch.²⁷ Historical versions of this dataset are

²⁶ 2000 was the first year where information on the availability of cable Internet access was included in the *Television and Cable Factbook*.

²⁷ Tariff #4 data from 2003 indicates that most (59%) of these 38,000 central office switches are located by themselves, while the other 41% are co-located in “wire centers,” which house two or more switches. These switches may belong to many different telecommunications providers.

available for purchase, allowing the user to recreate the level of DSL availability in 2000, 2001, and 2003. A digital technology infrastructure index is then created for every county (or city) by weighting the capability of various technologies in that county (or city) by the population level.²⁸ Unfortunately, in order to mesh this index with household data from the CPS, it must be aggregated to rural / urban status within a state.²⁹ Hence, the ultimate output from these data sources are the percentage of the population living in rural and urban areas of each state that have DCT infrastructure (either DSL or cable) available to them. Table 9 summarizes this information for the nine regions of the U.S. in 2000, 2001, and 2003.³⁰ Substantial differences in levels of infrastructure exist between rural and urban areas, but also between different regions of the U.S. Further differentiation exists among individual states (see Appendix B).

Table 9. Percent of U.S. Rural / Urban Population Living in Counties with DCT Infrastructure (Broken out by Region)

Region	DSL Infrastructure						Cable Infrastructure					
	2000		2001		2003		2000		2001		2003	
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban
Pacific	11.74	36.74	23.32	57.16	34.81	59.89	11.45	36.06	14.68	41.73	70.03	65.61
Mountain	0.94	7.48	5.00	7.90	13.65	13.59	1.35	14.39	1.21	15.14	27.09	73.86
West North Central	1.24	9.63	3.22	5.56	26.53	17.01	4.18	31.04	5.14	32.77	42.67	78.22
West South Central	4.86	33.70	8.99	51.12	45.05	66.93	1.64	13.98	1.97	16.28	45.56	82.47
East North Central	5.07	11.65	6.17	34.34	23.47	40.16	8.83	20.36	10.11	25.50	59.57	71.93
East South Central	3.62	32.06	5.57	47.13	64.20	70.67	2.80	28.33	4.61	30.53	24.63	68.57
South Atlantic	1.58	15.11	4.97	28.82	36.64	51.94	1.85	26.60	2.44	30.31	40.12	73.71
Middle Atlantic	0.47	1.77	1.04	3.11	6.34	15.26	4.89	25.20	4.89	28.21	22.34	75.87
New England	0.00	0.00	0.36	2.29	10.10	19.18	7.85	30.07	7.74	30.07	55.53	78.83
Totals	3.43	21.61	6.39	32.05	29.55	42.39	4.66	25.08	5.47	27.68	44.10	75.75

Sources: *Television and Cable Factbook*, NECA Tariff #4 Data

Several patterns are noticeable in Table 9. Looking first at DSL, the capacity has increased remarkably in the South - particularly in the rural areas. For all rural areas, DSL capacity increased from 6.39 percent in 2001 to 29.55 percent in 2003 (an impressive 362 percent increase). However, in the southern regions (South Atlantic, West South Central, and East South Central), rural households saw increases of 401

²⁸ Data on city / county population levels is taken from the 2000 census, provided by the Bureau of Labor Statistics.

²⁹ This aggregation potentially hides lower-level geographic differences in DCT infrastructure that may play a role in the rural – urban divides. This issue is further discussed in section 5.3.

³⁰ A map of these regions is shown in Figure 10.

percent, 637 percent, and 1,053 percent, respectively. This increase is consistent with BellSouth's aggressive deployment of DSL starting in late 2001 (Pinkham Group, 2002). Other rural areas, such as those in the Mountain, Middle Atlantic, and New England regions, continued to have relatively low DSL capacity through 2003. Additionally, the diffusion of DSL capacity has slowed for rural households in the Pacific and East North Central regions. While these regions had relatively high capacity in 2000, their growth rates did not keep pace with those for rural regions in the rest of the country. It is also interesting to note that the rural – urban discrepancy in capacity has shown different trends in various regions of the country. For instance, the Pacific region has consistently seen DSL capacity in rural areas lag behind that in urban areas by approximately 25 percentage points from 2000 through 2003. On the other hand, the rural population in the West North Central region has gone from lagging their urban counterparts in terms of DSL infrastructure capacity by 8 percentage points in 2000 to being 9 percentage points *higher* in 2003!

Cable Internet diffusion has also shown different trends in various regions of the country. Most of the diffusion occurred between 2001 and 2003, with capacity increasing by 39 percentage points in rural areas and 48 percentage points in urban areas. For all regions, rural areas experienced increases in capacity during this period, but none quite as dramatic as the West South Central region. Cable Internet capacity stood at less than 2 percentage points for the rural population in this region in 2001, and then skyrocketed to 46 percentage points in 2003 – a 2,213 percent increase! Similarly, in urban areas of the Mountain region cable Internet capacity increased from 15 percentage points to 74 percentage points over this period. The diffusion was less striking in rural areas of the East South Central region, with capacity increasing from 5 percentage points to 25 percentage points. In general, the rural – urban gap in cable Internet capacity has increased between 2000 and 2003 (from 20 percentage points to 32 percentage points), but significant variance exists within the country. The biggest change is in the Pacific region, where rural areas actually have *higher* rates of cable Internet capacity in 2003 than urban areas (70 percentage points to 66 percentage points). This is drastically different from 2000, when rural rates were 25 percentage points below urban rates. The East North Central region gap has been relatively consistent over this period, with rural

rates lagging urban rates by approximately 12 percentage points each year. Alternatively, the diffusion of cable Internet capacity in the Mountain region has occurred mainly in urban areas, as the urban – rural discrepancy increased from 13 percentage points in 2000 to 47 percentage points in 2003.

Other sources of data on DCT infrastructure also exist, such as FCC form 477. This form requires all high-speed providers with more than 250 subscribers to file information regarding the number of lines serviced and the zip codes where service is provided. Grubestic and Murray (2004) use form 477 data to explore changes in high-speed access (particularly, *competition* among high-speed providers) over the period from December 1999 to June 2001.³¹ Their analysis supports the current findings, as major metropolitan areas experienced an increasing amount of high-speed competition over this period, while rural areas lagged behind. Form 477, however, does not differentiate between various types of high-speed service (i.e. cable vs. DSL) and may fail to account for the presence of smaller high-speed providers that are located in rural areas. In contrast, because they include all existing central offices or cable systems, the *Television and Cable Factbook* and Tariff #4 datasets represent the most comprehensive source of information on high-speed providers currently available.³² Thus, while some market analysis suggests that at least 200 high-speed lines are necessary for a profitable investment (Paradyne, 2000), smaller high-speed providers do exist and are accounted for in these datasets.³³ However, it is important to note that one drawback of the data used in this dissertation is that its aggregation overstates the diffusion of high-speed providers, since not all rural areas included in a county will have access to the technology (due to the limitations such as the 3-mile radius for DSL discussed in the section 2.2).

The next four sections of this chapter discuss how the available data will be used to empirically model the rural - urban digital divide. Section 3.2 looks at the specification

³¹ Grubestic and Murray (2004) define "competition" based on the number of high-speed providers operating in a zip code.

³² The FCC is considering modifying form 477, in part due to concerns that these smaller providers are not accounted for. See Federal Register Vol 69 No. 103, May 27, 2004.

³³ In 2001, five out of the six cable Internet providers in Montana had less than 200 subscribers (Television and Cable Factbook, 2001).

of a model that incorporates the four factors associated with the divide. Section 3.3 discusses the choice among available binomial statistical models for analyzing the divide in general access. Section 3.4 specifies a model of the decision between dial-up, high-speed, and no access. Section 3.5 then discusses the methods available to quantify the roles played by the four factors in the evolution of the digital divide.

3.2 - Empirical Specification

The empirical model needs to account for the four factors associated with the rural - urban digital divide discussed in Chapter 2. In the remainder of this sub-section, alternative specifications appropriate for testing the contributions of the household characteristics, technology infrastructure, social networks, and temporal resistance to the emerging pattern of Internet adoption are presented. The hypothesized behavior of parameter estimates, including the sign and time trends, is also presented for each alternative. This information is summarized in Table 10.

Household characteristics

Similar to the theoretical model of utility discussed in section 2.3, the basic statistical model for the influence of household characteristics on Internet adoption is specified as

$$\begin{aligned}
 y_i^* &= U_i = X_i B + Z_i \delta + \varepsilon_i \\
 y_i &= 1 \quad \text{if} \quad y_i^* \geq 0 \\
 y_i &= 0 \quad \text{if} \quad y_i^* < 0
 \end{aligned} \tag{3}$$

where y_i^* is a latent measure of the benefits from residential Internet access relative to the costs of household i , y_i is the actual observation of household Internet access, X_i is a vector of household income and education levels, Z_i is a vector of other household characteristics, B is the parameter vector for household income and education levels, δ is the parameter vector for other household characteristics, and ε_i is the statistical model's error term.

Model hypotheses: Education and income levels are both expected to be positively related to the household propensity to access the Internet, $B > 0$. Other household

characteristics are also, as a group, expected to influence Internet access, $\delta \neq 0$. Further, neither set of parameters is expected to shift significantly over time, so no time subscripts are incorporated. Rural - urban differences in rates of Internet access are due to differences in variable levels, not parameter estimates (hence, it is expected that $B_R = B_U$ and $\delta_R = \delta_U$).

Early adopters

The early adopter framework introduces a temporal dimension into the statistical model of household Internet adoption. Specifically, income and education parameters are expected to vary over time. Thus, the modified statistical model becomes:

$$y_{it}^* = X_{it}B_t + Z_{it}\delta + \varepsilon_{it} . \quad (4)$$

Model hypotheses: The magnitude of the influence of higher income and education levels on the propensity of households to access the Internet is expected to decrease over time, $B_t < B_{t-1}$. As income and education levels are higher in urban areas, declines in parameter estimates over time are associated with a decline in the predicted gap in rural - urban residential Internet access.

Core to periphery diffusion

Under this framework information on the benefits and costs of home Internet access flows slowly from urban to rural areas. With this flow of information, the differential negative propensity for rural areas to access the Internet, *ceteris paribus*, declines over time. The statistical model now includes an indicator variable R_{it} for household i 's residence in a rural area in period t :

$$y_{it}^* = X_{it}B_t + Z_{it}\delta + R_{it}\gamma_t + \varepsilon_{it} \quad (5)$$

Model hypotheses: The parameter for the differential propensity for rural households to access the Internet, γ_t , is expected to be initially negative, but to decline in absolute value over time.

Table 10. Key Components of Alternative Residential Internet Adoption Models

Model Type	Key Variables	Variable Group Label	Parameter Label	Parameters Stable	Alternative Hypotheses
Household characteristics	Income Education	X	B	Yes	$B > 0$
	Race/ Ethnicity Age of head # of school age children Marital status of head Gender of head Member uses Internet at work	Z	δ		$\delta \neq 0$
Early adopters	Income Education	X_t	B_t	No, positive associations decline over time	$B_t < B_{t-1}$
Core to periphery diffusion	Rural (urban base)	R_t	γ_t	No, negative association declines over time	$\gamma_t < \gamma_{t-1}$
Technology infrastructure	Cable modem access DSL access	$D1_t$ $D2_t$	τ_1 τ_2	No, τ_1 and τ_2 positive association increases over time	$\tau_1, \tau_2 > 0$ $\tau_{1t} > \tau_{1t-1}$ $\tau_{2t} > \tau_{2t-1}$
Social Networks	Regional rate of residential Internet use	N_t	π_1	Yes	$\pi_1, \pi_2 > 0$
	Critical Mass	N_t^2	π_2		

Technology infrastructure differences

As discussed in the section detailing the data to be used (Section 3.2), information on DCT infrastructure is used to construct indices detailing the percentage of rural and urban residents in each state that have high-speed cable infrastructure ($D1_t$) and DSL infrastructure ($D2_t$) available to them in 2000, 2001, and 2003. These aggregate indices are included in the statistical model under the following specification.

$$y_{it}^* = X_{it}B_t + Z_{it}\delta + R_{it}\gamma_t + D1_{it}\tau_{1t} + D2_{it}\tau_{2t} + \varepsilon_{it} \quad (6)$$

Model hypotheses: The parameter estimates associated with cable and DSL infrastructure are τ_1 and τ_2 , respectively. While neither τ_1 and τ_2 are expected to be significant in the model for general access, both are expected to be positive in the model for high-speed access. Further, the influence of both τ_1 and τ_2 are expected to strengthen with the increased presence of audio and video applications requiring high-speed access, $\tau_{1t} > \tau_{1t-1}$ and $\tau_{2t} > \tau_{2t-1}$.

Social Networks

The influence of network externalities is tested through the inclusion of a measure of the regional rate of Internet access, N_t , and associated estimated parameter, π , in the statistical model. Additionally, to capture the 'critical mass' concept associated with social networks, a non-linear term (squared or logarithmic) for regional rates of access will be included. This leads to the full model specification:

$$y_{it}^* = X_{it}B_t + Z_{it}\delta + R_{it}\gamma_t + D1_{it}\tau_{1t} + D2_{it}\tau_{2t} + N_{it}\pi_1 + N_{it}^2\pi_2 + \varepsilon_{it} \quad (7)$$

Model hypotheses: $\pi_1, \pi_2 > 0$ and stable over time. The hypothesis for π_1 is that individual households in rural areas have lower Internet access propensities due to lower regional rates of Internet access. For π_2 , the hypothesis is that regional rates of Internet access have a larger impact after critical mass is reached (which is estimated to occur at the beginning of the dataset – 1997 –98), hence the inclusion of a squared term will have an even larger effect on areas with higher regional rates of access. Additionally, if both

π_1 and π_2 are positive, the total effect resulting from the inclusion of network externalities will take a concave up shape, where the resulting parameter values increase at an increasing rate. State-level statistics on rural / urban general and high-speed access (shown in Appendix C) will be used as the local adoption rate for household i in time period t , N_{it} .

The hypothesized parameter restrictions associated with each of the five sub-specifications will be examined, along with the implications of variable and parameter differences for the magnitude of the digital divide. The magnitude of the contribution of variables associated with each model to the observed rural - urban digital divide will then be examined by identifying the share of the divide associated with rural - urban differences in specific variables and the share associated with rural - urban differences in model parameter estimates. This decomposition is further discussed in section 3.5.

3.3 - Model Distribution Assumption

As discussed in the section on utility theory (section 2.6), the adoption of residential Internet access is a binary choice – the household either has Internet access (1) or it does not (0).³⁴ Due to the discrete nature of this decision, adoption will be estimated using a binomial variable statistical model (e.g. probit, logit, or linear probability model). The final choice of model is made by weighting the trade-offs between the specification restrictions associated with each model and the ease of parameter interpretation. For instance, the parameters of the linear probability model are easy to interpret, but the specification imposes the rather undesirable restriction that a change in an independent variable always changes the probability of the dependent variable by the same number of percentage points. As a result, the linear probability model allows for predictions outside the feasible [0,1] probability range. Logit and probit models restrict outcomes to the unit interval, but parameter values have non-linear interpretations that make decomposition of rural – urban differences difficult. This section provides a comparison

³⁴ The modeling choice discussed here is for the general divide in residential Internet access, with no distinction between dial-up and high-speed. The following section (3.4) discusses a model distinguishing between these types of Internet access.

of regression results and marginal effects for the logit, probit, and linear probability models from 2003 CPS data based on the independent variables in the empirical specification suggested in the previous section.³⁵ The results indicate that the logit model is the preferred specification.

A summary of these variable categories and names used in the various Internet access regressions is shown in Table 11. The results of the three binary choice models are then shown in Table 12. Interestingly, the results of all three models (linear probability, logit, and probit) are strikingly similar. With few exceptions, all variables have the same sign and significance level for the three different types of models.³⁶ Additionally, the predictive power of all three types of models is remarkably similar.

Table 11. Variable Summary (Preliminary Regressions)

Characteristic	Variable Name	Characteristic	Variable Name
Income Level		Race / Ethnicity	
Under \$5,000		White	
\$5,000 - \$7,499	faminc1	Black	black
\$7,500 - \$9,999	faminc2	Other	othrace
\$10,000 - \$12,499	faminc3	Hispanic	hisp
\$12,500 - \$14,999	faminc4	Household Composition	
\$15,000 - \$19,999	faminc5	Married	married
\$20,000 - \$24,999	faminc6	Headed by male	sex
\$25,000 - \$29,999	faminc7	Age of Head	peage
\$30,000 - \$34,999	faminc8	Age of Head^2	age2
\$35,000 - \$39,999	faminc9	1 child in house	chld1
\$40,000 - \$49,999	faminc10	2 children in house	chld2
\$50,000 - \$59,999	faminc11	3 children in house	chld3
\$60,000 - \$74,999	faminc12	4 children in house	chld4
Over \$75,000	faminc13	5+ children in house	chld5
Education		Regional Density	regdensity
No High School		Regional Density^2	regdensity2
High School	hs	Internet access at work	netatwork
Some College	scoll	Cable Internet Access	cableaccess
College Degree	coll	DSL Internet Access	dslaccess
Higher than Bachelor's	collplus	Retired	retired

Note: The characteristics without variable names are the "default" variables

³⁵ Only the most recent year of data was used (2003) in order to identify the preferred specification. The analysis underlying this task (comparison of marginal effects, percentage of predictions outside the unit interval) should not substantially vary between years.

³⁶ The standard errors for the LPM are White's heteroskedastic consistent errors due to the presence of heteroskedasticity in this model.

Table 12. LPM, Logit, and Probit Estimates for Internet Access in 2003

Dependant Variable: Internetuse			
Independent Variable	LPM	Logit	Probit
hs	0.1055 ***	0.6178 ***	0.3673 ***
scoll	0.2348 ***	1.2529 ***	0.7468 ***
coll	0.2770 ***	1.5514 ***	0.9167 ***
collplus	0.2923 ***	1.7096 ***	0.9920 ***
faminc1	-0.0159	-0.1566	-0.0881
faminc2	-0.0267	-0.2471 **	-0.1338 **
faminc3	-0.0054	-0.0627	-0.0337
faminc4	-0.0028	-0.0326	-0.0099
faminc5	0.0161	0.0902	0.0618
faminc6	0.0525 ***	0.2604 ***	0.1690 ***
faminc7	0.0724 ***	0.3231 ***	0.2091 ***
faminc8	0.1168 ***	0.5120 ***	0.3261 ***
faminc9	0.1706 ***	0.7238 ***	0.4554 ***
faminc10	0.2242 ***	0.9512 ***	0.5929 ***
faminc11	0.2545 ***	1.1038 ***	0.6853 ***
faminc12	0.2962 ***	1.3734 ***	0.8373 ***
faminc13	0.3171 ***	1.6686 ***	0.9923 ***
nm	0.0142	0.0757	0.0452
netatwork	0.0833 ***	0.5323 ***	0.3030 ***
black	-0.1185 ***	-0.6717 ***	-0.3942 ***
othrace	-0.0321 ***	-0.1976 ***	-0.1199 ***
hisp	-0.1212 ***	-0.6651 ***	-0.3941 ***
peage	0.0103 ***	0.0606 ***	0.0359 ***
age2	-0.0001 ***	-0.0008 ***	-0.0005 ***
sex	-0.0017	0.0114	0.0059
married	0.0933 ***	0.5449 ***	0.3206 ***
chld1	0.0358 ***	0.2517 ***	0.1504 ***
chld2	0.0377 ***	0.2994 ***	0.1739 ***
chld3	0.0234 **	0.1910 **	0.1137 ***
chld4	0.0173	0.1479	0.0888
chld5	0.0268	0.1944	0.1150
regdensity	0.8581 **	4.7270 *	2.8792 *
regdensity2	-0.4124	-2.0801	-1.3220
cableaccess	0.0003	-0.0033	0.0018
dslaccess	0.0143	0.0751	0.0425
retired	0.0230 **	0.1876 ***	0.1041 ***
constant	-0.3294 ***	-4.6344 ***	-2.7860 ***
Internetuse = 1			
Number of Observations	23,789	23,789	23,789
Percent Correctly Predicted	82.66	82.46	82.36
Internetuse = 0			
Number of Observations	16,383	16,383	16,383
Percent Correctly Predicted	66.86	67.30	67.32
Log-Likelihood Value	-	-19791.3	-19792.3
Pseudo R ²	0.325	0.273	0.273

***, **, and * indicate statistical significance at the p = 0.01, 0.05, and 0.10 levels, respectively

One area of interest to this analysis is the share of predictions that fall outside the unit interval for the Linear Probability Model. Out of 40,172 total observations, the LPM predicted that 1,947 had a probability of Internet access higher than one, and that 714 had

a probability of Internet access lower than zero. Hence, about 6.6% of the total predicted values fell outside the unit interval. Wooldridge (2002) suggests that if the LPM gives good estimates of the partial effects on the response probability near the center of the distribution, the advantage provided by an easily comprehensible coefficient may outweigh the potential for predictions that are outside the unit interval. To ascertain whether or not the LPM gives "good" estimates at the center of the distribution, the coefficients of the LPM are compared to the partial effects of the logit and probit models evaluated at the variable means.³⁷ These results are displayed in Table 13.

Table 13. Comparison of Partial Effects for LPM, Logit, and Probit

Independent Variable	LPM	Logit	Probit
hs	0.1055	0.1395	0.1366
scoll	0.2348	0.2659	0.2646
coll	0.2770	0.3034	0.3048
collplus	0.2923	0.3072	0.3095
faminc1	-0.0159	-0.0375	-0.0341
faminc2	-0.0267	-0.0596	-0.0520
faminc3	-0.0054	-0.0149	-0.0130
faminc4	-0.0028	-0.0077	-0.0038
faminc5	0.0161	0.0210	0.0235
faminc6	0.0525	0.0594	0.0632
faminc7	0.0724	0.0731	0.0777
faminc8	0.1168	0.1126	0.1185
faminc9	0.1706	0.1533	0.1607
faminc10	0.2242	0.1955	0.2041
faminc11	0.2545	0.2199	0.2300
faminc12	0.2962	0.2608	0.2706
faminc13	0.3171	0.3277	0.3302
nm	0.0142	0.0177	0.0173
netatwork	0.0833	0.1209	0.1133
black	-0.1185	-0.1642	-0.1549
othrace	-0.0321	-0.0474	-0.0466
hispanic	-0.1212	-0.1627	-0.1550
peage	0.0103	0.0143	0.0138
age2	-0.0001	-0.0002	-0.0002
sex	-0.0017	0.0027	0.0023
married	0.0933	0.1284	0.1229
chld1	0.0358	0.0578	0.0566
chld2	0.0377	0.0683	0.0653
chld3	0.0234	0.0439	0.0429
chld4	0.0173	0.0342	0.0336
chld5	0.0268	0.0446	0.0433
regdensity	0.8581	1.1130	1.1037
regdensity2	-0.4124	-0.4898	-0.5068
cableaccess	0.0003	-0.0008	0.0007
dslaccess	0.0143	0.0177	0.0163
retired	0.0230	0.0435	0.0395
constant	-0.3294	N/A	N/A

³⁷ See Appendix D for a discussion of how partial effects are derived for the logit and probit models, and how these values differ for continuous vs. discrete variables.

While the majority of significant variables (such as income, race, and age) have similar magnitudes for all three models, it is worth noting that the LPM provides lower marginal effect estimates of education when compared to the logit and probit models. For all levels of education, the increase in the probability of Internet access under the LPM is significantly lower than the results obtained for both the logit and probit models. The partial effect associated with the regional density term (the percent of households in a region with Internet access) is significantly lower under the LPM as well. Hence, the choice of distribution assumed may matter for the estimation of marginal effects. It should also be noted that the use of a large number of dummy variables is especially problematic for the LPM, based on its lack of precision as variables tend towards the tails of the distribution. In fact, one of the primary advantages of the logit / probit over the LPM is the decline in marginal effects at extreme values of the explanatory variables. Because most of the explanatory variables are dummies and hence evaluated *only* at their "extreme values," the LPM does not seem like the logical choice to evaluate these variables.³⁸ Given this problem, along with the fact that the LPM predicts that over 6 percent of the observations fall outside of the unit interval, evidence suggests that a logit or probit model might be better suited for this analysis. Most analysis on discrete outcomes that utilize logit and probit models suggest very little difference between the two (Capps and Kramer, 1985; Layton and Katsuura 2001). Therefore, as the logit equation has a closed form solution, it is employed as the preferred specification.

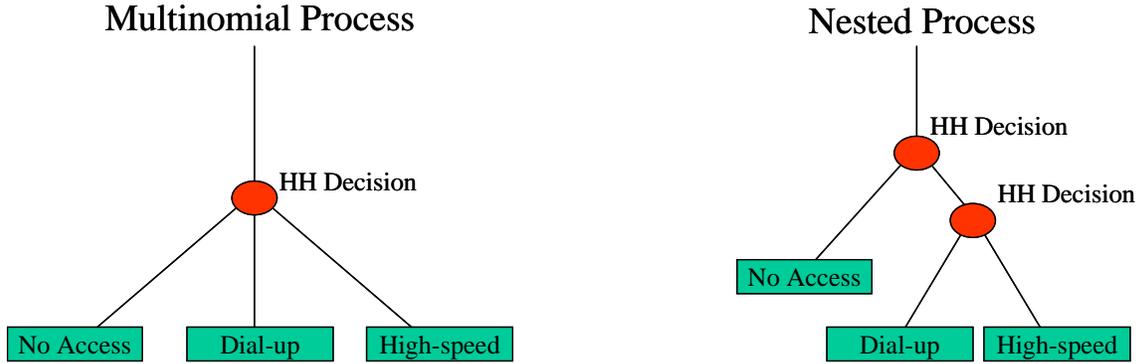
3.4 - Differentiating Dial-up and High-speed Access in the Adoption Decision

The previous section indicated that the access – no access decision will be modeled via a binary choice framework. A similar, but more complicated, model will be used to differentiate between dial-up and high-speed access (while still allowing households the decision to have "no access"). This model will be of particular interest due to the increasingly important role played by high-speed access documented in Chapters 1 and 2. Given the preference for the logit model in the framework for general access, two separate extensions to this model – the multinomial logit and nested logit –

³⁸ For binomial variables, 0 and 1 are essentially the "extreme values" of the distribution.

are discussed below. The difference between these two approaches is depicted in Figure 14.

Figure 14. Multinomial and Nested Decision Processes



Following the decision process depicted above, the multinomial logit model has three possible outcomes, which are indexed by the variable $j \in J = \{0, 1, 2\}$: no Internet access ($j = 0$), dial-up access ($j = 1$), and high-speed access ($j = 2$). Assume that the utility household i derives from alternative j (denoted U_{ij}) can be written as:

$$U_{ij} = V_{ij} + \varepsilon_{ij} \quad (8)$$

where V_{ij} can be modeled and ε_{ij} is an error term. The non-stochastic portion of the utility (V_{ij}) is dependent on both characteristics of the household (X_i) and characteristics of the alternative (Z_{ij}).³⁹ Hence we can re-write V_{ij} as:

$$V_{ij} = \beta_j' X_i + \gamma' Z_{ij} \quad (9)$$

where β_j' and γ' are the parameter vectors associated with X_i and Z_{ij} , respectively.

Note that the parameter vector associated with characteristics of the household (β_j') is specific to the alternative. Given the form of the logistic distribution, the probability that household i will choose alternative j is:⁴⁰

³⁹ X_i is a vector of household characteristics, while Z_j is a compressed vector of characteristics that vary by alternative, such as measures of telecommunications infrastructure and network externalities. Given that we are only discussing a single point in time, the temporal factors are not included.

⁴⁰ By using the logistic distribution we are implicitly assuming that the unknown terms are distributed according to a special form of the generalized extreme value (GEV) distribution (McFadden, 1981).

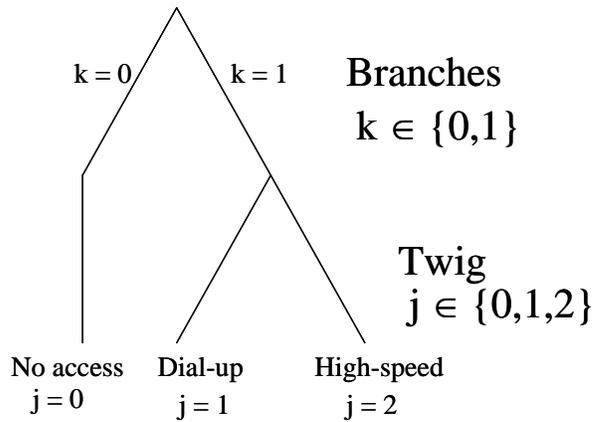
$$P_{ij} = \frac{\exp(V_{ij})}{\sum_{k \in J} \exp(V_{ik})} \quad \text{for all } j \in J \quad (10)$$

The distinctive characteristic of the multinomial logit model is that it assumes the independence of irrelevant alternatives (IIA). Simply stated, IIA implies that if only two choices existed (say, no access or dial-up access), then the addition of a third choice (high-speed access) would not change the ratios of probabilities of the first two choices. The nested logit model, however, allows the IIA restriction to be relaxed. To formally specify the nested logit model, some additional notation must be introduced. First, the household index i in equation (9) will be suppressed for simplicity, resulting in an expression for the observed utility of alternative j :

$$V_j = \beta_j' X + \gamma' Z_j \quad (11)$$

Second, the following "tree" structure is presented for this decision-making process:⁴¹

Figure 15. Nested Logit Tree Structure



Formally, the probability of a household selecting branch k and twig j is:

$$\Pr [\text{branch } k, \text{ twig } j] = P_{kj} = (P_{jk})(P_k). \quad (12)$$

The conditional probability is defined as

⁴¹ Although the tree structure suggests sequential decision making, this is not necessarily the case. The nested logit structure can accommodate a simultaneous decision regarding access and type. Knapp, White, and Clark (2001) indicate that if the decision making process is not known a priori, a sequential modeling structure should not be adopted.

$$P_{j|k} = \frac{\exp\left(\frac{1}{\tau_k} V_j\right)}{\sum_{l \in J^k} \exp(V_l)}, \quad (13)$$

where τ_k represents the degree of similarity between the alternatives in branch k . The marginal probability of selecting branch k is equal to

$$P_k = \frac{\exp(\tau_k IV_k)}{\sum_{k \in K} \exp(\tau_k IV_k)}. \quad (14)$$

For the k th branch, $IV_k = \ln \sum_{j \in J^k} \exp\left(\frac{1}{\tau_k} V_j\right)$, (15)

where IV stands for inclusive value and together with its parameter τ_k represents the feedback between the upper and lower levels of the tree. Inserting equations (13) and (14) into (12), the probability of selecting branch k and twig j is:

$$P_{kj} = \left(\frac{\exp\left(\frac{1}{\tau_k} V_j\right)}{\sum_{l \in J^k} \exp(V_l)} \right) \left(\frac{\exp(\tau_k IV_k)}{\sum_{k \in K} \exp(\tau_k IV_k)} \right) \quad (16)$$

For the degenerate branch containing no access, there is only one element $j \in J^k$. In this case, $IV_k = \ln(\exp(V_j)) = V_j$. As noted above, τ_k represents the degree of similarity between the alternatives in one nest. Hence, for the degenerate branch ($k = 0$), $\tau_k = 1$ because there is only one alternative in the nest. Note that the special case of $\tau_k = 1$ for all k collapses to the multinomial logit specification. Hence, allowing τ_k to vary between branches relaxes the IIA restriction associated with the multinomial model. In the above case, the dial-up and high-speed options are considered to be similar to each other.

