# Broadband Access and Content Consumption

**DRAFT**

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

June 2006

**Abstract**

*We use panel data describing Internet usage before and after broadband adoption to explore how access to broadband drives changes in content consumption. We motivate our analysis with a model that accounts for access speed, the bandwidth intensity of different types of content and the opportunity cost of time to explain usage changes. Faster download speeds generally increase consumption because deadweight losses incurred while waiting for content to load are significantly smaller. Furthermore, factoring preferences for content of differing bandwidth intensities into our model suggests large increases in consumption for individuals who may have spent little time online in a narrowband environment because the high bandwidth services they prefer require prohibitive amounts of time to download. Our data confirms that people who switch to broadband spend greater amounts of time online after controlling for time trends, with larger gains accruing to higher-bandwidth services. Consistent with our model, some of the largest usage gains come from individuals who comprised the lowest usage quintile when they were narrowband users, but greatly increased consumption of high-bandwidth applications after adopting broadband. These users consume content in greater quantities than users in neighboring quintiles, passing both the second and third quintiles in terms of absolute usage.*

Keywords: municipal broadband, digital divide, productivity, telecommunications policy

## I. Introduction

Arguments for municipal broadband provision have been difficult to advance because existing research on the adoption of broadband technologies has focused primarily on optimal policy measures to encourage widespread deployment, while assuming positive benefits. This study begins to address this gap by investigating the impacts of broadband access on consumption. We begin by presenting a model of how broadband access impacts Internet usage based on a utility framework, where consumers choose the optimal amount of time to spend on the Internet, taking into account the opportunity cost of their time and the time required to download content. As expected, our model suggests that increasing access speeds increases consumption for all users because of the falling opportunity costs of time for downloading a fixed amount of content. Providing broadband access has ambiguous effects on total usage time, however, because at the higher end of the usage spectrum, time saved while waiting for content to load may offset increased consumption of Internet content. We then explore how consumer preferences for content impact benefits derived from broadband access. Individuals who otherwise have strong preferences for the Internet bundle relative to other goods may spend relatively little time online if they prefer high bandwidth services that are prohibitively costly to download. This unmet demand may be satisfied by broadband adoption, which significantly reduces the time required to download high bandwidth applications.

We test these ideas using disaggregate Internet usage data from a panel of users in 2002 and 2004. Our analysis is based on exploring how broadband access changes the consumption patterns of members of our panel. Because we have connection speed information, we can compare before-and-after consumption data for users who switch from narrowband to broadband between 2002 and 2004, while using the behavior of non-switchers to control for confounding time trends. One of our challenges is that individuals select into the broadband pool. If people who switch to broadband have stronger preferences for online content than those who do not, they could have had higher usage numbers in 2004 even in the absence of broadband adoption. To account for selection issues, we use a difference-in-differences matching estimator to control for factors, such as preference for the Internet bundle or the opportunity cost of time, that may otherwise lead to biased estimates. We estimate increases in consumption resulting from broadband adoption, conditional on preferences for online content. We also explore how these effects relate to narrowband Internet usage to find the effects of providing higher bandwidth to users with different preferences. Finally, to tie these together with the bandwidth intensity of content, we investigate how these estimates break down by content category.

Our results suggest that increases in consumption are distributed in ways that may escape analyses that ignore the bandwidth intensity of content. While for the panel as a whole, usage declined, individuals who switched to broadband increased the amount of time they spent online in 2004. Much of these gains were from people who were relatively *light* Internet users in 2002, suggesting that broadband access provides significant benefits to users who spent little time with the Internet before having access to broadband. While greater relative increases in lower quintiles may just be part of a saturation story, in our data, the lowest quintile surpasses the two neighboring quintiles in *absolute* usage after broadband adoption. While users in the middle moderately increased their consumption across many different content types, those in the lowest quintile dramatically increase the time they spend on portals and sites that allow downloads of high-bandwidth content such as images, music, and online games, suggesting that for many users, broadband access satisfies unmet demand for high-bandwidth content that is not worth downloading in a narrowband environment.

This study contributes to the broadband literature in a number of ways. While other studies have looked at the demographic factors that influence the demand for high bandwidth access, we are interested in how broadband access affects consumption. We can address this question because of access to content level data about Internet usage for a panel of users before and after they have broadband access, avoiding common biases that contaminate cross-sectional studies. Furthermore, the disaggregate nature of our data allows us to test how broadband access affects consumption of different types of Internet content, giving us a more nuanced view of how increases in consumption are distributed. The rest of our paper proceeds as follows. This introductory section is followed by a review of other literature about the demand for broadband access. Section three presents a model of Internet usage that provides a framework for the rest of our study. Section four discusses our data and the empirical strategies we use to resolve endogeneity problems. We finish the paper with a discussion of the Results and Conclusions.

## II. Literature

While this is the first study to our knowledge that examines the effects of broadband provision on consumption choices, researchers have examined the digital divide and broadband adoption from a number of other angles. From a development perspective, interest in this topic stems from the potential for economic growth that has been linked with widespread broadband adoption. For example, having access to online information can mean better information about jobs, education, and health, as well as a more skilled and productive workforce. For information workers, broadband access can drive higher productivity through more efficient work practices, as well as better utilization of non-work hours. Furthermore, broadband access can lead to a higher quality of life through greater convenience, and increased involvement with community, ethnic, or civic organizations. Finally, for business and government organizations, widespread adoption of broadband can allow for more efficient provision of goods and services. Demand side studies of broadband availability have focused primarily on the availability of content and broadband pricing. In the INDEX project, for example, researchers used controlled experiments to gauge consumer price sensitivity to higher bandwidth. They find that consumers are willing to pay relatively little for higher bandwidth access, and suggest that this may in part be due to the lack of compelling applications available for broadband (Varian 2002). In the study closest to ours, Rappoport et al use click stream data from ten cities in 2001 to determine what demographic and usage factors distinguish narrowband households from broadband households (Rappoport et al 2001). They find that demographic characteristics alone do not provide a clear distinction between broadband households and narrowband households, and they look at usage statistics as well as price and the opportunity cost of time to explain broadband adoption. While their study looks at differences in usage patterns between narrowband and broadband subscribers, our paper addresses the fundamentally different question of how provision of broadband to existing narrowband users changes usage patterns. We are interested in the causal effect of broadband on information consumption, rather than characterizing a selection effect.

While many studies focus on broadband diffusion and adoption, the potential benefits from widespread broadband adoption have largely been hypothesized, primarily due to a lack of data. While there is case-based evidence of the benefits of municipal broadband roll outs, generalizable evidence is sparse. Ford and Koutkey examine a municipal broadband rollout in a Florida county. They use other similar counties in Florida as controls, and find that the county with wireless infrastructure subsequently experiences a greater growth rate than the comparison counties (2005). Lehr et al. use FCC data to attempt to establish a large-sample link between broadband roll out and measures of improved economic productivity, but note that they are hampered in their efforts by the lack of available data, and the probable lags between widespread deployment and observable benefits (2005). In this paper, we acknowledge the difficulty of translating broadband access to economic benefits, and focus instead on determining how broadband impacts information consumption choice at the level of the individual user. This allows us to make statements about who benefits from broadband access, and how the type of information that they consume might change. In the next section, we present a model that describes the Internet usage of utility-maximizing consumers. We draw heavily on a model recently presented by Goolsbee and Klenow, who suggest that Internet welfare calculations can best be determined not by its fixed price, but by the opportunity cost of the time people choose to spend with the product (2006).

