Broadband Adoption and Content Consumption

Lorin Hitt • Prasanna Tambe*
Wharton School, University of Pennsylvania
Philadelphia, PA 19014
lhitt@wharton.upenn.edu tambe@wharton.upenn.edu

Abstract

We explore how broadband access drives changes in the quantity and diversity of consumption of online content by using panel data that describes household Internet usage before and after broadband adoption. Our data suggests that on average, broadband adoption increases usage by over 1300 minutes per month. We also find that information consumption becomes more evenly distributed within the population, driven in part by post-adoption usage gains of almost 1800 minutes per month among individuals who were in the lowest usage quintile before adopting broadband. After adopting broadband, this pre-adoption lowest-usage quintile consumes content in greater quantities than users in neighboring quintiles, passing both the second and third quintiles in terms of absolute usage. This suggests that these users may have had strong preferences for high-bandwidth content that was too costly to consume in a narrowband environment. We also show that broadband adoption increases the variety of content that users consume although many of these gains appear to be associated with an increase in the variety of sites visited within previously visited content categories rather than an expansion in the types of content consumed.

Keywords: digital divide, broadband adoption, telecommunications policy, media, Internet

JEL Classification: L9, O1, O3

Thanks to valuable assistance and feedback from Janet McCabe, Andrea Meyer, Joel Waldfogel, Wharton Research Data Services, two anonymous reviewers, and participants at the 34th Annual Telecommunications Policy Research Conference, September 29-October 1, 2006, Arlington, Virginia, USA. * Corresponding author. email: tambe@wharton.upenn.edu. Tel: 215.898.1217. Fax: 215.898.3664. Mail: Suite 500, Jon M. Huntsman Hall, 3730 Walnut Street, Philadelphia, PA 19104.

I. Introduction

Policy makers, media companies, marketers, and economists have closely watched the deployment of broadband networks because access to these technologies can significantly change the types and amounts of information to which people have access. Developing a deeper understanding of how broadband adoption changes information consumption is of interest for several reasons. First, consumption of online content comes at the expense of other established media channels such as television or print publishing because broadband technologies greatly improve the level of access to websites featuring video, image, and music content. If broadband adoption is merely shifting the medium of content consumption -- acting as a substitute for more traditional entertainment media -- then the impacts on these industries should be considered when developing broadband policy. Second, broadband may increase consumers' access to product information, product reviews, and specialty retailers -- a fact of interest to marketers, businesses, and economists. Third, the distribution of usage within the citizenry may be of interest to policy makers who are interested in ensuring access to information about jobs, health, and the political process. Municipal broadband programs, for example, will do little to alleviate the "digital divide" between information haves and have-nots if the programs primarily increase consumption among individuals who were already heavy users of narrowband technologies. Therefore, developing an understanding of how broadband adoption drives changes in content consumption is an important step in understanding how the rollout of these networks will ultimately impact users.

The economics literature reflects the extensive public interest in broadband-related issues. Most existing studies, however, have focused on the availability and diffusion of broadband networks. By contrast, this study investigates the causal link between broadband adoption and changes in household usage patterns by using disaggregated Internet usage data from a panel of users sampled in 2002 and 2004. Because we have connection speed information and browsing behavior information, we can compare before-and-after usage data for individuals who switched from narrowband to broadband between 2002 and 2004, using the behavior of non-adopters to control for confounding influences through a differences-in-differences estimation approach. One of the challenges with a straightforward differences-in-differences approach, however, is that individuals self-select into the broadband pool. If people who switch to broadband have stronger preferences for online content than those who do not, they may have had higher usage numbers in 2004 even in the absence of broadband adoption. To account for these self-selection issues, we use a matching estimator to control for factors that may otherwise bias our estimates, such as intrinsic preferences for online content and differences in the opportunity costs of time.

We begin by estimating increases in usage (measured as time spent online) resulting from broadband adoption, and then explore how access to broadband networks changes the distribution of usage across adopters and how these changes relate to differences in preferences for content. We also investigate how broadband adoption affects the diversity of content that people consume, where diversity is measured by the number of different sites that users visit and the number of different types of content they access. We find that after accounting for the overall usage declines that occurred in our panel, broadband increased the amount of time spent online by over 1300 minutes per month. These increases lead to a more even distribution of usage across the population because larger increases in consumption come from individuals who were at the bottom of the usage spectrum when they only had narrowband access. Somewhat surprisingly, we find that absolute usage in this bottom group surpasses the absolute usage of the neighboring two quintiles after broadband adoption. Some of these shifts can be traced to large increases in the amount of time spent on sites that provide high-bandwidth content, such as entertainment, advertising, and downloading applications for images, music, and online games. Apparently, broadband access satisfies demand for high-bandwidth content that is prohibitively expensive to access in a narrowband environment where the opportunity costs of waiting for content to load are much higher. Finally, we explore how broadband impacts the diversity of content consumption, based on a classification of our website data into twenty-seven different categories such as news, business, entertainment, and sports. We find that users who adopt broadband visit a greater number of websites, but that the distribution of these visits across categories is less evenly spread. The rest of our paper proceeds as follows. In Section II, we review the existing literature on the consumer benefits of broadband, the digital divide, and household demand for broadband access. Section III discusses our data and empirical strategies, and Section IV describes our findings. Section V describes some of the challenges we faced with the data and presents some auxiliary analyses which confirm that our results are robust. The conclusions are presented in Section VI.

