“Learning to Surf: Spillovers in the Adoption of the Internet”

Michael R. Ward

University of Texas at Arlington

June, 2010

ABSTRACT

This paper develops an identification strategy to generate unbiased estimates of Internet usage spillovers using a unique data set of US households. I identify multiple potential sources of learning including those from the household’s locality, from educational Internet subsidies, from universities, and from word-of-mouth. I find general support for all four sources, but the locality and subsidy results are both more robust and larger. These findings have implications for policies to encourage Internet use as well as for identification strategies for the effects of the Internet on behavior.

JEL Codes: L86, D83, D62

Keywords: Internet, Learning, Spillovers

"Learning to Surf: Spillovers in the Adoption of the Internet"

I. Introduction

This paper explores spillover mechanisms at work in household decisions to use the Internet. The adoption of a new and complex technology by one agent often depends on what the user has learned from contact with those who have already adopted the technology. Four potential sources of learning are identified and tested. I construct measures associated with learning from local community Internet users, from K-12 educational Internet subsidies, from the nearby university communities, and from more distant social networks through word-of-mouth conversations. The locality and K-12 subsidies appear to be large while my estimates for university and word-of-mouth sources of learning are both smaller and less robust.

Identifying spillovers is complicated by similarly situated individuals experiencing similar and unobserved "shocks" to their adoption decisions. Measuring spillovers requires a methodology that insures that they can be identified independent of these unobserved common "shocks." I use a unique household level data set that allows me to implement two identification strategies. First, instrumental variables are constructed from demographic variables at the level of the spillover measure. Second, I measure changes in Internet usage at two points in time that implicitly control for unobserved time-invariant household characteristics. By construction, the instrumented endogenous variable is mostly uncorrelated with unobserved household-level time-varying shocks.

The Internet in particular is thought to be an important source for more efficient markets, more efficient production, and more efficient consumer decisions. Some growth and development literature identifies the role technology diffusion as central (Grossman and

Helpman,1991) and Parente & Prescott, 1994). Moreover, spillovers are related to network externalities which have been of interest in industrial organization (Economides,1996).

Network externalities could have large impacts on the dynamics of many industries where they are thought to be present. Users adoption decisions must take into account the future size of the network so as to avoid unpopular networks. Firms may inefficiently invest in proprietary networks from which they can earn rents. There is a growing literature of evidence of network externalities affecting many different industries. Some examples of network externalities are numerically controlled machines (Karshenas, and Stoneman, 1993), spreadsheets (Gandal, 1994), automated teller machines (Saloner and Shepard, 1995), prescription antiulcer drugs (Berndt, Pindyck, and Azoulay, 2003), electronic bank payments Gowrisankaran and Stavins, 2004) and automatic clearinghouse functions (Ackerberg and Gowrisankaran, 2006).

Independent of externalities, there is growing interest in determining the effects of Internet use, or the use of specific Internet applications, on users' behaviors (Ellison et al., 2007, Kendall, 2007). However, without a robust identification strategy, many studies finding an association between Internet usage and a behavior may suffer from possible reverse causation or selection bias. Finding external influences, such as measures related to various forms of local spillovers, provides possible instrumental variables for Internet adoption. Future studies could use similar measures to potentially identify variation in Internet usage related to the size of these spillovers that is not a result of the behavior under study. In this way, the causality from Internet usage to the behavior is more firmly established.

Spillovers occur when the decisions of one agent are affected by the independent decisions of those around her. This could occur because a more reticent adopter benefits from the information gleaned by more adventurous early adopters or because a market for related goods and services spurred by the early adopter is available to the later adopter. Network externalities are typically distinguished from spillovers by possible reciprocation in the latter (for example, "Show me how to do this on the Internet and I will increase your allowance"). While this distinction is important for policy implications, since possible reciprocation is not observed here, we will continue to refer to spillovers and only argue for the network externality interpretation in certain cases.

The analysis below identifies potential spillovers emanating from the E-Rate educational Internet subsidy program. Prior research has identified the effects of targeted policies on classroom Internet access. Puma et al. (2000) and Goolsbee and Guryan (2006) find that the E-Rate subsidies to schools increased the number of Internet enabled classrooms. However, Goolsbee and Guryan (2006) and Ward (2006) find modest, if any, evidence of student achievements related to the E-Rate. Below, the E-Rate funding is found to affect Internet use for households with school-aged children. Hoffman and Novak (2000) discuss the role of higher education with Internet adoption. Similarly, Goldfarb (2006) finds that attending university has had a significant effect on Internet use later in life. The analysis below allows for the possibility of proximity spillovers from universities to households in the county.