## IV. Framework

We consider a framework where the time consumers spend on the Internet is determined by the opportunity cost of their time including time waiting for content to download, and the utility derived from the Internet bundle relative to other goods. To account for these factors, we start with a model presented by Goolsbee and Klenow (2006) where consumers optimize over Internet usage and a composite bundle of all other purchased goods. To explore how access speed impacts total Internet usage, we incorporate into this model a parameter $\gamma$ ("*download speed*") that represents the additional fractional time spent while consumers wait for content to download, so that for each unit of consumption time, a net amount of time $1+\gamma$ is required, representing consumption plus download time. Then consumers maximize utility of the form

$$\theta(C_I^{\alpha_I} L_I^{1-\alpha_I})^{1-1/\sigma} + (1-\theta)(C_O^{\alpha_O} L_O^{1-\alpha_O})^{1-1/\sigma}$$

subject to a budget constraint

$$P_I C_I + F_I + P_O C_O = W(1 - L_1(1+\gamma) - L_O)$$

where $C_I$ is purchased Internet services, $L_I$ is the fraction of time spent consuming Internet content, $F_I$ is the fixed cost of the Internet and $L_I(1+\gamma)$ is the aggregate time spent downloading and consuming Internet content. $C_O$ and $L_O$ represent the amount and time spent on the composite, and $\theta$ scales the importance of the bundles. Then in an optimal choice set for consumers

$$L_I^* = \frac{1-\alpha_I}{1+\gamma+\Gamma(1+\gamma)^{\sigma(1-\alpha_I)+\alpha_I}}$$

where $\Gamma = \left(\frac{\lambda_I}{\lambda_O}\right)^{\sigma-1}\left(\frac{1-\theta}{\theta}\right)^{\sigma}$ and $\lambda_j = \left(\frac{P_j}{\alpha_j}\right)^{\alpha_j}\left(\frac{W}{1-\alpha_j}\right)^{1-\alpha_j}$ following the Goolsbee and Klenow notation. The impact of a change in bandwidth on consumption is

$$\frac{\partial L_I^*}{\partial \gamma} = -(1-\alpha_I)\frac{(\sigma(1-\alpha_I)+\alpha_I)\Gamma(1+\gamma)^{\sigma(1-\alpha_I)+\alpha_I-1}}{(1+\gamma+\Gamma(1+\gamma)^{\sigma(1-\alpha_I)+\alpha_I})^2}$$

which is always negative when $\sigma > 0$, which corresponds to a positive elasticity of substitution between the Internet bundle and other goods. Thus consumption will generally increase with bandwidth. The change in the total amount of time spent on the Internet when bandwidth increases is

$$\frac{\partial L_I^*(1+\gamma)}{\partial \gamma} = L_I^* - (1-\alpha_I)\frac{(\sigma(1-\alpha_I)+\alpha_I)\Gamma(1+\gamma)^{\sigma(1-\alpha_I)+\alpha_I}}{(1+\gamma+\Gamma(1+\gamma)^{\sigma(1-\alpha_I)+\alpha_I})^2}$$

which can be positive for large values of $L_I^*$. Intuitively, at the upper end of the usage spectrum, utility from usage is saturated, and the effect of broadband is to decrease the amount of time taken to consume a given amount of online content. At the lower end of the usage spectrum, these savings are outweighed by increases in consumption.

When download time is not a factor, conditional on other factors, consumption generally rises along with preferences for the Internet bundle relative to other goods. However, consumption will be suppressed when $\gamma$ is high, where $\gamma$ will be much higher for individuals who consume video or online games than for those who prefer news or financial information. If preferences among users are similar across the aggregate bundle, then after broadband adoption consumers adjust usage such that postbroadband usage will just be a monotonic transformation of pre-broadband usage. However, if users also significantly differ in their tastes for content, then the ranking order may change, as people who prefer high bandwidth content will now find greater returns from spending time online.

## IV. 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 individuals[1]. For each individual, we have demographic information, connection speed, and detailed session information for the entire month, including domain name level information for each web site visited, duration of visit, and number of pages viewed during the visit. Our analysis of how broadband affects content consumption is based on total usage as well as categorization of web sites, where our analysis is largely based on the duration distribution across different content types, as well as the increases in consumption of each of the different types of content. To provide a sense for this categorization, Table 4 lists categories as well as some of the most heavily visited sites in each category. Our demographic data include age, income, education, household size, census region, and whether or not a child is present in the house, where all demographic variables are coded in discrete levels. From this panel, we extracted the 5200 individuals who either retained narrowband access in both 2002 and 2004 *(stayers),* or who upgraded from narrowband in 2002 to broadband in 2004 (*adopters)*. We exclude from the panel users who had broadband during both years and those who switched from broadband in 2002 to narrowband in 2004. The individuals in the sample are taken from a population who had already conducted an online transaction by 2002, but had not yet adopted broadband, so we are cautious in interpreting our results.

Some descriptive statistics for the 2002 data are shown in Table 1, broken down by adopters and stayers. As expected, adopters in the population are slightly wealthier and younger. Significantly, they had much higher mean Internet usage in 2002, suggesting that adopters and stayers have different

[1] Source: Comscore Networks.

preferences for Internet usage, confirming that individuals select into the broadband pool because they expect higher benefits. We control for this factor when we estimate how broadband adoption impacts Internet usage. Figures 1 and 2 show the usage distributions across adopters and stayers during 2002 when both groups had narrowband access. Both distributions are heavily right-skewed, but the distribution of adopters has a fatter tail, suggesting that higher means in the adopter group are a result of many of the very heavy narrowband users in 2002 choosing to adopt broadband. Although the data is heavily right-skewed, we do not log-transform the duration variable for two reasons. First, because we rely on a non-parametric approach, our estimates should be unaffected by the distributional characteristics of the data. Secondly, we use total usage as a matching parameter in our matching estimator below. To the extent that log transformation makes large differences look like smaller differences, a log-transformation would result in the loss of information about the distance between observations.

For robustness and to ensure that changes to the duration variable are not simply reflecting the effects of "always-on" connections, we also explore the impacts of broadband adoption on an alternative dependent variable, page views. Table 2 includes figures for 2004 consumption by duration and page views. By both metrics, broadband users significantly increased consumption over the two year period. This effect is more pronounced when compared to individuals who stayed with narrowband, who decreased their Internet usage over the two year period. While it is tempting to interpret this as evidence linking broadband adoption with increased Internet usage, users who switched to broadband may differ from those who did not such that even in the absence of switching, they would have increased their usage numbers. Figures 3 and 4 show how mean Internet usage varies by age and income across the two year period. Internet usage drops between 2002 and 2004 for both groups in all categories, and there are suprisingly few trends within demographic category. Thus, it is unlikely that shifts within demographic are responsible for the observed increases in Internet usage. However, other unobservable factors, such as the effect of differences in preferences for content, may influence the size effects we see. In our empirical analysis below, we rule out some alternative explanations and derive estimates of the impacts of broadband usage that are less subject to confounding influences.