II. Literature

Researchers have examined broadband policy from a number of angles. From a commercial perspective, interest in this topic stems from the economic growth potential that has been linked with widespread broadband adoption (Crandall and Jackson, 2001). However, widespread broadband adoption also leads to a number of direct benefits for consumers. Access to online services can mean better information about jobs, education, and health (Brodie et al., 2000; Autor, 2001). Furthermore, broadband access can improve quality of life through greater convenience and increased involvement with civic, government, and community organizations (Norris, 2001). By improving access to information and decreasing the costs of communication, broadband promises to connect communities and provide substantial economic benefits to a large portion of the population. Many of these benefits, however, are contingent upon the adoption of broadband technologies. Though broadband penetration is on the rise, the distribution of availability is still a cause for concern among policy makers. A number of researchers have documented the emergence of a digital divide separating households that have regular access to online information from those that do not. If a divide is allowed to persist, broadband-related benefits will accrue disproportionately to certain segments of the population. The drivers of this divide have been explored in several studies. Researchers have collected evidence on the under-provision of broadband to rural households, economically disadvantaged regions, and areas populated primarily with racial minorities (Parker, 2000; Prieger, 2003; Hoffman and Novak, 1998). Studies have also explored how geography is related to consumer diffusion of broadband technologies using household data and diffusion analyses (Greenstein and Prince, 2006).

Because of the fixed costs involved with deploying these networks, availability of broadband technologies is often determined by estimates of local demand. Researchers have used different approaches to characterize the determinants of household demand for broadband access. In the INDEX project, researchers used controlled experiments to gauge consumers' willingness to pay for higher bandwidth. They find that the subjects in their study were willing to pay relatively little for higher bandwidth, but suggest that this may in part be due to the lack of compelling broadband applications that were available at the time of the experiment (Varian, 2002). Other studies have looked at demographic factors that influence household demand for bandwidth. Madden and Simpson (1996) show that socioeconomic variables are related to interest in network subscription, so that there is strong potential for a disadvantaged, information-poor class to develop. Rappoport et al. (2001) use clickstream data from ten cities to determine what demographic and usage factors distinguish narrowband households from broadband households. They find that demographic characteristics alone do not provide a clear distinction between broadband households and narrowband households, but that prior usage of narrowband services and the opportunity cost of time are good predictors of broadband adoption.

As broadband penetration continues to rise and governments step in to fill market gaps in broadband provision, some researchers have turned their attention from availability towards usage. The availability of various complements to broadband technologies, such as consumer equipment, social support, and skills, have been shown to drive large variations in household usage (Dimaggio and Hargittai, 2001), as have demographic and other individual factors (Kraut et al., 1996). In a recent study drawing on survey data, Goldfarb and Prince (2006) find that higher-income people are more likely to have adopted Internet access technologies, but conditional on adoption, lower-income people are likely to spend more time online. They also suggest that if provided with Internet access, non-adopters will use the Internet for many of the purposes intended by policy initiatives, such as telemedicine and e-government. Not all of the changes linked to consumer adoption of broadband, however, have been positive. A recent line of research, focused on the connection between illegal file-sharing activities and music sales, has connected broadband usage (through the use of these illegal file-sharing platforms) to a decline in music sales (Hong, 2004; Liebowitz, 2006). These studies generally infer the connection between broadband access and usage of high-bandwidth services rather than test it directly. In this study, we give an empirical foundation to some of these types of studies by looking at how broadband adoption drives changes in the consumption of different types of content, and we estimate the magnitudes of the overall usage increases driven by broadband adoption. In the next section, we describe our data set and the methods that we use in our analysis.

III. Data and Empirical Methods

A. Data

We use data from a panel consisting of the October 2002 and October 2004 disaggregated Internet usage of approximately 8100 households, including all of their website activity.[1] Although the disaggregate nature of the data is generally an asset, the data is captured at the household machine level and does not include workplace Internet usage by household members. Therefore, we are forced to make the assumption that the online activity that we observe is representative of the total Internet activity of that household. The data also include information on connection speed, from which we extracted the 5,497 households who either retained narrowband access in both 2002 and 2004 *(non-adopters)*, or who upgraded from narrowband in 2002 to broadband in 2004 (*adopters)*. Finally, we drop users who did not spend any time at all online in one of our two sample months, leaving us with 4,173 households. For some analyses we also include an additional 1,677 households that retained broadband access in both 2002 and 2004 (*maintainers*) and 628 households that downgraded from broadband in 2002 to narrowband in 2004 (*downgraders*).

For each of these households, we have demographic information, connection speed, and detailed session information for the entire month, including domain name level information for each website visited, duration of visit, and the number of pages viewed during the visit. In addition, websites are

[1] Source: comScore Networks. comScore Networks is a private market research firm that collects detailed data on household Internet usage. Each participating household in the comScore panel has an application installed on their computer which tracks Internet usage at a detailed level (web page domain location, time-of-access, duration-of-access) as well as other information about the machine. comScore utilizes these data to produce estimates of online usage patterns as well as for other studies of online user behavior conducted internally or by their clients. We were provided a subset of the fields in their panel through Wharton Research Data Services (WRDS).

grouped into one of twenty-seven different categories. To provide a sense for this categorization, Table 1 lists a representative website for each of these categories and reports the time spent online for each category by adopters and non-adopters. [Table 1 about here] These categories are available for the 2002 session data but not for the 2004 session data, so we construct a mapping of domain names to site categorizations using the 2002 data, and then use this mapping to apply categories to the 2004 data. In a later section, we describe robustness checks that we conduct to ensure that this mapping does not introduce errors into our data that may bias our results. Consistently across most categories, non-adopters spend less time online than adopters. The most popular category in our data set is portals, which includes the traditional portal sites (e.g., AOL, MSN, Yahoo!) as well as independent search engines (e.g., Google). Users also spend significant amounts of time on websites classified as entertainment, adult sites, advertising, and "helper" services that save passwords and information. The largest differences between the adopting and non-adopting populations exist in their consumptions of entertainment, sports and advertising. While most other categories are self-explanatory, we focus attention later in the paper on the analysis of the "Computer Applications" category, which are applications that allow users to download images, music, and other media. For many of these computer applications, such as file-sharing platforms, we do not observe the usage time spent after they have been downloaded, so download time is only a proxy for total usage (and therefore is likely to understate the usage of these applications).