It is widely accepted that consumer-to-consumer communication or 'word-of-mouth' has a significant impact on the marketplace. Word-of-mouth is believed to be a driver of competitive pricing, the formation of social movements, and the diffusion of innovations (Frenzen and Nakamoto, 1993). Nevertheless, measuring word-of-mouth effects has proved extremely difficult. Recently, word-of-mouth research enabled by the archiving of these inter-consumer conversations in the online context has demonstrated that television show ratings follow online conversations (Godes and Mayzlin, 2004) and that online consumer reviews affect product sales (Chevalier and Mayzlin, 2007). My measure of word-of-mouth is related to more telephone calls to places with more Internet usage and show that it is positively related to a household's Internet use.

Various policy implications might follow from this study's findings. Positive network externalities may indicate that the adoption rate is less than socially optimal that might justify public subsidies. The E-Rate is just one subsidy program that is shown to have a positive effect on the targeted audience. Likewise, one aspect of universities shown here is increased Internet usage by households exposed to the universities. This, along with other potential positive university related externalities, may justify the current public support for higher education. However, it is not clear whether these programs represent greater or less than the optimal "Pigovian" subsidy.

II. Modeling Internet Spillovers

The probability that a household will use the Internet is modeled as a function of various spillover measures and household characteristics. Spillovers are identified to emanate from four potential sources: local Internet usage, E-Rate funding, university exposure and word-of-mouth. Locality based learning is measured by the effect of the share of Internet users in the household's county on usage decisions. This is constructed by simply calculating the share of households in a county and quarter that say they use the Internet. E-Rate based learning is measured by recent funding in the county and, especially, E-Rate funding interacted with the presence of school-aged children in the household. This is constructed by calculating the per capita level of E-Rate funding in all school districts in the county and constructing an indicator variable for children aged 6-18. University based learning is measured by the share of the county population enrolled at a university. This is constructed as the ratio of college enrollments in four year schools in the county to the county's total population. Word-of-mouth based learning is measured by the average share of Internet users in states where a household's telephone calls are terminated. For a subset of households that provide telephone calling information, this is calculated by identifying the state called for each long distance call made and averaging the Internet usage shares across these states.

In addition to these variables, I control for many different household demographic variables. These characteristics include the usual demographics, including age, income, education, socio-economic status, household size, children, race, and occupation. It is likely that Internet use is more valuable for larger households, younger households, those with more education and those that are wealthier.

While the dataset I use is rich in detail, it is likely that remaining unobserved household and county characteristics also affect the decision to use the Internet. For example, the clustering of high technology firms may attract households with unobservable technologically sophistication. Likewise, unobservable Internet Service Provider price and quality may induce more or fewer households to subscribe. Identifying plausible causal effects from these selection effects is at the crux of the identification strategy. Below, I denote households with the subscript i, the county a household resides with the subscript j, and the time period with t. Superscripts O and U denote observable and unobservable variables.

$$\Pr\left(Internet_{it} = 1\right) = \alpha CountySpill_{ijt}^{O} + \beta demog_{it}^{O} + County_{ijt}^{U} + demog_{it}^{U} + \varepsilon_{it}$$

Either unobservable household or county-level variables could simultaneously affect the spillover measure and the Internet adoption decision. If so, they would lead to biased estimates of the spillover effects.

One way to address these potential sources of biases is to treat county-level spillover variables as endogenous and use an instrumental variables estimator. Otherwise, estimation bias could emerge from spillover measures being correlated with unobserved differences in county characteristics. For example, price reductions or the rollout of high speed Internet service in a county will simultaneously increase the probability of all households using the Internet. Such an unmeasured county-level effect will lead to an upward bias in estimates of spillover effects. Instrument set I employ primarily includes the demographic variables described above aggregated to the county level for the current time period. In this way, the variation in the projection of the county-level spillover measures on these county-level demographic variables from the first-stage will be due only to observable county demographics and will be mostly free of any contribution from unobserved shocks.

$$\Pr\left(Internet_{it}\right) = \alpha \widehat{CountySpill}_{ijt}^{O} + \beta demog_{it}^{O} + County_{ijt}^{U} + demog_{it}^{U} + \varepsilon_{it}$$
$$CountySpill_{ijt}^{O} = \gamma \overline{demog}_{ijt}^{O} + \nu_{ijt}$$

These county-level characteristics are used to identify the measures of spillovers for the county independent of any unobserved time-varying county effects. As a consequence, this projection will be largely uncorrelated with unobserved county shocks in the second-stage. This method was used successfully in Goolsbee and Klenow (2002).

In addition, it is possible to exploit changes in household decisions in different time periods. The data allow for the sampling of a subset of households at two different points in time. Thus, the model can be estimated based on first-differences of all variables.

$$\Pr(\Delta Internet_{it}) = \alpha \Delta \widehat{County}Spill_{ijt}^{O} + \beta \Delta demog_{it}^{O} + \Delta County_{ijt}^{U} + \Delta demog_{it}^{U} + \Delta \varepsilon_{it}$$
$$\Delta CountySpill_{ijt}^{O} = \gamma \Delta \overline{demog}_{ijt}^{O} + \Delta \nu_{ijt}$$

Doing so eliminates any potential bias due to time-invariant unobservable household characteristics. Many changes in key demographic characteristics are observable and changes in these observed demographic characteristic are included in the specification. It may still be possible for bias to emerge from unobserved time-varying household characteristics. For example, bias could be due to a large change in the negotiated union contract that changes total compensation but not reported incomes from, say, a more generous benefits package. Multiple households could substitute their own income toward Internet use affecting both the spillover measures and the household Internet usage decision county simultaneously.

