Journal of Mathematics and Statistics 2 (1): 328-333, 2006
ISSN 1549-3644

Adults' Demand for the Internet use in the USA: An Empirical Approach

[1,3]Ismail H. Genc, [2]Hasan Sahin and [3]Robert W. Stone
[1]School of Business and Management, American University of Sharjah, P.O. Box 26666
Sharjah, United Arab Emirates
[2]Department of Economics, The Faculty of Political Sciences, Ankara University, Ankara, Turkey
[3]Department of Economics, Finance and Information Systems, College of Business and Economics
University of Idaho, Moscow, Idaho 83844-3178, USA

Abstract: Academic literature regarding the factors influencing Internet use is presented. Based on this literature, a theoretical model of the demand for the use of the Internet is developed. The estimation of the demand for Internet use in the U.S. is performed using logit and probit. The model includes explanatory variables such as gender, race, income, age, educational level, marital status, parenthood, employment status and student status. Features related to the Internet such as familiarity with and the need to use the Internet and the degree of Internet availability at home are also included. The empirical study uses survey data containing a relatively large number of respondents. The results indicate no gender or a racial digital divide in the use of the Internet. Employment is found to negatively correspond to Internet use. Marital status has no significant impact on Internet use. The study concludes with an interpretation of the empirical results as well as directions for future research.

Key words: Econometrics standpoint, economic theory, econometric techniques, internet use

INTRODUCTION

The hallmark of the fast moving change in the United States economy has been in part due to the incorporation of information technology and in particular the Internet into business operations. One result of these changes include global increases in sales by businesses to consumers as well as to other businesses (e-commerce). Furthermore, using the corporate WWW-site has become a natural method of conducting business and an aspect of daily life. As a result, significant debate as to the ability of e-commerce to transform conventional business operations has developed. Within this debate, some individuals have argued that e-commerce has create a new digital economy. However, the economics of the new economy are not quite incorporated into the mainstream of professional economic discussions. One notable exception is a symposium sponsored by the Federal Reserve Bank of Kansas City in 2001: Economic Policy for the Information Economy.

The new economy impacts, directly or indirectly, people from all walks of life mainly through providing alternatives to the brick and mortar business model. E-commerce is perhaps the most touted benefit of the Internet and has implications for consumers as well as producers/sellers. The conduct of business using the Internet also has potential for additional taxation opportunities and is producing a related policy debate. This policy debate has two basic forms in the United States. At the federal level, the interest is in the practices of taxing intangibles such as digital content, particularly in an international setting. At the state level the interest is in the impact on sales taxation. Additional issues include taxation on e-commerce crossing boundaries of tax jurisdictions applying value added taxes.

There are also privacy issues arising between business owners, customers and the government as to what data are private and what are not private. Every transaction over the Internet provides the possibility of extensive personal information to be revealed either voluntarily or involuntarily. This provides the possibility for privacy abuse by unscrupulous transaction participants. These issues relate to a larger concern of the trust the consumer has in an electronic environment. Key factors in developing trust in the e-commerce environment and hence encourage consumer use, include security risks and privacy issues.

As the preceding discussion indicates, there are many factors influencing Internet use and e-commerce or what has been referred to as the new digital economy. A comprehensive study of this new economy within the boundaries of a single paper is impossible. The study presented below focuses on the determinants (social factors) of Internet use in the United States using well known econometric tools. The results have policy implications through identifying the characteristics of Internet use and non-use in order to address the current and future gap between the users and non-users of the Internet (the digital divide).

Corresponding Author: Ismail H. Genc, School of Business and Management, American University of Sharjah, P.O. Box 26666, Sharjah, United Arab Emirates

The theoretical model: There are a variety of factors influencing Internet usage. From the perspective of e-commerce, there have been studies examining the linkages of behavioral, psychological and demographic variables to Internet use. The characteristics of the system have also been studied. Examples include usability[1], the role of graphical enhancements[2] and the information infrastructure.

The characteristics of individuals have been shown to impact Internet use. The frequency and focus of Internet use is influenced by the user's age[3,4] and gender. The educational level of the individual also influences their use of the Internet[3,4]. Other variables influencing Internet use include race and/or culture[3-5] and employment status.