### *B. Empirical Strategy*

In this section, we describe our strategy for isolating the effect of broadband on information consumption behaviors. Because we have usage data before and after the adoption decision, simple differences 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. Having data on a group that does not adopt broadband, however, helps us to control for these trends, and a difference-in-differences (DID) estimator allows us to separate out these effects and separately identify the effects of broadband adoption. The success of this estimator, however, relies on some restrictive identifying assumptions. In the absence of the treatment, the time trends for the treated and control groups must be the same. Because we are using non-randomized data, however, a concern is that the changes in usage over the two year period may have differed between the two groups in the absence of treatment. For example, Table 1 shows that 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, and even in the absence of the treatment, may have consumed content at different rates than those in the untreated group.

To address this issue, we use a difference-in-differences matching estimator, used heavily in the program evaluation literature, and orginally proposed in Heckman et al (1995). In our context, a matching estimator will identify the treatment effect when the following assumption is satisified -- conditional on observable matching parameters, the change in Internet usage for either group would have been the same in the absence of treatment. In large samples, therefore, the time trend differences between matched pairs average out, leaving only the effect of the treatment, and allowing us to mimic an experiment where treatment is randomly assigned if we can find a set of observable parameters to satisfy unconfoundedness. To motivate our choice of parameters, we argue that broadband adoption depends on an individual's utility for online content and their opportunity cost of time, represented in our data by income. We also test to see if preferences for different types of content are a stronger driver of broadband adoption in our population. While other unobserved factors such as geographic availability may also influence broadband adoption, we assume that these remaining factors are independent of total usage, and therefore do not bias our estimates. Although an individual's preference for online content is unobserved, we do observe pre-broadband outcomes for both groups, which are functions of preference only because members of both groups had only narrowband access in 2002. Then 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. Our unconfoundedness assumption is satisfied if pre-broadband outcomes control for heterogeneous unobserved preferences for online content. By using pre-treatment usage to infer preference for online content, we implicitly make assumptions that preferences for individuals are stable over time.

For our matching estimator, we draw upon the nearest neighbor algorithm outlined in Abadie and Imbens (2001). 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 lagged outcomes for the treated individual and $z_0$ is the equivalent vector for the untreated individual. Then all untreated individuals are sorted by distance score, and the closest neighbors are chosen, with replacement. The average of the three closest neighbors are used to estimate the time trends for the treated individual. 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

$$\hat{\alpha}_{DDM} = \frac{1}{n_1} \sum_{i \in I_1 \cap S_P} \left\{ (Y_{1ti} - Y_{0t'i}) - \sum_{j \in I_0 \cap S_P} W(i,j)(Y_{0tj} - Y_{0t'j}) \right\}$$

where $\hat{\alpha}_{DDM}$ represents the causal effect of broadband adoption on Internet usage. Finally, we choose two adjustments to the simple matching estimator, also discussed in Abadie and Imbens. Because inexact matches produce biases, we use the bias-adjusted form of the matching estimator. Secondly, we use robust errors to account for heteroskedasticity across usage intensity. To assess differences in consumption by category, we use a similar approach, matching individuals based on demographics and prior total usage, but also prior consumption of that category type to control for category-specific preferences. In this set of analyses, however, we also control for the possibility of complementarities between different types of content. We present our estimates below.

## IV. Results

### *A. Overall Usage*

Table 6 show estimates of how broadband access impacts overall usage using the matching estimator described above. On average, broadband was responsible for 1700 minutes of increased usage per month. Although we control for preferences with the matching estimator, this is a *greater* number than the simple estimates from Table 2 would indicate. This suggests that increases among heavy users are not driving our results. To check the robustness of our results to measurement error caused by individuals with "always-on" connections, we explore the impact of broadband using page views instead of duration as the dependent variable. Table 6 shows the results of our analysis on page views. Like duration, broadband adoption also leads to significant increases in the number of pages viewed.

Interpreting these effects as returns to 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 preferences for a particular type of content. We therefore examine the mean usage by category over our matched sample. If increased usage due to broadband adoption is reflecting a preference for a certain type of content, then the mean usage in our matched sample will be higher for certain types of content that are associated with broadband adoption. Table 5 shows the result of our comparison across the matched sample. While there are slight differences across categories, in our matched sample, the distribution of usage across adopters and stayers appears to be similar, suggesting that it is unlikely that preferences for particular types of content are being spuriously attributed to changes resulting from broadband adoption.

### *B. Increases by Pre-Adoption Usage*

The results presented above suggest that the individuals at the top of the distribution are not experiencing the largest increases from broadband adoption. This is intuitive because if the consumption of individuals at the top end is already saturated, broadband adoption may drive 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, broken down by 2002 usage. Table 6 shows the results of our nearest neighbor matching estimator when the data is broken down by quintile. Interestingly, the most significant gains in Internet usage come from the quintile that used the Internet the least in 2002. Because this table shows differences, however, this result may be driven by the fact that low intensity users of the Internet simply had more room to increase usage, while users in higher quintiles are already be saturated. Thus, Table 7 shows the total usage in 2004 by quintile for individuals who adopted broadband. Individuals who used the internet the least in 2002 and who switched to broadband showed the largest gains, passing both the second and third quintiles in total Internet usage in 2004. This suggests that broadband adoption may be impacting these groups in fundamentally different ways.

To understand the dynamics behind these consumption data, we break down content consumption by quintile. Table 8 shows the increases by quintile in each different category. The large increases in Internet usage for the first quintile are dominated by increases in the use of portals, computer applications, and advertisements, where computer applications include utilities that allow downloading of images and music. For the first quintile, these numbers are much larger than the equivalent observations for neighboring quintiles, and by contrast, the highest quintile saw modest increases or decreases in these categories. To put these numbers in better perspective, we show total usage numbers for each of these categories in 2004 for broadband switchers and stayers. Figures 6 and 7 illustrate dramatic differences between consumption of computer applications between the lowest quintile users who adopted broadband, and those who did not. Figures 8 and 9 show similar figures for portals and advertising, suggesting that the time required to download high-bandwidth applications suppresses the usage of certain Internet users in a narrowband environment. When these users adopt broadband, they spend disproportionate amounts of time downloading content or applications that stream content, such as weather, music, or entertainment, to the desktop, suggesting high utility for the Internet as a high-bandwidth channel.