III. Internet Use, E-Rate, College Enrollment, and Word-of-Mouth Data

Internet usage data come from, TNS Telecoms ReQuest® Market Monitor, a large US survey of households conducted between 1999:3 and 2001:4. This is a time period when Internet usage was not universal and many potential users in the US were just discovering the potential of using the Internet. This period predated Facebook, YouTube, and iTunes. Most Internet applications were related to E-commerce (via Amazon, Expedia) or communication (Email rather than instant messaging). The education establishment also saw the potential of the Internet for educational enhancement and made early investments in the technology. Annual data for E-

Rate funding of school and library Internet access subsidies for the various school districts in the US were collected from the Universal Service Administrative Company (USAC). College enrollment data were collected from the US Department of Education's Integrated Postsecondary Education Data System (IPEDS). Finally, a subset of the surveyed households provided detailed telephone call data from which a 'word-of-mouth' measure.

The primary dataset is from a unique national panel that surveys households on their use of consumer communications and electronics products and services. The panel, TNS Telecoms' ReQuest® Market Monitor, samples about 30,000 households every quarter from the 48 continental states and D.C. These households are selected from a panel of over a quarter-million participating households. I have available to me 10 quarters of the panel running from third quarter of 1999 through the fourth quarter of 2001. Most observations in different quarters represent repeated cross-sections of different households. However, about 20% of households are resampled each quarter allowing examination of changes in behaviors related to changes in demographic characteristics of a household.

The dataset contains responses to a mail survey that includes a long list of household demographic measures relating to age, income, race, education, occupation, etc. These generate a substantial number of dummy variables for each variable. Those used here include ten categories for the age of the respondent, seven for education level, five for race/ethnicity, 16 for income level, eight for general socioeconomic status, eight for the presence and number of children of different ages, five for household size, and 13 each for the occupation category of male and female householders. Of particular interest are questions relating to Internet use and household composition. The Internet related questions used here ask whether anyone in the household uses the Internet. From the household composition questions, it is possible to determine if the household contains a school aged child (6-18 years) or a college student living either at home or away at school. The county in which the household resides is one of the smallest consistently measured geographic identifiers and is used to merge these data with other sources.

Locality spillovers are measured from the fraction of households using the Internet in the county. This variable is constructed by aggregating the household data on Internet use to the county level for each of the 10 quarters. To avoid small sample problems, the analysis includes only households if the county had at least 200 households in the TNS data over the 10 quarters. Because the TNS data include over a quarter-million observations, 281 counties meet this threshold. Similarly, the 78 household demographic variables are aggregated to the county-quarter level generating 78 variables measuring the fraction of households with that demographic characteristic. Finally, the changes in these 78 variables between samples are used as instrumental variables for the change in county-level Internet use.

E-Rate data are available via download from USAC[1]. These data include, among other variables, the funding commitment levels for each public school district in the US since the program began in 1998. Not all districts receive E-Rate subsidies each year and, because subsidy rates are increase with the number of low income students in schools being served, per capita amounts received may vary across areas. Districts were matched to counties based on the zip code of the district offices. The analysis below uses the per capita amounts and, as with Goolsbee and Guryan (2006) and Ward (2007), includes lag values of E-Rate funding to capture the possibility of learning over time. For each household, I merge the county's per capita E-Rate funding level for the previous year. E-Rate funding is specifically targeted toward children

[1]Universal Service Administrative Company <www.sl.universalservice.org/>.

through subsidies to schools. Because of this, I interact the per capita E-Rate funding level with the presence of school aged children in the household. This would measure the intended effect of the E-Rate program. In addition, specifications include E-Rate funding uninteracted with the presence of school aged children in the household. This could measure the effect of an unintended spillover of the E-Rate program to households without school aged children.

The college related spillover hypothesis discussed above is that greater exposure to colleges and universities leads a household to be more likely to use the Internet. Students, faculty and staff have been among the earlier adopters of Internet technology. A spillover from greater exposure to these individuals could lead even non-university related households to use the Internet. The size of these spillovers are measured as proportional to the fraction of the county population enrolled in college. University enrollment, as used here, should be thought of as a proxy for a larger and more active university community in general and not merely more students. Enrollment data were collected from the US Department of Education's Integrated Postsecondary Education Data System (IPEDS).[2] From the universe of all schools, I selected the nearly 2,300 baccalaureate degree granting institutions where Internet exposure typically has been more intense. The schools included in the sample had enrollments totaling about 10 million students during the sample period. The college or university's city was mapped into its county and the enrollments of multiple schools in a county were aggregated to a county total. These colleges and universities are contained within about 900 of the more than 3,200 counties in the US. Table 1 lists counties with the highest enrollment to population ratios.