The above specification has been shown to be consistent with random utility maximization if all τ_k lie in the unit interval (McFadden, 1981; Hensher and Greene, 2000).⁴² Furthermore, this condition can be relaxed for *local* consistency with random utility maximization (Borsch-Supan, 1990). Testing the IIA restriction in the

⁴² It is worth noting that there is a discrepancy between this specification and the type implemented by various software packages such as STATA. See Heiss (2002) for a full discussion.

multinomial model and comparing the predictive ability between the two models provides two different ways for determining which specification is appropriate (Hausman and McFadden, 1984). Inherent similarities exist between dial-up and high-speed Internet access – they both provide some type of Internet access, and can both be used to accomplish many of the same tasks. Hence, from a consumer (household) standpoint, it is unlikely that these two choices are viewed as completely independent of each other. Rather, households probably view high-speed access as an "upgraded" version of dial-up access, with a subset of those preferring to have some type of access opting for the "upgraded" version. In other words, the introduction of high-speed access likely drew subscribers more heavily from those with dial-up access than those with no access. Therefore, the nested logit model is hypothesized to be the appropriate specification.

3.5 - Assessing the Importance of the Four Factors

Thus far two distinct models have been identified: a model for general access and a model distinguishing between dial-up and high-speed access. The preferred specifications for these models are the logit and nested logit, respectively. This section discusses how these models will be used to assess the importance of the four previously discussed factors affecting Internet access.

General Access: Logit Model

Table 10 explicitly demonstrated how each potential factor will be accounted for in the models for both general and high-speed access. To determine the contribution of each factor to the rural – urban digital divide, a decomposition technique will be employed. The decomposition technique used will be a non-linear version of the Oaxaca-Blinder (Oaxaca, 1973; Blinder, 1973) decomposition, due to the non-linear nature of the logit model.⁴³ However, this decomposition method is valid only for a single point in time. As such, only the effects of household characteristics, infrastructure differences, and network externalities can be evaluated in any given year. Including temporal

⁴³ This non-linear version of the Oaxaca-Blinder decomposition is similar to the procedure used in Mills and Whitacre (2003) and Fairlie (2003).

resistance to adoption requires introducing a time dimension to the decomposition, and is addressed in a manner similar to the technique for a single year.

The standard (linear) Oaxaca-Blinder decomposition of the general rural – urban digital divide in residential Internet access can be expressed as:

$$\bar{Y}^U - \bar{Y}^R = (\bar{X}^U - \bar{X}^R)\hat{\beta}^U + \bar{X}^R(\hat{\beta}^U - \hat{\beta}^R) \quad (17)$$

where \bar{Y}^G is the average value of Internet access, \bar{X}^G is a row vector for average values of the independent variables, and $\hat{\beta}^G$ is a vector of coefficient estimates for rural / urban status G . Following Fairlie (2003), the decomposition for a non-linear equation, such as $Y = F(X\hat{\beta})$, can be written as:

$$\bar{Y}^U - \bar{Y}^R = \left[\sum_{i=1}^{N^U} \frac{F(X_i^U \hat{\beta}^U)}{N^U} - \sum_{i=1}^{N^R} \frac{F(X_i^R \hat{\beta}^U)}{N^R} \right] + \left[\sum_{i=1}^{N^R} \frac{F(X_i^R \hat{\beta}^U)}{N^R} - \sum_{i=1}^{N^R} \frac{F(X_i^R \hat{\beta}^R)}{N^R} \right] \quad (18)$$

where N^G is the sample size for rural / urban status G . This equation applies various coefficient estimates to the two distinct groups of explanatory variables, X_i^U and X_i^R .

Equivalently, the decomposition can be written as:

$$\bar{Y}^U - \bar{Y}^R = \left[\sum_{i=1}^{N^U} \frac{F(X_i^U \hat{\beta}^R)}{N^U} - \sum_{i=1}^{N^R} \frac{F(X_i^R \hat{\beta}^R)}{N^R} \right] + \left[\sum_{i=1}^{N^R} \frac{F(X_i^U \hat{\beta}^U)}{N^U} - \sum_{i=1}^{N^U} \frac{F(X_i^U \hat{\beta}^R)}{N^U} \right] \quad (19)$$

This first term on the right hand side of equations (18) and (19) represents the part of the digital divide due to group differences in the distributions of the explanatory variables X . The choice of which set of parameters to use (either $\hat{\beta}^U$ in (18) or $\hat{\beta}^R$ in (19)) is the essence of the familiar "index problem" to the Oaxaca-Blinder decomposition, and can be the source of significantly different results. Some studies suggest weighting the parameters by using coefficient estimates from a pooled sample of the two groups (Neumark, 1988; Oaxaca and Ransom, 1994). This approach is valid when the "weighted average" access rates are considered exemplary of the rates that would exist in the absence of a digital divide (Oaxaca and Ransom, 1994).

In order to calculate the contributions from individual explanatory variables included in this first term, we must be able to "replace" a single rural characteristic (for example, education level) with its urban counterpart. Hence, a one-to-one mapping of the rural and urban samples is needed to establish an urban counterpart for each rural

observation. In order to create such a mapping, predicted probabilities are calculated for all observations (both rural and urban) using all independent variables. Since the sample size for urban households is larger than the sample size for rural households, a sub-sample of urban households is drawn equal in size to the rural sample. This sampling procedure will clearly affect \bar{Y}^U and X_i^U , since both are dependent on the households included in the sample. However, results from the entire urban sample can be approximated by bootstrapping a large number of random urban samples, mapping these samples to the rural sample, and averaging the results of the decomposition.

Returning now to the decomposition procedure for two individual samples, both should be ranked by predicted probability. Hence, rural households that have characteristics placing them high (low) in their distribution are matched with urban households that have characteristics placing them high (low) in their distribution. To accomplish the decomposition, let X_1, X_2 , and $X_3 \in X$ be the three distinct non-temporal categories of independent variables listed above: X_1 represents household characteristics, X_2 represents telecommunications infrastructure, and X_3 represents network externalities. Using coefficient estimates $\hat{\beta}$ from a pooled sample of both rural and urban households, the independent contribution of X_1 to the digital divide can be expressed as:⁴⁴

$$\frac{1}{N^R} \sum_{i=1}^{N^R} F(\hat{\beta}_0 + X_{1i}^U \hat{\beta}_1 + X_{2i}^U \hat{\beta}_2 + X_{3i}^U \hat{\beta}_3) - F(\hat{\beta}_0 + X_{1i}^R \hat{\beta}_1 + X_{2i}^U \hat{\beta}_2 + X_{3i}^U \hat{\beta}_3). \quad (20)$$

Similarly, the independent contribution of X_2 can be expressed as:

$$\frac{1}{N^R} \sum_{i=1}^{N^R} F(\hat{\beta}_0 + X_{1i}^U \hat{\beta}_1 + X_{2i}^U \hat{\beta}_2 + X_{3i}^U \hat{\beta}_3) - F(\hat{\beta}_0 + X_{1i}^U \hat{\beta}_1 + X_{2i}^R \hat{\beta}_2 + X_{3i}^U \hat{\beta}_3), \quad (21)$$

and a similar expression can be written for X_3 . Hence, the contribution of each group of variables to the gap equals the change in average predicted probability from replacing the rural distribution with the urban distribution for that group of variables while holding the distributions of the other groups constant.⁴⁵ As noted in Fairlie (2003), this technique is

⁴⁴ Note that since a pooled sample is used to obtain coefficient estimates, the decomposition uses weighted averages of the parameter estimates shown in equations (18) and (19).

⁴⁵ Because of the non-linear form assumed by the use of the logit model, the contributions of X_1, X_2 , and X_3 depend on values of the other variables. Hence, the order of how the variables enter equations (18) and (19) may affect their individual contributions to the rural-urban digital divide. To account for this, the order in which variables enter the analysis will be varied, and the results will be compared.

useful because the sum of the contributions from the individual groups will be equal to the total contribution from all variables in the sample.

It is important to note that equations (20) and (21) deal *only* with the first term of the decomposition shown in equations (18) and (19). The second term in equations (18) and (19) above represents the portion of the gap due to rural - urban differences in underlying parameters, and is not affected by differences in explanatory variables. Differences in the three non-temporal categories of independent variables discussed above, along with this "other" portion, make up the entire rural - urban digital divide in any given year. Thus, this framework will be useful in determining the roles played by the three non-temporal categories for any given time period.

To determine the impact of the temporal resistance to adoption, the above model must be converted to two time periods, t and $t-1$. While most analysis of inter-temporal decomposition has focused on linear functional forms (Le and Miller, 2004), the logistic form of this specification requires a different technique. Instead of focusing on differences in characteristics between rural and urban areas, the focal point of the following analysis is differences in characteristics between time periods. Hence, equation (18) must be reconstructed to incorporate a temporal dimension:

$$\bar{Y}_t - \bar{Y}_{t-1} = \left[\sum_{i=1}^{N_t} \frac{F(X_{it}^* \hat{\beta}_t)}{N_t} - \sum_{i=1}^{N_{t-1}} \frac{F(X_{it-1}^* \hat{\beta}_t)}{N_{t-1}} \right] + \left[\sum_{i=1}^{N_{t-1}} \frac{F(X_{it-1}^* \hat{\beta}_t)}{N_{t-1}} - \sum_{i=1}^{N_{t-1}} \frac{F(X_{it-1}^* \hat{\beta}_{t-1})}{N_{t-1}} \right] \quad (22)$$

In the same way that equations (18) and (19) are equivalent expressions for the rural – urban divide for a given year, the difference in access between years t and $t-1$ can also be expressed as:

$$\bar{Y}_t - \bar{Y}_{t-1} = \left[\sum_{i=1}^{N_t} \frac{F(X_{it}^* \hat{\beta}_{t-1})}{N_t} - \sum_{i=1}^{N_{t-1}} \frac{F(X_{it-1}^* \hat{\beta}_{t-1})}{N_{t-1}} \right] + \left[\sum_{i=1}^{N_{t-1}} \frac{F(X_{it}^* \hat{\beta}_t)}{N_t} - \sum_{i=1}^{N_{t-1}} \frac{F(X_{it}^* \hat{\beta}_{t-1})}{N_t} \right] \quad (23)$$

X_{it}^* is a vector of characteristics for household i at time t , and N_t is the total number of households in the sample at time t (both rural and urban). This allows X_{it}^* to take on characteristics that are hypothesized to change in importance over time, such as education

and income levels (X_{1it}^*), rural – urban status (X_{2it}^*), and levels of telecommunications infrastructure (X_{3it}^*).⁴⁶ Having performed this type of decomposition for a single year, this extension implies that the first term in equations (22) and (23) represents differences in Internet access between years t and $t-1$ due to differences in the distributions of the explanatory variables X_{it}^* . The procedure outlined above focuses on this first term, and can be used to decompose the differences between years associated with changes in variables, particularly X_{1it}^* and X_{3it}^* .⁴⁷ However, for the inter-temporal extension, the analysis is also interested in the *second* term in equations (22) and (23), which deals with the impact of parameter shifts over time. Because both sets of characteristics in this term use the same time period, the mapping between groups that was performed in the initial version of this decomposition is not necessary.⁴⁸ To set up the decomposition, individual logit models are run for the time periods t and $t-1$, resulting in the parameters $\hat{\beta}_t^*$ and $\hat{\beta}_{t-1}^*$. The independent contribution of changes in $\hat{\beta}_1^*$ between time periods t and $t-1$ can then be identified in a manner similar to equation (20) above:

$$\frac{1}{N_{t-1}} \sum_{i=1}^{N_{t-1}} F(\hat{\beta}_{0t}^* + X_{1it-1}^* \hat{\beta}_{1t}^* + X_{2it-1}^* \hat{\beta}_{2t}^* + X_{3it-1}^* \hat{\beta}_{3t}^*) - F(\hat{\beta}_{0t}^* + X_{1it-1}^* \hat{\beta}_{1t-1}^* + X_{2it-1}^* \hat{\beta}_{2t}^* + X_{3it-1}^* \hat{\beta}_{3t}^*). \quad (24)$$

Similarly, the contribution of $\hat{\beta}_2^*$ can be written as:

$$\frac{1}{N_{t-1}} \sum_{i=1}^{N_{t-1}} F(\hat{\beta}_{0t}^* + X_{1it-1}^* \hat{\beta}_{1t}^* + X_{2it-1}^* \hat{\beta}_{2t}^* + X_{3it-1}^* \hat{\beta}_{3t}^*) - F(\hat{\beta}_{0t}^* + X_{1it-1}^* \hat{\beta}_{1t}^* + X_{2it-1}^* \hat{\beta}_{2t-1}^* + X_{3it-1}^* \hat{\beta}_{3t}^*). \quad (25)$$

Note that the difference being evaluated in equations (22) and (23) is the average difference in Internet access between years t and $t-1$, and has nothing to do with rural or urban differences. The analysis just described decomposes the effects of various parameter shifts on why average Internet access rates have changed over time. For example, if rural - urban status has an effect on Internet access over time (as hypothesized

⁴⁶ Note that these three groups of variables are all hypothesized to have unstable parameters in Table 10.

⁴⁷ While education and income levels (X_{1it}^*) and telecommunications infrastructure (X_{3it}^*) may vary significantly from year to year, rural-urban status (X_{2it}^*) is not expected to shift significantly over time.

⁴⁸ However, the time period of the characteristics used does matter. Particularly, equation (22) uses characteristics from time period $t-1$, while equation (23) uses characteristics from time period t . As a sensitivity check, the characteristics in both time periods will be used and the decomposition results will be compared.

by the core to periphery diffusion in Table 10), the change in $\hat{\beta}_2^*$ over time will have a significant contribution to equations (22) and (23). It is important to remember that only one term on the right-hand side of equations (22) and (23) is being decomposed in this manner – in this case, the second term. As noted above, the first term represents the portion of the yearly difference due to variation in household characteristics over time, and can be decomposed in a manner similar to the decompositions performed for a single year. However, the inter-temporal decomposition is also concerned with the importance of the parameters associated with these characteristics. Hence, differences in Internet access rates between the two time periods can be broken into a “characteristic differences” term and a “parameter differences” term. Resulting decompositions from both terms can be compared to derive the most important factors for changes in Internet access over time. Additionally, the same issues hold for this analysis as they did for the analysis in a single time period – namely, the effect of the sample drawn and the order in which the explanatory variables or parameters are introduced.⁴⁹ These issues are addressed in the same manner as above – repeating the sampling process for a large number of draws, and varying the order in which the variables or parameters enter the analysis.

High-speed vs. Dial-up Access: Nested Logit Model

Given the more complicated form of the nested logit specification, a decomposition technique similar to that suggested for the logit model is not proposed. Instead, a generalized extension of Nielson's (1998) decomposition technique will be implemented to isolate the impact of rural-urban parameter estimate differences. Equation (16) dealing with the nested logit probability of choosing alternative j is rewritten as:

⁴⁹ The effect of the sample drawn will only be an issue for the first term in equations (22) and (23), since decomposing the second term does not require the sampling procedure to be used. Also note that the first term uses changes in variables, while the second term uses changes in parameters to complete the decomposition.

$$P_{kj} = \left(\frac{\exp\left(\frac{1}{\tau_k} V_j\right)}{\sum_{l \in J^k} \exp(V_l)} \right) \left(\frac{\exp(\tau_k IV_k)}{\sum_{k \in K} \exp(\tau_k IV_k)} \right) = F_j[X, \beta, \tau] \quad (26)$$

since utility (V_j) is expressed in terms of X and β . The associated log-likelihood function then becomes:

$$\ln(L) = \sum_{i=1}^{N^U} \sum_{j=0,1,2} S_{ij} \ln F_j[X_i^U, \beta, \tau] + \sum_{i=1}^{N^R} \sum_{j=0,1,2} S_{ij} \ln F_j[X_i^R, (\beta + \delta), \tau] \quad (27)$$

where $S_{ij} = 1$ when household i chooses alternative j , and is 0 otherwise. As in the general logit model above, the subscript $G=(U,R)$ represents the metropolitan status of household i , N^G is the total number of households having status G , and X_i^G is a vector of characteristics for household i with status G . Maximization of equation (27) gives estimated parameter vectors for urban households ($\hat{\beta}$) and the associated shift for rural households ($\hat{\delta}$). In order to assess the roles that the three non-temporal factors play, the following three probabilities are simulated:

$$\hat{P}_{uj} = \frac{\sum_{i=1}^{N^U} F_j[X_i^U, \hat{\beta}, \hat{\tau}]}{N^U} \quad (28)$$

$$\hat{P}_{rj} = \frac{\sum_{i=1}^{N^R} F_j[X_i^R, (\hat{\beta} + \hat{\delta}), \hat{\tau}]}{N^R} \quad (29)$$

$$\hat{P}_{rj}^0 = \frac{\sum_{i=1}^{N^R} F_j[X_i^R, \hat{\beta}, \hat{\tau}]}{N^R} \quad (30)$$

\hat{P}_{uj} and \hat{P}_{rj} are the average probabilities of having Internet access equal to type j for urban and rural households, and will yield the averages displayed in Table 6 for rural and urban areas of the U.S. \hat{P}_{rj}^0 is not a probability that has an empirical counterpart; however, it is the most important simulation in this decomposition technique. \hat{P}_{rj}^0 is the average probability of having Internet access equal to type j for rural households using the parameter vector associated with urban households. Essentially, this probability uses

rural characteristics, but urban parameter vectors. This simulated probability allows us to split the total difference between rural and urban Internet access for type j into two distinct components:

$$\hat{P}_{uj} - \hat{P}_{rj} = (\hat{P}_{uj} - \hat{P}_{rj}^0) + (\hat{P}_{rj}^0 - \hat{P}_{rj}) \quad (31)$$

Equations (28) and (29) indicate that the first term on the right-hand side of equation (31) uses urban parameters for both rural and urban households, and hence isolates differences in *attributes* (or characteristics) between households in rural and urban areas. Similarly, equations (29) and (30) indicate that the second term $(\hat{P}_{rj}^0 - \hat{P}_{rj})$ isolates differences in underlying *parameters* between the rural and urban groups.

By changing the vectors X_i^U and X_i^R to include different factors associated with the digital divide, the importance associated with each factor can be assessed in the term $(\hat{P}_{uj} - \hat{P}_{rj}^0)$. To begin the analysis, the initial model will use $X_i^U = X_i^R = 1$, so that no household characteristics are included. In this model, \hat{P}_{uj} will equal \hat{P}_{rj}^0 . Thus, equation (31) simplifies to $(\hat{P}_{rj}^0 - \hat{P}_{rj})$ and the parameter $\hat{\delta}$ will account for all rural – urban differences in the various types of access. Next, models will include a factor (for instance, education levels of all households), so that X_i^U does not equal X_i^R . Hence, when \hat{P}_{rj}^0 is calculated, the education characteristics of rural households are used, along with the parameter vector associated with urban households. Thus, the term $(\hat{P}_{uj} - \hat{P}_{rj}^0)$ will indicate how much of the divide for type j is due to differences in education levels between the two areas. The "leftover" portion of the divide still associated with the parameter vector $\hat{\delta}$ should become smaller if the explanatory variables used are a factor in the divide. Later, additional explanatory variables can be added to the vectors X_i^U and X_i^R . If these variables are significant components of the rural-urban digital divide, the term $(\hat{P}_{uj} - \hat{P}_{rj}^0)$ should continue to have more explanatory power, while the term $(\hat{P}_{rj}^0 - \hat{P}_{rj})$ should continue to decrease in importance. Table 14 displays this sequential decomposition of the nested logit model in tabular form.

Table 14. Decomposition of Nested Logit Specification

Specification	Explanatory Variables	Variables Added	$(\hat{P}_{uj} - \hat{P}_{rj})$	
			$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$
			Differences in Attributes	Differences in Parameters
(1)	Constant term		0	rural intercept (δ)
(2)	(1) + Education Levels	X_E	X_E	rural intercept (δ)
(3)	(2) + Income Levels	X_I	$X_E + X_I$	rural intercept (δ)
(4)	(3) + Other Household Characteristics	X_O	$X_E + X_I + X_O$	rural intercept (δ)
(5)	(4) + Network Externalities	X_N	$X_E + X_I + X_O + X_N$	rural intercept (δ)
(6)	(5) + DCT Infrastructure	X_T	$X_E + X_I + X_O + X_N + X_T$	rural intercept (δ)

Hypotheses: $\delta(1) > \delta(2) > \delta(3) > \delta(4) > \delta(5) > \delta(6)$ for dial-up and high-speed access
 Importance of X_T for high-speed access > Importance of X_T for dial-up access

It is important to note that the more complicated specification of the nested logit model apparent in $F_j[X, \beta, \tau]$ (equation (26)) discourages the distinct decomposition of the effects of each component, such as the decomposition performed on the logit model for general access. However, observing the changes to equation (31) as the dependent variables are varied will result in an "intuitive" analysis of the importance of each factor. One issue with this technique is the sensitivity of these results to the ordering in which the dependent variables enter the analysis (due to the non-linearity of the nested logit functional form). To account for this, several orderings of specifications (2) through (6) in Table 14 will be performed, and the resulting decompositions will be compared.

Similar to the model for general access, the method for isolating the contribution of various factors for no access / dial-up / high-speed access can be altered to incorporate a temporal dimension. In particular, the log likelihood function expressed in equation (27) can be rewritten as

$$\ln(L) = \sum_{i=1}^{N_t} \sum_{j=0,1,2} S_{ij} \ln F_j[X_{it}, \beta^*, \tau^*] + \sum_{i=1}^{N_{t-1}} \sum_{j=0,1,2} S_{ij} \ln F_j[X_{it-1}, (\beta^* + \delta^*), \tau^*] \quad (32)$$

where X_{it} and X_{it-1} are vectors of explanatory variables for time periods t and $t-1$, respectively. Hence, maximization of equation (32) results in parameters for time period t ($\hat{\beta}^*$) and the associated shift for time period $t-1$ ($\hat{\delta}^*$). Similar to the preceding

discussion, the differences in the types of Internet access between periods t and $t-1$ can be recovered from the following simulations:

$$\hat{P}_{jt} = \frac{\sum_{i=1}^{N_t} F_j[X_{it}, \hat{\beta}^*, \hat{\tau}^*]}{N_t} \quad (33)$$

$$\hat{P}_{jt-1} = \frac{\sum_{i=1}^{N_{t-1}} F_j[X_{it-1}, (\hat{\beta}^* + \hat{\delta}^*), \hat{\tau}^*]}{N_{t-1}} \quad (34)$$

$$\hat{P}_{jt-1}^0 = \frac{\sum_{i=1}^{N_{t-1}} F_j[X_{it-1}, \hat{\beta}^*, \hat{\tau}^*]}{N_{t-1}} \quad (35)$$

where \hat{P}_{jt} is the probability of type j access in time period t , \hat{P}_{jt-1} is the probability of type j access in time period $t-1$, and \hat{P}_{jt-1}^0 is the probability of access equal to type j for households in time period $t-1$ using the parameter vector associated with households in time period t . The total difference in type j Internet access between time periods t and $t-1$ can then be written as:

$$\hat{P}_{jt} - \hat{P}_{jt-1} = (\hat{P}_{jt} - \hat{P}_{jt-1}^0) + (\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}) \quad (36)$$

and decomposed in the same way as rural – urban differences. Hence, the vector X_{it} and X_{it-1} will be varied to include different explanatory variables, and the effect that these changes have on the two components $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$ and $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$ will be observed. Similar to above, the time intercept vector (and hence $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$) should decrease in explanatory power as more explanatory variables are added to the analysis and account for an increasing portion of the inter-temporal differences. Table 15 is an inter-temporal counterpart to Table 14, and shows how the nested logit will be "decomposed" into the various components affecting types of Internet access over time.

Table 15. Inter-temporal Decomposition of Nested Logit Specification

	$X_{it} = X_{it-1}$		$(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	$(\hat{P}_{jt} - \hat{P}_{jt-1})$	$(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$
Specification	Explanatory Variables	Variables Added	Differences in Attributes		Differences in Parameters
(1)	Constant term		0		time intercept (δ^*)
(2)	(1) + Education Levels	X_E	X_E		time intercept (δ^*)
(3)	(2) + Income Levels	X_I	$X_E + X_I$		time intercept (δ^*)
(4)	(3) + Other Household Characteristics	X_O	$X_E + X_I + X_O$		time intercept (δ^*)
(5)	(4) + Network Externalities	X_N	$X_E + X_I + X_O + X_N$		time intercept (δ^*)
(6)	(5) + DCT Infrastructure	X_T	$X_E + X_I + X_O + X_N + X_T$		time intercept (δ^*)

Hypotheses: $\delta^*(1) > \delta^*(2) > \delta^*(3) > \delta^*(4) > \delta^*(5) > \delta^*(6)$ for dial-up and high-speed access
 Importance of X_T for high-speed access > Importance of X_T for dial-up access

The nature of this decomposition requires the use of the full specification (i.e. specification (6) in Table 15) in order to evaluate the role played by changing parameters over time. Recalling the form of equations (33) through (36), the analysis described above will tell us how much of the inter-temporal gap is due to differences in characteristics. The remainder of each access-specific gap is then due to changes in parameters. To complete the decomposition, parameters from each group of explanatory variables from time t will be incrementally replaced with those from time $t-1$. The resulting change in access rates can then be attributed to the change in that parameter. Hence, the roles played by parameter shifts such as those associated with core-to-periphery diffusion and early adopter characteristics will be addressed by this analysis.

To summarize, for any particular year, the roles of the three non-temporal factors (household characteristics, technology infrastructure, and social networks) in the general digital divide will be determined using a non-linear decomposition technique similar in spirit to the Oaxaca-Blinder decomposition. Additionally, differences in general Internet access rates between years will be decomposed into the roles played by characteristics and parameters that vary over time periods. Assessing the roles of the various factors in any given year and calculating what drives changes in inter-temporal access rates should provide important insights for the formulation of policy prescriptions to effectively address the general rural – urban digital divide. Moving to the specification that allows for distinction between high-speed and dial-up access, the complexity of the nested logit

equations discourages use of a decomposition technique similar to that used for the simple logit model. However, an intuitive decomposition is still accomplished by incrementally adding explanatory variables to the analysis and seeing how the resulting probabilities are affected for each type of access. A similar examination of inter-temporal differences allows for evaluation of the factors that have had the largest effects over time. Ascertaining the relative importance of all relevant factors is crucial for the formation of policy prescriptions that will be effective in closing the rural - urban digital divide for all types of access.

Chapter 4: Results

The first section of Chapter 4 discusses the general logit model results, presenting the parameter estimates and significance of the variables included in the final specification. The section also identifies which variables display significant temporal shifts and hence have become either more or less important in determining Internet access over time. The results of a logit model that allows for differences in parameter estimates between rural and urban areas in 2003 are then discussed, with particular focus on the rural parameter vector which shows how household characteristics affect Internet access differently. Section 2 decomposes the general rural – urban digital divide for each year of data into the contributions from the factors outlined previously. Section 3 extends this analysis to decompose the increase in access rates over time into the roles played by both shifting characteristics and shifting parameters. Section 4 then presents the results of the nested logit model, focusing on the parameter interpretation of the significant variables included in the analysis. Section 5 uses this nested logit specification to decompose the no access – dial-up access – high-speed access divides into contributions from various factors. Finally, section 6 performs an inter-temporal decomposition on the nested logit model to determine the roles played by various characteristics and parameters as access rates shift over time.

4.1 - General Logit Model Results

The final full specification for the simple yes-no access decision is:

$$y_{it}^* = X_{it}B_t + Z_{it}\delta + R_{it}\gamma_t + D1_{it}\tau_{1t} + D2_{it}\tau_{2t} + N_{it}\pi + \varepsilon_{it} \quad (36)$$

where X_{it} is a vector of household income and education levels at time t , Z_{it} is a vector of other household characteristics, R_{it} denotes rural status in time t , $D1_{it}$ and $D2_{it}$ are measures of cable and DSL access in time t , N_{it} is a measure of network externalities in

⁵⁰ Note that this specification differs from the one in equation (7) in that no critical mass term N_{it}^2 is included. This is due to the lack of significance of the critical mass term, and the reduction in explanatory power of the regional density term (N_{it}) when N_{it}^2 is included.

time t for household i ; $\beta_t, \delta, \lambda_t, \tau_{1t}, \tau_{2t}$, and π are the associated parameter vectors, and ε_{it} is the associated error term.

Table 16 displays the logit results under this specification for the year 2003 in the first column. Columns for 2001 and 2000 then show coefficient “shifts” in these years compared to 2003. This allows us to observe whether or not the shifts in parameter estimates of independent variables over time have been significant in explaining Internet access. As a general note, a significant positive (negative) coefficient on a discrete variable indicates an increase (decrease) in the probability of Internet access *relative to the base group for that variable* for the year 2003. Again, coefficients from 2000 and 2001 represent shifts from this 2003 base. For instance, a significant positive coefficient on “high-school” in 2003 would imply that households headed by individuals with this level of education are more likely to have Internet access than are households headed by individuals with the base group value (less than high-school education in this case). Meanwhile, a significant positive coefficient for “high-school” in 2001 would imply that this level of education was more important in 2001 than in 2003, since the 2001 coefficient represents a shift from its 2003 counterpart. Note that the base groups are all explicitly identified in Table 11.

Table 16. Logit Results for General Internet Access (2000 - 2003)

	2003		2001 Shifts		2000 Shifts	
	Coefficient	S.E.	Coefficient	S.E.	Coefficient	S.E.
hs	0.6185	0.0510 ***	-0.0050	0.0729	0.0481	0.0788
scoll	1.2538	0.0528 ***	-0.1115	0.0749	0.0244	0.0806
coll	1.5520	0.0605 ***	-0.1291	0.0856	0.0177	0.0894
collplus	1.7105	0.0729 ***	-0.1608	0.1011	-0.0005	0.1035
faminc1	-0.1569	0.1116	-0.0648	0.1667	-0.0922	0.1836
faminc2	-0.2470	0.1155 **	0.0343	0.1698	0.0627	0.1860
faminc3	-0.0628	0.1045	0.1326	0.1514	-0.0686	0.1673
faminc4	-0.0333	0.1072	-0.0094	0.1553	0.2580	0.1647
faminc5	0.0902	0.0952	-0.0133	0.1381	0.1189	0.1466
faminc6	0.2605	0.0915 ***	0.0366	0.1314	-0.0114	0.1405
faminc7	0.3232	0.0900 ***	0.1319	0.1306	0.1637	0.1385
faminc8	0.5116	0.0902 ***	0.2099	0.1309	0.1879	0.1390
faminc9	0.7233	0.0936 ***	0.0084	0.1341	-0.0293	0.1419
faminc10	0.9508	0.0887 ***	0.0128	0.1282	0.0121	0.1357
faminc11	1.1027	0.0925 ***	0.0632	0.1319	0.0031	0.1394
faminc12	1.3726	0.0933 ***	0.0285	0.1340	-0.0063	0.1408
faminc13	1.6671	0.0905 ***	0.1351	0.1298	0.0593	0.1368
nm	0.0620	0.0541	-0.0268	0.0748	0.0522	0.0746
netatwork	0.5325	0.0395 ***	-0.0433	0.0535	-0.4710	0.0555 ***
black	-0.6728	0.0482 ***	-0.0712	0.0672	-0.1484	0.0709 **
othrace	-0.1981	0.0704 ***	0.2189	0.1020 **	0.0606	0.1001
hisp	-0.6647	0.0524 ***	-0.0238	0.0767	-0.0634	0.0782
age	0.0606	0.0059 ***	-0.0167	0.0081 **	-0.0181	0.0085 **
age2	-0.0008	0.0001 ***	0.0002	0.0001 *	0.0001	0.0001
sex	0.0118	0.0306	-0.0337	0.0426	-0.0137	0.0433
married	0.5449	0.0339 ***	0.0349	0.0474	-0.0751	0.0476
chld1	0.2515	0.0468 ***	-0.0448	0.0639	-0.3308	0.0672 ***
chld2	0.2992	0.0511 ***	0.0467	0.0698	-0.3316	0.0708 ***
chld3	0.1913	0.0742 **	0.0746	0.1014	-0.3274	0.1072 ***
chld4	0.1484	0.1313	0.0030	0.1702	0.0940	0.1920
chld5	0.1944	0.2075	-0.1061	0.2826	-0.3202	0.2310
regdensity	2.3540	0.2799 ***	0.0441	0.3780	0.5573	0.3791
cableacces	0.0094	0.0831	-0.1135	0.1203	0.0183	0.1262
dslaccess	0.0730	0.0608	-0.0053	0.0851	0.0348	0.0909
retired	0.1876	0.0575 ***	-0.1243	0.0814	-0.1762	0.0866 **
constant	-3.9747	0.2463 ***	0.2804	0.3223	0.0549	0.3210

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10$, 0.05 , and 0.01 levels, respectively. 2001 and 2000 coefficients represent shifts on 2003 coefficients.