Related to Internet use is how the Internet is used by the individual. These factors include where access occurs (at work or at home), the motivation for using the Internet (e.g., use as a student and which software applications are used) and the amount of time it is used[6,7]. Additionally, marital status appears to influence Internet use[3,6,7].

The literature discussed above implies theoretical relationships among Internet usage and the discussed variables. Grouping these variables into broad classifications, these classifications are the user's age, educational level, gender, income level, race, marital status, Internet access and the motivation to use the Internet (number of children in the household, access at work and use as a student). These theoretical relationships are expressed in equation 1. It is this theoretical model that is empirically tested.

INTERNET USE=i(AGE, EDUCATIONAL LEVEL, GENDER, INCOME LEVEL, INTERNET ACCESS, MARITAL STATUS, RACE, NUMBER OF CHILDREN AT HOME, USE AT WORK, USE AS A STUDENT) (1)

The data: The data set used in the study is from the Pew Internet & American Life Project collected by the University of Pennsylvania Pew Internet Research Center. The specific data set is located at http://www.pewinternet.org/datasets/ under the "4/30/01-- March 2000 Survey Data." The description of the data reads as "extensive tracking of Internet use basics, email use, family connectedness and social capital." Table 1 provides descriptive statistics of the variables in the data set along with explanations regarding their coding for use in this study. The usable number of observations is quite large for all variables, ranging from 1685 to 2238. The unequal number of observations across variables is due to the lack of response from some survey participants to certain questionnaire items. The questionnaire items forming the variables for the study are shown in Table 1. Several of these items were reconfigured to create variables needed for the study. These variables are also shown in The typical survey respondent is the head of a household whose average age is nearly 40 years old. This age is consistent with findings reported by Kehoe *et al.*[8]. The typical survey respondent had no less than a high school education. The Kehoe et al survey indicated a slightly higher educational level among its Internet users' with 87.7% of the respondents having some college experience. While the two respondent groups are not identical, they are similar. One potential explanation for the difference in educational attainment between the two surveys could be due to the fact that Kehoe et al's survey focused exclusively on Internet users while the survey used here included non-users of the Internet.

Almost half of the respondents are women which helps clarify contradictions found in earlier studies. For example, in a 1994 survey Pitkow and Kehoe[9] found that only 5%-10% of the Internet users were female. Kehoe et al's survey puts the percentage of Internet using women to be almost 50%. Hargittai[10] does not find a statistical difference between male and female adult users of the Internet where the user ratio in both genders is approximately 50%. Ruud[11] reports that almost half of the labor force is women. Likewise, slightly half of those who took the survey are in some sort of a marital relationship, which is consistent with the value reported by Kehoe et al. The percentage of respondents with children had a similar distribution. The average respondent's income level was between 40 to 50 thousand dollars annually. Kehoe et al identified the average income level in the upper 50 thousands range. While about 80% of the respondents reported being employed, 20% of the respondents were students. The majority of the respondents have more than one phone line available at home for Internet and other users. Whites comprise the largest portion of the survey respondents. This finding is a replication of Kehoe et al's Internet users. Furthermore, respondents appear to use the Internet for activities such as emailing and shopping.

The empirical model: The empirical study investigates the factors determining individuals' Internet use. The factors examined include gender, age, race, marital status, educational level, income level, the number of telephone lines and children, use of the Internet at work and at home, whether the individual is a student or is working. The price of Internet access is excluded from the estimations as a regular theoretical demand model would suggest since the most common pricing scheme of these services is a flat rate charge[7,12-14].

The relationships among the independent variables were examined. The correlations among these variables are shown in Table 2.