## VI. Data Issues and Alternative Explanations

While the level of usage detail in our data is a considerable asset, our data also has several important limitations. The first is that 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. While ninety percent of the 2002 data has been assigned categories, this method results in classification of about seventy percent of the 2004 data. The difference between these two numbers can primarily be attributed to sites that are new between 2002 and 2004. To address the ability of this categorization process and mapping method to answer our questions, we consider that errors in categorization may be of several types. The first type of categorization error affects categories in both 2002 and 2004, and occurs when sites are placed in the wrong category, are left unclassified, or are restricted to one type of category even though the website includes functions that belong in many different categories. These types of errors will affect the accuracy of the distribution across categories when it affects categories in a systematic way that does not disappear in large samples, such that some categories are overrepresented or underrepresented in the sample. For example, if adult websites are put in the unclassified category in a disproportionate way because of ambiguity of domain name, the duration numbers spent on adult web sites will be underrepresented in both 2002 and 2004, leading to errors in distribution across categories. While this problem is difficult to address, it is mitigated by the fact that our empirical methods emphasize differences in distributions between the two years, not levels. Thus if there is minimal growth in duration spent on adult websites between the two years, this type of categorization error differences out. However, if a particular miscategorized site is responsible for considerable usage growth as a result, for example, of providing particularly high-quality broadband ready content, then it may still produce errors because we interpret results at the category level, and this web site will contribute to increases in the wrong category. For example, our results may suggest that broadband adoption leads to an increase in time spent on entertainment web sites, when these numbers are actually driven by a site that should belong in portals. However, cursory inspection of website categorizations, such as those represented in Table 2, suggest that the websites that have been categorized are accurately placed into categories. Furthermore, we are careful to investigate significant results at the underlying disaggregate web site level before interpreting our results.

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 possibilities for categorization error, we cross-check our classifications using the DMOZ Open Directory Project categorization schema, available online[2]. 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. Then if web sites that appear between 2002 and 2004 in our data fall disproportionately into certain groups, this should show up in differences between the distributions in the DMOZ categorizations of the two groups. For example, if news sites appeared at a much greater rate than other types of sites between 2002 and 2004, their point value in the DMOZ distribution of uncategorized sites would be much higher than in the distribution of categorized sites. To the extent that the two distributions look similar, our mapping method does not introduce severe categorization biases.

Table 4 shows the results of our cross-categorization, broken down by sites that are in our categorized group, and sites that are not. Column 1 shows the DMOZ category name. Column 2 shows the number of websites categorized by our primary mapping scheme that were categorized into each of

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

the respective DMOZ categories, and Column 3 expresses these numbers as percentages of total sites categorized both by our primary mapping and also by 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. This test is an effective one only if any categorization error introduced by the DMOZ process is uncorrelated with the error introduced by our own categorization methods. Because it is human edited and kept current, the DMOZ category mapping may be weighted more heavily towards current web sites. The mapping derived from the Comscore categories, on the other, is weighted more towards web sites in the 2002 data because that is when the categories were assigned. The biases from the different methods, therefore, should be uncorrelated. These tests suggest that while our mapping methods do not categorize all of the 2004 websites, they do not introduce categorization errors in a systematic way that is likely to bias the results.

While the categorization of sites generally conform to expectations, there are a few points worth considering. All sites, even if multi-purpose, are put into only one category. This can lead to difficulties in interpretation for 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. Additionally, 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 increases in a particular type of content, we take a detailed look at the disaggregate site level behavior of our panel to ensure that it is not driven by these types of categorization problems. Secondly, 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. While these applications continuously pull content to the desktop, individuals do not actually visit their websites by opening a browser and entering a URL. Because these applications still provide information that is consumed by their users, however, we treat them the same way as other websites. An additional difficulty with the data worth noting is that because it is collected at the machine level, we are unable to distinguish Internet usage by individual. Furthermore, if users spend significant time on the Internet at work or at other locations, our data represents only a portion of the individual's total information consumption. Thus our analysis makes the assumption that the Internet usage that we observe is representative of total Internet usage for that individual.

## VII. Discussion

A major concern for policymakers is ensuring that access to information technologies is widely available to encourage continued economic growth. Specifically, online content is an important way to access information, goods, and services, and as more companies require remote work, employees who want to maximize productivity must be part of the digital network. Individuals without access to these information channels may find themselves at a disadvantage in coming years, and the social cost of such a divide may be large. An important question to consider is whether governments should step in where the private sector has fallen short, and part of the answer to this question derives from understanding what people will do with broadband once they have it. Accordingly, the goal of this study was to provide evidence of how providing broadband access impacts information consumption and who benefits most.

In this study, we explored which type of Internet user benefits most from broadband. Significantly, much of the consumption gains comes from individuals who were in the lowest quintile of usage when they were narrowband users, suggesting that broadband satisfies unmet demand in a certain population. While these results are conditional on adoption, they illustrate that benefits from broadband are not limited to the gains that might be predicted from a model considering only total Internet usage. To identify some of the characteristics of this demand, we also examined consumption by content type and found that broadband access increases usage of some types of content, such as portals, entertainment, and news, more than others, suggesting that many individuals in this group have little use for available online content, but given higher bandwidth, gain utility from using the Internet as a channel for downloading and exchanging other types of content. In future work, we hope to obtain data allowing us to more precisely relate broadband adoption to the tradeoff between download time and consumption time.

## References


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

Chinn, M. and Fairlie, R. (2004) "The Determinants of the Global Digital Divide: A Cross Country Analysis of Computer and Internet Penetration", Working Paper, 2004.

The $500 Billion Dollar Opportunity: The Potential Economic Benefit of Widespread Diffusion of Broadband Internet Access, Robert Crandall and Charles Jackson

Dewan, S., Ganley, D., and Kraemer, K. "Across the Digital Divide: A Cross-Country Analysis of the Determinants of IT Penetration", Working Paper, 2004.

Goolsbee, A. (2006) "The Value of Broadband and the Deadweight Loss of Taxing New Technology", Working Paper, January 2006

Goolsbee, A. and Klenow, P. (2006). "Valuing Consumer Products By the Time Spent Using Them: An Application to the Internet", January 2006.

Hausman, Jerry A., Sidak, J. Gregory and Singer, Hal J. (2001), "Cable Modems and DSL: Broadband Internet Access for Residential Customers" . American Economic Association Papers & Proceedings, Vol. 91.

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**, No. 4, pp. 605-654.

Ford, G. (2005). "Does Municipal Supply of Communications Crowd-Out Private Communications Investment? An Empirical Study", *Applied Economic Studies*.

Ford, G. and Koutkey, T. (2005). "Broadband and Economic Development : A Municipal Case Study from Florida. *Applied Economic Studies*.

Hoffman, D. L. and Novak, T. P. (1998), "Bridging the Racial Divide on the Internet", *SCIENCE*, 280, April 17, 390-391.

Lehr, W., Osorio, C., Gillet, S., and Sirbu, M. "Measuring Broadband's Economic Impact", Working Paper (2005).

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

Rappoport, Paul, Kridel, Donald, and Taylor, Lester. "The Demand for Broadband: Access, Content, and the Value of Time", Working Paper, 2001.