Most of the independent variables have the hypothesized signs and significance in 2003. In particular, the parameter values for education are positive, and increase as the level of education increases. This implies that, relative to a household headed by an individual with no high school education, higher levels of education increase the relative odds of a household having Internet access. Similarly, the parameter values for income are significantly positive after income reaches \$20,000 (faminc6). All income parameters should be interpreted relative to the income base group – in this case, households making

less than \$5,000 per year. These parameters increase in value as the income level rises, meaning that the propensity of Internet access increases with income. Additionally, the presence of Internet access at work (netatwork), having a married household head, and the presence of one, two, or three children all positively impact the probability of Internet access. The significant positive coefficient on regdensity indicates that local connectivity rates are important in the Internet access decision, with higher local rates resulting in increased probability of access for a household. On the opposite end of the spectrum, households headed by individuals that are Black, another non-White race, or Hispanic are less likely to have Internet access.

Several variables are notably lacking significance in 2003. First, the availability of cable (cableaccess) and DSL (dslaccess) are not significant. This is also true for the rural status of the household (nm). This lack of significance implies that, after controlling for other variables such as education, income, and other household characteristics, the level of DCT infrastructure available to and rural / urban status of the household do not influence the probability of Internet access.

Table 16 also indicates that very few of the variables show significant shifts over time. In fact, only the age coefficient is significant when interacted with time dummies for *both* 2001 and 2000. The coefficients associated with these interaction terms are negative, which, when coupled with the positive coefficients in 2003, indicates that age has increased in importance as time has progressed.⁵¹ Similarly, the coefficients for netatwork, chld1, chld2, chld3, and retired are all negative and significant when interacted with 2000 dummy variables, indicating that the influence of these characteristics on residential Internet access has increased with time. The increased importance of access at work (netatwork) may indicate a rising recognition of the type of tasks Internet access can accomplish. Additionally, the increased importance of children (chld1, chld2, chld3) may reflect an increasing reliance on the Internet for homework-related activities, as many public school systems have encouraged their students to take advantage of this information source. The coefficient for a Black household head is also negative in 2000; however, it is shifted from a negative coefficient in 2003, meaning that

⁵¹ Age has increased in importance over time because the 2000 and 2001 parameters for age are “shifts” from the 2003 parameter. These negative shifts from a positive original parameter reduce the impact in that year; hence, the overall impact of age in 2000 or 2001 is lower than its impact in 2003.

the possession of this racial characteristic is becoming *less* of a factor over time. The most noticeable trend in Table 16, however, is the lack of significance for the vast majority of variables when interacted with time dummies. Hence, the significance of characteristics such as education, income, and availability of DCT infrastructure in determining Internet access has not varied substantially over the three years represented by this data. This lack of significant impact over time implies that hypotheses involving shifting parameters over time (early adopter, core to periphery diffusion, and increasing importance of DCT infrastructure) may not hold empirically.⁵² These hypotheses are tested in section 4.3.

Due to the lack of several variables from the data in 1997 and 1998 (namely DCT infrastructure and information on children), a separate regression was run for these years. Table 17 reports the results of this regression in a manner similar to those for 2000 – 2003: that is, the column for 1998 represents the base, while coefficients for 1997 are shifts from this base.

⁵² Recall that increasing importance of DCT infrastructure was only hypothesized to occur for the model involving high-speed access, and that DCT infrastructure was not expected to play a role in the model for general access.

Table 17. Logit Results for General Internet Access (1997 – 1998)

	1998		1997 Shifts	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.6004	0.0773 ***	0.3606	0.1640 **
scoll	1.2651	0.0770 ***	0.6020	0.1620 ***
coll	1.5987	0.0806 ***	0.5408	0.1661 ***
collplus	1.7292	0.0857 ***	0.5904	0.1716 ***
faminc1	-0.1581	0.1673	-0.2904	0.2621
faminc2	-0.0309	0.1629	-0.0149	0.2534
faminc3	0.0292	0.1487	-0.3975	0.2407 *
faminc4	-0.0938	0.1541	-0.1878	0.2417
faminc5	0.0907	0.1311	-0.2070	0.2040
faminc6	0.1339	0.1260	-0.0881	0.1881
faminc7	0.3630	0.1233 ***	-0.2015	0.1857
faminc8	0.5087	0.1228 ***	-0.3281	0.1831 *
faminc9	0.6457	0.1225 ***	-0.3450	0.1833 *
faminc10	0.8748	0.1176 ***	-0.4497	0.1748 **
faminc11	1.1195	0.1185 ***	-0.6032	0.1765 ***
faminc12	1.2238	0.1200 ***	-0.4417	0.1778 **
faminc13	1.5330	0.1178 ***	-0.5645	0.1741 ***
nm	-0.0344	0.0492	0.0465	0.0798
netatwork	0.1535	0.0400 ***	1.0487	0.0587 ***
black	-0.8543	0.0627 ***	0.0519	0.1026
othrace	-0.0127	0.0763	-0.3213	0.1176 ***
hisp	-0.6984	0.0719 ***	0.0968	0.1179
age	0.0413	0.0073 ***	-0.0247	0.0118 **
age2	-0.0007	0.0001 ***	0.0001	0.0001
sex	0.1458	0.0333 ***	0.1697	0.0539 ***
married	0.3225	0.0364 ***	-0.1483	0.0574 **
regdensity	2.9500	0.2720 ***	1.5302	0.5467 ***
retired	-0.0653	0.0741	0.3288	0.1270 **
constant	-4.1989	0.2138 ***	-0.3574	0.3439

Note: *, **, and *** indicate statistically significant differences from zero at the p = 0.10, 0.05, and 0.01 levels, respectively. 1997 coefficients represent shifts on 1998 coefficients.

The majority of the coefficients from the 1997 – 1998 regression follow a pattern similar to those displayed for its 2000 – 2003 counterpart. Higher education and income levels remain positively associated with Internet access. A Black household head reduces the probability of Internet access relative to a White household head, and a Hispanic household head reduces the probability of access relative to a non-Hispanic household head. However, shifts from these 1998 parameters are much more significant than was the case for the 2000 – 2003 data. In particular, the coefficients for education levels are all positive and significant when interacted with a 1997 dummy variable, while higher levels of income are all negative and significant when interacted with the same dummy variable. This implies that, as hypothesized in section 3.2, education levels were more

important in the early days of adoption, with these effects diminishing over the following year. However, the significant negative coefficients associated with income levels imply that income became *more* important as time continued, which is counter to the “early adopter” hypothesis.⁵³ A multitude of other shifts are significant, including the presence of Internet access at work (netatwork), a male household head (sex), and levels of local connectivity (regdensity), all of which became less important in 1998 than they were in 1997. On the other side of the spectrum, the age of the household head (age) and whether the head is married (married) became more important in 1998 than they were in 1997. It is worth noting that similar to the regression performed for the 2000 – 2003 time period, the rural – urban status of the household (nm) is not significant.

To further test for differences that may exist in propensities of rural and urban households to access the Internet, a separate specification is utilized. This specification allows parameter estimates to differ between rural and urban areas by including a rural interaction term for each explanatory variable. Hence, the resulting rural parameter coefficient represents a "shift" on the urban coefficient caused by the rural location of a household. Table 18 displays the results of this specification for the year 2003, while similar results for 1997 through 2001 are shown in Appendix E.

⁵³ Recall that neither of the "early adopter" hypotheses held for the 2000 – 2003 time period, as no shifts in education or income were significant.

Table 18. Logit Regression for Urban – Rural Internet Access (2003)

Variables	Urban		Rural	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.6316	0.0589 ***	-0.0511	0.1174
scoll	1.2938	0.0606 ***	-0.1728	0.1235
coll	1.5728	0.0681 ***	-0.0804	0.1514
collplus	1.7406	0.0813 ***	-0.1692	0.1896
faminc1	-0.1753	0.1307	0.1246	0.2526
faminc2	-0.2844	0.1312 **	0.1621	0.2779
faminc3	-0.0838	0.1214	0.1190	0.2391
faminc4	-0.0873	0.1247	0.2373	0.2458
faminc5	0.0397	0.1100	0.2170	0.2215
faminc6	0.2219	0.1046 **	0.1858	0.2157
faminc7	0.2480	0.1034 **	0.3446	0.2108
faminc8	0.4509	0.1032 ***	0.2803	0.2134
faminc9	0.6481	0.1065 ***	0.3352	0.2236
faminc10	0.9219	0.1012 ***	0.1274	0.2112
faminc11	1.0212	0.1052 ***	0.3915	0.2209 *
faminc12	1.3627	0.1064 ***	0.0238	0.2227
faminc13	1.6484	0.1023 ***	0.0324	0.2229
netatwork	0.5212	0.0439 ***	0.0721	0.1008
black	-0.6767	0.0513 ***	0.0292	0.1535
othrace	-0.1740	0.0790 **	-0.1719	0.1652
hisp	-0.6911	0.0557 ***	0.2762	0.1579 *
peage	0.0570	0.0066 ***	0.0164	0.0143
age2	-0.0008	0.0001 ***	-0.0002	0.0001
sex	0.0349	0.0348	-0.1131	0.0729
married	0.5169	0.0387 ***	0.1249	0.0806
chld1	0.2659	0.0527 ***	-0.0712	0.1144
chld2	0.3081	0.0583 ***	-0.0392	0.1218
chld3	0.2160	0.0865 **	-0.1078	0.1668
chld4	0.3029	0.1537 **	-0.6483	0.2835 **
chld5	0.1827	0.2313	0.0989	0.5352
regdensity	2.2890	0.3516 ***	0.4919	0.6075
cableaccess	0.0418	0.1102	-0.1075	0.1716
dslaccess	0.0767	0.0664	0.0217	0.1773
retired	0.1323	0.0662 **	0.2481	0.1331 *
constant	-3.8844	0.1174 ***	-0.5059	0.5410
Log-likelihood	-19769.9			

Note: ***, **, and * represent statistically significant differences from zero at the $p = 0.01$, 0.05 , and 0.10 levels, respectively. Rural coefficients represent shifts on urban coefficients.

The significant parameters for urban households in Table 18 coincide with those for the entire population during the year 2003 in Table 16. This is expected, given the lack of significance for most of the estimated rural area parameter shifts. The lack of significance of these rural interaction terms indicates that, for the most part, the impact of

the characteristics displayed in Table 18 on a household's propensity for Internet access does not vary between rural and urban areas. The only significant rural shifts in 2003 are for households with income levels between \$50,000 and \$60,000 (faminc11), a Hispanic household head (hisp), four children (chld4), and retired household heads (retired). Interpreting these shifts depends on the signs (and values) of both the rural and urban parameters. For instance, the 'faminc11' parameter is positive and significant for urban households, and is positive and marginally significant for the rural shift. Thus, this level of income creates a greater propensity for Internet access in rural areas than urban areas relative to the base income level. Looking now at the parameters associated with having a Hispanic household head (hisp), the rural parameter is positive and is shifted from a negative urban coefficient. Hence, rural households headed by a Hispanic individual show a higher propensity to access the Internet than their urban counterparts (although the overall effect of a Hispanic household head is still negative for rural households).⁵⁴ Similarly, the presence of four children in a household (chld4) generates different effects in rural and urban areas. However, the negative effect seen in rural areas is large enough to outweigh the positive shift seen in urban areas, and therefore chld4 is the only variable that results in a different qualitative influence on the propensity for access in one category (urban - positive) than for the other (rural - negative). The last significant rural shift, retired, is positive, indicating an even stronger propensity for Internet access for rural households with a retired household head. While these four significant parameters indicate that some minor differences exist in the propensities for Internet access between rural and urban areas, the majority of the rural parameter shifts in Table 18 are not significant. As results for 1997, 1998, 2000 and 2001 in Appendix E show, this general pattern of non-significant rural interaction terms also holds for the remaining years included in the study.⁵⁵ Thus, on the whole, rural – urban differences in the way that various characteristics influence Internet access are not likely to drive the digital divide.

⁵⁴ The overall effect of having a rural Hispanic household head (since the rural parameter is a shift on the urban parameter) is the summation of both the urban and rural parameters, which is negative in this case.

⁵⁵ One notable exception to this statement occurs in 1997, when education levels in rural areas demonstrated positive shifts from their urban counterparts, indicating a difference in the influence of education between rural and urban areas for this year.

The logit specifications discussed above provide some preliminary intuition on the importance of various factors to the general access decision. Income and education levels are positive and significant, but the early adopter framework only seems to hold for education levels in the very early years of the analysis. Other household characteristics do appear to have an effect on the probability of Internet access, with the importance of some characteristics varying over time, particularly age. For instance, the importance of age appears to be increasing over time, as does the influence of children. The importance of core to periphery diffusion (nm) and technology infrastructure (cableaccess and dslaccess) seem to be minimal, as none of these variables are statistically significant in any of the regressions. Furthermore, very few rural parameter shifts were significant when differences in propensities associated with specific characteristics were tested (Table 18), suggesting that the influence of various characteristics on Internet access is very similar between rural and urban areas. Finally, social networks (regdensity) are positively associated with Internet access, but their importance may be diminishing over time. The following sections provide a more formal assessment of the importance of individual factors to the rural – urban digital divide, and how their contribution has shifted over time.

4.2 - Decomposition of the General Digital Divide

Table 19 presents the results of the non-linear decomposition for general Internet access.⁵⁶ The first two lines of Table 19 indicate the percentage of rural and urban households with Internet access, and the third shows the resulting "digital divide" for each of the five years of CPS data. The remainder of the table reports the individual contributions from rural – urban differences in education, income, other household characteristics, network externalities, and DCT infrastructure variables.⁵⁷ Note that since no information on DCT infrastructure is available for 1997 and 1998, the decomposition performed in these years does not include the DCT variables.

⁵⁶ Unless noted otherwise, all decompositions use 1,000 random samples of urban households. The results did not dramatically shift when either 100 or 10,000 random samples were used, lending confidence to the estimates.

⁵⁷ The decompositions in Table 19 are estimated using coefficients from a "pooled" logit regression including both rural and urban households. Estimates from logit regressions using coefficients only from rural or urban samples are shown in Appendix F.

Table 19. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access, 1997 - 2003

	Year				
	2003	2001	2000	1998	1997
Urban Household Internet Access Rate	0.6123	0.5657	0.4667	0.3044	0.1769
Rural Household Internet Access Rate	0.4840	0.4294	0.3290	0.1805	0.0908
Rural / Urban Gap	0.1283	0.1363	0.1377	0.1239	0.0861
Contributions from Rural / Urban Differences in:					
Education Levels	0.0279	0.0229	0.0264	0.0219	0.0103
	21.7%	16.8%	19.2%	17.7%	12.0%
Income Levels	0.0434	0.0489	0.0487	0.0381	0.0173
	33.8%	35.9%	35.4%	30.8%	20.1%
Other Household Characteristics	0.0010	-0.0001	-0.0021	0.0024	0.0207
	0.8%	-0.1%	-1.5%	1.9%	24.0%
Network Externalities	0.0376	0.0466	0.0553	0.0402	0.0273
	29.3%	34.2%	40.2%	32.4%	31.7%
DCT Infrastructure	0.0035	-0.0001	0.0049		
	2.7%	-0.1%	3.6%		
All included variables	0.1134	0.1182	0.1332	0.1026	0.0756
	88.4%	86.7%	96.7%	82.8%	87.8%

Note: Percentages indicate the contribution of each group of variables to the rural / urban gap for that year.

The difference between rural and urban general Internet access rates ranges from 8.6 percentage points in 1997 to 13.8 percentage points in 2000. As expected, differences in education and income levels between urban and rural areas explain a large portion of this gap. Lower levels of education in rural areas account for between 12 and 22 percent of the gap, while lower levels of income account for between 20 and 36 percent of the gap. Differences in network externalities (as measured by regional rates of access) also play an important role, as they make up between 29 and 40 percent of the gap in a given year. With the exception of 1997, other household characteristics do not have much explanatory power, consistently making up less than 2 percent of the gap. Given the results of the general logit model (discussed in section 4.1), however, the minimal contribution of those factors grouped under “other household characteristics” could mask significant offsetting contributors, such as the positive impact of children in the household or the negative impact of a Black or Hispanic household head. Rural – urban

differences in DCT infrastructure explain very little, comprising only between –0.1 to 4 percent of the gap for the years where these variables are included. The decompositions indicate that group differences in all of the included variables explain between 83 and 97 percent of the gap in general access. These high numbers reveal that only a relatively small portion of the gap (between 3 and 17 percent) is left unexplained by rural – urban differences in the included variables.

Due to the use of different specifications for the 2000 – 2003 and 1997 – 1998 time periods, we cannot say anything about the contributions of various groups over the entire 1997 – 2003 period.⁵⁸ However, it is interesting to note that for the 2000 – 2003 time period, the contribution of differences in education levels slightly *increased* over time (rising from accounting for 19.2 percent to 21.7 percent of the divide), which is contrary to what was expected under the early adopter hypothesis. Since differences in education levels in rural and urban areas remained largely unchanged, education’s robustness as a determinant of residential Internet access largely accounts for its continued strong contribution to the rural – urban digital divide. The contribution of income levels is consistent at around 34 – 35 percent for this time frame. For the 1997 – 1998 time frame, neither component of the early adopter framework holds, as both education and income levels accounted for lower portions of the divide in the earlier year. A more detailed look into the inter-temporal aspect of the decomposition is provided in section 4.3. This inter-temporal decomposition looks at the roles of both changing rural – urban *characteristics* and *parameter estimates* over time.

Ordering of variables

As noted in section 3.5 above, the non-linearity of the logit model implies that the results may be sensitive to the order in which the variables are included. To explore this issue, Table 20 reverses the order of the explanatory variables. The majority of the estimates are very similar to those obtained with the original ordering; however, several estimates are noticeably different. In 2000, for example, the role played by education jumps from 19 percent under the initial orderings to 26 percent when the orderings are reversed, and

⁵⁸ The inclusion of additional variables for the 2000 – 2003 specification alters the significance of the original variables used in the 1997 – 1998 specification.

other household characteristics shift from explaining –2 percent of the divide to explaining –9 percent. Similar changes occur for the 1997 and 1998 results, as the role of education increases by about 10 to 15 percentage points and the impact of other household characteristics declines by 10 to 20 percentage points. The total contribution remains the same in all years because the sum of the individual contributions necessarily equals the first term on the right-hand side of equations (18) and (19).

Table 20. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access, 1997 – 2003 (Reverse Ordering)

	Year				
	2003	2001	2000	1998	1997
Urban Household Internet Access Rate	0.6123	0.5657	0.4667	0.3044	0.1769
Rural Household Internet Access Rate	0.4840	0.4294	0.3290	0.1805	0.0908
Rural / Urban Gap	0.1283	0.1363	0.1377	0.1239	0.0861
Contributions from Rural / Urban Differences in:					
Education Levels	0.0304	0.0280	0.0360	0.0343	0.0224
	23.7%	20.5%	26.1%	27.7%	26.0%
Income Levels	0.0412	0.0476	0.0492	0.0383	0.0183
	32.1%	34.9%	35.7%	30.9%	21.3%
Other Household Characteristics	0.0015	-0.0020	-0.0130	-0.0136	0.0033
	1.2%	-1.5%	-9.4%	-11.0%	3.8%
Network Externalities	0.0374	0.0449	0.0561	0.0436	0.0316
	29.2%	32.9%	40.7%	35.2%	36.7%
DCT Infrastructure	0.0029	-0.0003	0.0049		
	2.3%	-0.2%	3.6%		
All included variables	0.1134	0.1182	0.1332	0.1026	0.0756
	88.4%	86.7%	96.7%	82.8%	87.8%

Note: Percentages indicate the contribution of each group of variables to the rural / urban gap for that year.

As Fairlie (2003) notes, the sensitivity of the order in which the variables are introduced is dependent on the initial location in the logistic distribution and the movement inflicted by switching distributions of rural and urban characteristics. Fairlie suggests experimenting with the ordering in order to verify the robustness of the results. While there are $5! = 120$ different orderings for 2000, 2001, and 2003 and $4! = 24$ different orderings for 1997 and 1998; approximately 10 different orderings were attempted for each year, with all estimates lying in the intervals created by Tables 19 and

20. It is interesting to note that under this reversed ordering, the early adopter hypothesis appears to hold for both income and education over the years 2000 – 2003. The contribution of these characteristics has decreased over time. This is not true for the 1997 – 1998 time period, where the early adopter hypothesis once again appears to fail.⁵⁹ In general, although some differences exist when the ordering is varied, the dominant factors remain the same in all years and their impact is not greatly diminished.

The estimation procedure underlying these decomposition results assume that equation (36) is correctly specified. However, it is feasible that both the DCT infrastructure measures and the network externality measure might be endogenous – that is, an individual household's probability of adoption may impact the amount of DCT infrastructure or the local access rates in an area (as opposed to vice versa). If this is the case, then endogeneity does exist and the endogenous variables are correlated with the error term. This may result in biased parameter estimates for the variables in question. This problem is typically handled via the instrumented variables approach, where the variables in question are regressed on an "instrument" that, in the best case scenario, is highly correlated with DCT infrastructure or network externalities but *not* the error terms. Effectively testing for the existence of endogeneity requires the use of such an instrument (Hausman, 1978). Unfortunately, finding an instrument is difficult, given the lack of externality-related questions in the CPS questionnaire (such as percentage of friends with home access, or amount of interaction with neighbors) and the stand-alone nature of the infrastructure data. In particular, information on cost is notably missing from all data sources. The costs to cable and phone companies for providing access to an area and the cost of monthly access in that area would likely serve as useful instruments due to their probable correlation with DCT infrastructure and local adoption rates, respectively. This lack of relevant instruments limits the treatment options for the potential endogeneity.

It is intuitive that cable and phone companies take into account the expected subscriber rates in an area (i.e. the probability of adoption) before investing in high-speed infrastructure in that area. However, it is reassuring to note that alternative data used by

⁵⁹ Section 4.3, which deals with the inter-temporal digital divide, provides additional insight into the early adopter hypothesis.

Song (2005) suggests that households are not likely to move solely to get Internet access.⁶⁰ This implies that DCT infrastructure may in fact be an exogenous variable, and that concern over its endogeneity may not be warranted. The issue with the network externality proxy is not exactly the same. An alternative way of conceptualizing the network externality issue is to realize that this proxy is actually measuring regional differences in access. Association with these access propensities may arise from any number of factors underlying regional access rates *other than* network externalities (for instance, education and income levels, infrastructure levels, or the cost of access). Hence, the proxy used for "network externalities" may not be a very precise way of measuring this concept.

To address the endogeneity and lack of precision concerns about the network externality measure, separate decompositions were run after removing N_{it} from equation (36). The results are shown in Table 21 and lend confidence to the above results – namely the importance of education and income differences to the general digital divide. Furthermore, the role of differences in DCT infrastructure does not increase dramatically when the network externality proxy is removed from the decomposition.

⁶⁰ Song uses survey data from the UCLA Center for Communication Policy meshed with FCC form 477 data for 1999 and 2001. He derives access measures for both zip codes and counties, and infers that the lack of differences between these measures indicates that household movement to achieve access is not a serious problem.

Table 21. Decomposition of Rural – Urban Digital Divide in General Residential Internet Access (No Network Externality Term), 1997 - 2003

	Year				
	2003	2001	2000	1998	1997
Urban Household Internet Access Rate	0.6123	0.5657	0.4667	0.3044	0.1769
Rural Household Internet Access Rate	0.4840	0.4294	0.3290	0.1805	0.0908
Rural / Urban Gap	0.1283	0.1363	0.1377	0.1239	0.0861
Contributions from Rural / Urban Differences in:					
Education Levels	0.0285	0.0232	0.0273	0.0222	0.0109
	22.2%	17.0%	19.8%	17.9%	12.6%
Income Levels	0.0463	0.0503	0.0493	0.0358	0.0157
	36.1%	36.9%	35.8%	28.9%	18.2%
Other Household Characteristics	0.0031	-0.0001	-0.0121	-0.0141	0.0028
	2.4%	-0.1%	-8.8%	-11.4%	3.2%
DCT Infrastructure	-0.0053	0.0022	0.0081		
	-4.2%	1.6%	5.9%		
All included variables	0.0725	0.0755	0.0725	0.0438	0.0293
	56.5%	55.4%	52.7%	35.4%	34.0%

Note: Percentages indicate the contribution of each group of variables to the rural / urban gap for that year.

The results of the decomposition suggest that differences in education, income, and network externalities (when included) are the driving force behind the divide in general access. Based on the ordering used, these three variables account for between 65 and 100 percent of the divide for any given year. On the other hand, for the three years where data on DCT infrastructure is available, the results indicate that differences in this infrastructure never accounts for more than 6 percent of the divide. Hence, addressing the digital divide in general access will require policies that deal with the broader inequities in education and income, and promote community or localized access rates. Policies that focus solely on technology or DCT infrastructure will *not* address the primary causes of the general divide. Further analysis is needed to understand the impact of these variables on high-speed access.

4.3 - Inter-temporal Decomposition of the General Digital Divide

As noted in section 3.5, inter-temporal decomposition of the digital divide is concerned with both the role played by changing characteristics and the role played by

changing parameters. The rural – urban decomposition for a single year focuses only on the portion associated with different characteristics, which allows the specifications in equations (18) and (19) to be converted into a "most likely" scenario by using the results of a pooled regression.⁶¹ This is not the case for the inter-temporal decompositions, so the results must be expressed for *both* equation (22) and (23). Table 22 summarizes the results for both of these specifications for the time period between 1997 and 2003. In order to accomplish this decomposition, the set of regressed variables had to be the same for both 1997 and 2003. Due to the lack of data on DCT infrastructure and number of children in the household in 1997, these variables were removed from the regression.

Table 22. Summary of Inter-temporal Decomposition for General Internet Access, 1997 - 2003

	Equation	
	(22)	(23)
2003 Household Internet Access Rate	0.5874	0.5874
1997 Household Internet Access Rate	0.1595	0.1595
1997 / 2003 Gap	0.4279	0.4279
Contributions from Characteristic shifts	0.2105	0.3181
	49.2%	74.3%
Contributions from Parameter shifts	0.2174	0.1098
	50.8%	25.7%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in characteristics or parameters

Table 22 indicates that the percentage of households with Internet access rose by 43 percent from 1997 to 2003. It is apparent that the choice of base year for the decomposition has a significant impact on the results, as shifting characteristics over time make up 49 percent of this increase when equation (22) is used, while they make up 74 percent of the increase when equation (23) is used. Similarly, shifting parameters make up 51 percent of the increase under equation (22) and 26 percent of the increase under equation (23). Table 23 breaks these results down into the individual contributions of shifts in various characteristics and parameters. As noted above, since 1997 does not

⁶¹ This “index problem” is discussed in section 3.5, which also provides references such as Oaxaca and Ransom (1994) for the use of a pooled regression. Note that the direct (non-pooled) results from equations (18) and (19) are included in Appendix E

have any available data on DCT infrastructure, this variable was not included in the decomposition. Also, the effect of the age parameter shift is identified separately from the remaining "other household" parameters due to its surprisingly large impact on the results.

Table 23. Individual Contributions of Characteristics and Parameters to the Inter-temporal Decomposition in General Internet Access, 1997 - 2003

Individual Contributions from: Characteristic Shifts	Equation		Parameter Shifts	Equation	
	(22)	(23)		(22)	(23)
Education Levels	0.0114 2.7%	0.0065 1.5%	Education	-0.0808 -18.9%	-0.0685 -16.0%
Income Levels	0.0286 6.7%	0.0214 5.0%	Income	0.0684 16.0%	0.0604 14.1%
Other Household Charact	0.0086 2.0%	0.0269 6.3%	Age	0.1864 43.6%	0.2046 47.8%
Network Externalities	0.1619 37.8%	0.2633 61.5%	Other Household	-0.0035 -0.8%	-0.0046 -1.1%
			Network Externalities	-0.0474 -11.1%	-0.2049 -47.9%
			Intercept	0.0943 22.0%	0.1228 28.7%
All Included Variables	0.2105 49.2%	0.3181 74.3%	All Included Parameters	0.2174 50.8%	0.1098 25.7%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in individual groups of characteristics or parameters

As Table 23 shows, changing levels of characteristics between 1997 and 2003 consistently raised household Internet access rates. However, most of the contribution of changes in characteristics came from higher levels of network externalities, which were responsible for between 38 and 62 percent of the increase over time. By contrast, increasing education levels only accounted for between 2 to 3 percent of the increase in rates between 1997 and 2003, while increasing income levels accounted for between 5 and 7 percent of the increase.⁶² Shifts in the *parameters* associated with education and income characteristics had a number of countervailing effects. Recalling the form of equations (22) and (23), we can surmise that negative percentage shifts capture parameters that were more important in 1997 than in 2003, and positive percentage shifts

⁶² The categorical nature of the CPS income data prevents income from being measured in real terms. Hence, these nominal measures of income may slightly skew the contribution of income over time.

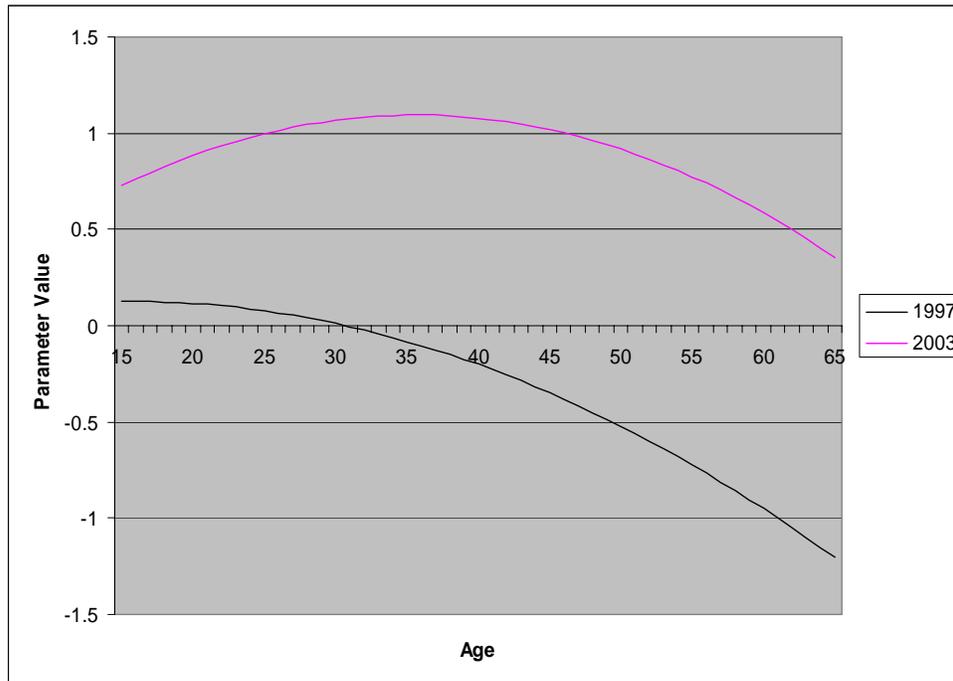
reflect parameters that increased in importance between 1997 and 2003. For example, differences between the 1997 and 2003 education parameters actually *increased* the temporal shift by around 16 – 19 percent (a negative percentage explained), implying that education parameters were more important for Internet access in 1997 than in 2003.⁶³ By contrast, differences in the income parameters accounted for around 15 percent of the increase in access rates.⁶⁴ Hence, shifts in education and income parameters somewhat offset each other. Similarly offsetting shifts are found for the network externalities and age parameters. Parameters for network externalities were much larger in 1997 than in 2003, and differences in these parameters contribute to an expected decrease in residential Internet access by between 11 and 48 percent.⁶⁵ Meanwhile, differences in the age parameters account for between 44 and 48 percent of the increase. This dramatic impact of the age parameter is surprising, especially given the negligible role for parameters involving other household characteristics (around –1 percent). Figure 16 displays the age – adoption propensity profiles resulting from the linear and quadratic age terms of the regression in 1997 and 2003.