Table 1: Descriptive Statistics of Individual Variables

Variable	Question	Mean	Values of Variable	Observations
AGE	AGE: D3	38.95	The age of the respondent as reported in the survey	2194
EDU	EDUC: D4	4.74	=1 if no education or grades 1-8, =2 if grades 9-11, =3 if grade 12, =4 if above high school with no college, =5 if some college, =6 if college graduate, =7 if graduate education	2225
EXPERIENCE	Q12	2.74	=1 if started going on line within the last six months, =2 if started going on line a year ago, =3 if started going on line 2 or 3 years ago, =4 if started going on line more than 3 years ago	1685
INTUSE	Q6	0.76	=1 if uses internet, =0 if not	
GENDER	D1: SEX	0.49	=0 if female, =1 if male	2238
INCOME	D11	4.99	=1 if less than $10K, =2 if between $10K & $20K, =3 if between $20K & $30K, =4 if between $30K & $40K, =5 if between $40K & $50K, =6 if between $50K & $75K, =7 if between $75K & 100K, =8 if more than $100K	1842
LINES	D9	1.54	=The number of phone lines available at home	2222
MARRIED	D7	0.58	=1 if Married or Living as married, =0 otherwise	2223
PARENT	D2	0.41	=1 if parent , =0 if not	2234
RACE	D6	0.82	=1 if white, =0 otherwise	2185
STU	D8A	0.20	=1 if part time or full time student, =0 otherwise	2217
WHYINT	Q17_1	2.33	=sum of the 'yes=1' answers to this question	2238
WORK	D8	0.80	=1 if employed part time or full time, =0 otherwise	2225

The original question survey is mentioned under "Question." The "Values of Variable" is the coding used in this study which does not necessarily correspond to the original survey's coding system. However, there could be a more direct correspondence between this study's coding and the original survey's coding if a variables is marked with an * in the "Values of Variable" column. The number of observations is given under the "Observations" column.

Table 2: The upper diagonal correlation matrix of the independent variables

	AGE	EDU	GENDER	INCOME	LINES	MARRIED	PARENT	RACE	STU	WORK
AGE	1.00	0.16	-0.01	0.19	0.03	0.29	-0.07	0.10	-0.36	-0.24
EDU		1.00	0.02	0.29	0.03	0.11	-0.02	0.04	-0.02	0.05
GENDER			1.00	0.14	0.00	0.03	-0.08	0.02	0.03	0.08
INCOME				1.00	0.20	0.36	0.06	0.07	-0.14	0.06
LINES					1.00	0.05	-0.01	-0.06	0.01	0.00
MARRIED						1.00	0.30	0.10	-0.22	-0.05
PARENT							1.00	-0.06	-0.16	0.07
RACE								1.00	-0.09	-0.03
STU									1.00	-0.05
WORK										1.00

The correlations between pairs of variables of common samples.

These correlations are quite low satisfying a requirement for the estimation techniques employed in the study. These correlations also indicate the direction of the linear relationships among the explanatory variables.

The dependent variable, INTUSE, is the answer to the questionnaire item "Do you use the Internet?" Since the potential answers are yes or no, the required empirical tools to use are binary choice models. Specifically, logit and probit estimation techniques for modeling the relation at hand were used. The model to be estimated is shown in equation 2.

INTUSE=i(AGE, EDU, GENDER, INCOME, LINES, MARRIED, RACE, PARENT, STU, WORK) (2)

The exact form of the function i is determined by the estimation method. One method used implies the logistic function (logit), the second method implies the standard normal distribution (probit).

RESULTS

The model presented in equation (2) was estimated using three empirical techniques, logit, probit and ordinary least squares (OLS). These results are presented in Table 3. It needs to be noted first that all the slopes from both the logit and probit model estimations as well as the coefficients from the OLS estimation are consistent in terms of magnitude. It is within this context that the remaining empirical results are discussed below.

There was no gender difference regarding accessing online services identified in the results. Similarly, the race variable coefficient was not a statistically significant at a 5% level of significance. Furthermore, redefining the race variable into