---

[2]Integrated Postsecondary Education Data System <nces.ed.gov/ipeds/>.

College experiences allow for both direct effects and indirect spillover effects. First, there could be a direct effect for household children attending college. The household level data allow for test to distinguish between the effects of college students living at home versus away at school. Variables are constructed indicating a college student at home or away. Second, independent of this, there could be an effect from living in a community with a higher fraction of college students. This could be because a member of the household becomes is more likely to become affiliated with a college when enrollment rises or because household members have more contact with those who are affiliated with a college when it grows.

Finally, as a companion to the TNS survey data, TNS conducted a ‘bill harvesting’ program. A subsample of survey participants submitted their telecommunications bills to TNS for detailed data entry. Between 10 to 20 percent of the sample submitted long distance telecommunications bills that list all calls made during a month. From these calls, I was able to map the area code and called number prefix into the state in which the call was terminated. For each call, I merged the terminated state’s average Internet usage from the TNS survey for the quarter. The final word-of-mouth measure is the average of this state-level Internet usage across all calls made that month. I hypothesize that households with higher word-of-mouth values are more likely to have discussed Internet usage than households with lower values and that these discussions could have lead to Internet usage decisions.

Table 2 reports summary statistics for the key variables in levels and the changes across resamples. Over the sample, almost two-thirds of households said that someone in the household used the Internet but that rose substantially between resamples. Just over a quarter had school-aged children. E-Rate subsides came to just over $2 per person. Almost one-third had a college student living at home but only 4% had a college student away at school. College enrollment averaged about 3% of the county population. The weighted average Internet share among those called was about 60% and rose between resamples.

IV. Empirical Results

I present results for the cross-section and the changes in cross sections separately. The determinants of the change household level Internet use are estimated using a simple Linear Probability Model (LPM) rather than a more standard Probit. This eases the computational burden, especially for the IV estimator with many exclusion restrictions. Also, it allows for comparability in results when the dependent variable is binary (cross-section) or when it can take on three values (changes in cross-section).

A. Results from Cross-Sectional Observations

Table 3 reports the results from an Ordinary Least Squares (OLS) and two Instrumental Variables (IV1 an IV2) estimators of the determinants of Internet Use. The OLS estimator treats all independent variables as exogenous. The IV estimators attempt to address the possible selection bias issues discussed above. The IV1 specification treats the Internet Share in the County, E-Rate Funding in the County, E-Rate Funding interacted with School-Aged Children as endogenously determined. The IV2 specification expands the set of variables assumed to be endogenously determined to also include all county level variables, in this case adding the University Share of County Population.

Table 3 reports coefficient estimates only for the variables of interest and suppresses most coefficients of demographic variables.[1] First, the Internet share in the county is positively associated with individual Internet usage. This result is evident in all three specifications and provides support from general locality learning. Consistent with unobserved selection, the IV coefficient estimates are smaller in magnitude but still highly statistically significant. Second, E-Rate funding in the previous year is associated with greater Internet Usage but only for households with school-aged children. In fact, E-Rate funding uninteracted is associated with lower Internet usage. This result could be due to selection bias since E-Rate funding levels increase in school districts with more poor families. That is, unmeasured increases in the number of more poor families lead to both more E-Rate funding and less Internet usage. Overall, however, this provides support for E-Rate funding meeting its policy objective of encouraging Internet usage among school aged children. Third, households with college students are more likely to use the Internet, more so if the student lives at home rather than away at school. This in itself is consistent with universities being at the forefront of Internet adoption. Controlling for this, living in a county with more college students is also associated with more Internet usage. This is consistent with spillover learning from the university community to the residents of the college town.

Table 4 reports similar coefficient estimates for the subsample of households that provided long distance calling information. Again, this sample is only one-sixth the size of the entire cross-sectional sample and yields some differences in coefficient estimates from table 3. First, the coefficient estimates for the Internet share in the county are virtually unchanged. Second, while the E-Rate funding level estimates have the same sign as before, they are no

[1] In general, Internet usage is more prevalent among households with higher income, with younger households, in later time periods, with larger households, and with white-collar occupations.

longer statistically significant. Third, the college-student results have the same qualitative interpretation as table 3. Fourth, households that call more often to states with higher Internet usage tend to more often use the Internet. This is consistent with non-locality based word-of-mouth learning. It is plausible that having regular conversations with a existing Internet user eases the transition to Internet use.