**Table 1: 2002 Mean Demographics by Group**

| Variable | Adopters | Stayers |
|---|---|---|
| **income** | 4.33 | 4.20 |
| **age** | 6.68 | 6.98 |
| **child present** | 0.45 | 0.44 |
| **household size** | 3.04 | 2.94 |
| **education** | 2.70 | 2.70 |
| **usage** | 2881.0 | 2199.2 |
| **N** | 512 | 4985 |

**Figure 1: Distribution of October 2002 Usage for Non-Adopters**

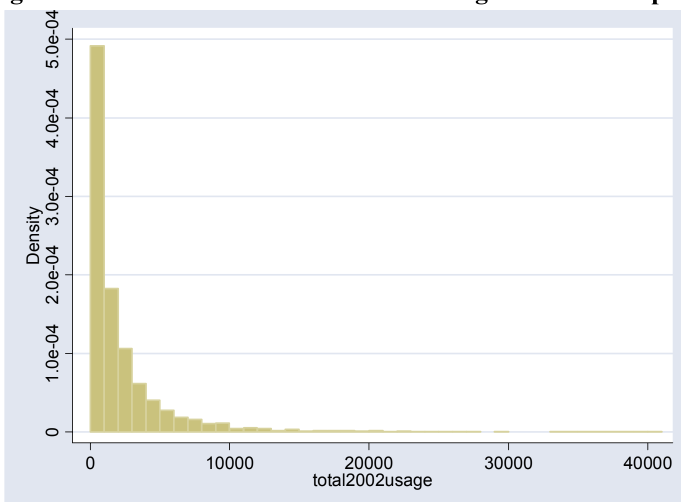


**Figure 2: Distribution of October 2002 Usage for Adopters**

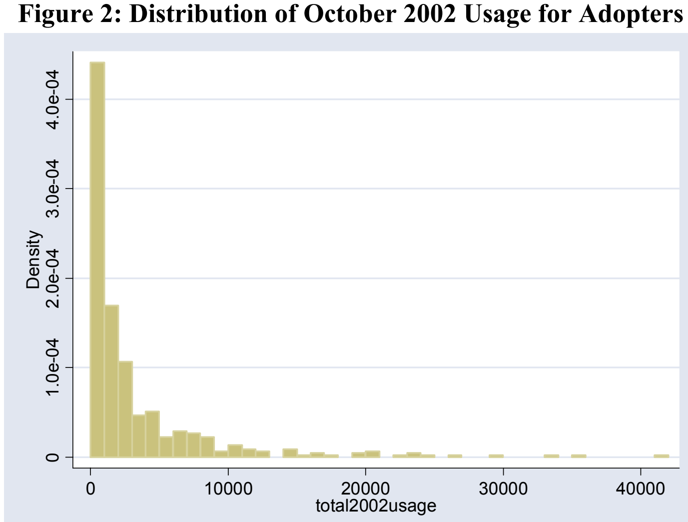

**Table 2: Usage Comparison by Group**

| Group | 2002 Duration | 2004 Duration | Difference | 2002 Pages | 2004 Pages | Difference |
|---|---|---|---|---|---|---|
| **Non-Adopters** | 2199.2 | 1767.8 | -431.4 | 1502.9 | 1174.8 | -328.1 |
| **Adopters** | 2881.0 | 3823.1 | 942.1 | 2072.1 | 2549.5 | 477.4 |