Table 3: Estimation results

	LOGIT		PROBIT		LINEAR
Variables	Coefficient	Slope	Coefficient	Slope	Coefficient
Constant	0.21327 (0.57)		0.169965 (0.79)		0.579925 (10.06)
Age	-0.0415 (-7.72)	-0.00608	-0.02441 (-7.95)	-0.00653	-0.00682 (-8.19)
Edu	0.366949 (8.35)	0.053726	0.214926 (8.50)	0.057476	0.06226 (9.12)
Gender	0.179279 (1.46)	0.026249	0.106645 (1.50)	0.028519	0.030927 (1.61)
Income	0.247098 (6.59)	0.036178	0.142824 (6.68)	0.038194	0.037275 (6.67)
Lines	0.303441 (3.50)	0.044428	0.155627 (3.45)	0.041618	0.037395 (3.34)
Married	-0.08235 (-0.58)	-0.01206	-0.03586 (-0.44)	-0.00959	-0.00484 (-0.22)
Race	0.20865 (1.88)	-0.01988	0.166190 (1.84)	-0.02066	0.050543 (0.04)
Parent	-0.12583 (-0.96)	-0.01842	-0.07732 (-1.01)	-0.02068	-0.01714 (-0.83)
Student	0.142434 (0.80)	0.020854	0.081103 (0.80)	0.021689	0.018898 (0.71)
Work	-0.60561 (-3.37)	-0.08867	-0.35131 (-3.42)	-0.09395	-0.086 (-3.21)
F(bx)	0.146413		0.267421		1
Log likelihood	-851.7966		-969.5373		
Restr. Log likelihood	-969.5373		236.1327		
LRI	12.14%		12.18%		

The dependent variable is Intuse. Numbers in parentheses are the z-statistics. f(bx) refers to the numerical value of the distribution function evaluated at the mean values of the explanatory variables. The number of included observations is 1782.

Table 4: The prediction success table of the logit model

		Actual	
		0	1
Predicted	0	78	50
	1	339	1315

The number of right predictions is 0.139E+04, which corresponds to a 0.78171 percentage of right predictions. The naïve model's percentage of right predictions is 0.76599.

Table 5: The prediction success table of the probit model

		Actual	
		0	1
Predicted	0	71	48
	1	346	1317

The number of right predictions is 0.139E+04, which corresponds to a 0.7789 percentage of right predictions. The naïve model's percentage of right predictions is 0.76599.

subcategories of Hispanics and blacks did not change the original finding of statistical non-significance for this variable. Hence, the results did not identify a digital divide among people from different racial backgrounds regarding accessing Internet services.

The results indicated, however, a digital divide across different income categories. All three estimation techniques found a statistically significant coefficient on income with about a 4% marginal, positive impact on the decision to access the Internet. Hargittai[10] similarly reports that the higher level of family income leads to higher levels of Internet usage among adults in the United States.

The empirical results also indicated that parenthood is not a significant factor in Internet access. This result may sound contrary to what one might think at first, but it actually makes sense given that the respondents to the questionnaire items are more likely to be the parent in the household rather than the child. In this setting, it is generally more probable that the children in the household have a priority of using the Internet rather than their parents. Yet, the survey respondents were more likely the household parents. Hence, in all models, the parent coefficient was not statistically significant. The empirical results also showed that the age variable was negatively related with Internet use. This result is probably explained by the same observation as the parenthood result. This explanation being that the household children generally have more interest in using the Internet than the older parents in the household. Yet, the parents completed the survey. Thus, the age coefficient was significantly negative in all models, although its marginal effect is less than 1%.

The empirical results also showed that more educated respondents were more likely to use the Internet. The result was indicated by the statistically significant and positive coefficient on the education variable in all models. None of the estimated models produced a statistically significant coefficient on the marital status variable. The result suggests that marital status was not a determining factor as to household Internet use. In other words, married and single people are equally likely to use the Internet.

An important empirical result was the negative coefficient on the variable measuring Internet use at work (WORK). All three estimated functional forms indicated approximately a 9% negative marginal effect on the willingness of individuals to use the Internet while working. A potential explanation is that a person not working (unemployed or retired) has more free time to "surf" the Internet.

One might claim that the day the respondent took the survey matters because a working person is less likely to use the Internet if the survey is taken over the weekend. Several counter arguments can be made, however. First, the large number of respondents surveyed over a relatively long time frame makes the time bias insignificant. That is to say, it is unlikely that all the employed people were surveyed over the weekend verses the work week. Second, possibly strict rules at work may discourage workers from attempting to use the Internet at work. Furthermore, unemployed people may make use of free Internet access to the Internet via public libraries. The zero marginal price of such Internet access should also make this argument questionable once someone has paid the flat fee. It is worth mentioning that Kehoe et al found that Internet usage from work decreased over time as observed by

surveys conducted at different time periods. This may indicate a tightening of Internet access at work over time as companies encourage workers' attention only to their jobs.