B. *Results from Changes in Cross-Sectional Observations*

While the above results provide support for the main hypotheses, some are problematic. In particular the E-Rate results suggest that E-Rate funding levels are not exogenous to county characteristics. To address this issue, I estimate the above equations in first-differences for the nearly 37,000 households that are re-sampled. This is equivalent to including household level fixed effects so that differences across households or counties that are time-invariant will not contribute to the coefficient estimates. Coefficient estimates are based solely on variation over time for a household.

The dependent variable is no longer a binary variable. In this case, along with no change in Internet usage (=0), we observe both Internet adoption (=1) and the more rare de-adoption (=-1) decisions.[2] All specifications include changes in household demographic variables between resamples but, again, most are not reported. Many changes in individual demographic variables are not significant but they are jointly significant in all specifications at standard confidence levels. Table 5 replicates table 3 for this sample. Again, I report OLS, IV1 and IV2 specifications where now changes in county-level variables are used as instruments. Note that the

[2] The qualitative results are largely unchanged when a multinomial or an ordered logit estimator is employed.

goodness-of-fit measure, R-squared, is much smaller. This is because time-invariant household determinants are differenced from the data. For the most part, this specification yields estimates more consistent with expectations. The first result, that an increase in the Internet share in the county has a positive and significant effect on a household's Internet use, is consistent with implications from table 3.

Also consistent with table 3, these estimates indicate that increases in E-Rate funding are associated with Internet adoption but only for households with school-aged children. In particular, E-Rate funding is estimated to have no effect on households without school-aged children. This is further evidence that the cross-sectional negative correlation was due to cross-sectional variation in E-Rate funding being related to cross-sectional variation in county demographics. These specifications also suggest that, independent of E-Rate funding, when a child in a households becomes of school-age, the household is more likely to use the Internet.

There are some differences in the college related coefficient estimates from table 3. First, while having a new college student at home is associated with Internet use, having a new college student away at school has no effect. Second, the magnitude of this college student at home coefficient is about one quarter the size as its comparable estimate in table 3. Third, the change in the enrollment share is not different from zero when the variable is treated as endogenous in the IV2 specification. This could be because due to weak instruments. The instrument set, changes in county demographic variables, may not be related to changes in university enrollments.

For completeness, in table 6 I also report coefficient estimates for the subset of resampled observations that also provided long distance calling information. Because this sample includes less than 4,000 of the original 132,000 observations, it is less likely that it is representative. Still, the estimates provide support for general locality based learning hypothesis. Households with children than become of school-age are much more likely to use the Internet. However, E-Rate funding for households with school-aged is now only marginally significant in the OLS specification. The magnitude of college student at home estimate is now similar to those in table 3 but the college based learning result is rejected. Finally, a change in calling patterns toward states with more Internet usage is associated more Internet usage.

C. *Interpretation*

Across almost all specifications and samples there is support for locality based learning. Households are more likely to use the Internet when more other households in the county use the Internet. Some of the effect from the OLS estimates appears to be due to unobserved selection bias, but this result remains in IV estimates and in when household, and thus county-level, fixed effects are accounted for. The IV estimates of the general locality spillovers suggest that an 11% increase in county-level Internet usage, about one standard deviation in the sample, leads to a 3% increase in the probability that any household's will use the Internet. Support for this form of learning is not surprising since it is similar to that found for computer use (Goolsbee and Klenow, 2002).

While there is less support for learning due to E-Rate funding, there is still substantial support from estimates in table 5. Independent of an E-Rate effect, there is 17 to 19% more Internet usage when school-aged children become present in a household. With household fixed effects, time-invariant differences across counties (more lower income families) does not induce the negative correlation between overall Internet usage and E-Rate. Households without school-aged children are unaffected by E-Rate funding while those with children increase their Internet usage substantially. These findings suggest that the E-Rate program is succeeding in promoting Internet usage among school-aged children but is not generating spillovers beyond these families. The mean level of E-Rate funding is $2.32 per capita with a standard deviation of $3.81. Thus, a one standard difference in funding is associated with a 5 to 6% change in the probability that a household with school-aged children will use the Internet. Note that this last value is about one-third the size of the direct effect of a new school-aged child in the household.

University enrollment appears to affect household Internet usage, but perhaps only if the college student lives at home. A household member in college but living at home is estimated to increase the probability that a household uses the Internet by 4-5%. Independent of this, the university related spillover measure is significant only it is treated as exogenous. However, in this case, we are likely to have weak instruments, the main ones being county-aggregates of changes in household demographic indicators. It seems plausible that a change in a household's Internet decisions and changes in university importance are largely independent of unobserved county-level shocks. In this case, the estimates suggest that a one standard deviation change in university enrollment share leads to 0.5 to 1.0% change in Internet usage.

The evidence for word-of-mouth spillovers from telephone calling patterns is somewhat problematic to interpret. On the one hand, the estimates from tables 4 and 6 are always positive and statistically significant. On the other hand, the sub-samples these estimates are drawn from are much smaller and possible non-representative. With these caveats, an increase of 0.076 in the average Internet share of the called states, about one standard deviation, leads to about a 2 to 3% increase in Internet usage.