**Figure 3: Mean Usage By Income**

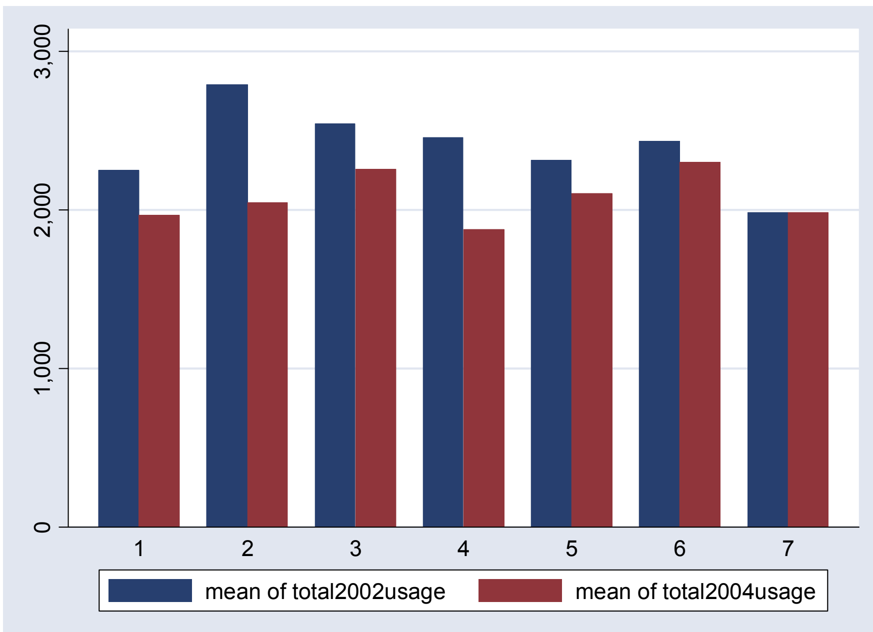

**Figure 4: Mean Usage by Age**

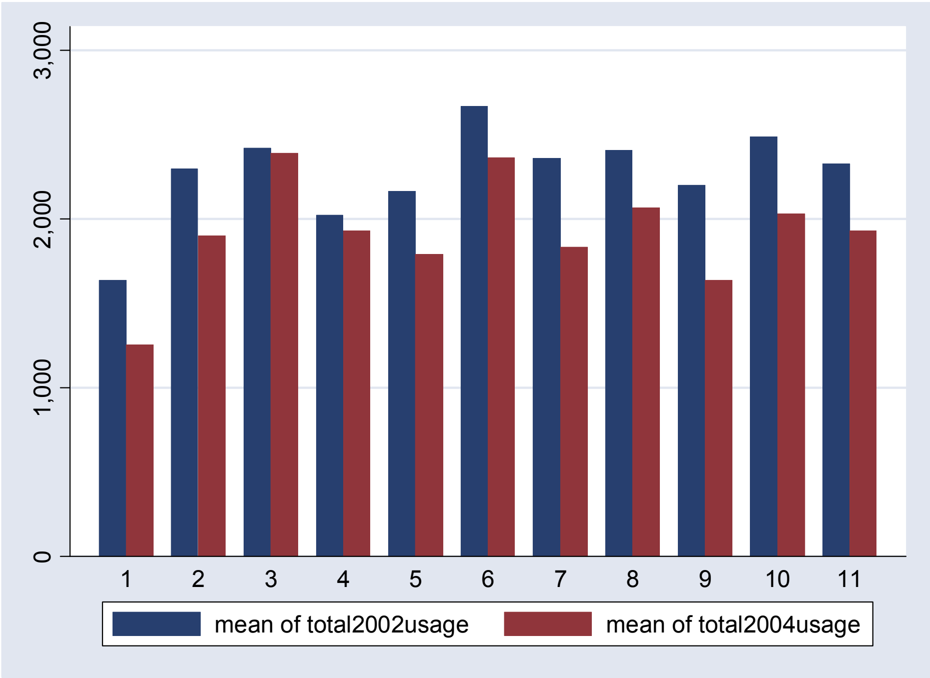

**Table 3: Representative Top Sites by Category**

| Category | Top Sites in Category | | | |
|---|---|---|---|---|
| **Adult** | sextracker.com | nastydollars.com | datecam.com | voyeurweb.com |
| **Careers** | careerbuilder.com | bepaid.com | careercast.com | jobsearch.org |
| **Society** | blackplanet.com | match.com | americansingles.com | ezboard.com |
| **Computer Apps** | hotbar.com | speedbit.com | morpheus.com | cometsystems.com |
| **Education** | ed.gov | fsu.edu | blackboard.com | umn.edu |
| **Entertainment** | neopets.com | slingo.com | windowmedia.com | igl.net |
| **Health** | webmd.com | ediets.com | healthandage.com | menshealth.com |
| **Finance** | realtor.com | bankofamerica.com | schwab.com | fidelity.com |
| **Regional** | cleveland.com | floridalink.com | nj.com | al.com |
| **Business** | microsoft.com | mediaplex.com | 180solutions.com | goback.com |
| **Home & Living** | allrecipes.com | hgtv.com | petfinder.org | bhg.com |
| **News** | weatherbug.com | myweather.net | cnn.com | weather.com |
| **Portals** | yahoo.com | msn.com | aol.com | google.com |
| **Reference** | digitalcity.com | about.com | gohip.com | theuseful.com |
| **Shopping** | peoplepc.com | amazon.com | nextag.com | mypoints.com |
| **Sports** | sportsline.com | nascar.com | mlb.com | trapshooters.com |
| **Travel** | expedia.com | orbitz.com | milesource.com | hotwire.com |
| **Services** | gator.com | whenu.com | passport.com | passport.net |
| **International** | aif.ru | bannerbank.ru | hobby.ru | sandesh.com |
| **Automotive** | autotrader.com | yachtworld.com | cars.com | boats.com |
| **Auction** | ebay.com | honesty.com | ebaymotors.com | auctionworks.com |
| **Ads** | atdmt.com | atwola.com | realmedia.com | trafficmp.com |
| **Market Research** | npdor.com | mysurvey.com | 4at1.com | surveyspot.com |
| **Other** | waol.exe | centrport.net | akamai.net | liveperson.net |
| **Government** | tx.us | noaa.gov | ny.us | pa.us |
| **Unclassified** | mywebsearch.com | freeslots.com | coalregion.com | blast.net |

**Table 4: Matched Sampled Estimator - Overall Usage**

| | Duration | | | Page Views | | |
|---|---|---|---|---|---|---|
| | Adopters | Stayers | Diff | Adopters | Stayers | Diff |
| 2002 | 1743.8 | 1722.8 | | | | |
| 2004 | 3555.9 | 1564.9 | | | | |
| Diff | 1812.1 | -157.9 | **1757.16** | | | **842.06** |

**Figure 5: 2002 % Usage by Category**

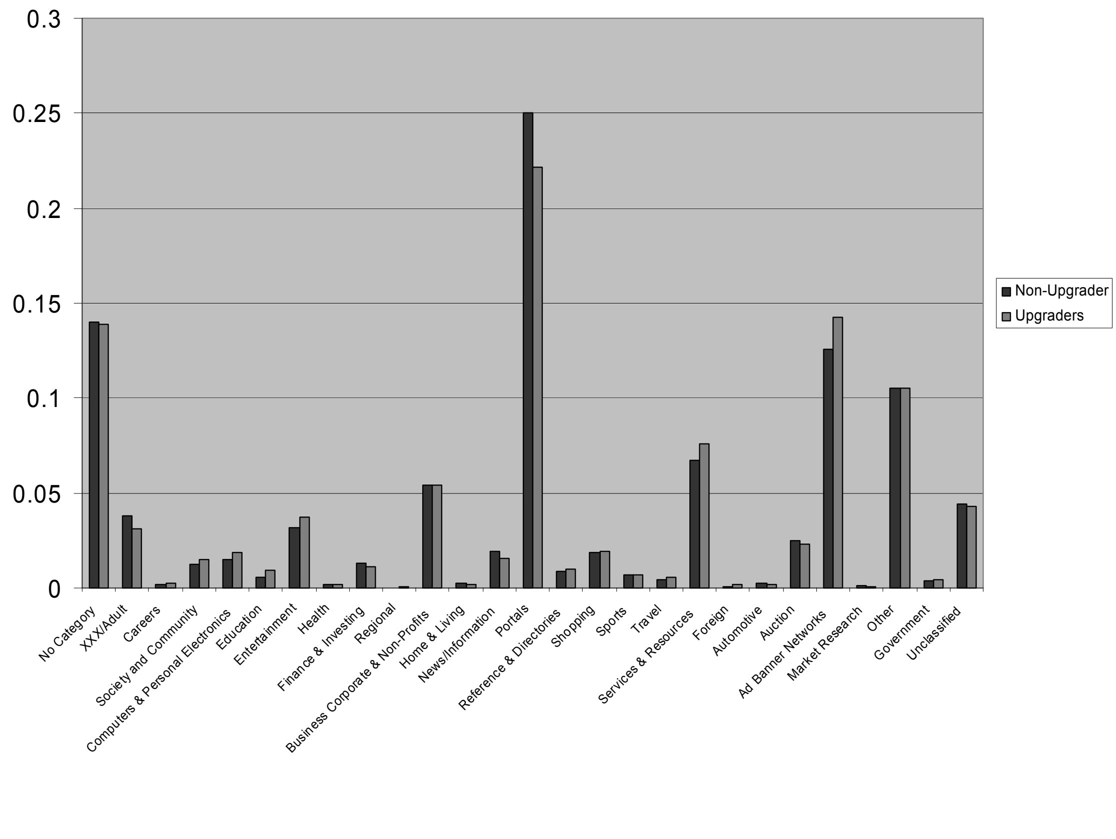

**Table 5: Comparison of Matched Samples by Category**

| Category | Adopters | Stayers | Category | Adopters | Stayers |
|---|---|---|---|---|---|
| Adult | 54.14 | 71.76 | Reference | 15.2 | 16.4 |
| Careers | 3.30 | 4.31 | Shopping | 31.4 | 37.8 |
| Society | 13.75 | 21.48 | Sports | 7.0 | 12.0 |
| Computer Apps | 29.43 | 27.23 | Travel | 7.51 | 9.12 |
| Education | 12.3 | 11.0 | Services | 143.1 | 121.4 |
| Entertainment | 48.47 | 53.13 | International | 1.07 | .74 |
| Health | 2.99 | 3.74 | Automotive | 2.71 | 5.38 |
| Finance | 15.21 | 20.68 | Auction | 20.7 | 40.5 |
| Regional | .31 | .58 | Ads | 202.2 | 196.1 |
| Business | 89.8 | 91.0 | Market Research | .91 | 1.7 |
| Home | 3.16 | 4.35 | Other | 214.0 | 179.8 |
| News | 25.9 | 28.2 | Government | 5.77 | 8.06 |
| Portals | 447.5 | 446.5 | Unclassified | 75.8 | 79.2 |