Our demographic data include age, income, education, household size, census region, and whether or not a child is present in the house. All demographic variables are coded in discrete levels. When multiple users are present in the household, demographic information is based on the head of the associated household. Descriptive statistics for the 2002 data are shown in Table 2, grouped by adopters and non-adopters. For each demographic variable, Table 2 also shows Pearson chi-squared statistics, suggesting that broadband adoption is only significantly associated with age and household size. [Table 2 about here] Table 3 presents a logit analysis of how broadband adoption is affected by demographic variables. Users are more likely to shift from narrowband to broadband if they are younger, and if they are in larger households.

Table 4 presents consumption changes in terms of both usage minutes and page views for 2002 and 2004 for each of the four categories of households (non-adopters, adopters, downgraders, and maintainers). The core of our sample is formed by users who stayed with narrowband in both time periods (non-adopters), and users who switched from narrowband to broadband (adopters). For comparison, we also include usage numbers for users who had broadband in both time periods (maintainers), and those who downgraded from broadband to narrowband (downgraders).

Of these four groups, the only group that increased its average usage is the group composed of users who upgraded from narrowband to broadband, while usage dropped for both groups that maintained the same connection speed in both sample periods. Significantly, users who switched from narrowband in 2002 to broadband in 2004 had much higher mean Internet usage compared to those who did not switch. This suggests that they had stronger intrinsic preferences for Internet usage, and confirms that individuals select into the broadband pool because they expect higher benefits. To ensure that changes to the duration variable are not simply reflecting the effects of "always-on" connections, we also consider the impacts of broadband adoption on an alternative dependent variable, page views. [Table 4 about here] Table 4 shows that by either consumption metric, users who switched to broadband significantly increased information consumption over the two-year period. This effect is more pronounced when compared to individuals in the other three groups, who decreased their aggregate usage over the two-year period. Figure 1 examines changes in the aggregate consumption of content by individuals who adopted broadband between 2002 and 2004, segmented by 2002 usage quintiles. [Figure 1 about here] First, the figure suggests that total usage in 2002 was concentrated among the heaviest usage quintile but became more evenly distributed in 2004 because of reductions in time spent online among the heaviest users and increases at the bottom end of the distribution. Second, it shows unexpectedly large increases in usage in the lowest quintile of usage after broadband adoption. These formerly extremely light-usage households pass the neighboring quintiles in total usage after adopting broadband. Although it is tempting to interpret these trends as evidence linking broadband adoption with increased Internet usage, a number of other influences, such as time trends, may be driving these changes. In our empirical analysis below, we rule out the effects of time trends, opportunity cost explanations, and intrinsic preferences for online content to isolate the impact of broadband adoption.

B. Empirical Strategy

In this section, we describe our strategy for isolating the effect of broadband on information consumption behaviors. We have usage data before and after the adoption decision, so differences within groups (adopters versus non-adopters) give us some information about how broadband adoption impacts overall usage. However, because our panel extends over a two-year window, these effects will be confounded with other general time trends over the two years that could change usage pattern irrespective of connection speed. In addition, our data is not a true experiment, but an observation of consumption differences that follow households' voluntary choice to adopt broadband. The control group we use for our comparison is the set of households who, in 2002, had narrowband (i.e., dial-up) connections but had not yet adopted broadband. In general, this group may be different than an equivalent group of narrowband users in 2006 or other years, because connection speed is determined by an endogenous choice between narrowband and broadband, given the state of the technologies at the time. Although this may influence the generalizability of some of our specific estimates, the trends that we derive using this data should be extensible to other contexts. As shown in Table 4, online usage for eventual broadband adopters was considerably higher even before adopting broadband, suggesting that individuals who switched to broadband have stronger preferences for online content. To the extent that broadband adopters generally have a greater demand for content, it is reasonable to believe that these users may have changed their usage patterns in different ways from non-adopters, confounding any simple comparisons of the adopter and non-adopter group.

To address this issue, we use a difference-in-differences matching estimator, commonly found in the program evaluation literature, and originally proposed in Heckman et al. (1997). Matching estimators seek to identify the effects of a treatment by examining changes in observational units in the treatment population as compared to changes in matched observational units in the untreated population. This type of analysis assumes that the matching criteria, used to find equivalent units in both populations, are chosen such that the two populations would change in identical ways over time, absent the treatment. In our context, this is equivalent to assuming that the change in Internet usage for broadband adopters if they had remained with narrowband would have been the same as the changes in usage in the matched narrowband user population. In large samples, the time trend differences between matched pairs average out, leaving only the effect of the treatment (the difference-in-differences), and allowing us to mimic an experiment where treatment is randomly assigned. However, this result is dependent on a correct choice of matching parameters. To motivate our choice of parameters, we follow prior literature (e.g., Rappaport, et al. 2002) and argue that broadband adoption depends largely on an individual's utility for online content and their opportunity cost of time, where the opportunity cost of time is represented in our data by income.[2] In our analysis below, we also consider that preferences for different types of content may be a driver of broadband adoption in our population, and we control for these preferences accordingly. Finally, we also match on household size. Because our data are collected at the machine level, larger households are likely to have more people using the machine, which may increase usage numbers.