⁶³ Again, this decrease in education parameters over time is consistent with the early adopter hypothesis.

⁶⁴ This increase in income parameters over time is contrary to the early adopter hypothesis.

⁶⁵ This large variance (11 to 48 percent) reflects the fact that equation (22) uses characteristics from 1997, while equation (23) uses characteristics from 2003. The network externalities proxy was much higher in 2003 than in 1997, reflecting a large change in the percentage of the gap explained.

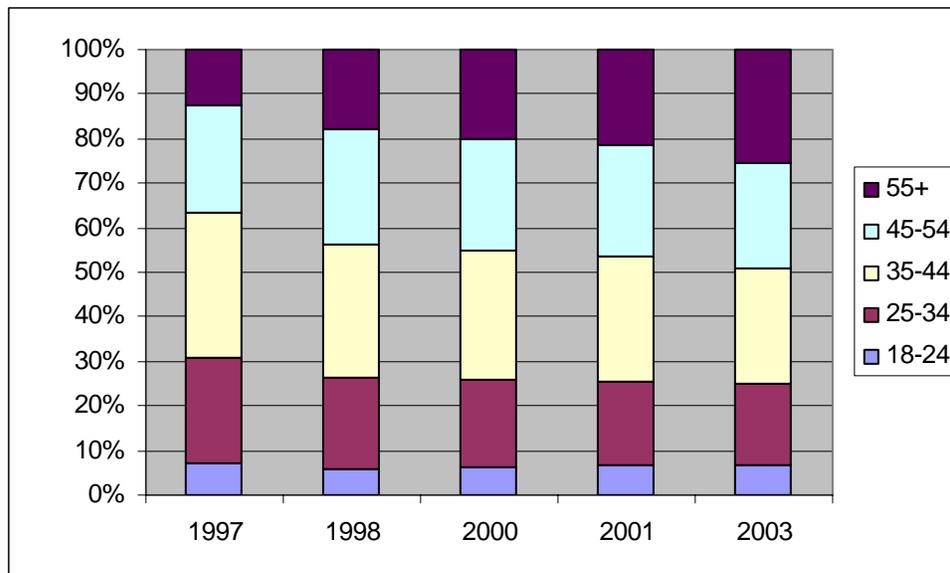
Figure 16. Age Parameter Values from 1997 and 2003 Regressions



Upon closer inspection, the quadratic form approximated by the 2003 parameters reaches its maximum at a later age than the form for the 1997 parameters, and results in significantly higher parameter estimates for the mean household head age, which is around 48 years in both 1997 and 2003. In fact, the resulting age parameter values from the 1997 regression are only positive for household heads under the age of 31. Hence, this large discrepancy in resulting parameter values makes the results of the decomposition a bit more intuitive. In essence, we can affirm that the Internet became more “age-friendly” over this period, with the shifting positive impact of age contributing a great deal to the increase in access rates. Descriptive statistics on the age of household heads with Internet access reinforce this notion, as older heads comprise a much higher percentage of all households with access as time progresses (Figure 17). Thus, while the hypothesized “core-to-periphery” diffusion does not seem to be supported by the data, a different type of diffusion appears to be taking place – from younger to older

households.⁶⁶ The increasing prevalence of older household heads with Internet access may lead one to suspect that the rural – urban gap may simply dissipate over time, since rural households have an age distribution that is more skewed to the right than their urban counterparts. However, closer inspection of the data reveals that the primary discrepancy between the two distributions occurs above the age of 75 (11.8 percent of rural households were over this age in 2003, compared with only 8.5 percent of urban households). At ages above 75, the parameter value shown in Figure 16 turns negative, implying that the differences in age actually tend to increase the rural – urban gap in Internet access. Hence, the shifting age and age2 parameters will *not* lead to a reduction in the rural – urban divide over time.

Figure 17. Age Profile of Household Heads with Internet Access



Source: CPS Computer and Internet Use Supplements, 1997, 1998, 2000, 2001, 2003.

To explore the impact of shifting levels of DCT infrastructure over time, the decomposition is performed for the period between 2000 and 2003. Rates of cable Internet and DSL availability are now included in the regressions, allowing the results to capture the effects of increasing levels of DCT infrastructure availability over time. It is

⁶⁶ As Tables 16 and 17 show, the parameters associated with the "nm" variable are insignificant in each year, indicating that rural status has not been decreasing in importance over time as the core-to-periphery hypothesis suggests.

important to note that this decomposition is being performed *even though* the parameters for DCT infrastructure are not significant in these regressions. Table 24 summarizes the results for the specifications in equations (22) and (23) for the time period 2000 – 2003.

Table 24. Summary of Inter-temporal Decomposition for General Internet Access, 2000 - 2003

	Equation	
	(22)	(23)
2003 Household Internet Access Rate	0.5874	0.5874
2000 Household Internet Access Rate	0.4394	0.4394
2000 / 2003 Gap	0.1480	0.1480
Contributions from Characteristic shifts	0.0783	0.0891
	52.9%	60.2%
Contributions from Parameter shifts	0.0697	0.0589
	47.1%	39.8%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in characteristics or parameters

Household Internet access rates rose by 15 percent from 2000 to 2003. Depending on the specification used, shifting characteristics accounted for between 53 and 60 percent of this increase, while shifting parameters accounted for between 47 and 40 percent. Table 25 displays the role played by DCT infrastructure as well as other individual variables. Again, the age parameter shift is broken out of the "other household characteristics" category due to its large effect.

Table 25. Individual Contributions of Characteristics and Parameters to the Inter-temporal Decomposition in General Internet Access, 2000 - 2003

Individual Contributions from: Characteristic Shifts	Equation		Parameter Shifts	Equation	
	(22)	(23)		(22)	(23)
Education Levels	0.0060 4.1%	0.0038 2.6%	Education	-0.0044 -3.0%	-0.0043 -2.9%
Income Levels	0.0103 6.9%	0.0099 6.7%	Income	-0.0087 -5.9%	-0.0084 -5.7%
Other Household Charact	0.0029 2.0%	-0.0033 -2.2%	Age	0.0932 63.0%	0.0926 62.6%
Network Externalities	0.0553 37.3%	0.0715 48.3%	Other Household	0.0435 29.4%	0.0508 34.3%
DCT Infrastructure	0.0038 2.6%	0.0071 4.8%	Network Externalities	-0.0426 -28.8%	-0.0575 -38.9%
			DCT Infrastructure	-0.0018 -1.2%	-0.0048 -3.2%
			Intercept	-0.0095 -6.4%	-0.0095 -6.4%
All Included Variables	0.0783 52.9%	0.0891 60.2%	All Included Parameters	0.0697 47.1%	0.0589 39.8%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in individual groups of characteristics or parameters

Similar to the results for the 1997 – 2003 time period, higher levels of network externalities are the dominant variable from a characteristic standpoint, explaining between 37 and 48 percent of the increase in residential Internet access rates. Additionally, the rising education and income levels continue to contribute between 3 and 7 percent of the temporal increase. The added DCT infrastructure variable, which accounts for higher levels of infrastructure between 2000 and 2003, explains only between 3 and 5 percent of the increased Internet access rates. This is not surprising, since these parameters were not statistically significant in either year. Turning to the parameter shifts, the availability of DCT infrastructure is more important in determining Internet access in 2000 than it is in 2003, which is why differences in these parameters actually increases the inter-temporal difference. However, the contribution of this DCT infrastructure parameter shift is minimal, accounting for between –1 and –3 percent of the difference. This minimal contribution is consistent with the hypothesis that DCT infrastructure does not play a major role in the general (yes – no) access decision. Once again the higher age parameter for 2003 is extremely important in explaining why Internet access rose between 2000 and 2003, accounting for around 63 percent of the

increase in access rates. Conversely, other parameters have different effects for this decomposition than they did for the 1997 – 2003 decomposition. For example, the shifting parameters for other household characteristics (particularly the increased importance associated with having children in the household) explain around 30 percent of the increase in access rates between 2000 and 2003. However, changing these parameters did not explain any of the increase in access rates between 1997 and 2003, indicating that parameters for other household characteristics became more important between 1997 and 2000.⁶⁷ Similarly, the parameters associated with income are less important in 2003 than they were in 2000, resulting in a decline in the percentage of the increase in access rates explained. This is the opposite of what occurred in the 1997 – 2003 decomposition. Hence, between 2000 and 2003, propensities associated with education, income, and network externalities became less important in determining Internet access, while propensities associated with other household characteristics (such as the number of children in the household) became more important.⁶⁸

Ordering of Variables

Similar to the procedures performed for the single-year decompositions, the sensitivity of the order in which the variables and parameters are introduced is now explored. Table 26 shows the individual contributions of the characteristics and parameter shifts between 1997 and 2003, with the original order reversed. Table 27 provides a similar assessment for the 2000 – 2003 time period.

⁶⁷ Note, however, that information on the number of children in the household was not present in the 1997 dataset.

⁶⁸ This decreasing importance of education and income is consistent with the early adopter hypothesis.

Table 26. Individual Contributions of Characteristics and Parameters to the Inter-temporal Decomposition in General Internet Access, 1997 – 2003, Order Reversed

Individual Contributions from: Characteristic Shifts	Equation		Parameter Shifts	Equation	
	(22)	(23)		(22)	(23)
Education Levels	0.0142 3.3%	0.0127 3.0%	Education	-0.0509 -11.9%	-0.0808 -18.9%
Income Levels	0.0322 7.5%	0.0176 4.1%	Income	0.0462 10.8%	0.0745 17.4%
Other Household Charact	0.0049 1.2%	0.0136 3.2%	Age	0.1634 38.2%	0.2149 50.2%
Network Externalities	0.1592 37.2%	0.2742 64.1%	Other Household	-0.0098 -2.3%	-0.0207 -4.8%
			Network Externalities	-0.0537 -12.5%	-0.2101 -49.1%
			Intercept	0.1223 28.6%	0.1320 30.8%
All Included Variables	0.2105 49.2%	0.3181 74.3%	All Included Parameters	0.2174 50.8%	0.1098 25.7%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in individual groups of characteristics or parameters

Table 27. Individual Contributions of Characteristics and Parameters to the Inter-temporal Decomposition in General Internet Access, 2000 – 2003, Order Reversed

Individual Contributions from: Characteristic Shifts	Equation		Parameter Shifts	Equation	
	(22)	(23)		(22)	(23)
Education Levels	0.0074 5.0%	0.0095 6.4%	Education	-0.0044 -3.0%	-0.0045 -3.0%
Income Levels	0.0099 6.7%	0.0104 7.0%	Income	-0.0083 -5.6%	-0.0085 -5.7%
Other Household Charact	0.0028 1.9%	-0.0088 -6.0%	Age	0.0897 60.6%	0.0875 59.1%
Network Externalities	0.0543 36.7%	0.0708 47.9%	Other Household	0.0446 30.1%	0.0485 32.8%
DCT Infrastructure	0.0039 2.6%	0.0073 4.9%	Network Externalities	-0.0409 -27.6%	-0.0508 -34.3%
			DCT Infrastructure	-0.0018 -1.2%	-0.0045 -3.0%
			Intercept	-0.0092 -6.2%	-0.0088 -5.9%
All Included Variables	0.0783 52.9%	0.0891 60.2%	All Included Parameters	0.0697 47.1%	0.0589 39.8%

Note: Percentages indicate the amount of the inter-temporal gap due to changes in individual groups of characteristics or parameters

Reversing the order of the two inter-temporal decompositions leaves the results virtually unchanged. In fact, the percent of the inter-temporal shift explained by

individual characteristic or parameter shifts never changes by more than 5 percent. Various orderings were used to verify the robustness of these results. Regardless of the ordering used, the results are within one standard deviation of the original results. Hence, the results of the original decomposition can be considered relatively robust to changes in order.

When considering the results of the inter-temporal decomposition, it is important to note that, in general, rates of Internet access have been increasing over time. Meanwhile, the proxy for network externalities is essentially a regional measure of average access rates in a given year. Hence, the network externality measure is, in part, picking up this upward trend in Internet access. The results are therefore likely to overstate the importance of this measure. Intuitively, the question underlying this inter-temporal analysis is, "What is causing the increase in residential Internet access rates – changes in characteristics or changes in parameters?" The specification in equation (22) suggests that the increase is about evenly split between the two, while specification (23) suggests that changes in characteristics accounts for a higher portion.⁶⁹ However, the dominant contributing factor for the characteristic portion is the network externality measure. In fact, Table 4 shows that most other attributes (education and income levels, racial / ethnic composition of households) have not changed significantly over time. The proxy for network externalities, on the other hand, has shown a significant increase between 1997 and 2003.

The measure of network externalities used in this study is very blunt and may not capture the true essence of the externality concept. Ideally, to create a true measure of "network externalities," the actual impact that local access rates have on their neighboring households would have to be removed from other factors associated with these access rates – in particular, levels of DCT infrastructure, underlying education and income factors, and costs of access that are embedded in local access rates. This ideal measure of network externalities would then *not* include any type of association with other factors, and differences in this measure over time could be used in the decomposition procedure

⁶⁹ Note that in Tables 25 and 26, equation (22) attributes 49.2 and 52.9 percent (about half) of the increase to characteristic shifts, while equation (23) attributes a larger portion (74.3 and 60.2 percent) to characteristic shifts.

above. Unfortunately, creating this ideal measure of network externalities is not possible due to a lack of relevant data. Removing the portion of the local access rate not related to “network externalities” would require information specific to individual households pertinent to the concept of "network externalities", such as the percentage of their friends with home access or the amount of interaction they have with their neighbors. These additional variables could be used to separate the actual "network externalities" from their association with other variables.

This shortcoming in the network externality measure tempers the conclusions that can be drawn from the inter-temporal decomposition results. Hence, the remaining discussion focuses on the impact of various shifting parameters. The most surprising result is the dramatic impact of the age parameter. In the earlier years of the analysis, the quadratic form approximated by the age parameters turned negative at a relatively early age (31 in 1997), while in 2003 the quadratic form was still positive for 65-year old household heads. The dramatic shift in the age profile of Internet users (Figure 17) is indicative of the changes that took place over this period, as households headed by older individuals made up increasingly larger shares of those households with access. Furthermore, the (questionable) importance of higher levels of local access rates noted above is somewhat tempered by a decrease in the relevance of the parameters associated with these rates. Specifically, a high regional rate of local access was much more strongly associated with Internet access in the early years of the analysis than in 2003. Thus, as local rates of access have been climbing over the years, the importance associated with regional variations in these rates has been declining – perhaps indicating decreasing returns of network externalities or the mitigation of initial inequalities in access. It is worth reiterating that the effect of the parameter shift associated with DCT infrastructure, when included, is minimal. In fact, shifts in the DCT infrastructure parameter account for the smallest change in Internet access over time of any variable included in this analysis.

The results provide varying degrees of support for the hypotheses regarding the adoption and diffusion theories discussed in the previous chapters (and summarized in Table 10). In particular, the “early adopter” hypothesis that education and income should

become less important over time finds mixed support from the inter-temporal decomposition results. For the 1997 – 2003 time period, only education follows this pattern, as the association between education and Internet access weakens over time. The association between income and Internet access actually strengthens over this period. For the 2000 – 2003 time period, both income and education follow the early adopter hypothesis, but the effects of these shifting parameters is minimal when compared to the effects from other parameters, such as those for age and network externalities. The “core to periphery” diffusion hypothesis that rural status should become less important over time is not supported by the data, as the rural term lacks significance in all years of the analysis. Hence, even in the initial adoption phase, rural status was not an important factor in determining Internet access after accounting for differences in household characteristics.

While the importance of income, education, and rural status to Internet access were hypothesized to *decrease* over time, the presence of DCT infrastructure was not hypothesized to play a role for the general Internet access decision. The estimation results provide support for this hypothesis. Further, shifts in the DCT infrastructure parameter account for the smallest change in Internet access over time of any variable included in the analysis. Finally, network externalities were hypothesized to have stable parameters. This hypothesis does not hold, as parameters for network externalities were more important in the early adoption phase, suggesting that wider regional diffusion has taken place.

4.4 - Nested Logit Model Results

The preceding results dealt with a simple yes / no access decision by the household. We now turn to the results for the nested decision process that differentiates between dial-up and high-speed access. Table 28 below displays the results for a nested logit regression including only constant terms and education levels as the independent variables for 2003.⁷⁰ The nature of the nested logit model requires that one type of access be chosen as the “default.” For the purposes of this analysis, dial-up access was selected. *As such, all coefficients for a characteristic group should be interpreted as relative to a*

⁷⁰ Other groups of explanatory variables will be added to the model as this section progresses.

"default household" – that is, one with the default characteristic value and dial-up access. Hence, for the regression involving education levels, this “default household” is one that is headed by an individual with no high school education and that has dial-up access. Furthermore, as detailed in section 3.5, rural parameter shifts for each household characteristic are included in order to decompose the impact of specific characteristics on the rural – urban gap in specific types of Internet access.

Table 28. Nested Logit Results for Education (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	1.6630 ***	-0.9288 **	0.0535 **	-0.6986 **
hs	-1.1620 ***	0.2106	-0.0133	-0.0802
scoll	-1.8816 ***	0.4375 ***	0.0480	0.1049
coll	-2.4821 ***	0.8156 ***	-0.0182	-0.2315 **
collplus	-2.8031 ***	0.8488 ***	0.2033 **	-0.0477 **
IV - no	1			
IV - yes	0.9708 *			
Log-likelihood	-38975.1			

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10$, 0.05 , and 0.01 levels, respectively. For the inclusive value (IV), they indicate a statistically significant difference from one.

The first column in Table 28 presents the parameter values associated with the impact of the characteristic on the probability of a household having *no access* relative to the default household. For instance, the significant negative coefficient associated with high school (hs) indicates that relative to the default household with no high school education, a household headed by someone with a high school education would be *less likely* to have no access than dial-up access. The second column shows the effect of this variable on the probability of *high-speed access* relative to the default household. Therefore, the significant positive coefficient associated with some college (scoll) indicates that relative to the default household, having a household head with some college education *increases* the probability of high-speed access. The third and fourth columns interact rural dummy variables with each characteristic, and report the effects of these “rural shifts” on the probabilities of no access and high-speed access, respectively. There are several significant rural shifts in Table 28, indicating that the impact of education on access

decisions differs between rural and urban areas. For instance, the positive coefficient in column 3 associated with more than a bachelor's degree (collplus) indicates that rural location will diminish the negative impact of this level of education.⁷¹ Similarly, the negative coefficients associated with a bachelor's degree (coll) and more than a bachelor's degree (collplus) in column 4 indicate that rural status lowers the effects of these education levels on high-speed access relative to the education effects in urban areas.

The coefficients for the constant terms can be interpreted as a *ceteris paribus* preference for no access or high-speed access relative to dial-up access. The positive coefficient associated with the no access constant term implies that, for a household headed by an individual with no high school education, the probability of no access is higher than the probability of dial-up access. Alternatively, the negative coefficient on the high-speed constant term implies that the probability of high-speed access is less than the probability of dial-up access for this same household. Significant shifts on these constant terms exist for households in rural areas, indicating that households in rural areas have a higher base propensities for no access and lower base propensities for high-speed access after controlling for education.

The results above are specific to the initial specification for the nested logit model – that is, one that only includes education levels and constant terms as the explanatory variables. To account for the individual impact of other variables, the remainder of this section incrementally adds groups of characteristics to the analysis. One issue of concern is whether the nested logit model is preferred to the multinomial model in each case. Section 3.5 discussed two separate methods for determining which specification is preferred. The first method is a test of the independence of irrelevant alternatives (IIA) assumption. This is essentially a *t*-test of whether or not the inclusive value coefficient equals unity (which it would in a multinomial model). The inclusive value (IV) coefficient indicates the degree of similarity between alternatives in a nest. Note that for the no access decision, the value of this coefficient is necessarily unity, since only a single alternative exists within the nest. The value of the IV coefficient for the “yes” decision deals with the similarity between dial-up and high-speed access. This

⁷¹ It is important to remember that these rural coefficients are *shifts* from the coefficients in the first two columns. Thus, the positive coefficient for collplus under rural - no access informs us that rural status decreases the negative coefficient for collplus in column one.

coefficient equals 0.9708 in Table 28, which is statistically different from unity at the $p = 0.10$ level for this dataset, and thus weakly supports the use of the nested logit model. The second method for testing the nested and multinomial specifications is a comparison of their predictive abilities. Table 29 summarizes the two methods for comparing the multinomial and nested logit model when education levels are included. A similar table will be presented for each of the nested logit specifications (which add explanatory variables to the analysis) in this section.

Table 29. Nested and Multinomial Logit Model Comparison - I (2003)

	Percent Correctly Predicted						IV parameter <i>t</i> -test
	No access		Dial-up access		High-speed access		Parameter
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education</i>							
Nested	60.64	70.13	72.04	59.50	0.00	0.00	0.9708 *
Multinomial	60.64	70.13	72.04	59.50	0.00	0.00	
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: * indicates statistically significant differences from unity at the $p = 0.10$ level.

Although both methods perform equally well in terms of predictions, the fact that the IV parameter is in the unit interval and is statistically different from zero provides some support for the nested logit model. It is interesting that neither model predicts that any households will opt for high-speed access. The nominal amount of explanatory variables (only four education levels are included in the analysis) likely contributes to the poor predictive ability of the models, as well as the similarities in predictive power between them.

Education and Income

In order to perform the decomposition detailed in section 3.5, various groups of independent variables must be added to the nested logit specification. Table 30 summarizes the results when income levels are added to the education levels and constant term.

Table 30. Nested Logit Results for Education and Income (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	2.3914 ***	-0.4905 ***	-0.1230	-1.2776 **
hs	-0.7923 ***	0.1360	0.0126	-0.0542
scoll	-1.3891 ***	0.3228 *	0.1113	0.1466
coll	-1.6597 ***	0.5673 *	0.0117	-0.1208 *
collplus	-1.7710 ***	0.5129 ***	0.2138 **	0.0958
faminc1	0.0609	-0.4276	0.2593	0.9829
faminc2	0.1255	-0.5305	0.1637	0.4931
faminc3	-0.1016	-0.6040	0.2529	0.7378
faminc4	-0.1679	-0.6825	0.0679	0.7625
faminc5	-0.3531 **	-0.7637	0.1025	1.0150
faminc6	-0.5869 ***	-0.6402	0.0247	0.5903
faminc7	-0.6540 ***	-0.4736	-0.2569	0.3263
faminc8	-0.9371 ***	-0.6439	-0.1049	0.6294
faminc9	-1.2733 ***	-0.8180	-0.0132	1.3414
faminc10	-1.5290 ***	-0.5194	-0.0723	0.5380
faminc11	-1.6498 ***	-0.3789	-0.2971	0.4414
faminc12	-2.0810 ***	-0.3325	0.1623	0.7202
faminc13	-2.2807 ***	0.1245 ***	0.0822	0.4150 **
IV - no	1			
IV - yes	1.0633 **			
Log-likelihood	-36086.2			

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10, 0.05, \text{ and } 0.01$ levels, respectively. For the inclusive value (IV), they indicate a statistically significant difference from one.

The “default household” relevant to the above results has a household head with no high school education, an income level of less than \$5,000, and dial-up access. The significant coefficients in the first column indicate that relative to this default household, higher levels of education and income *decrease* the probability of no access. Furthermore, the coefficients in column two indicate that while most higher education levels significantly *increase* the probability of high-speed access relative to dial-up, only the highest level of income (faminc13) is significant. Further, very few rural shifts are significant for this regression. When the household head has more than a bachelor’s degree (collplus), rural status diminishes the negative impact observed for this education level on the likelihood of no access. Similarly, rural status increases the positive impact on high-speed access observed for urban households with yearly income levels over \$75,000 (faminc13),

suggesting that income barriers to high-speed access may be higher in rural areas. The only other characteristic-induced rural shift that is significant is a reduction in the impact that a household head with a bachelor's degree (coll) has on high-speed access.

The constant terms for no access and high-speed access maintain significant positive and negative coefficients, respectively. This once again indicates that for the default household, no access has a higher probability than dial-up access, and dial-up has a higher probability than high-speed, *ceteris paribus*. Additionally, the rural shift in the constant term is significantly negative for high-speed access, revealing that rural status significantly lowers the propensity to have high-speed access relative to dial-up, *ceteris paribus*. Table 31 shows how the nested logit model compares to its multinomial counterpart for this specification. The largest differences in predictive power between the two models occur in the rural dial-up and high-speed access categories. The multinomial model again fails to predict that any rural households will have high-speed access; however, the nested logit model correctly predicts about 14 percent of the 1,300 households in this category. On the other hand, the multinomial model correctly predicts 63 percent of the rural households with dial-up access as opposed to 49 percent for the nested model. The IV parameter is again statistically significantly different from one, although it is outside the unit interval. The fact that the IV parameter is larger than one does not necessarily conflict with the conditions for utility maximization (Borsch-Supan, 1990). As expected, the predictive power of nearly all categories is improved under this new specification. When compared to Table 29, Table 31 has higher percentages correctly predicted for the no access and high-speed access categories. Although the percent correctly predicted for dial-up access fell, this is likely due to the ability of the updated specification to differentiate between dial-up and high-speed, compared to the specification in Table 29 that predicted no household would have high-speed access. Furthermore, because the models underlying Table 28 and Table 30 are hierarchically nested, a likelihood ratio test can be used to determine if the addition of income is significant in explaining the type of access selected. The null hypothesis that the income parameters are equal to zero is rejected at the $p = 0.01$ level.

Table 31. Nested and Multinomial Logit Model Comparison - II (2003)

Percent Correctly Predicted	No access		Dial-up access		High-speed access		IV parameter <i>t</i> -test Parameter
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education + Income</i>							
Nested	73.19	82.93	50.79	48.73	32.05	13.75	1.0633 **
Multinomial	74.13	78.26	49.86	63.14	32.05	0.00	
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: ** indicates statistically significant differences from unity at the $p = 0.05$ level.

Education, Income, and Other Household Characteristics

The next group of variables added to the specification is “other household characteristics” such as number of children, race, sex, and age. The results for this nested logit model are shown in Table 32.

Table 32. Nested Logit Results for Education, Income, and Other Household Characteristics (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	3.2233 ***	0.0863	0.3134 *	-0.6627 **
hs	-0.5959 ***	0.0824	-0.0337	-0.1099
scoll	-1.2168 ***	0.2291	0.0966	0.0697
coll	-1.4273 ***	0.4127 *	-0.0912	-0.1557
collplus	-1.5939 ***	0.4187 **	0.0422 ***	0.0178
faminc1	-0.0534	-0.4418	0.2354	0.8595
faminc2	0.0156	-0.4748	0.2376	0.5232
faminc3	-0.1758	-0.5098	0.1868	0.4808
faminc4	-0.1900	-0.5503	0.0963	0.4697
faminc5	-0.3255	-0.6136	0.1043	0.6455
faminc6	-0.4936 ***	-0.5644	0.1154	0.4414
faminc7	-0.4951 ***	-0.4275	-0.0613	0.2992
faminc8	-0.7472 ***	-0.5752	0.0800	0.5836
faminc9	-0.9783 ***	-0.7188	0.1017	1.1520
faminc10	-1.1886 ***	-0.4616	0.1758	0.4775
faminc11	-1.2577 ***	-0.3275	-0.0446	0.4573
faminc12	-1.5756 ***	-0.2867	0.3135	0.7196
faminc13	-1.6796 ***	0.1876 **	0.1802	0.4327 **
netatwork	-0.4693 ***	0.1337 **	-0.0864 **	0.0261 **
black	0.6345 ***	-0.1809 **	0.1756 **	0.2304
othrace	0.1971	0.1327 *	0.2557	0.5642
hisp	0.6380 ***	-0.0882 **	-0.2191	-0.1172
peage	-0.0639 **	-0.0132	-0.0206	-0.0237
age2	0.0008	0.0000	0.0003	0.0003
sex	0.0290	0.1833	0.0403	-0.1677
married	-0.5228 **	-0.0364	-0.1393	-0.2260
chld1	-0.2640 ***	-0.0145	0.0969	0.3097 **
chld2	-0.3067 ***	-0.0045	0.0435	0.2143 **
chld3	-0.2460 **	-0.1076	0.1175	0.3067
chld4	-0.2998	-0.2500	0.3607	0.0845
chld5	-0.1929	-0.2891	-0.3068	-0.0467
retired	-0.1625 ***	-0.0979 *	-0.1461 **	0.3429
IV - no	1			
IV - yes	0.9565 **			
Log-likelihood	-34624.8			

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10, 0.05,$ and 0.01 levels, respectively. For the inclusive value (IV), they indicate a statistically significant difference from one.

Note that these other household characteristics are slightly different from education and income levels in that most variables are binary as opposed to categorical. Thus, the “default household” does NOT possess these characteristics: the head does *not* have access at work, he / she is *not* Black, *not* another race (other than white), *not* Hispanic,

not male, *not* married, and *not* retired. In this way we can interpret the effect of these characteristics on various types of access.

The education and income results are consistent with the previous specification: higher education and income levels reduce the likelihood of no access, and increase the likelihood of high-speed access. Internet access at work is a significant factor for both no access and high-speed access, with its presence having similar effects as those for education and income (decreasing the probability of no access, increasing the probability of high-speed). The presence of a Black or Hispanic household head increases the probability of no access, and decreases high-speed probability, implying that even after controlling for a multitude of other characteristics, racial and ethnic factors still play a role in the type of access adopted. The presence of a married household head or between one and three children in the household decreases the probability of no access, but interestingly has no significant effect on high-speed access. Finally, a retired household head decreases the probability of both no access and high-speed access.

The number of rural parameter shifts that are significant actually increases for this specification. Internet access at work is particularly important for rural households, as the rural interaction term reinforces the general effects of this variable (decreasing the probability of no access, increasing the probability of high-speed access). The positive rural interaction for Black household heads term fortifies the positive relationship between urban households with this characteristic and no access. Interestingly, the shift from having one or two children in the household is significantly positive in determining high-speed access for rural households, even though this shift is from a non-significant original coefficient in column 2. Rural status for a retired household head further decreases the probability of no access, but is not significant for high-speed access. The only constant term that is significant in this specification is for no access, indicating that other factors have accounted for some of the default *ceteris paribus* preference for dial-up access over high-speed access that existed in the previous specifications. However, rural shifts to these constant terms are significant for both no access and high-speed access. These significant constant interaction terms indicate a remaining decreased propensity for high-speed and an increased propensity for no access in rural areas. This is different than the model for general access (Table 18), where the rural shift was not

significant. As Table 33 indicates, the nested logit specification outperforms the multinomial in terms of high-speed predictions for both rural and urban households. The IV parameter is significantly different from one, lending additional support for the nested specification. The introduction of other household characteristics to the specification did not dramatically change the percent correctly predicted for either model, although the rural dial-up access predictions did experience an increase. A likelihood ratio test between the models displayed in Table 30 and Table 32 reveals that the inclusion of other household characteristics is warranted, as the null hypothesis that all other household characteristic parameters equal zero is rejected at the $p = 0.01$ level.

Table 33. Nested and Multinomial Logit Model Comparison - III (2003)

Percent Correctly Predicted	No access		Dial-up access		High-speed access		IV parameter <i>t</i> -test Parameter
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education + Income + Other</i>							
Nested	73.54	79.46	54.57	64.22	32.82	13.75	0.9565 **
Multinomial	73.58	79.42	53.62	64.77	28.67	1.19	
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: ** indicates statistically significant differences from unity at the $p = 0.05$ level.

Education, Income, Other Household Characteristics, and Network Externalities

To capture the effect of network externalities, local rates of access are added to the nested logit specification. Each household has a rate for the three feasible types of access – none, dial-up, and high-speed. A summary of these rates by rural – urban status within a state is available in Appendix C. Because the measure for network externalities is continuous, the resulting coefficients for this variable can be interpreted as shifts in the probability of type j access (relative to dial-up access) when the local rate of type j access increases.