**Table 6: Usage Increases By Quintile**

| Quintile | 2002 Usage | Usage Increase |
|---|---|---|
| 1 | 75.87 | 2663.32**<br>(1037.16) |
| 2 | 432.32 | 548.28**<br>(219.77) |
| 3 | 1058.32 | 1487.50**<br>(657.24) |
| 4 | 2255.54 | 1464.79**<br>(608.92) |
| 5 | 7486.76 | 1317.47<br>(746.7)* |

**Figure 6: Total 2004 Usage By 2002 Usage Quintile for Broadband Users**

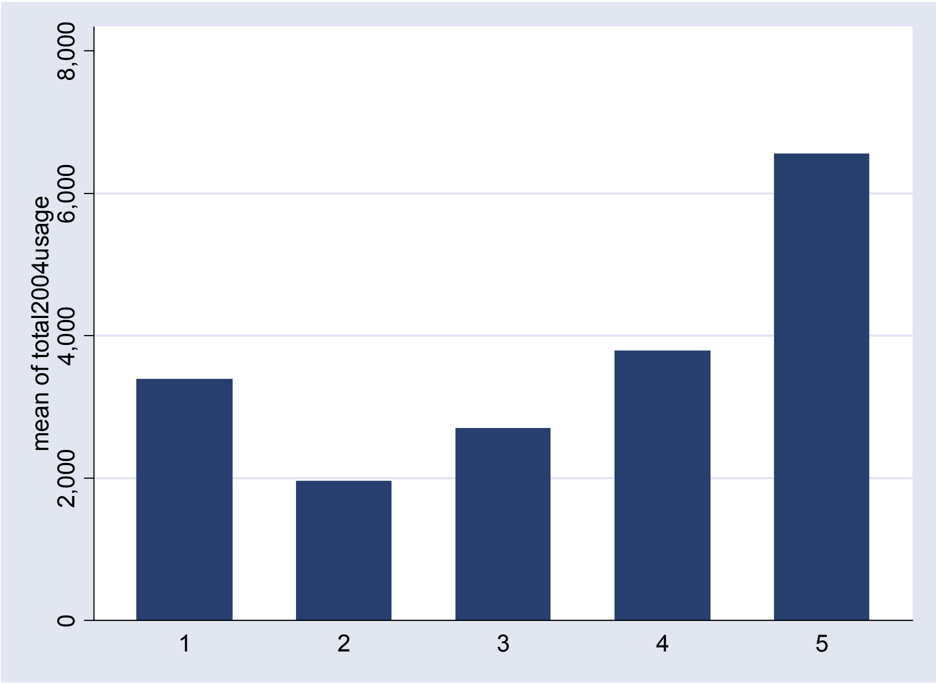

**Table 7: Usage Changes by Category**

| Content Type | 2002 Usage (Adopters) | Difference-In-Differences |
|---|---|---|
| News | 40.6 | 54.33**<br>(22.07) |
| Portals | 581.03 | 403.84**<br>(165.79) |
| Business | 142.3 | 19.00**<br>(8.09) |
| Entertainment | 98.69 | 104.46**<br>(24.98) |
| Adult | 81.49 | 0.26<br>(5.00) |
| Sports | 17.96 | 11.39**<br>(3.89) |
| Careers | 6.24 | 1.33<br>(1.43) |
| Education | 24.35 | 8.67**<br>(3.80) |
| Health | 5.40 | 0.66<br>(0.57) |
| Computer Apps | 48.96 | 157.83<br>(92.78) |
| Government | 10.87 | 2.81<br>(1.32) |
| Services | 198.93 | 127.50<br>(68.54) |
| Shopping | 50.68 | 14.20<br>(8.78) |
| Ad Networks | 373.61 | 212.38**<br>(53.11) |

**Table 8: Usage Changes By Category By 2002 Usage Quintile**

| | (1) | (2) | (3) | (4) | (5) |
|---|---|---|---|---|---|
| adult | -3.6 | -7.0 | 30.2 | -3.6 | 58.0 |
| careers | 0.9 | -0.9 | 0.3 | -0.4 | -3.2 |
| society | 7.4 | -2.1 | 7.3 | 36.7 | -52.4 |
| computer apps | 486.5 | 158.2 | 47.8 | 37.8 | -33.7 |
| entertainment | 9.1 | 6.2 | 52.9 | 34.7 | -60.6 |
| finance | 15.1 | 2.6 | 2.7 | 2.6 | 0.1 |
| health | -0.3 | 1.7 | -0.4 | -3.1 | -1.0 |
| business | 52.7 | 201.2 | 48.8 | 21.4 | 320.8 |
| shopping | 37.4 | 6.6 | 25.5 | 15.6 | -19.3 |
| reference | 7.3 | 1.0 | -5.3 | -7.3 | -12.9 |
| auction | 8.6 | 37.7 | 7.8 | 22.3 | 97.4 |
| international | 0.0 | 0.1 | -4.5 | 0.4 | 4.9 |
| market research | -1.3 | -2.5 | -3.9 | -2.4 | -24.4 |
| other | 60.0 | -10.3 | 17.2 | 13.5 | -104.6 |
| education | 27.3 | 18.1 | 14.1 | -5.2 | -28.8 |
| regional | -0.1 | -0.6 | -0.6 | -2.7 | -0.2 |
| home | 2.6 | 3.2 | 6.0 | 1.0 | -3.7 |
| sports | 1.6 | 4.8 | -4.5 | 40.8 | 13.9 |
| travel | 13.6 | 0.0 | 2.3 | -3.7 | -2.8 |
| services | 179.5 | 15.5 | 116.8 | 76.4 | 178.9 |
| automotive | 1.1 | 0.0 | 0.2 | 2.0 | -1.2 |
| government | 4.9 | 1.1 | -5.0 | -3.4 | 3.3 |
| classified | 16.6 | 6.1 | 0.3 | 9.3 | -32.4 |
| portals | 567.5 | 79.1 | 299.8 | 685.2 | 570.5 |
| news | 45.2 | 2.2 | 244.5 | -0.5 | 26.9 |
| ads | 394.3 | 57.6 | 188.5 | 245.4 | 65.6 |