Although we do not directly observe users' utility for online content, we do observe the quantity of pre-broadband usage for both groups, which provides indications of preferences because both our "treated" and "untreated" groups had the same access (narrowband) in 2002. Identification in our model therefore relies on the assumption that any two individuals who spend the same amount of time online, conditional on income, should have similar preferences for online content. By including pre-broadband Internet usage as a covariate along with income, we condition on preferences for broadband and account for the large observed differences in pre-broadband usage observed in Table 1. By using pre-treatment usage to infer preference for online content, we implicitly rely on the assumption that preferences for individuals are stable over time. This assumption may be problematic if life-changes occur in households

[2] The use of income as a proxy for the opportunity cost of time is a common approach in empirical modeling and follows a theoretical argument originally attributed to Becker (1993).

in our sample that significantly change the propensity to adopt broadband as well as preferences for content consumption. For example, students who move into the workforce may experience changes in preferences for both of these goods, clouding our estimates of the direct impact of broadband on content consumption. Although our data only allow us to address changes relating to income and household composition, we include a variety of tests later in our analysis to rule out the possibility that our results are not being driven by these types of changes.

For our matching estimator, we utilize the nearest neighbor algorithm outlined in Abadie et al. (2001) and implemented in STATA as the "nnmatch" procedure. For each treated individual, a distance score is computed for all possible untreated neighbors $\|z_1 - z_0\|$ where z_1 is the vector of demographic variables including income and prior period outcomes for the treated individual, and z_0 is the corresponding vector for the untreated individual. All untreated individuals are sorted by distance score, and the closest neighbors are matched to each treated individual, with replacement. This approach allows us to directly compare the outcomes of individuals with and without broadband access who had similar Internet usage patterns in 2002. The full form of the estimator can be written

$$\hat{\alpha}_{DDM} = \frac{1}{N}\sum_{i=1}^{N} \{(Y_{1ti} - Y_{0t'i}) - (Y_{0tj} - Y_{0t'j})\}$$

where $\hat{\alpha}_{DDM}$ represents the causal effect of broadband adoption on Internet usage, (Y_{1ti}-Y_{0ti}) represents the difference in a treated individual i, and (Y_{0tj}-$Y_{0t'j}$) represents the observed differences in an untreated matched observation j. These differences are summed over all of the individuals in the treated group, and then averaged over the total number of observations N. Finally, we choose two adjustments to the simple matching estimator (Abadie et al., 2001). First, because inexact matches produce biases in finite-samples, we use a bias-adjusted form of the matching estimator. Intuitively, while the standard matching estimator computes the treatment effects from average differences in matched outcomes, the bias-corrected estimator adjusts these amounts by estimates of how much of these differences are due to inexact matches in the matching parameters. That is, we adjust for bias-inducing variations associated with inexact matches across the covariates. Second, we use robust errors to account for heteroskedasticity across usage intensity. To assess differences in consumption by content category, we use a similar approach, matching individuals based on demographics and prior total usage, but we also use prior consumption of that category type to control for category-specific preferences. We present our estimates below.

IV. Results

A. Overall Usage

Table 5 shows estimates of how broadband access impacts overall usage using four alternative sets of demographic variables with the matching estimator described above. Column 1 shows the estimates of the impact of broadband adoption on consumption when matching is conducted on 2002 usage, income, and household size, where household size is included because data is collected at the machine level, and household size is needed to control for the number of people using each computer. [Table 5 about here] The coefficient estimates suggest that broadband adopters increase consumption by about 1300 minutes per month over non-adopters, representing an increase in usage of about 40%. The differences in usage between adopter and non-adopter groups is statistically significant (t=3.06, p<.01).[3] Column 2 shows the adoption impact estimates in which demographic variables from both 2002 and 2004 are included as matching parameters. Inclusion of variables from both years eliminates the possibility that simultaneous increases in broadband and content consumption are simply reflecting the choices of households that experienced important demographic changes over the time period, such as changes to household composition or income. The estimates in Column 2 are essentially unchanged from Column 1, suggesting that demographic changes are not driving these effects. Column 3 shows the estimates when additional demographic variables are included. With the full set of demographic variables, the estimate of the impact of broadband on content consumption increases to over 1370 minutes per month (t=3.61, p<.01). For all of these estimates, although we control for total 2002 usage with the matching estimator, our

[3] The t-statistics and significance levels in the remainder of this paragraph and the following paragraph represent the hypothesis test that usage time was unchanged.

estimates are greater than a simple comparison of the averages of the two groups (see Table 4), suggesting that increases among heavier users are not driving our results. To check the robustness of our results to measurement error caused by individuals with "always-on" connections, we test the impact of broadband using number of pages viewed rather than duration as the dependent variable (Table 5). As with duration, broadband adoption also leads to significant increases in the number of pages viewed. However, the coefficient estimate on broadband adoption jumps significantly when demographic variables are included, from just over 600 additional page views per month when demographic variables are omitted (t=1.52), to over 900 page views per month with demographic variables included (t=2.7, p<.01).

Interpreting these effects as being driven by broadband adoption will be misleading if matching on pre-broadband outcomes does not adequately control for unobservable preferences for online content. If preference for a particular content type is confounded with broadband adoption, usage increases may simply represent stronger preferences for a particular type of content. In Column 4 of Table 5 we present results of our matching estimator where we control for preferences for the content categories that are responsible for the heaviest usage. The estimate on the impact of broadband for usage increases slightly, to just under 1400 minutes per month (t=3.41 p<.01), as does the estimate for page views, which rises to about 1200 page views per month (t=4.06, p<.01). Thus, it does not appear that our changes in usage are attributable simply to differences in preferences between our matched populations.