Table 34. Nested Logit Results for Education, Income, Other Household Characteristics, and Network Externalities (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	0.9501 **	-0.7353	0.3572	-1.2335
hs	-0.6333 ***	0.0425	0.0515	0.0137
scoll	-1.3320 ***	0.1963	0.1578	0.1759
coll	-1.6717 ***	0.3802 *	0.1301	-0.0311
collplus	-1.8401 ***	0.3935 **	0.1846	0.1194
faminc1	0.2804	-0.3924	-0.3403	1.0813
faminc2	0.3915	-0.4097	-0.2966	0.5624
faminc3	0.1988	-0.4468	-0.2992	0.9074
faminc4	0.2112	-0.5054	-0.4394	1.0235
faminc5	0.0825	-0.4885	-0.4206	1.0528
faminc6	-0.1209	-0.3726	-0.3351	0.7114
faminc7	-0.1828	-0.2226	-0.4421	0.4409
faminc8	-0.3504 ***	-0.3697	-0.4413	0.7854
faminc9	-0.5109 ***	-0.5322	-0.6147	1.4179
faminc10	-0.8468 ***	-0.2668	-0.2584	0.6519
faminc11	-0.9854 ***	-0.1260	-0.5068	0.6591
faminc12	-1.3349 ***	-0.1015	-0.1848	0.9227
faminc13	-1.7786 ***	0.3519 **	-0.0667	0.6418 **
netatwork	-0.5647 ***	0.1355 **	-0.0556 *	0.0058 **
black	0.7194 ***	-0.1804 *	-0.1192	0.3184
othrace	0.1316	0.0866	0.0429	0.6098
hispanic	0.7164 ***	-0.1149 *	-0.2645	-0.0729
peage	-0.0537 **	-0.0053	-0.0154	-0.0180
age2	0.0008	-0.0001	0.0002	0.0003
sex	-0.0813	0.1831	0.1661	-0.2184
married	-0.5038 **	-0.0351	-0.1009	-0.1984
chld1	-0.2578 ***	-0.0266	-0.0018	0.3861
chld2	-0.3013 ***	-0.0328	-0.0099	0.3051
chld3	-0.1797 **	-0.1428	0.0365	0.4062
chld4	-0.2247	-0.2991	0.5847	0.2673
chld5	-0.1157	-0.2370	-0.0925	-0.1892
retired	-0.1214 ***	-0.0810	-0.2978	0.3574
rate	2.4722 ***	2.2696 ***	0.4868 *	2.7628 ***
IV - no	1			
IV - yes	0.7992 ***			
Log-likelihood	-17704.5			

The majority of the variables included in the previous analysis maintain their significance and sign; however, the inclusion of these regional rates of access removes almost all other significant rural shifts! That is, the inclusion of these regional rates of access seems to explain a large portion of the access-specific digital divides. The positive coefficients

for “rate” in all four columns indicate that as the local rate of a particular type of access increases, the probability of a household having that type of access increases as well (relative to the probability of dial-up access assumed for the default household). The significant coefficients in the third and fourth columns imply that rates have a different impact on rural households than they do on urban households. The fact that these rural coefficient shifts are positive implies that rural areas may be particularly prone to clusters of adoption, with network externalities possibly having a larger impact in rural areas than they do in urban areas.

As indicated above, perhaps the most noticeable difference in the results when network externalities are included is the dramatic reduction in the number of “rural shift” parameter estimates that are significant. While shifts for collplus, chld1, chld2, Black, and retired were all significant in Table 32; they lose their significance once the “rate” variable is included. This dramatic reduction implies that rural differences in these factors may be associated with regional variations in rates.

As Table 35 shows, the nested logit specification greatly outperforms its multinomial counterpart when the proxy for network externalities is included. In particular, the dial-up predictions under the nested model are roughly 20 percentage points higher than the multinomial model, and the nested logit high-speed predictions are much higher as well. A significant increase in the percent correctly predicted is observed once the network externalities variable is introduced to the analysis, and the dramatic increase in the log likelihood between Table 32 and Table 34 indicates that network externalities should be included in the specification (likelihood ratio test is significant at the $p = 0.01$ level).

Table 35. Nested and Multinomial Logit Model Comparison - IV (2003)

Percent Correctly Predicted	No access		Dial-up access		High-speed access		IV parameter <i>t</i> -test Parameter
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education + Income + Other + Network Externalities</i>							
Nested	74.56	91.34	71.59	88.29	50.72	15.75	0.7992 ***
Multinomial	73.47	79.65	53.65	65.26	35.20	2.23	
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: *** indicates statistically significant differences from unity at the $p = 0.01$ level.

Income, Education, Other Household Characteristics, Network Externalities, and DCT Infrastructure

The final variable added to the analysis is a measure of DCT infrastructure availability for the household. This measure essentially captures the percentages of area households (summarized by rural – urban status within a state) that have DSL and cable Internet access available to them. Section 3.1 discusses the data and methodology used to derive this measure, and a summary of the results is available in Appendix B. The resulting coefficients (shown in Table 36) indicate the relative shift in the probability of no access or high-speed access for an increase in the rate of DCT infrastructure availability.

Table 36. Nested Logit Results for Education, Income, Other Household Characteristics, Network Externalities, and DCT Infrastructure (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	1.0559 **	-1.1814	0.0771	-0.7472
hs	-0.6386 ***	0.0526	0.0503	0.0050
scoll	-1.3375 ***	0.2028	0.1515	0.1731
coll	-1.6842 ***	0.3871 *	0.1335	-0.0376
collplus	-1.8546 ***	0.4027 **	0.1853	0.1080
faminc1	0.2913	-0.3959	-0.3608	1.0742
faminc2	0.4053	-0.4137	-0.3092	0.5577
faminc3	0.2085	-0.4363	-0.3117	0.8844
faminc4	0.2256	-0.5061	-0.4576	1.0108
faminc5	0.0957	-0.4796	-0.4391	1.0317
faminc6	-0.1136	-0.3570	-0.3463	0.6927
faminc7	-0.1795	-0.2086	-0.4451	0.4246
faminc8	-0.3407 ***	-0.3615	-0.4537	0.7753
faminc9	-0.4997 ***	-0.5148	-0.6360	1.3911
faminc10	-0.8418 ***	-0.2525	-0.2627	0.6280
faminc11	-0.9836 ***	-0.1142	-0.5125	0.6383
faminc12	-1.3348 ***	-0.0872	-0.1925	0.8989
faminc13	-1.7958 ***	0.3647 **	-0.0634	0.6222
netatwork	-0.5698 ***	0.1357 **	-0.0558	0.0088 *
black	0.7308 ***	-0.1901 *	-0.1170	0.3095
othrace	0.1435	0.0680	0.0162	0.6222
hisp	0.7381 ***	-0.1386 *	-0.2782	-0.0619
peage	-0.0538 **	-0.0043 *	-0.0149	-0.0187
age2	0.0008	-0.0001	0.0002	0.0003
sex	-0.0882	0.1855	0.1726	-0.2229
married	-0.5024 **	-0.0349	-0.0991	-0.1952
chld1	-0.2558 ***	-0.0279	-0.0107	0.3900
chld2	-0.3010 ***	-0.0305	-0.0167	0.3049
chld3	-0.1757 **	-0.1397	0.0225	0.4133
chld4	-0.2211	-0.2889	0.5778	0.2635
chld5	-0.1249	-0.2058	-0.0894	-0.2093
retired	-0.1223 ***	-0.0789	-0.3006	0.3487
rate	2.7352 ***	2.2806 ***	0.5027 *	2.8191 **
dslaccess	-0.1544	0.2099	0.0402	-0.1538
cableaccess	-0.2525	0.4210	0.3713	-0.5937
IV - no	1			
IV - yes	0.8920 ***			
Log-likelihood	-17632.0			

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10$, 0.05 , and 0.01 levels, respectively. For the inclusive value (IV), they indicate a statistically significant difference from one.

The parameter estimates for DCT infrastructure (both DSL and cable) are not significant in the regression. In particular, the high-speed parameters associated with these variables

lack significance for both rural and urban areas. This is noteworthy due to the hypothesized importance of DCT infrastructure to the high-speed access decision. Further, the addition of these variables results in very few changes from the previous model. All variables that were significant in the model run without DCT infrastructure (Table 34) maintain the same level of significance in this updated regression (Table 36). Although a likelihood ratio test indicates that the DCT infrastructure parameters should not be restricted to zero, the lack of any significant changes in the model results implies that the added variables (DCT infrastructure) are not important factors in the context of this nested decision-making process. This finding is reinforced by the predictions shown in Table 37, which barely changed from those shown in Table 35.

Table 37. Nested and Multinomial Logit Model Comparison - V (2003)

Percent Correctly Predicted							IV parameter <i>t</i> -test
	No access		Dial-up access		High-speed access		Parameter
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education + Income + Other + Network Externalities + DCT Infrastructure</i>							
Nested	74.65	91.55	71.89	88.15	51.29	15.75	
Multinomial	73.68	81.26	53.98	64.59	34.85	2.58	0.8920 ***
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: *** indicates statistically significant differences from unity at the $p = 0.01$ level.

Table 38 summarizes the variables included in each model discussed in this section, along with relevant IV parameter *t*-tests and predictive capabilities. The results support the choice of the nested logit specification. In each case the IV parameter is significant and the multinomial specification is outperformed in terms of predictive ability. Furthermore, the impact of adding variables to the analysis can be observed in terms of increased predictive power between various specifications, as well as through the results of the likelihood ratio tests discussed above.

Table 38. Nested and Multinomial Logit Model Comparison - Summary (2003)

	Percent Correctly Predicted						IV parameter <i>t</i> -test Parameter
	No access		Dial-up access		High-speed access		
	Urban	Rural	Urban	Rural	Urban	Rural	
<i>Education</i>							
Nested	60.64	70.13	72.04	59.50	0.00	0.00	0.9708 *
Multinomial	60.64	70.13	72.04	59.50	0.00	0.00	
<i>Education + Income</i>							
Nested	73.19	82.93	50.79	48.73	32.05	13.75	1.0633 **
Multinomial	74.13	78.26	49.86	63.14	32.05	0.00	
<i>Education + Income + Other</i>							
Nested	73.54	79.46	54.57	64.22	32.82	13.75	0.9565 **
Multinomial	73.58	79.42	53.62	64.77	28.67	1.19	
<i>Education + Income + Other + Network Externalities</i>							
Nested	74.56	91.34	71.59	88.29	50.72	15.75	0.7992 ***
Multinomial	73.47	79.65	53.65	65.26	35.20	2.23	
<i>Education + Income + Other + Network Externalities + DCT Infrastructure</i>							
Nested	74.65	91.55	71.89	88.15	51.29	15.75	
Multinomial	73.68	81.26	53.98	64.59	34.85	2.58	0.8920 ***
# Observations	11,385	4,998	10,932	4,033	7,524	1,300	

Note: *, **, and *** indicate statistically significant differences from unity at the $\rho = 0.10, 0.05,$ and 0.01 levels, respectively.

4.5 - Decomposition of the Nested Logit Model

The previous section discussed the significance of coefficients resulting from various nested logit runs. These runs provided some ideas on which characteristics are important in determining the magnitude of the digital divide with respect to the two types of access. In order to explicitly test the importance of these characteristics, a decomposition of the nested logit model is performed as described in section 3.5. This technique determines a synthetic type j access rate (denoted \hat{P}_{ij}^0) by assigning urban parameter vectors to rural households. Hence, \hat{P}_{ij}^0 is the average probability of type j Internet access for rural households when urban parameter vectors are used (this is more clearly demonstrated in equation (30)). This methodology breaks out the total rural – urban difference in type j access rates into a component associated with differences in characteristics ($\hat{P}_{ij} - \hat{P}_{ij}^0$) and a component associated with differences in parameters

$(\hat{P}_{ij}^0 - \hat{P}_{ij})$. By altering the variables included in the analysis, the impact of various characteristics can be inferred. The technique is presented in detail in section 3.5.

Table 39 below indicates the results of this decomposition for the years 2003, 2001, and 2000. The first two lines of Table 39 show the urban (\hat{P}_{uj}) and rural (\hat{P}_{rj}) average rates of type j access for a given year, with the third line showing the “digital divide” for the relevant type of access. One group of explanatory variables is introduced at a time, starting with education levels. The decomposition is performed by running a nested logit regression using the relevant group of explanatory variables on ONLY urban households. The resulting parameter vector is then applied to rural households to create synthetic type j access rates (\hat{P}_{rj}^0). These rates are displayed for each combination of explanatory variables used, along with the total percentage of the type j gap that is accounted for by rural-urban differences in the explanatory variables.

Table 39. Nested Logit Decomposition Results

Rates of Access	2003			2001			2000		
	$j = 0$	$j = 1$	$j = 2$	$j = 0$	$j = 1$	$j = 2$	$j = 0$	$j = 1$	$j = 2$
Urban \hat{P}_{uj}	0.3877	0.3623	0.2500	0.4343	0.4475	0.1181	0.5324	0.4142	0.0533
Rural \hat{P}_{rj}^{syn}	0.5161	0.3718	0.1122	0.5706	0.3864	0.0430	0.6706	0.3058	0.0236
Delta ($\hat{P}_{uj} - \hat{P}_{rj}$)	-0.1284	-0.0095	0.1378	-0.1363	0.0611	0.0751	-0.1382	0.1084	0.0298
Explanatory Variables									
<i>Education</i>									
\hat{P}_{rj}^0	0.4593	0.3339	0.2068	0.5054	0.3972	0.0974	0.6262	0.3345	0.0432
% Explained ($\hat{P}_{uj} - \hat{P}_{rj}^0$)	56%	-299%	31%	52%	82%	28%	68%	74%	34%
Remainder ($\hat{P}_{rj}^0 - \hat{P}_{rj}$)	44%	399%	69%	48%	18%	72%	32%	26%	66%
<i>Education + Income</i>									
\hat{P}_{rj}^0	0.4985	0.3194	0.1821	0.5510	0.3628	0.0862	0.6456	0.3165	0.0378
% Explained ($\hat{P}_{uj} - \hat{P}_{rj}^0$)	86%	-452%	49%	86%	139%	43%	82%	90%	52%
Remainder ($\hat{P}_{rj}^0 - \hat{P}_{rj}$)	14%	552%	51%	14%	-39%	57%	18%	10%	48%
<i>Education + Income + Other HH Characteristics</i>									
\hat{P}_{rj}^0	0.4757	0.3372	0.1871	0.5249	0.3874	0.0877	0.6210	0.3408	0.0382
% Explained ($\hat{P}_{uj} - \hat{P}_{rj}^0$)	69%	-264%	46%	66%	98%	40%	64%	68%	51%
Remainder ($\hat{P}_{rj}^0 - \hat{P}_{rj}$)	31%	364%	54%	34%	2%	60%	36%	32%	49%
<i>Education + Income + Other HH Characteristics + Network Externalities</i>									
\hat{P}_{rj}^0	0.5174	0.3379	0.1447	0.5715	0.3604	0.0681	0.6798	0.2940	0.0216
% Explained ($\hat{P}_{uj} - \hat{P}_{rj}^0$)	101%	-257%	76%	101%	143%	67%	107%	111%	107%
Remainder ($\hat{P}_{rj}^0 - \hat{P}_{rj}$)	-1%	357%	24%	-1%	-43%	33%	-7%	-11%	-7%
<i>Education + Income + Other HH Characteristics + Network Externalities + DCT Infrastructure</i>									
\hat{P}_{rj}^0	0.5316	0.3401	0.1284	0.5731	0.3585	0.0682	0.6855	0.2862	0.0283
% Explained ($\hat{P}_{uj} - \hat{P}_{rj}^0$)	112%	-234%	88%	102%	146%	66%	111%	118%	84%
Remainder ($\hat{P}_{rj}^0 - \hat{P}_{rj}$)	-12%	334%	12%	-2%	-46%	34%	-11%	-18%	16%

Note: Percentages indicate the contribution of the regressed group of variables to the rural - urban gap for each type of access

Note that in 2003, the rates of dial-up access are extremely similar between rural and urban households. Looking at similar rates for 2001 and 2000, it is apparent that the dial-up gap has been shrinking while the high-speed gap has been growing. This is an important point – with the increasing prevalence of high-speed access, the digital divide is becoming directly linked to such access. The similar dial-up rates between rural and urban areas in 2003 result in volatile percentages explained when the decomposition is performed. Since significant dial-up gaps exist in 2001 and 2000, these decomposition results are easier to interpret.

As more variables are added to the analysis, the percentage of the type *j* gap explained by the included variables typically becomes larger (Table 39). This is intuitive, because the inclusion of more explanatory variables captures the effects that rural – urban differences in these variables have on the likelihood of type *j* access. For instance, the initial decomposition focused only on the differences in education levels between rural and urban households. Accounting for these education differences explains 56 percent of the no access gap and 31 percent of the high-speed gap in 2003. Once differences in income levels were added to the analysis, Table 39 indicates that 86 percent of the no access gap and 49 percent of the high-speed gap were explained. These dramatic increases imply that income differences between rural and urban households are an important part of the gap in various access rates.⁷² Similar results are seen in the rural – urban gaps from 2000 and 2001, with differences in education levels consistently accounting for 50 – 70 percent of the no access divide, 70 – 80 percent of the dial-up divide, and around 30 percent of the high-speed divide.⁷³ Once income differences are added to the analysis, the decomposition accounts for 80 – 90 percent of the no access divide, 90 – over 100 percent of the dial-up divide, and 40 – 50 percent of the high-speed divide.

The inclusion of differences in other household characteristics actually decreases the percentage of each type of gap explained, but this decrease is expected. In general,

⁷² It is important to note, however, that the increase in the percentage explained when a variable group is *not* explicitly due solely to that variable group (due to the non-linearity of the nested logit model). This is further discussed in section 3.5.

⁷³ In 2000 and 2001 the rural – urban gaps in dial-up access are 11 and 6 percent, respectively. These are significantly larger than the less than 1 percent gap seen in 2003, allowing for easier interpretation of the decomposition results.

characteristics in this category that lead to higher rates of Internet access (such as having a White household head, being married, or having at least one child) are disproportionately found in *rural* households (see Table 4 for a comparison of average values for these variables between rural and urban households). Hence, including these characteristics will tend to increase the synthetic rates (\hat{P}_{rj}^0), which in turn will shrink the amount of the rural – urban gap explained. This is why a decrease is found in each sample year in the percentage explained when other household characteristics are included in the analysis.

Perhaps the most dramatic increase in the percentage explained for each type of access occurs when the measures of network externalities are included. In each year, the percentage of each type of access explained increases by approximately 30 – 50 percentage points after the inclusion of network externalities. This dramatic increase provides additional evidence that the likelihood of type *j* access for an individual household is affected by regional variations in access rates. On the other hand, the inclusion of differences in DCT infrastructure increases the percentage of the high-speed gap explained by less than 12 percentage points in each of the three years included in the analysis. This increase is small compared to other changes (such as the inclusion of education or network externalities) and is not based on statistically significant parameter shifts for rural areas (Table 36). Similarly, small increases (or even reductions) in the percent correctly predicted after including the DCT infrastructure variables for all three type *j* divides are seen in 2001 and 2000.

Ordering of Variables

Similar to the general (non-nested) results discussed in section 4.2, the order in which the variables enter the analysis is important. To account for this, Table 40 presents the results when the order in which the variables enter the analysis is reversed.

Table 40. Nested Logit Decomposition Results (Order Reversed)

		2003			2001			2000		
		<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2	<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2	<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2
Rates of Access		None	Dialup	Highspeed	None	Dialup	Highspeed	None	Dialup	Highspeed
Urban	\hat{P}_{uj}	0.3877	0.3623	0.2500	0.4343	0.4475	0.1181	0.5324	0.4142	0.0533
Rural	\hat{P}_{rj}	0.5161	0.3718	0.1122	0.5706	0.3864	0.0430	0.6706	0.3058	0.0236
Delta	$(\hat{P}_{uj} - \hat{P}_{rj})$	-0.1284	-0.0095	0.1378	-0.1363	0.0611	0.0751	-0.1382	0.1084	0.0298
Explanatory Variables										
<i>DCT Infrastructure</i>										
	\hat{P}_{rj}^0	0.3985	0.3999	0.2417	0.4405	0.4447	0.1148	0.5434	0.4058	0.0508
% Explained	$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	8%	396%	6%	5%	5%	4%	8%	8%	8%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$	92%	-296%	94%	95%	95%	96%	92%	92%	92%
<i>DCT Infrastructure + Network Externalities</i>										
	\hat{P}_{rj}^0	0.4585	0.3331	0.1875	0.4850	0.4018	0.0752	0.6004	0.3459	0.0428
% Explained	$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	55%	-307%	45%	37%	75%	57%	49%	63%	35%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$	45%	407%	55%	63%	25%	43%	51%	37%	65%
<i>DCT Infrastructure + Network Externalities + Other HH Characteristics</i>										
	\hat{P}_{rj}^0	0.4492	0.3405	0.1957	0.4805	0.4106	0.0852	0.5924	0.3511	0.0435
% Explained	$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	48%	-229%	39%	34%	60%	44%	43%	58%	33%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$	52%	329%	61%	66%	40%	56%	57%	42%	67%
<i>DCT Infrastructure + Network Externalities + Other HH Characteristics + Income</i>										
	\hat{P}_{rj}^0	0.5014	0.3383	0.1502	0.5685	0.3785	0.0724	0.6739	0.2988	0.0285
% Explained	$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	89%	-253%	72%	98%	113%	61%	102%	106%	83%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$	11%	353%	28%	2%	-13%	39%	-2%	-6%	17%
<i>DCT Infrastructure + Network Externalities + Other HH Characteristics + Income + Education</i>										
	\hat{P}_{rj}^0	0.5316	0.3401	0.1284	0.5731	0.3585	0.0682	0.6855	0.2862	0.0283
% Explained	$(\hat{P}_{uj} - \hat{P}_{rj}^0)$	112%	-234%	88%	102%	146%	66%	111%	118%	84%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj})$	-12%	334%	12%	-2%	-46%	34%	-11%	-18%	16%

Note: Percentages indicate the contribution of the regressed group of variables to the rural - urban gap for each type of access

Under this new ordering, the first variable to enter the decomposition is DCT infrastructure. Replacing urban levels of DCT infrastructure with those for rural areas explains between 4 and 8 percent of the divides in no access and high-speed access for the three years under analysis.⁷⁴ Consistent with the results under the initial ordering, one of the largest jumps in the percentage explained occurs when network externalities are added. In particular, for the no access (*j*=0) divide, the percentage explained increases by between 32 and 47 percentage points after the inclusion of network externalities. Similarly, the percentage of the high-speed (*j*=2) divide explained jumps by 27 – 53 percent in each year. Additionally, the percentage of the dial-up (*j*=1) divide explained increases by 70 percentage points in 2001 and by 58 percentage points in 2000. Thus, varying the order in which network externalities are introduced to the analysis does not

⁷⁴ Recall that the dial-up rates for 2003 are remarkably similar for rural and urban areas, leading to considerable volatility in the percentage of the gap explained.

change the dramatic impact network externalities have on the percentage of the gap explained. The same is true for the insertion of other household characteristics and income levels. Similar to the results displayed under the initial ordering, including other household characteristics reduces the percentage of the gap explained of all three types of divides. Likewise, the results from adding income under this ordering are analogous to those obtained under the initial ordering. When income is included, the percentage of the no access divide that is explained by differences in rural and urban variables increases by between 51 and 64 percent. Similarly, the percentage of the dial-up divide explained increases by around 50 percent and the percentage of the high-speed divide that is explained increases by between 17 and 50 percent. Thus far, introducing network externalities, other household characteristics, and income has resulted in similar outcomes under both orderings. This is not the case for education. While the initial decomposition results suggested that differences in education between rural and urban areas accounted for over 50 percent of each of the three types of divides, the inclusion of education under the reordering alters the no access and high-speed divides by less than 16 percent in all years.

Thus, changing the order that the characteristics enter the analysis does have an effect on the magnitude of the resulting percentages of the rural – urban gap explained. This "ordering effect" is particularly notable in the reduced role of education differences and the increased role of DCT infrastructure differences (for 2000 and 2001) under the reordering. However, the reordering had little effect on the increase in the percentage of the gap explained when income and network externalities were introduced into the analysis. Accounting for income and network externality differences between rural and urban areas consistently had large impacts on all three types of digital divides. A number of other reorderings were attempted in an effort to verify this result, including allowing each group of variables to be the first one entered in the analysis. These results (from entering only a single group of variables) are displayed in Table 41 below.

Table 41. Nested Logit Decomposition Results (Single Explanatory Variables)

	2003			2001			2000		
	<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2	<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2	<i>j</i> = 0	<i>j</i> = 1	<i>j</i> = 2
Rates of Access	None	Dialup	Highspeed	None	Dialup	Highspeed	None	Dialup	Highspeed
Urban \hat{P}_{uj}^0	0.3877	0.3623	0.2500	0.4343	0.4475	0.1181	0.5324	0.4142	0.0533
Rural \hat{P}_{rj}^0	0.5161	0.3718	0.1122	0.5706	0.3864	0.0430	0.6706	0.3058	0.0236
Delta $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	-0.1284	-0.0095	0.1378	-0.1363	0.0611	0.0751	-0.1382	0.1084	0.0298
Explanatory Variables									
<i>Education</i>									
\hat{P}_{rj}^0	0.4593	0.3339	0.2068	0.5054	0.3972	0.0974	0.6262	0.3345	0.0432
% Explained $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	56%	-299%	31%	52%	82%	28%	68%	74%	34%
Remainder $(\hat{P}_{rj}^0 - \hat{P}_{rj}^0)$	44%	399%	69%	48%	18%	72%	32%	26%	66%
<i>Income</i>									
\hat{P}_{rj}^0	0.4792	0.3271	0.1937	0.5337	0.3764	0.0899	0.6031	0.3536	0.0432
% Explained $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	71%	-371%	41%	73%	116%	38%	51%	56%	34%
Remainder $(\hat{P}_{rj}^0 - \hat{P}_{rj}^0)$	29%	471%	59%	27%	-16%	62%	49%	44%	66%
<i>Other HH Characteristics</i>									
\hat{P}_{rj}^0	0.3999	0.3642	0.2359	0.4502	0.4384	0.1113	0.5370	0.4123	0.0507
% Explained $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	10%	20%	10%	12%	15%	9%	3%	2%	9%
Remainder $(\hat{P}_{rj}^0 - \hat{P}_{rj}^0)$	90%	80%	90%	88%	85%	91%	97%	98%	91%
<i>Network Externalities</i>									
\hat{P}_{rj}^0	0.4987	0.3213	0.1589	0.5598	0.3810	0.0751	0.6638	0.3055	0.0306
% Explained $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	86%	-431%	66%	92%	109%	57%	95%	100%	76%
Remainder $(\hat{P}_{rj}^0 - \hat{P}_{rj}^0)$	14%	531%	34%	8%	-9%	43%	5%	0%	24%
<i>DCT Infrastructure</i>									
\hat{P}_{rj}^0	0.3985	0.3999	0.2417	0.4405	0.4447	0.1148	0.5434	0.4058	0.0508
% Explained $(\hat{P}_{uj}^0 - \hat{P}_{rj}^0)$	8%	396%	6%	5%	5%	4%	8%	8%	8%
Remainder $(\hat{P}_{rj}^0 - \hat{P}_{rj}^0)$	92%	-296%	94%	95%	95%	96%	92%	92%	92%

Note: Percentages indicate the contribution of the regressed group of variables to the rural - urban gap for each type of access

In general, these results reinforce the finding that levels of income and network externalities explain the largest percentage of the rural – urban digital divide. However, the large percentages of the no access and dial-up divides attributed to differences in network externalities may be cause for concern. In fact, replacing rural levels of network externalities with those for urban areas explains over 85 percent of the no access divide in each year, and over 100 percent of the dial-up divide in 2000 and 2001. The rates for the high-speed divide (between 57 and 76 percent) also seem to be on the high side. This is not the first time that the regional access rate proxy for network externalities has yielded questionable results. Recall that under the general inter-temporal decomposition discussed in section 4.3, the blunt measure used for access rates likely overstates the importance of this variable. A similar exaggeration may be occurring here. Since the proxy for network externalities is simply a measure of the no access / dial-up / high-speed rates for a given aggregate area, regressing a household's access decision on this variable captures a large component of the variables underlying those rates. For example, an area

with a large percentage of high-speed users likely has relatively high education and income levels. Those variables are then inherent to the measure of network externalities. Similarly, local rates of high-speed access may indicate some measure of DCT infrastructure capacity, and the proxy for network externalities may be capturing some of the effects of such capacity. Correcting this problem requires controlling for these other variables, which is done in the final specification when all variables are included. Therefore, the results obtained when the proxy for network externalities is used as the sole dependent variable are likely misleading. Table 42 displays the results of a logit model when the measure for network externalities is excluded for 2003, and Table 43 compares decomposition results for this specification to others included above. In general, the percentage of the various gaps explained drops significantly when the network externality term is removed.

Table 42. Nested Logit Results for Education, Income, Other Household Characteristics, and DCT Infrastructure (No Network Externalities) (2003)

Variables	Urban		Rural	
	None	Highspeed	None	Highspeed
constant	2.9175 **	-0.2790	0.5349 *	-0.3503 **
hs	-0.6010 ***	0.0709	-0.0078	-0.1164
scoll	-1.2323 ***	0.2193	0.1244	0.0644
coll	-1.4439 ***	0.3995 *	-0.0615	-0.1688
collplus	-1.6084 ***	0.4085 **	0.0677 **	0.0088
faminc1	-0.0446	-0.4325	0.2530	0.8443
faminc2	0.0217	-0.5109	0.1994	0.5116
faminc3	-0.1669	-0.5118	0.1907	0.4934
faminc4	-0.1847	-0.5750	0.0827	0.4925
faminc5	-0.3217	-0.6386	0.1257	0.7170
faminc6	-0.4936	-0.5743	0.1431	0.4404
faminc7	-0.4794	-0.4216	-0.0808	0.2628
faminc8	-0.7289 ***	-0.5747	0.0511	0.5825
faminc9	-0.9740 ***	-0.7378	0.1319	1.2265
faminc10	-1.1769 ***	-0.4634	0.1679	0.4782
faminc11	-1.2367 ***	-0.3228	-0.1003	0.4278
faminc12	-1.5721 ***	-0.2913	0.3212	0.7353
faminc13	-1.6832 ***	0.1809 **	0.1897	0.4570 **
netatwork	-0.4684 ***	0.1365 **	-0.1012 **	0.0071 *
black	0.6330 ***	-0.1938 *	0.0943 *	0.3186
othrace	0.1854	0.1129	0.2850	0.6599
hisp	0.6310 ***	-0.1075 *	-0.2294	-0.0882
peage	-0.0626 **	-0.0139 *	-0.0225	-0.0160
age2	0.0008	0.0000	0.0003	0.0002
sex	0.0333	0.1858	0.0345	-0.1852
married	-0.5200 **	-0.0385	-0.1392	-0.2298
chld1	-0.2701 ***	-0.0167	0.1231	0.3458 **
chld2	-0.3056 ***	0.0007	0.0465	0.1947 **
chld3	-0.2488 **	-0.1069	0.1373	0.3111
chld4	-0.3007	-0.2333	0.4329	0.0919
chld5	-0.1983	-0.2502	-0.2893	-0.0924
retired	-0.1444 ***	-0.0929	-0.2019	0.3461
dslaccess	0.0950	0.1285	0.1659	-0.3500
cableaccess	0.2840	0.4684	-0.3846 **	-0.4959
IV - no	1			
IV - yes	0.9025 ***			
Log-likelihood	-34588.4			

Note: *, **, and *** indicate statistically significant differences from zero at the $p = 0.10$, 0.05 , and 0.01 levels, respectively. For the inclusive value (IV), they indicate a statistically significant difference from one.