**Figure 7: Broadband Adopters, 2004 Computer Apps by Quintile**

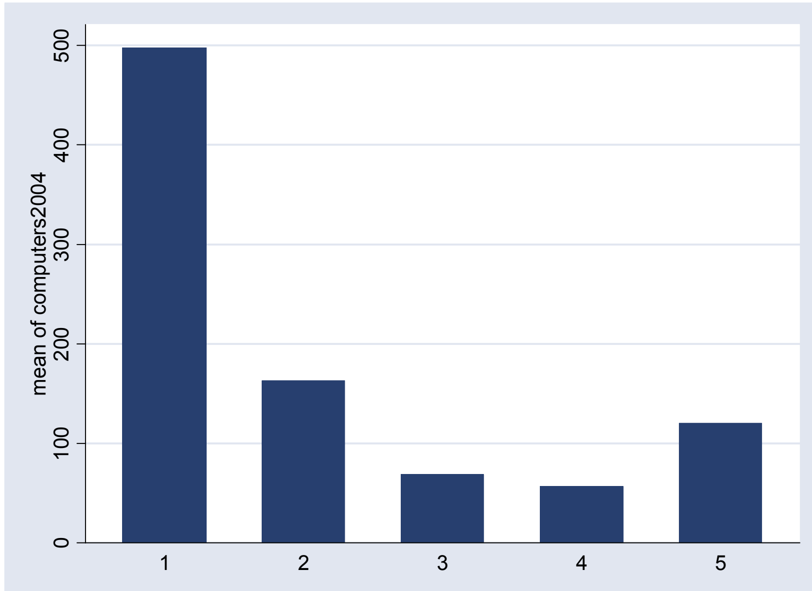


**Figure 8: Non-Adopters, 2004 Computer Apps by Quintile**

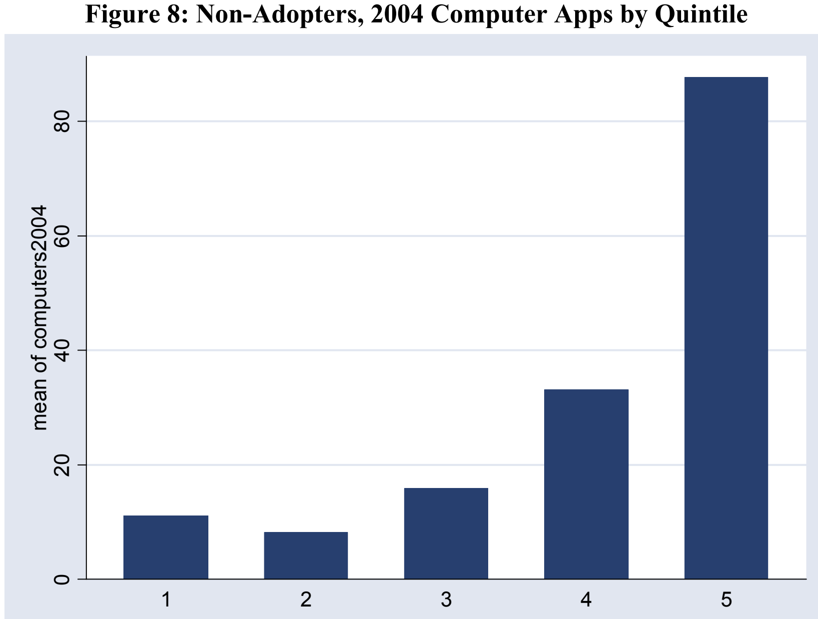

**Figure 9: Broadband Adopters**

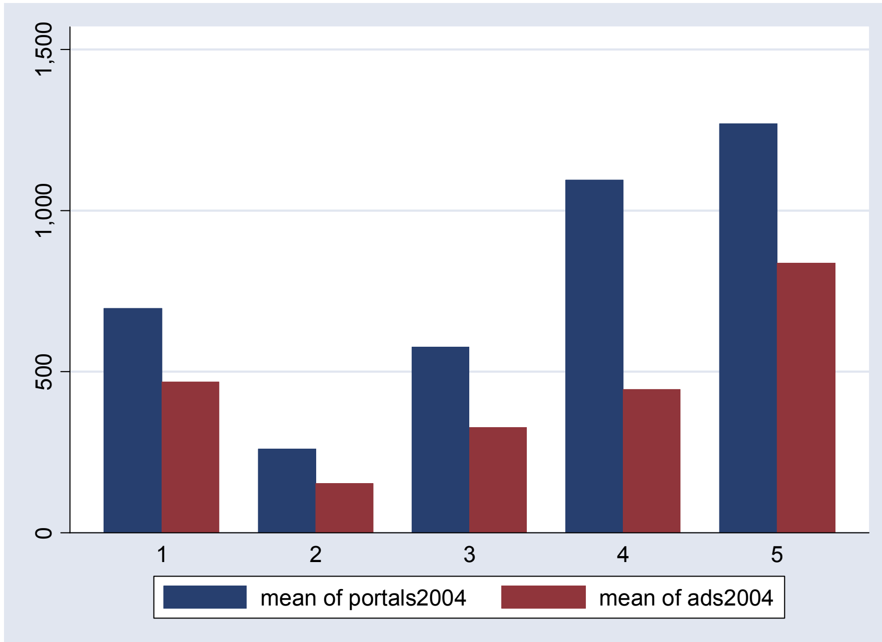


**Figure 10: Stayers**

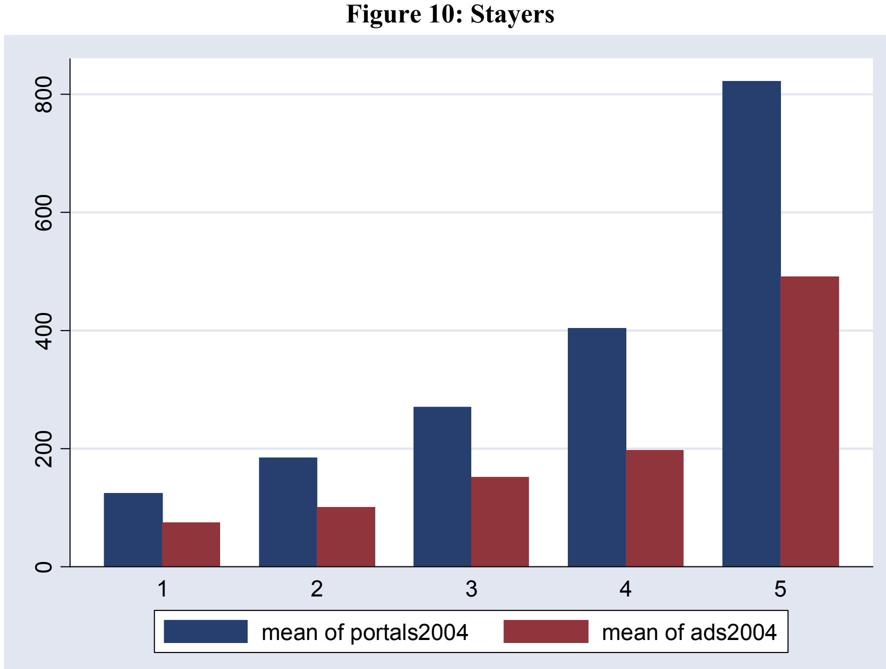

**Table 4: DMOZ Categorization**

| 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 |