B. Increases by Pre-Adoption Usage

The results described above demonstrate that broadband adoption drives significant increases in consumption. The results also suggest that the heaviest users are not experiencing the largest increases from broadband adoption. This is somewhat intuitive because if consumption by individuals at the upper end of the consumption spectrum is already saturated, then broadband adoption will allow them to decrease the time spent on the Internet while consuming the same amount of content. To get a better understanding of these numbers, however, we look at the distributional characteristics of consumption increases, segmented by quintiles of 2002 usage, and using our matching estimator to control for other trends. [Table 6 about here]   Table 6 shows the results of our nearest neighbor matching estimator when the data is broken down by quintile. Interestingly, some of the most significant gains in Internet usage come from the lowest 2002 quintile. For this group, usage jumps by over 1700 minutes per month over comparable non-adopters, representing a 2200% increase in usage, by far the largest percentage increase of any group, and a test of the hypothesis that usage in this population remains unchanged yields significant results ($t=2.04$, $p<.05$). Furthermore, in our sample, members of this group surpass neighboring quintiles in absolute usage. To ensure that these results are not being driven by changes within a particular age or income group (such as college students transitioning to their first job), Table 7 presents usage increases for the lowest-usage quintile that have been further segmented into age and income segments of at least 10 observations per cell. [Table 7 about here] Although the relatively small number of observations per cell produces higher standard errors, the absence of trends in the estimates suggest that our results are not being driven by any one age or income group.

To further understand the dynamics behind these data, we break down consumption of different types of content by quintile of total 2002 usage. Table 8 shows the increases by quintile in each different category. [Table 8 about here] Large increases in Internet usage in the first (previously lightest-user) quintile are dominated by increases in the use of portals, computer applications, and advertisements, where computer applications include utilities that allow downloading of images, music, and other high-bandwidth content. For the first quintile, these numbers are larger than the equivalent observations for neighboring quintiles, and by contrast, we observe modest increases or even decreases in these categories for higher quintiles.

C. Content Diversity

We also tested how broadband adoption impacts the diversity of content that users consume. We used two measures, the number of sites visited during the October time period, and a category concentration index that measures how users distributed their time across the different content categories, computed as the sum of the squares of the time-share of each category. Thus, a reduction in the value of this index suggests that consumers are spreading their time over a greater variety of content types. Table 9 reports our findings. [Table 9 about here] We find that broadband adoption leads users to visit about twelve new sites per month, although these results have limited statistical significance (the hypothesis that the number of sites visited is unchanged is t=1.46, p<.15). Furthermore, we find an increase in the category concentration index, suggesting that the percentage of time that users spend within favorite categories is increasing (t=1.65, p<.1). This is consistent with an increase in the consumption of high-bandwidth content, because users can consume greater amounts of time on those sites that deliver high-bandwidth content.

V. Data Issues

Although the level of detail in our data offers several advantages, some important limitations are worth noting. As mentioned above, website categorizations are available for the 2002 session data but not for the 2004 session data. We address this gap by constructing a mapping of domain names to site categorizations using the 2002 data, and then use this mapping to apply categories to the 2004 data. Although ninety percent of the 2002 domains had been assigned categories, this method results in classification of only about seventy percent of the 2004 domains. The difference between these two numbers can be primarily attributed to new sites created between 2002 and 2004. Although this should not influence our total usage analysis, it may bias the category-level analysis and may also inject random errors into our content diversity measures, making it harder to draw clear contrasts. To reduce the possibility of error, we individually inspected the website classifications and also corroborated the category mapping results by inspecting the changes at the website level.

A second type of categorization error may occur when the mapping process introduces distributional errors that differ across the two years. If websites that appeared between 2002 and 2004 belong disproportionately to a particular content type, then the distribution across categories in 2004 may be inaccurate, producing spurious results. To verify that our methods are not overly sensitive to these types of errors, we cross-checked our classifications using the DMOZ Open Directory Project categorization schema, available online.[4] The DMOZ Open Directory project is an attempt to leverage the online community to categorize websites, similar to the way in which Yahoo! originally created their Internet index. We begin by dividing all websites into two groups: those that have categories, and those that do not, and then assigning DMOZ categories to both groups. If new websites in our 2004 data fall disproportionately into certain groups, this will show up in differences between the distributions in the DMOZ categorizations of the two groups.

Table 10 shows the results of our cross-categorization, comparing the numbers and percentages of websites that were categorized or missing from the 2002 categorization. Column 1 shows the DMOZ category name. [Table 10 about here] Column 2 and 3 shows, respectively, the number and percentage of websites categorized by our primary mapping scheme that were categorized into each of the respective DMOZ categories. Columns 4 and 5 show the comparable numbers for sites that were not categorized by our primary mapping but were categorized by DMOZ. If large biases are introduced by our mapping, we should expect to see large differences in the relative percentages of our two mappings, indicating that our mapping table did a poor job of capturing that type of content. Column 6 shows the differences in the two mappings. With the exception of the World category, the distributions are similar, suggesting that between 2002 and 2004, new websites entered at a rate that reflected their overall distribution.

Although the categorization of sites generally conforms to expectations, there are a few points worth considering. First, all sites, even if multi-purpose, are put into a single category. This can cause difficulties in interpretation of traffic to sites like microsoft.com, placed in the Business category, where the high number of site visits may correspond to downloads of patches or applications from that site. In addition, a few sites from the Business and Other categories are businesses that run content networks related to advertising or marketing, so although the domain is reported as receiving heavy traffic, users are actually responding to services being run at those sites. Thus, in cases where we make statements indicating that broadband adoption leads to an increase in usage of a particular type of content, we look at

[4] Available at http://www.dmoz.org

the disaggregate site level behavior of our panel to ensure that it is not driven by these types of categorization problems. Second, many of the sites that have the highest traffic numbers are sites which allow individuals to download content to their desktop, such as games or weather information. Although these applications continuously pull content to the desktop, individuals do not actually visit those websites by opening a browser and entering a website address. Because these applications still provide information that is consumed by their users, however, we treat them the same as other websites for the purposes of this study.