Table 43. Nested Logit Decomposition Results (Comparison with Model Excluding Network Externalities)

Rates of Access		2003			2001			2000		
		$j = 0$	$j = 1$	$j = 2$	$j = 0$	$j = 1$	$j = 2$	$j = 0$	$j = 1$	$j = 2$
Urban	\hat{P}_{uj}^j	0.3877	0.3623	0.2500	0.4343	0.4475	0.1181	0.5324	0.4142	0.0533
Rural	\hat{P}_{rj}^j	0.5161	0.3718	0.1122	0.5706	0.3864	0.0430	0.6706	0.3058	0.0236
Delta	$(\hat{P}_{uj}^j - \hat{P}_{rj}^j)$	-0.1284	-0.0095	0.1378	-0.1363	0.0611	0.0751	-0.1382	0.1084	0.0298
<i>Education + Income + Other HH Characteristics</i>										
	\hat{P}_{rj}^0	0.4757	0.3372	0.1871	0.5249	0.3874	0.0877	0.6210	0.3408	0.0382
% Explained	$(\hat{P}_{uj}^j - \hat{P}_{rj}^0)$	69%	-264%	46%	66%	98%	40%	64%	68%	51%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj}^j)$	31%	364%	54%	34%	2%	60%	36%	32%	49%
<i>Education + Income + Other HH Characteristics + DCT Infrastructure (No Network Externalities)</i>										
	\hat{P}_{rj}^0	0.4662	0.3607	0.1731	0.5148	0.3751	0.0798	0.6097	0.3198	0.0360
% Explained	$(\hat{P}_{uj}^j - \hat{P}_{rj}^0)$	61%	-17%	56%	59%	118%	51%	56%	87%	58%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj}^j)$	39%	117%	44%	41%	-18%	49%	44%	13%	42%
<i>Education + Income + Other HH Characteristics + Network Externalities + DCT Infrastructure</i>										
	\hat{P}_{rj}^0	0.5316	0.3401	0.1284	0.5731	0.3585	0.0682	0.6855	0.2862	0.0283
% Explained	$(\hat{P}_{uj}^j - \hat{P}_{rj}^0)$	112%	-234%	88%	102%	146%	66%	111%	118%	84%
Remainder	$(\hat{P}_{rj}^0 - \hat{P}_{rj}^j)$	-12%	334%	12%	-2%	-46%	34%	-11%	-18%	16%

Note: Percentages indicate the contribution of the regressed group of variables to the rural - urban gap for each type of access

4.6 - Decomposition of the Inter-Temporal Nested Logit Model

A similar decomposition technique is employed to determine the most important characteristics affecting the access decision over time. Table 44 shows the general (both rural and urban) household rates of type j access in 2003 (denoted \hat{P}_{jt}) and 2000 (\hat{P}_{jt-1}). The first row indicates that 41 percent of households had no access, 36 percent of households had dial-up access, and 22 percent of households had high-speed access in 2003. The second row shows similar rates of access in 2000, while the third row indicates the temporal percentage point change for each type of access. The most dramatic changes occurred in no access and high-speed access, with no access rates dropping by 15 percentage points and high-speed access rising by 17 percentage points over this period. Dial-up access rates actually decreased slightly (by 2.8 percent) over this period. Again, as larger changes occurred in no access and high-speed access rates, measures of the percentage of the shift explained by individual sets of variables are less volatile. Hence, the remainder of this section will focus on the results obtained for the no access and high-speed access temporal shifts.⁷⁵ In order to decompose the change in access rates over time, a synthetic type j access rate (denoted \hat{P}_{jt-1}^0) is created by

⁷⁵ Note that the term "shift" is used for the inter-temporal decomposition, since changes over time are not truly access "gaps."

assigning 2003 parameter vectors to households with characteristics from the year 2000 (as indicated by equation (35) in section 3.5). This allows for the total temporal change in access rates ($\hat{P}_{jt} - \hat{P}_{jt-1}$) to be broken into a component associated with differences in characteristics ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$) and a component associated with differences in parameters ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$).

Table 44. Inter-temporal Nested Logit Decomposition Results

	$j = 0$	$j = 1$	$j = 2$
Rates of Access	No Access	Dialup	High-speed
2003 P_{jt}	0.4127	0.3641	0.2233
2000 \hat{P}_{jt-1}	0.5591	0.3921	0.0488
Delta ($P_{jt} - \hat{P}_{jt-1}$)	-0.1465	-0.0280	0.1744
Explanatory Variables			
<i>Education</i>			
\hat{P}_{jt-1}^0	0.4216	0.3597	0.2186
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	6%	-16%	3%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	94%	116%	97%
<i>Education + Income</i>			
\hat{P}_{jt-1}^0	0.4277	0.3577	0.2145
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	10%	-23%	5%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	90%	123%	95%
<i>Education + Income + Other HH</i>			
\hat{P}_{jt-1}^0	0.4295	0.3594	0.2110
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	11%	-17%	7%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	89%	117%	93%
<i>Education + Income + Other HH + Network Externalities</i>			
\hat{P}_{jt-1}^0	0.4782	0.3730	0.1488
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	45%	32%	43%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	55%	68%	57%
<i>Education + Income + Other HH + Network Externalities + DCT Infrastructure</i>			
\hat{P}_{jt-1}^0	0.4884	0.3809	0.1306
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	52%	60%	53%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	48%	40%	47%

Note: Percentages indicate the contribution of the regressed group of variables to the temporal gap for each type of access

The decomposition begins by running a nested logit model on the type of Internet access for 2003, using only education levels as the dependent variables. The resulting

parameters are then combined with education levels from the year 2000 to determine synthetic rates for the three types of access (\hat{P}_{jt-1}^0). These synthetic rates detail the impact of changing education levels over this time period. Data from Table 4 implies that education levels have risen slightly between 2000 and 2003; which, according to Table 44, accounts for 6 percent of the shift in no access and 3 percent of the shift in high-speed access.

Continuing with the decomposition, a separate nested logit model on 2003 data is run using education and income levels as the dependent variables. New synthetic rates of access (\hat{P}_{jt-1}^0) arise when parameters from this regression are combined with education and income levels from 2000. Table 44 indicates that higher levels of education and income in 2003 account for 10 percent of the shift in no access and 5 percent of the shift in high-speed access. Note that these percentages have increased from their values when only education levels were included (6 and 3 percent, respectively), indicating that income differences do play a minor role in explaining the changes in access rates over time. This pattern continues when other household characteristics are introduced to the nested logit specification. However, the slight increase in the percentage explained that is encountered when other household characteristics are included in the decomposition process is likely the product of countervailing effects. In particular, higher levels of racial and ethnic diversity in 2003 likely increase the various divides, while higher numbers of children in the household and higher rates of access at work in 2003 likely decrease the divides.

It is important to note that over the period 2000 to 2003, the only household characteristics that showed significant change were local access rates (Table 4). As was the case with the rural – urban decomposition, introducing this measure for network externalities to the analysis resulted in the largest increase in the percentage explained. The percentage of the shift in no access explained by the decomposition jumps from 11 percent to 45 percent when network externalities are introduced, while a similar increase is seen for the shift in high-speed access (7 percent to 43 percent). This dramatic increase indicates that the proxy for network externalities explains a large portion of the changes in rates over time; however (as noted in section 4.3), the inability to separate the network

externality measure from other underlying factors likely overstates the importance of this variable. The inclusion of DCT infrastructure increases the percentage explained of the no access and high-speed access temporal changes by 7 and 10 percentage points, respectively. However, this result is tempered by the fact that this increase in the percentage explained stems from DCT variables that are not statistically significant in the nested logit models for both 2000 and 2003.

Ordering of Variables

The non-linearity of the nested logit model implies that the order in which the independent variables are introduced into the analysis is important. To account for this, Table 45 presents the results of the inter-temporal decomposition after reversing the order in which the variables enter the analysis.

Table 45. Inter-temporal Nested Logit Decomposition Results (Order Reversed)

	$j = 0$	$j = 1$	$j = 2$
Rates of Access	No Access	Dialup	High-speed
2003 P_{jt}	0.4127	0.3641	0.2233
2000 \hat{P}_{jt-1}	0.5591	0.3921	0.0488
Delta $(P_{jt} - \hat{P}_{jt-1})$	-0.1465	-0.0280	0.1744
Explanatory Variables			
<i>DCT Infrastructure</i>			
\hat{P}_{jt-1}^0	0.4265	0.3528	0.1950
% Explained $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	9%	-40%	16%
Remainder $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$	91%	140%	84%
<i>DCT Infrastructure + Network Externalities</i>			
\hat{P}_{jt-1}^0	0.5048	0.3545	0.1264
% Explained $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	63%	-34%	56%
Remainder $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$	37%	134%	44%
<i>DCT Infrastructure + Network Externalities + Other HH</i>			
\hat{P}_{jt-1}^0	0.5076	0.3565	0.1295
% Explained $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	65%	-27%	54%
Remainder $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$	35%	127%	46%
<i>DCT Infrastructure + Network Externalities + Other HH + Income</i>			
\hat{P}_{jt-1}^0	0.4952	0.3692	0.1264
% Explained $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	56%	18%	56%
Remainder $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$	44%	82%	44%
<i>DCT Infrastructure + Network Externalities + Other HH + Income + Education</i>			
\hat{P}_{jt-1}^0	0.4884	0.3809	0.1306
% Explained $(\hat{P}_{jt} - \hat{P}_{jt-1}^0)$	52%	60%	53%
Remainder $(\hat{P}_{jt-1}^0 - \hat{P}_{jt-1})$	48%	40%	47%

Note: Percentages indicate the contribution of the regressed group of variables to the temporal gap for each type of access

The results of the reordering are relatively similar to the initial results. Increases in DCT infrastructure between 2000 and 2003 account for 9 percent of the decrease in no access rates and 16 percent of the increase in high-speed access rates, which are comparable to the 7 and 10 percent increases seen under the initial ordering. Once again, the largest increases in the percentages explained (54 percent increase for no access, 40 percent increase for high-speed) are observed when changes in network externalities are added to the analysis. The inclusion of income and education differences over time actually reduces the percent of 2000 to 2003 changes explained by the model, suggesting that their minimal contributions under the initial ordering are not robust results. Even

when all five categories are included, only about half of the no access and high-speed shifts are explained (consistent with the fact that most variables did not change over this period). This indicates that parameter differences between 2000 and 2003 play an equally important role in explaining the inter-temporal shifts. Acknowledging the problem with the network externalities measure implies that the contribution of characteristic differences is probably overestimated in the inter-temporal analysis. Intuitively, the increasing adoption propensities are more likely to be driven by changing relationships between households and the probability of Internet access than by changing characteristics over time. The following analysis uses the full final specification (Education + Income + Other HH Characteristics + Network Externalities + DCT Infrastructure) and incrementally replaces 2003 parameters with those from 2000 in an effort to ascertain which parameter shifts dominate the remaining portion of the change in access rates over this period. Table 46 displays these results.

Table 46. Inter-temporal Nested Logit Decomposition - Contributions of Parameter Shifts

	$j = 0$	$j = 1$	$j = 2$
Rates of Access	No Access	Dialup	High-speed
2003 P_{jt}	0.4127	0.3641	0.2233
2000 \hat{P}_{jt-1}	0.5591	0.3921	0.0488
Delta ($P_{jt} - \hat{P}_{jt-1}$)	-0.1465	-0.0280	0.1744
Replacing Explanatory Variables			
<i>DCT Infrastructure + Network Externalities + Other HH + Income + Education</i>			
\hat{P}_{jt-1}^0	0.4884	0.3809	0.1306
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	52%	60%	53%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	48%	40%	47%
Replacing Parameters			
<i>Inclusive Value (IV) Parameter</i>	0.5076	0.3665	0.1259
% Explained	65%	9%	56%
<i>+ Education</i>	0.4982	0.4072	0.0946
% Explained	58%	154%	74%
<i>+ Income</i>	0.4879	0.4191	0.0930
% Explained	51%	197%	75%
<i>+ Other HH Characteristics</i>	0.6150	0.3099	0.0750
% Explained	138%	-194%	85%
<i>+ Network Externalities</i>	0.6759	0.2352	0.0889
% Explained	180%	-461%	77%
<i>+ DCT Infrastructure</i>	0.6726	0.2444	0.0828
% Explained	177%	-428%	81%
<i>+ Constant</i>	0.5591	0.3921	0.0488
% Explained	100%	100%	100%

Note: Percentages indicate the contribution of the regressed group of variables and replaced parameters to the temporal gap for each type of access

Table 46 begins where Table 44 and Table 45 left off: with the percentage of the inter-temporal changes in access rates explained when all explanatory variables are included. This process generated synthetic rates of access (\hat{P}_{jt-1}^0) by combining characteristics from 2000 with parameter values from 2003. Hence, when variables from 2003 are replaced with those from 2000, 52 and 53 percent of the no access and high-speed access shifts are explained, respectively. The next step in this decomposition is to replace the current parameters in use (those from 2003) with those from 2000 and

observe how this affects the synthetic rate generated. Replacing only the inclusive value parameter has a slight impact on the percentage of the changes in access rates explained, increasing the percentages for both the no access (to 65 percent) and high-speed (to 56 percent) shifts in access rates. This increase in the percentage of the high-speed shift explained indicates that the IV parameter became more important over time – implying that people became increasingly cognizant of the difference between high-speed and dial-up access. Using the 2000 education parameters in place of those for 2003 actually decreases the amount of the no access shift explained, while increasing the percentage of the high-speed shift explained (by nearly 20 percentage points). This indicates that the relationship between education and high-speed access has become *more* important over time, which is contrary to any type of early adopter hypothesis. However, the relationship between education and no access has become less important over time. Replacing the 2003 income parameters with those from 2000 has a similar impact – the percentage of the no access shift explained decreases, while the percentage of the high-speed shift explained increases (by a single percentage point). Thus the early adopter hypothesis appears to fail for both education and income in terms of high-speed access. A dramatic increase in the percentages of the various shifts explained occurs when other household characteristic parameters are substituted. In particular, the percentage of the no access shift explained skyrockets from 51 percent to well over 100 percent. Additionally, the percentage of the high-speed shift explained increases by 10 percentage points. It appears that the parameters associated with other household characteristics might be particularly important in explaining the no access shift. Upon closer examination, it is the shifting value of the age parameters that account for the majority of the large increases seen under this category. In fact, changing only the age parameters and leaving the remaining "other household characteristic" parameters at their 2003 values still increases the percentage of the no access shift explained to over 100 percent, and increases the percentage of the high-speed shift explained by 6 percentage points. Thus, similar to the results seen in the inter-temporal decomposition for general access, it can be argued that high-speed access is becoming more age friendly. Additionally, shifting age parameter values contribute a significant amount to the decrease in "no access" rates seen over this period. While older household heads had a much stronger

association with a lack of access in 2000, this relationship diminished by 2003. It is also worth noting that replacing the parameter associated with rural status had a negligible impact on the results. To observe this result, only the "nm" parameter was replaced and the remaining "other household characteristic" parameters remained at their 2003 values. The resulting synthetic rates of access were not statistically significantly different for any of the three divides, indicating that the core-to-periphery hypothesis is not supported for this model. That is, the rural status of a household did not become any less important in the adoption decision as time progressed.

The replacement of the network externality parameters had a significant impact on the shifts, further increasing the percentage of the no access shift explained, while decreasing the percentage of the high-speed shift explained. This decrease indicates that the network externalities parameters have become less important (in terms of the high-speed shift) as time progresses, which is similar to the results seen for the inter-temporal decomposition performed on general access in section 4.3. This result is interesting, in that it may signify that a different type of diffusion is occurring. The diminishing importance of the network externalities proxy indicates that the effects of regional variation in high-speed access rates have been decreasing over time. Hence, as residential access becomes more prevalent, regional variation in access rates has become less of a factor in the individual household adoption decision.

When DCT infrastructure parameters from 2003 are replaced with those from 2000, the percentage of the no access and high-speed shifts explained vary only marginally. Thus the relationship between DCT infrastructure and the various types of access does not seem to be significantly changing over this time period. One other result worth noting is the dramatic impact of the constant term. Prior to replacing the constant term from 2003, the percentages of the no access and high-speed shifts explained were 177 and 81 percent, respectively. After their inclusion, both percentages necessarily converged to 100 percent. This indicates a rather large *ceteris paribus* shift in preferences away from no access and towards high-speed access. Thus, the constant terms are essentially capturing trends in access rates – this result is somewhat intuitive given the large decrease in no access rates and increase in high-speed rates over this period.

The results demonstrated in Table 46 are reinforced when the order that the parameters are introduced is varied. These results (where the parameters are substituted in reverse order) are shown in Table 47. The inclusion of the constant parameters from 2000 dramatically reduces the percentage of the no access shift explained, while dramatically increasing the percentage of the high-speed access shift explained. These changes confirm the *ceteris paribus* preference shifts (due to trends in access rates) identified under the initial ordering. Changing the DCT infrastructure parameters again had minimal effects on both the no access and high-speed shifts, while replacing the network externalities parameters had a similar effect to that observed under the initial ordering. In particular, the percentage of the high-speed shift explained declined, confirming that these parameters have become less important over time. A dramatic increase in the percentage of the no access shift explained (from 26 to 110 percent) was seen when other household characteristics parameters were replaced. These parameters also had a significant impact on the high-speed shift, increasing the percentage explained from 82 to 89 percent. Changing the income parameters had a relatively small impact on both the no access and high-speed shifts; however, changing the education parameters increased the percentage of the high-speed shift explained by 10 percent. This is consistent with the education results observed under the initial ordering.

Table 47. Inter-temporal Nested Logit Decomposition - Contributions of Parameter Shifts (Order Reversed)

	$j = 0$	$j = 1$	$j = 2$
Rates of Access	No Access	Dialup	High-speed
2003 P_{jt}	0.4127	0.3641	0.2233
2000 \hat{P}_{jt-k}	0.5591	0.3921	0.0488
Delta ($P_{jt} - \hat{P}_{jt-k}$)	-0.1465	-0.0280	0.1744
Replacing Explanatory Variables			
<i>DCT Infrastructure + Network Externalities + Other HH + Income + Education</i>			
\hat{P}_{jt-1}^0	0.4884	0.3809	0.1306
% Explained ($\hat{P}_{jt} - \hat{P}_{jt-1}^0$)	52%	60%	53%
Remainder ($\hat{P}_{jt-1}^0 - \hat{P}_{jt-1}$)	48%	40%	47%
Replacing Parameters			
<i>Inclusive Value (IV) Parameter</i>	0.5076	0.3665	0.1259
% Explained	65%	9%	56%
+ <i>Constant</i>	0.4013	0.5339	0.0648
% Explained	-8%	607%	91%
+ <i>DCT Infrastructure</i>	0.4014	0.5400	0.0587
% Explained	-8%	629%	94%
+ <i>Network Externalities</i>	0.4512	0.4681	0.0807
% Explained	26%	372%	82%
+ <i>Other HH Characteristics</i>	0.5742	0.3575	0.0683
% Explained	110%	-24%	89%
+ <i>Income</i>	0.5644	0.3688	0.0669
% Explained	104%	17%	90%
+ <i>Education</i>	0.5591	0.3921	0.0488
% Explained	100%	100%	100%

Note: Percentages indicate the contribution of the regressed group of variables and replaced parameters to the temporal gap for each type of access

The results demonstrate that shifting levels of access rates over the 2000 – 2003 time frame are approximately equally due to two distinct factors: (1) changes in levels of household characteristics and network externalities (although the proxy for network externalities sheds doubt on the role played by this variable), and (2) shifting relationships between characteristics and access decisions (particularly *ceteris paribus* shifts in preferences for no access and high-speed access). Other results of interest are the relatively large role played by increasing levels of DCT infrastructure in the high-speed access shift (although the DCT infrastructure parameters are not statistically

significant), the importance of shifting age parameters in explaining the no access and high-speed access shifts, and the minimal contributions of shifting income and DCT infrastructure parameters. The conclusions and policy implications stemming from the decompositions performed in this chapter are discussed in the following chapter, along with a discussion of the major limitations of the analysis and areas for further research.

Chapter 5: Policy Implications, Limitations, and Conclusions

The results reported in Chapter 4 decomposed two distinct models of Internet access: one dealing with the general (yes – no) access decision, and one dealing with the nested (none – dial-up – high-speed) access decision.⁷⁶ The first section in this chapter distills policy implications from the general decomposition results, emphasizing both the lack of empirical support for infrastructure-oriented policies and the potential benefits from fostering externalities associated with increasing Internet access rates. The second section focuses on policy implications associated with the nested decomposition results, paying particular attention to past policy initiatives associated with high-speed access. The third section looks at the limitations of the study that give rise to potential future research. The final section provides a brief overarching conclusion.

5.1 – Policy Implications for General Access

At the heart of the discussion on the rural – urban digital divide is whether public intervention to close the divide is warranted. The potential benefits from Internet access are undeniable – commerce, education, and entertainment opportunities, as well as prospects for enhanced social interaction, are all available in today's on-line environment. Existing inequalities in rural and urban household economic well-being (demonstrated in Table 4) may be aggravated by the digital divide. The data presented in this dissertation indicates that the rural – urban divide has persisted over the period 1997 – 2003. This persistence over time provides perhaps the strongest argument that public policies are required to address the divide.

The generation of policies that effectively deal with the rural – urban digital divide must be based on an understanding of the underlying factors. The decomposition results reported in section 4.2 indicate that rural – urban differences in education, income, and regional access rates (“network externalities”) account for between 65 and 100 percent of the general digital divide in any given year. The results also consistently point to the minimal contribution of differences in DCT infrastructure. Furthermore, section 4.3 reports that increases in general access rates over time are due to a number of

⁷⁶ Recall that each model (general and nested) was decomposed twice: once with respect to the rural - urban divide, and once with respect to inter-temporal shifts in access rates.

counteracting forces, including the reduction of age barriers to Internet access. Several policy implications arise from these results. First and foremost, policies that solely address low levels of DCT infrastructure in rural areas will not be effective in reducing the general divide. When data on cable Internet and DSL availability are included, differences in these types of infrastructure between rural and urban areas never account for more than 4 percent of the divide. Hence, even if infrastructure levels in rural areas rivaled those in urban areas, only a small percentage of the digital divide would disappear. Secondly, the lack of support for the "early adopter" and "core-to-periphery" hypotheses indicates that the differential negative propensity for rural areas to access the Internet will not simply dissolve over time. However, the diminishing role of network externalities does indicate that there has been some dispersion in the importance of local access rates in determining individual access. The diffusion of access to older households is also a positive sign for rural areas (whose household heads are typically 2-3 years older than their urban counterparts), but age differences are not a major cause of the general rural – urban divide. **Thus, initiatives that address broader inequities in rural income and education levels will likely have the largest effect in reducing the general divide.**

It is important to understand that the ideology behind the promotion of income and education-oriented policies deals with the *externalities* that arise from increased levels of Internet access (and use). In 2003, over 40 percent of U.S. households had no Internet access. These individuals are essentially left out of the emerging digital culture, and it is difficult for them to interact with and benefit others that are already on-line. Conversely, individuals that are participating in the on-line environment are providing a number of benefits to other Internet users, typically by contributing to the wealth of information that the Internet offers. Question and answer forums are becoming increasingly common on the web, dealing with topics ranging from car repair to career advice. Additionally, most users depend on the Internet for a variety of information and reflect on that information during their everyday conversations with others. Hence, raising the access rates of historically lagging groups (those with lower income and lower education levels) will be beneficial to *all other users*, which is the primary motivation for government intervention. In essence, this intervention is warranted as the promotion of a

public good of Internet information, whose consumption is non-rivalrous and (with appropriate policies) non-excludable. Furthermore, as the results in this dissertation have shown, efforts to encourage Internet access rates should be targeted towards households with lower income and education levels.

The actual form and weight applied to policies on these topics is subject to debate. The results demonstrate that lower levels of income and education seen in rural areas consistently account for approximately 1/3 and 1/5 of the digital divide, respectively. Several policies that deal with these factors are outlined below.

Income

As Mills and Whitacre (2003) note, public support for policies dealing with general income discrepancies is limited at best. Hence, policies purporting to raise Internet access rates via income levels should include components that are “access-specific.” For instance, simply providing subsidies to all households without Internet access is a policy action that is not likely to find much public support or to be cost effective. Rather, the provision of additional income *conditional on* a household demonstrating that they are accessing and using the Internet (perhaps through completion of monthly on-line questionnaires) would be more apt to satisfy critics who believe the subsidy might be used inappropriately. Additionally, a means test may be implemented to determine whether a household would be eligible for this subsidy. The results shown in Table 16 indicate that household income levels must reach \$20,000 (faminc6) before becoming statistically significant in the adoption decision. Thus, households with income levels below this threshold should be the first targets of the subsidy. As Table 4 indicates, rural areas have a disproportionately larger share of households under this income level, implying that these subsidies are likely to be provided in rural areas, and thus have an impact on the rural – urban digital divide. Recall that the principal rationale for implementing *any* type of public policy is to take advantage of the externalities generated by Internet access. Access-specific subsidies to low-income households would therefore

ensure that this group (predominantly located in rural areas) contribute to the on-line culture by requiring some type of Internet activity.⁷⁷

Education

Intuitively, affecting Internet access rates by raising education levels is more difficult than affecting them via income levels. In particular, a significant amount of time and effort are required to increase the education level of a household head. For example, moving from "high school degree" to "college degree" would typically take at least four full years and a large financial investment, leaving education levels of Americans in their 30s and up essentially fixed. A number of studies have pushed for public policies that address education discrepancies between rural and urban areas based on equity grounds; however, the strong contribution of education differences to the digital divide in general Internet access – has never been explicitly documented in this literature. This contribution should be stressed when policies dealing with education differences are proposed. Furthermore, there is an inherent inter-generational effect that occurs when education is addressed, due to the strong inter-generational transmission of educational attainment. For this reason, policies dealing with education will likely be more effective in reducing the inter-generational digital divide than will policies addressing current income differences.

Investing in education-oriented policies simply to capture the social externalities associated with Internet access is not realistic. Rather, the results should be thought of as another dimension of the benefits from closing the rural – urban education gap. Policies should attempt to ensure that investments are made at a socially, rather than privately, optimal level. Increasing funding for community colleges should be stressed, as the marginal effects of some college education on the likelihood of Internet access are close to those for a bachelor's degree (Table 13). Additionally, community colleges typically offer introductory courses on computer and Internet use, which would be useful to any interested individuals (regardless of education level). Public investments have already had success in creating equal Internet access among children ages 6 to 17 across income

⁷⁷ Note that there is a (somewhat blurred) distinction between the provision of free access and the subsidy discussed here. In effect the subsidy is attempting to influence the household decision on Internet access, while the access provision simply gives access to a household regardless of whether or not they want it.

and racial groups in the nation's public schools (Newburger 2001, NTIA 2002). This success will likely reduce the role of education differences in the rural – urban digital divide for future generations. Similar investments to promote the educational attainment of today's household heads would increase the likelihood of access for households headed by an individual with less than a college degree. While not cost effective solely from the standpoint of closing the rural – urban general digital divide, such investments should point out this byproduct while addressing education gaps in general.

Network Externalities

While policies for income and education are stressed to promote the benefits of externalities resulting from Internet access, the results discussed herein indicate that local access rates are also important in the household adoption decision. However, the proxy for these network externalities is open to criticism in this analysis. In particular, it is very likely that the proxy captures other underlying factors in the adoption decision, thus overstating the importance of network externalities. This issue will be discussed further under section 5.3, where the limitations of the study are explored. The following section discusses the policy implications for network externalities, but these policies are proposed with caution due to the problems with the network externality measure noted above.

Increasing levels of network externalities is somewhat of a “chicken-and-the-egg” argument, as raising local rural rates of access will by definition increase the percentage of rural households accessing the Internet, and should decrease the rural – urban divide. However, the results demonstrate that local access rates are a significant determinant of the individual household's adoption decision after differences in household characteristics are controlled for in the model. This indicates that individual households are largely influenced by the access decisions of those directly surrounding them. This is, in essence, the positive social externality generated by individual household access. Policies that focus on the promotion of access in localized rural areas are, therefore, likely to have substantial spillovers. Such policies include digital villages or subsidized area user groups like community technology centers. Historically, digital villages have been promoted through private sector contributors such as Hewlett Packard

(<http://grants.hp.com/us/digitalvillage/>), and have been successful in introducing the Internet to individuals unfamiliar with such technology. Similar results have been obtained through community technology centers funded by the government.⁷⁸ The implication of such government-run community technology centers could have a dramatic impact on local access rates, particularly if participation were linked to the income subsidies discussed above. Recipients of the previously discussed access-specific income subsidies might be required to participate in the technology center. This would simultaneously increase their knowledge of the Internet while allowing them to pass on their own information and influence the adoption propensities of their neighbors. It should be noted, however, that the influence of these network externalities has been decreasing as rates of access have increased over time. Regional variation in access rates still influenced household decisions in 2003, but this effect has been decreasing since 2000. This reduction in network externalities may be associated with diffusion of information on Internet access from core to periphery.⁷⁹

5.2 – Policy Implications for High-speed Access

As the data in Table 39 suggests, the rural – urban digital divide has essentially shifted to one of high-speed access.⁸⁰ This is an important point, as policies dealing with the digital divide must now focus specifically on this type of access. Recent history has seen the introduction of a number of congressional bills that center on the provision of high-speed infrastructure in rural or low-income areas (Kruger, 2005). These policies have pushed for the provision of DCT infrastructure through grants (the House Agricultural Appropriations for FY2004 contained \$9 million for broadband grants), tax credits (included in the Broadband Internet Access Act, which has been introduced to congress 3 times), or other measures (guaranteed broadband service loans were made

⁷⁸ Approximately \$10 million was funded for such centers in FY2004 under the Department of Education (Kruger, 2005).

⁷⁹ “Core to periphery diffusion” as discussed in section 4.1 only dealt with the metro / non-metro status of a household. Using local access rates (network externalities) instead of a 0/1 measure allows for more insight into the diffusion of access.

⁸⁰ The dial-up divide shrank from 11 percent in 2000 to a negative percentage in 2003. Meanwhile, the high-speed divide increased from 3 percent to 14 percent over these years.

available through the Farm Security and Rural Investment Act of 2002).⁸¹ The results of this dissertation, however, suggest that such policies do not address the dominant factors underlying the high-speed divide. Rather, the decomposition results reported in section 4.5 indicate that much like the general divide, tackling the high-speed divide requires addressing differences in income and levels of network externalities. Replacing rural levels of income and network externalities with those from urban areas consistently generate synthetic high-speed (as well as dial-up) access rates that are significantly higher than actual rural rates.

Addressing the rural – urban divide in high-speed residential Internet access should therefore be linked to policies that deal with income inequalities and promote “localized” high-speed user groups. Suggested policies that deal with these issues are the same as those for general access presented in section 5.1, including the provision of income subsidies and the insertion of community technology centers. Some may argue that technology centers in rural areas will have little impact on residential high-speed access rates if there is no infrastructure available for such access; however, the presence of these centers, along with the suggested income subsidies, would be a persuasive argument for cable and DSL providers to invest in these areas. Additionally, it must be remembered that the results reported here categorize *only* cable and DSL as DCT infrastructure; in reality, unwired technology such as wireless and satellite systems may further diminish the role that “wired” infrastructure plays. The minimal contribution of differences in DCT infrastructure between rural and urban areas does not mean that future policies should completely forsake promoting infrastructure in rural areas. It simply implies that other factors – namely, differences in levels of income and network externalities – are potentially more important in determining high-speed access rates and need to be part of the policy portfolio. Moreover, as the inter-temporal results discussed in section 4.6 reveal, higher levels of DCT infrastructure do play a role in increasing high-speed access rates over time; however, this role is much smaller than that played by increasing levels of network externalities.⁸² While the inter-temporal results indicate that

⁸¹ See Kruger (2005) for a comprehensive list of federal assistance programs dealing with broadband, and their obligations in FY2004.