The results also indicated no differences in student status encouraging use of the Internet. In other words, personal status as a student does not make him or her more likely to use the Internet. Finally, all the models supported the proposition that the number of available phone lines at home is a significant determinant in one's decision to use the Internet. The marginal impact of the number of phone lines is about 4% according to all models.

Statistical checks on the estimated models: In order to evaluate the robustness and validity of the estimation techniques, statistical tests were performed. The objective was to avoid any aberrations in the research efforts. Now we present the statistical findings regarding the empirical analysis to assess their sensitivity and evaluate plausible alternative hypotheses. It should also be noted that the linear model, OLS, is not appropriate for this type of analysis. It was used in this empirical analysis only for comparative purposes. However, the logit and probit models warrant further analyses of their respective estimations. As far as the logit estimation was concerned, the value of the unrestricted log likelihood (the logarithmic value of the likelihood function evaluated with no restrictions imposed on the coefficients) was -851.80. On the other hand, the restricted log likelihood function had a value of -969.54. The restriction was that the coefficients on all variables except the constant were restricted to equal zero. Thus, the likelihood ratio, which is defined to be "the negative of the twice the difference between the restricted and unrestricted likelihood values," turns out to be 235.48. With 10 degrees of freedom, the significance level associated with this statistic is 0.00, concluding that a null hypothesis of insignificance of all slope coefficients is clearly rejected. Similarly, the unrestricted log likelihood value from the probit model was -851.47 and for the restricted model it was –969.54. the resulting likelihood ratio was 236.13 which rejects the hypothesis of the insignificance of all the slope coefficients in the model.

In summary, these tests indicated that there was no statistical evidence to claim that the slope coefficients in either the logit or probit models were altogether irrelevant. An additional measure of goodness of fit was also calculated. This statistic is a similar to a coefficient of determination (R-square). This statistic is defined in equation 3.

$$LRI=1-UL/RL \quad (3)$$

where UL is the unrestricted log likelihood function. RL is the maximized value of the restricted function with only a constant term involved. For the logit model, this statistic was 12.14% and for the probit model it was 12.18%. These results confirm the results from the log likelihood tests.

Another standard evaluation technique found in the literature is to compare the performance of the logit and probit models to a naïve model. The naïve model uses each data point and predicts a one for the dependent variable if it is more likely to occur than finding a zero for the dependent variable otherwise the naïve model predicts a zero for the dependent variable. As shown in Tables 4 and 5, such a specification produced 76.60% correct predictions of ones and zeros for this data set. The logit model estimation correctly predicted 1315 of 1365 of the ones and 78 of 417 of the zeros. That amounted to 78.17% correct predictions for the logit model, which was approximately 1.6% larger than the naïve model's predictive ability. Similarly, the probit model outperformed the naïve model with 77.90% correct predictions. The probit model correctly predicted 1317 of 1365 of the ones and 71 of 417 of the zeros. These results indicated an improvement in the estimation process by the use of the logit and probit models compared to the naïve model.

CONCLUSION

An adult's Internet use in the United States was studied using a data set with a large number of participants. The analysis used econometric techniques well-known in the literature. Several interesting conclusions were drawn from the results. Neither a gender nor a racial digital divide was found to impact Internet use by adults in the United States. This might be a point of policy interest in allocating resources in efforts to make the Internet accessible to all individuals. On the other hand, income level was found to determine Internet use. The result could focus attention in designing government policies regarding access and use of the Internet by considering the income distribution of potential Internet users. Furthermore, this result could encourage finding cost efficient ways to bring Internet access and training to financially stressed areas. However, comprehensive analyses of these issues are beyond the scope of this study.

It was also found that students were equally likely to use the Internet as other potential users. This may well constitute supporting evidence for the commonality of Internet use. In the past, Internet access and hence potential use was more widely available in educational institutions making students more likely candidates to engage in Internet activities. Furthermore, educated people have the tendency to use the Internet more than individuals with lower educational attainment levels. The negative coefficient on the work variable indicated that, in general, employees do not use the Internet at work. The empirical results also suggested that parenthood, marital status and age have either no or minor impacts on Internet use.