V. Conclusion

This paper develops a methodology to identify household Internet usage spillovers. Using a large panel of US households from 1999-2002, I find support for the existence of multiple sources of spillovers. The results are strongest for locality spillovers and E-Rate learning, but there is also support for learning from universities and from word-of-mouth. The existence of these spillovers could have affected the speed and patterns of adoption. The existence of spillovers for Internet adoption suggest the existence of similar spillovers for specific Internet applications, such as social networking, blogging or e-commerce applications, may exhibit similar spillovers. If so, these spillovers could play a role in the development of markets for these applications. In addition, these spillovers suggest the potential existence of spillovers for more advanced technologies such as wireless, broadband or mobile Internet access. If so, they could play a role in the diffusion of these technologies as recommended by the US National Broadband Plan (Federal Communications Commission, 2010).

In particular, the E-Rate findings suggest a level of effectiveness of these subsidies in promoting Internet usage among the programs' targeted audience. While others have found this program did generate Internet able classrooms, because there is little other evidence that the program affected test scores it has not been clear that it encouraged Internet use (Puma, 2000, Goolsbee and Guryan, 2006, and Ward, 2006). Perhaps the affect of Internet use is not in school performance but will manifest itself in a more technologically savvy population. If so, this population may be more likely to adopt new Internet based applications and succeed in ways not measured by test scores.

Of methodological interest, I find evidence that might help identify the effects of the Internet on behaviors. Studies that link Internet usage and a particular behavior suffer from interpretation problems. It is not clear whether Internet usage lead to the behavior or whether those that engage in a behavior that might benefit from Internet intermediation select into Internet users. The spillover measures I identify here are, by definition, exogenous to household level choices. These then could be used as instrumental variables that identify variation in Internet usage independent of selection issues. In this way, one might plausibly identify causality versus selection in the interpretation of correlations between Internet use and the behavior under investigation. For example, as children age, their attendance in school, especially one receiving E-Rate funding, will tend to increase their Internet use independent of the behavior. Likewise, as the child ages further and attends college (while living at home), Internet usage increases. Is variation in the individual behavior linked to variation in E-Rate exposure or to variation in college exposure?

References

Ackerberg , Daniel A. and Gautam Gowrisankaran (2006), “Quantifying equilibrium network externalities in the ACH banking industry” *RAND Journal of Economics* 37, 738-61.

Berndt, Ernst R., Pindyck, Robert S., and Azoulay, Pierre. (2003) “Consumption Externalities and Diffusion in Pharmaceutical Markets: Antiulcer Drugs,” *Journal of Industrial Economics,* Vol. 51, pp. 243-270.

Chevalier, J. and D. Mayzlin (2006). "The Effect of Word-of-Mouth on Sales: Online Book Reviews." *Journal of Marketing Research* 43:3, 345-54.

Economides, Nicholas. “Economics of Networks.” *International Journal of Industrial Organization* 14 (1996): 673–700.

Ellison, N. B., Steinfield, C., & Lampe, C. (2007). The benefits of Facebook "friends:" Social capital and college students' use of online social network sites. *Journal of Computer-Mediated Communication*, *12*(4), article 1. http://jcmc.indiana.edu/vol12/issue4/ellison.html

Federal Communications Commission, (2010). "Connecting America: The National Broadband Plan." <http://download.broadband.gov/plan/national-broadband-plan.pdf>.

Frenzen, J. and K. Nakamoto (1993). "Structure, Cooperation, and the Flow of Market Communication." *Journal of Consumer Research* 20, 360-75.

Gandal, Neil. (1994) “Hedonic Price Indexes for Spreadsheets and an Empirical Test for Network Externalities.” *Rand Journal of Economics* 25: 160–70.

Godes, D. and D. Mayzlin (2004). "Using Online Conversations to Study WOM Communication." *Marketing Science* 23(4) 545-60.

Goldfarb, Avi. (2006). "The (Teaching) Role of Universities in the Diffusion of the Internet". *International Journal of Industrial Organization* 24(2), 203-225.

Goolsbee, Austan and Jonathan Guryan (2006),"The Impact of Internet Subsidies in Public Schools," *Review of Economics and Statistics*, 88, 336-47.

Goolsbee, Austan and Peter J. Klenow (2002),"Evidence of Learning and Network Externalities in the Diffusion of Home Computers," *Journal of law and Economics*, 45, 317-43.

Gowrisankaran, Gautam, and Stavins, Joanna. (2004) "Network Externalities and Technology Adoption: Lessons from Electronic Payments." *RAND Journal of Economics* 35, 260-76.

Grossman, Gene M., and Helpman, Elhanan. *Innovation and Growth in the Global Economy.* Cambridge, Mass.: MIT Press, 1991.