VI. Discussion

The Internet is an important channel through which to access information about goods and services. As broadband availability and adoption increases, we should ask if these technologies are having the desired impact on usage and information access. Accordingly, the goal of this study was to provide evidence on how providing broadband access impacts information consumption and which users benefit the most. We first estimated the impacts of broadband on quantity of usage, and found that adoption leads to an average increase of over 1300 minutes per month. We then explored which users benefit most from broadband. Significantly, we find that the greatest increases in consumption come from individuals who were in the lowest quintile of usage when they were narrowband users. This finding suggests that broadband satisfies unmet demand in certain populations. In our sample, these users experienced an increase in usage of almost 1800 minutes per month, illustrating that benefits from broadband are not limited to previously heavy users. These results may be of interest to decision makers who want to ensure that broadband rollouts bring access to information to as wide a population as possible.

We also examined consumption by content type and found that broadband access drives increases in usage of some types of content, such as portals, entertainment, and news, more than others. Increases in these categories suggest that the large usage increases that may be driven by broadband adoption may have important implications for the consumption of other forms of national and local media. The direction of these effects, however, is ambiguous, because although some forms of online content, such as entertainment videos, may be substitutes for television programs, some television programs may in fact exhibit complementarities with other types of online content. Further research in this area requires a finer level of data describing the geography of users and their consumption choices among other media.

References

Abadie, A., Drukker, D., Herr D. and Imbens G. 2001. Implementing Matching Estimators for Average Treatment Effects in Stata, *The Stata Journal* **1**, 1-18.

Autor, D. 2001. Wiring the Labor Market. *The Journal of Economic Perspectives* **15 (1)**, 25-40.

Becker, G. S. 1993. *Human Capital: A Theoretical and Empirical Analysis*. Chicago: University of Chicago Press.

Brodie, M., Flournoy, R. E., Altman, D. E., Blendon, R. J., Benson, J. M., and Rosenbaum, M. D. 2000. Health Information, the Internet, and the Digital Divide. *Health Affairs* **19 (6)**, 255-265.

Crandall, R. and Jackson, C. 2001. *The $500 Billion Dollar Opportunity: The Potential Economic Benefit of Widespread Diffusion of Broadband Internet Access*. Washington, D.C.: Criterion Economics.

Dimaggio, P. and Hargittai, E. 2001. *From the Digital Divide to Digital Inequality: Studying Internet Use as Penetration Increases*. Princeton University Working Paper, Summer 2001.

Goldfarb, A. and Prince, J. 2006. *Internet Adoption and Usage Patterns are Different : Implications for the Digital Divide*. Cornell University Working Paper.

Heckman, J., Ichimura, H., and Todd, P. 1997. Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Programme. *Review of Economic Studies* **64 (4)**, 605-654.

Hong, S. 2006. *The Effect of Digital Technology on the Sales of Copyrighted Goods: Evidence from Napster*. Working Paper, University of Illinois at Urbana-Campaign.

Greenstein, S. and Prince, J. 2006. *The Diffusion of the Internet and the Geography of the Digital Divide in the United States*. Cornell University Working Paper.

Hoffman, D. L. and Novak, T. P. 1998. Bridging the Racial Divide on the Internet. *Science*, **280**, 390-391.

Kraut, R., Scherlis, W., Mukhopadhyay, T., Manning, J., and Kiesler, S. 1996. HomeNet: A Field Trial of Residential Internet Services. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*.

Liebowitz, S. 2006. *Testing File-Sharing's Impact By Examining Record Sales in Cities*. Working Paper, University of Texas at Dallas.

Madden, G. and Simpson, M. 1996. A Probit Model of Household Broadband Service Subscription Intentions: A Regional Analysis. *Information Economics and Policy* **8 (3)**, 249-267.

Norris, P. 2001. Digital Divide: Civic Engagement, Information Poverty, and the Internet Worldwide. Cambridge: Cambridge University Press.

Parker, E. B. 2000. Closing the Digital Divide in Rural America. *Telecommunications Policy* **24**, 281-290.

Prieger, J. 2003. The Supply Side of Digital Divide: Is There Equal Availability in the Broadband Internet Access Market?, *Economic Inquiry* **41(2)**, 346-363.

Rappoport, P., Kridel, D., and Taylor, L. 2002. The Demand for Broadband: Access, Content, and the Value of Time, in R. Crandall and J. Alleman ed., *Broadband: Should We Regulate High-Speed Internet Access?* Washington D.C.: AEI- Brookings Joint Center for Regulatory Studies.

Table 1: October 2002 Minutes by Category

Category	Representative Site	Adopters	Non-Adopters	T-statistic*
Adult	*voyeurweb.com*	**83.0**	**65.4**	**1.01**
Careers	*careerbuilder.com*	**4.3**	**4.2**	**0.09**
Society	*match.com*	**59.1**	**35.6**	**1.54**
Computer Applications	*hotbar.com*	**30.4**	**24.3**	**0.75**
Education	*fsu.edu*	**17.7**	**14.2**	**0.74**
Entertainment	*neopets.com*	**179.9**	**96.9**	**2.29**
Health	*webmd.com*	**5.5**	**3.2**	**2.06**
Finance	*fidelity.com*	**44.8**	**30.1**	**1.30**
Regional	*cleveland.com*	**1.46**	**0.6**	**1.69**
Business	*microsoft.com*	**77.2**	**48.7**	**2.72**
Home & Living	*allrecipes.com*	**5.2**	**5.8**	**0.19**
News	*cnn.com*	**22.3**	**24.2**	**0.30**
Portals	*yahoo.com*	**454.2**	**412.5**	**0.48**
Reference	*about.com*	**15.4**	**13.6**	**0.59**
Shopping	*amazon.com*	**56.9**	**48.7**	**0.50**
Sports	*sportsline.com*	**30.3**	**12.3**	**3.06**
Travel	*expedia.com*	**14.7**	**13.7**	**0.31**
Services	*gator.com*	**103.6**	**81.4**	**1.64**
International	*sandesh.com*	**1.9**	**0.3**	**3.47**
Automotive	*cars.com*	**7.9**	**8.5**	**0.19**
Auction	*ebay.com*	**69.8**	**61.3**	**0.38**
Ads	*realmedia.com*	**151.1**	**108.5**	**2.00**
Market Research	*mysurvey.com*	**4.6**	**2.6**	**1.77**
Other	*akamai.net*	**162.2**	**114.3**	**1.96**
Government	*ny.us*	**11.1**	**8.4**	**1.47**
Unclassified	*freeslots.com*	**87.6**	**95.6**	**0.40**
N		**366**	**3807**	