⁸² Recall, however, the network externalities measure may capture some of the DCT infrastructure effects, thus understating the importance of infrastructure.

the impact of network externalities has been decreasing over time, these externalities continue to be an important contributor to the high-speed divide. The inter-temporal results did not find any support for early adopter or core-to-periphery hypotheses, further indicating that government action may be necessary in order to address the rural – urban high-speed divide.⁸³

Recall that the ultimate rationale for government intervention is to capture positive externalities that would not result from individual household choices. The existence of these externalities also suggests that the market may not provide optimal levels of service, since suppliers are more likely to supply if there are more users, and consumers are more likely to demand if there are more people to interact with or ways to use the technology. This is particularly true for high-speed access due to the expenses involved in providing infrastructure and the multitude of on-line experiences available to high-speed users. The results of this dissertation suggest that the best policies to reach households with lower access rates (for the purposes of this study, those in rural areas) will induce demand, namely by subsidizing access and promoting community networks.

Gillett, Lehr, and Osorio (2004) develop a taxonomy for potential broadband policies involving local governments. Four types of initiatives are discussed, based on whether the role of the government is that of a (1) demand stimulator / aggregator, (2) rule-maker, (3) financier, or (4) infrastructure developer. The results discussed earlier suggest that a combination of the first and third roles (demand stimulator and financier) would be the most effective in reducing the high-speed digital divide. Particularly, stimulating demand through policies such as community technology centers or community information services (the nearby Blacksburg Electronic Village is a good example) would likely promote local network externalities. Since *regional* variations in Internet access rates are used as the measure of network externalities in the analysis, local or state-level policies may be the most effective way to address this factor (by implementing changes at the regional level). Additionally, providing subsidies to low-income potential high-speed adopters in rural areas would address income-based discrepancies. As noted previously, these subsidies should do more than simply provide

⁸³ The core-to-periphery hypothesis deals explicitly with the role of the "rural" term in the specification, however, as noted, the diminishing role of network externalities indicates that the impact of local access rates has been diffusing to some extent.

access to rural households. An example of the shortcomings of this type of subsidy can be found in LaGrange, Georgia; where WebTV equipment and service was given away for a one-year period. However, the free Internet service was conditional on the household having cable service, something that many of the households did not have (Youtie, 2002). Additionally, the WebTV devices provided had no print or downloading capability. More general income subsidies would allow households to make decisions on which equipment to purchase (recall that the policies discussed above would require proof of Internet use in order for the subsidy to continue) as they see fit. A means test may not be necessary to ensure that these subsidies reach the proper households, since Table 30 indicates that household income must be quite high (\$75,000 – faminc13) before the level of income becomes statistically significant in the high-speed adoption decision. Raising household income to this threshold simply to address the high-speed divide is not a realistic policy prescription. Instead, clearly indicating that the (smaller) subsidy provision is conditional on high-speed access should attract rural households whose primary rationale for *not* having high-speed access is income-related.

5.3 – Limitations and Areas for Future Research

Several limitations have been brought up during the course of this analysis. Perhaps the largest limitation is the problematic nature of the proxy used for network externalities. A number of the decomposition results pointed towards the dramatic impact of these externalities, both from a rural – urban standpoint and from an inter-temporal perspective. However, the household measure used for network externalities – the state-level rural or urban rates of access where the household is located – has several significant drawbacks. The first drawback is that this measure is quite aggregate. While it is reasonable to think that a household's adoption decision might be influenced by other households in a relatively close region, rural / urban status within a state may be pushing the boundary of a relevant region. In effect, the aggregate nature of the measure implies that the adoption decision of a rural household in southern California is influenced by the adoption rates of rural households in northern California, which are over 700 miles away. The second drawback of this measure is that it most likely captures underlying factors in the adoption decision. For example, areas with high access rates (and thus high levels of

network externalities) likely have high levels of factors that influence the adoption decision, potentially including DCT infrastructure levels that are not adequately measured in this study. Thus, including this crude network externality measure in a regression will likely overstate their importance. Unfortunately, the data required to separate the true impact of network externalities from other factors (such as information on the amount and type of interaction with their neighbors) are not included in the Current Population Surveys.

Another limitation of this analysis is the measure of DCT infrastructure. Data on cable Internet and DSL availability exist at the county and city level, respectively. However, the current analysis assumes that if cable Internet is available in some part of the county, then all parts of that county have cable Internet available to them. This will not always be a correct assumption, particularly for rural areas in a county. Thus, the measure for DCT infrastructure likely overstates its availability and may understate its importance in the various specifications. Furthermore, the infrastructure data must be aggregated to rural / urban status within a state in order to mesh with the household survey data. This aggregation further limits the capability of the empirical models to derive the role of infrastructure in the various divides. Finally, only cable and DSL are included as measures of DCT infrastructure. While these technologies made up 99 percent of the high-speed market in 2003, other technologies such as satellites and wireless connections may become more prevalent as time progresses. In fact, Philadelphia is attempting to provide wireless access throughout the entire city through their "Wireless Philadelphia" project (Levy, 2005).⁸⁴ The roles of these wireless technologies are not evaluated under the analysis presented here.

Future research on this topic would benefit from lower levels of geographic information on households (such as counties or zip codes). Historically this information has been available on a limited number of datasets through the Census Bureau, but only after research proposals have specifically been selected to obtain access to the confidential data. Currently, the computer use supplement is not one of those datasets.

⁸⁴ It is interesting to note that cable and telecommunications companies are fighting to prevent such municipal wireless service (Levy, 2005).

Obtaining this lower geographic level of data would allow household information to be meshed with more accurate measures of DCT infrastructure, since these measures are available at the city and county level. It would also allow for more "local" measures of network externalities to be used in the analysis.

Future research should also focus on the shifting nature of the digital divide from dial-up to high-speed access. 2003 was the first year where rural dial-up rates actually exceeded those for urban areas. As high-speed access becomes more prevalent, researchers should keep a close eye on the factors affecting the rural – urban divide in high-speed access. It may be the case that DCT infrastructure becomes more important as knowledge regarding the benefits from high-speed access diffuses to households with lower income and education levels. Along this same vein, determining the factors affecting Internet "use" as opposed to simply "access" would be valuable. While datasets may be limited, any insight regarding how rural and urban areas differ in their actual use of the Internet would be beneficial to understanding differences in underlying costs and benefits of access. If current patterns persist, rural areas may continue to exceed urban areas in terms of dial-up access, but fall further behind in terms of high-speed access. This may lead to a digital divide in Internet "use" since a greater variety and a larger quantity of activities can be accomplished via a high-speed connection.

Finally, future research should attempt to determine the efficacy of relevant public policies, such as the implementation of a community technology center or the provision of income subsidies contingent upon Internet use. This will likely require datasets gathered both before and after the policy is enacted. Such panel data would allow for any changes in household characteristics to be controlled and hence determine the impact of the policy on access rates in an area. However, the requirement of pre- and post- policy datasets will likely limit such analysis to smaller geographic regions, such as counties. This type of research is critical to understanding the impact of various policies on potential adopters.

5.4 – Concluding Remarks

The Internet has risen from relatively obscurity to an essential part of life for the majority of U.S. households in less than 10 years. During this time, a gap in the access

rates of rural and urban households has evolved and persisted. Additionally, recent data indicates that the rate of high-speed residential connections is increasing dramatically, along with a rural – urban access disparity. This dissertation has decomposed the rural – urban gaps in general and high-speed access into the roles played by a number of contributing factors. The results suggest that the gaps are driven by differences in levels of income, education, and network externalities between rural and urban households. Perhaps more importantly, differences in levels of DCT infrastructure (namely DSL and cable Internet capacity) do *not* explain a large portion of the gaps. To date, the majority of the policies purporting to deal with the digital divide have focused on the provision of infrastructure to rural areas. Upon closer examination, it appears that these policies have been dealing with a relatively unimportant aspect of the divide.

Future policies should focus on fostering positive externalities by increasing the propensity of access for those households with historically low rates – those with low income and education levels. These policies will inherently address the rural – urban digital divide since low income and education levels are disproportionately found in rural areas. Examples of policies dealing with these contributors were discussed in sections 5.1 and 5.2. Effective implementation of such policies will likely require a combination of federal and local actions. Once enacted, the effects of any policies should be monitored, and additional research should be conducted to determine if the divide is actually being “bridged.”

References

- Barnett, A. and D. Kaserman. 1998. "The Simple Welfare Economics of Network Externalities and the Uneasy Case for Subscriber Subsidies." *Journal of Regulatory Economics*. 13: 245-254.
- Ben-Akiva, M. and S. Lerman. 1985. *Discrete Choice Analysis: Theory and Application to Travel Demand*. Cambridge, MA: MIT Press.
- Bimber, B. 2000. "Measuring the Gender Gap on the Internet." *Social Science Quarterly*. 81: 868-876.
- Blinder, A. 1973. "Wage Discrimination: Reduced Form and Structural Variables." *Journal of Human Resources*. 8: 436-455.
- Borsch-Supan, A. 1990. "On the Compatibility of Nested Logit Models with Utility Maximization." *Journal of Econometrics*. 43: 373-388.
- Brown, L.A. 1981. *Innovation Diffusion: A New Perspective*. New York: Methuen.
- Capps, O and R.A. Kramer. 1985. "Analysis of Food Stamp Program Participation Using Qualitative Choice Models." *American Journal of Agricultural Economics*. February: 49-59.
- Children's Partnership. 2000. "Online Content for Low Income and Underserved Americans: The Digital Divide's New Frontiers. The Children's Partnership: www.childrenspartnership.org
- Ciccone, A. and R.E. Hall. 1996. "Productivity and the Density of Economic Activity." *American Economic Review*. 86(1): 54-70.
- Compaine, B. M., ed. 2001. *The Digital Divide: Facing a crisis or Creating a Myth?* London: MIT Press.
- Competitive Broadband Coalition (CBC). 1999. *Setting the Record Straight: The Fallacies and Realities of the Broadband Debate*. Available at <http://www.competitivebroadband.org/1041/> (Oct. 25, 1999).
- Cooper, D. 2002. "What is the Future for Universal Service in a Broadband World?" *USTA Telecom Executive Magazine*. September / October.
- Cooper, M and G. Kimmelman. 1999. "The Digital Divide Confronts the Telecommunications Act of 1996." Consumer Federation of America: Washington D.C.

- Crandall, R., C. Jackson, and H. Singer. 2003. *The Effect of Ubiquitous Broadband Adoption on Investment, Jobs, and the U.S. Economy*. Criterion Economics. Downloaded Feb 2004 from http://www.criterioneconomics.com/docs/ubiquitous_broadband_adoption.pdf
- Cremer, J. 2000. "Network Externalities and Universal Service Obligation on the Internet." *European Economic Review*. 44: 1021-1031.
- Cummings, J. and R. Kraut. 2002. "Domesticating computers and the Internet." *The Information Society*. 18: 221-231.
- De Castro, E. and C. Jensen-Butler. 2003. "Demand for Information and Communication Technology-based Services and Regional Economic Development." *Papers in Regional Science*. 82: 27-50.
- Downes, T. and S. Greenstein. 1998. "Do Commercial ISPs Provide Universal Access?" Mimeo. Department of Economics, Tufts University.
- Drabenstott, M. 2001. "New Policies for a New Rural America." *International Regional Science Review*. 24, 1:3-15.
- Encyclopedia Britannica*. 2004. Volume 13. "Modems," pp. 130-132.
- Fairlie, R. 2003. "An Extension of the Blinder-Oaxaca Decomposition Technique to Logit and Probit Models." *Economic Growth Center: Yale University*. Paper Number 873.
- Faulhaber, G. and C. Hogendorn. 2000. "The Market Structure of Broadband Telecommunications." *The Journal of Industrial Economics*. 48: 305-329.
- Feder, G., R. Just, and D. Zilberman. 1985. "Adoption of agricultural innovations in developing countries: A survey." *Economic Development and Cultural Change*. 33: 255-298.
- Federal Communications Commission. 2000. *Deployment of Advanced Telecommunications Capability: Second Report*. FCC 00-290. Available at http://www.fcc.gov/Bureaus/Common_Carrier/Orders/2000/fcc00290.pdf
- Federal Communications Commission. 2002. *Deployment of Advanced Telecommunications Capability: Third Report*. FCC 02-330. Available at http://www.fcc.gov/Bureaus/Common_Carrier/Orders/2002/fcc02330.pdf
- Federal Communications Commission – Industry Analysis and Technology Division. 2003. *High Speed Services for Internet Access: Status as of June 30, 2003*. Available at <http://www.fcc.gov/wcb/iatd/comp.html>

- Finke, M. and S. Huston. 2003. "Factors Affecting the Probability of Choosing a Risky Diet." *Journal of Family and Economic Issues*. 24:3, 291-303.
- Foreman, R. 2002. "For Whom the bell alternatives toll: Demographics of residential facilitiesbased telecommunications competition in the United States." *Telecommunications Policy* 26: 573-587.
- Forestier, E., J. Grace, and C. Kenney. 2002. "Can information and communication technologies be pro-poor?" *Telecommunications Policy*. 26: 623-646.
- Fox, W and S. Porca. 2001. "Investing in Rural Infrastructure." *International Regional Science Review*. 24:1, 103-133.
- Gabe, T. and J. Abel. 2002. "Deployment of Advanced Telecommunications Infrastructure in Rural America: Measuring the Digital Divide." Department of Resource Economics and Policy: University of Maine.
- Georgia Institute of Technology. 1998. *10th Gvu WWW User Survey*. Downloaded from http://www.cc.gatech.edu/gvu/user_surveys/survey-1998-10/ on 7-7-04.
- Gillett, S., W. Lehr, and C. Osorio. 2004. "Local Government Broadband Initiatives." *Telecommunications Policy*. 28: 537-558.
- Glasmeier, A. and L. Wood. 2003. *Broadband Internet Service in Rural and Urban Pennsylvania: A Common Wealth or Digital Divide?* Center for Rural Pennsylvania: Harrisburg, PA.
- Goolsbee, A. and P. Klenow. 2002. "Evidence on Learning and Network Externalities in the Diffusion of Home Computers." *Journal of Law and Economics*. 45: 317 – 343.
- Graham, S. and A. Aurigi. 1997. "Virtual Cities, Social Polarization, and the Crisis in Urban Public Space." *Journal of Urban Technology* 4:19-52.
- Greenman, C. 2000. *Life in the Slow Lane: Rural Residents Are Frustrated by Sluggish Web Access and a Scarcity of Local Information Online*. New York Times. May 18, p. D1.
- Greenstein, S. 2000. "Building and Delivering the Virtual World: Commercializing Services for Internet Access." *The Journal of Industrial Economics* 48:8, 391-411.
- Greenstein, S. and M. Lizardo. 1999. "Determinants of the Regional Distribution of Information Technology Infrastructure in U.S.," in (Eds) Dale Orr and Tom Wilson, *The Electronic Village: Public Policy Issues of the Information Economy*. C.D. Howe Institute, Toronto, Canada.

- Griliches, Z. 1957. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica* 25:4, 501 – 522.
- Grimes, S. 1992. "Exploiting Information and Communications Technologies for Rural Development." *Journal of Rural Studies* 8:3, 269-278.
- Grimes, S. 2000. "Rural areas in the information society: diminishing distance or increasing learning capacity?" *Journal of Rural Studies* 16, 13-21.
- Grubestic, T. 2003. "Inequities in the Broadband Revolution." *Annals of Regional Science* 37: 263 – 289.
- Grubestic, T. and A. Murray. 2004. "Waiting for Broadband: Local Competition and the Spatial Distribution of Advanced Telecommunication Services in the United States." *Growth and Change*. 35: 2, 139-165.
- Hargittai, E. 2003. "The Digital Divide and What To Do About It." Book chapter to appear in *New Economy Handbook* edited by Derek C. Jones. San Diego, CA: Academic Press. Downloaded from <http://www.eszter.com/research/pubs/hargittai-digitaldivide.pdf> on 4-10-04.
- Hausman, J. 1978. "Specification Tests in Econometrics." *Econometrica*. 46: 1251-1271.
- Hausman, J. and D. McFadden. 1984. "Specification Tests for the Multinomial Logit Model." *Econometrica*. 52,5: 1219 – 1240.
- Heiss, F. 2002. "Specification(s) of Nested Logit Models." *The Stata Journal*. 2:3, 227-252.
- Hensher, D. and W. Greene. 2000. "Specification and Estimation of the Nested Logit Model: Alternative Normalizations." Mimeo, New York University.
- Hite, J. 1997. "The Thunen model and the new economic geography as a paradigm for rural development policy." *Review of Agricultural Economics* 19 (2), 230 - 240.
- Hobbs, V. and J. Blodgett. 1999. *The Rural Differential: An Analysis of Population Demographics Areas Served by Rural Telephone Companies*, P99-8. Rural Policy Research Institute, Columbia, MO. Downloaded from <http://www.tprc.org/ABSTRACTS99/hobbspap.pdf> on 9-10-04.
- Horrigan, J.B. 2001. *Online Communities: Networks that Nurture Long-Distance Relationships and Local Ties*. Pew Internet & American Life Project. <http://pewinternet.org/>

- Horrigan, J.B. 2004. *PEW Internet Data Project Memo: Home Broadband Adoption has Increased 60% in the past year and use of DSL Lines is Surging*. Pew Internet & American Life Project. <http://pewinternet.org/>
- Jevons, W.S. 1911. *Theory of Political Economy* (4th Ed). London: MacMillan.
- Knapp, T, N. White, and D. Clark. 2001. "A Nested Logit Approach to Household Mobility." *Journal of Regional Science*. 41,1: 1-22.
- Kruger, L. 2005. "Broadband Internet Access and the Digital Divide: Federal Assistance Programs." *Congressional Research Service – The Library of Congress*.
- Layton, A. and M. Katsuura. 2001. "Comparison of Regime Switching, Probit and Logit Models in Dating and Forecasting U.S. Business Cycles." *International Journal of Forecasting*. 17: 403-417.
- Le, A. and P. Miller. 2004. "Inter-temporal Decompositions of Labour Market and Social Outcomes." *Australian Economic Papers*. March: 10 –20.
- Levy, Steven. 2005. "Pulling the Plug on Local Internet." *Newsweek (Technology and Science Section)*. July 18.
- Madden, M. 2003. *America's On-line Pursuits: The changing picture of who's on-line and what they do*. Pew Internet & American Life Project. <http://pewinternet.org/>
- Mahajan, V., and R. Peterson. 1985. *Models for innovation diffusion*. Beverly Hills, CA: Sage.
- Mahler, A. and E. Rogers. 1999. "The Diffusion of Interactive Communication Innovations and the Critical Mass: the Adoption of Telecommunications Services by German Banks." *Telecommunications Policy* 23: 719-740.
- Malecki, E.J. 2003. "Digital Development in rural areas: potentials and pitfalls." *Journal of Rural Studies* 19: 201-214.
- Malecki E.J. 2002a. "Local competition in telecommunications in the United States: Supporting conditions, policies, and impacts." *Annals of Regional Science* 36: 437-454.
- Malecki, E.J. 2002b. "Telecommunications Competition since 1996: Is It Happening? If So, Where?" presented at the 2002 North American Regional Science Conference, San Juan, Puerto Rico, November 2002.

- Malecki, E.J. and C. Boush. 2003. "Telecommunications Infrastructure in the Southeastern United States: Urban and Rural Variation." *Growth and Change* 34: 109-129.
- Marshall, A. 1920. *Principles of Economics: An Introductory Volume. 8th Edition.* London: MacMillan.
- Mason, R. and T. Valletti. 2001. "Competition in Communication Networks: Pricing and Regulation." *Oxford Review of Economic Policy.* 17:3, 389 – 415.
- Mason, S. and K. Hacker. 2003. "Applying Communication Theory to Digital Divide Research." *IT & Society.* 1, 5: 40 – 55.
- McConnaughey, J., C. Nila, and T. Sloan. 1995. *Falling through the net: A survey of the "have nots" in rural and urban America.* Washington, DC: U.S. Department of Commerce, National Telecommunications and Information Administration.
- McConnaughey, J., and W. Lader. 1998. *Falling Through the Net II: New Data on the Digital Divide.* National Telecommunications and Information Administration, <http://www.ntia.doc.gov/ntiahome/net2/falling.html>
- McFadden, D. 1981. Econometric Models of Probabilistic Choice. In *Structural Analysis of Discrete Data and Econometric Applications*, eds. C.F. Manski and D.L. McFadden, 198-272. Cambridge, MA: MIT Press.
- Mills, B.F. and B. Whitacre. 2003. "Understanding the Non-Metropolitan – Metropolitan Digital Divide." *Growth and Change.* 34,2: 219-243.
- Moss, M. and S. Mitra. 1998. "Net Equity: A Report on Income and Internet Access." *Journal of Urban Technology* 5:23-32.
- National Cable and Telecommunications Association. 2004. *Industry Overview.* Downloaded from <http://www.ncta.com/Docs/pagecontent.cfm?pageID=96> July 2004.
- National Telecommunications and Information Administration and Economics Statistics Administration. 1999. *Falling Through the Net: Defining the Digital Divide.* U.S. Department of Commerce: Washington, D.C.
- National Telecommunications and Information Administration and Economics Statistics Administration. 2000. *Falling Through the Net: Towards Digital Inclusion.* U.S. Department of Commerce: Washington, D.C.
- National Telecommunications and Information Administration and Economics Statistics Administration. 2002. *How Americans Are Expanding Their Use of the Internet.* U.S. Department of Commerce: Washington, D.C.

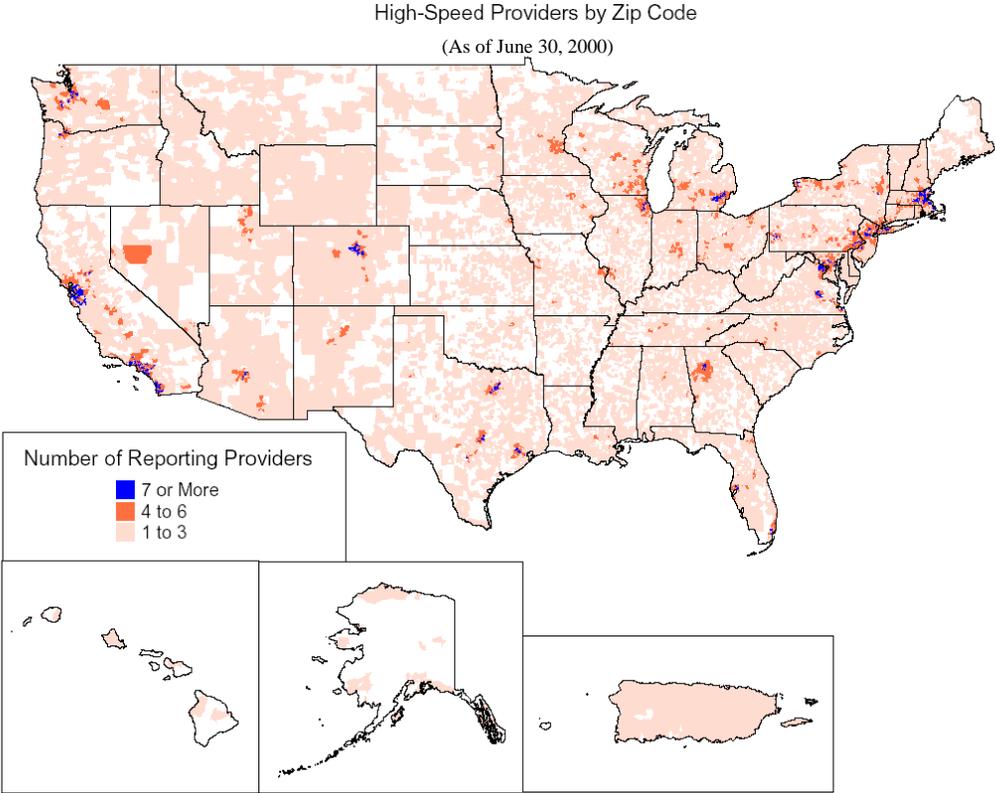
- National Telecommunications and Information Administration (NTIA) and Rural Utilities Service (RUS). 2000. *Advanced Telecommunication in Rural America: The Challenge of Bringing Broadband Service to All Americans*. U.S. Department of Commerce: Washington, D.C.
- Neumark, D. 1988. "Employers Discriminatory Behavior and the Estimation of Wage Discrimination." *Journal of Human Resources*. 23: 279-295.
- Newberger, E.C. 2001. *Home Computers and Internet Use in the United States*. Special Study P23-207. Washington, DC: U.S. Department of Commerce.
- Nicholas K. 2002. Stronger than barbed wire: How geo-policy barriers construct rural Internet access, in L.F. Cranor and S. Greenstein, eds. *Communications Policy and Information Technology: Promises, Problems, Prospects*. Cambridge, MA: MIT Press, pp. 299-316.
- Nielson, H.S. 1998. "Discrimination and Detailed Decomposition in a Logit Model." *Economics Letters*. 61: 115-120.
- Oaxaca, R. 1973. "Male-Female Differentials in Urban Labor Markets." *International Economic Review*. 14: 693-709.
- Oaxaca, R and M. Ransom. 1994. "On Discrimination and Decomposition of Wage Differentials." *Journal of Econometrics*. 61: 5-21.
- Paradyne Corp. 2000. The DSL Sourcebook. Online document available at www.paradyne.com/sourcebook_offer/sb_1file.pdf.
- Parker, E. B. 2000. "Closing the digital divide in rural America." *Telecommunications Policy*. 24: 281-290.
- Pinkham Group. 2002. "DSL Deployment Analysis of RBOCs and Independent LECs in Metropolitan and Rural Areas – Q4 2001." Available at http://www.dslprime.com/a/pinkham_deployment.htm
- Prieger, J. 2003. "The supply side of the digital divide: Is there equal availability in the broadband Internet access market?" *Economic Inquiry*. 41,2: 346 - 363.
- Rogers, E.M. 1960. *Diffusion of Innovations, First Edition*. New York: Free Press.
- Rogers, E.M. 2003. *Diffusion of Innovations, Fifth Edition*. New York: Free Press.
- Rogers, E. M., and Shoemaker, F. F. 1971. *Communication of innovations: A cross-cultural approach (2nd ed.)*. New York: Free Press.

- Rose, R. 2003. *Oxford Internet Survey Results*. The Oxford Internet Institute: The University of Oxford, UK.
- Rural Utilities Service (RUS). 2000. *Exparte Comments in the Matter of Annual Assessment of Competition in Markets for the Delivery of Video Programming*. U.S. Department of Agriculture: Washington, D.C.
- Rural Utilities Service (RUS). 2004. *2003 Rural Utilities Service Annual Report*. U.S. Department of Agriculture: Washington, D.C.
- Schultz, T.W. 1975. "The Value of the Ability to Deal with Disequilibria." *Journal of Economic Literature*. 13, 3: 827-846.
- Smale, M., R. Just, and H. Leathers. 1994. "Land Allocation in HYV Adoption Models: An Investigation of Alternative Explanations." *American Journal of Agricultural Economics*. 76, 3: 535-546.
- Smith, J. and F. Welch. 1989. Black Economic Progress after Myrdal. *Journal of Economic Literature*. 27: 519 – 564.
- Song, T. 2005. "The Computer and Internet Technology Adoption and The Urban – Rural Digital Divide: Does High-Speed Internet Access Matter?" Unpublished chapter of dissertation, Iowa State University.
- Stigler, G. 1950. "The Development of Utility Theory." *The Journal of Political Economy*. 58:4, 307 – 327.
- Strover, S. 1999. *Rural Internet Connectivity*, P99-13. Rural Policy Research Institute, Columbia, MO.
(http://www.rupri.org/pubs/archive/reports/1999/P99_13/index.html)
- Strover, S. 2001. "Rural Internet Connectivity." *Telecommunications Policy* 25: 331-347.
- Strover, S. 2003. "The prospects for broadband deployment in rural America." *Government Information Quarterly* 20: 95-106
- Strover, S. and L. Berquist. 1999. "Telecommunication infrastructure development: The state and local role." Columbia, MO, The Rural Policy Research Institute.
- Tarde, G. 1903. *The Laws of Imitation*, translated by E.C. Parsons with introduction by F.Giddings, New York, Henry, Holt and Co.
- TeleGeography, Inc. 2003. U.S. Internet Geography 2003: Domestic Internet Statistics and Commentary. Washington, DC: TeleGeography Inc.

- Telecommunications Industry Association. 2003. *The Economic and Social Benefits of Broadband Deployment*. Arlington, VA. Downloaded from www.tiaonline.org/policy/broadband/Broadbandpaperoct03.pdf Feb 2004.
- Townsend, A. 2001. "The Internet and the Rise of the New Network Cities, 1969 - 1999." *Environment and Planning B: Planning and Design*, 28: 39-58.
- Warf, B. 2001. "Segueways into Cyberspace: Multiple Geographies of the Digital Divide." *Environment and Planning B: Planning and Design* 28: 3-19.
- Warren Publishing Inc. 2000. *Television and Cable Factbook, Cable and Services*. Number 68: Cable Volumes 1-2. Washington, D.C.
- Warren Publishing Inc. 2001. *Television and Cable Factbook, Cable and Services*. Number 69: Cable Volumes 1-2. Washington, D.C.
- Warren Publishing Inc. 2003. *Television and Cable Factbook, Cable and Services*. Number 71: Cable Volumes 1-2. Washington, D.C.
- Warschauer, M. 2002. "Reconceptualizing the Digital Divide." *First Monday* 7 (7).
- Wellington, A. 1993. "Changes in the Male / Female Wage Gap, 1976-85." *Journal of Human Resources*. 28: 383-411.
- Wooldridge, J.M. 2002. *Econometric Analysis of Cross Section and Panel Data*. Cambridge: MIT Press.
- Youtie, J. 2000. "Field of dreams revisited: economic development and telecommunications in LaGrange, Georgia." *Economic Development Quarterly* 14: 146-153.
- Youtie, J., P. Shapira, and G. Lauderman. 2002. "Transitioning to a knowledge economy: The LaGrange Internet TV Initiative." Paper presented to Telecommunications Policy Research Conference, Alexandria, VA. Retrieved from <http://intel.si.umich.edu/tprc/papers/2002/100/LGIAI-TPRC-2002.pdf>
- Zolnierek, J., J. Eisner and E. Burton. 2001. "An empirical examination of entry patterns in local telephone markets." *Journal of Regulatory Economics* 19, 2: 143-159.