Hoffman, Donna L., & Novak, Thomas P. 2000. The Growing Digital Divide: Implications for an Open Research Agenda. In Erik Brynjolfsson & Brian Kahin, eds. Understanding the Digital Economy: Data, Tools, Research. The MIT Press: Cambridge MA, pp. 245-260.

Kendall, Todd D., "Pornography, Rape, and the Internet," (May 2007) working paper <http://people.clemson.edu/~tkendal/internetcrime.pdf>.

Karshenas, Massoud, and Stoneman, Paul L. (1993) "Rank, Stock, Order, and Epidemic Effects in the Diffusion of New Process Technologies: An Empirical Model." *Rand Journal of Economics* 24: 503–28.

Puma, Michael E., Duncan Chaplin, Andreas D. Pape, *E-Rate Program and the Digital Divide: A Preliminary Analysis from the Integrated Studies of Educational Technology*, The Urban Institute, Washington, D.C. Sept., 2000. http://www.urban.org/UploadedPDF/410385_Standards.pdf.

Saloner, Garth, and Shepard, Andrea. (1995) “Adoption of Technologies with Network Effects: An Empirical Examination of the Adoption of Automated Teller Machines.” *Rand Journal of Economics* 26: 479–501.

Ward, Michael R., "The Effects of the E-Rate Internet Subsidies in Education" (March 2006). Available at SSRN: <papers.ssrn.com/sol3/papers.cfm?abstract_id=940092>.

Table 1

Counties with the Highest Percent College Enrollment

State	County	Major Institution	2004 Popl.	2004 Enroll.	Percent Enroll.
MI	Isabella	Mich State U.	61,812	44,836	72.5%
WA	Whitman	Wash State U.	40,164	23,241	57.9%
VA	Charlottesville City	U. of Virginia	41,950	23,341	55.6%
VA	Harrisonburg City	James Madison U.	37,974	16,108	42.4%
ID	Madison	BYU Idaho	27,450	11,555	42.1%
WY	Albany	U. of Wyoming	31,976	13,207	41.3%
IL	McDonough	Western IL U.	33,360	13,558	40.6%
SD	Brookings	S. Dakota State U.	27,631	10,884	39.4%
MS	Lafayette	U. of Mississippi	37,366	14,497	38.8%
MS	Oktibbeha	Miss. State U.	41,245	15,934	38.6%
ID	Latah	U. of Idaho	34,442	12,824	37.2%
IL	Jackson	Southern IL U. -Carbondale	59,873	21,589	36.1%
KS	Riley	Kansas State U.	64,615	23,151	35.8%
OK	Payne	OK State U.	66,897	23,819	35.6%
NC	Watauga	Appalachian St. U.	41,497	14,653	35.3%
VA	Montgomery	Virginia Tech	81,225	27,619	34.0%
GA	Clarke	U. of Georgia	98,961	33,405	33.8%
IA	Story	Iowa State U.	78,315	26,380	33.7%
OH	Athens	Ohio U.	61,588	20,143	32.7%
IN	Monroe	Indiana U.	118,199	37,821	32.0%
WV	Monongalia	West Virginia U.	81,533	25,255	31.0%
PA	Centre	Penn. State U.	135,028	41,289	30.6%
TX	Brazos	Texas A&M	146,353	44,435	30.4%
GA	Bulloch	GA Southern U.	54,250	16,100	29.7%
KS	Douglas	U. of Kansas	96,464	26,980	28.0%
IN	Tippecanoe	Purdue U.	145,981	40,108	27.5%
NY	Tompkins	Cornell U.	97,296	19,518	20.1%

Enrollment figures from IPEDS data.

Table 2

Summary Statistics of Key Variables

	Cross-Section		Changes	
Variable	Mean	Std. Dev	Mean	Std. Dev
HH Internet Use	0.640	0.480	0.073	0.364
County Internet Use	0.613	0.122	0.090	0.114
HH School-Aged Child	0.276	0.447	0.011	0.104
County E-Rate Funding Prev. Yr. ($1,000 per capita)	0.0023	0.0038	0.0019	0.0038
County E-Rate Funding Prev. Yr. ($1,000 per capita) × HH School-Aged Child	0.0007	0.0023	0.0004	0.0020
HH College Student at Home	0.294	0.456	0.019	0.137
HH College Student Away	0.037	0.188	0.011	0.105
County Fraction Enrolled in College	0.029	0.030	0.002	0.016
HH Word-of Mouth[1]	0.595	0.092	0.099	0.076

Cross-section is based on 132,712 observations and changes are based on 36,642 observations.
[1]Word-of-mouth statistics are based on 23,293 and 3,849 observations respectively.