*T-Statistics test the hypothesis of no difference between group means.

Table 2: 2002 Demographics

	Proportion of Adopters	Proportion of Non-Adopters
Income		
< 15,000	.038	.052
15,000-24,999	.118	.118
25,000-34,999	.161	.160
35,000-49,999	.197	.207
50,000-74,999	.260	.258
75,000-99,999	.115	.112
> 100,000	.112	.093
	$X^2(6) = 2.72, p < .84$	
Age		
18-20	.016	.018
21-24	.052	.038
25-29	.074	.049
30-34	.090	.083
35-39	.077	.079
40-44	.107	.100
45-49	.128	.179
50-54	.197	.156
55-59	.087	.102
60-64	.087	.088
65-74	.085	.108
	$X^2(10) = 16.89, p < .07$	
Children Present		
Yes	.448	.444
No	.552	.556
	$X^2(1) = .02, p < .88$	
Household Size		
1	.057	.099
2	.399	.366
3	.210	.231
4	.194	.176
5	.107	.079
6+	.033	.050
	$X^2(5) = 13.58, p < .02$	
Education		
Less than High School	.008	.008
High School Diploma	.128	.133
Some College but no degree	.344	.339
Associate Degree	.265	.279
Bachelors Degree	.156	.147
Graduate Degree	.098	.095
	$X^2(5) = 0.55, p < .99$	
N	366	3807

Table 3: Logit Analysis of the Demographic Determinants of Broadband Adoption

	Broadband Adoption
Income	0.041
	(0.036)
Age	-0.039
	(0.021)
Education	0.003
	(0.049)
Household Size	0.051
	(0.056)
Children	-0.093
	(0.146)
Constant	-2.366
	(0.270)**
Observations	4173
* significant at 5%; ** significant at 1%, standard errors in parentheses	

Table 4: October 2002 and October 2004 Usage by Minutes and Page Views

2002 to 2004 Speed	2002 Usage (Minutes)	2004 Usage (Minutes)	Difference	2002 Page Views	2004 Page Views	Difference	N
Non-Adopter	2451.4	1973.7	-477.7	1567.2	1227.3	-339.9	3807
Upgrader	3231.0	3906.4	675.4	2106.0	2552.3	446.3	336
Downgrader	3189.4	2306.5	-882.9	2025.0	1712.3	-312.7	628
Maintainer	3217.9	2487.1	-730.8	2082.2	1548.7	-533.5	1677

Non-Adopters had narrowband connections in 2002 and in 2004. *Upgraders* had narrowband in 2002 but adopted broadband in 2004. *Downgraders* switched from broadband in 2004 to narrowband in 2002. *Maintainers* had broadband in both years.

Figure 1: October 2002 vs. October 2004 Usage for Adopters, Organized by 2002 Usage Quintile

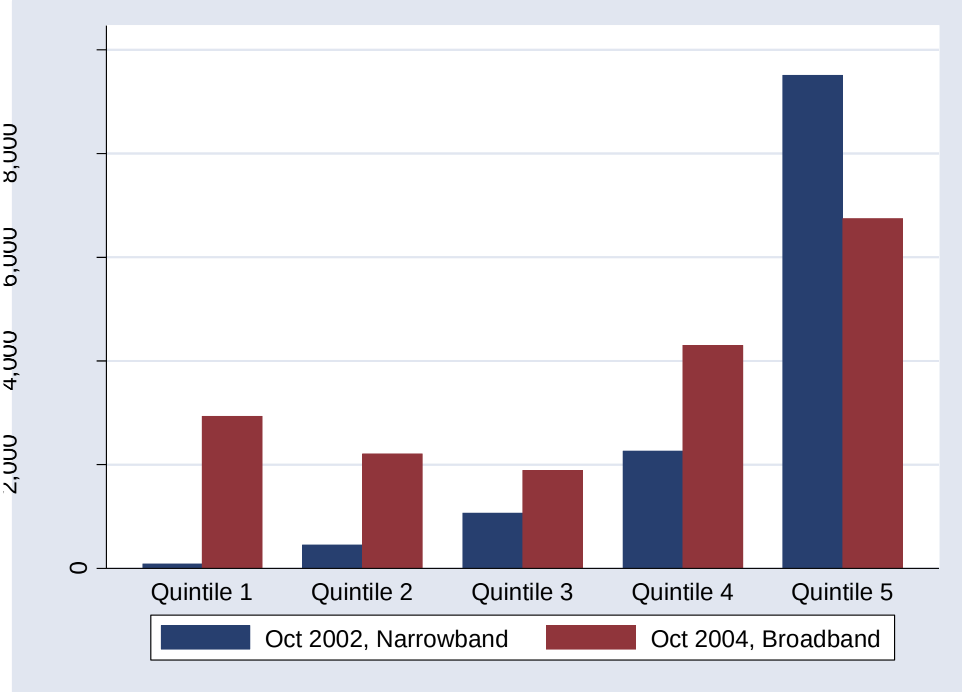

Table 5: Matching Estimators - Impact of Broadband Adoption on the Overall Usage of Adopters