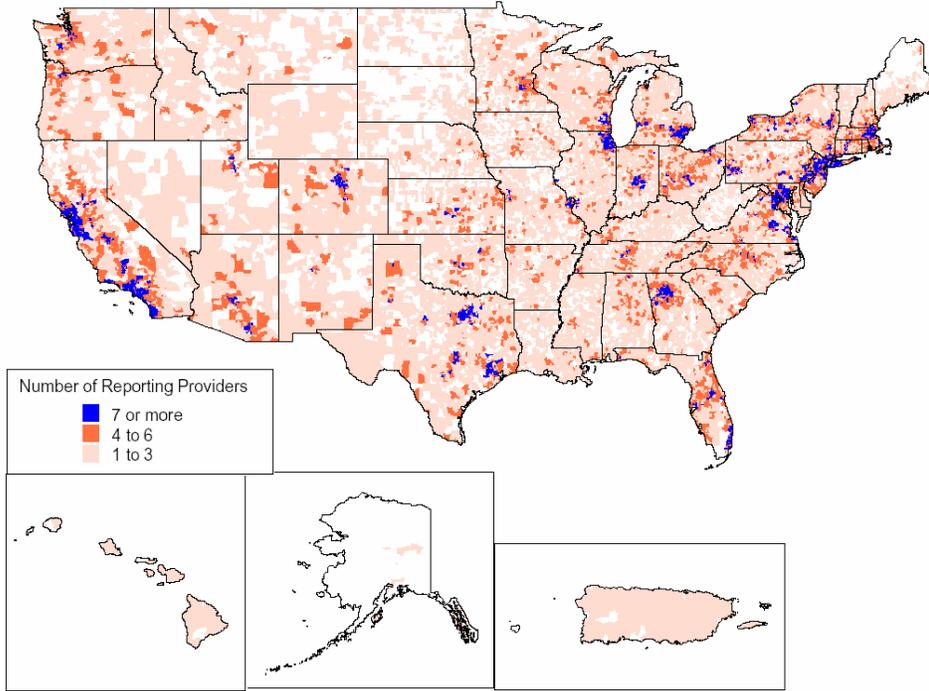
Appendix A

Diffusion of High-Speed Providers



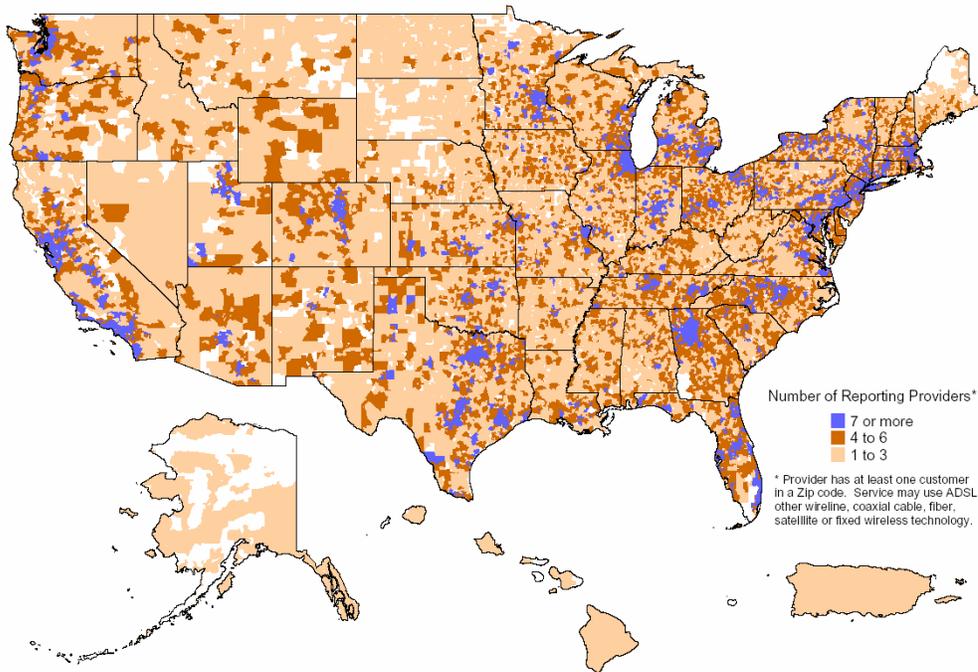
Source: FCC Form 477 Report dated October 2000

High-Speed Providers by Zip Code
(As of June 30, 2001)



Source: FCC Form 477 Report dated February 2002

High-Speed Providers by Zip Code
(As of June 30, 2003)



Source: FCC Form 477 Report dated December 2003

Appendix B
State-level Rates of DCT Infrastructure Capacity (Cable and DSL)
2000, 2001, and 2003

These numbers represent the percentage of rural / urban population within each state that had DSL or Cable access within their city (for DSL) or county (for Cable) of residence.

	2000		DSL 2001		2003		2000		Cable 2001		2003		Region
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	
Maine	0.00	0.00	1.07	1.32	6.10	5.76	11.41	15.33	10.76	15.33	57.61	87.92	New England
New Hampshire	0.00	0.00	0.00	0.00	6.86	1.46	11.42	32.85	11.42	32.85	71.52	73.16	
Vermont	0.00	0.00	1.10	5.58	8.01	5.58	0.00	0.00	0.00	0.00	45.68	53.81	
Massachusetts	0.00	0.00	0.00	6.84	0.00	9.40	0.00	27.19	0.00	27.19	58.39	92.24	
Rhode Island	0.00	0.00	0.00	0.00	0.00	19.55	0.00	70.80	0.00	70.80	0.00	76.32	
Connecticut	0.00	0.00	0.00	0.00	39.62	73.32	24.25	34.24	24.25	34.24	100.00	89.56	
New York	0.00	2.70	1.67	3.61	7.18	30.44	0.00	13.10	0.00	13.64	12.00	96.78	Middle Atlantic
New Jersey	0.00	0.00	0.00	0.00	0.00	3.61	0.00	32.85	0.00	32.85	0.00	73.16	
Pennsylvania	1.40	2.62	1.44	5.72	11.85	11.73	14.68	29.67	14.68	38.16	55.03	57.65	
Ohio	13.00	6.90	14.15	26.85	39.63	41.59	4.62	22.01	6.02	22.01	72.97	70.59	East North Central
Indiana	3.13	16.86	5.49	47.80	27.17	61.35	18.51	14.59	18.70	18.65	80.86	56.79	
Illinois	4.67	32.75	4.67	54.32	12.09	39.20	1.46	22.68	6.13	40.57	25.12	85.79	
Michigan	4.55	0.68	4.55	18.61	10.26	20.62	18.21	40.42	18.21	45.16	47.57	70.56	
Wisconsin	0.00	1.03	1.98	24.10	28.20	38.05	1.37	2.11	1.51	1.08	71.35	75.91	
Minnesota	0.85	0.00	1.96	5.04	9.38	9.43	5.76	4.03	5.90	4.16	32.32	100.00	
Iowa	0.07	0.00	0.66	0.74	15.95	9.36	0.44	53.20	3.17	53.84	37.73	73.16	
Missouri	6.37	19.13	6.36	24.48	33.02	46.69	0.00	3.59	0.07	7.04	18.38	29.13	
North Dakota	0.00	0.00	1.71	0.00	47.77	1.78	0.52	30.10	0.78	30.10	48.42	87.67	
South Dakota	0.00	0.00	9.07	0.00	28.36	0.79	17.38	42.77	19.13	45.78	59.36	100.00	
Nebraska	0.00	0.00	1.41	0.00	9.70	1.34	0.00	76.12	0.29	76.23	48.29	81.34	
Kansas	1.37	48.25	1.37	8.68	41.56	49.65	5.17	7.48	6.64	12.21	54.20	76.24	
Delaware	0.00	0.00	0.00	9.48	0.00	9.48	0.00	80.18	0.00	80.18	19.08	90.10	South Atlantic
Maryland	0.00	0.00	0.00	16.09	0.00	17.48	0.00	40.95	0.00	40.95	41.44	68.52	
DC	0.00	0.00	0.00	0.00	0.00	100.00	0.00	0.00	0.00	0.00	0.00	100.00	
Virginia	9.89	15.47	7.51	27.37	45.32	30.77	5.64	31.92	5.64	36.22	23.77	59.52	
West Virginia	0.00	0.00	0.00	0.33	6.71	39.65	0.00	0.00	0.00	21.34	36.81	61.88	
North Carolina	0.52	40.50	24.32	67.58	78.47	85.43	1.24	8.17	1.24	8.17	55.07	91.62	
South Carolina	3.82	24.35	4.33	45.86	59.78	61.44	8.45	15.33	8.45	15.89	93.14	57.68	
Georgia	0.00	29.08	2.16	51.68	70.86	75.17	1.30	46.11	6.62	46.47	29.74	53.68	
Florida	0.00	26.61	6.41	41.00	68.58	48.01	0.00	16.75	0.00	23.56	62.04	80.37	
Kentucky	7.78	19.20	8.91	35.84	60.90	53.88	0.73	60.14	0.73	60.47	15.67	70.24	
Tennessee	1.87	54.07	5.90	66.16	70.64	89.30	2.82	39.01	2.82	41.43	26.44	84.14	East South Central
Alabama	3.65	31.97	6.32	49.19	51.51	66.80	3.38	6.72	7.65	8.44	28.36	63.69	
Mississippi	1.16	23.02	1.16	37.35	73.75	72.70	4.26	7.45	7.23	11.78	28.05	56.23	
Arkansas	0.00	16.64	7.59	42.40	36.02	67.90	5.60	6.10	5.60	13.32	65.17	100.00	West South Central
Louisiana	3.56	38.63	6.90	52.98	74.39	74.14	0.00	33.47	0.00	34.54	47.30	70.16	
Oklahoma	0.00	56.88	3.36	50.51	30.91	56.63	0.88	0.00	0.88	0.91	31.18	93.67	
Texas	15.89	22.63	18.12	58.59	38.88	69.04	0.10	16.33	1.40	16.35	38.59	66.06	
Montana	0.00	0.00	14.47	0.00	25.44	26.51	0.49	0.00	0.49	0.00	10.74	55.94	Mountain
Idaho	2.56	5.80	2.56	5.80	8.66	5.80	0.32	0.00	0.32	0.00	49.45	54.01	
Wyoming	0.00	0.00	0.00	0.00	1.56	0.00	3.56	0.00	3.56	0.00	24.85	92.16	
Colorado	0.00	0.00	12.81	0.00	26.29	3.73	0.00	13.74	0.00	17.01	4.72	74.34	
New Mexico	0.00	0.00	0.00	0.00	21.12	0.00	0.00	28.81	0.00	28.81	12.00	70.53	
Arizona	0.00	0.09	0.00	0.09	8.02	9.49	2.19	0.00	1.10	2.68	48.95	100.00	
Utah	0.00	0.00	4.44	0.00	12.44	0.44	0.00	2.98	0.00	2.98	0.18	46.37	
Nevada	4.96	53.97	5.70	57.34	5.70	62.78	4.22	69.61	4.22	69.61	65.79	97.50	
Washington	11.09	16.47	25.01	23.60	39.34	27.05	44.51	31.72	44.51	52.32	84.62	73.60	
Oregon	6.73	16.49	20.93	18.25	40.38	28.46	1.13	29.50	17.27	34.89	77.09	78.41	
California	11.74	75.68	27.37	83.34	31.88	83.33	10.24	23.02	10.24	25.37	54.35	64.39	
Alaska	0.00	0.00	14.15	85.53	33.30	85.53	1.38	0.00	1.38	0.00	100.00	13.19	
Hawaii	29.16	75.07	29.16	75.07	29.16	75.07	0.00	96.07	0.00	96.07	34.07	98.44	

Appendix C

State-level Rates of Dial-up and High-speed Access: 2000, 2001, and 2003

	Dial-up						High-speed						Region
	2000		2001		2003		2000		2001		2003		
	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	Rural	Urban	
Maine	37.03	37.72	47.12	46.40	44.22	39.90	3.32	9.86	6.49	16.08	14.37	26.64	New England
New Hampshire	59.88	46.27	52.90	45.63	41.42	34.49	3.52	8.71	9.90	19.04	24.03	36.39	
Vermont	45.31	51.72	46.76	49.47	45.12	40.79	1.42	7.00	7.33	17.50	13.53	34.08	
Massachusetts	53.10	41.16	61.66	42.17	53.17	33.33	0.00	8.40	3.19	15.72	19.66	30.40	
Rhode Island	N/A	35.44	N/A	44.73	N/A	31.22	N/A	5.96	N/A	12.56	N/A	27.96	
Connecticut	N/A	51.46	N/A	45.02	N/A	36.37	N/A	4.85	N/A	14.55	N/A	30.21	
New York	36.66	37.46	36.14	41.55	36.52	31.72	5.59	5.38	6.52	14.55	21.42	27.12	Middle Atlantic
New Jersey	N/A	45.24	N/A	48.94	N/A	35.08	N/A	5.97	N/A	13.42	N/A	31.62	
Pennsylvania	37.85	42.00	44.14	44.31	39.18	38.34	0.73	4.07	8.63	9.34	9.38	23.44	
Ohio	32.81	40.25	51.61	44.57	42.87	37.88	1.50	4.91	3.00	9.25	9.17	20.82	East North Central
Indiana	34.10	40.64	40.29	49.08	43.08	40.77	3.67	4.36	2.88	6.58	6.01	14.95	
Illinois	31.02	42.57	40.25	45.75	34.01	36.51	0.00	3.32	0.59	7.95	11.90	22.09	
Michigan	41.46	40.12	43.05	43.37	36.32	33.96	2.45	6.13	1.82	11.78	5.78	24.88	
Wisconsin	30.05	42.83	44.21	47.72	42.14	37.54	2.48	2.29	5.14	7.02	12.90	24.63	
Minnesota	27.94	45.18	39.69	53.98	36.75	45.69	3.20	5.62	3.64	9.96	14.21	24.15	West North Central
Iowa	29.02	40.98	42.40	47.19	38.32	42.55	2.60	6.23	6.67	13.24	14.41	25.90	
Missouri	32.71	41.90	48.01	43.33	35.68	41.30	1.78	7.61	4.67	10.13	7.76	22.39	
North Dakota	30.82	40.65	39.35	44.16	40.00	31.63	1.42	3.47	1.43	9.51	14.05	25.29	
South Dakota	31.92	40.73	32.79	47.54	30.80	32.48	1.97	4.39	8.57	14.28	19.31	30.85	
Nebraska	26.70	41.64	30.08	40.37	37.44	34.00	3.97	5.57	3.30	15.76	9.62	31.89	
Kansas	33.02	44.41	43.65	41.78	40.35	29.07	1.12	7.50	2.71	15.62	12.84	30.75	
Delaware	39.29	46.39	38.67	49.97	44.02	43.23	6.19	6.50	1.33	9.70	8.82	19.31	South Atlantic
Maryland	N/A	41.13	N/A	52.00	N/A	41.51	N/A	5.11	N/A	11.20	N/A	24.18	
DC	N/A	36.71	N/A	37.33	N/A	36.83	N/A	4.15	N/A	8.51	N/A	24.24	
Virginia	30.07	48.02	41.81	52.99	43.18	45.46	0.93	3.50	5.65	8.63	11.19	24.00	
West Virginia	25.74	37.61	34.83	42.88	35.48	34.97	2.03	5.04	2.48	6.58	10.53	21.61	
North Carolina	31.04	35.93	31.47	41.14	35.18	32.39	3.93	4.14	5.97	8.90	11.91	24.63	
South Carolina	25.35	30.48	34.76	41.63	31.32	30.13	0.00	5.21	1.43	10.97	9.90	20.77	
Georgia	17.90	42.67	28.09	48.39	29.41	36.59	0.92	3.70	3.13	10.07	10.94	25.61	
Florida	33.08	39.54	44.01	45.26	30.93	37.51	2.95	6.16	1.13	12.03	4.97	22.20	
Kentucky	26.35	41.58	42.43	46.44	39.37	45.06	3.72	6.01	2.65	3.91	8.54	19.47	
Tennessee	23.12	41.28	27.53	40.59	29.11	30.57	3.94	5.23	9.33	11.39	11.61	25.41	
Alabama	24.37	38.73	23.59	38.49	29.82	32.16	1.24	2.65	4.58	6.66	9.26	19.14	
Mississippi	20.17	31.58	27.12	46.97	31.68	30.98	5.53	4.28	7.36	6.25	8.39	17.71	
Arkansas	21.63	29.16	28.81	38.05	28.82	33.07	1.92	2.84	3.53	4.92	13.27	14.61	West South Central
Louisiana	31.62	30.63	38.31	36.29	35.08	34.25	0.94	5.03	1.64	7.46	3.26	14.44	
Oklahoma	24.45	36.63	30.60	47.98	34.61	34.15	4.44	7.47	2.97	11.58	10.55	17.88	
Texas	24.47	38.06	33.50	41.23	34.53	35.08	1.84	5.55	1.07	11.79	8.30	22.39	
Montana	41.80	38.11	46.90	47.12	40.77	38.90	1.60	4.04	2.74	6.42	9.24	14.90	Mountain
Idaho	38.77	47.25	48.62	45.97	45.88	44.21	1.99	3.75	5.03	13.22	13.39	18.03	
Wyoming	43.07	45.56	44.03	56.40	44.22	48.04	1.99	1.48	6.76	6.63	13.38	14.41	
Colorado	40.75	49.05	43.80	50.19	51.08	40.34	2.64	5.89	6.22	11.71	10.00	27.03	
New Mexico	26.84	36.48	31.25	49.73	39.77	42.06	1.41	6.67	2.70	3.18	4.37	10.76	
Arizona	30.22	40.31	36.37	40.08	32.17	31.76	0.00	7.40	6.55	15.45	16.81	25.87	
Utah	47.20	46.37	40.91	44.56	39.33	46.65	3.20	5.33	5.38	15.08	14.42	23.84	
Nevada	47.30	34.56	47.72	48.59	37.84	32.77	2.09	6.20	3.21	8.04	9.01	28.82	
Washington	27.23	47.91	42.41	50.58	40.12	36.00	3.63	6.53	1.17	15.89	13.66	32.91	Pacific
Oregon	40.64	56.32	44.71	49.96	44.74	46.27	2.40	3.96	8.30	12.49	12.58	22.53	
California	42.54	43.48	35.25	44.20	40.93	36.09	4.72	6.51	4.98	15.63	10.50	29.30	
Alaska	46.97	55.90	53.79	52.07	50.17	40.60	3.66	7.60	8.85	19.17	19.59	30.31	
Hawaii	39.15	40.46	44.84	34.09	38.81	23.41	0.00	8.63	7.30	26.29	27.09	32.27	

Appendix D

Derivation of Partial Effects for Logit and Probit Models⁸⁵

Logit and probit models deal with the probability of an independent variable y being equal to 1. These models take the form

$$P(y = 1 | x) = G(x\beta) \equiv p(x)$$

where x is a $1 \times K$ vector of dependant variables with the first element assumed to be unity, and β is $K \times 1$ vector of parameters. $G(\cdot)$ is restricted to the unit interval $[0,1]$. These models are often referred to as index models because the probability of $y = 1$ is only a function of x through the index $x\beta = \beta_1 + \beta_2 x_2 + \dots + \beta_K x_K$. Hence, the function $G(\cdot)$ maps the index into the unit interval.

For the specific cases of the logit and probit, we can define $G(\cdot)$ as follows:

Probit (derived from standard normal distribution)

$$G(z) \equiv \Phi(z) \equiv \int_{-\infty}^z \phi(v) dv$$

where $\phi(z)$ is the standard normal density

$$\phi(z) = (2\pi)^{-1/2} \exp(-z^2 / 2)$$

Logit (derived from standard logistic distribution)

$$G(z) \equiv \Lambda(z) \equiv \exp(z) / [1 + \exp(z)]$$

The partial effects for both logit and probit are derived in the following manner:

$$\frac{\partial p(x)}{\partial x_j} = g(x\beta)\beta_j, \quad \text{where } g(z) \equiv \frac{dG}{dz}(z)$$

Hence, the partial effect of x_j on $p(x)$ depends on both β_j and how x is affected through $G(\cdot)$. Estimating the effects of the variables x_j on $p(x)$ is performed differently for continuous or binary x_j .

For Continuous x_j : $\Delta \hat{P}(y = 1 | x) \approx [g(x\hat{\beta})\hat{\beta}_j] \Delta x_j$

For Binary x_j : $\Delta \hat{P}(y = 1 | x) = G[\hat{\beta}_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_{j-1} x_{j-1} + \hat{\beta}_j + \hat{\beta}_{j+1} x_{j+1} + \dots + \hat{\beta}_K x_K] - G[\hat{\beta}_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_{j-1} x_{j-1} + \hat{\beta}_{j+1} x_{j+1} + \dots + \hat{\beta}_K x_K]$

Hence, the binary specification looks at the change in probability associated with varying x_j from 1 to 0.

Typically, the other variables are evaluated at their sample averages to estimate the effect on the "average" household.

⁸⁵ The above discussion was taken primarily from Wooldridge (2002), chapter 15 – Discrete Response Models

Appendix E
Logit Regressions for Rural – Urban Internet Access (1997, 1998, 2000, 2001)

Table 18 presented logit results on Internet access in 2003 when parameter estimates for various explanatory variables were allowed to differ between rural and urban areas. This appendix presents similar results for the years 1997, 1998, 2000, and 2001.

1997

Variables	Urban		Rural	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.7870	0.1535 ***	1.1673	0.4892 **
scoll	1.7301	0.1503 ***	0.9808	0.4878 **
coll	2.0121	0.1530 ***	0.8706	0.4969 *
collplus	2.1711	0.1565 ***	1.0810	0.5049 **
faminc1	-0.4031	0.2188 *	-0.2755	0.5690
faminc2	-0.0885	0.2157	0.2729	0.5032
faminc3	-0.3631	0.2102 *	0.0142	0.4866
faminc4	-0.3478	0.2107 *	0.3707	0.4622
faminc5	-0.1056	0.1718	-0.0186	0.4204
faminc6	0.0351	0.1531	0.0892	0.3822
faminc7	0.1306	0.1528	0.1875	0.3768
faminc8	0.1411	0.1487	0.2423	0.3716
faminc9	0.2771	0.1486 *	0.1458	0.3828
faminc10	0.4283	0.1408 ***	-0.0304	0.3661
faminc11	0.4794	0.1424 ***	0.2292	0.3686
faminc12	0.7474	0.1420 ***	0.2721	0.3763
faminc13	0.9591	0.1389 ***	0.0167	0.3708
netatwork	1.1845	0.0458 ***	0.1701	0.1302
black	-0.7920	0.0839 ***	-0.1303	0.3442
othrace	-0.3459	0.0923 ***	0.2558	0.3366
hisp	-0.5886	0.0965 ***	-0.2986	0.3660
peage	0.0143	0.0101	0.0180	0.0275
age2	-0.0005	0.0001 ***	-0.0002	0.0003
sex	0.3293	0.0455 ***	-0.0990	0.1265
married	0.1629	0.0475 ***	0.0926	0.1364
regdensity	3.8773	0.5265 ***	3.3063	1.1445 ***
retired	0.2870	0.1120 **	-0.1534	0.2948
constant	-4.2446	0.2888 ***	-1.8550	0.8130 **
Log-likelihood	-12095.9			

Note: ***, **, and * represent statistically significant differences from zero at the $p = 0.01$, 0.05 , and 0.10 levels, respectively. Rural coefficients represent shifts on urban coefficients.

1998

Variables	Urban		Rural	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.6078	0.0872 ***	-0.0431	0.1884
scoll	1.2594	0.0866 ***	0.0205	0.1899
coll	1.5915	0.0901 ***	0.0426	0.2035
collplus	1.7376	0.0951 ***	-0.1377	0.2278
faminc1	-0.2552	0.1858	0.6688	0.4409
faminc2	-0.0618	0.1771	0.3110	0.4617
faminc3	-0.0326	0.1643	0.4699	0.3979
faminc4	-0.1134	0.1713	0.2434	0.4038
faminc5	-0.0130	0.1455	0.6547	0.3540 *
faminc6	0.0665	0.1381	0.4737	0.3491
faminc7	0.2917	0.1348 **	0.4896	0.3439
faminc8	0.4363	0.1339 ***	0.4833	0.3446
faminc9	0.5869	0.1335 ***	0.4154	0.3466
faminc10	0.7543	0.1280 ***	0.7481	0.3324 **
faminc11	1.0527	0.1288 ***	0.4494	0.3356
faminc12	1.1424	0.1301 ***	0.5644	0.3437
faminc13	1.4749	0.1274 ***	0.3884	0.3411
netatwork	0.1715	0.0432 ***	-0.1486	0.1140
black	-0.8360	0.0652 ***	-0.2691	0.2477
othrace	0.0126	0.0813	-0.2355	0.2250
hisp	-0.6801	0.0754 ***	-0.2290	0.2455
peage	0.0372	0.0079 ***	0.0319	0.0211
age2	-0.0006	0.0001 ***	-0.0003	0.0002
sex	0.1552	0.0362 ***	-0.0741	0.0919
married	0.2995	0.0396 ***	0.1645	0.1033
regdensity	2.4414	0.3184 ***	2.0807	0.5865 ***
retired	-0.0796	0.0812	0.1188	0.1974
constant	-3.8837	0.2344 ***	-1.6897	0.5838 ***
Log-likelihood	-16707.9			

Note: ***, **, and * represent statistically significant differences from zero at the $p = 0.01$, 0.05 , and 0.10 levels, respectively. Rural coefficients represent shifts on urban coefficients.

2000

Variables	Urban		Rural	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.6586	0.0678 ***	0.0381	0.1494
scoll	1.2301	0.0683 ***	0.2377	0.1522
coll	1.5098	0.0730 ***	0.3276	0.1714 *
collplus	1.6684	0.0810 ***	0.1834	0.1968
faminc1	-0.2219	0.1664	-0.1125	0.3498
faminc2	-0.2015	0.1662	0.0537	0.3459
faminc3	-0.0723	0.1480	-0.2940	0.3154
faminc4	0.1893	0.1426	0.1369	0.2989
faminc5	0.1786	0.1281	0.1178	0.2624
faminc6	0.2418	0.1214 **	0.0107	0.2553
faminc7	0.4891	0.1195 ***	-0.0446	0.2549
faminc8	0.6744	0.1200 ***	0.0964	0.2554
faminc9	0.6830	0.1206 ***	0.0099	0.2613
faminc10	0.9502	0.1162 ***	0.0318	0.2508
faminc11	1.0902	0.1173 ***	0.0699	0.2571
faminc12	1.3612	0.1183 ***	-0.0046	0.2636
faminc13	1.7332	0.1149 ***	-0.0292	0.2596
netatwork	0.0823	0.0423 *	-0.1675	0.1078
black	-0.8266	0.0544 ***	0.0637	0.1828
othrace	-0.0892	0.0767	-0.4109	0.2229 *
hisp	-0.7246	0.0611 ***	-0.0838	0.1980
peage	0.0383	0.0069 ***	0.0256	0.0161
age2	-0.0007	0.0001 ***	-0.0002	0.0002
sex	0.0097	0.0339	-0.0858	0.0793
married	0.4385	0.0369 ***	0.1782	0.0884 **
chld1	-0.0994	0.0540 *	0.1209	0.1190
chld2	-0.0224	0.0547	-0.0421	0.1246
chld3	-0.1547	0.0861 *	0.1277	0.1931
chld4	0.2320	0.1541	0.0846	0.3543
chld5	-0.0893	0.1105	-0.2658	0.2832
regdensity	2.6155	0.3377 ***	0.7857	0.5302
cableaccess	0.0053	0.0964	0.4126	0.4572
dslaccess	0.1044	0.0676	0.3547	0.6874
retired	0.0332	0.0726	-0.1184	0.1605
constant	-3.6178	0.2445 ***	-1.1119	0.4837 **
Log-likelihood	-18176.9			

Note: ***, **, and * represent statistically significant differences from zero at the $p = 0.01$, 0.05 , and 0.10 levels, respectively. Rural coefficients represent shifts on urban coefficients.

2001

Variables	Urban		Rural	
	Coefficient	S.E.	Coefficient	S.E.
hs	0.6164	0.0598 ***	-0.0030	0.1211
scoll	1.1420	0.0606 ***	0.0268	0.1254
coll	1.4396	0.0681 ***	-0.1152	0.1491
collplus	1.5281	0.0779 ***	0.1824	0.1769
faminc1	-0.2002	0.1447	-0.0631	0.2790
faminc2	-0.3049	0.1434 **	0.4118	0.2893
faminc3	-0.0548	0.1277	0.4782	0.2504 *
faminc4	-0.0575	0.1299	0.0569	0.2603
faminc5	0.0682	0.1151	0.0100	0.2338
faminc6	0.2460	0.1080 **	0.2012	0.2218
faminc7	0.4271	0.1082 ***	0.1017	0.2228
faminc8	0.6822	0.1084 ***	0.1446	0.2240
faminc9	0.7354	0.1096 ***	-0.0588	0.2281
faminc10	0.9306	0.1053 ***	0.1249	0.2212
faminc11	1.1393	0.1068 ***	0.0898	0.2256
faminc12	1.3867	0.1090 ***	0.0191	0.2318
faminc13	1.7991	0.1053 ***	-0.0343	0.2268
netatwork	0.4646	0.0403 ***	0.1255	0.0895
black	-0.7730	0.0493 ***	0.2800	0.1505 *
othrace	0.0758	0.0815	-0.4561	0.2053 **
hisp	-0.6947	0.0593 ***	0.1118	0.1865
peage	0.0397	0.0063 ***	0.0243	0.0134 *
age2	-0.0006	0.0001 ***	-0.0002	0.0001
sex	-0.0239	0.0335	0.0022	0.0721
married	0.5454	0.0376 ***	0.1915	0.0800 **
chld1	0.1910	0.0493 ***	0.1227	0.1042
chld2	0.3219	0.0539 ***	0.1541	0.1145
chld3	0.2850	0.0793 ***	-0.0416	0.1611
chld4	0.1195	0.1227	0.1691	0.2569
chld5	0.2179	0.2226	-0.4624	0.4540
regdensity	2.2523	0.3478 ***	0.5449	0.5133
cableaccess	-0.1128	0.0907	0.1235	0.3736
dslaccess	0.0568	0.0612	0.2765	0.4303
retired	0.0312	0.0656	0.1481	0.1346
constant	-3.4354	0.2587 ***	-1.2132	0.4528 ***
Log-likelihood	-20687.6			

Note: ***, **, and * represent statistically significant differences from zero at the $p = 0.01$, 0.05 , and 0.10 levels, respectively. Rural coefficients represent shifts on urban coefficients.

Appendix F

Results from Logit Regressions – Rural and Urban Samples

As noted in section 3.5, the choice of which parameter set to use can affect the results of a linear or non-linear decomposition. The primary results reported in section 4.1 are based on “pooled” regressions, which include both rural and urban households. This Appendix reports similar results when only rural (or urban) samples are used. The pooled results discussed in section 4.1 are also shown for comparison purposes.

Internet access rates	Contributions from Differences in:	RURAL PARAMS			URBAN PARAMS			POOLED PARAMS		
		Rate	Decrease	% of Gap	Rate	Decrease	% of Gap	Rate	Decrease	% of Gap
2003	Urban	61.23			61.23			61.23		
	Education		2.60	20.3%		2.83	22.1%		2.79	21.8%
	Income		4.03	31.4%		4.44	34.6%		4.34	33.8%
	Other HH Characteristics		0.46	3.6%		-0.02	-0.1%		0.10	0.8%
	Network Externalities		4.55	35.5%		3.68	28.7%		3.76	29.3%
	DCT Infrastructure		0.02	0.1%		0.53	4.1%		0.35	2.7%
	All Variables		11.66	90.9%		11.46	89.3%		11.34	88.4%
	Rural	48.40			48.40			48.40		
2001	Urban	56.57			56.57			56.57		
	Education		2.27	16.6%		2.28	16.7%		2.29	16.8%
	Income		4.33	31.8%		4.98	36.5%		4.89	35.9%
	Other HH Characteristics		0.12	0.9%		0.01	0.1%		-0.01	-0.1%
	Network Externalities		5.30	38.9%		4.31	31.6%		4.66	34.2%
	DCT Infrastructure		1.71	12.5%		-0.07	-0.5%		-0.01	-0.1%
	All Variables		13.73	100.7%		11.51	84.5%		11.82	86.7%
	Rural	42.94			42.94			42.94		
2000	Urban	46.67			46.67			46.67		
	Education		3.25	23.6%		2.56	18.6%		2.64	19.2%
	Income		4.57	33.2%		4.84	35.2%		4.87	35.4%
	Other HH Characteristics		-0.85	-6.2%		-0.10	-0.7%		-0.21	-1.5%
	Network Externalities		6.42	46.6%		5.01	36.4%		5.53	40.1%
	DCT Infrastructure		2.87	20.8%		0.44	3.2%		0.49	3.5%
	All Variables		16.26	118.1%		12.76	92.6%		13.31	96.7%
	Rural	32.90			32.90			32.90		
1998	Urban	30.44			30.44			30.44		
	Education		2.15	17.3%		1.87	15.1%		2.19	17.7%
	Income		3.55	28.6%		3.62	29.2%		3.81	30.7%
	Other HH Characteristics		-0.24	-1.9%		-0.12	-0.9%		0.24	1.9%
	Network Externalities		6.26	50.5%		3.53	28.5%		4.02	32.5%
	DCT Infrastructure		11.72	94.6%		8.90	71.8%		10.26	82.8%
	All Variables		18.05					18.05		
1997	Urban	17.69			17.69			17.69		
	Education		1.11	12.9%		0.78	9.0%		1.03	12.0%
	Income		1.66	19.3%		1.61	18.7%		1.73	20.1%
	Other HH Characteristics		2.20	25.6%		1.82	21.1%		2.07	24.1%
	Network Externalities		4.57	53.1%		2.49	29.0%		2.72	31.6%
	DCT Infrastructure		9.54	110.9%		6.70	77.8%		7.56	87.8%
	All Variables		9.08					9.08		
	Rural	9.08			9.08			9.08		