The research presented above is extendable in several ways. A few of the extensions are data related. A more divergent populace, especially regarding age and race, could provide additional insights on Internet use with regard to these variables. Furthermore, as new data become available it might be possible to observe if the findings are sensitive to response times of the respondents. A time series or a panel data set would better track changes in attitudes toward Internet use as individuals pass through various stages of their lives. Specifically, data from several surveys could be combined or pooled to conduct a panel data analysis for this purpose. The intensity of demand for Internet services in terms of time spent using the Internet makes use of a very different strand of econometric techniques, duration analysis. An alternative modeling issue is the frequency of Internet use.

In conclusion, the context of these caveats, the study illustrated interesting peculiarities of the Internet economy in the United States from an econometrics standpoint. The results might be of interest to policy makers in both government and the business world. The employed analysis brings a new dimension to the simple statistical methods based studies largely found in the literature. These estimations suffer from biases injected in the analysis due to omitted variables. This analysis avoided such troubles by employing a multivariate analysis with reasonable functional specifications based on economic theory.

ACKNOWLEDGEMENTS

We would like to thank Todd Chavez for the initial guidance in locating possible data sources. Genc gratefully acknowledges the support by a Summer Grant from CBE, UI in 2001 & a Small Travel Grant of UI Research Office in Summer 2002.

REFERENCES

1. Sears, A., 2000. Introduction: Empirical Studies of WWW Usability. Intl. J. Human-Computer Interaction, 12: 167-171.
2. Sears, A.J. A. Jacko and E.M. Dubach, 2000. International aspects of world wide web usability and the role of high-end graphical enhancements. Intl. J. Human-Computer Interaction, 12: 241-261.
3. Wasserman, I.M. and M. Richmond-Abott, 2005. Gender and the internet: Causes of variation in access, level and scope of use. Social Sci. Quart., 86: 252-271.
4. McGerty, L.J., 2000. Nobody lives only in Cyberspace: Gendered subjectivities and domestic use of the internet. CyberPsychology & Behavior, 3: 895-899.
5. Bellman, S., E.J. Johnson, S.J., Kobrin and G.L. Lohse, 2004. International differences in information privacy concerns: A global survey of consumers. Information Society, 20: 313-324.
6. Hendrix, E., 2005. Permanent injustice: Rawls theory of justice and the digital divide. Educational Technology & Society, 8: 63-68.
7. Anderson, B., C. Gale, M.L.R. Jones and A. McWilliam, 2002. Domesticating broadband-what consumers really do with flat-rate, always-on and fast internet access. BT Technol. J., 20: 103-114.
8. Kehoe, C., J. Pitkow, K. Sutton, G. Aggarwal and J.D. Rogers, 1999. Results of GVU's Tenth World Wide Web User Survey, Graphics Visualization and Usability College of Computing Georgia Institute of Technology Atlanta, GA, USA. Available at http://www.gvu.gatech.edu/user_surveys/survey-1998-10/tenthreport.html.
9. Pitkow, J. and C. Kehoe, 1996. Emerging trends in the world wide web user population. Communications of the ACM, 39: 106-108.
10. Hargittai, E., (Forthcoming). The Digital Divide and What to do About it. In New Economy Handbook. Edited by Derek C. Jones. San Diego, CA: Academic Press.
11. Ruud, P.A., 2000. An Introduction to Classical Econometric Theory. Oxford University Press, Oxford.
12. Varaiya, P., 1999. Demand and provisioning of quality-differentiated internet access. Talk delivered at the University of Maryland (Nov., 12). Available at http://www.ee.umd.edu/Lecture/Fall99/Pravin_Varaiya.html.
13. Scalise, K., 1999. Internet congestion caused by flat rate pricing and waste, UC Berkley Researchers Find. News Release. (May 20), Available at http://www.berkeley.edu/news/media/releases/99legacy/5-20-1999.html
14. Richardson, C., 1996. Australia's Peak Demand for Internet Bandwidth: A Modeling and Forecasting Methodology. Research Report No: 3, La Trobe University Online Media Program.