	Difference in Differences Estimates Combined With			
	(1)	(2)	(3)	(4)
	Matching on Income, Usage, and Household Size	Matching on Income, Usage, Household Size Data from 2002 and 2004	Matching on Income, Usage, Household Size and Demographic Variables	Matching on Income, Usage, Household Size, and Content Profile**
Increased Minutes Online	1311.8 (429.1)	1318.0 (455.2)	1371.1 (379.9)	1395.5 (410.2)
Increased Page Views	624.1 (409.1)	859.1 (320.0)	914.4 (338.4)	1198.7 (292.8)

Robust Standard errors in parentheses. Estimates are bias-adjusted and represent the change in adopter usage resulting from broadband adoption. Control group is the matched population of non-adopters. **Individuals are matched on consumption of content by category. Matching occurs on largest usage categories including portals, business, ads, services, entertainment, and computers.

Table 6: Matching on Income, 2002 Usage, HH Size – Impact of Broadband on Usage by Quintile

2002 Usage Quintile	Mean 2002 Usage (Minutes)	Differences (Minutes)	N
1	78.3	1775.6 (866.9)	61
2	449.2	976.8 (530.3)	66
3	1051.0	500.8 (398.6)	63
4	2271.3	3077.0 (803.2)	82
5	9399.0	214.6 (1339.5)	94

Standard errors estimated in parentheses.

Table 7: Lowest Quintile Usage Increases by Income and Age Group

Income	Mean 2002 Usage (Minutes)	Differences (Minutes)	N
< 34,999	89.4	1323.5 (1455.3)	18
35,000-49,999	83.8	-588.60 (2025.6)	13
50,000-74,999	95.4	5429.0 (2486.1)	15
> 75,000	67.5	728.90 (410.3)	15
Age	**Mean 2002 Usage (Minutes)**	**Differences (Minutes)**	**N**
< 34	88.0	2574.0 (1318.2)	13
35-44	89.6	3207.6 (3924.8)	19
45-54	85.1	-235.7 (1484.7)	14
> 55	70.7	1309.7 (596.5)	15

Standard errors in parentheses. Age is associated with head of household. Cells were chosen to retain a minimum of ten observations in each category.

Table 8: Matching on Income, Usage, and Household Size – Increases by Category and by 2002 Total Usage Quintile

	2002 Total Usage Quintile				
	(1)	(2)	(3)	(4)	(5)
Adult	10.9	8.4	19.9	30.9	50.5
Careers	1.2	0.9	2.7	2.8	2.7
Society	15.8	19.2	43.3	18.6	100.6
Computer apps	70.1	11.9	34.4	40.8	132.5
Entertainment	38.9	46.1	41.1	170.7	111.4
Finance	20.4	27.3	19.4	56.8	39.8
Health	1.7	1.6	0.4	5.0	6.0
Business	83.4	57.2	60.0	86.7	364.5
Shopping	36.5	42.6	55.9	111.5	76.8
Reference	2.5	3.3	1.6	11.3	15.1
Auction	33.4	145.3	100.4	56.2	1047.3
International	0.0	0.0	0.0	0.0	10.0
Market research	1.1	0.3	0.8	2.3	12.0
Other	77.3	78.7	22.2	48.2	72.3
Education	5.9	25.9	7.9	17.7	19.4
Regional	0.1	0.1	0.0	0.5	2.6
Home	3.3	11.3	3.9	6.2	5.7
Sports	4.0	4.7	12.9	32.1	58.2
Travel	6.5	6.2	9.6	15.6	22.6
Services	35.3	53.1	36.7	183.4	134.0
Automotive	3.3	5.0	2.2	9.9	6.3
Government	18.2	12.8	4.5	21.5	9.4
Unclassified	20.2	18.2	14.3	45.0	55.0
Portals	438.9	283.9	339.6	666.6	926.3
News	16.6	84.4	16.2	32.8	64.9
Ads	202.6	66.8	62.3	110.3	595.5

Table 9: Matching on Income, Usage, Household Size - Content Diversity Measures

	Broadband Impact
Number of Sites Visited	12.4 (8.5)
Category Concentration	.071 (.043)

NOTE –Robust standard errors are shown in parentheses. Category Concentration is sum of the squares of the time-shares of each category. Impact estimates represent Sample Average Treatment Effect on the Treated (SATT).

Table 10: Validating Classification Using DMOZ Categories

DMOZ Category	Categorized	% (of Total)	Missing	% (of Total)	Difference
Arts	12125	10.02	2734	6.99	3.03
Shopping	3765	3.11	2891	7.39	-4.28
Science	2674	2.21	974	2.49	-0.28
Games	2978	2.46	645	1.65	0.81
Business	4260	3.52	2754	7.04	-3.52
Computers	7214	5.96	2340	5.98	-0.02
Health	1971	1.63	1013	2.59	-0.96
Sports	4137	3.42	774	1.98	1.44
World	36690	30.31	8976	22.94	7.37
Society	8987	7.43	2549	6.51	0.91
News	826	0.68	410	1.05	-0.37
Home	1474	1.22	482	1.23	-0.01
Regional	20367	16.83	7877	20.13	-3.30
Recreation	5362	4.43	1275	3.26	1.17
Kids and Teens	2018	1.67	482	1.23	0.44
Adult	1333	1.10	901	2.30	-1.20
Reference	4855	4.01	2050	5.24	-1.23
Total	121036	100	39127	100	0