Table 3

Cross-Sectional Regression of Internet Usage

	OLS	IV1	IV2
Internet Share in County	0.491***	0.269***	0.269***
	(0.011)	(0.015)	(0.015)
School-Aged Children	0.010	-0.190***	-0.181***
	(0.014)	(0.041)	(0.041)
E-Rate Funding in County	-1.189***	-29.547***	-28.494***
	(0.377)	(5.155)	(5.147)
E-Rate Funding x School-Aged Children	1.334**	90.065***	86.156***
	(0.667)	(16.879)	(16.879)
College Student at Home	0.170***	0.168***	0.168***
	(0.004)	(0.005)	(0.005)
College Student Away	0.091***	0.090***	0.090***
	(0.005)	(0.006)	(0.006)
College Enrollment Share in County	0.143***	0.213***	0.311***
	(0.038)	(0.042)	(0.066)
Observations	132,712	132,712	132,712
R-squared	0.32	0.21	0.22

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10
Household level demographics are included in all specifications but most coefficients are not reported.
IV1 - Instrumented variables include Internet Share in County and E-Rate variables.
IV2 - Instrumented variables include Internet Share in County, E-Rate variables and College Enrollment Share in County.

Table 4

Cross-Sectional Regression of Internet Usage with Calling Information

	OLS	IV1	IV2
Internet Share in County	0.531***	0.298***	0.262***
	(0.027)	(0.034)	(0.046)
School-Aged Children	0.024	0.065	0.090
	(0.019)	(0.058)	(0.058)
E-Rate Funding in County	-0.896	-10.602*	-8.225
	(0.938)	(5.663)	(5.712)
E-Rate Funding x School-Aged Children	1.949	31.205	19.839
	(1.707)	(20.430)	(20.812)
College Student at Home	0.196***	0.200***	0.199***
	(0.011)	(0.011)	(0.011)
College Student Away	0.109***	0.110***	0.109***
	(0.013)	(0.013)	(0.013)
College Enrollment Share in County	0.210**	0.295***	0.595***
	(0.092)	(0.095)	(0.147)
Internet Share in Called States	0.106**	0.222***	0.398**
	(0.052)	(0.054)	(0.180)
Observations	23,293	23,293	23,293
R-squared	0.33	0.32	0.33

Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.10$
Household level demographics are included in all specifications but most coefficients are not reported.
IV1 - Instrumented include Internet Share in County and E-Rate variables.
IV2 - Instrumented include Internet Share in County, E-Rate variables, College Enrollment Share in County and Internet Share in Called States.

Table 5

Re-Sampled Observations Cross-Sectional Regressions of Internet Usage

	OLS	IV1	IV2
ΔInternet Share in County	0.356***	0.178***	0.167***
	(0.018)	(0.027)	(0.028)
ΔSchool-Aged Children	0.166**	0.139*	0.143*
	(0.074)	(0.076)	(0.076)
ΔE-Rate Funding in County	0.030	-0.330	-0.096
	(0.587)	(1.652)	(1.659)
ΔE-Rate Funding x School-Aged Children	4.737***	14.403***	14.808***
	(1.321)	(3.176)	(3.189)
ΔCollege Student at Home	0.047**	0.041**	0.041**
	(0.019)	(0.019)	(0.019)
ΔCollege Student Away	-0.017	-0.024	-0.025
	(0.017)	(0.017)	(0.017)
ΔCollege Enrollment Share in County	0.370***	0.327***	-0.303
	(0.117)	(0.116)	(0.377)
Observations	36,642	36,642	36,642
R-squared	0.02	0.01	0.01

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Changes in household level demographics are included in all specifications but most coefficients are not reported.
IV1 - Instrumented include Internet Share in County and E-Rate variables.
IV2 - Instrumented include Internet Share in County, E-Rate variables and College Enrollment Share in County.

Table 6

Re-Sampled Observations Cross-Sectional Regressions of Internet Usage

with Calling Information

	OLS	IV1	IV2
ΔInternet Share in County	0.384***	0.206**	0.128
	(0.058)	(0.089)	(0.106)
ΔSchool-Aged Children	0.636***	0.637***	0.615***
	(0.225)	(0.222)	(0.217)
ΔE-Rate Funding in County	-0.717	0.217	-0.896
	(1.827)	(4.704)	(4.847)
ΔE-Rate Funding x School-Aged Children	7.815*	13.124	13.458
	(4.264)	(9.828)	(9.911)
ΔCollege Student at Home	0.137*	0.135*	0.136*
	(0.080)	(0.080)	(0.080)
ΔCollege Student Away	-0.094	-0.104*	-0.104*
	(0.059)	(0.060)	(0.060)
ΔCollege Enrollment Share in County	0.301	0.265	0.110
	(0.353)	(0.349)	(1.104)
ΔInternet Share in Called States	0.236***	0.316***	0.651***
	(0.088)	(0.109)	(0.244)
Observations	3,849	3,849	3,849
R-squared	0.03	0.03	0.02

Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.10.
Changes in household level demographics are included in all specifications but most coefficients are not reported.
IV2 - Instrumented include Internet Share in County and E-Rate variables.
IV1 - Instrumented include Internet Share in County, E-Rate variables, College Enrollment Share in County and Internet Share in Called States